Title	材料発見における意思決定支援のための不確実性を明らかにするための証拠に基づくフレームワーク
Author(s)	HA, MINH QUAN
Citation	
Issue Date	2023-09
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/18770
Rights	
Description	Supervisor:DAM, Hieu Chi, 先端科学技術研究科, 博士

Doctoral Dissertation

An evidence-based framework to reveal uncertainties for decision support in materials discovery

Minh-Quyet HA

Supervisor: Hieu-Chi DAM

Graduate School of Advanced Science and Technology

Japan Advanced Institute of Science and Technology

Knowledge science

September, 2023

JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY

Abstract

Co-Creative Intelligence research area
Graduate School of Advanced Science and Technology

Knowledge science

An evidence-based framework to reveal uncertainties for decision support in materials discovery

by Minh-Quyet HA

Searching for new materials is challenging and multifaceted, with countless potential candidates to explore. While data-driven approaches have shown promise in narrowing down the search space, their effectiveness is limited by certain factors. This doctoral thesis addresses these limitations and proposes a data-driven framework for quantitatively measuring the similarities between materials while accounting for uncertainty. The framework focuses on two scenarios: binary properties and continuous properties. For binary properties, the material similarity is assessed by evaluating the differences in compositions and their impact on the property of interest. For continuous properties, the similarity is determined by examining the correlation between occurrences of specific physical properties in the materials. The goal is to identify materials that exhibit similar behavior and characteristics, despite variations in composition. This framework provides a basis for informed decision-making in materials discovery and offers insights into the underlying mechanisms governing material behavior. This research contributes to more reliable and robust material discovery processes by incorporating uncertainty in similarity measurements. The proposed framework offers a valuable tool for materials science researchers, aiding in identifying and understanding materials with desirable properties and paving the way for advancements in materials discovery and development.

Keywords: Materials discovery, Decision-making, Data-driven approach, Similarity measurement, Dempster-Shafer theory

Declaration of Authorship

I, Minh-Quyet HA, declare that this thesis titled, "An evidence-based framework to reveal uncertainties for decision support in materials discovery" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

“Whenever someone creates something with all of their heart, then that creation is given a soul.”

Hayao Miyazaki

Acknowledgements

I would like to express my heartfelt appreciation to Professor Dam Hieu Chi for giving me the opportunity to engage in scientific research and experience the satisfaction that comes with it. Collaborating with him has taught me important lessons on professionalism, clarity of thought, devotion to science, and especially enjoying the journey.

I would also like to extend my gratitude to my lab colleagues, specifically Dr. N.V. Cuong and Dr. N.D. Nguyen, for their unwavering support and insightful discussions that have aided me in completing my work.

Lastly, I would like to dedicate this thesis to my family, whose constant encouragement and guidance have been indispensable in my journey. I will always cherish their assistance.

Contents

List of Figures

List of Tables

List of Abbreviations

HEA	**H**igh **E**ntropy **A**lloys
DST	**D**empster **S**hafer **T**heroy
ERS	**E**vidence-based **R**ecommender **S**ystem
eRSM	**E**vidential **R**egresion-based **S**imilarity **M**easure

List of Symbols

$\mathcal{D}$	Dataset	
$m()$	Mass function	
Ω	Frame of discernment	
$\mathbf{x}$	Representation vector of a data instance	
y	Target variable of a data instance	
$\mathbf{x}_i$	Representation vector of a data instance indexed i	
y_i	Target variable of a data instance instance indexed i	
ω	angular frequency	rad

For/Dedicated to/To my…

Chapter 1

Introduction

1.1 Materials discovery

Throughout the history of humankind, extending from the Stone Age to the contemporary era, the evolution of humanity has been inextricably intertwined with our utilization of materials and tools (Hong et al., 2021). It is indisputable that the discovery of new materials has served as a cornerstone in shaping human civilization when the most advanced material of a given era frequently emerges as a defining factor of historical periods. This is evident in naming historical periods, such as the Stone Age, Bronze Age, Iron Age, or Silicon Age, in which primary prehistoric phases of our civilization are named after the engineering materials of the era rather than our linguistic abilities, social interactions, or economic achievements. This emphasis on materials underscores the progressively accelerated pace at which they have transformed society. To illustrate, while the Stone Age encompasses millions of years (from 2.6 million years ago to 3,300 BC), subsequent ages are delineated in tens of millennia, with the duration of each era successively diminishing. This pattern persisted until the onset of the Industrial Revolution, a period characterized by the rapid introduction of innovative processes and materials. As a result, the engineering age during this time can be measured in mere decades. Ultimately, this escalating pace highlights the crucial role of material discoveries in propelling the progress and development of human civilization.

The journey from discovery to practical use is crucial in advancing technology and improving society. The duration it takes for an innovative material to move from its initial discovery in the lab to its application in the industry can significantly vary, depending on the material and its intended use. Nonetheless, it is widely recognized that this process can be time-consuming and may take anywhere from 10 to 20 years or longer (White, 2012). The timeline is influenced by various factors such as the complexity of the material, the establishment of scalable manufacturing techniques, comprehensive assessment of safety and efficacy, regulatory endorsements, and market acceptance. Consequently, materials crucial to tackling some of society's most pressing challenges might have already been discovered, yet they are still awaiting implementation in manufactured products. For instance, graphene, a single layer of carbon atoms arranged in a hexagonal lattice structure, forms a two-dimensional (2D) material with outstanding mechanical, electrical, and thermal properties. The material was discovered in 2007 by Geim and Novoselov, who were awarded the Nobel Prize for their groundbreaking work. However, despite its exceptional properties and potential applications, graphene's deployment is hampered by high production costs.

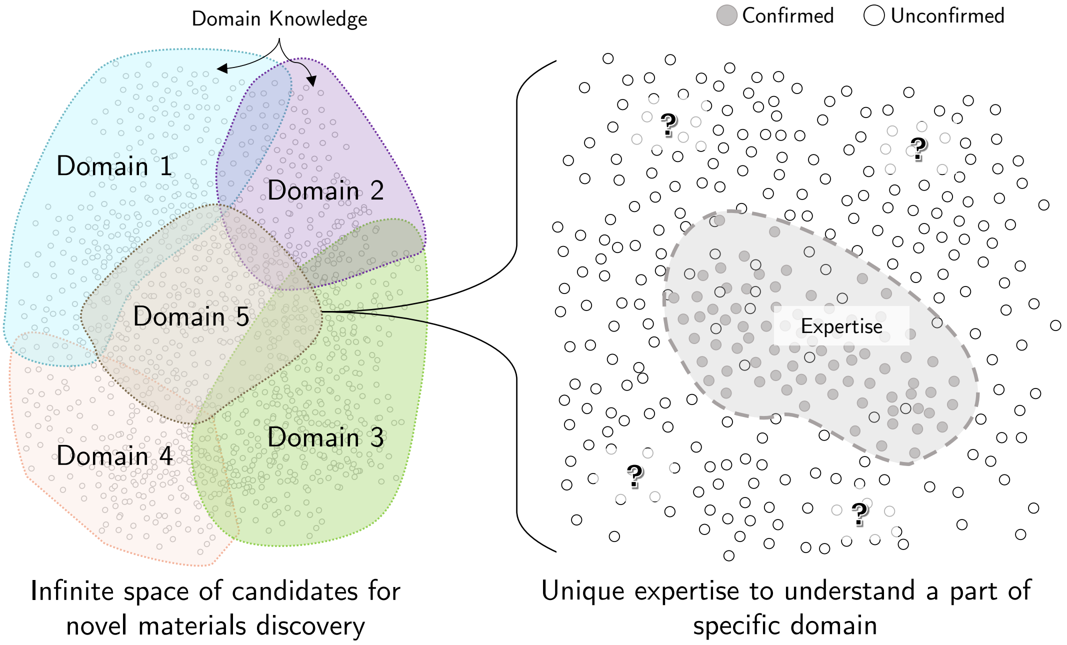


FIGURE 1.1: Challenge of discovering new materials from an endless space of unconfirmed candidates.

1.1.1 Materials informatics: a new paradigm for discovering new materials

Pursuing materials discovery poses a formidable challenge as researchers navigate through an extensive array of candidates have not confirmed yet (Fig. 1.1). The most crucial task in this ambitious journey is identifying and selecting materials that merit further exploration and expensive validations. However, materials scientists face a significant limitation: their expertise typically extends to specific domains, leaving a vast pool of potential candidates beyond their immediate knowledge. For instance, a materials scientist specializing in ceramics may encounter difficulties when attempting to discover novel ceramic materials with enhanced mechanical properties. While their expertise in ceramics is extensive, their knowledge of polymers, which possess unique mechanical characteristics, might be limited. However, correlations or relationships between specific ceramic compositions and polymer structures could provide valuable insights into the design of new materials. By embracing interdisciplinary collaboration and seeking insights from other domains, such as polymers, materials scientists can broaden their understanding of materials and enhance the efficiency of materials discovery. By leveraging collective expertise and exploring uncharted territory, researchers can identify promising materials with superior properties, ultimately propelling scientific advancement to new frontiers.

Throughout human history, *to discover advanced materials*, the progression of materials science has been closely intertwined with scientific and technological advancements. As identified by Agrawal (Agrawal and Choudhary, 2016), four distinct scientific paradigms have emerged within materials science: empirical science, model-based theoretical science, computational science, and data-driven science. Initially, the field predominantly relied on empirical approaches to acquire knowledge about material extraction, purification, and processing. However, as time went on, with the vigorous development of mathematics, including algebra and differential calculus, theoretical physics built a solid foundation for deductive studying and

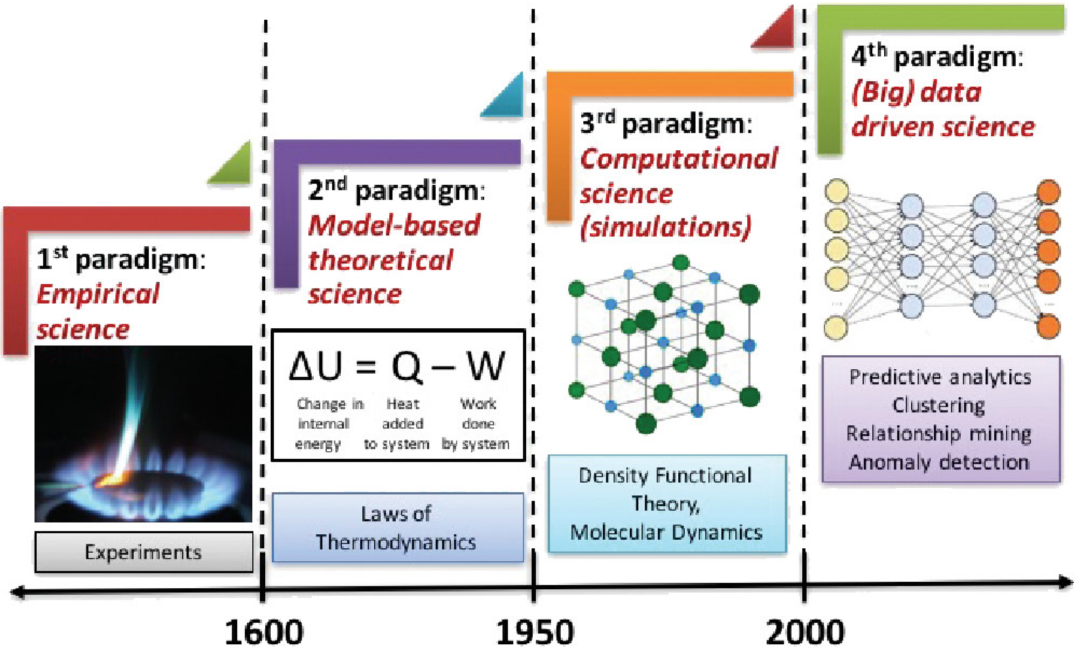


FIGURE 1.2: The four paradigms of materials science: empirical, theoretical, computational, and data-driven. (Agrawal and Choudhary, 2016)

modeling natural behaviors. Consequently, the research field of materials shifted towards theoretical models and generalizations characterized by mathematical equations, such as the laws of thermodynamics. Subsequently, the rise of computational science allowed scientists to simulate complex real-world phenomena using sophisticated algorithms and computer programs, as exemplified by the density functional theory and molecular dynamics simulations. In recent decades, the vast amount of data generated by experiments and simulations have given rise to the fourth paradigm—(big) data-driven science—which unifies theory, experimentation, and computation/simulation. This paradigm has gained traction in materials science, leading to a new field of materials informatics. Figure 1.2 illustrates these four paradigms of materials science. By bridging the gap between domain expertise and exploring uncharted territory, such approaches hold the key to accelerating materials discovery and unlocking new frontiers in scientific advancement.

Several promising projects, such as the Materials Genome Initiative[1] (MGI) or Materials research by Information Integration (MI^2I), have been launched to accelerate the process of discovering and applying these new materials. Such projects encompass both computational methodologies based on high performance computing and high throughput experiments founded on combinatorial techniques. The Materials Genome Initiative (MGI) is a continuation of a movement that began in the 1980s, focusing on designing materials using computation (Potyrailo et al., 2011). The MGI takes this further, using advanced computational capabilities and high-throughput experimental techniques to revolutionize the materials development process. By integrating experimental and computational tools, as well as digital data, the MGI aims to speed up the process of discovering and deploying new materials, leading to a higher number of novel materials being brought to market in a shorter timeframe (Drosback, 2014). This will lead to more incredible innovation and faster development of novel materials to meet the needs of various industries and society

---

[1] About the Materials Genome Initiative, https://www.mgi.gov

as a whole. Machine learning has shown great promise in various fields, and materials science is no exception. In this domain, numerous challenges addressed by data-driven approaches hinge on the effective utilization of existing material data to predict new materials' properties and comprehend the underlying physicochemical mechanisms (Kailkhura et al., 2019).

1.1.2 Challenge in material Informatics

From an engineering perspective, developing a data-driven model capable of rapidly and accurately predicting the physical properties of potential materials from accumulated data can expedite the material development process. By employing a data-driven model for in-silico material screening, the number of candidates necessitating expensive calculations and experimental validation can be reduced. When sufficient independent supervised data is available from the distribution of target material data, state-of-the-art data-driven techniques can be employed to build a highly accurate predictive model. However, given that materials research and development aims to create materials with superior properties, the distribution of target prediction data may differ significantly from the distribution of the original training data. This discrepancy raises concerns about the ability of data-driven models for extrapolation, to predict the physical properties of novel materials accurately.

In the past decade, research using data-driven approaches in materials science has grown significantly, with each study employing a variation of the materials informatics workflow (Saal et al., 2013; Ward et al., 2016; Li et al., 2020). Data-driven methods have gained recognition as powerful tools for accelerating material discovery by reducing the number of candidates to investigate. However, these methods face three critical challenges in the materials research field:

1. **Quality of material data:** The data quality often limits the effectiveness of data-dependent methods. Experimental or computational datasets for materials might need essential information, such as missing or heavily biased data since researchers tend to report successful materials while overlooking unsuccessful attempts. For example, suppose a dataset only contains high-performing materials for a specific purpose without data on low-performing or failed materials. In that case, the machine learning (ML) model may not accurately capture the complete range of material properties and could overestimate new materials' performance.

2. **Inconsistency of material datasets:** Integrating non-comparable datasets derived from various calculation methods or experiments presents a significant challenge for data-driven approaches. Inconsistencies may arise between two datasets of the same material property, collected using different experimental setups or computational methods, leading to discrepancies in reported values. This inconsistency makes it difficult for ML models to learn a reliable relationship between material descriptors and properties.

3. **Difficulty designing suitable material descriptors:** Data-driven approaches often rely on descriptors derived from condensed matter theory and parameters fitted from experimental or computational data to predict potential materials. Descriptors calculated from constituent atomic properties are commonly used, but assessing similarities between materials with varying numbers of compositions remains challenging. For instance, a descriptor that accurately captures the electronic properties of binary alloys may not be suitable for

ternary or more complex alloys, leading to practical limitations in predicting material properties.

In material science, data-driven approaches are faced with significant challenges due to limitations in design descriptors, lack of data, and inconsistent data. These limitations introduce uncertainty into the prediction process, impacting the learned models' accuracy and reliability. Addressing these issues requires innovative approaches to enhance the quality and quantity of material data and develop robust descriptors that capture the intricate relationships between composition, structure, and properties.

1.2 Decision-making in material discovery

Decision-making is an integral part of our daily lives, as we frequently encounter situations that require us to make choices. Some of these decisions are relatively straightforward, such as deciding what to wear in the morning or choosing a meal from a restaurant menu. In contrast, other choices are more complex and require careful consideration, like deciding on a career path, making an investment, or choosing a medical treatment. This fundamental life process has garnered considerable interest and has been the subject of numerous studies since prehistory, leading to a plethora of perspectives, assessment methods, and theoretical models across scientific disciplines and subdisciplines (Thaler and Sunstein, 2009; Mcfall, 2015).

When discovering new materials, decision-making involves choosing candidates for experimental validation from an infinite pool. It's crucial to make effective decisions to allocate limited resources properly. With numerous potential candidates and costly experiments, it's essential to develop rational strategies for decision-making. One crucial aspect is identifying the scenarios that require decision-making, such as selecting candidates for experimental validation, material prioritization, or identifying exploration opportunities. By defining these scenarios, researchers can establish clear objectives and criteria for material evaluation and optimizing resource allocation. This approach reduces the risk of wasted resources, increasing the chances of uncovering novel materials or gaining valuable insights. Ultimately, strategic decision-making plays a critical role in advancing our understanding of materials science and driving progress. It is imperative for researchers to distinguish between risk and uncertainty when making decisions, as suggested by Knight (Knight, 1921). This differentiation is of utmost importance.

1.2.1 Scenarios of decision-making

According to Knight, risk and uncertainty are fundamental concepts that shape the processes, providing a framework for understanding and managing the complexities associated with making choices under uncertain conditions. Therefore, distingusihing between risk and uncertainty is crucial in understanding the complexities of decision-making in various contexts, including the business world. As Knight described, risk refers to situations where we do not know the specific outcome but can accurately measure the odds. It arises when there are multiple possible outcomes, each associated with a *certain* level of probability. In this context, decision-makers can assess and quantify the potential consequences and make informed choices based on the probability of various outcomes. This process often involves weighing each option's potential benefits and drawbacks, considering the likelihood of success and the potential consequences of failure. Probability is a measure of the likelihood of a specific outcome or event occurring, expressed as a number between 0 (impossible) and 1 (certain). In decision-making, probability provides a quantitative way to assess and compare the chances of different outcomes, helping decision-makers make more informed choices under uncertain conditions. Probability theory plays a significant role in developing models and tools to analyze and manage risk in decision-making processes. Decision-makers can use probabilistic models to assess the potential impact of various factors, estimate the likelihood of specific outcomes, and optimize their choices accordingly.

Contrasting with risk, uncertainty pertains to situations where acquiring all essential information to determine probabilities of outcomes or even identify possible outcomes accurately is unachievable. This renders predicting future outcomes exceedingly difficult, if not outright impossible. Previous research suggests that many real-world problems exhibit ill-structured and challenging-to-formalize qualities, with humans facing cognitive limitations when processing the complexities arising from such information (Simon, 1990). This leads to the concept of bounded rationality, where individuals make decisions based on limited information and cognitive capacity. Consequently, decision-makers often resort to heuristics and approximate strategies when uncertain (Gigerenzer and Gaissmaier, 2011; Brighton and Gigerenzer, 2012; Selten, Pittnauer, and Hohnisch, 2012). Although not always optimal, these methods can offer satisfactory and pragmatic solutions in scenarios where comprehensive information and boundless processing capabilities are absent. Nevertheless, it is important to acknowledge that using mental shortcuts may give rise to biases in judgement under uncertainty, potentially leading to errors in decision-making.

Although some economists argue that this distinction is overblown, as real-world events are often so complex that they involve "true uncertainty" rather than risk, Knight's distinction remains relevant in examining decision-making behavior in various domains, such as finance. An instance of this would be investment banks that once relied on their accurate risk assessments, assuming they were working under conditions of Knightian risk and could determine the chances of future outcomes. However, they later realized the insufficiency of their assessments and that they were operating under Knightian uncertainty conditions. This change in perspective could result in them being hesitant to make trades or offer capital, thereby contributing to the further slowdown of the economy. In various fields, including finance or material discovery, Knight's differentiation between risk and uncertainty sheds light on the intricacies of decision-making processes. Distinguishing between the two can have a significant impact on the decisions made by both individuals and organizations. The key to Knight's differentiation lies in assessing how well one comprehends the possible *outcomes* of their decisions and the *probabilities* of each outcome.

Building on Knight's distinction, Meder (Meder, Le Lec, and Osman, 2013) has introduced two most popular scenarios of decision-making that individuals can encounter: *decision-making under risk* and *decision-making under uncertainty* (Fig. 1.3). The first scenario refers to situations where the distribution of the outcomes is known, meaning we have an idea of what could happen and how likely each outcome is. In the first scenario, decisions under risk, decision-makers do not know the certain outcomes of their decisions; however, they still assess probabilities to potential consequences. Despite the risk, decision-makers can evaluate their options by considering each outcome's probabilities and potential utility, allowing them to derive preferences in their choices (Neumann, Morgenstern, and Rubinstein, 1944). For instance,

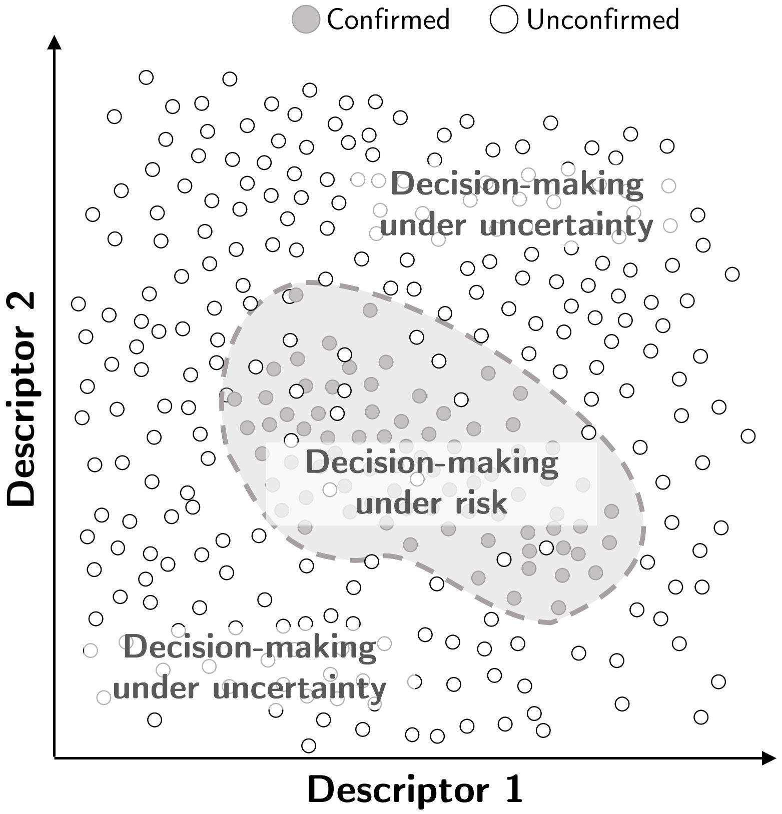


FIGURE 1.3: Illustration of decision-making scenarios in materials discovery: selecting next candidates for computational or experimental validation under uncertainty and risk. (Meder, Le Lec, and Osman, 2013)

random factors, such as the presence of bacteria or wild yeasts, can introduce off-flavors or spoilage present in the fermentation process and introduce unexpected consequences for wine production. Consequently, winemakers must consider the risks associated with various factors in the fermentation process to make rational decisions. By understanding and managing these risks, winemakers can optimize their production processes, minimize potential losses, and improve the overall quality of productions. This method of making decisions when facing risks is useful in different industries and scenarios, where decision-makers aim to balance risk management with informed choices that are in line with their objectives.

In contrast, the second scenario involves situations where we are unaware of the distribution of outcomes, leaving us uncertain about what could happen and how likely each outcome is. This uncertainty makes it challenging to predict material behavior and select candidates for validation in materials science. This scenario is expected when the number of confirmed materials is relatively small compared to the large pool of candidates. For instance, the discovery of high-entropy alloys (HEAs), which are disordered alloys of multiple elements, typically more than four, requires predictive models to assess the likelihood of forming a HEA phase (Yeh et al., 2004). However, the reliability of these models is limited due to the scarcity of confirmed materials data (Ha et al., 2021). The small number of available materials data makes it challenging to assess the likelihood of desired material properties accurately. Therefore, developing robust models to address this type of uncertainty and predict material behavior in the context of HEAs is a crucial area of research in

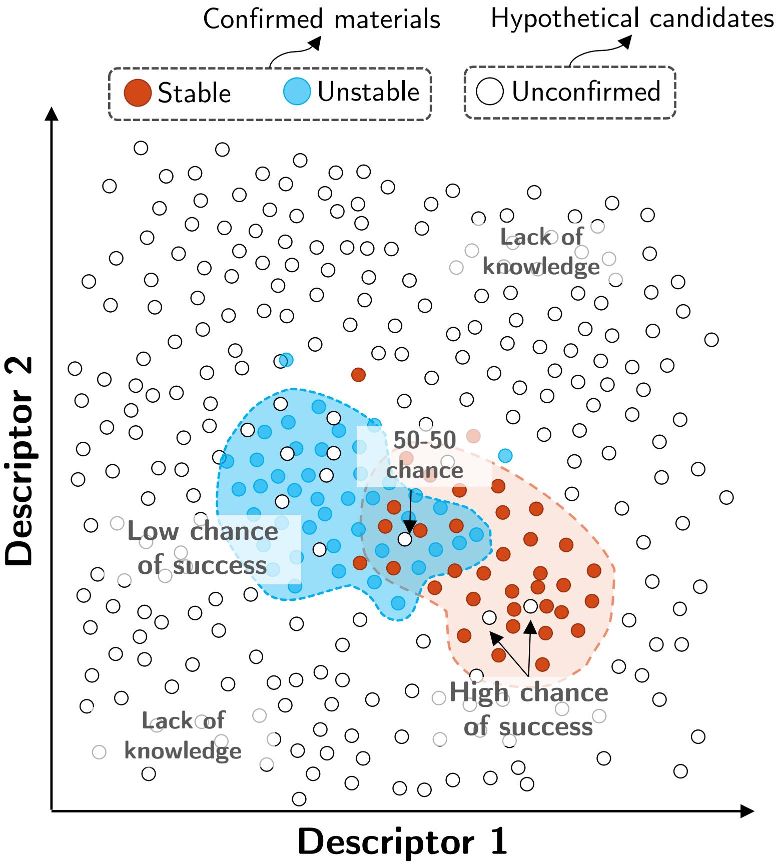


FIGURE 1.4: Illustration of exploitation and exploration criteria in each scenarios of decision-making in materials discovery.

materials science.

1.2.2 Trade-offs between exploitation and exploration in each scenario

Researchers must balance two crucial decision-making criteria in materials discovery: *exploitation* and *exploration*. Exploitation involves prioritizing candidates with a higher chance of success, using existing knowledge and models to narrow the search space efficiently. This criteria optimizes resource utilization and increases the chances of finding materials that meet predetermined criteria. On the other hand, exploration goes beyond existing knowledge to uncover new materials and phenomena with untapped potential. By embracing exploration, researchers can drive innovation and push the boundaries of materials science, leading to significant advancements. Balancing the criteria of exploitation and exploration is essential for making informed decisions, optimizing research efforts, and accelerating the pace of materials discovery.

When making decisions under risk, we may have enough information to assess the likelihood of outcomes accurately. In such cases, we prioritize selecting candidates with a high chance of success in discovering new materials. Based on existing knowledge and models, these candidates have a higher certainty in their predicted properties, as shown in the red region of Figure 1.4. On the other hand, if we are uncertain about predicting the properties of a material, we favor candidates with a 50-50 chance of success. However, it's important to note that selecting such candidates with high uncertainty does not necessarily reduce the uncertainty in prediction. This is because it may arise from random factors not adequately captured by the available descriptor spaces. The upcoming section will discuss the specific challenges related to this issue in detail (Section 2.1).

In situations where uncertainty reigns, relying solely on the exploitation criteria is insufficient as we lack sufficient information to predict a decision's outcome accurately. Instead, it is imperative to concentrate on selecting candidates that will increase our knowledge and understanding of the materials space. We can effectively reduce uncertainty and make better-informed decisions by actively exploring and gathering more data. This approach is crucial in providing valuable insights, enabling us to refine our models and ultimately enhance our ability to discover new materials.

1.3 The rising need for interpretable data-driven methods in materials discovery

Materials science is transforming as data-driven approaches become increasingly integral to the field. These approaches leverage advanced statistical and machine learning algorithms to analyze extensive datasets, revealing hidden patterns and relationships. This shift aligns seamlessly with the ongoing digitization of materials data, offering new avenues for expediting materials discovery. Interpretable data-driven approaches are emerging as particularly invaluable. Beyond merely designing algorithms for learning from data, these approaches focus on constructing systems that reason, elucidate their reasoning, and evolve. Such dynamic systems promise to enhance decision-making in materials discovery, particularly crucial in managing the inherent uncertainty in materials data and streamlining the path to discoveries.

A critical facet of these data-driven approaches is interpretability, which bridges the multifaceted nature of data-driven models and the instincts of materials scientists. Interpretability is indispensable because it enables materials scientists to understand how these models make predictions or reach conclusions. In a field where experimental and computational validation can be exceedingly costly, having an interpretable model allows for a more informed decision-making process, even in the presence of uncertainties. Without interpretability, materials scientists might rely on a "black box," which provides no understanding or insight into the studied materials. This would severely limit the scientist's ability to extrapolate beyond known materials and inhibit innovation.

Material similarity is central to interpretable data-driven approaches in materials science. The concept of similarity is rooted in analogy-based reasoning, where established principles are applied to novel scenarios. This cognitive process is intimately connected with the ability to identify patterns in data that suggest similarities in material compositions and structures. Materials science has recognized the importance of similarities in compositions and structures, as materials with analogous attributes often exhibit comparable properties. However, exploring advanced materials introduces staggering diversity in compositions and structures, presenting opportunities for unprecedented physical properties.

The challenge lies in the diversity and novelty, the sheer number of variables, and the complexity of relationships between compositions, structures, and properties. Discerning material similarities becomes vital for hypothesizing mechanisms behind structure-property relationships, yet the multitude of variables makes this task arduous. Therefore, the data-driven approaches developed for materials science are essential in this complex landscape. Incorporating material similarity and analogy-based reasoning within data-driven strategies is effective in tackling challenges and bridging the intuition-data divide. Consequently, this fusion is paving the way for groundbreaking discoveries in advanced materials, with interpretability as the keystone for informed and innovative advancements.

1.3.1 Analogy-based reasoning: Bridging the gap between human intuition and data-driven approach

Traditionally, scientific intuition has played a pivotal role in identifying and comprehending novel materials. A prime example of this intuition in action is employing analogy-based reasoning, a cognitive process that scientists frequently utilize to apply established principles to new situations. This intuitive approach to reasoning is crucial in experimental materials science and frequently results in fortuitous discoveries.

In materials science, analogy-based reasoning requires carefully identifying similarities between materials. This process relies heavily on human intuition, explicit and implicit knowledge, and abductive reasoning. We must recognize similarities among compositions, structures, and desired properties to transfer knowledge from available systems (source analogs) to unexplored ones (target analogs). However, determining the similarity between materials requires a deep understanding of the correlations between compositions, structures, and outcome properties that determine the physical property of interest. With many features available to describe a material, identifying the relevant attributes that define similarity is a complex and challenging task that varies depending on the target property. This complexity adds an additional layer of difficulty to the identification of material analogs.

Moreover, this form of intuitive, analogy-based reasoning fits naturally with the quest for an interpretable data-driven approach in materials science. Data-driven models strive to identify underlying patterns and structures driving input-output relationships, mirroring the process of finding connections between known and novel materials. By integrating analogy-based reasoning into data-driven models, we can bring the often elusive 'black box' of these models into the light, turning them into transparent, intuitive systems that resonate with a materials scientist's intuitive approach [11]. This inductive reasoning can significantly support high-stakes decisions in materials discovery. It aligns data-driven methodologies with the mindset and goals of materials scientists, making these models more relatable and trusted tools in materials science.

1.4 Research objectives

In the previous section, I highlighted the various decision-making scenarios encountered in the material discovery process and emphasized the importance of identifying the appropriate scenarios to facilitate rational decision-making. Furthermore, considering that human decision-making behavior is sensitive to heuristics and biases, it is crucial to quantitatively evaluate our understanding of the underlying mechanisms in material discovery. Researching new materials is an arduous task due to the immense number of potential candidates that scientists must sort through. While data-driven approaches have shown promise in narrowing down the candidate space, they also face several limitations that hinder their effectiveness:

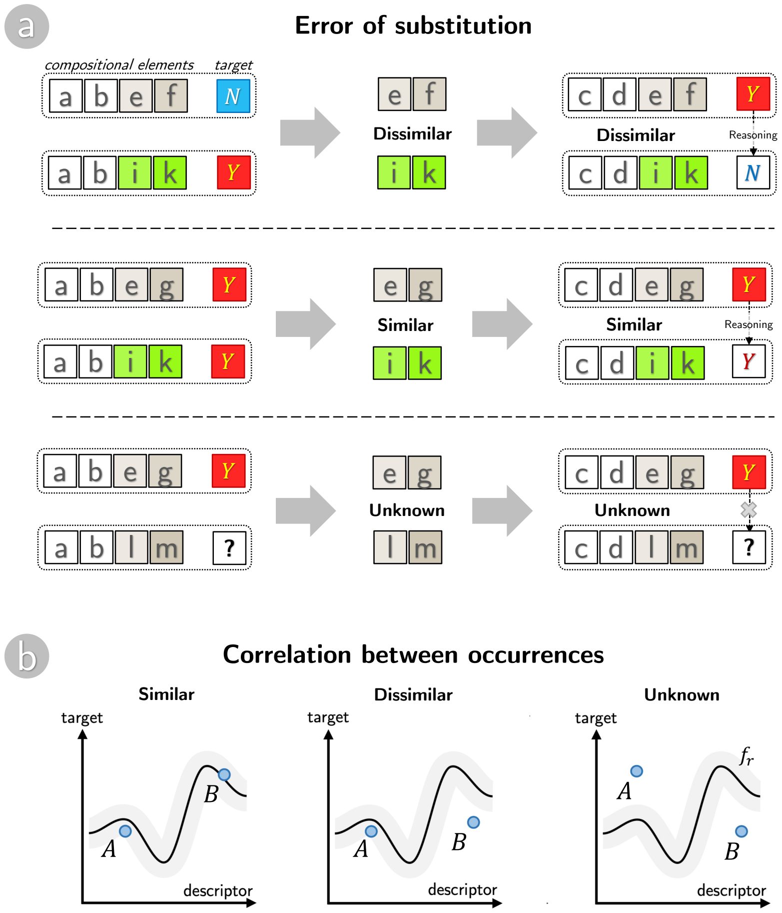


FIGURE 1.5: Data-driven framework for measuring similarities between materials with respect to correlations of their compositions and a binary property (a) or a continuous property (b).

- Designing appropriate descriptors for materials is a challenging task, as it requires capturing the essential features that govern material properties.
- The availability of experimental data is often limited and biased, which can introduce uncertainties and inaccuracies in the analysis.
- Computational data suffers from inconsistencies, as different calculation methods can yield varying results for the same materials.
- Accurately evaluating prediction uncertainty, even when achieving high prediction accuracy is difficult, remains a significant challenge in material discovery research.

To address these limitations, I have developed a data-driven framework that quantitatively measures the similarities between materials while accounting for uncertainty. The similarities between materials are assessed based on the correlation of their compositions and physicochemical properties of interest. Specifically, the framework is designed to focus on two different scenarios, depending on the type of property being examined:

- For binary property (Error of substitution): the similarity is determined by evaluating whether the differences in compositions of materials allow for substitution without affecting the property of interest (Fig. 1.5 a).

- For continuous property (Correlation between occurrences): the similarity is determined by examining the correlation between occurrences of specific physical properties observed in the materials. Materials in the dataset are deemed to be alike if their physical characteristics can be traced back to equivalent underlying mechanisms. Conversely, materials are categorized as dissimilar if the same mechanisms cannot account for their properties (Fig. 1.5 b).

The purpose of this analysis is to identify materials that display similar traits and behaviors, even with variations in their composition. This information can be applied to make decisions in materials discovery and gain insight into the hidden mechanisms that govern the behavior of these materials. Our proposed methods are highly effective, as demonstrated by the creation of a recommender system for exploring new materials (Chapter 3) and utilizing the material similarities to uncover underlying mechanisms (Chapter 4). Additionally, we apply the framework to create a map of candidates that shows the degree of knowledge about the material space, which helps decision-makers select candidates for further validation (Chapter 5).

Chapter 2

Concept of uncertainty in machine learning

In this chapter, we will explore two primary machine learning (ML) techniques based on Bayesian and Dempster-Shafer (DS) theories that assist learners in understanding their level of uncertainty when making predictions. These techniques vary in the type of predictions they produce and how they represent uncertainty. Therefore, it is crucial to examine whether these methods can distinguish between aleatoric and epistemic uncertainty and measure the degree of uncertainty in terms of aleatoric and epistemic uncertainties. In addition, the difficulty in specifying epistemic uncertainty in these approaches stems from the broader challenge in probability theory to define the lack of knowledge. This issue will be more detailed in the subsequent section.

In the first part of this chapter, we explore the intricacies of Bayesian machine learning techniques, such as Bayesian Neural Networks and Gaussian Processes. These methods offer a convenient way to represent and measure uncertainty by utilizing probability distributions to encode uncertainty in model parameters. However, it is essential to note that they may not be able to completely capture the concept of epistemic uncertainty or lack of knowledge due to their dependence on prior knowledge, which can be subjective and may not reflect the true state of knowledge.

The Dempster-Shafer theory, also known as evidence theory, is a more versatile version of the Bayesian approach, especially for situations that involve incomplete or imperfect information. Instead of assigning probabilities to individual elements, this theory assigns non-negative weights to subsets of possibilities, called the frame of discernment. The DS theory can handle both aleatoric and epistemic uncertainties by combining evidence from different sources and allowing for various types of uncertainty representations. However, implementing and computing the DS theory may be more complex, especially when handling large sets of possibilities or evidence. To learn more about these evidence-based approaches, please refer to the third section of the chapter.

2.1 Chance versus imprecision: Aleatory and Epistemic uncertainties

In life and across different industries like finance, engineering, and healthcare, uncertainty is unavoidable and plays a significant role in decision-making. To handle uncertainties better, it is crucial to differentiate between two main types: aleatory uncertainty, which arises from randomness, and epistemic uncertainty, which comes

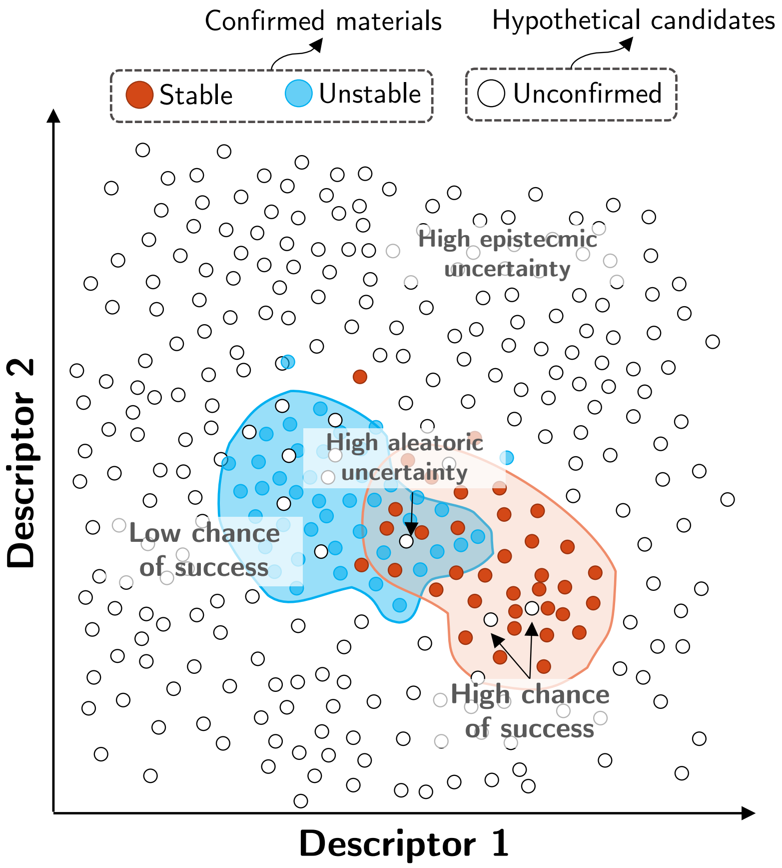


FIGURE 2.1: Illustrations of uncertainty concept for distinguishing scenarios of decision-making in materials discovery. (Hüllermeier and Waegeman, 2021)

from incomplete knowledge. This section of my dissertation will explore the distinctions between these two types of uncertainty and their impact on decision-making and risk management.

The utilization of probability theory has long been the standard approach for evaluating uncertainty. It is widely regarded as the most efficient method for handling uncertainty in fields such as statistics and machine learning. In the realm of probabilistic modeling, relying on a single probability distribution through traditional methods is insufficient in distinguishing between two distinct types of uncertainty: aleatoric and epistemic (Hora, 1996; Der Kiureghian and Ditlevsen, 2009). *Aleatoric* uncertainty, also known as statistical uncertainty, refers to the natural random factors that bring about randomness or variability in the results of an experiment. (Fig. 2.1). Coin flipping is a perfect example of aleatoric uncertainty. The stochastic element involved in the results of tossing a coin cannot be reduced by repeating the attempt unless one possesses the abilities of Laplace's demon. Therefore, accurately representing this uncertainty is a difficult task. Due to the complexity of this process, even the most precise model can only offer the likelihood of two probable outcomes - heads or tails, without a definite answer.

In contrast, *epistemic* (also known as systematic) uncertainty originates from a scantiness of knowledge regarding the optimal model (Fig. 2.1). Epistemic uncertainty is a type of uncertainty that arises from the agent or decision-maker's lack of knowledge, rather than any random phenomenon. Unlike uncertainty due to randomness, epistemic uncertainty can be minimized by acquiring additional information. For instance, the question, "Can you please tell me the meaning of the Swahili word ’kichwa’? Does it refer to the head or the tail?" might prompt the same uncertainty as a coin flip, but the nature of the uncertainty is different. It can be easily resolved with additional knowledge. In this context, epistemic uncertainty refers to the part of the total uncertainty that can be lessened, while aleatoric uncertainty accounts for the part that cannot be minimized. Recognizing the distinction between aleatoric and epistemic uncertainties is essential for developing appropriate models and strategies in various domains, as it allows decision-makers to focus on the aspects of uncertainty that can be reduced and managed more effectively.

In the application of data-driven approaches to material discovery, it is common for the distinct types of uncertainty, aleatoric and epistemic, to be improperly identified. Sometimes, making a distinction might appear unnecessary, especially when we need to make a decision or prediction, and the origin of uncertainty is immaterial. Bayesians often use this viewpoint to support a probabilistic approach and classical Bayesian decision theory. Nevertheless, it is essential to note that this scenario does not always apply. For example, in some situations, the final decision can be delayed or refused, like in a classification process that includes a rejection option (Chow, 1970; Hellman, 1970). Additionally, actions can be explicitly taken to reduce uncertainty, as in active learning (Aggarwal et al., 2014). In these cases, recognizing the difference between aleatoric and epistemic uncertainties becomes more relevant, as it can help guide the decision-making process more effectively. For example, strategies to reduce epistemic uncertainty may involve collecting more data or refining the learning algorithm, while managing aleatoric uncertainty might require the development of more robust models or incorporating domain knowledge. By distinguishing between the two types of uncertainty and understanding their implications, researchers and decision-makers in material discovery can better allocate resources, develop targeted strategies, and ultimately make more informed decisions.

Motivated by scenarios that emphasize the importance of distinguishing between aleatoric and epistemic uncertainties, Senge et al. explicitly address this distinction in their data-driven approach to decision support (Senge et al., 2014). They propose methods to measure such uncertainties and illustrate the effectiveness of their approach in the field of medical decision-making. Similarly, Kull and Flach suggest using trustworthiness maps to differentiate between predicted probability scores and the level of uncertainty in those predictions (Kull and Flach, 2014). They illustrate this distinction using a weather forecasting example, wherein the aleatoric uncertainty refers to the 50% chance of rain, while the uncertainty in the 20% estimate is related to epistemic uncertainty. In a more practical context, Varshney and Alemzadeh discuss the importance of understanding epistemic uncertainty in AI-assisted systems, using the example of a self-driving car accident that resulted in the driver’s death (Varshney and Alemzadeh, 2017). They attribute the failure of cars to exceptionally infrequent circumstances and emphasize the importance of epistemic uncertainty in such systems. Such examples underscore the value of differentiating between aleatoric and epistemic uncertainties, as doing so can provide more accurate and trustworthy representations of uncertainty in decision-making processes. By accounting for both types of uncertainty, researchers and practitioners can develop more reliable models and predictions, ultimately leading to better-informed decisions in various fields, including material discovery, healthcare, and AI-assisted systems.

In recent years, the discrepancy between aleatoric and epistemic uncertainties has gained attention in the field of ML (Kendall and Gal, 2017). One of the main concerns with deep learning models, particularly neural networks, is their limited awareness of their own competence. Experiments on image classification have revealed instances where trained models fail in specific situations, despite being extremely confident in their predictions. It is crucial to note that these models lack resilience and are highly susceptible to being misled by "adversarial examples" (Papernot and McDaniel, 2018). It is surprising how even small changes in the input can greatly affect the predictions of a model. This makes it vulnerable to adversarial attacks, not just in images but also in natural language text and other data types. To improve the model's performance and protect it from such attacks, researchers are now incorporating the difference between aleatoric and epistemic uncertainty in deep learning models. This approach aims to enhance the model's understanding of its own capabilities and increase its reliability and trustworthiness, particularly in applications that require safety and precision, such as autonomous vehicles, healthcare, and other AI-driven systems.

To differentiate between aleatoric and epistemic uncertainty, one can examine whether the uncertainty can be reduced with additional information (Hüllermeier and Waegeman, 2021). Aleatoric uncertainty is the fundamental part of the uncertainty that arises from the unpredictable correlation of input instances (x) and outcomes (y). The relationship between x and y can be seen in the probability of one variable occurring given the occurrence of the other, known as conditional probability. Aleatoric uncertainty arises from the inherent randomness of the process and cannot be reduced by gathering more data. On the other hand, epistemic uncertainty arises when there is insufficient knowledge about the optimal model, such as uncertainty about a model's parameters. Unlike aleatoric uncertainty, epistemic uncertainty can be decreased by obtaining more information, refining the model, or gathering more data. This type of uncertainty represents the limitations in our understanding of the underlying processes or relationships, and it can be minimized as our knowledge and modeling capabilities improve. At first glance, this characterization seems clear, but it becomes less straightforward upon further examination. What does it mean for uncertainty to be "reducible"?

One obvious way to obtain more information is through additional training data ($\mathcal{D}$), which can decrease the learner's uncertainty while maintaining the same learning problem settings, such as the instance space ($\mathcal{X}$), output space ($\mathcal{Y}$), hypothesis space ($\mathcal{H}$), and joint probability ($\mathcal{P}$) on $\mathcal{X} \times \mathcal{Y}$. However, this is not always the case in practice. An example of how a learner can enhance instance descriptions is by adding extra features. This integration of the current instance space ($\mathcal{X}$) results in the creation of a new instance space ($\mathcal{X}'$). This alteration may affect the level of uncertainty. In the given illustration of Figure 2.2 a, we can see the overlap of two class distributions of stable and unstable materials in a two-dimensional space defined by descriptor 1 and descriptor 2. This overlap creates aleatoric uncertainty in the overlapping region of the instance space. Incorporating the data into a higher-dimensional space (achieved by adding a third descriptor) makes the two classes discernible, and the uncertainty can be resolved (Fig. 2.2 b). Typically, when data is placed in a higher-dimensional space, the level of aleatoric uncertainty decreases. However, it also leads to an increase in epistemic uncertainty. This is because the process of fitting a model becomes more complex and requires a greater amount of data.

Previous example demonstrates that the two uncertainties (aleatoric and epistemic) are not fixed concepts but context-dependent, relying on the specific setting

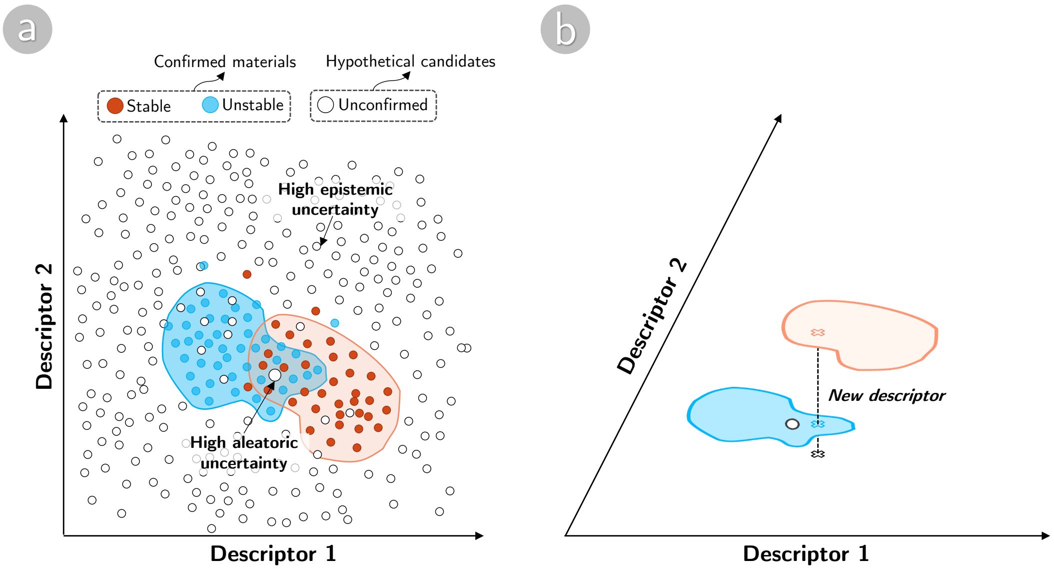


FIGURE 2.2: (a) Illustration of the high aleatoric (the overlapping region between stable and unstable materials) and epistemic uncertainties. (b) Aleatoric uncertainty is reducible by introducing a third descriptor.(Hüllermeier and Waegeman, 2021)

of problem $(\mathcal{X}, \mathcal{Y}, \mathcal{H}, \mathcal{P})$. Altering the circumstance can then modify the origins of uncertainty. It is within the realm of possibility for aleatoric uncertainty to convert into epistemic uncertainty, and the reverse can also occur. As a result, when learners have the ability to modify settings, it can lead to uncertainties that may create confusion between the two types of uncertainty, ultimately affecting their accurate measurement. According to Der Kiureghian and Ditlevsen, the connection between aleatoric and epistemic uncertainty can only be clearly established within a specific analytical framework. They assert that what might be categorized as aleatory uncertainty in one model may be considered as epistemic uncertainty in another model. (Der Kiureghian and Ditlevsen, 2009).

In a decision-making context, overall uncertainty can arise from both aleatoric and epistemic uncertainties. Knowing and understanding the differences between them can assist decision-makers in devising effective strategies to address each type of uncertainty. Revisiting the five decision-making scenarios mentioned earlier; we find that the second and third scenarios, decision-making under risk and black-swan events, exhibit high aleatoric uncertainty and low epistemic uncertainty. On the other hand, the remaining scenarios, decision-making under Knightian uncertainty and radical uncertainty, demonstrate high epistemic uncertainty. By distinguishing between these two types of uncertainty, decision-makers can better comprehend the sources of variability and imprecision in their systems, allowing them to create suitable strategies to tackle them for accelerate the process of material discovery and material development. To mitigate aleatoric uncertainty, decision-makers often revise their models reprenseting hypothesis about nature mechanisms, while addressing epistemic uncertainty usually involves gathering additional information.

2.2 Representation of uncertainty by using probability theory

2.2.1 Scoring, calibration, and ensembling to enhance probability estimation

Several ML methods can be employed to induce probabilistic predictors, which are hypotheses h that produce probability estimates $p_h(.|x) = p(.|x,h)$ instead of point predictions $h(x) \in \mathcal{Y}$. For classification tasks, this involves predicting a sole conditional probability $p_h(y|x) = p(y|x,h)$ for each label $y \in \mathcal{Y}$. Meanwhile, regression tasks involve predicting a density function $p(.|x,h)$ on $\mathbb{R}$. Such estimators can be trained discriminatively, i.e., as a mapping $x \rightarrow p(.|x)$, or generatively, which basically implies estimating a joint distribution on $\mathcal{X} \times \mathcal{Y}$. There are two types of approaches to data analysis: parametric and non-parametric. Parametric approaches assume specific families of probability distributions, while non-parametric approaches do not make such assumptions. Examples of these approaches include logistic and linear regression for classical statistical techniques, Bayesian networks and Gaussian processes for Bayesian approaches, and multifarious methods in the domain of (deep) neural networks. These probabilistic models provide a powerful way to model uncertainty in machine learning tasks by generating complete probability distributions on Y, enabling more informed decision-making and a better understanding of the underlying uncertainties.

To train probabilistic predictors, the usual approach is to minimize loss functions that ensure the predicted conditional probabilities are correct. Proper scoring rules take a significant part in this process, and the log-loss is a well-known example of such rules, as mentioned in Gneiting's work (Gneiting and Raftery, 2007). These loss functions are essential in guiding the learning process and ensuring accurate probability estimates from the predictor. In certain cases, probability estimates are obtained through simple frequentist techniques like Naïve Bayes or nearest neighbor classification. These methods rely on counting occurrences and calculating relative frequencies to generate probability estimates. Although not as sophisticated as proper scoring rule methods, they can be effective in specific situations and provide a baseline for advanced probabilistic prediction approaches.

Some predictive methods generate "pseudo-probabilities" that may not be entirely accurate. Additionally, certain methods produce natural scores that express a level of confidence but cannot be considered probabilities. An example of this is the usage of the distance from the separating hyperplane in support vector machines, which indicates the level of confidence in the prediction obtained from the hyperplane. To address this, scaling or calibration methods transform these scores into proper, well-calibrated probabilities by learning a mapping from scores to the unit interval. This mapping can then be applied as a post-processing step to the predictor's output. (Flach, 2017). There are various methods of calibration such as binning (Zadrozny and Elkan, 2001), isotonic regression (Zadrozny and Elkan, 2002), logistic scaling (Platt et al., 1999), as well as the use of Venn predictors (Johansson et al., 2018). The field of calibration remains an active area of research, with ongoing efforts to develop and refine techniques that ensure the resulting probability estimates are more accurate and reliable. These calibrated probabilities can provide a more meaningful measure of uncertainty and improve decision-making processes in various applications.

In machine learning, ensemble learning methods like bagging and boosting have become increasingly popular. This is because they have the ability to enhance the

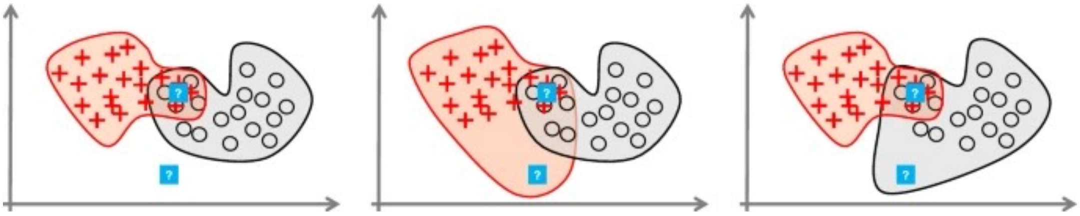


FIGURE 2.3: When the hypothesis space is very complex and can produce almost any decision boundary, areas with few training examples tend to have high epistemic uncertainty. In the left picture, the lower query point could be classified as either red (middle picture) or black (right picture), while the second query is located in a region where the two class distributions overlap, leading to high aleatoric uncertainty. (Hüllermeier and Waegeman, 2021)

precision of point predictions. In addition, these methods generate a vast set of estimators $(h_1, \ldots, h_M)$ instead of a single one, which may lead to the desire to generate probability estimates through fundamental frequentist principles. In basic classification, each prediction for a class can be seen as a vote indicating a preference. These preferences can be expressed as probabilities using relative frequencies. However, it is important to know that this method can result in biased probabilities requiring further calibration. In addition, tree-based methods like random forests (Kruppa et al., 2014; Breiman, 2001), are particularly significant in this field. These methods create ensembles of decision trees, leveraging the wisdom of the crowd to produce more accurate and robust predictions. While ensemble learning methods can provide better predictions, it is crucial to ensure that the resulting probability estimates are well-calibrated to deliver meaningful measures of uncertainty and to support informed decision-making processes in various applications.

Indeed, standard probability estimation is a useful approach for representing uncertainty in predictions. Nonetheless, it fails to explicitly differentiate between two types of uncertainty. Most techniques in this group focus on aleatoric uncertainty, which is the inherent randomness or noise present in the data. Aleatoric uncertainty arises due to the variability in the observed data, and it is generally considered irreducible. On the other hand, epistemic uncertainty arises due to insufficient knowledge or information regarding the fundamental process or associations between variables. Reducing epistemic uncertainty is possible by either acquiring additional data or enhancing the model's assumptions and structure.

In order to quantify epistemic uncertainty, an intuitive idea is to consider generative models. Usually, these methods analyze the densities of $p(x)$ to identify if input points are situated in areas with high or low density, where the latter serves as a substitute for high epistemic uncertainty. (Fig. 2.3). Traditional methods like kernel density estimation or Gaussian mixtures can be used to estimate the density $p(x)$. However, newer methods for density estimation are constantly emerging in the field of machine learning. Such recent approaches are isolation forests (Liu, Ting, and Zhou, 2008), auto-encoders (Goodfellow, Bengio, and Courville, 2016), and radial basis function networks (Hossein Zadeh Bazargani and Mac Namee, 2020). These methods demonstrate that density estimation is also prominent in anomaly and outlier detection. Usually, a threshold of the density $p(x)$ is assigned to identify whether a data point is considered as an outlier. A prime example is the utilization of auto-encoders to create a compressed version of the input data and then scrutinize the reconstruction error to establish its position within the distribution. These techniques are considered semi-supervised outlier detection methods due to their independence from annotated outliers for training.

Several semi-supervised outlier detection techniques are influenced by the concept of one-class support vector machines (SVMs). These SVMs create a hyperplane that distinguishes outliers from other regular data points in a latent space with higher dimensionals. (Khan and Madden, 2014). Variations of this approach exist, such as fitting a hypersphere instead of fitting a hyperplane for the data (Tax and Duin, 2004). Using generative models and density estimation techniques makes it possible to better quantify and understand epistemic uncertainty in various applications. Detecting outliers can be compared to a classification scenario where the option to discard data is available. For example, in situations where a classifier decides not to predict a label in areas with low density (Perello Nieto et al., 2016). Outliers are not tolerated when predicting data as they can significantly affect accuracy. This approach prioritizes epistemic uncertainty above all else. It's important to note that papers on classification using a reject option often rely on reasoning based on conditional class probabilities, represented by $p(y|x)$, and specific utility scores. Many methods focus on estimating aleatoric uncertainty. However, it is crucial to understand that different papers may have varying interpretations of uncertainty. To adequately address uncertainty, it's essential to differentiate between epistemic and aleatoric uncertainty and identify their sources.

In addition, epistemic and aleatoric uncertainty can be estimated by generative models in a simple way. The difference between such two uncertainties is known as conflict, which means overlapping distributions, and ignorance, which refers to sparsely populated regions with no evidence for any class (Hullermeier and Brinker, 2008). Similarly to other methods that have been mentioned, generative models also come with drawbacks. Implementing semi-supervised outlier detection methods can pose some challenges, such as adequately establishing the threshold to identify outliers and selecting an appropriate model class for the generative model. Density estimation, a complex problem, often requires a considerable amount of data to ensure precise results. It can be difficult to accurately determine the appropriate model class when working with a small sample size, resulting in increased model uncertainty. Despite these challenges, generative models can still provide valuable insights into epistemic and aleatoric uncertainties, especially when combined with other approaches or when the sample size is sufficiently large. By considering the strengths and weaknesses of different methods, researchers can develop more effective strategies for handling uncertainty in machine learning tasks.

In summary, when using standard probability estimation techniques, it is essential to recognize their limitations in distinguishing between aleatoric and epistemic uncertainty. To better understand and quantify the different types of uncertainty, alternative approaches, such as Bayesian methods or Dempster-Shafer theory, may be more appropriate. These methods can help provide a more nuanced understanding of uncertainty, thereby supporting more informed decision-making processes in various applications.

2.2.2 Bayesian inference

Some deep learning techniques use probabilistic approaches to estimate the likelihood of an individual model output or adjust the calibration after the fact (Jiménez-Luna, Grisoni, and Schneider, 2020a). Various techniques involve treating neural networks as Bayesian models, where a prior distribution is considered over their learnable weights. Inference is then performed over their posterior distribution using different methods, such as Markov chain Monte Carlo (Zhang et al., 2020) or variational inference (Graves, 2011). Gal and Ghahramani, 2016 suggested using dropout regularization to perform approximate Bayesian inference (Kendall and Gal, 2017), which was later extended to compute both epistemic (caused by model misspecification) and aleatoric uncertainty (inherent to data noise). Batch normalization can also make similar approximations (Teye, Azizpour, and Smith, 2018).

Mean-variance estimation is another approach where a neural network is designed to output both a mean and variance value (Nix and Weigend, 1994). To train the model, a negative Gaussian log-likelihood loss function is utilized. Additionally, some approaches involve making Gaussian distributional assumptions about prediction errors and using asymptotic approximations, like the delta technique (Chryssolouris, Lee, and Ramsey, 1996; Hwang and Ding, 1997). These methods aim to provide more reliable and robust predictions by accounting for both aleatoric and epistemic uncertainties in deep learning models.

The Bayesian approach is an absolute necessity for deep learning techniques that deal with uncertainty, particularly in relation to neural networks. These methods are specifically tailored for neural networks as a model class. When utilizing a standard neural network for classification, the final layer commonly produces a probability distribution on the set of classes $\mathcal{Y}$ for a given query $x \in \mathcal{X}$, utilizing techniques like softmax. This distribution is based on a set of classes. For regression problems, a distribution like a Gaussian is placed over a point prediction $h(x) \in \mathbb{R}$, which is typically regarded as the distribution expectation. While neural networks can execute maximum likelihood inference and produce probabilistic predictions, they are incapable of providing information about the confidence in these probabilities. In other words, neural networks merely capture aleatoric uncertainty and not epistemic uncertainty (Hüllermeier and Waegeman, 2021).

2.3 Representation of ignorance: Sets versus Distributions

Methods based on probability theory offer various ways to represent aleatoric and epistemic uncertainty. However, these methods might not be entirely accurate when it comes to representing total ignorance. In situations of complete ignorance, probability-based methods often rely on the assumption of maximizing entropy, which implies that the probability is uniformly distributed. The uniform distribution represents a state of maximum uncertainty, where every outcome is considered equally likely. While this assumption may serve as a convenient starting point in some cases, it may not accurately represent the actual state of knowledge or lack thereof in many real-world situations.

Knowledge in traditional probabilistic modeling and Bayesian inference is conveyed via a distribution p on Ω. However, it can be challenging to determine how to modify or "dilute" this knowledge. Belief is distributed among elements in a unit mass fixed in terms of the total amount of belief in $\omega \in \Omega$. Unlike sets, adding or removing a candidate ω does not affect the plausibility of other candidates. This is because increasing the weight of one alternative ω means decreasing the weight of another alternative ω by the same amount. Measuring uncertainty for probability distributions, such as Shannon entropy for finite Ω, primarily capture the distribution's shape (its "peakedness" or non-uniformity) and inform about the predictability of a random experiment's outcome (Hullermeier and Brinker, 2008). From this perspective, these measurements are more closely connected to aleatoric uncertainty.However, the set-based method, where knowledge is represented through a subset $C \subseteq \Omega$ of potential options, is thought to be more effective in estimating epistemic uncertainty. By explicitly representing the set of plausible candidates, it more directly reflects the lack of knowledge about the actual underlying model or state, offering a more intuitive way to express uncertainty in situations where knowledge is incomplete or uncertain.

Indeed, some researchers consider the representation of ignorance or lack of knowledge using probability distributions problematic (Dubois, Prade, and Smets, 1996). A significant problem arises when probability theory uses a uniform distribution to represent complete ignorance, as it fails to differentiate between having precise probabilistic knowledge about a random event and not having any knowledge due to an incomplete experiment description. This limitation was highlighted by Ronald Fisher, who argued that *not knowing the chance of mutually exclusive events and knowing the chance to be equal represent two different states of knowledge*.

Recognizing the impact of probability, one must acknowledge that its measure-theoretic foundation and the risk of addiction can present significant challenges. Notably, it's essential to understand that the uniform distribution, even after reparametrization, may not maintain its consistency. A uniform distribution on a parameter ω does not decode to a uniform distribution on $1/\omega$, even though ignorance about ω should imply ignorance about $1/\omega$. Similarly, expressing ignorance about a cube's length x using a uniform distribution on an interval $[l, u]$ does not result in a uniform distribution of x^3 on$[l^3, u^3]$, which implies some degree of informedness about its volume. These problems raise questions about using the uniform prior distributions, often interpreted as representing epistemic uncertainty in Bayesian inference, and whether they are the most appropriate way to model ignorance. There is a belief that one probability distribution is not enough to represent uncertain knowledge, resulting in the creation of alternative versions of standard probability theory. These include imprecise probability (Walley, 1991), evidence theory(Shafer, 1976; Kennes, R, and Smets, 1994), and possibility theory (Dubois and Prade, 2012). The aim of these formalisms is to seamlessly combine the distinct characteristics of sets and distributions.

Uncertainty representation and reasoning are essential in various fields, such as artificial intelligence, engineering, and the natural sciences. A common approach to describe uncertain situations involves using a reference set Ω, also known as the frame of discernment (Shafer, 1976). The set Ω comprises all hypotheses or precise information that must be distinguished in a particular context. The elements $\omega \in \Omega$ are exhaustive and mutually exclusive, with one of them, $\omega^\star$, representing the truth. Various examples can illustrate the concept of the frame of discernment. In the case of coin tossing, $\Omega = \{H, T\}$, where H represents heads, and T represents tails. In predicting the result of a football match, $\Omega = \{win, loss, tie\}$. In estimating the parameters (expected value and standard deviation) of a normal distribution from data, $\Omega = \mathbb{R} \times \mathbb{R}_+$, where $\mathbb{R}$ represents real numbers and $\mathbb{R}_+$ represents positive real numbers. In subsequent discussions, we will assume that Ω is a discrete set (finite or countable) for simplicity and to avoid measure-theoretic complications. This assumption allows for easier exposition and understanding of the concepts related to uncertainty representation and reasoning.

The assumption of Ω exhaustiveness is only sometimes applicable in certain situations. For instance, in problems related to machine learning classification, it is possible that not all classes are known beforehand, or new classes may come up in the future (Liang, Li, and Srikant, 2018; DeVries and Taylor, 2018). This scenario is referred to as the "open world assumption," in contrast to the "closed world assumption," where only possibility options is accounted for Ω (Deng, 2014). Though this distinction may seem technical, it has significant implications for handling and representing uncertain information, particularly in the context of the empty set. Assuming a closed world means that an empty set is not valid information. However, if you assume an open world, an empty set may indicate that the natural state $\omega^\star$ is outside of Ω. This concept makes managing uncertain information in fields like artificial intelligence and machine learning more complex. Therefore, it is crucial for researchers and practitioners to carefully examine their assumptions regarding the comprehensiveness of their frame of discernment and the potential consequences of such assumptions on the handling of uncertainty.

There are two main ways to express uncertain information about $\omega^\star$: subsets and distributions. When we use a subset $C \subseteq \Omega$, we are suggesting that $\omega^\star$ is included in that set. This method of expressing information or knowledge allows us to differentiate between possible values and those that are definitely not an option. This approach is suitable for capturing uncertainty that arises due to imprecision, as seen in examples where incomplete information about a numerical quantity is specified as an interval $C = [l, u]$.

In contrast, a distribution provides a weight, known as $p(\omega)$, to each element ω. This weight represents a degree of belief. This approach is a suitable extension of the set-based method. Any subset C can be described by its indicator function $\mathbb{I}_C$ on Ω, assuming no restrictions on the weights. This function assigns a weight of 1 to each $\omega^\star$ element in C and 0 to $\omega^\star$ elements not in C. However, this view is not valid for the specifically important case of probability distributions. The reason is that probability distributions adhere to specific constraints, such as normalization (summing up to 1). Consequently, a probability distribution cannot fully capture the same sense of uncertainty as a set-based representation. For example, the current approach to handling uncertainty may not consider the level of confidence in a model or knowledge about a situation, known as epistemic uncertainty. Alternative frameworks like imprecise probabilities, evidence theory, and possibility theory have been developed to address this limitation. These frameworks aim to combine set-based and distribution-based representations of uncertainty to capture better the different types of uncertainties that may occur in different contexts.

2.4 Dempster–Shafer theory

Quantifying and representing uncertainty are essential aspects of decision-making and reasoning in many fields, including artificial intelligence, engineering, and natural sciences. Several theories have been proposed to handle uncertainty, including probability theory, evidence theory (Shafer, 1976; Kennes, R, and Smets, 1994), imprecise probability (Walley, 1991), and possibility theory (Dubois and Prade, 2012). Evidence theory, also known as Dempster-Shafer theory, has gained prominence due to its expressive representation and ability to handle partial and conflicting information. This essay discusses the advantages of evidence theory and why it is often chosen over its alternatives, namely imprecise probability and possibility theory.

One of the significant advantages of evidence theory is its explicit representation of uncertainty. Evidence theory employs essential probability assignments (BPAs) and mass functions to represent uncertain knowledge. These representations allow mass assignment to all subsets ω of the frame of discernment Ω, excluding the empty set ($\omega \neq \emptyset$). This flexibility in mass assignment enables evidence theory to capture both partial and conflicting information, which is often difficult to

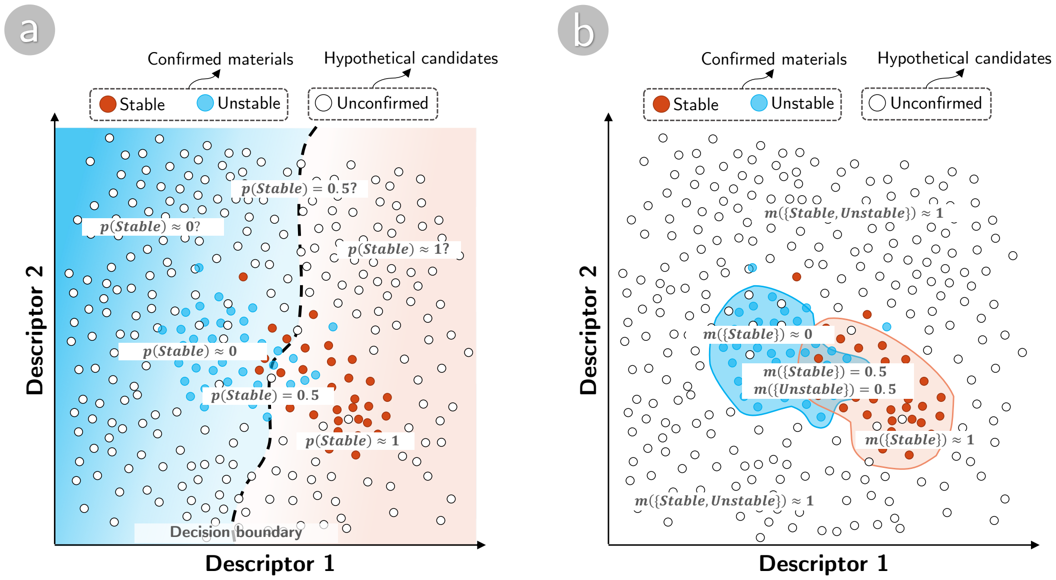


FIGURE 2.4: Representation of ignorance by using Bayesian theory (a) and Dempster-Shafer theory (b).

achieve with other uncertainty representation frameworks like imprecise probability and possibility theory. Evidence theory is particularly well-suited for handling ignorance or incomplete information (Fig. 2.4). Unlike probability theory, which models complete ignorance using a uniform distribution, evidence theory explicitly represents ignorance by assigning mass to the subsets of the frame of discernment. This feature is advantageous when dealing with situations where there needs to be more knowledge or complete information about the problem. In contrast, imprecise probability and possibility theory may need help to capture and represent ignorance as effectively as evidence theory.

Evidence theory is an excellent method for systematically combining evidence from multiple sources. With Dempster's rule of combination, one can effectively combine evidence while considering any potential conflicts and uncertainties. This approach is beneficial when dealing with evidence from sources that may not be entirely reliable. While imprecise probability and possibility theory are viable options for combining evidence, Dempster's rule of combination in evidence theory is more robust and systematic. Furthermore, evidence theory provides belief and plausibility functions that aid in interpreting and comprehending the uncertainty related to a specific event or outcome. Belief and plausibility functions provide lower and upper bounds on the probability of events, respectively, facilitating decision-making processes by making it easier to understand the implications of uncertain information. In comparison, imprecise probability and possibility theory may not provide as straightforward or easy-to-understand measures of uncertainty as evidence theory does.

In summary, evidence theory presents several advantages over imprecise probability and possibility theory, making it a popular choice for representing and reasoning with uncertainty in various applications. Its expressive representation, ability to handle ignorance, a systematic combination of evidence from multiple sources, and interpretable uncertainty measures all contribute to its superiority over other uncertainty representation frameworks. However, it is essential to note that choosing the best framework ultimately depends on the specific problem and the nature of the uncertainties involved. In some cases, imprecise probability or possibility theory may be more appropriate for the task. Nonetheless, evidence theory remains a versatile and powerful tool for representing uncertainty and should be carefully considered when dealing with partial or conflicting information from multiple sources.

2.4.1 Overview

Recognizing the value of the Dempster-Shafer theory, commonly known as evidence theory, is crucial as it is often considered a superior approach to representing and merging beliefs in comparison to traditional probability theory. The main ideas behind Dempster–Shafer theory are: 1) using subjective probabilities for a related question to obtain beliefs for a given question, and 2) applying Dempster's rule to combine independent, evidence-based beliefs (Shafer, 1976). Degrees of belief in this framework are represented as belief functions, also known as mass functions assigning masses to sets of possibilities rather than single events. This approach has the advantage of encoding evidence in favor of propositions more naturally and providing a more expressive representation of uncertainty (Smets, 1993).

Shafer's formalism, or evidence theory, provides a framework for representing belief and uncertainty about propositions by assigning belief and plausibility values as intervals. Based on the available evidence, these intervals capture the lower and upper bounds of the possibility of a proposition being true. In Shafer's framework, a set of possibilities, known as the frame of discernment Ω, is considered. A subset of this frame represents a possible outcome. Subjective probabilities, or masses, are assigned to all frame subsets, with non-zero masses typically assigned to a restricted number of sets called focal elements.

In Dempster–Shafer theory, belief functions assign masses to all nonempty subsets of the Ω, frame of discernment, meaning they operate on the power set of the states. For instance, consider a situation with two possible states. In this case, a belief function would assign mass to the first state, the second state, both states. This way of representing uncertainty allows for a more nuanced representation of epistemic plausibilities, sometimes leading to answers that contradict those derived using probability theory. Dempster–Shafer theory is often used as a method of sensor fusion, combining information from multiple sensors or sources to obtain a more accurate and reliable estimation. The theory's appeal lies in its ability to encode evidence naturally in favor of propositions and its flexibility in handling different types of uncertainties.

The belief in a hypothesis is determined by the combined mass of all its subsets. Belief, also known as Bel, represents the degree of evidence supporting a proposition and can range between 0 (no evidence) and 1 (certainty). On the other hand, plausibility serves as the maximum likelihood of the hypothesis being accurate. It is one minus the sum of the masses of all sets whose intersection with the hypothesis is empty. Plausibility is related to Belief through the equation $Pl(p) = 1 - Bel(\neg p)$, and it also ranges from 0 to 1. Plausibility measures the extent to which evidence supporting a proposition's negation affects the ability to believe in it.

Suppose there is a belief that the cat inside the box has a 50% chance of being dead. However, evidence also indicates a 20% chance that the cat is alive. The remaining 30% is considered as epistemic uncertainty, which means we cannot be sure whether the cat is dead or alive. The difference between these two subsets represents the aleatoric uncertainty from the randomness factors in the hypothesis about the cat's life. Shafer's formalism can help us better understand uncertainty and how it relates to evidence in complicated situations.

In this scenario, there are two possible outcomes in the frame of discernment: $\Omega_{cat} = \{"Alive", "Dead"\}$. From these outcomes, we establish three subsets that represent hypotheses about the cat's life: "*Alive*", "*Dead*", and "*Either*". The "*neither*" situation is automatically assigned a value of zero as a result of the closed-world assumption. The hypotheses "*Alive*" and "*Dead*" are orthogonal, and their probabilities are 0.2 and 0.5, respectively. These probabilities may correspond to the dependability of a "Live/Dead Cat Detector" signal. The "*Either*" hypothesis includes all possibilities and ensures that the sum of all probabilities adds up to 1. Belief in "*Alive*" and "*Dead*" is equivalent to their corresponding probabilities, as they have no subsets. Conversely, the belief for "*Either*" is the total of all three probabilities (*Alive*, *Dead*, and *Either*), as "*Alive*" and "*Dead*" are both subsets of "*Either*". The plausibility for "*Alive*" is $1 - m(Dead) = 0.5$, and the plausibility for "Dead" is $1 - m(Alive) = 0.8$. Alternatively, the plausibility for "*Alive*" is $m(Alive) + m(Either)$, and the plausibility for "*Dead*" is $m(Dead) + m(Either)$. Finally, the plausibility for "*Either*" is the sum of $m(Alive)$, $m(Dead)$, and $m(Either)$. The universal outcomes ("*Either*") will always have 100% belief and plausibility, acting as a sort of checksum. This example demonstrates how the Dempster-Shafer theory can calculate belief and plausibility values for various hypotheses. This representation captures the uncertainty and interplay between evidence in the system, providing a more expressive representation of uncertainty than traditional probability distributions.

Dempster-Shafer theory is a helpful framework for combining beliefs from different sources using various fusion operators. Dempster's rule of combination (Dempster, 1968) is a widely recognized fusion operator that combines belief constraints from different independent belief sources. In merging clues or personal preferences, one can utilize probability masses to gauge the level of disagreement between different sources of belief. Additionally, various fusion operators can be employed to replicate other situations, such as the cumulative fusion of beliefs from independent sources through the use of the cumulative fusion operator.

Dempster's rule of combination is sometimes seen as an approximate generalization of Bayes' rule. While traditional Bayesian methods require specifying priors and conditionals, Dempster-Shafer theory allows for the expression of ignorance in situations where prior probabilities cannot be assigned (Denœux, Dubois, and Prade, 2020). This distinction between risk and ignorance has been a topic of extensive debate among statisticians and economists, with contrasting views from notable figures such as Kenneth Arrow, and Frank Knight (Knight, 1921; Arrow, 1971).

2.4.2 Formal definition

Let define a *frame of discernment* Ω, which is the universe set representing all possible outcomes of a system, the question under consideration. Then, the power set of Ω is the set of all subsets of Ω. For example, reconsider the example of Schrodinger's cat:

$$\Omega_{cat} = \{"Alive", "Dead"\}$$

$$2^{\Omega_{cat}} = \{\emptyset, \{"Alive"\}, \{"Dead"\}, \{"Alive", "Dead"\}\}$$

The power set $2^{\Omega_{cat}}$ in the theory of evidence denotes actual states of the problem where the proposition is true. The theory of evidence assigns a belief mass to each element in the power set through a mass function.

$$m : 2^{\Omega_{cat}} \rightarrow [0,1], m(\emptyset) = 0$$

The mass of a given member of a power set, represented as $m(A)$ where A belongs to $2^{\Omega_{cat}}$, expresses the amount of relevant and available evidence supporting the claim that the actual state belongs to A but not to any particular subset of A. The value of $m(A)$ is exclusive to the set A and does not make any additional claims about its subsets, each of which has its own mass. Based on these mass assignments, belief (or support) and plausibility are two non-additive continuous measures that can define the upper and lower bounds of a probability interval. This interval contains the precise probability of a set of interests in the classical sense.

$$bel(A) \leq P(A) \leq \mathrm{pl}(A)$$

, where $bel(A) = \sum_{B|B\subseteq A} m(B)$ expresses the belief for a set A and is defined as the sum of all the masses of subsets of the set of interest, meanwhile, the plausibility $pl(A) = \sum_{B|B\cap A\neq\emptyset} m(B)$ is the sum of all the masses of the sets B that intersect the set of interest A. The two measures are related to each other as follows: $pl(A) = 1 - \mathrm{bel}(\overline{A})$.

In specific scenarios, merging two sets of probability mass assignments becomes imperative. When multiple sources convey their beliefs through belief constraints such as hints or preferences, we utilize Dempster's rule of combination (Dempster, 1967; Dempster, 1968) to serve as the suitable fusion operator. Using a normalization factor, this rule extracts shared beliefs from multiple sources and disregards any conflicting (non-shared) beliefs. However, employing this rule in situations other than combining belief constraints has been criticized, such as fusing different belief estimates from multiple sources that must be accumulated cumulatively rather than as constraints. In cumulative fusion, all probability masses from various sources are incorporated in the resulting belief, ensuring no probability mass is neglected.

Specifically, the combination (called the joint mass) is calculated from the two sets of masses m_1 and m_2 in the following manner:

$$m_{1,2}(\emptyset) = 0$$

$$m_{1,2}(A) = (m_1 \oplus m_2)(A) = \frac{1}{1-K} \sum_{B\cap C=A\neq\emptyset} m_1(B)m_2(C)$$

, where $K = \sum_{B\cap C=\emptyset} m_1(B)m_2(C)$ is a measure of the amount of conflict between the two mass sets. The normalization factor, 1 - K, effectively disregards any conflict by allocating the mass associated with the conflict to the empty set. Consequently, the combination rule for evidence may lead to counterintuitive outcomes, as will be demonstrated subsequently.

The following example illustrates how Dempster's rule can yield intuitive results when used for preference fusion, even in situations with high conflict. Imagine two friends, Alice and Bob, planning to watch a movie at the cinema one evening, with only three options available: X, Y, and Z. Alice has a 0.99 probability preference for film X and a 0.01 probability preference for film Y. Bob has a 0.99 probability preference for film Z and a 0.01 probability preference for film Y. When combining their preferences using Dempster's rule of combination, their joint preference results in a 1.0 probability for film Y, as it is the only film they both agree on.

Dempster's rule can still yield sensible outcomes despite conflicting beliefs. Consider the scenario where Alice is fully certain of film X and Bob is certain of film Z.

When applying Dempster's rule to combine their preferences, the result is undefined, indicating an absence of any resolution. Consequently, they opt not to go to the cinema that evening. However, the meaning of preference as probability is ambiguous: if it refers to the likelihood of watching film X tonight, it assumes the fallacy of the excluded middle where the probability of not watching any films is 0.

In 1979, Lotfi Zadeh presented an example that highlights the unexpected consequences of Dempster's rule when there is a high level of disagreement (Zadeh, 1984). Let's imagine that two doctors, who are equally reliable, are examining a patient. The first doctor believes that the patient has a brain tumor with a probability (basic belief assignment, or mass of belief) of 0.99 or meningitis with a probability of 0.01. The second doctor believes that the patient has a concussion with a probability of 0.99 and meningitis with a probability of 0.01. When we apply Dempster's rule to combine these two sets of belief masses, the result is $m(meningitis) = 1$, which implies that meningitis is diagnosed with 100% confidence. This outcome is contrary to common sense since both doctors agree that there is only a very small chance that the patient has meningitis. This example has motivated many research studies that aim to either establish a solid justification for Dempster's rule and the foundations of the Dempster-Shafer theory or to expose the inconsistencies of the theory.

In the following scenario, we can see how Dempster's rule can lead to unexpected outcomes despite a low level of disagreement. To illustrate, let's imagine two doctors with differing opinions on a patient's condition. One doctor believes the patient has a 99% chance of having a brain tumor and only a 1% chance of having meningitis. The other doctor thinks the patient has a 99% chance of having a brain tumor and a 1% chance of having a concussion. When we apply Dempster's rule to determine the likelihood of a brain tumor, the resulting calculation is as follows:

$$m(\text{brain tumor}) = \text{Bel}(\text{brain tumor}) = 1$$

Based on the results, there is strong evidence that supports the likelihood of a brain tumor diagnosis, which both doctors have agreed is highly probable. This agreement is due to the similarity of the evidence presented by both doctors. Given this information, it is reasonable to assume that:

$$m(\text{brain tumor}) < 1 \text{ and } \text{Bel}(\text{brain tumor}) < 1,$$

since the existence of non-zero belief probabilities for alternative diagnoses indicates a lack of endorsement for the brain tumor diagnosis.

Chapter 3

Quantifying material similarity with uncertainty with respect to binary properties

3.1 Introduction

In the past decade, research using data-driven approaches in materials science has grown significantly, with each study employing a variation of the materials informatics workflow. Data-driven methods have gained recognition as powerful tools for accelerating material discovery by reducing the number of candidates to investigate. However, these methods face three critical challenges in the materials research field:

1. **Quality of material data:** The data quality often limits the effectiveness of data-dependent methods. Experimental or computational datasets for materials might need essential information, such as missing or heavily biased data since researchers tend to report successful materials while overlooking unsuccessful attempts. For example, suppose a dataset only contains high-performing materials for a specific purpose without data on low-performing or failed materials. In that case, the machine learning (ML) model may not accurately capture the complete range of material properties and could overestimate new materials' performance.

2. **Inconsistency of material datasets:** Integrating non-comparable datasets derived from various calculation methods or experiments presents a significant challenge for data-driven approaches. Inconsistencies may arise between two datasets of the same material property, collected using different experimental setups or computational methods, leading to discrepancies in reported values. This inconsistency makes it difficult for ML models to learn a reliable relationship between material descriptors and properties.

3. **Difficulty designing suitable material descriptors:** Data-driven approaches often rely on descriptors derived from condensed matter theory and parameters fitted from experimental or computational data to predict potential materials. Descriptors calculated from constituent atomic properties are commonly used, but assessing similarities between materials with varying numbers of compositions remains challenging. For instance, a descriptor that accurately captures the electronic properties of binary alloys may not be suitable for ternary or more complex alloys, leading to practical limitations in predicting material properties.

Considering the challenges associated with material data, material scientists must make decisions under uncertainty, which includes both aleatoric and epistemic uncertainties. Therefore, developing data-driven methods that can effectively distinguish and accurately measure these uncertainties is crucial. To solve these issues and pay attention to predicting whether particular combinations of elements exhibit properties of interest, we adopted the Dempster–Shafer theory (Dempster, 1968; Shafer, 1976; Denœux, Dubois, and Prade, 2020), referred to as the evidence theory, to develop a descriptor-free recommender system, called evidence-based recommender system (ERS), for exploring potential advanced material. The proposed system measures the similarity between materials with respect to the correlation of their compositions and binary target property. It then utilizes the similarity to decide on selecting the following candidates for evaluation.

Details of the proposed system are introduced in section 3.2. The applications of the proposed system to explore high-entropy alloys (HEA) and solvent mixtures that exfoliate graphite to produce graphene are shown in sections 3.3 and 3.4, respectively. Note that the HEA and solvent mixture are represented by a set of items (elements or solvents).

3.2 Evidence-based recommender system (ERS)

In high-entropy alloys and solvent mixtures, the combinations are represented similarly, with alloys being combinations of elements and solvent mixtures being solvents. Generally, each material A in dataset $\mathcal{D}$ is represented by a set of components. The binary property of interest y_A for the material A can be either Y or $\neg Y$ (not Y).

3.2.1 Methodology

Our proposed recommendation system is based on the widely used elemental substitution method, which is used to synthesize various materials. This method involves replacing an element or group of elements with a similar chemical counterpart in order to maintain the properties of the target material. However, determining the similarity between the chemical functions of alloy metal combinations or solvent combinations to find potential HEAs can be challenging. To overcome this challenge, we examine each observed combination to compare the contribution of its constituent elements to the target property of interest (i.e. forming HEA phase or exfoliating graphite). This evidence is then used to generate hypotheses about the properties of candidates. Our ERS involves three main steps:

1. **Measure the similarity between element combinations:** All the pieces of evidence obtained from the data are modeled and combined to conclude the similarity between the element combinations by using evidence theory.
2. **Evaluate the hypothesis on the properties of the substituted alloys:** The pieces of evidence for the substituted alloys are modeled and combined to evaluate the hypothesis about the target property by using evidence theory.
3. **Rank substituted alloys:** The candidates are ranked according to various criteria based on the combined evidence of their target properties.

Each combination of elements A in data set $\mathcal{D}$ is represented by a set of its components. The property of interest y_A for the combinations A (alloy), which can be either Y or $\neg Y$ (not Y). We first assess the similarity between element combinations

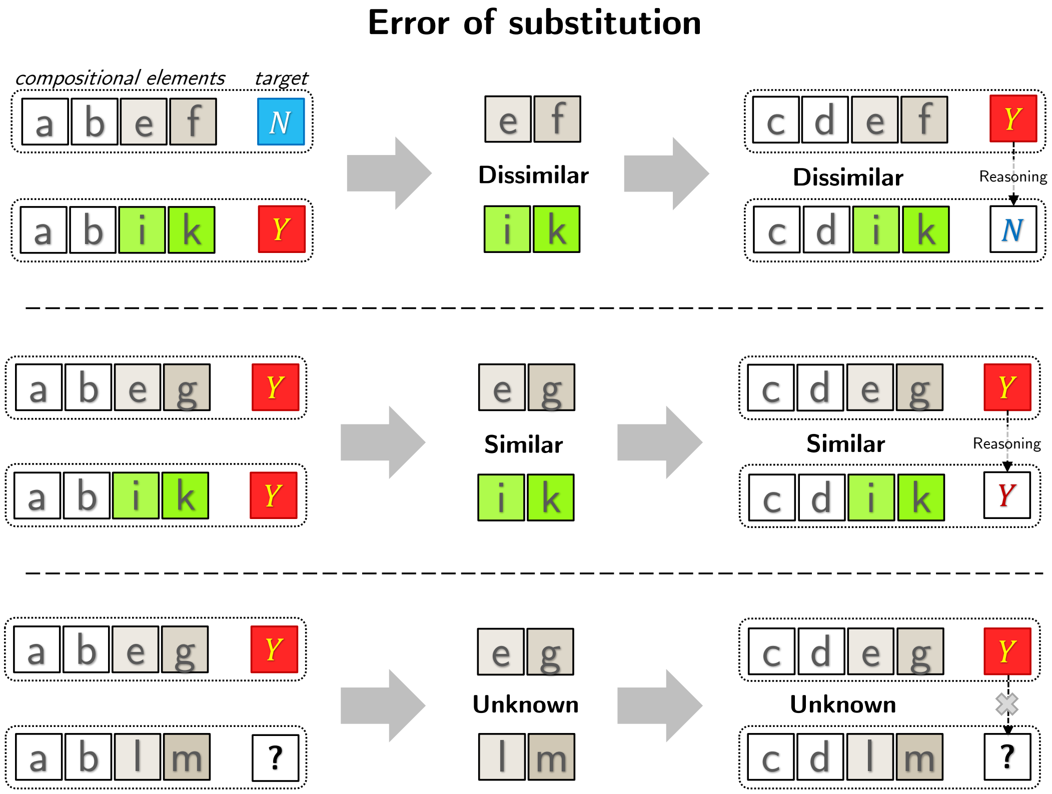


FIGURE 3.1: Illustration of the similarity between materials with respect to the correlation of compositions and binary properties of interest.

by applying the evidence theory to model and combine all pieces of evidence collected form the dataset $\mathcal{D}$.

Various forms show the similarity between objects, as stated by Tversky (Tversky, 1977). These include rating pairs, object sorting, associations, substitutability, and correlation between occurrences. In this discussion, we explore the solid-solution formability for element combinations and how substitutability measures their similarity (Fig. 3.1). Each nondisjoint pair of materials $A_i = \{a, b, e, f\}$ and $A_j = \{a, b, i, k\}$ in $\mathcal{D}$ is a source of evidence for measuring the substitutability between two element combinations $C_t = A_i - (A_i \cap A_j) = A_i - A_j = \{e, f\}$ and $C_v = A_j - (A_i \cap A_j) = A_j - A_i = \{i, k\}$. The nonempty intersection set $A_i \cap A_j = \{a, b\}$ is considered as the context for the similarity measurement. If $y_{A_i} = y_{A_j}$ then C_t and C_v are substitutable, otherwise C_t and C_v are not substitutable.

To model evidence about the similarity between any pair of element combinations, we first define a frame of discernment (Shafer, 1976) $\Omega_{sim} = \{similar, dissimilar\}$ containing all possible values. The evidence collected from alloys A_i and A_j is then represented by a mass function (Shafer, 1976) (or a basic probability assignment), $m^{C_t,C_v}_{A_i,A_j}$, which assigns probability masses to all the nonempty subsets of Ω_{sim} (i.e., $\{similar\}$, $\{dissimilar\}$, and $\{similar, dissimilar\}$), as follows:

$$m^{C_t,C_v}_{A_i,A_j}(\{similar\}) = \begin{cases} \alpha & \text{if } y_{A_i} = y_{A_j} \\ 0 & \text{otherwise} \end{cases}, \tag{3.1}$$

$$m^{C_t,C_v}_{A_i,A_j}(\{dissimilar\}) = \begin{cases} \alpha & \text{if } y_{A_i} \neq y_{A_j} \\ 0 & \text{otherwise} \end{cases}, \tag{3.2}$$

$$m^{C_t,C_v}_{A_i,A_j}(\{similar, dissimilar\}) = 1 - \alpha \tag{3.3}$$

Note that the masses assigned to $\{similar\}$ and $\{dissimilar\}$ indicate the degrees of belief exactly committed to A_i and A_j to support the similarity and dissimilarity between C_t and C_v, respectively. The weight assigned to subset $\{similar, dissimilar\}$ expresses the degree of belief that A_i and A_j provide no information about the similarity (or dissimilarity) between C_t and C_v. Here, the parameter α is determined by an exhaustive search ($0<\alpha<1$) for the best cross-validation score (Section 3.2.3). We retain some degree of uncertainty $(1 - \alpha)$ about the similarities collected from each piece of evidence for dealing with the inconsistencies in the data set. The sum of the masses assigned to all three nonempty subsets of Ω_{sim} is 1.

Suppose that we can collect multiple pieces of evidence from $\mathcal{D}$ to compare two element combinations C_t and C_v, all obtained mass functions corresponding to those pieces of evidence are then combined using the Dempster rule of combinations (Dempster, 1968) to assign the final mass $m^{C_t,C_v}_{\mathcal{D}}$ (Section 3.2.2). Similar analyses are performed for all pairs of element combinations of interest to obtain a symmetric matrix M consisting of all the similarities between them ($M[t,v] = M[v,t] = m^{C_t,C_v}_{\mathcal{D}}(\{similar\})$).

For hypothesizing whether a potential combinations of elements (or solvents) A_{new} forms an HEA phase (or exfoliate graphite), we apply the substitution method using the obtained matrix M. We replace a combination of elements, C_t, in an existing alloy, A_k, ($C_t \subset A_k$) with a combination of elements, C_v, adequate to obtain alloy A_{new} showing a property (label $y_{A_{new}}$) similar to that of A_k (label y_{A_k}). On the basis of the label of A_k and the similarity between C_t and C_v, the basic beliefs on the label of A_{new} are quantified (Figure 1 b). If C_t and C_v are substitutable (non-substitutable), this serves as a piece of evidence that the labels of A_{new} and A_k are the same (different).

To model evidence about existence of HEA phase in a particular alloy, we first define a frame of discernment (Shafer, 1976) $\Omega_Y = \{Y, \neg Y\}$.The evidence collected from A_k, C_t, and C_v is then represented by the mass function $m^{A_{new}}_{A_k,C_t\leftarrow C_v}$, which assigns probability masses to all the nonempty subsets of Ω_Y (i.e., $\{Y\}$, $\{\neg Y\}$, and $\{Y, \neg Y\}$), as follows:

$$m^{A_{new}}_{A_k,C_t\leftarrow C_v}(\{Y\}) = \begin{cases} M[t,v] & \text{if } y_{A_k} = Y \\ 0 & \text{otherwise} \end{cases}, \tag{3.4}$$

$$m^{A_{new}}_{A_k,C_t\leftarrow C_v}(\{\neg Y\}) = \begin{cases} M[t,v] & \text{if } y_{A_k} = \neg Y \\ 0 & \text{otherwise} \end{cases}, \tag{3.5}$$

$$m^{A_{new}}_{A_k,C_t\leftarrow C_v}(\{Y, \neg Y\}) = 1 - M[t,v], \tag{3.6}$$

Note that the masses assigned to $\{Y\}$ and $\{\neg Y\}$ reflect the levels of confidence whereby A_k and the substitution of C_v for C_t support the probabilities that A_{new} is or is not an HEA, respectively. The mass assigned to subset $\{Y, \neg Y\}$, expresses the probability that A_k, C_t, and C_v provide no information about the property of A_{new}. The sum of the probability masses assigned to all three nonempty subsets of Ω_Y is 1.

We assume that for a specific hypothetical alloy, A_{new}, we can collect pieces of evidence about its properties from $\mathcal{D}$ (pair of A_k and the corresponding substitution to obtain A_{new} from A_k). The obtained mass functions for A_{new} are then combined using the Dempster rule (Dempster, 1968) to obtain a final mass function $m^{A_{new}}$ (Section 3.2.2). Similar analyses are performed for all the possible alloys (A_{new}) that are not included in the observed data. We then use the final value of $m_{\mathcal{D}}^{A_{new}}(\{Y\})$ for sorting the ranking of recommendation. The proposed recommender system considers the alloys with a higher value of $m_{\mathcal{D}}^{A_{new}}(\{Y\})$ to have the greater potential of having property of interest Y.

3.2.2 Combining multiple pieces of evidence

We assume that we can collect q pieces of evidence from $\mathcal{D}$ to compare a specific pair of element combinations, C_t and C_v. If no evidence is found, the mass function $m_{none}^{C_t,C_v}$ is initialized, which assigns a probability mass of 1 to subset $\{similar, dissimilar\}$. $m_{none}^{C_t,C_v}$ models the condition under which no information about the similarity (or dissimilarity) between C_t and C_v is available. Any two pieces of evidence a and b modeled by the corresponding mass functions $m_a^{C_t,C_v}$ and $m_b^{C_t,C_v}$ can be combined using the Dempster rule (Dempster, 1968) to assign the joint mass $m_{a,b}^{C_t,C_v}$ to each subset ω of Ω_{sim} (i.e. $\{similar\}$, $\{dissimilar\}$, or $\{similar, dissimilar\}$) as follows:

$$\begin{aligned} m_{a,b}^{C_t,C_v}(\omega) &= \left(m_a^{C_t,C_v} \oplus m_b^{C_t,C_v}\right)(\omega) \\ &= \frac{\sum\limits_{\forall \omega_k \cap \omega_h = \omega} m_a^{C_t,C_v}(\omega_k) \times m_b^{C_t,C_v}(\omega_h)}{1 - \sum\limits_{\forall \omega_k \cap \omega_h = \emptyset} m_a^{C_t,C_v}(\omega_k) \times m_b^{C_t,C_v}(\omega_h)}, \end{aligned} \tag{3.7}$$

where ω, ω_k and ω_h are subsets of Ω_{sim}. Note that the Dempster rule is commutative and yields the same result by changing the order of $m_a^{C_t,C_v}$ and $m_b^{C_t,C_v}$. All the q obtained mass functions corresponding to the q collected pieces of evidence from $\mathcal{D}$ are then combined using the Dempster rule to assign the final mass $m_{\mathcal{D}}^{C_t,C_v}$ as follows:

$$m_{\mathcal{D}}^{C_t,C_v}(\omega) = \left(m_1^{C_t,C_v} \oplus m_2^{C_t,C_v} \oplus \cdots \oplus m_q^{C_t,C_v}\right)(\omega). \tag{3.8}$$

Multiple pieces of evidence about the label of each new alloy are combined using the similar manner. We assume that for a specific hypothetical alloy, A_{new}, we can collect pieces of evidence about its properties from $\mathcal{D}$ (pair of A_k and the corresponding substitution to obtain A_{new} from A_k). If no evidence is found, $m_{none}^{A_{new}}$ is initialized and a probability mass of 1 is applied to set $\{Y, \neg Y\}$. $m_{none}^{A_{new}}$ models the condition that no information about the label of A_{new} can be obtained from $\mathcal{D}$. The obtained mass functions for A_{new} are then combined using the Dempster rule (Dempster, 1968) to obtain a final mass function $m_{\mathcal{D}}^{A_{new}}$ on Ω_Y.

3.2.3 Tuning hyper-parameter of the ERS

Because data sets used in this work are the output of calculation prediction methods, we add some degree of uncertainty α in the mass function which models similarity evidence. In each data set, we use grid search to determine the α that best reproduced the alloy labels in the data set (achieving best cross-validation score). The search space of α is from 0.01 to 0.9 with a step of 0.01. However, the relative magnitudes of (degree of belief HEA) and (degree of belief not HEA) are almost unchanged. In summary, the absolute value of alpha has little effect on the final result of the recommender system.

3.2.4 Illustrative examples

The following examples provide explanations of how the evidence theory work to learn the similarity and infer the HEA formation for new element combinations, identifying equiatomic alloys.

Example 1: Suppose we have collected four pairs of alloys from experiments. Three of those pairs are alloys that both form HEA phase: $pair_1 = (\{A^1, B^1, C^1, D\}, \{A^1, B^1, C^1, E\})$; $pair_2 = (\{A^2, B^2, C^2, D\}, \{A^2, B^2, C^2, E\})$; and $pair_3 = (\{A^3, B^3, C^3, D\}, \{A^3, B^3, C^3, E\})$. The fourth pair $pair_4 = (\{A^4, B^4, C^4, D\}, \{A^4, B^4, C^4, E\})$ is different from the other three, in which $\{A^4, B^4, C^4, D\}$ forms HEA phase while $\{A^4, B^4, C^4, E\}$ does not form HEA phase. We consider each pair as a source of evidence support that $\{D\}$ is similar to $\{E\}$ in term of substitutability to form the HEA phase. Each evidence is modeled using mass function as follows:

$$\begin{aligned}
&m^{\{C\},\{D\}}_{pair_1}(\{similar\}) = 0.1,\\
&m^{\{C\},\{D\}}_{pair_1}(\{dissimilar\}) = 0,\\
&m^{\{C\},\{D\}}_{pair_1}(\{similar, dissimilar\}) = 0.9
\end{aligned}$$

$$\begin{aligned}
&m^{\{C\},\{D\}}_{pair_2}(\{similar\}) = 0.1,\\
&m^{\{C\},\{D\}}_{pair_2}(\{dissimilar\}) = 0,\\
&m^{\{C\},\{D\}}_{pair_2}(\{similar, dissimilar\}) = 0.9
\end{aligned}$$

$$\begin{aligned}
&m^{\{C\},\{D\}}_{pair_3}(\{similar\}) = 0.1,\\
&m^{\{C\},\{D\}}_{pair_3}(\{dissimilar\}) = 0,\\
&m^{\{C\},\{D\}}_{pair_3}(\{similar, dissimilar\}) = 0.9
\end{aligned}$$

$$\begin{aligned}
&m^{\{C\},\{D\}}_{pair_4}(\{similar\}) = 0,\\
&m^{\{C\},\{D\}}_{pair_4}(\{dissimilar\}) = 0.1,\\
&m^{\{C\},\{D\}}_{pair_4}(\{similar, dissimilar\}) = 0.9
\end{aligned}$$

The three pieces of evidence are combined using the Dempster′ rule of combination to accumulate the believe that $\{D\}$ is similar to $\{E\}$:

$$\begin{aligned}
&m^{\{C\},\{D\}}(\{similar\}) = 0.25,\\
&m^{\{C\},\{D\}}(\{dissimilar\}) = 0.075,\\
&m^{\{C\},\{D\}}(\{similar, dissimilar\}) = 0.675
\end{aligned}$$

Next, if we observed (included in the data) that the HEA phase exists for alloy $\{G, H, I, D\}$, the ERS (which focuses on finding some chance for discovering new combination of elements that the HEA phase exist and ignores the belief regarding $\neg HEA$) will consider that there is some believe that the HEA phase also exists for $\{G, H, I, E\}$ (by substituting $\{D\}$ with $\{E\}$). The evidence is modeled using mass function as follows:

$$
\begin{aligned}
&m^{\{G,H,I,E\}}_{\{G,H,I,D\},\{D\}\leftarrow\{E\}}(\{\neg HEA\}) = 0,\\
&m^{\{G,H,I,E\}}_{\{G,H,I,D\},\{D\}\leftarrow\{E\}}(\{HEA\}) = m^{C,D}(\{similar\}) = 0.25,\\
&m^{\{G,H,I,E\}}_{\{G,H,I,D\},\{D\}\leftarrow\{E\}}(\{HEA, \neg HEA\}) = 1 - m^{C,D}(\{similar\}) = 0.75
\end{aligned}
$$

Example 2: In a same manner but for an extrapolative recommendation: if the HEA phases exist for all the alloys in the three following pairs: $pair_1 = (\{A^1, B^1, C\}, \{A^1, B^1, D, E\})$, $pair_2 = (\{A^2, B^2, C\}, \{A^2, B^2, D, E\})$, $pair_3 = (\{A^3, B^3, C\}, \{A^3, B^3, D, E\})$. In the fourth pair $pair_4 = (\{A^4, B^4, C\}, \{A^4, B^4, D, E\})$, $\{A^4, B^4, C\}$ forms HEA phase while $\{A^4, B^4, D, E\}$ does not form HEA phase. The algorithm will accumulate the believe that $\{C\}$ is similar to $\{D, E\}$ as follows:

$$
\begin{aligned}
&m^{\{C\},\{D,E\}}(\{similar\}) = 0.25,\\
&m^{\{C\},\{D,E\}}(\{dissimilar\}) = 0.075,\\
&m^{\{C\},\{D,E\}}(\{similar, dissimilar\}) = 0.675
\end{aligned}
$$

Consequently, if we observed (included in the data) that the HEA phase exists for $\{G, H, I, C\}$, the algorithm (which focuses on finding some chance for discovering new combination of elements that the HEA phase exist and ignores the belief regarding $\neg HEA$) will consider that there is some believe that the HEA phase also exists for $\{G, H, I, D, E\}$ (by substituting $\{C\}$ with $\{D, E\}$).

$$
\begin{aligned}
&m^{\{G,H,I,D,E\}}_{\{G,H,I,C\},\{C\}\leftarrow\{D,E\}}(\{\neg HEA\}) = 0,\\
&m^{\{G,H,I,D,E\}}_{\{G,H,I,C\},\{C\}\leftarrow\{D,E\}}(\{HEA\}) = m^{C,D}(\{similar\}) = 0.25,\\
&m^{\{G,H,I,D,E\}}_{\{G,H,I,C\},\{C\}\leftarrow\{D,E\}}(\{HEA, \neg HEA\}) = 1 - m^{C,D}(\{similar\}) = 0.75
\end{aligned}
$$

3.3 Case study 1: Exploring high-entropy alloys

Multi-principle element alloys (MPEAs), also known as high-entropy alloys (HEAs) when comprising five or more elements, represent a novel development concept of alloy Yeh et al., 2004; Cantor et al., 2004; Senkov et al., 2015. These alloys consist of multiple elements forming highly disordered solid-solution phases. Since their discovery, MPEAs, and HEAs have garnered significant interest in the scientific community due to their promising properties and potential applications Rickman et al., 2019; Tsai and Yeh, 2014. These alloys exhibit exceptional strength-to-weight ratios, tensile strengths, and corrosion and oxidation resistance. For the sake of consistency with the published data utilized in this study, the term "HEA" refers to random alloys comprising equiatomically combined elements that form a solid-solution phase.

Approaches to recommend specific element combinations likely to form single-phase HEAs vary between deductive and inductive methods based on distinct concepts. In the deductive approach, understanding the mechanisms governing HEA formation or starting with quantum-mechanical equations derived from extensive first-principles calculations is necessary. Previous HEA research hypothesized that configurational entropy-induced stabilization results in the formation of single-phase solid solutions. However, this hypothesis holds only for certain multicomponent alloys, as most have experimentally demonstrated the formation of multiple phases (GUO and LIU, 2011). While significant attention has been devoted to understanding the mechanisms driving HEA stability, the key factors governing the formation of single-phase HEAs remain unknown (Zhang et al., 2016). Using first-principles calculations to construct phase diagrams for multicomponent alloys can predict which alloys will form solid solutions. Still, the computational demands of this method, which involves energy calculations for numerous configurations and the implementation of statistical mechanical models, are significant (Huhn and Widom, 2013). Therefore, exploring HEAs through first-principles calculations becomes crucial.

Inductive screening methods have been developed, employing descriptors derived from condensed matter theory and fitted parameters based on available experimental data to predict possible HEAs or their structural phases (Zhang et al., 2008; Ye et al., 2015; Tsai, 2016; Tsai et al., 2019; Huang, Martin, and Zhuang, 2019). However, the inductive approach relies on sufficient and balanced data to ensure prediction accuracy, which is often challenging due to limited or biased experimental data skewed towards positive results (George, Raabe, and Ritchie, 2019; Konno et al., 2021). Additionally, quantitatively evaluating prediction uncertainty remains an ongoing challenge, even in cases of high prediction accuracy. Designing appropriate material descriptors to represent alloys with varying numbers of elements poses another obstacle. Descriptors calculated from atomic properties of constituent elements, such as mean, variance, and differences in atomic sizes, are commonly used (Pham et al., 2016a; Kobayashi et al., 2017; Tamura et al., 2017; Seko et al., 2017; Huang, Martin, and Zhuang, 2019; Kobayashi, 2021). However, accurately assessing similarity or dissimilarity between alloys with different compositions is mathematically complex, leading to limitations in the results obtained through data-driven approaches using these descriptors (Pham et al., 2016a; Nguyen et al., 2018a). One potential solution is describing the alloy using one-hot vectors of constituent elements. However, this approach presents its own challenge of designing an appropriate metric in this vector space (Pham et al., 2017).

The Dempster-Shafer theory is a superior approach to the Bayesian method for dealing with incomplete information and insufficient data. This approach is particularly beneficial for solving material data problems. Evidence theory assigns non-negative weights to subsets of possibilities, known as the "frame of discernment," instead of individual elements like the Bayesian approach. This allows for the modeling, collection, and combination of evidence from multiple alloy data without the need for material descriptors. Consequently, this system can confidently suggest advanced materials by learning from multiple observed materials with fewer constituent elements.

3.3.1 Alloys data sets

In our evaluation tests, we use eight data sets consisting of binary, ternary, quaternary, and quinary alloys comprising multiple equiatomically combined elements. The datasets include data from experiments and calculations. The alloys contained in the data sets comprise $\mathcal{E} = \{$ Fe, Co, Ir, Cu, Ni, Pt, Pd, Rh, Au, Ag, Ru, Os, Si, As, Al, Tc, Re, Mn, Ta, Ti, W, Mo, Cr, V, Hf, Nb, and Zr$\}$. Figure 3.2 shows the proportion of 27 elements in the data sets. Any alloy included in the following data sets is considered as HEA if its disordered transition temperature is lower than its melting point.

- $\mathcal{D}_{\text{ASMI16}}$: The order-disorder transition temperatures (T_c^{exp}) and melting temperatures (T_m^{exp}) of the alloys are both experimentally evaluated (Okamoto, Schlesinger, and Mueller, 2016). All of the alloys contained in $\mathcal{D}_{\text{ASMI16}}$ show an

TABLE 3.1: Summary of the eight alloy data sets used in evaluation experiments. No. alloys: number of alloys included in each data set; No. HEAs: number of the alloys confirmed or estimated to form HEA phase in each data set; No. candidates: number of possible alloys generated using the set of all elements in the data sets. The "HEA rate" is the ratio of No. HEA to No. alloys, whereas the "Observation rate" is the ratio of No. alloys to No. candidates.

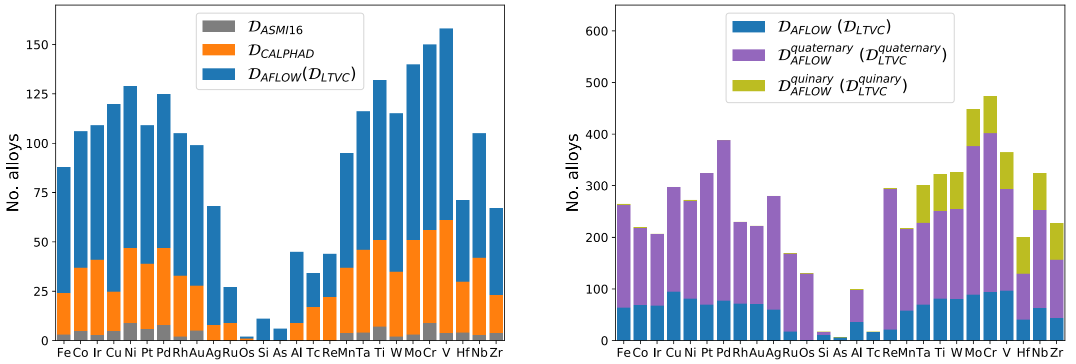
Data set	No. alloys	No. HEAs (HEAs rate)	No. candidates (Observation rate)
$\mathcal{D}_{\text{ASMI16}}$ [1]	45 binary alloys	45 (100%)	351 (13%)
$\mathcal{D}_{\text{CALPHAD}}$ [2]	243 ternary alloys	243 (100%)	2925 (9%)
$\mathcal{D}_{\text{AFLOW}}$ [3]	117 binary alloys	60 (51%)	351 (33%)
	441 ternary alloys	234 (53%)	2925 (15%)
$\mathcal{D}_{\text{LTVC}}$ [4]	117 binary alloys	58 (49%)	351 (33%)
	441 ternary alloys	148 (33%)	2925 (15%)
$\mathcal{D}_{\text{AFLOW}}^{\text{quaternary}}$ [3]	1,110 quaternary alloys	754 (68%)	17,550 (6%)
$\mathcal{D}_{\text{LTVC}}^{\text{quaternary}}$ [4]	1,110 quaternary alloys	480 (43%)	17,550 (6%)
$\mathcal{D}_{\text{AFLOW}}^{\text{quinary}}$ [3]	130 quinary alloys	129 (99%)	80,730 (0.16%)
$\mathcal{D}_{\text{LTVC}}^{\text{quinary}}$ [4]	130 quinary alloys	91 (70%)	80,730 (0.16%)

[1] Okamoto, Schlesinger, and Mueller, 2016; [2] Zhang et al., 2014; [3] Nyshadham et al., 2017; [4] Lederer et al., 2018;

order-disorder transition temperature below their melting temperature ($T_c^{\text{exp}} < T_m^{\text{exp}}$).

- $\mathcal{D}_{\text{CALPHAD}}$: The order-disorder transition temperatures ($T_c^\star$) and melting temperatures ($T_m^\star$) of the alloys are both predicted using calculated-phase-diagram (CALPHAD) calculations (Senkov et al., 2015; Alman, 2013; Zhang et al., 2014) based on the temperatures for some binary alloys (three possible for each ternary alloy) found in the Thermo-Calc software SSOL5 database Andersson et al., 2002. Similar to the $\mathcal{D}_{\text{ASMI16}}$ data set, the $\mathcal{D}_{\text{CALPHAD}}$ data set only contains the alloys satisfying $T_c^\star < T_m^\star$.

- $\mathcal{D}_{\text{AFLOW}}$, $\mathcal{D}_{\text{AFLOW}}^{\text{quaternary}}$, and $\mathcal{D}_{\text{AFLOW}}^{\text{quinary}}$: The order-disorder transition temperatures (T_c^{AFLOW}) of the alloys contained in these data sets are estimated using the automatic flow (AFLOW) convex-hull database (Nyshadham et al., 2017). The



(a) **(b)**

FIGURE 3.2: Proportions of 27 elements in $\mathcal{D}_{\text{ASMI16}}$, $\mathcal{D}_{\text{CALPHAD}}$, $\mathcal{D}_{\text{AFLOW}}$, $\mathcal{D}^{\text{quaternary}}_{\text{AFLOW}}$, $\mathcal{D}^{\text{quinary}}_{\text{AFLOW}}$, $\mathcal{D}_{\text{LTVC}}$, $\mathcal{D}^{\text{quaternary}}_{\text{LTVC}}$, and $\mathcal{D}^{\text{quinary}}_{\text{LTVC}}$ data sets.

melting temperatures T_m^{exp} and $T_m^{\star}$ are applied to the binary and ternary alloys, respectively. The alloy is considered as an HEA if $T_c^{\text{AFLOW}} < T_m^{\text{exp}}$ for binary alloys and $T_c^{\text{AFLOW}} < T_m^{\star}$ for ternary, quaternary, and quinary alloys).

- $\mathcal{D}_{\text{LTVC}}$, $\mathcal{D}^{\text{quaternary}}_{\text{LTVC}}$, and $\mathcal{D}^{\text{quinary}}_{\text{LTVC}}$: These data sets contain the same alloys as those contained in data sets $\mathcal{D}_{\text{AFLOW}}$, $\mathcal{D}^{\text{quaternary}}_{\text{AFLOW}}$, and $\mathcal{D}^{\text{quinary}}_{\text{AFLOW}}$, respectively. However, the properties of the alloys contained in these data sets are predicted using the method of Lederer, Toher, Vecchio, and Curtarolo (LTVC) (Lederer et al., 2018). *Ab-initio* calculations are used to estimate the order-disorder transition temperatures (T_c^{LTVC}) of the alloys contained in these data sets. In addition, the T_m values are the same as those of the alloys contained in the AFLOW data sets. Any alloy in these data sets is predicted as an HEA if $T_c^{\text{LTVC}} < T_m^{\text{exp}}$ for binary alloys and $T_c^{\text{LTVC}} < T_m^{\star}$ for ternary, quaternary, and quinary alloys

Note that $\mathcal{D}_{\text{ASMI16}}$ and $\mathcal{D}_{\text{CALPHAD}}$ contain only experimental confirmed and calculated HEAs, respectively. Therefore, although we assume that the properties of all other binary or ternary alloys (not included in the data set) remain unconfirmed, we do not assume that such alloys do not form HEA phase.

3.3.2 Materials descriptors

Descriptors play a crucial role in developing a recommender system for exploring potential new high-entropy alloys (HEAs). In this research, we employ various descriptors to represent the raw data of alloys, represented as combinations of elements. Several descriptors have been studied in materials informatics to represent compounds (Seko, Togo, and Tanaka, 2018). For this work, we utilize a compositional descriptor (Seko et al., 2017), a rating matrix representation (Seko et al., 2018), and binary elemental descriptors (Seko, Togo, and Tanaka, 2018).

The compositional descriptor represents an alloy using a set of 135 features, including means, standard deviations, and covariances of atomic representations that constitute the alloy. This descriptor can be applied to both crystalline and molecular systems. We adopted 15 atomic representations: (1) atomic number, (2) atomic mass,(3) period and (4) group in the periodic table, (5) first ionization energy, (6) second ionization energy, (7) Pauling electronegativity, (8) Allen electronegativity, (9) van der Waals radius, (10) covalentradius, (11) atomic radius, (12) melting point, (13) boiling point, (14) density, and (15) specific heat. However, the compositional descriptor struggles to differentiate compounds with different numbers of atoms since it treats the atomic representations as data distributions. Consequently, the compositional descriptor is not suitable for cases involving extrapolation in the number of components.

The rating matrix representation, a descriptor-free approach, has demonstrated robust performance in recommendation systems across various data sets in the machine learning community. Seko et al. employed this representation to build a recommender system for exploring chemically relevant compositions (Seko et al., 2018). In their work, a composition data set is transformed into two feature sets corresponding to users and items in a user-item rating matrix. Missing element ratings are approximated based on feature similarity provided by the representation. To build a recommender system for HEAs, we define the candidate alloys as AB, where A and B represent the elemental components. We introduce two types of matrix representations for the eight alloy data sets, decomposing an alloy into two elementary components:

- *Type 1*: $|A| \in 1, 2$ and $|B| \in 1, 2, 3$. The number of possible components for A and B are 378 and 3303, respectively. The size of the rating matrix is (378×3303).
- *Type 2*: $|A| = 1$ and $|B| \in 1, 2, 3, 4$. The number of possible components for A and B are 27 and 20853, respectively. The size of the rating matrix is (27×20853).

Binary elemental descriptors represent the presence or absence of chemical elements using binary digits. The number of binary elemental descriptors corresponds to the number of element types included in the training data. In our work, the alloy data sets comprise 27 kinds of elements; thus, an alloy is described by a 27-dimensional binary vector with elements of either one or zero.

3.3.3 Learning about the similarity between elements

By applying the proposed ERS to the $\mathcal{D}_{\text{ASMI16}}$, $\mathcal{D}_{\text{CALPHAD}}$, $\mathcal{D}_{\text{AFLOW}}$, and $\mathcal{D}_{\text{LTVC}}$ data sets (Table 3.1), we assess the similarity between the $\mathcal{E}$ elements and all the possible binary combinations obtained therein. Figure 3.3 (a, b, c, and d) show the M_{ASMI16}, M_{CALPHAD}, M_{AFLOW}, and M_{LTVC} similarity matrices obtained for all the $\mathcal{E}$ elements in the first four experiments. These similarity matrices are then properly transformed into distance matrices to which Ward's hierarchical agglomerative clustering (Murtagh and Legendre, 2014) can be applied to construct the corresponding hierarchically clustered structures of these elements (Figure 3.3 e, f, g, and h).

The similarity matrix M_{ASMI16} reveals three distinct element groups (Figure 3.3 e) consisting of Ti, V, Cr, Mn, Zr, Nb, Mo, Hf, Ta, and W; Fe, Co, Ni, Cu, Rh, Pd, Ir, Pt, and Au; and Al, Ag, Tc, Si, Ru, As, Re, and Os, where the first two groups correspond to the early and late transition metals, respectively. Given the similar physical and chemical properties of these elements, the high degree of similarity between the elements within the same group, as revealed by the ERS, is rational. Interestingly, the matrix M_{ASMI16} shows a remarkable similarity between Mn (an earlier transition metal) and Au (a late transition metal). Furthermore, the similarity matrix M_{ASMI16} indicates none of the belief about the similarity among the elements in the third group and between the elements of the third group and the other two groups because the binary alloys contained in $\mathcal{D}_{\text{ASMI16}}$ do not contain these elements

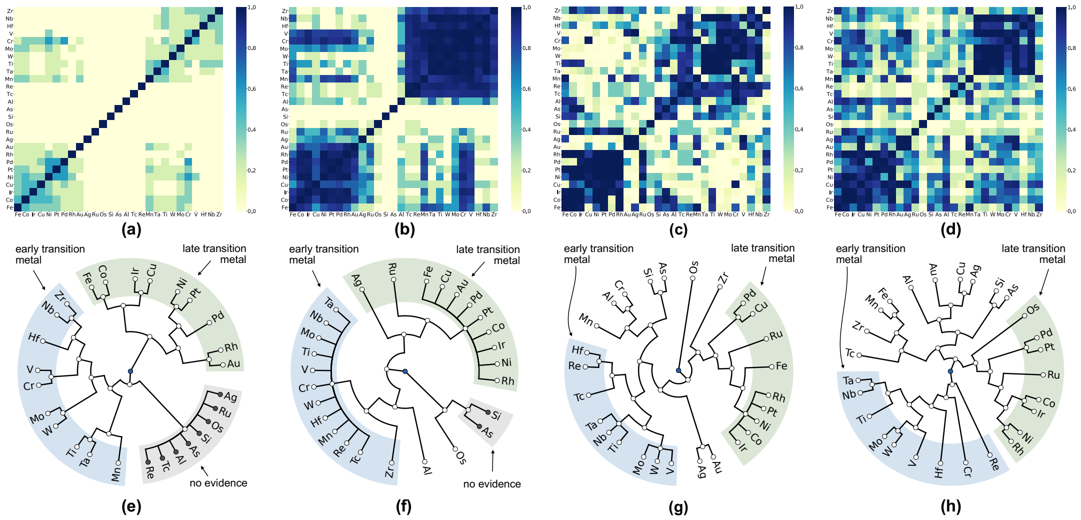


FIGURE 3.3: Visualization of similarities between elements. Top: Heat maps for similarity matrices (a) M_{ASMI16}, (b) M_{CALPHAD}, (c) M_{AFLOW}, and (d) M_{LTVC}. Each matrix element is the probability mass that the similarity mass function of the corresponding element pair is assigned to subset $\{similar\}$ of Ω_{sim}. These matrix elements indicate the degree of belief learned from the similarity data of the corresponding element pairs. In these figures, the degrees of belief are illustrated using colormap. Bottom: Hierarchically clustered structures of all elements in $\mathcal{E}$ constructed using hierarchical agglomerative clustering and these similarity matrices (e) M_{ASMI16}, (f) M_{CALPHAD}, (g) M_{AFLOW}, and (h) M_{LTVC} data sets. The blue, green, and gray regions indicate groups of early and late transition metals, and elements without similarity evidence, respectively.

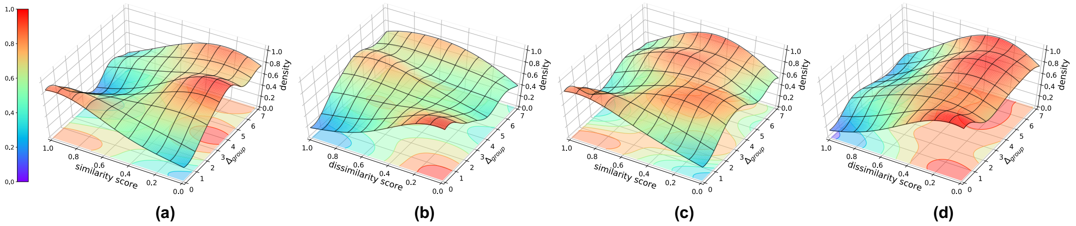
(Figure 3.2 a). Therefore, no evidence of similarities can be collected from $\mathcal{D}_{\mathrm{ASMI16}}$ for these elements.

The similarity matrix M_{CALPHAD} also reveals three somewhat modified element groups (Figure 3.3 f) compared to those obtained from $\mathcal{D}_{\mathrm{ASMI16}}$. Because $\mathcal{D}_{\mathrm{CALPHAD}}$ contains some Tc- and Re-containing alloys, these elements join the group of early transition metals. Similarly, $\mathcal{D}_{\mathrm{CALPHAD}}$ contains more Ag- and Au-containing alloys, and these elements join the group of late transition metals. Therefore, only Al, Si, As, and Os remain in the third group. Although no evidence of any similarities between Si and As can be collected from $\mathcal{D}_{\mathrm{CALPHAD}}$ (Figure 3.2 a), Os and Al are somewhat similar to the first and second groups, respectively.

In contrast, it is difficult to divide all the elements contained in $\mathcal{E}$ into groups according to the matrix M_{AFLOW}. However, some characteristic groups of metallic elements are distinct. Although two distinct groups of early or late transition metals are observed(Figure 3.3 g), there are some notable differences between these results obtained from experiments with $\mathcal{D}_{\mathrm{CALPHAD}}$ and $\mathcal{D}_{\mathrm{AFLOW}}$. The similarity matrix learned from $\mathcal{D}_{\mathrm{AFLOW}}$ shows that Au and Ag are very similar (Figure 3.3 c). Furthermore, both are similar to V, Mn, and Al but not to other late transition metals (Figure 3.3 g). Mn is also similar to Tc, Re, and Cr but not to the other early transition metals. However, Tc and Re are somewhat similar to the other early transition metals. Furthermore, Zr is somewhat similar to the late transition metals, but different from the early transition metals. Clearly, these results are different from that obtained from $\mathcal{D}_{\mathrm{CALPHAD}}$ owing to the difference between the predicted label (HEA or $\neg HEA$) for

TABLE 3.2: Comparison of the properties of alloys containing Zr in the two datasets $\mathcal{D}_{\text{ASMI16}}$ and $\mathcal{D}_{\text{CALPHAD}}$ to those predicted in $\mathcal{D}_{\text{LTVC}}$ and $\mathcal{D}_{\text{AFLOW}}$.

	$\mathcal{D}_{\text{ASMI16}}$ 4 alloys	$\mathcal{D}_{\text{CALPHAD}}$ 19 alloys
# in agreement with $\mathcal{D}_{\text{AFLOW}}$	4	10
# disagreement with $\mathcal{D}_{\text{AFLOW}}$	0	9
# in agreement with $\mathcal{D}_{\text{LTVC}}$	3	10
# disagreement with $\mathcal{D}_{\text{LTVC}}$	1	9



(a) (b) (c) (d)

FIGURE 3.4: Correlation between pairwise similarity and difference in group index (Δ_{group}) of elements. Sub-figures illustrate the distribution of pairwise similarities, which are obtained from (a, b) $\mathcal{D}_{\text{AFLOW}}$ and (c, d) $\mathcal{D}_{\text{LTVC}}$ data sets, according to the Δ_{group} of these element pair. Colormap illustrates the estimated density of the distribution of pairwise similarity.

the Zr-containing alloys recommended based on CALPHAD and AFLOW calculations, as listed in Table 3.2. Al, Si, and As are all similar to each other and to Fe and Co (Fig. 3.3 c). However, Al is similar to V, Cr, and Mn but not to Ti, whereas Si and As are very similar to Ti but not to V or Cr.

In addition, the similarity matrix M_{AFLOW} does not show any similarity between Os and any of the other elements because very few Os-containing alloys are contained in the data set (Figure 3.2 a). Furthermore, the similarity matrices M_{LTVC} and M_{AFLOW} are approximately similar. However, the hierarchically clustered structure constructed from $\mathcal{D}_{\text{LTVC}}$ indicates that Cu, Ag, and Au form a distinct subgroup (Figure 3.3 h).

Figure 3.4 shows the correlation between the pairwise similarities learned from the $\mathcal{D}_{\text{AFLOW}}$ and $\mathcal{D}_{\text{LTVC}}$ data sets and the corresponding difference between the periodic-table group index obtained for each of the transition metal pairs contained in $\mathcal{E}$. Clearly, the elements showing the same periodic-table group index ($\Delta_{group} = 0$) tend to show high similarity scores (Figure 3.4 a and c) and low dissimilarity scores (Figure 3.4 b and d). Therefore, the elements in the same group similarly contribute to HEA formation and are substitutable for each other. However, it should be noted that several pairs of elements have a similarity with a low degree of belief even though they belong to the same groups, i.e. {(Ti, Zr), (Cu, Ag), (Fe, Ru)} in $\mathcal{D}_{\text{AFLOW}}$ and {(Ti, Zr), (Mn, Re), (Ni, Pd)} in $\mathcal{D}_{\text{LTVC}}$ (Figure 3.3 c and d).

Furthermore, as the difference in the group index increases from 0 to 4, the similarity between the elements decreases ($\Delta_{group} : 0 \to 4$). The results learned from the $\mathcal{D}_{\text{AFLOW}}$ and $\mathcal{D}_{\text{LTVC}}$ data sets both show that the elements are the least similar when the difference between their group indices is three or four. However, the elements become slightly more similar as Δ_{group} increases from 5 to 7, which is consistent with

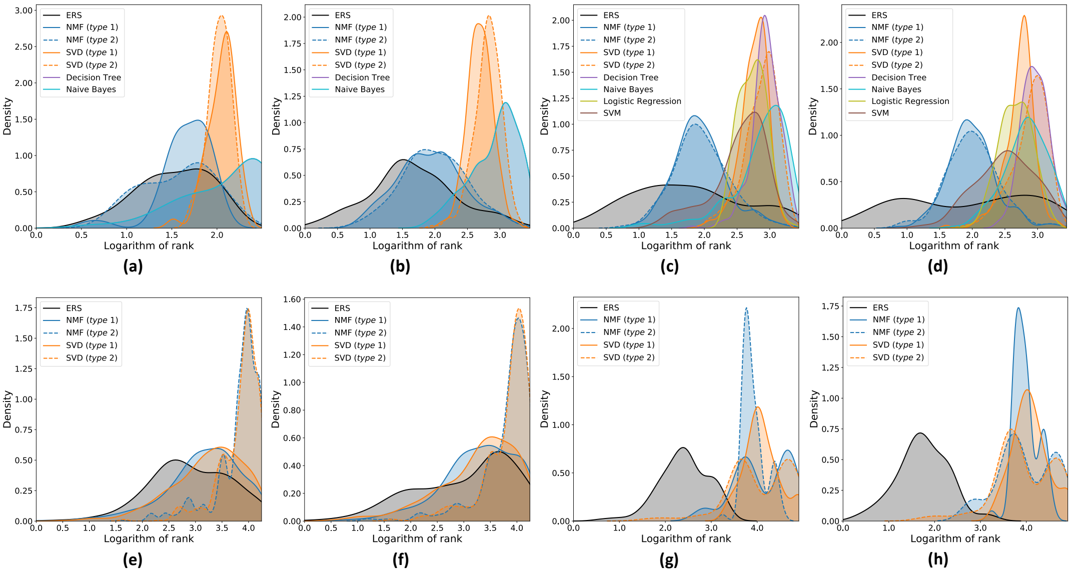


FIGURE 3.5: Evaluation of HEA-recommendation capability. Probability density functions of the rank of the HEAs in the test sets in (a) $\mathcal{D}_{\mathrm{ASMI16}}$, (b) $\mathcal{D}_{\mathrm{CALPHAD}}$, (c) $\mathcal{D}_{\mathrm{AFLOW}}$, (d) $\mathcal{D}_{\mathrm{LTVC}}$, (e) $\mathcal{D}^{\mathrm{quaternary}}_{\mathrm{AFLOW}}$, (f) $\mathcal{D}^{\mathrm{quaternary}}_{\mathrm{LTVC}}$, (g) $\mathcal{D}^{\mathrm{quinary}}_{\mathrm{AFLOW}}$, and (h) $\mathcal{D}^{\mathrm{quinary}}_{\mathrm{LTVC}}$ experiments. The ranks of HEAs in the test sets are expressed on a base-10 logarithmic scale. The HEAs with higher ranking order are recommended materials with a firmer belief in the formation of the HEA phase.

the domain knowledge about the differences between early and late transition metals.

3.3.4 Evaluation of recommendation capability by cross-validation

We apply k-fold cross-validation to the $\mathcal{D}_{\mathrm{ASMI16}}$, $\mathcal{D}_{\mathrm{CALPHAD}}$, $\mathcal{D}_{\mathrm{AFLOW}}$, and $\mathcal{D}_{\mathrm{LTVC}}$ data sets to assess the HEA-recommendation capabilities of the ERS, the four matrix-based recommender systems (NMF and SVD, each one with two types of matrix representations) (Seko et al., 2018). These two matrix representations (*type 1* and *type 2*) decompose an alloy into two elementary components A and B with different sizes (Section 3.3.2). We also compare the ERS with the four supervised-learning-method-based (i.e., decision tree, Naïve-Bayes, logistic-regression, and SVM) recommender systems.

The learned similarity matrix is used to rank all the alloys contained in the test set and all the possible combinatorial alloys other than those used to train the similarity matrix. The resulting alloy rankings are then used to evaluate the HEA-recommendation performance. We designed a virtual experiment that sequentially identifies the alloys on the basis of the order in which they were previously ranked. To evaluate the HEA-recommendation capability of the proposed ERS, we monitor the rank of HEAs in the test set and the HEA recall depending on the number of trials required to identify all possible HEAs. The detailed experimental conditions are shown in the section 3.3.7.

Figure 3.5 (a–d) illustrate the distributions of the HEA ranks of the test set recommended by the different systems. The HEAs in the test set are generally recommended with higher rank using the ERS (i.e., the ERS rank distributions are on the left of the curves for the other systems). Consequently, the ERS can significantly reduce the number of trials required to recover the HEAs in the test set compared to the competitor systems. Only in the experiment with $\mathcal{D}_{\mathrm{ASMI16}}$, the distributions of the rank using the ERS and NMF (*type 2*) are somewhat similar (Fig. 3.5 a). We also monitor the dependence of the HEA recall ratio on the number of trials required to measure the HEA-recommendation performance of the ERS quantitatively. In summary, the ERS outperforms the other systems in recalling one-half and three-quarters of the HEAs in the test set (Section A.1). However, the ERS cannot reliably recall the remaining one-quarter of the HEAs because insufficient evidence is available in the training data to make inferences about the remaining HEAs. Interestingly, in the $\mathcal{D}_{\mathrm{ASMI16}}$ and $\mathcal{D}_{\mathrm{CALPHAD}}$ experiments, the supervised-method-based recommender systems either approximately randomly selected possible HEAs (Naïve Bayes and decision tree) or could not rank any at all (logistic regression and SVM) because these data sets contain only positively labeled HEAs.

3.3.5 Evaluation of recommendation capability by extrapolation

The cross-validation experiments show the recommendation systems based on supervised learning methods, including SVMs (Hearst, 1998), logistic regression (LaValley, 2008), decision trees (Quinlan, 1986), and Naïve-Bayes (Yager, 2006) have much lower recommendation performance. These results are attributed to the inappropriate assessment of the similarity between alloys with different numbers of compositions (Section 3.3.2). Therefore, to evaluate the HEA-recommendation capability when extrapolating the number of components, we focus on comparing the performances of the ERS with those of matrix-based recommender systems. The detailed experimental settings are shown in the section 3.3.7.

Figure 3.5 (e–h) illustrate the distributions of the recommended HEA rank of the quaternary and quinary HEAs in the test set that are extrapolated using recommender systems. The obtained results show that the ERS outperforms the capability of the competitor systems for recommending quaternary HEAs (Fig. 3.5 e and f) and substantially outperforms the capability of the other systems for recommending quinary HEAs (Fig. 3.5 g and h). Interestingly, in the experiments with $\mathcal{D}_{\mathrm{LTVC}}^{\mathrm{quinary}}$ and $\mathcal{D}_{\mathrm{AFLOW}}^{\mathrm{quinary}}$, the numbers of quinary HEAs in the test set, and those found in the top 100 and top 1,000 HEA candidates recommended by the ERS, are much larger than those predicted by the competitor systems. These numbers are very high because the two data sets only contain quinary alloys of the early transition metals. Much of the evidence of the similarities between these element combinations can be collected from the corresponding data sets containing binary, ternary, and quaternary alloys (Figure 3.2 b). Moreover, to recall 50 and 75% of the quinary HEAs from these data sets, approximately 10-100 fewer trials are required by the ERS than by the NMF and SVD-based recommender systems. The results of experiments monitoring the dependence of the HEA recall ratio on the number of trials required are listed in detail in Section A.2. In the absence of sufficient evidence, the answer of the system, regarding a mixture of many types of elements, will retain a large degree of uncertainty ($m(\{HEA, \neg HEA\}) \approx 1$).

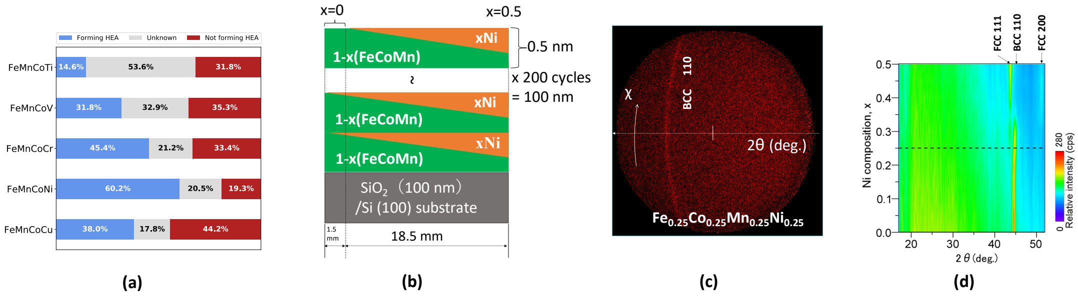


FIGURE 3.6: Recommendation and experimental validation for thin film of FeCoMnNi HEA. (a) Recommended candidates for Fe-Co-based HEAs containing first-transition-series elements: FeMnCoTi, FeMnCoV, FeMnCoCr, FeMnCoNi, and FeMnCoCu. (b) Schematic illustration of the sample, which includes 200 cycles of 0.5 nm spread film, was fabricated on SiO_2/Si (100) substrate using the combinatorial method. Each spread film consists of a 0.25 nm FeCoMn sublayer and a 0.25 nm 1-x(FeCoMn)-xNi sublayer. (c) 2D-XRD image of $Fe_{0.25}Co_{0.25}Mn_{0.25}Ni_{0.25}$ thin film measured by changing the incident angle of X-rays. (d) Heat map shows the dependence of the X-ray diffraction intensity of 1-x(FeCoMn)-xNi films on Ni composition and diffraction angle θ.

3.3.6 Synthesis of recommended FeMnCo-based HEAs

Fe-Co-based film soft-magnetic materials have attracted interest from device community and will be applied to improve the performance of next-generation high-power devices (Silveyra et al., 2018). Therefore, we focus on Fe-Co-based quaternary alloys containing the first transition-series elements. We combine all evidence collected from all the data sets to recommend quaternary Fe-Co-based HEAs for experimental validation.

Figure 3.6 a shows the recommended possible magnetic quaternary HEAs containing Fe, Mn, and Co. Clearly, FeMnCoNi is the only HEA candidate recommended with a belief higher than 0.5. Although FeMnCoCr and FeMnCoCu are HEA candidates recommended with the next highest belief, some uncertainty still remains as to their potential as HEAs. Therefore, we chose FeMnCoNi as the target HEA candidate for the experimental validation. Figure 3.6 b shows the sample structure of the FeMnCoNi. The composition film layer consists of three layers. One is a single FeCoMn layer with a thickness of 0.25 nm. The other layers are composition spread film formed by FeCoMn and Ni layers (Please refer to section B for further information).

Figure 3.6 c shows a 2D-XRD image of a region of the $Fe_{0.25}Co_{0.25}Mn_{0.25}Ni_{0.25}$ alloy annealed at 400°C. A reflection attributed to the (110) plane of the BCC crystal structure appears in the ring pattern at $2\theta = 44.7°$ (PDF 03-065-7519 Gates-Rector and Blanton, 2019). Note that out-of-plane XRD measurements were also performed to identify the crystal structure in more detail, as shown in Figure B.2 (a), indicating the formation of a polycrystalline film. Reportedly, the BCC crystal structure of the FeCoMn alloy is stable (Snow et al., 2018), and previous reports have mentioned that FeCoMnNi alloy has an face-centered cubic FCC structure in high temperature synthesized bulk; however, detailed information is still not available (Wu et al., 2014; Cui et al., 2018). Therefore, to investigate the stability of the crystal structure, the effect of Ni doping on the crystal structure was analyzed based on the heat map generated from the X-ray diffraction patterns of FeCoMn films prepared with various Ni contents (Fig 3.6 d). For an Ni content above 0.3, the FCC structure is also observed at $2\theta = 43.5^\circ$, corresponding to the (111) reflection [Figure B.2 b] (PDF 03-065-5131 Gates-Rector and Blanton, 2019). These results suggest that the $Fe_{0.25}Co_{0.25}Mn_{0.25}Ni_{0.25}$ HEA shows a BCC structure. In our experiment, the BCC structure of the starting material, FeCoMn, is considered as an essential reason for which the thin films produced by this method tend to be in the BCC phase.

3.3.7 Experimental settings for evaluation of recommendation capability

Experimental settings for cross-validation

Cross-validated testing accuracy rates of our method when considered as a supervised learning method are 80% and 75% in $\mathcal{D}_{\text{AFLOW}}$, and $\mathcal{D}_{\text{LTVC}}$ data sets, respectively, which are almost at the same level with those in the previous study (Huang, Martin, and Zhuang, 2019). However, our work pays more attention toward calculating the recall, which is the percentage of the total HEAs correctly classified. This recall value is a more appropriate evaluation measure compared to supervised learning accuracy for finding new combinations of elements having HEA phases.

Because the $\mathcal{D}_{\text{ASMI16}}$ data set only contains binary alloys, we can learn a similarity matrix between the elements from a training set sampled from $\mathcal{D}_{\text{ASMI16}}$. By applying the proposed process for recommending substituted alloys, we can rank all the possible binary alloys other than those in the training set. A total of 351 hypothetical binary alloys showing equivalent components can be generated from the 27 elements in $\mathcal{E}$, 45 of which are contained in $\mathcal{D}_{\text{ASMI16}}$. Because no information is available for the other 306 alloys, they are ranked by the constructed model. We apply 9-fold cross-validation to $\mathcal{D}_{\text{ASMI16}}$. A total of 40 out of the 45 alloys in $\mathcal{D}_{\text{ASMI16}}$ are used as the training set, and the remaining 5 alloys are used as the test set to evaluate the HEA recall rate. The model learned from the 40 alloys in the training set is then used to rank the other 311 alloys, including the 5 in the test set. This cross-validation is repeated 100 times so that the HEA-recommendation performance can be reliably calculated.

Because the $\mathcal{D}_{\text{CALPHAD}}$ data set only contains ternary alloys, we can learn a similarity matrix between the elements or binary combinations thereof from a training set sampled from $\mathcal{D}_{\text{CALPHAD}}$. We can build a model to rank all the possible ternary alloys other than those in the training set. There are 2,925 hypothetical ternary alloys showing equivalent components that can be generated from the 27 elements in $\mathcal{E}$, 243 of which are contained in $\mathcal{D}_{\text{CALPHAD}}$. Because no information is available for the other 2,682 alloys, they are ranked by the constructed model. We apply 9-fold cross-validation to $\mathcal{D}_{\text{CALPHAD}}$ and use 216 of the 243 alloys in $\mathcal{D}_{\text{CALPHAD}}$ as the training set. The remaining 27 alloys in $\mathcal{D}_{\text{CALPHAD}}$ are used as the test set to evaluate the HEA recall rate. The model learned from the 216 alloys in the training set is used to rank the other 2,709 alloys, including the 27 in the test set. This cross-validation is also repeated 100 times to ensure the reliable evaluation of the HEA-recommendation performance.

In contrast, the $\mathcal{D}_{\text{ASMI16}}$, $\mathcal{D}_{\text{CALPHAD}}$, $\mathcal{D}_{\text{AFLOW}}$, and $\mathcal{D}_{\text{LTVC}}$ data sets contain both binary and ternary alloys. Owing to the information obtained from both types of alloys, we can learn a similarity matrix between the various elements, elements and binary combinations thereof, and binary element combinations obtained from the training set sampled from $\mathcal{D}_{\text{AFLOW}}$ and $\mathcal{D}_{\text{LTVC}}$. We can build a model to rank all the possible candidates for binary and ternary alloys other than those in the training set.

There are 3,276 hypothetical binary and ternary alloys showing equivalent components that can be generated from the 27 elements in $\mathcal{E}$, 558 of which are contained in $\mathcal{D}_{\mathrm{AFLOW}}$. Because no information is available for the other 2,718 alloys, they are ranked by the constructed model. We apply 9-fold cross-validation to $\mathcal{D}_{\mathrm{AFLOW}}$ and use 496 of the 558 alloys in $\mathcal{D}_{\mathrm{AFLOW}}$ as the training set. The remaining 62 alloys in $\mathcal{D}_{\mathrm{AFLOW}}$ are used as the test set to evaluate the HEA recall rate. The model learned from the 496 alloys in the training set is used to rank the other 2,780 alloys including the 62 in the test set. The same evaluation method is applied to $\mathcal{D}_{\mathrm{AFLOW}}$.

A similar experiment is conducted with the $\mathcal{D}_{\mathrm{LTVC}}$ data set to evaluate the HEA-recommendation performance of the proposed ERS. Note that although the $\mathcal{D}_{\mathrm{LTVC}}$ data set contains the same alloys as the $\mathcal{D}_{\mathrm{AFLOW}}$ one, the target properties of the alloys are dissimilar because the values are estimated using different computation methods (Nyshadham et al., 2017; Lederer et al., 2018).

It should be noted that owing to the computational cost, these experiments do not use the selected alloys (i.e., those in the test set) to improve the accuracy of the HEA recommendation model for the next trial. A recommendation model based on the results of previous trials may work more accurately.

Experimental settings for evaluation of extrapolation capability

Because $\mathcal{D}_{\mathrm{AFLOW}}$ contains both binary and ternary alloys, we can learn the similarities between the various elements and binary combinations thereof. Consequently, we can apply the ERS to $\mathcal{D}_{\mathrm{AFLOW}}$ to rank the 17,550 quaternary alloys comprising the 27 elements contained in $\mathcal{E}$. Additionally, $\mathcal{D}_{\mathrm{AFLOW}}$ and $\mathcal{D}_{\mathrm{AFLOW}}^{\mathrm{quaternary}}$ are both used to build a recommender system that ranks all the possible candidates (i.e., 80,730 alloys) for synthesizing quinary HEAs. The 754 quaternary HEAs in $\mathcal{D}_{\mathrm{AFLOW}}^{\mathrm{quaternary}}$ and 129 quinary HEAs in $\mathcal{D}_{\mathrm{AFLOW}}^{\mathrm{quinary}}$ are used to monitor the HEA recall rate for recommending quaternary and quinary HEAs, respectively. Moreover, similar experiments are conducted on the $\mathcal{D}_{\mathrm{LTVC}}$, $\mathcal{D}_{\mathrm{LTVC}}^{\mathrm{quaternary}}$, and $\mathcal{D}_{\mathrm{LTVC}}^{\mathrm{quinary}}$ data sets to evaluate the HEA-recommendation performance of the ERS.

3.4 Case study 2: Solvent screening for efficient chemical exfoliation of graphite

Geim and Novoselov demonstrated graphene production by peeling graphite with Scotch tape Novoselov, 2004. One advantage of graphene over other carbon nanomaterials is its direct production from graphite through a top-down approach. This indicates that the cost of graphene could be comparable to that of graphite if the production process becomes highly efficient. Currently, large-scale graphene production relies primarily on the graphene oxide (GO) route (Toh et al., 2014; Eigler et al., 2013; Feng et al., 2013; Zhu et al., 2010; Gu et al., 2018; Dreyer et al., 2010). In this route, graphite is oxidized by potassium permanganate in a sulfuric acid medium Hummers and Offeman, 1958, resulting in GO with various oxygen-containing functional groups introduced onto graphene sheets. The thermal or chemical reduction of GO eventually yields graphene, specifically referred to as reduced GO. However, the GO route has drawbacks, including the generation of significant amounts of harmful waste and the introduction of non-negligible chemical and structural defects on graphene sheets (Jiang et al., 2011; Xu et al., 2011; Larciprete et al., 2011; Liu et al., 2011; Ton et al., 2018).

As a promising alternative for mass-producing higher-quality graphene, liquid-phase exfoliation without relying on chemical energy has gained attention (Nicolosi et al., 2013; Cui et al., 2011; Qian et al., 2009; Hao et al., 2008). Higher quality in this context refers to graphene with fewer layers and reduced defect density. This approach requires physical energy for exfoliation and a solvent for stabilizing the produced graphene sheets as a dispersion. Cavitation, generated by ultrasonication (Lotya et al., 2010; Khan et al., 2010; Shen et al., 2015), jet cavitation (Shen et al., 2011; Wang, Shen, and Yi, 2019; Lin, Wu, and Liu, 2018; Yi et al., 2011), and high-pressure homogenization (Nacken et al., 2015; Qi et al., 2017), is a commonly employed physical energy for exfoliation. Various solvents have been reported as effective for the liquid-phase exfoliation of graphite. Among them, N-methyl pyrrolidone (Hasan et al., 2007; al., 2008b), 1,2-dichlorobenzene (Hamilton et al., 2009; Sahoo et al., 2013), and benzylamine (Cai et al., 2012; Economopoulos et al., 2010) are popular solvents initially used to disperse other carbon nanomaterials, and have proven effective in producing high-quality graphene. Additional solvents include N,N-dimethylformamide (al., 2008a), dimethyl sulfoxide (Du et al., 2013), 1,3-dimethyl-2-imidazolidinone, N,N-dimethylacetamide, and γ-butyrolactone (al., 2008b). Early efforts to find exfoliating solvents focused on producing high-quality graphene with higher yields. Generally, the yield and quality of graphene are not solely determined by the choice of solvents. Prolonged ultrasonication enhances graphene yield but at the cost of defect formation, while higher-speed centrifugation provides thinner graphene sheets at the expense of yield.

Nevertheless, the yield in the literature typically ranges from 0.1 to 1 miligram (mg) of graphene per milliliter (mL) of solvent (Hasan et al., 2007; al., 2008b; Hamilton et al., 2009; Sahoo et al., 2013; Cai et al., 2012; Economopoulos et al., 2010; al., 2008a; Du et al., 2013). Another focus of solvent exploration is the development of a green process. The addition of surfactants or polymers compensates for differences in surface tension (Lotya et al., 2010; Bourlinos et al., 2009; Guardia et al., 2011; Liang and Hersam, 2010; Wajid et al., 2012; Patole et al., 2010; Li et al., 2013), enabling graphite exfoliation even in water. However, the persistence of additives during solvent removal remains an important issue to address. Solvent exploration also aims to diversify the solvent library, providing suitable solvents for specific applications and understanding the molecular features that determine the yield and quality of produced graphene. Many exfoliating solvents have surface tensions close to that of graphene (40 mJ/m˘^{2}) (Bissantz, Kuhn, and Stahl, 2010), yet this does not fully explain the significant yield deviations observed among solvents with comparable surface tension (Shen et al., 2015; al., 2008b). Other features, such as the Hildebrand and Hansen solubility parameters, also present similar problems (Shen et al., 2015), resulting from the inability to describe molecular interactions based on macroscopic parameters. In 2009, Bourlinos et al. reported several new exfoliating solvents (Bourlinos et al., 2009). They hypothesized that electron-withdrawing groups enhance π-π interaction through charge transfer, based on the finding of perfluorinated aromatic molecules as exfoliating solvents. They also suggested that electron-donating groups may play a similar role, as evidenced by pyridine's usability as an exfoliating solvent. However, these features, combined with surface tension matching, still do not completely explain the inability of solvents such as aniline, pyrrole, and thiophene.

This section describes our research endeavors in further exploring solvents for the liquid-phase exfoliation of graphite under ultrasonication. We conducted a comprehensive solvent screening process while concurrently developing a novel machine learning method based on the evidence theory, which can effectively learn

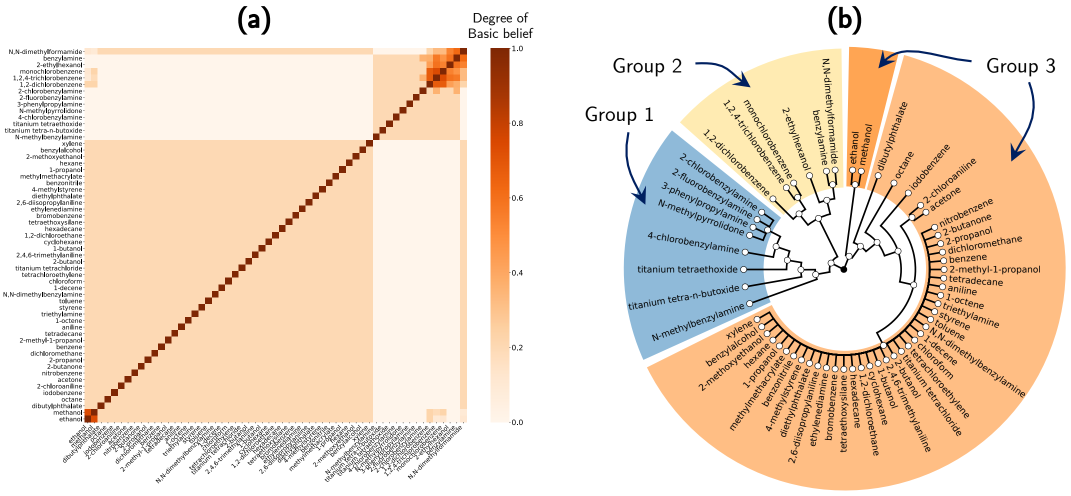


FIGURE 3.7: a) Heat map of the similarity matrix for the 57 solvents, where the color represents the degree of basic belief to be similar. b) Hierarchical cluster structure that is constructed from the similarity matrix using the hierarchical agglomerative clustering method.

from the experimental data obtained. As a result, we identified several new exfoliating solvents that showed promising potential. Notably, the combination of chlorobenzene and benzylamine, known as chlorobenzylamine, exhibited the highest graphene yield among the solvents tested. Additionally, we observed that titanium alkoxide, despite having a much lower surface tension than graphene, still yielded good results. However, the most significant finding of our study is the importance of utilizing a synergy between different functional groups, mainly through solvent mixtures. For instance, the chlorobenzene/benzylamine mixture yielded a higher graphene yield compared to using chlorobenzylamine alone. These findings highlight the potential of leveraging diverse solvent combinations to enhance the exfoliation process and improve graphene production.

3.4.1 Solvent dataset

To evaluate the similarity between solvents in the data set $\mathcal{D}$, we employed machine learning based on the Dempster-Shafer theory of evidence (Shafer, 1976). The solvents, both single solvents (Table 3.3) and their mixtures (Table 3.4), were classified into two classes based on a threshold value of the graphene yield: exfoliating ($\geq$ 0.1 mg mL^{-1}) or non-exfoliating ($<$ 0.1 mg mL^{-1}) solvents (Nu Thanh Ton et al., 2020). These solvents that exfoliate and do not exfoliate are notated by labels EXF and $\neg EXF$, respectively.

3.4.2 Learning similarity between solvents.

Figure 3.7 a illustrates the similarity matrix obtained for the 57 solvents considered in this study, representing their contribution to the exfoliation of graphite as single solvents and solvent mixtures. Each element of the matrix represents the degree of a fundamental belief in the similarity between two solvents based on their properties. To construct a hierarchical cluster structure, the similarity matrix was transformed into a distance matrix, and a hierarchical agglomerative clustering method Nielsen, 2016 was applied (Figure 3.7 b).

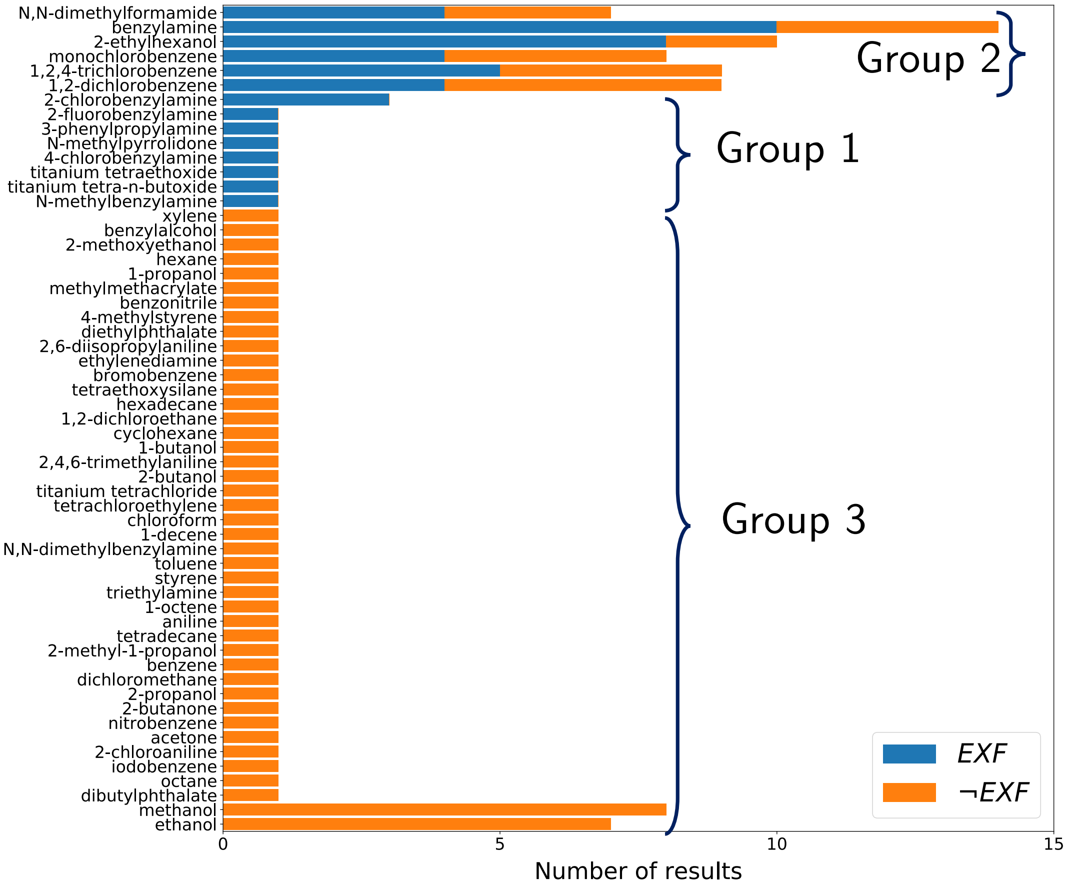


FIGURE 3.8: Numbers of exfoliating and non-exfoliating results (EXF and $\neg EXF$) for each of the 57 solvents. Note that both single solvents and solvent mixtures were taken into account.

The resulting hierarchical cluster structure revealed the presence of three distinct groups of solvents. The first group included solvents such as titanium tetra-*n*-butoxide, 4-chlorobenzylamine, N-methylpyrrolidone, 3-phenylpropylamine, 2-chlorobenzylamine, titanium tetraethoxide, N-methylbenzylamine, and 2-fluorobenzylamine. The second group comprised solvents like 2-ethylhexanol, benzylamine, N,N-dimethylformamide, monochlorobenzene, 1,2-dichlorobenzene, and 1,2,4-trichlorobenzene. The third group encompassed the remaining solvents. In Figure 3.8, the numbers of exfoliating and non-exfoliating results are presented for each specific solvent, considering both single solvents and solvent mixtures. It is evident that the first group corresponds to the group of strong single solvents with graphene yields above 0.1 mg mL^{-1} even in their single form. Conversely, the third group corresponds to the group of single solvents with graphene yields below 0.1 mg mL^{-1} (mostly yielding no graphene). The obtained similarity matrix clearly indicates a substantial similarity between ethanol and methanol. This finding aligns with the fact that none of the solvent mixtures containing these two solvents achieved a graphene yield above 0.1 mg mL^{-1}. The second group comprises both strong and weak exfoliating solvents, with graphene yields above and below 0.1 mg mL^{-1}, respectively. Interestingly, the similarity matrix reveals a significant similarity among monochlorobenzene, 1,2-dichlorobenzene, and 1,2,4-trichlorobenzene (Figure 3.7). Additionally, N,N-dimethylformamide was found to be similar to 2-ethylhexanol and benzylamine in terms of their properties.

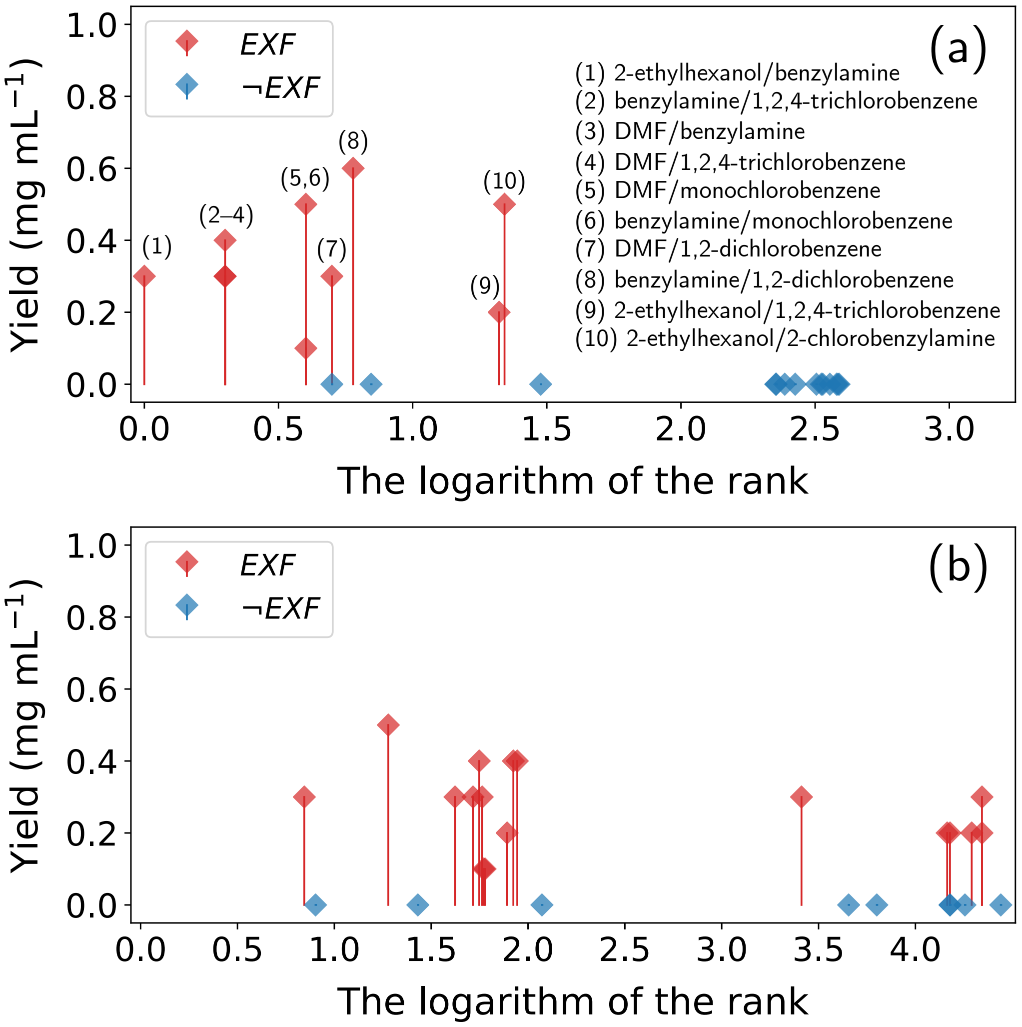


FIGURE 3.9: a) Stem plot of 24 screened binary solvent mixtures. b) Stem plot of 26 new ternary solvent mixtures that are randomly selected for the sake of experimental validation. The red and blue diamonds indicate experimental validation results corresponding to exfoliating and non-exfoliating mixtures, respectively.

These three solvents exhibited similar behavior when used as solvent mixtures (Figure 3.8).

3.4.3 Recommending potential solvent mixtures.

To evaluate the model's capability, leave-one-out cross-validation was conducted to recommend new solvent mixtures based on machine learning. Each binary solvent mixture (Runs 58-81 in Table 3.4) was used for testing while the remaining solvents were used to learn the similarity. The properties of unconfirmed binary mixtures were also predicted simultaneously. The results of the testing were used to rank all possible binary solvent mixtures (Figure 3.9 a). Out of the ten screened binary solvent mixtures with graphene yields above 0.1 mg mL^{-1}, eight were ranked in the top ten by the proposed method, while the remaining two were ranked at 21st and 22nd. On the other hand, the 11 binary solvent mixtures out of the 14 with exfoliation yields below 0.1 mg mL^{-1} had rankings ranging from 200th to 400th. The remaining three mixtures were ranked in the top fifty. Solvent mixtures that lacked information about their properties started with a ranking of 400th. These results clearly demonstrate the effectiveness of the proposed method in recommending potential binary solvent mixtures.

Furthermore, the model was used to recommend new potential ternary solvent mixtures by ranking all possible combinations that could be generated from the 57 solvents. To assess the significance of the ranking, 14 ternary solvent mixtures ranked within the top 150, and 12 mixtures with lower rankings were selected for additional experiments. Notably, the 12 mixtures with lower rankings had a low probability mass assigned to the class EXF. Since most solvent mixtures had limited data, they were assigned low rankings. Thus, a low-ranked solvent mixture did not necessarily indicate a lower capability of graphite exfoliation but rather a higher degree of uncertainty in inductively estimating its exfoliation capability based on the observed data.

Figure 3.9 b overviews the rankings and observed graphene yields for the 26 selected ternary solvent mixtures. The results are summarized in Table 3.5. Out of the 14 ternary solvent mixtures ranked within the top 150, 11 exhibited a graphene yield above 0.1 mg mL^{-1}. However, the remaining three mixtures did not demonstrate the capability for graphite exfoliation. Conversely, out of the 12 low-ranked ternary solvent mixtures, six showed a high graphene yield, while the other six mixtures did not exhibit the ability to exfoliate graphite. These experimental validation results highlight the effectiveness of the proposed recommendation model based on the Dempster-Shafer theory of evidence. Notably, the presence of non-solvents did not hinder the graphene yield if the other two solvents were chosen appropriately (Runs 103, 108, and 112). Furthermore, the co-existence of a Lewis base and acid did not impair their capabilities (Run 94). These findings demonstrate the power of machine learning in expanding the exploration range beyond human biases acquired from limited experiences.

3.5 Contributions and limitations

Applying inductive approaches typically requires sufficient and balanced data to ensure prediction accuracy. However, material data is often scarce or heavily biased towards positive results (Table 3.1). Building a prediction model with such limited data and a strong skew toward positive results is challenging. Moreover, conflicts within and between material datasets are additional obstacles that inductive approaches must overcome. Thus, quantitative assessment of the prediction uncertainty is crucial. The ERS offers an advantage in handling these situations, and its effectiveness has been demonstrated in applications for exploring high-entropy alloys and solvent mixtures. Instead of forcibly merging data from multiple datasets, our system treats each dataset as an evidence source. It then combines the evidence to draw reasonable final conclusions for recommending HEAs or solvent mixture, with the epistemic and aleatoric uncertainties being quantitatively evaluated.

However, to serve the purpose of screening element combinations forming HEA phases, the ERS focuses on the fundamental question of whether the HEA phase exists. Therefore, we design a frame of discernment $\Omega_{HEA} = HEA, \neg HEA$ to model the existence of HEA phases with mass functions. As a result, the ERS does not address essential questions regarding the structure and other properties of the HEAs. However, by redesigning the frame of discernment to reflect additional properties of interest, we can construct a model that can recommend potential alloys forming HEA phases with desirable properties.

Additionally, the evidence-based recommender system has certain limitations regarding the target property type. Owing to its foundation in Dempster–Shafer theory, the system is not equipped to directly model continuous properties, such as evaluating the yield of graphene in a solvent. Rather than directly modeling the continuous property, we are required to categorize the property into discrete segments. This approach may lead to potential loss of information or granularity in the evaluation process. To address this issue, we can utilize the transferable belief functions theory, which Philippe Smets proposed. This theory offers an alternative framework for dealing with continuous properties and associated uncertainties more flexibly and comprehensively. By incorporating the transferable belief functions theory into the evidence-based recommender system, we can better accommodate continuous properties and enhance the system's overall applicability to a broader range of materials science problems while preserving the nuanced details of the continuous property under investigation.

TABLE 3.3: Screening results for single solvents. (Nu Thanh Ton et al., 2020)

Run	Solvent	Functional groups	Surface tension (mJ m^{-2})	Yield[a] (mg mL^{-1})
1	N-Methylpyrrolidone	Cycloalkylamide	40.79	0.2
2	N,N-Dimethylformamide	Alkylamide	36.42	< 0.1
3	Benzylamine	R–NH$_2$, aromatic	38.82	0.2
4	1,2-Dichlorobenzene	Aryl halide (Cl)	36.61	0.1
5	N-Methylbenzylamine	R$_2$–NH, aromatic	n.d.	0.3
6	N,N-Dimethylbenzylamine	R$_3$–N, aromatic	38.81	0
7	3-Phenylpropylamine	R–NH$_2$, aromatic	37.6	0.3
8	2,6-Diisopropylaniline	Ar–NH$_2$	33.9	< 0.1
9	2,4,6-Trimethylaniline	Ar–NH$_2$	33.92	< 0.1
10	Aniline	Ar–NH$_2$	43.40	0
11	Triethylamine	R$_3$–N	20.22	0
12	Ethylenediamine	R–NH$_2$	42	0
13	Benzonitrile	Nitrile, aromatic	38.79	0
14	Nitrobenzene	Nitro, aromatic	46.34	0
15	1,2,4-Trichlorobenzene	Aryl halide (Cl)	39.1	0.2
16	Monochlorobenzene	Aryl halide (Cl)	33.60	0.2
17	Bromobenzene	Aryl halide (Br)	36.50	0
18	Iodobenzene	Aryl halide (I)	39.70	0
19	Chloroform	Alkyl halide (Cl)	27.50	0
20	1,2-Dichloroethane	Alkyl halide (Cl)	33.30	0
21	Dichloromethane	Alkyl halide (Cl)	26.50	0
22	Tetrachloroethylene	Alkenyl halide (Cl)	31.74	0
23	2-Chlorobenzylamine	R–NH$_2$, aryl halide (Cl)	42.0	0.5
24	4-Chlorobenzylamine	R–NH$_2$, aryl halide (Cl)	42.0	0.4
25	2-Fluorobenzylamine	R–NH$_2$, aryl halide (F)	n.d.	0.3
26	2-Chloroaniline	Ar–NH$_2$, aryl halide	43.66	0
27	Benzylalcohol	Alcohol	39.00	0
28	2-Ethylhexanol	Alcohol	28.00	0.1
29	2-Methoxyethanol	Alcohol, ether	30.84	< 0.1
30	Methanol	Alcohol	22.70	0
31	Ethanol	Alcohol	22.10	0
32	1-Propanol	Alcohol	23.75	0
33	2-Propanol	Alcohol	23.00	0
34	1-Butanol	Alcohol	24.93	0
35	2-Butanol	Alcohol	23.0	0
36	2-Methyl-1-propanol	Alcohol	23.0	0
37	Acetone	Ketone	25.20	0
38	2-Butanone	Ketone	23.97	0
39	Methylmethacrylate	Ester	28	0
40	Diethylphthalate	Aromatic, ester	37.5	0
41	Dibutylphthalate	Aromatic, ester	34	0
42	Styrene	Aromatic hydrocarbon	32.3	0
43	4-Methylstyrene	Aromatic hydrocarbon	n.d.	0
44	Benzene	Aromatic hydrocarbon	28.88	0
45	Xylene	Aromatic hydrocarbon	30.10	0
46	Toluene	Aromatic hydrocarbon	28.40	0
47	1-Octene	Aliphatic hydrocarbon	21.76	0
48	1-Decene	Aliphatic hydrocarbon	24	0
49	Octane	Aliphatic hydrocarbon	21.62	0
50	Hexane	Aliphatic hydrocarbon	18.43	0
51	Tetradecane	Aliphatic hydrocarbon	26.56	0
52	Hexadecane	Aliphatic hydrocarbon	27.47	0
53	Cyclohexane	Aliphatic hydrocarbon	24.95	0
54	Titanium tetra-*n*-butoxide	Titanium alkoxide	28.0	0.3
55	Titanium tetraethoxide	Titanium alkoxide	23.1	0.2
56	Titanium tetrachloride	Titanium halide (Cl)	n.d.	0
57	Tetraethoxysilane	Silicone alkoxide	22.8	0

[a] The screening was performed in two steps: When the supernatant looked clear or grayish, the graphene yield was regarded to be 0 or < 0.1 mg mL˘1, respectively; When black, the graphene concentration was quantified based on UV/Vis measurements.

TABLE 3.4: Screening results for solvent mixtures. (Nu Thanh Ton et al., 2020)

Run	Solvent mixture[a]	Surface tension[b] (mJ m^{-2})	Yield (mg mL^{-1})
58	Benzylamine/1,2-dichlorobenzene	37.73	0.6
59	Benzylamine/monochlorobenzene	36.11	0.5
60	Benzylamine/1,2,4-trichlorobenzene	38.95	0.4
61	2-Ethylhexanol/2-chlorobenzylamine	35.90	0.5
62	2-Ethylhexanol/benzylamine	34.37	0.3
63	2-Ethylhexanol/1,2,4-trichlorobenzene	34.18	0.2
64	N,N-Dimethylformamide/benzylamine	37.42	0.3
65	N,N-Dimethylformamide/monochlorobenzene	35.20	0.1
66	N,N-Dimethylformamide/1,2-dichlorobenzene	36.49	0.3
67	N,N-Dimethylformamide/1,2,4-trichlorobenzene	37.45	0.3
68	Monochlorobenzene/1,2-dichlorobenzene	35.01	0
69	Monochlorobenzene/1,2,4-trichlorobenzene	36.07	0
70	1,2-Dichlorobenzene/1,2,4-trichlorobenzene	37.79	0
71	Ethanol/N,N-dimethylformamide	28.26	< 0.1
72	Ethanol/benzylamine	27.93	< 0.1
73	Ethanol/1,2,4-trichlorobenzene	27.53	0
74	Ethanol/monochlorobenzene	26.30	0
75	Ethanol/2-ethylhexanol	23.70	0
76	Methanol/N,N-dimethylformamide	27.41	0
77	Methanol/benzylamine	26.83	0
78	Methanol/monohlorobenzene	25.81	0
79	Methanol/1,2-dichlorobenzene	26.36	0
80	Methanol/1,2,4-trichlorobenzene	26.73	0
81	Methanol/2-ethylhexanol	23.79	0
82	2-Ethylhexanol/benzylamine/1,2,4-trichlorobenzene	35.98	0.8
83	2-Ethylhexanol/benzylamine/1,2-dichlorobenzene	35.18	0.7
84	2-Ethylhexanol/benzylamine/monochlorobenzene	34.08	0.5
85	2-Ethylhexanol/benzylamine/2-chlorobenzylamine	37.02	0.5
86	Ethanol/1,2-dichlorobenzene/benzylamine	30.11	0
87	Methanol/1,2-dichlorobenzene/benzylamine	29.03	0
88[c]	Aluminum isopropropoxide/1,2-dichlorobenzene	n.d.	< 0.1

[a] Equivolume mixtures of two or three solvents.

[b] The surface tension of a solvent mixture was derived based on the mole-fraction-weighted average of the surface tension of individual solvents.

[c] 500 mg of aluminum isopropoxide in a solid state was dissolved in 5.0 mL of 1,2-dichlorobenzene, and the resultant mixture was used for the exfoliation of graphite.

TABLE 3.5: Experimental validation of prediction results.

No.	Solvent mixture	ERS results			Yield[a] (mg mL^{-1})
		High	Low	Not sure	
89	Monochlorobenzene/2-ethylhexanol/N,N-dimethylformamide	0.95	0.00	0.05	0.3
90	2-Ethylhexanol/2-chlorobenzylamine/N,N-dimethylformamide	0.92	0.00	0.08	0.0
91	Monochlorobenzene/1,2-dichlorobenzene/benzylamine	0.65	0.33	0.02	0.5
92	Titanium tetra-*n*-butoxide/2-ethylhexanol/benzylamine	0.59	0.00	0.41	0.0
93	Monochlorobenzene/2-ethylhexanol/N-methylbenzylamine	0.20	0.00	0.80	0.3
94	2-Chlorobenzylamine/2-ethylhexanol/3-phenylpropylamine	0.20	0.00	0.80	0.3
95	2-Chlorobenzylamine/titanium tetraethoxide/benzylamine	0.20	0.00	0.80	0.4
96	2-Chlorobenzylamine/2-ethylhexanol/4-chlorobenzylamine	0.20	0.00	0.80	0.3
97	2-Chlorobenzylamine/2-ethylhexanol/titanium tetraethoxide	0.20	0.00	0.80	0.1
98	2-Chlorobenzylamine/2-ethylhexanol/titanium tetra-*n*-butoxide	0.20	0.00	0.80	0.1
99	Monochlorobenzene/2-ethylhexanol/N-methylpyrrolidone	0.20	0.00	0.80	0.2
100	Monochlorobenzene/benzylamine/titanium tetra-*n*-butoxide	0.20	0.00	0.80	0.4
101	Monochlorobenzene/2-ethylhexanol/4-chlorobenzylamine	0.20	0.00	0.80	0.4
102	2-Chlorobenzylamine/methanol/2-ethylhexanol	0.00	0.91	0.09	0.0
103	Monochlorobenzene/benzylamine/benzylalcohol	0.00	0.00	1.00	0.3
104	Monochlorobenzene/2-ethylhexanol/xylene	0.00	0.00	1.00	0.0
105	2-Chlorobenzylamine/2-methoxyethanol/benzylamine	0.00	0.00	1.00	0.0
106	2-Chlorobenzylamine/iodobenzene/benzylamine	0.00	0.00	1.00	0.2
107	2-Chlorobenzylamine/2-ethylhexanol/dichloromethane	0.00	0.00	1.00	0.0
108	2-Chlorobenzylamine/2-ethylhexanol/2-propanol	0.00	0.00	1.00	0.2
109	2-Chlorobenzylamine/2-ethylhexanol/acetone	0.00	0.00	1.00	0.0
110	Monochlorobenzene/benzylamine/2,6-diisopropylaniline	0.00	0.00	1.00	0.0
111	Bromobenzene/1,2,4-trichlorobenzene/benzylamine	0.00	0.00	1.00	0.2
112	Monochlorobenzene/benzylamine/nitrobenzene	0.00	0.00	1.00	0.3
113	Monochlorobenzene/benzylamine/2-chloroaniline	0.00	0.00	1.00	0.2
114	Monochlorobenzene/2-ethylhexanol/toluene	0.00	0.00	1.00	0.0

[a] The experimentally obtained yield.

Chapter 4

Quantifying material similarity with uncertainty with respect to continuous properties

4.1 Introduction

Machine learning has immense potential in materials science, particularly in the discovery of new materials and understanding their properties. In many materials science problems, data-driven approaches are used to leverage existing material data to predict the properties of new materials and uncover underlying physicochemical mechanisms (Kailkhura et al., 2019). One intuitive and interpretable data-driven approach is analogy-based inductive reasoning, which involves inferring the properties of a new instance based on the information from observed instances that are most similar to it (Tenenbaum, 1996; Tenenbaum, Silva, and Langford, 2000; Yang et al., 2014; Chen et al., 2019). By applying analogy-based models, researchers can explain the reasoning process behind predictions and gain insights into the physicochemical mechanisms underlying the observed properties (Letham et al., 2015; Rudin, 2019). In materials science, researchers have successfully addressed various problems by systematically capturing analogies in composition or structure between materials that exhibit similar physicochemical properties (Goldsmith et al., 2017; Ramprasad et al., 2017; Nguyen et al., 2018b; Nguyen et al., 2019). These analogy-based models provide valuable insights into the relationships between materials, enabling more accurate predictions and targeted experimentation. As a result, the discovery of new materials with desired properties can be accelerated, the understanding of material behavior can be enhanced, and the field of materials science can advance.

In condensed matter physics, a discipline rooted in fundamental principles, it is essential to uncover the physical mechanisms underlying specific material properties. However, despite the discovery of numerous new materials with exceptional properties, accurately quantifying material similarities to reveal the underlying physicochemical mechanisms remains a challenge. This challenge arises from the fact that material properties are often interpreted based on relative criteria and physicochemical concepts. Superconductivity in materials provides a prime example of this complexity. The Bardeen-Cooper-Schrieffer (BCS) theory of superconductivity (Bardeen, Cooper, and Schrieffer, 1957) successfully describes the microscopic mechanisms, attributing superconductivity to electron-phonon interactions. However, alternative mechanisms, such as electron-electron interactions, are believed to drive the superconductivity observed in high-T_C cuprates. Classifying the superconducting mechanism of materials is not straightforward, as multiple mechanisms can contribute cooperatively to enhance the critical temperature (T_C). Achieving a consensus on the origin of superconductivity is challenging due to the intricate interplay of mechanisms. Despite these classification difficulties, inductively quantifying material similarities and grouping similar materials using available observational data can help uncover the underlying physical mechanisms. By harnessing data-driven approaches and machine learning techniques, researchers can gain insights into material similarities, leading to a deeper understanding of physicochemical mechanisms and potentially expediting the discovery of materials with desired properties.

Indeed, inductive reasoning with inefficient similarity assessment can introduce aleatoric uncertainty, leading to challenges such as misidentification of outliers and difficulties in explaining underlying physicochemical mechanisms using single models (Hüllermeier and Waegeman, 2021; Seko, Togo, and Tanaka, 2018). To address these challenges, it is necessary to conduct an exhaustive examination of all possible hypotheses about the unknown physicochemical mechanisms to accurately assess the similarity between materials, considering predefined material descriptors. Furthermore, similarity measures in materials science are often context-dependent, and they need to be adapted to adequately capture the specific phenomena under study (Tversky, 1977; Goldstone, Medin, and Halberstadt, 1997). The context-dependent nature of similarity implies that the measure must account for uncertainty arising from the context or the measurement process itself, particularly in situations where material data are limited and heavily biased. It is crucial to consider the uncertainty associated with the data and the measurement conditions to ensure robust and reliable similarity assessments. Moreover, it is essential to note that similarities from different contexts may not be directly comparable when integrating data to conclude material similarity. The integration process should account for the varying contexts and the associated uncertainties to make meaningful and accurate assessments. These factors contribute to the challenges faced when applying data-driven approaches to materials science, requiring careful consideration and appropriate methodologies to overcome them.

To address these challenges and extract knowledge efficiently from the data, we propose a novel approach that shifts the focus from measuring similarities between materials to quantitatively measuring the confidence in their similarities. This approach utilizes the Dempster-Shafer theory (Shafer, 1976; Denœux, Dubois, and Prade, 2020; Dempster, 1967), also known as evidence theory, to develop an evidential regression-based similarity measurement (eRSM). The objective is to identify subgroups of materials where the learned models exhibit high correlations between descriptors and the target property of the constituent materials. Further analysis of the models describing these subgroups provides valuable insights to extract, interpret, and understand the underlying physical mechanisms. The Dempster-Shafer theory can be seen as a generalization of the Bayesian approach, specifically designed to handle problems involving incomplete and insufficient information. It is well-suited for addressing material data problems (Nu Thanh Ton et al., 2020; Ha et al., 2021). In our approach, the measure of similarity refers to whether the observed physical properties of the materials under study can be explained by the same hidden mechanism that has not yet been revealed. In other words, we consider a pair of materials in the dataset as similar if the same underlying mechanism can describe their physical properties; otherwise, the pair of materials is considered dissimilar.

To implement the eRSM, we first generate numerous hypothetical mechanisms by randomly selecting subsets of data instances and constructing regression models for each subset. Each of these regression models serves as a source of evidence

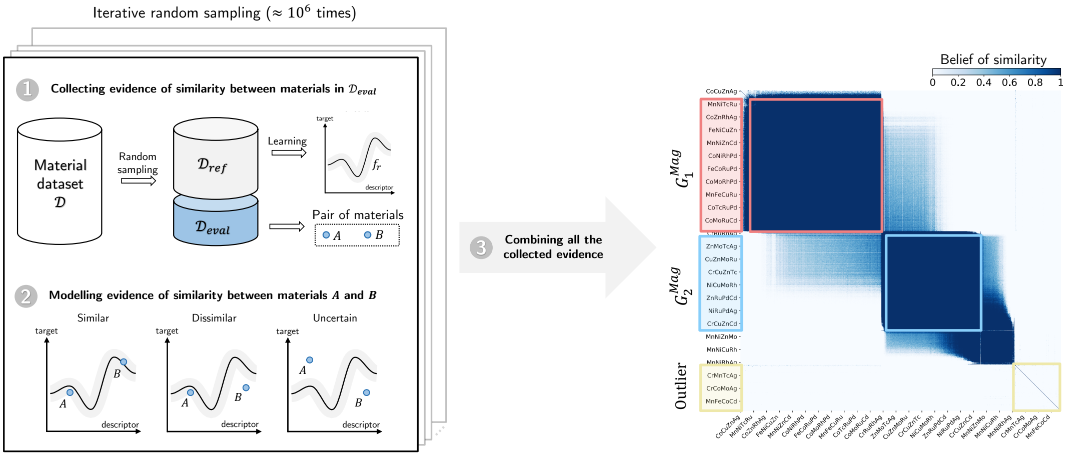


FIGURE 4.1: Workflow of the similarity measurement for materials with respect to their composition and a continuous property.

regarding the similarities between materials. We then employ the Dempster-Shafer theory, which provides a framework for modeling and combining the uncertainty of evidence to integrate the collected pieces of evidence and draw conclusions about the similarities between materials. The eRSM consists of three main steps as follows (Fig. 4.1):

1. *Collect sources of evidence:* Hypothetical mechanisms, which are represented by reference function f_r, are collected from a dataset by applying regression analysis with single or mixture models. The functions are used as sources of evidence to rationalize the similarity states of materials.

2. *Model similarity evidence:* An appropriate mass function is designed to model the obtained evidence within the framework of the evidence theory.

3. *Combine pieces of evidence:* Dempster's rule of combination is used to integrate the pieces of the evidence.

The steps of the eRSM are explained in detail in Section 4.2. Regarding the framework of the evidence theory, the essential contributions of the eRSM are collecting sources of evidence about the similarities between materials from datasets and designing suitable mass functions to model the pieces of evidence rationally. The effectiveness of obtained similarities using the eRSM for subdividing alloys from datasets into homogenous subgroups is supported by experiments on 1) a dataset of binary alloys with their Curie temperature as a target property (Section 4.4); and 2) two dataset of quaternary alloys with their magnetization (Section 4.4.2) and Curie temperature (Section 4.4.3) as the target properties. Further analysis of the detected subgroups to interpret the underlying physical mechanisms is shown in Section 4.4.4

4.2 Evidential regression-based similarity measurement (eRSM)

We consider a dataset $\mathcal{D}$ consisting of p data instances. We assume that a data instance with index i in $\mathcal{D}$ is described by n predefined descriptors and is represented by an n-dimensional numerical vector, $\boldsymbol{x}_i = \left(x_i^1, x_i^2, \ldots, x_i^n\right) \in \mathbb{R}^n$. The target property of the data instance $\boldsymbol{x}_i$ is $y_i \in \mathbb{R}$. Thereafter, the dataset $\mathcal{D} = \left\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2) \ldots (\boldsymbol{x}_p, y_p)\right\}$ is represented using a $(p \times (n+1))$ matrix. In this

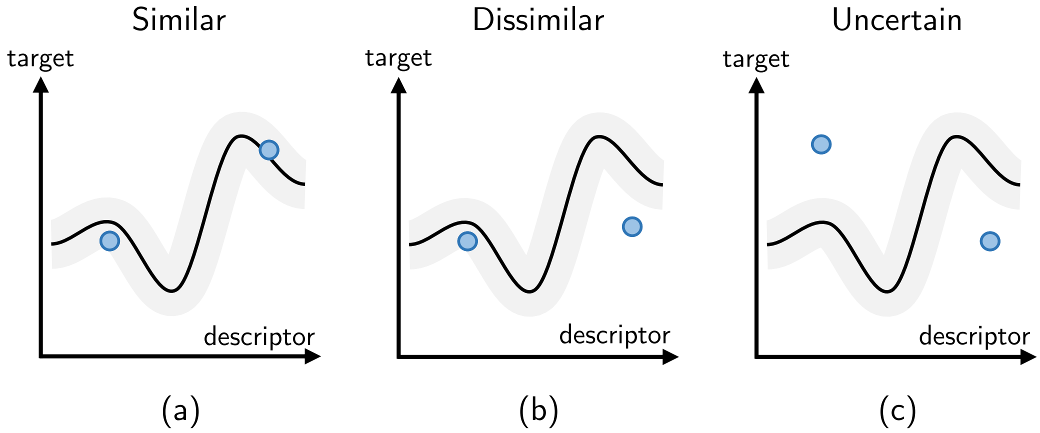


FIGURE 4.2: Illustrative figures of the three possible similarity states between two data instances (blue circles), including similar (a), dissimilar (b), and uncertain (c), considering a referential regression model f_r (black line). The gray region is the interval that determines whether a data instance can be considered to have been generated by regression model f_r.

study, we consider that $\mathcal{D}$ may contain pairs of data instances $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, where $\boldsymbol{x}_i \approx \boldsymbol{x}_j$; however, the value of y_i is far from y_j.

4.2.1 Collecting sources of similarity evidence

We perform random subset sampling of the data instances without replacement to collect a large amount of evidence of the similarity between pairs of data instances in $\mathcal{D}$. Considering each sample, we obtain two datasets: the reference dataset, $\mathcal{D}_{ref}$, and the evaluation dataset, $\mathcal{D}_{eval}$ ($\mathcal{D}_{ref} \cap \mathcal{D}_{eval} = \emptyset$ and $\mathcal{D}_{ref} \cup \mathcal{D}_{eval} = \mathcal{D}$). Considering $\mathcal{D}_{ref}$, we can generate a single or multiple reference functions $f_r : \mathbb{R}^n \to \mathbb{R}$ using a Gaussian process (GP) (Williams and Rasmussen, 1996) or a mixture of Gaussian processes (MGP) (Lázaro-Gredilla, Van Vaerenbergh, and Lawrence, 2012), respectively. This study applies GP- or MGP-based models instead of other nonlinear regression models such as kernel ridge regression (Vovk, 2013), random forest regression (Breiman, 2001), or artificial neural networks (Jain, Mao, and Mohiuddin, 1996) because GP or MGP can quantify the uncertainty of its prediction without introducing any other statistical validation. The sampling ratios of $\mathcal{D}_{ref}$ from $\mathcal{D}$ are fixed at 0.3 and 0.7 for the experiments with GP and MGP, respectively. Each reference function f_r is considered as a source to provide pieces of evidence for the similarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$ in $\mathcal{D}_{eval}$. The function f_r is not used to provide any information about the similarities between the data instances in $\mathcal{D}_{ref}$ or between a data instance in $\mathcal{D}_{ref}$ and a data instance in $\mathcal{D}_{eval}$. This is to exclude self-evaluation to ensure the objectivity of the evidence. Regarding a reference function f_r, we consider the state of the similarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$ as:

- Similar: Both data instances can be considered to have been generated by the function f_r (Fig. 4.2 a).
- Dissimilar: Only one of the data instances can be considered to have been generated by the function f_r (Fig. 4.2 b).

- Uncertain: Neither of the data instances can be considered to have been generated by the function f_r (Fig. 4.2 c). The uncertain state indicates that f_r does not provide any information about the similarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$.

To quantitatively evaluate whether $(\boldsymbol{x}_i, y_i)$ can be considered to have been generated by the regression function f_r, we use the likelihood $p(O_i|f_r)$, the probability of event O_i that a data instance $(\boldsymbol{x}_i, y_i)$ is observed, considering f_r. The likelihood $p(O_i|f_r)$ is modeled using a normal distribution with mean and standard deviation depending on the predicted target value $\hat{y}_i = f_r(\boldsymbol{x}_i)$ and the corresponding standard error $\sigma_{\boldsymbol{x}_i}$ by f_r, respectively. This is expressed as:

$$p(O_i|f_r) = \begin{cases} 1 & \text{if } \Delta_i \leq 3\,\bar{\sigma} \\ 2 \times \int_{\Delta_i - 3\,\bar{\sigma}}^{+\infty} \mathcal{N}\left(u|0, \alpha\, \sigma_{\boldsymbol{x}_i}\right) du & \text{otherwise} \end{cases}, \tag{4.1}$$

where $\Delta_i = |y_i - \hat{y}_i| = |y_i - f_r(\boldsymbol{x}_i)|$ is the deviation from the true to the predicted target values of data instance i using f_r, and $\bar{\sigma}$ is the average of the predictive standard error of all the data instances in $\mathcal{D}_{ref}$. α is the hyperparameter used to adjust the condition that restricts the data instances belonging to the function f_r. In other words, the interval that determines the probability that a data instance $(\boldsymbol{x}_i, y_i)$ belongs to f_r is $\alpha\,\sigma_{\boldsymbol{x}_i}$, and if the data instance falls outside this interval, it is determined that it does not belong to f_r. By increasing or decreasing the value of the parameter α, the condition for determining whether a data instance $(\boldsymbol{x}_i, y_i)$ belongs to f_r is relaxed or tightened, making $p(O_i|f_r)$ larger or smaller, respectively. Optimal values of α can be chosen using statistical criteria and appropriate validation methods; however, we set $\alpha = 2$ for all experiments in this work to reduce model complexity. We consider $p(O_i|f_r)$ as the probability that $(\boldsymbol{x}_i, y_i)$ is generated by f_r, and $p(\overline{O}_i|f_r) = 1 - p(O_i|f_r)$ is the probability that $(\boldsymbol{x}_i, y_i)$ is not generated by f_r. Figure 4.3 illustrates the process of modeling the probability $p(O_i|f_r)$.

Events where $(\boldsymbol{x}_i, y_i)$ or $(\boldsymbol{x}_j, y_j)$ is generated by the function f_r are independent events. Therefore, considering the function f_r, we can evaluate the joint probabilities of observing:

- Both data instances:

$$p(O_i, O_j|f_r) = p(O_i|f_r) \times p(O_j|f_r); \tag{4.2}$$

- Only one of the data instances:

$$\begin{aligned} &p(O_i, \overline{O}_j|f_r) + p(\overline{O}_i, O_j|f_r) \\ &\quad = p(O_i|f_r) \times p(\overline{O}_j|f_r) + p(\overline{O}_i|f_r) \times p(O_j|f_r); \end{aligned} \tag{4.3}$$

- Neither of the data instances:

$$\begin{aligned} &p(\overline{O}_i, \overline{O}_j|f_r) = p(\overline{O}_i|f_r) \times p(\overline{O}_j|f_r) \\ &\quad = 1 - p(O_i, O_j|f_r) - p(O_i, \overline{O}_j|f_r) - p(\overline{O}_i, O_j|f_r). \end{aligned} \tag{4.4}$$

4.2.2 Modeling evidence by mass functions

Considering the Dempster–Shafer theory framework (Shafer, 1976), we begin by defining the frame of discernment Ω. Let $\Omega = \{s, ds\}$ be the universal set representing the similarity states of any two data instances $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$. s and ds

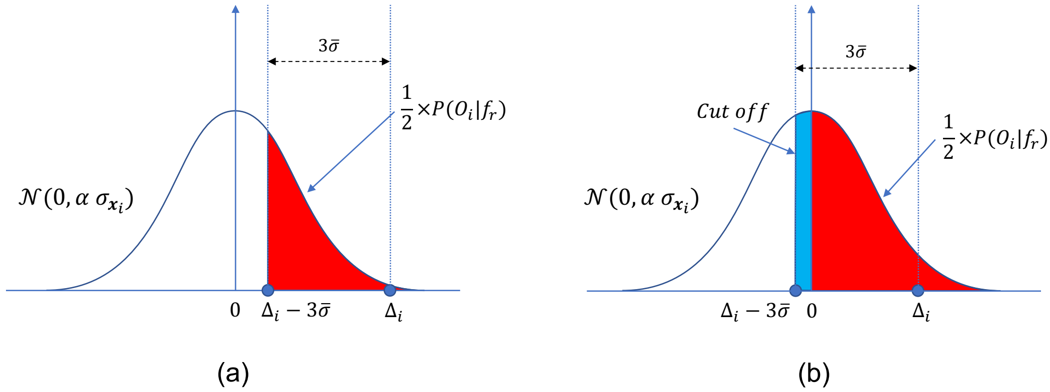


FIGURE 4.3: Illustration of probability $P(O_i|f_r)$ that a new data instance $(\boldsymbol{x}_i, y_i)$ is observed, considering f_r. We model the probability using a normal distribution with mean of zero and standard deviation of $\alpha\, \sigma_{\boldsymbol{x}_i}$, which are derived by f_r. Subplot (a) shows the probability $p(O_i|f_r)$ as deviation ($\Delta_i = |y_i - \hat{y}_i| = |y_i - f_r(\boldsymbol{x}_i)|$) from the true to the predicted target values of data instance i using f_r is less than $3\,\bar{\sigma}$. The $\bar{\sigma}$ is the average of the predictive standard error of all the training data instances. In contrast, subplot (b) illustrates modeling the probability as Δ_i exceeds $3\,\bar{\sigma}$. In each subplot, red area under the curve indicate $\frac{1}{2} \times P(O_i|f_r)$. Blue area in subplot (b) show cut off of the probability. In other words, we consider the probability $P(O_i|f_r)$ equal to 1 as its prediction error less than $3\,\bar{\sigma}$.

denote the similarity and dissimilarity states between the two data instances, respectively.

According to the Dempster–Shafer theory, the evidence of the similarity states between these two data instances is represented by a mass function $m^{i,j}$ (or a basic probability assignment) (Shafer, 1976). This assigns probability masses to all the nonempty subsets of Ω ($\mathcal{X} = \{\{s\}, \{ds\}, \{s, ds\}\}$). It is defined as follows:

$$m^{i,j} : \mathcal{X} \to [0, 1] \text{ with } \sum_{E \in \mathcal{X}} m(E) = 1. \tag{4.5}$$

The masses assigned to $\{s\}$ and $\{ds\}$ reflect the degrees of belief exactly committed to the evidence to support the similarity and dissimilarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$, respectively. The weight assigned to $\{s, ds\}$ expresses the degree of belief that the evidence provides no information about the similarity (or dissimilarity) between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$.

Therefore, the mass function $m^{i,j}_{f_r}$, which models a piece of evidence of the similarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$ collected from f_r, is defined as follows:

$$m^{i,j}_{f_r}(\{s\}) = \frac{p(O_i, O_j|f_r)}{\gamma_{i,j}} \tag{4.6}$$

$$m^{i,j}_{f_r}(\{ds\}) = \frac{p(O_i, \overline{O}_j|f_r) + p(\overline{O}_i, O_j|f_r)}{\gamma_{i,j}} \tag{4.7}$$

$$m^{i,j}_{f_r}(\{s, ds\}) = 1 - \frac{1}{\gamma_{i,j}} + \frac{p(\overline{O}_i, \overline{O}_j|f_r)}{\gamma_{i,j}}, \tag{4.8}$$

where $\gamma_{i,j} = (e^{\frac{\bar{\sigma}}{\Delta_y}} + 1) \times (\frac{\sigma_{\boldsymbol{x}_i}}{\bar{\sigma}} + 1) \times (\frac{\sigma_{\boldsymbol{x}_j}}{\bar{\sigma}} + 1)$ is a discounting factor (Shafer, 1976; Smets, 1993), which describes the unreliability of evidence about the similarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$ collected from a source of evidence f_r. Δ_y is the variation range of the target variable y in the dataset $\mathcal{D}$. The smaller $\bar{\sigma}$ is relative to Δ_y, the more reliable the learned regression function f_r is. Also, when $\sigma_{\boldsymbol{x}_i}$ and $\sigma_{\boldsymbol{x}_j}$ are smaller than $\bar{\sigma}$, f_r can provide reliable evidence for the relationship between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$. By contrast, when $\sigma_{\boldsymbol{x}_i}$ and $\sigma_{\boldsymbol{x}_j}$ are large compared to $\bar{\sigma}$, f_r cannot provide reliable evidence for the relationship between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$. A detailed explanation of each component in $\gamma_{i,j}$ is provided in Section D.1.

4.2.3 Dempster's rule in combining evidence

Assuming that we can collect q pieces of evidence from $\mathcal{F}_r = \{f_r^1, \ldots, f_r^q\}$, a set of q reference functions is generated from $\mathcal{D}$ to evaluate the similarity between a pair of data instances with indices i and j. According to the Dempster–Shafer theory framework, any two pieces of evidence collected from the reference functions f_r^l and f_r^k, which are modeled by the corresponding mass functions $m_{f_r^l}^{i,j}$ and $m_{f_r^k}^{i,j}$, respectively, can be combined using the Dempster rule of combination to assign the joint mass $m_{\{f_r^l, f_r^k\}}^{i,j}$ to each nonempty subset E of Ω as follows:

$$\begin{aligned} m_{\{f_r^l, f_r^k\}}^{i,j}(E) &= (m_{f_r^l}^{i,j} \oplus m_{f_r^k}^{i,j})(E) \\ &= \frac{\sum\limits_{E_t \cap E_v = E} m_{f_r^l}^{i,j}(E_t) \times m_{f_r^k}^{i,j}(E_v)}{1 - \sum\limits_{E_t \cap E_v = \emptyset} m_{f_r^l}^{i,j}(E_t) \times m_{f_r^k}^{i,j}(E_v)}, \end{aligned} \tag{4.9}$$

where E, E_t, and E_v are nonempty subsets of Ω. Dempster's rule is commutative and associative.

Based on Dempster's rule, the obtained mass functions corresponding to the q pieces of evidence are combined to assign the final mass $m_{\mathcal{F}_r}^{i,j}$ as follows:

$$m_{\mathcal{F}_r}^{i,j}(E) = \left(m_{f_r^1}^{i,j} \oplus m_{f_r^2}^{i,j} \oplus \cdots \oplus m_{f_r^q}^{i,j} \right)(E). \tag{4.10}$$

We perform similar analyses for all pairs of data instances in $\mathcal{D}$ to construct symmetric matrices M comprising the similarities ($M[i,j] = M[j,i] = m_{\mathcal{F}_r}^{i,j}(\{s\})$) between them. Thereafter, the obtained matrix is applied for further unsupervised data mining analysis, such as clustering or data visualization.

4.3 Case study 1: Revealing the similarity between transition-rare earth metal binary alloys

4.3.1 Binary dataset

Binary alloys dataset $\mathcal{D}_{binary}$ (Hieu-Chi, 2023): A material dataset containing 100 transition-rare earth metal binary alloys, comprising nickel (Ni), manganese (Mn), cobalt (Co), or iron (Fe), and the corresponding Curie temperatures (T_C). This dataset was collected from the Atomwork database of the National Institute of Materials Science (Villars et al., 2004; Xu, Yamazaki, and Villars, 2011). Each binary alloy in $\mathcal{D}_{binary}$ is represented using seven descriptors: (1,2) the atomic number of transition metal

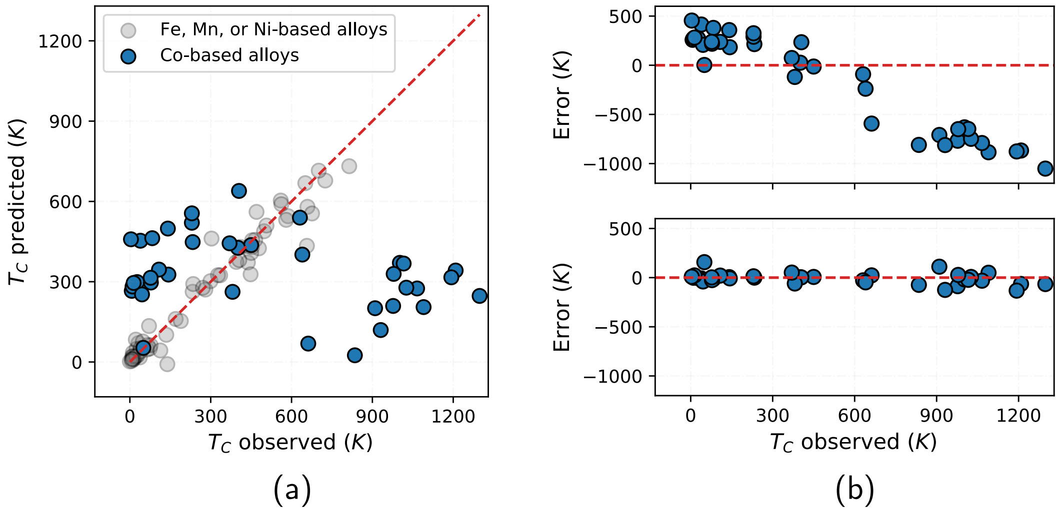


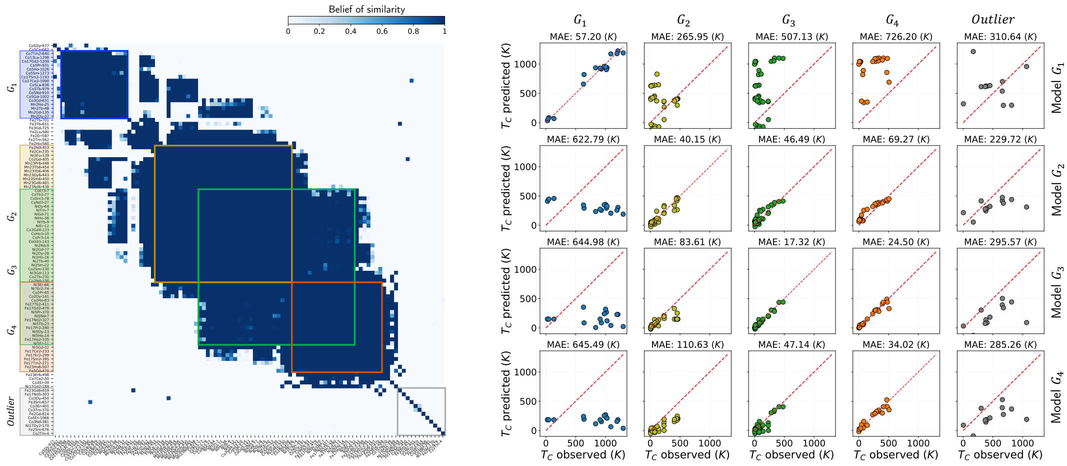
FIGURE 4.4: (a) Observed and predicted Curie temperature of alloys in the dataset $\mathcal{D}_{binary}$ using model generated for nickel (Ni), iron (Fe), and manganese (Mn)-based alloys. The blue and gray points indicate cobalt (Co)-based alloys and alloys of other transition metals (Ni, Fe, Mn), respectively. (b) Prediction error of Co-based alloys when excluding (top) or including (bottom) data of other Co-based alloys to the training dataset.

(Z_T) and rare-earth (Z_R) constituents; (3) projection of the spin magnetic moment onto the total angular moment of the $4f$ elections ($J_{4f}\left(1-g_j\right)$); (4, 5) covalent radius (r_{covT}) and first ionization (IP_T) of the transition metal; (6, 7) concentration of the transition metal (C_T) and rare-earth metal (C_R). The selection of these seven descriptors has been discussed in detail in previous studies (Nguyen et al., 2018b; Dam et al., 2018).

4.3.2 Assessment of the similarity between transition-rare earth metal binary alloys based on mechanisms of Curie temperature

In the first experiment, we demonstrate the versatility of the eRSM in detecting outliers and identifying mixtures of mechanisms. We apply the eRSM to assess the similarities between 100 transition rare earth metal binary alloys comprising nickel (Ni), manganese (Mn), cobalt (Co), or iron (Fe) based on their Curie temperatures. We construct a regression model using a Gaussian process by considering the data instances in $\mathcal{D}_{binary}$. This regression model achieves high prediction accuracy, as indicated by an R^2 score of 0.963 and a mean absolute error (MAE) of 40 (K) in ten-fold cross-validation. However, the reliability of this nonparametric regression model in subsequent exploratory predictions is not guaranteed due to the relatively small number of observable alloys compared to the number of possible alloys.

Figure 4.4 a illustrates the results of the exploratory prediction of the Curie temperature for Co-based binary alloys in $\mathcal{D}_{binary}$ using a Gaussian process regression model constructed from the data of binary alloys of Ni, Mn, and Fe. The regression model trained on the data of Ni, Mn, and Fe alloys demonstrates high prediction accuracy in ten-fold cross-validation, with an R^2 value of 0.946 and an MAE of 35 (K). However, the model tends to underestimate the Curie temperature of Co-based alloys with high values while overestimating the Curie temperature of other Co-based alloys. Significantly, when additional data of Co-based alloys are included in



(a) (b)

FIGURE 4.5: (a) Heatmap illustrating the similarity matrix M_{binary} extracted for all the data instances in the $\mathcal{D}_{binary}$. (b) Confusion matrices measuring the regression-based similarities between alloys in four groups G_1-G_4 and the dissimilarities between the models generated for alloys in different groups.

the training set, the prediction error for Co-based alloys is reduced (Figure 4.4 b). This observation supports the hypothesis that the underlying mechanisms governing the Curie temperature differ between Co-based alloys and alloys of other transition metals. The eRSM enables the identification and characterization of mixture mechanisms within this dataset, providing valuable insights into the similarities and differences between materials.

After applying the eRSM on the dataset $\mathcal{D}_{binary}$, we obtain a similarity matrix M_{binary} that reveals moderately high similarity values among the data instances (Figure 4.5 a). This indicates that most of the data instances can be effectively regressed by a relatively smooth function, which aligns with the high prediction accuracy observed in ten-fold cross-validation for all the alloys in the dataset. To facilitate the exploratory data analysis and avoid false intuitions or misunderstandings, we group the alloys in $\mathcal{D}_{binary}$ based on their high similarities within each group. It is important to note that an alloy can belong to multiple groups simultaneously or not belong to any group. By applying a graph-based clustering method, we identify four distinct groups of alloys denoted as G_1, G_2, G_3, and G_4, which exhibit high intra-group similarities exceeding 0.7 (Figure 4.5 a). However, the similarities between alloys in group G_1 and those in G_2, G_3, and G_4 are significantly dissimilar. Additionally, a small group of alloys (a gray region in Figure 4.5 a) shows distinct differences from all the other alloys and can be considered outliers. The remaining alloys are not assigned to any specific group, indicating uncertainty in the clustering analysis results.

To quantitatively evaluate the validity of the analysis process, we train regression models for T_C using the data from each of the four groups G_1, G_2, G_3, and G_4, and assess their prediction accuracy on these groups. The confusion matrix in Figure 4.6 summarizes the correlation between the observed and predicted T_C values by the four learned regression models. The diagonal plots represent the cross-validation results of the models trained on the alloys from the respective groups. The off-diagonal plot shows the correlation between the observed T_C values and the predictions made by the model trained on alloys from the other groups. The results confirm the high

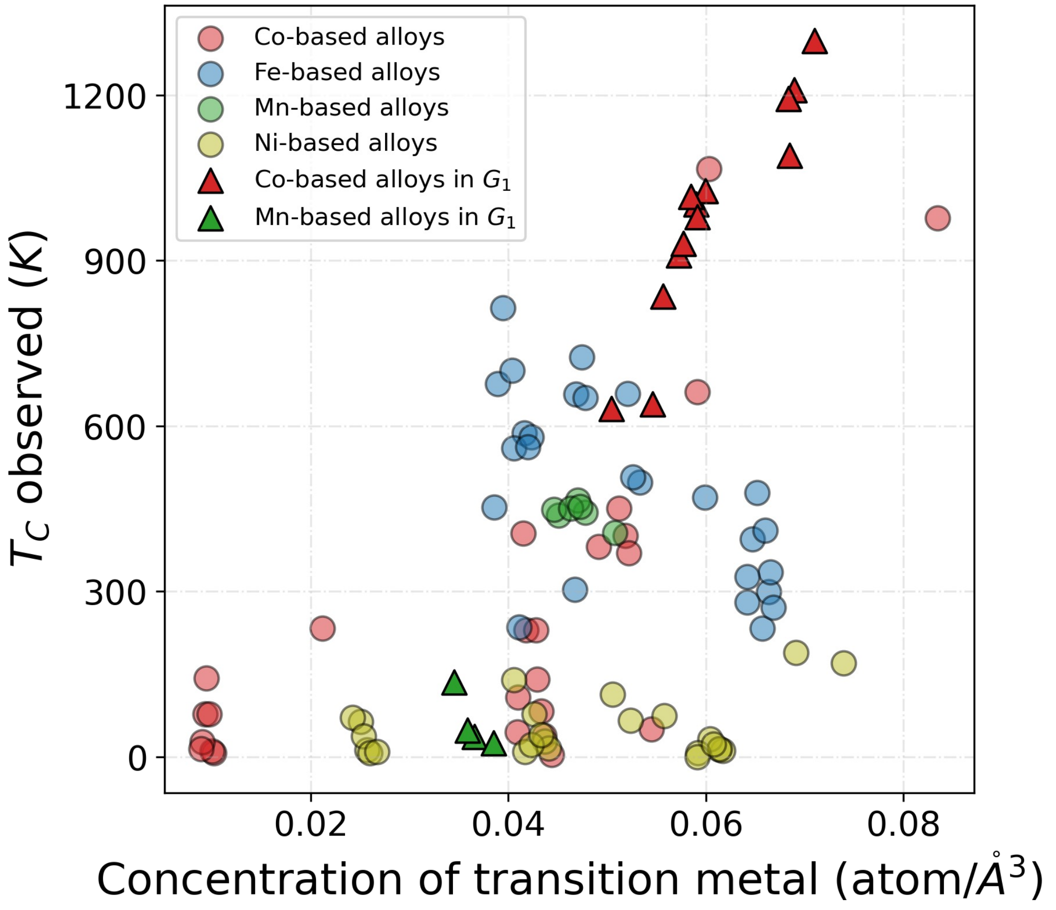


FIGURE 4.6: Dependence of T_C on the concentration of the transition metal (C_T) in alloys. Red, blue, green and yellow scatters indicate alloys containing cobalt (Co), iron (Fe), manganese (Mn), and nickel (Ni). Alloys in G_1 are highlighted by triangles.

intra-group similarity among the alloys in groups G_1, G_2, G_3, and G_4, respectively, as well as the dissimilarity between the four groups. Furthermore, the alloys considered outliers exhibit significant dissimilarity from all the other groups. These findings suggest that the physical mechanisms governing the T_C of alloys in group G_1 may differ from those in groups G_2, G_3, and G_4. However, it remains challenging to determine the specific differences in the mechanisms of T_C among the alloys in groups G_2, G_3, and G_4.

Moreover, when considering the alloys in G_1, a strong linear correlation is observed between T_C and the concentration of transition metals in the alloys, with a Pearson correlation coefficient of 0.95 (Figure 4.6, triangle scatters). This finding aligns with the observations of previous research (Dam et al., 2018) that investigated all binary alloys of transition metals and rare earth metals in $\mathcal{D}_{binary}$. The range of T_C is found to correlate with the transition metals' composition ratio. Furthermore, out of the 17 alloys in G_1, 13 are Co-based alloys with high Curie temperatures ($T_C > 600$ K). In contrast, most of the other Co-based alloys in $\mathcal{D}_{binary}$ have lower Curie temperatures ($T_C < 500$ K) and are assigned to groups G_2, G_3, and G_4. These results are consistent with the observation that the regression model for Fe-, Mn-, and Ni-based alloys tend to underestimate the T_C of Co-based alloys with high T_C and overestimate the T_C of the remaining Co-based alloys (Figure 4.4 a).

Additionally, we conducted experiments on synthetic toy datasets that include outliers and multiple mechanisms to evaluate the performance of the eRSM in assessing similarity. The detailed results of these experiments are summarized in Appendix C. In summary, the eRSM demonstrates its effectiveness in assessing the similarity between data instances and successfully detects outliers and mixtures of mechanisms. These results validate the efficiency of the proposed similarity measure.

4.4 Case study 2: Revealing the similarity between quaternary alloys with respect to Curie temperature and Magnetization

4.4.1 Quaternary alloy dataset

Quaternary high-entropy alloys datasets $\mathcal{D}_{quaternary}$ (Hieu-Chi, 2023): A material dataset contains 990 equiatomic quaternary high-entropy alloys, which comprise 14 transition metals $\{Ag, Cd, Co, Cr, Cu, Fe, Mn, Mo, Ni, Pd, Rh, Ru, Tc, Zn\}$, and the corresponding calculated magnetizations and Curie temperatures in the BCC phase. The dataset was collected from an original dataset of $147,630$ equiatomic quaternary high-entropy alloys calculated using Korringa-Kohn-Rostoker coherent approximation method (Fukushima et al., 2022). Each alloy in $\mathcal{D}_{quaternary}$ is represented using 135 compositional descriptors, including the means, standard deviations, and covariance of the atomic representations of their constituent elements (Seko, Togo, and Tanaka, 2018) and four categorical features indicating the elements comprising the quaternary alloy. We use 15 atomic representations to generate these 135 descriptors: (1) atomic number, (2) atomic mass, (3) period and (4) group in the periodic table, (5) first ionization energy, (6) second ionization energy, (7) Pauling electronegativity, (8) Allen electronegativity, (9) van der Waals radius, (10) covalent radius, (11) atomic radius, (12) melting point, (13) boiling point, (14) density, and (15) specific heat.

It is desirable to reduce the number of descriptors both to reduce the computational cost of modeling and, in some cases, to improve model performance. We adopt a statistics-based feature selection method as a filter to remove redundancy descriptors. This method evaluates the Pearson coefficient between pairs of descriptors and removes descriptors that exhibit strong relationships with the other ones. For instance, given a pair of descriptors d_i and d_j, we remove d_i if the Pearson coefficient between d_i and d_j exceeds 0.8 and otherwise. Table 4.1 shows 36 selected descriptors, which are used to learn regression models in this study.

TABLE 4.1: Descriptors for modeling expert.

Features	Definition
$\mu_{\text{atomic number}}$	Mean of atomic numbers of elements comprising the alloy
$\mu_{\text{group index}}$	Mean of group index of elements comprising the alloy
$\mu_{\text{second ionization energies}}$	Mean of second ionization energies of elements comprising the alloy

Continued on next page

Table 4.1 – *Continued from previous page*

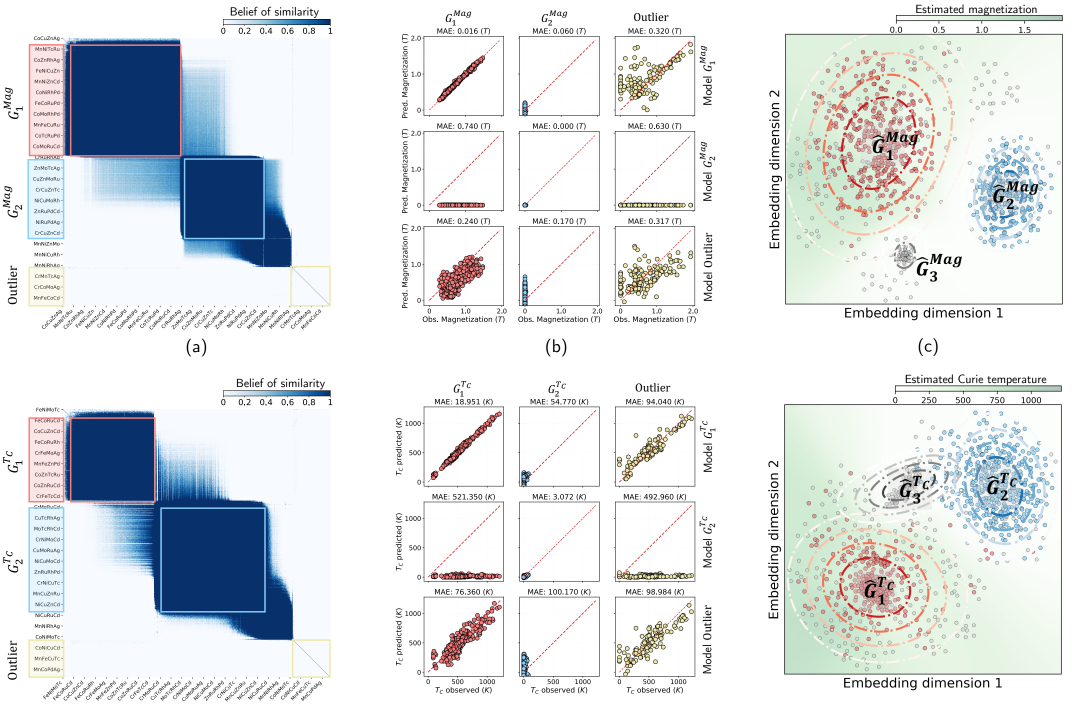
Features	**Definition**
$\mu_{\text{pauling electronegativity}}$	Mean of pauling electronegativity of elements comprising the alloy
$\mu_{\text{allen electronegativity}}$	Mean of allen electronegativity of elements comprising the alloy
$\mu_{\text{covalent radius cordero}}$	Mean of covalent radius cordero of elements comprising the alloy
$\mu_{\text{atomic radius}}$	Mean of atomic radius of elements comprising the alloy
$\sigma_{\text{atomic number}}$	Standard deviation of atomic number of elements comprising the alloy
$\sigma_{\text{group index}}$	Standard deviation of group index of elements comprising the alloy
$\sigma_{\text{first ionization energies}}$	Standard deviation of first ionization energies of elements comprising the alloy
$\sigma_{\text{second ionization energies}}$	Standard deviation of second ionization energies of elements comprising the alloy
$\sigma_{\text{pauling electronegativity}}$	Standard deviation of pauling electronegativity of elements comprising the alloy
$\sigma_{\text{allen electronegativity}}$	Standard deviation of allen electronegativity of elements comprising the alloy
$\sigma_{\text{van der Waals radius}}$	Standard deviation of van der Waals radius of elements comprising the alloy
$\sigma_{\text{covalent radius cordero}}$	Standard deviation of covalent radius cordero of elements comprising the alloy
$\sigma_{\text{atomic radius}}$	Standard deviation of atomic radius of elements comprising the alloy
$\sigma_{\text{melting points}}$	Standard deviation of melting points of elements comprising the alloy
$\text{cov}^{\text{atomic number}}_{\text{group index}}$	Covariance between atomic number and group index of elements comprising the alloy
$\text{cov}^{\text{atomic number}}_{\text{second ionization energies}}$	Covariance between atomic number and second ionization energies of elements comprising the alloy
$\text{cov}^{\text{atomic number}}_{\text{pauling electronegativity}}$	Covariance between atomic number and pauling electronegativity of elements comprising the alloy
$\text{cov}^{\text{atomic number}}_{\text{allen electronegativity}}$	Covariance between atomic number and allen electronegativity of elements comprising the alloy

Continued on next page

Table 4.1 – *Continued from previous page*

Features	**Definition**
$\mathrm{cov}^{\text{atomic number}}_{\text{covalent radius cordero}}$	Covariance between atomic number and covalent radius cordero of elements comprising the alloy
$\mathrm{cov}^{\text{group index}}_{\text{second ionization energies}}$	Covariance between group index and second ionization energies of elements comprising the alloy
$\mathrm{cov}^{\text{group index}}_{\text{allen electronegativity}}$	Covariance between group index and allen electronegativity of elements comprising the alloy
$\mathrm{cov}^{\text{group index}}_{\text{covalent radius cordero}}$	Covariance between group index and covalent radius cordero of elements comprising the alloy
$\mathrm{cov}^{\text{group index}}_{\text{atomic radius}}$	Covariance between group index and atomic radius of elements comprising the alloy
$\mathrm{cov}^{\text{first ionization energies}}_{\text{second ionization energies}}$	Covariance between first and second ionization energies of elements comprising the alloy
$\mathrm{cov}^{\text{first ionization energies}}_{\text{atomic radius}}$	Covariance between first ionization energies and atomic radius of elements comprising the alloy
$\mathrm{cov}^{\text{second ionization energies}}_{\text{pauling electronegativity}}$	Covariance between second ionization energies and pauling electronegativity of elements comprising the alloy
$\mathrm{cov}^{\text{second ionization energies}}_{\text{allen electronegativity}}$	Covariance between second ionization energies and allen electronegativity of elements comprising the alloy
$\mathrm{cov}^{\text{second ionization energies}}_{\text{covalent radius cordero}}$	Covariance between second ionization energies and covalent radius cordero of elements comprising the alloy
$\mathrm{cov}^{\text{second ionization energies}}_{\text{atomic radius}}$	Covariance between second ionization energies and atomic radius of elements comprising the alloy
$\mathrm{cov}^{\text{pauling electronegativity}}_{\text{allen electronegativity}}$	Covariance between pauling electronegativity and allen electronegativity of elements comprising the alloy
$\mathrm{cov}^{\text{pauling electronegativity}}_{\text{covalent radius cordero}}$	Covariance between covalent radius cordero and allen electronegativity of elements comprising the alloy
$\mathrm{cov}^{\text{pauling electronegativity}}_{\text{atomic radius}}$	Covariance between pauling electronegativity and atomic radius of elements comprising the alloy

Continued on next page



(d) (e) (f)

FIGURE 4.7: (a,d) Heatmaps illustrating the similarity matrices $M^{Mag}_{quaternary}$ (a) and $M^{T_C}_{quaternary}$ (d) extracted from datasets $\mathcal{D}^{Mag}_{quaternary}$ and $\mathcal{D}^{T_C}_{quaternary}$, focusing on mechanisms of magnetization and T_C, respectively. (b,e) The confusion matrix summarizes the differences between the magnetization (b) or T_C (e) mechanisms of alloys in extracted groups. (c,f) Visualization of quaternary alloys in the two-dimensional embedding spaces constructed by applying the T-distributed Stochastic Neighbor Embedding (t-SNE) to $M^{Mag}_{quaternary}$ (c) and $M^{T_C}_{quaternary}$ (f). Red, blue, and gray contours indicate gaussian models $\hat{G}_1^{Mag}$ ($\hat{G}_1^{T_C}$), $\hat{G}_2^{Mag}$ ($\hat{G}_2^{T_C}$), and $\hat{G}_3^{Mag}$ ($\hat{G}_3^{T_C}$), respectively, learned by using the Gaussian Mixture Models (Lindsay, 1995) in the embedding space focusing on mechanisms of magnetization (T_C). In addition, red and blue points in sub-figures b and c (e and f) indicate the alloys in G_1^{Mag} ($G_1^{T_C}$) and G_2^{Mag} ($G_2^{T_C}$), respectively.

Table 4.1 – *Continued from previous page*

Features	**Definition**
$\mathrm{cov}^{\text{covalent radius cordero}}_{\text{atomic radius}}$	Covariance between covalent radius cordero and atomic radius of elements comprising the alloy

4.4.2 Assessment of the similarity between quaternary high-entropy alloys based on mechanisms of magnetization

In the next experiment, we apply the eRSM to assess the similarities between 990 quaternary high-entropy alloys comprising 14 transition metals in the dataset $\mathcal{D}^{Mag}_{quaternary}$

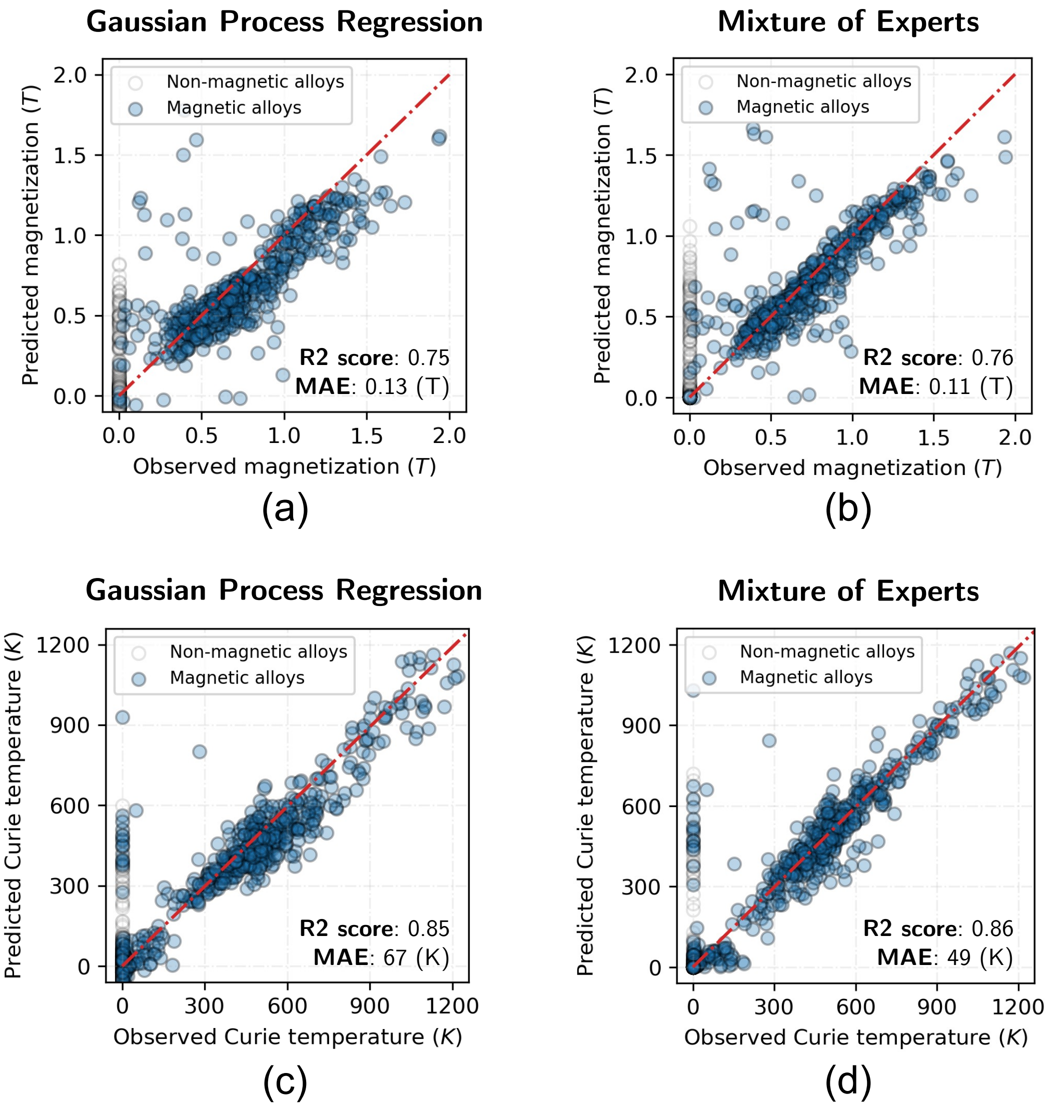


FIGURE 4.8: Prediction accuracies for magnetization (a, b) and Curie temperature (c, d) of the alloys with *10*-fold cross-validations. Prediction validation results with single gaussian process regression models for magnetization and Curie temperature are shown in sub-figures (a) and (c), respectively. Prediction validation results with mixtures of expert models for magnetization and Curie temperature are shown in sub-figures (b) and (d), respectively. Blue and white circles indicate magnetic alloys (finite magnetization) and non-magnetic alloys (zero magnetization), respectively.

based on their magnetization. We aim to construct an optimal Gaussian process regression model using the designed descriptors to predict the magnetization of these alloys. However, the Gaussian process regression model performs poorly, with an R^2 score of 0.75 and an MAE of 0.13 (T) in ten-fold cross-validation. This suggests that the magnetization of these alloys may not be accurately described by a single model in the designed descriptor space. It indicates the presence of outliers or a mixture of models for the magnetization properties of these alloys within the descriptor space.

Applying the eRSM to the dataset, we obtain a similarity matrix $M^{Mag}_{quaternary}$ that reveals two core groups of alloys, denoted as G_1^{Mag} and G_2^{Mag}, showing high intra-group similarities above 0.5 (Figure 4.7a). Some of the alloys in G_1^{Mag} are similar to those in G_2^{Mag}; however, the majority exhibit apparent dissimilarities. Additionally, a small group of alloys (Figure 4.7 a, yellow region) shows dissimilarities with the rest and can be considered outliers. The remaining alloys in $\mathcal{D}_{quaternary}^{Mag}$ do not exhibit significant similarities with the alloys in groups G_1^{Mag} and G_2^{Mag}, and therefore they are not assigned to any group.

To validate the obtained results quantitatively, we trained three regression models using data from each group, G_1^{Mag}, G_2^{Mag}, and outliers. We monitored the prediction accuracy of the three learned regression models for data in all the groups. The confusion matrix summarizing the correlations between the observed and predicted values of the target variable using the learned regression models is shown in Figure 4.7 c. The diagonal plots illustrate the ten-fold cross-validation results of the models learned from these three groups of alloys. In contrast, the off-diagonal plot shows the correlation between the observed magnetization and the predictions made by the model learned from the alloys of the other groups. The results confirm the intra-group similarity of the alloys in groups G_1^{Mag} and G_2^{Mag}, respectively, as well as the dissimilarity between the two groups and the intra-group dissimilarity of the alloys considered as outliers. Specifically, we observe that group G_2^{Mag} consists of ferrimagnetic alloys or alloys with magnetization relatively smaller (magnetization < 0.1 (T)) compared to the alloys in group G_1^{Mag}. On the other hand, when using the data from G_1^{Mag}, we can construct a Gaussian process regression model with high prediction accuracy, yielding an R^2 score of 0.992 and an MAE of 0.016 (T) in the *ten*-fold cross-validation.

Therefore, we can use the information of the constituent elements of each alloy to predict which group it belongs to in advance (Ha et al., 2021) and apply an appropriate regression model to improve prediction accuracy for the alloys. We combine the similarity measured by using the eRSM with the Jaccard similarity coefficient (Murphy, 1996) and apply the T-distributed Stochastic Neighbor Embedding (Maaten and Hinton, 2008) (t-SNE) to construct a two-dimensional embedding map (Fig. 4.7 c). Details of the combination method are shown in Section D.2. As a result, we can easily distinguish the alloys in groups G_1^{Mag} (red) and G_2^{Mag} (blue) when they form two separate regions with high density in the embedding space. We apply a Gaussian mixture model (Lindsay, 1995) (GMM) on the embedding space to identify groups and calculate the probability of an alloy belonging to a particular identified group. Alloys in different groups are treated differently by using a mixture of experts (Pham et al., 2016b) (MoE) approach. Figure 4.8 a and b show a reduction of the proposed mixture of experts in MAE of 18% compared with result of the single model, from 0.13 (T) to 0.11 (T). Further analysis shows that applying the obtained similarities in MOE improves the prediction accuracy for magnetic alloys (Fig. 4.9 a).

4.4.3 Assessment of the similarity between the quaternary high-entropy alloys based on mechanisms of Curie temperature

Considering this experiment, the target data are the same as in the previous section ($\mathcal{D}_{quaternary}$); however, the physical property of interest is T_C. A regression model can be constructed using a Gaussian process. This shows a rather high prediction accuracy in *ten*-fold cross-validation with an R^2 score of 0.85 and an MAE of 67 (K). We also observe two distinguishable groups of quaternary alloys in the dataset

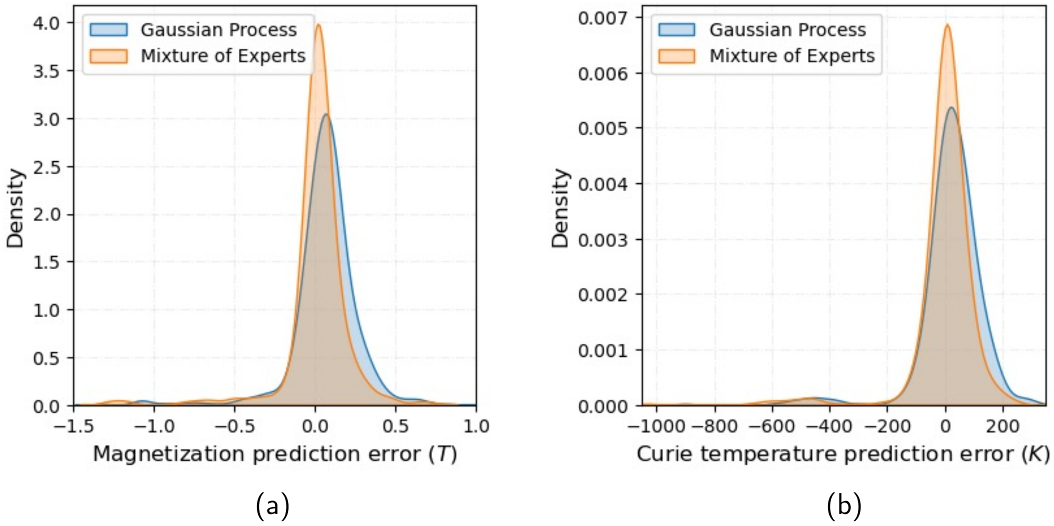


FIGURE 4.9: Error of prediction (observed value - predicted value) for magnetization (a) and curie temperature (b) of the magnetic alloys in $\mathcal{D}_{quaternary}$. Blue and orange lines in these plots indicate prediction errors of models learned by using Gaussian Process and Mixture of Experts, which used the similarities measured by the eRSM, respectively.

$\mathcal{D}^{T_C}_{quaternary}$ when applying the eRSM. Figure 4.7 d illustrates the similarity matrix $M^{T_C}_{quaternary}$ with two groups of alloys denoted as $G^{T_C}_1$ and $G^{T_C}_2$, showing high intra-group similarities and exceeding 0.5. Some of the alloys in $G^{T_C}_1$ are similar to those in $G^{T_C}_2$. Nonetheless, the others exhibit apparent dissimilarities, which is consistent with the observation of two high-density regions (red) in the embedding map of $M^{T_C}_{quaternary}$ (Fig. 4.7 e). Furthermore, a small group of alloys (Fig. 4.7 d, yellow region) showed dissimilarities with all the others and could be considered as outliers. The remaining alloys do not show apparent similarities with alloys in groups $G^{T_C}_1$ and $G^{T_C}_2$; thus, they are not assigned to any group.

Following the same analysis procedure as in the previous section, we trained regression models for Curie temperature using data from each of the three groups $G^{T_C}_1$, $G^{T_C}_2$, and outliers and monitored their prediction accuracy on these groups. Figure 4.7 f shows the confusion matrix that summarizes the obtained results. The diagonal plots illustrate the *ten*-fold cross-validation results of the models learned from these three groups of alloys. The off-diagonal plot shows the correlation between the observed Curie temperature and the predictions made by the regression model learned from the alloys of the other groups. We can also confirm the intra-group similarity of the alloys in groups $G^{T_C}_1$ and $G^{T_C}_2$, respectively, dissimilarity between the two groups, and intra-group dissimilarity of the alloys considered as outliers. Specifically, we observe that the Curie temperatures of approximately all the alloys in group $G^{T_C}_2$ have a low T_C, which is 0 (K) or relatively smaller than that of the other alloys. Furthermore, using the data in $G^{T_C}_1$, we can construct a Gaussian process regression model with a high prediction accuracy with an R^2 score of 0.985 and an MAE of 19 (K) in the *ten*-fold cross-validation.

Therefore, we utilize the similarity information to design descriptors for quaternary alloys due to the effectiveness of the data for detecting the mixture of multiple

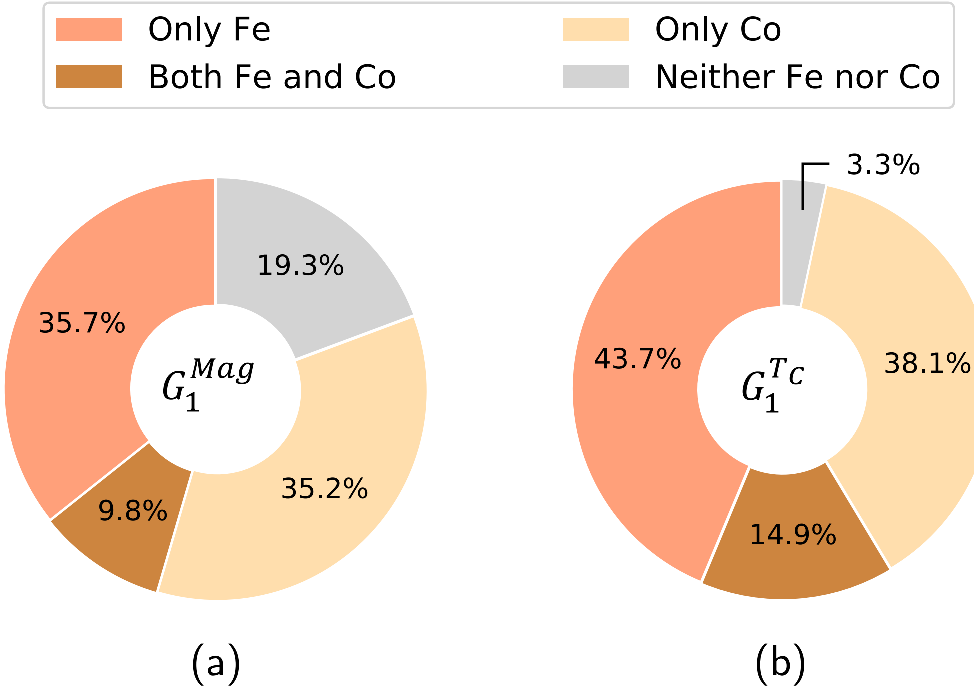


FIGURE 4.10: Proportions of quaternary alloys containing Fe or Co in group G_1^{Mag} (a) and $G_1^{T_C}$ (b).

mechanisms in the dataset. We apply similar method (Section D.2) as in the previous experiment to construct a two-dimensional embedding map (Fig. 4.7 f) and then learn a mixture of experts to predict Curie temperature of quaternary alloys in the dataset $\mathcal{D}^{T_C}_{quaternary}$. The proposed mixture of models exhibits higher prediction accuracy than the single model in *10*-folds cross-validations (Fig. 4.8 c and d). The MAE of the proposed mixture of expert reduces approximately 36%, from 67 (K) to 49 (K).

4.4.4 Discussion of the obtained similarities between materials and the associated physical mechanisms

In the experiments conducted on the datasets $\mathcal{D}^{Mag}_{quaternary}$ and $\mathcal{D}^{T_C}_{quaternary}$, which focus on magnetization and T_C respectively, we observed that these datasets provide a clear example where magnetization and T_C are cases sensitive to finite or zero values. As demonstrated in Sections 4.4.2 and 4.4.3, the prediction accuracy is low when considering a single regression model for the entire dataset. In this section, we shift our attention to analyzing the extracted alloy groups G_1^{Mag}, G_2^{Mag}, $G_1^{T_C}$, and $G_2^{T_C}$ to identify underlying patterns.

Figure 4.10 reveals that Fe and Co, which possess significant spin moments and exhibit ferromagnetic interactions with many elements, are dominant elements in alloys belonging to groups G_1^{Mag} (a) and $G_1^{T_C}$ (b). Furthermore, when considering the proportion of quaternary alloys with fixed two of their four constituent elements across the four extracted groups (G_1^{Mag}, G_2^{Mag}, $G_1^{T_C}$, and $G_2^{T_C}$), we observe that the proportion of Fe-containing and Co-containing alloys in groups G_1^{Mag} (a) and $G_1^{T_C}$ is significantly higher compared to the other groups (Fig. 4.11). Consequently, prediction models constructed using data from alloys in G_1^{Mag} or $G_1^{T_C}$ are better suited

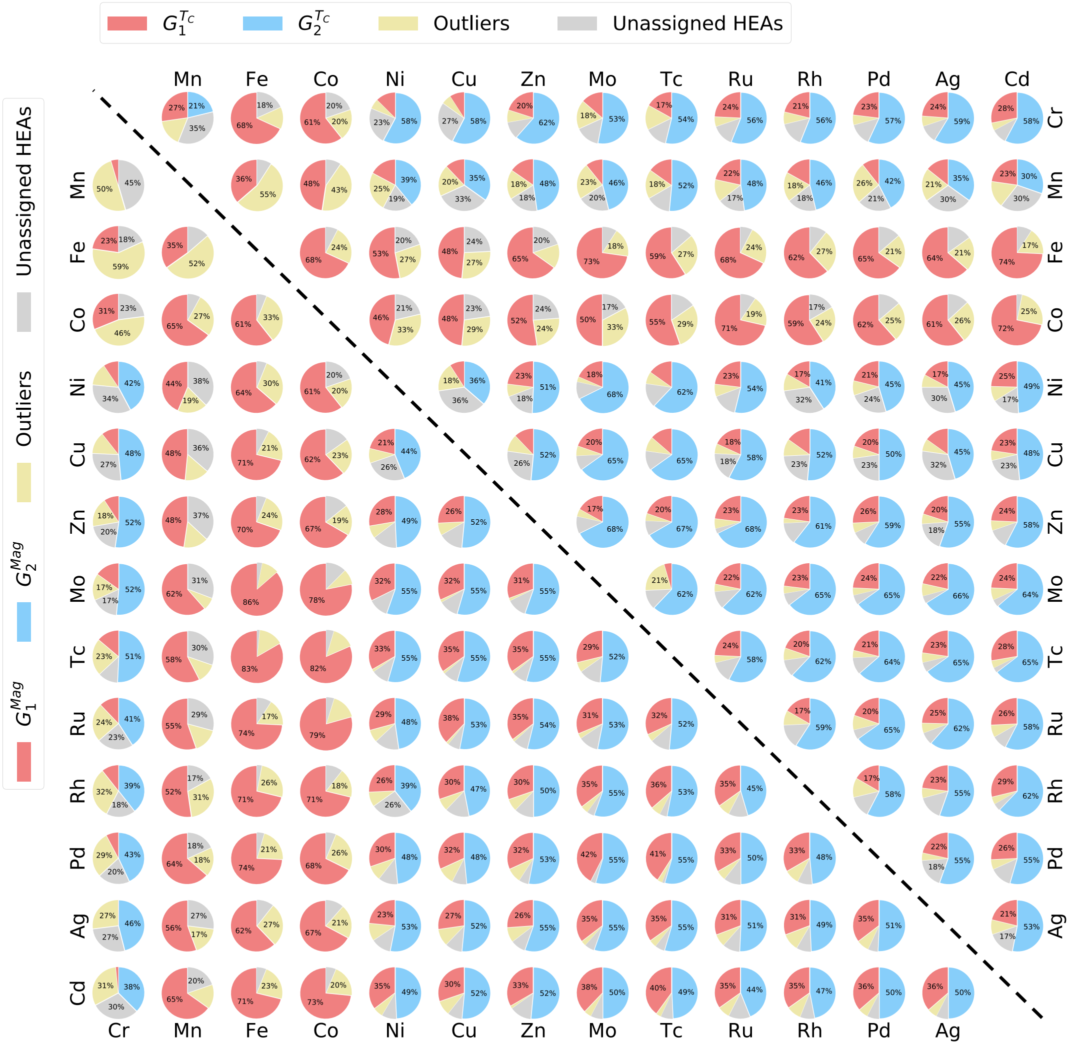


FIGURE 4.11: Effect of coexistence of the 14 transition metals on magnetization and Curie temperature mechanisms. Each pie chart results from quaternary alloys containing the respective element pair. They show the percentages of alloys that follow the magnetization mechanisms (lower-left triangle) and Curie temperature mechanisms (upper-right triangle), as extracted by the eRSM. Red and blue areas indicate the percentages of alloys whose magnetization and T_C are finite (G_1^{Mag} and $G_1^{T_C}$) and zero (G_2^{Mag} and $G_2^{T_C}$), respectively. Yellow areas indicate the percentages of alloys that are detected as outliers. By contrast, gray regions indicate the fractions of alloys not assigned to the extracted groups.

for predicting magnetization or T_C of alloys containing these elements, respectively. The remaining Fe-X and Co-X alloys are considered outliers in terms of the extracted mechanisms or unassigned high-entropy alloys that do not fall into any of these mechanisms. Conversely, Mn-X alloys exhibit behavior similar to Fe-X and Co-X alloys in the context of magnetization mechanisms. However, for Curie temperature, Mn-X alloys are categorized in group $G_2^{T_C}$ representing low T_C, distinct from the other groups. Notably, among the Fe-X and Co-X alloys, the percentage of Fe-Mn and Co-Mn alloys considered outliers of the mechanisms extracted from $G_1^{T_C}$ is relatively high, accounting for 55% and 43% respectively (Fig. 4.11).

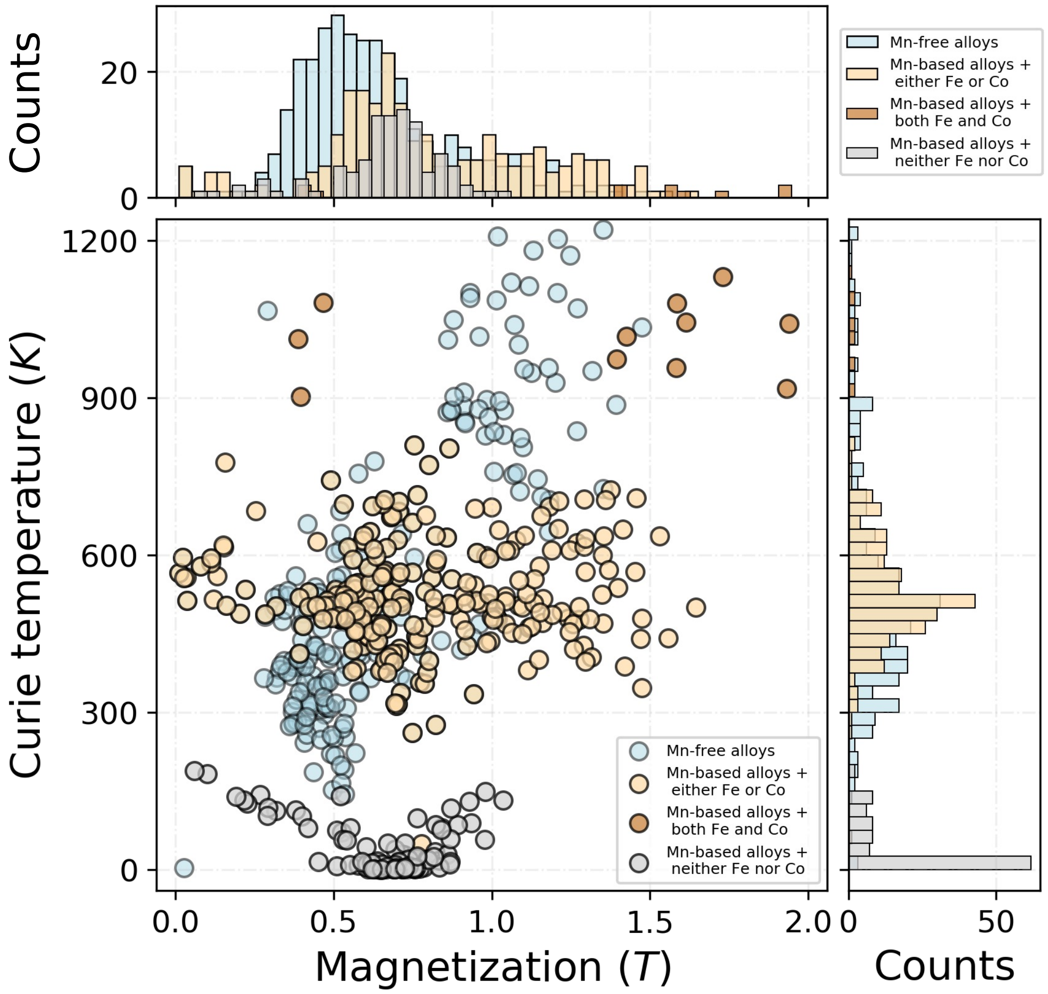


FIGURE 4.12: Correlation between magnetization (T) and Curie temperature (K) of quaternary alloys with non-zero magnetization and non-zero Curie temperature in datasets $\mathcal{D}^{Mag}_{quaternary}$ and $\mathcal{D}^{T_C}_{quaternary}$. Marginal plots show histogram of the properties of the alloys.

For further investigation, we organized the raw data of the quaternary alloys by focusing on the presence or absence of Mn. Figure 4.12 displays the correlation between magnetization and Curie temperature for 556 (56%) alloys with non-zero properties. Out of the total 990 data instances, 413 (42%) correspond to alloys with zero values for both T_C and magnetization, while there are twenty-one (2%) alloys with zero T_C but finite magnetization. We made several observations based on these findings. Alloys containing all three elements, Mn, Fe, and Co, exhibit high Curie temperatures ($T_C > 900$ (K)). On the other hand, alloys containing pairs of Mn-Fe or Mn-Co show moderate Curie temperatures. In contrast, Mn-containing alloys without Fe or Co display low Curie temperatures ($T_C < 250$ (K)). Furthermore, these trends do not correlate significantly between magnetization and Curie temperature. However, a clear positive correlation between magnetization and Curie temperature can be observed in the group of Mn-free alloys.

To interpret these results, we considered a hypothesis regarding the origin of the observed data. The estimated magnetization is determined by the sum of all local magnetic moments divided by the unit volume. These local magnetic moments are influenced by the spin configurations of atomic sites, which stabilize the structure of alloys. Conversely, the T_C can be estimated from the spin-spin exchange energy given a specific structure and spin configuration. First-principles calculations and exhaustive calculations in high-entropy alloys have revealed that early transition metals and late transition metals often exhibit antiferromagnetic interactions (Akai et al., 1990; Fukushima et al., 2022). Mn, being situated between early and late transition metals, requires cautious consideration of the spin configuration (ferromagnetic or antiferromagnetic) in Mn-containing alloys, especially in high-entropy alloys where elements can stochastically occupy the same atomic site. Based on this consideration, we hypothesize that alloys containing Mn may follow different rules for magnetization compared to those grouped into G_2^{Mag}. Conversely, Mn-containing alloys may follow the same rules as alloys grouped into $G_2^{T_C}$ regarding T_C, albeit with a spin configuration that yields magnetization. While further analysis is beyond the scope of this paper and will not be discussed here, these preliminary findings indicate promising avenues for future investigation.

4.5 Contributions and limitations

In this study, we developed a method called the evidetial regression-based similarity measure (eRSM) that can rationally transform material data from multiple sources into evidence of similarities between materials and combine the evidence to conclude the similarities. The extracted similarity-dissimilarity information has significant potential for application in the subgroup discovery of materials. We demonstrated the effectiveness of the eRSM in detecting homogenous subgroups of materials using two experiments on two datasets of magnetic materials.

Our method successfully revealed differences in the mechanisms of the Curie temperature of Co-based binary alloys when applied to a dataset of 100 transition-rare earth metal binary alloys comprising Ni, Mn, Co, and Fe. Furthermore, we explored the mechanisms of ferrimagnetic and low Curie temperature alloys from the magnetic dataset of calculated quaternary alloys. These results illustrate the potential of our method in extracting valuable information for describing and interpreting the underlying physical mechanisms in material datasets.

In addition, this approach can identify material subgroups in data that exhibit homogeneities in multivariate statistical relationships, providing hints for experts to adjust the current instance space $\mathcal{X}$ by analyzing and elucidating the underlying physical mechanisms. Therefore, we can reduce the effect of the aleatoric uncertainty in each decision. According to Hüllermeier and Waegeman, aleatoric and epistemic uncertainties are not fixed concepts, but rather depend on the specific setting (instance space $\mathcal{X}$, output space $\mathcal{Y}$, hypothesis space $\mathcal{H}$, and joint probability $\mathcal{P}$ on $\mathcal{X} \times \mathcal{Y}$). Changing the context can alter the sources of uncertainty, enabling aleatoric uncertainty to become epistemic uncertainty and thus reducible. This distinction between the two types of uncertainty becomes less clear when the learner is allowed to modify the setting, making their quantification more complex. Der Kiureghian and Ditlevsen also share this perspective on the relationship between aleatoric and epistemic uncertainty. They contend that these concepts can only be unambiguously defined within a particular analytical model, and uncertainties classified as aleatory in one model might be considered epistemic in another (Der Kiureghian and Ditlevsen, 2009). One possible approach to address aleatoric uncertainty is to extend the description of instances by incorporating additional features, effectively replacing the current instance space $\mathcal{X}$ with another space $\mathcal{X}'$. This change could impact the uncertainty and contribute a more accurate understanding of materials and their properties.

While the proposed method provides valuable information for understanding the underlying mechanisms in datasets, it requires human involvement and further analysis of the detected groups to design additional features for the instance space. Moreover, our method's application faces computational resource challenges when the number of data instances increases. For example, the experiment on the quaternary alloy dataset was conducted on a cluster of 288 cores, 2.3 TB RAM, CPU Intel(R) Xeon(R) CPU E5-2697A v4 @ 2.60GHz, and it took 2.5 months to complete.

The eRSM method shows great promise in transforming material data into evidence of similarities between materials, subgroup discovery, and elucidating underlying mechanisms. However, it is essential to address the challenges in computational resources and further involve human expertise in analyzing the detected groups. Future research should focus on improving the method's scalability and automating the analysis of detected subgroups for more efficient discovery of material properties and mechanisms. By overcoming these challenges, our method can significantly contribute to materials science and accelerate the discovery of new materials with desired properties.

Chapter 5

Evidence-based method for visualizing materials with respect to physical properties

5.1 Introduction

In the previous chapters of my thesis, I proposed data-driven methods to measure the similarities between materials by considering the correlations between their compositions and target properties. These methods involve a rational transformation of materials data into evidence of similarities between materials. The evidence is then modeled and combined to conclude the similarities between materials based on their correlations with compositions and properties of interest. For binary properties, we employed the errors of substitutions as a measure of similarity. This approach quantifies the similarity between materials based on the extent to which their compositions align with each other. On the other hand, we used the correlations of occurrences for continuous properties as a measure of similarity. This measurement captures the degree of correlation between the compositions of materials and their corresponding properties. These similarity measurements have demonstrated their capability to quantitatively assess the similarities between materials while considering the associated uncertainty. Furthermore, they have been successfully applied in exploring new materials, aiding in identifying materials with desired properties based on their similarities to existing materials.

However, rationalizing the estimated uncertainty in predictions through the structure of the query molecule or constituent elements of materials remains a challenging task (Kendall and Gal, 2017; Ha et al., 2021). In practice, attributing prediction failure to a specific material is often performed manually and relies on human intuition. Nonetheless, human decision-making under uncertainty is often subject to biases and heuristics (Tversky and Kahneman, 1974), which can negatively impact the decision-making process and impede the discovery of advanced materials (Gigerenzer and Gaissmaier, 2011). Given those predictions from black-box models, such as deep learning methods, are difficult to interpret and analyze due to their lack of transparency (Koh and Liang, 2017; Rudin, 2019), Explainable Artificial Intelligence (XAI) has recently garnered considerable attention (Linardatos, Papastefanopoulos, and Kotsiantis, 2020; Barredo Arrieta et al., 2020). Explainability refers to the capacity to elucidate why an artificial intelligence model has reached a specific decision or prediction (Barredo Arrieta et al., 2020).

To increase trust and comprehension in AI models used for safety-critical tasks, it's important to incorporate explainability that aligns with human intuition, according to experts (Linardatos, Papastefanopoulos, and Kotsiantis, 2020). In the field of material property predictions, researchers have made progress in understanding model behaviors by analyzing material graphs, compounds, atoms, or feature representations (Rodríguez-Pérez and Bajorath, 2021; Rao et al., 2022; Jiménez-Luna, Grisoni, and Schneider, 2020b). To better understand the reasons behind prediction failures, it's desirable to rationalize estimated uncertainty through material structures or components. This approach can help identify unrecognized functional groups or rare chemical structures in the dataset, and also assist in determining out-of-domain materials. These techniques are being explored in active learning (Li et al., 2019; Nguyen et al., 2022) and drug discovery (Soleimany et al., 2021; Eyke, Green, and Jensen, 2020; Gubaev, Podryabinkin, and Shapeev, 2018). By developing explainable models that can rationalize uncertainty concerning material structures, researchers can improve the understanding of the sources of uncertainty and further optimize the decision-making process in materials discovery.

Nowadays, embedding maps have emerged as a popular method for explaining the predictions of machine learning models, making these explanations more transparent and intuitive for human understanding. Embeddings represent complex, high-dimensional data in lower-dimensional spaces, making it easier to visualize and interpret the relationships between data points. For instance, in natural language processing, word embeddings like Word2Vec (Mikolov et al., 2013) and GloVe (Pennington, Socher, and Manning, 2014) have revolutionized the way text is analyzed by representing words as continuous vectors in a lower-dimensional space. These embeddings capture semantic and syntactic relationships between words, enabling more advanced text analysis and understanding. Similarly, material embeddings have been developed to represent materials in a continuous lower-dimensional space, capturing meaningful relationships and similarities between materials based on their composition, structure, or properties (Jaeger, Fulle, and Turk, 2018; Tshitoyan et al., 2019; Nguyen et al., 2022). These embeddings can be obtained directly or indirectly through various techniques, such as autoencoders (Kingma and Welling, 2019; Jaeger, Fulle, and Turk, 2018; Tshitoyan et al., 2019), or dimensionality reduction methods (Borg, 1997; Kruskal, 1964; McInnes, Healy, and Melville, 2020). By utilizing embeddings, researchers can gain a more comprehensive understanding of the intricate relationships between materials. Embeddings allow for the visualization of material data and the identification of factors that contribute to uncertainties in material discovery. Moreover, embeddings provide a means to input material data into machine learning models, resulting in highly accurate and easily interpretable predictions of material properties. Additionally, embedding maps enable researchers to explore the material space visually, identify trends and patterns, and uncover previously unknown relationships between materials. This empowers researchers to focus on the most promising candidates, thus reducing the time and resources required for material discovery.

Moreover, constructing such maps to visualize data plays a crucial role not only in the initial and but also in final stages of data-driven studies in computational physics, chemistry, and materials science (Cheng et al., 2020). It enables researchers to understand complex datasets' structure and relationships better, ultimately enhancing their ability to make informed decisions and identify meaningful patterns. In the initial stage, low-dimensional maps provide a condensed view of the dataset, revealing underlying patterns such as clusters, outliers, and correlations (Isayev et al., 2015; Ceriotti, 2019). This allows researchers to gain insights from visual inspections, helping them formulate hypotheses and plan further analyses. Advanced visualization techniques like PCA, MDS, t-SNE, and UMAP can be used to create these low-dimensional representations, preserving essential characteristics of the data while reducing its complexity. Visualization is essential for effectively communicating results to a broader audience during the final stage. Clear and informative visualizations enable researchers to convey their findings, making it easier for others to understand the significance and implications of the study. This facilitates knowledge dissemination and promotes collaboration within the scientific community, ultimately accelerating the pace of discovery and innovation. In conclusion, data visualization is vital to data-driven studies in computational physics, chemistry, and materials science. It serves as both an initial exploration tool and a final communication medium, allowing researchers to extract meaningful insights, develop a deeper understanding of complex datasets, and effectively share their findings with others.

The primary objective of this chapter is to address the disconnect between quantitative evaluations and the underlying rationale behind them, empowering scientists to gain insights into the reasoning process and fostering a more profound understanding of the mechanisms governing material properties. This is of paramount importance, as it helps researchers to not only make informed decisions but also to develop a comprehensive grasp of the driving forces behind the properties of interest. By incorporating these explanations into the data-driven method, we aim to refine the decision-making process, enabling scientists to navigate the complex landscape of materials discovery with increased confidence and precision. Furthermore, integrating these explanations into the data-driven model will bolster the overall reliability and adoption of the recommender system within the materials science community, as it addresses the critical need for transparency and interpretability in the field (Goldsmith et al., 2017; Yang and Li, 2023). By striking a balance between quantitative evaluations and their rationale, this study seeks to provide researchers with an indispensable tool that not only facilitates the discovery of novel materials but also promotes a deeper comprehension of the underlying mechanisms, ultimately contributing to the advancement of materials science as a whole.

This chapter presents a data-driven approach that visualizes quantitative uncertainty evaluations into transparent and interpretable maps for supporting scientists in rational decision-making. By expressing both aleatoric and epistemic uncertainties in a manner that is intuitive for human understanding, our method enables scientists to make more informed decisions, reducing the influence of biases that might otherwise lead to suboptimal choices. Furthermore, the embedding maps provide valuable information to reveal insights into the underlying mechanisms involved in the properties of materials. Our approach consists of three main components: 1) measuring the distance between materials, 2) constructing the embedding map based on the calculated distances, and 3) inferring the properties and uncertainties of new materials. We demonstrate the effectiveness of our approach through various case studies, illustrating how increased transparency and understanding contribute to more effective materials discovery and a better comprehension of the driving mechanisms behind material properties. In Section 5.2 of the chapter, we explain the components of the proposed method in detail. We discuss the various techniques and algorithms used to measure distances between materials and construct the embedding maps and the methods employed to infer properties and uncertainties for new materials. Through the case studies, we highlight the practical applications of our approach and showcase its potential for assisting scientists in the materials discovery process. Overall, developing this data-driven approach for visualizing uncertainties in material discovery promises to facilitate better decision-making, enabling scientists to explore new materials more effectively and gain a deeper understanding of the underlying mechanisms governing material properties. By offering a transparent and interpretable means of representing uncertainty, our method can become an indispensable tool for researchers in the quest for new materials and unraveling the complex relationships between materials and their properties.

5.2 Methodology

We consider a set $\mathcal{E}$ of n elements and a corresponding dataset $\mathcal{D}^{pool}$ that contains m possible combinatorial materials comprised of the elements in $\mathcal{E}$. The materials in $\mathcal{D}^{pool}$ are represented by sets of their constituent elements that form the materials. For each material $A \in \mathcal{D}^{pool}$, the categorical variable y_A denotes the property of interest. The y_A is assigned the label "*Unknown*" if we do not know the state of the property of material A. In other words, the property of A has not been calculated or experimentally determined yet. Therefore, we denote $\mathcal{D}^{observed}$ ($\subseteq \mathcal{D}^{pool}$) and $\mathcal{D}^{unobserved}$ ($\subseteq \mathcal{D}^{pool}$) as datasets containing materials with property $y_A \neq Unknown$ and $y_A = Unknown$, respectively. It should be noted that $\mathcal{D}^{observed} \cup \mathcal{D}^{unobserved} = \mathcal{D}^{pool}$ and $\mathcal{D}^{observed} \cap \mathcal{D}^{unobserved} = \emptyset$.

In order to create a visual representation of the materials in $\mathcal{D}^{pool}$, we first establish a way to measure the distance between them based on their properties and composition, specifically the property y_A. This distance measurement allows us to determine how similar or dissimilar materials are to each other. We then use dimensionality reduction techniques to generate a map based on these calculated distances. These techniques aim to maintain the relationships between materials in the reduced space as closely as possible, resulting in a meaningful representation of the original data. Finally, we use nearest neighbor methods within the map to aid in material synthesis decisions. By identifying materials that are close together in the map, we can infer that they have similar properties or composition. This information can be used to guide the selection of materials for further investigation, prioritizing those that are the most promising based on their position in the map. By using the embedding map as a decision-making tool, scientists can make more informed choices when it comes to material synthesis, ultimately leading to more efficient and effective discovery of new materials.

5.2.1 Distance measurement between materials

One of the most popular approaches to assess the dissimilarity between two sets is to analyze the differences and commonalities in their constituent elements (Tversky, 1977). Jaccard distance is one of the widely used measurements that implements the form of dissimilarity between the sets (LEVANDOWSKY and WINTER, 1971). The Jaccard distance measures dissimilarity between finite sample sets and is defined as the complement of the Jaccard similarity coefficient, which is a ratio of the intersection sizes and union of the two sets. Mathematically, given two material combinations A and B ($\in \mathcal{D}^{pool}$), the Jaccard distance between these materials is calculated as follows:

$$d_{jaccard}(A_i, A_j) = 1 - \frac{|A_i \cap A_j|}{A_i \cup A_j}, \tag{5.1}$$

where $|A_i \cap A_j|$ represents the size of the intersection between sets A_i and A_j, and $|A_i \cup A_j|$ represents the size of the union of the two sets. The Jaccard distance ranges between 0 and 1, with a value of 0 indicating that the two sets are identical (i.e., they share all constituent elements) and a value of 1 indicating that the two sets have no common elements.

Moreover, given a property of interest, the materials that exhibit difference properties should be considered more dissimilar and vice versa. For detail, considering the three materials A, B and C, such that $d_{jaccard}(A, B) = d_{jaccard}(A, C) \neq 0$, if $y_A = y_B \neq y_C$ we should be considered the material A is much similar to B than C. Therefore, we propose a distance metric d_{ers} with respect to the property of interest, as follows:

$$d_{ers}(A, B) = m^{C_t, C_v}_{\mathcal{D}^{observed}}(\{dissimilar\}) + \frac{m^{C_t, C_v}_{\mathcal{D}^{observed}}(\{similar, dissimilar\})}{2}, \tag{5.2}$$

where $C_t = A - (A \cap B)$, $C_v = B - (A \cap B)$, and the $m^{C_t, C_v}_{\mathcal{D}^{observed}}$ is the combined mass function about the similarity in terms of substitutability between C_t and C_v. The mass function is modeled and combined from evidence collected from the dataset $\mathcal{D}^{observed}$. Details of the method are introduced in section 3.2 of chapter 3.

Finally, the distance between the two materials A and B is derived by integrate the $d_{jaccard}(A, B)$ and $d_{ers}(A, B)$, as follow:

$$d(A, B) = \sqrt{d_{jaccard}(A, B) \times d_{ers}(A, B)}. \tag{5.3}$$

It should be noted that if the substitution operator between the two materials A and B is unobserved in the dataset $\mathcal{D}^{observed}$, the $m^{C_t, C_v}_{\mathcal{D}^{observed}}(\{similar, dissimilar\})$ equal to one and considers no information about the dissimilarity between the material A and B ($d_{ers}(A, B) = 1$). Therefore, the combined distance $d(A, B) = d_{jaccard}(A, B)$.

Similar analyses are performed for all pairs of materials in $\mathcal{D}^{pool}$ to obtain a symmetric matrix M that comprises all of the distances between them ($M[a, b] = M[b, a] = d(A, B)$).

5.2.2 Dimensionality reduction methods

In order to preserve the distance between materials, we utilize manifold learning, a non-linear dimensionality reduction approach, to create a lower-dimensional embedding map. This involves using algorithms that recognize that the dimensionality of many datasets is artificially high. Visualizing high-dimensional datasets can be challenging, whereas data in two or three dimensions can be easily plotted to show the inherent structure of the data. To help visualize the structure of the dataset, the dimension must be reduced. In our study, we utilized two different methods of manifold learning: multi-dimensional scaling (MDS) and t-distributed Stochastic Neighbor Embedding (t-SNE).

Multi-dimensional Scaling (MDS): MDS is a classical dimensionality reduction technique that strives to represent high-dimensional data in a lower-dimensional space while preserving the pairwise distances between data points as closely as possible (Borg, 1997; Kruskal, 1964). The primary goal of MDS is to produce a low-dimensional representation of the data where the Euclidean distances between points in the reduced space closely resemble the original distances in the high-dimensional space. MDS is particularly well-suited for visualizing dissimilarity data, such as distances between cities or similarities between texts. However, MDS might face difficulties in capturing complex non-linear relationships in the data, and its computational efficiency may decrease as the number of instances increases.

For detail, we start from the distance matrix $M_{\mathcal{D}^{pool}}$ containing the pairwise distance between materials in $\mathcal{D}^{pool}$, which are obtained from the section 5.2.1. The objective of the MDS is finding representation vectors $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$ for m materials in $M_{\mathcal{D}^{pool}}$ that minimize the following loss function:

$$loss_{MDS} = \sum_{i<j} M_{\mathcal{D}^{pool}}[i,j] - \hat{M}_{\mathcal{D}^{pool}}[i,j], \tag{5.4}$$

where $\hat{M}_{\mathcal{D}^{pool}} = ||\mathbf{x}_i - \mathbf{x}_j||^2$ is the Euclidean distances between the representation vectors of the materials on the embedding map.

t-Distributed Stochastic Neighbor Embedding (t-SNE): t-SNE is a non-linear dimensionality reduction technique introduced by van der Maaten and Hinton in 2008 (Maaten and Hinton, 2008). It is designed to preserve local structures in the data, making it particularly effective for visualizing high-dimensional data with intricate relationships. t-SNE works by minimizing the divergence between two probability distributions: one representing pairwise similarities in the high-dimensional space and the other representing pairwise similarities in the lower-dimensional space. The critical innovation of t-SNE is its use of the t-distribution in the lower-dimensional space, which mitigates the crowding problem that may arise in other techniques. The crowding problem in t-SNE (t-Distributed Stochastic Neighbor Embedding) derives from the "curse of dimensionality." When data points are reduced from a high-dimensional space to a lower-dimensional space (e.g., 2D or 3D for visualization), the distances between points change. In high-dimensional spaces, points are generally far apart, whereas, in lower-dimensional spaces, points are closer together. This distance discrepancy can make preserving the data's original structure challenging when embedded in a lower-dimensional space. t-SNE is specifically designed to mitigate the crowding problem by modeling pairwise similarities between points with probability distributions. A Gaussian distribution is used in high-dimensional space, whereas in lower-dimensional space, a t-distribution is used. The t-distribution has heavier tails than the Gaussian distribution, allowing t-SNE better model the pairwise similarities in the lower-dimensional space.

For detail, we start from the distance matrix $M_{\mathcal{D}^{pool}}$ containing the pairwise distance between materials in $\mathcal{D}^{pool}$, which are obtained from the section 5.2.1. t-SNE first computes probabilities p_{ij} that are proportional to the materials A_i and A_j in the dataset $\mathcal{D}^{pool}$, as follows:

$$p_{ij} = \frac{p_{i|j} + p_{j|i}}{2m}, \tag{5.5}$$

where $p_{i|j}$ is the conditional probability that A_j would pick A_i as its neighbor if neighbors were picked in proportion to their probability density under a Gaussian centered at A_j. The conditional probability is defined as follow:

$$p_{j|i} = \frac{\exp(-M_{\mathcal{D}^{pool}}[i,j]/2\sigma_i^2)}{\sum_{k\neq i}\exp(-M_{\mathcal{D}^{pool}}[i,k]/2\sigma_i^2)}. \tag{5.6}$$

To ensure that the entropy of the conditional distribution matches a predetermined entropy, the Gaussian kernel's bandwidth, denoted as σ_i, is adjusted using the bisection method. This modification enables the bandwidth to conform to the data density, with smaller σ_i values applied in areas of higher density within the data space.

t-SNE is used to learn d-dimensional representation vectors $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$ of the m materials (with $\mathbf{x}_i \in \mathbb{R}^d$ and d typically chosen as 2 or 3) that reflects the similarities p_{ij} as well as possible. To achieve this objective, a comparable technique is employed to gauge the similarities (q_{ij}) between two points (x_i and x_j) on the embedding map. Precisely, for cases where $i \neq j$, q_{ij} is defined as follows:

$$q_{ij} = \frac{(1 + \|\mathbf{x}_i - \mathbf{x}_j\|^2)^{-1}}{\sum_k \sum_{l \neq k} (1 + \|\mathbf{x}_k - \mathbf{x}_l\|^2)^{-1}}. \tag{5.7}$$

A heavy-tailed Student t-distribution (with one degree of freedom, equivalent to a Cauchy distribution) is utilized to measure similarities between low-dimensional points to model dissimilar objects far apart in the map. The locations of the points $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$ are determined by minimizing the non-symmetric Kullback-Leibler (KL) divergence of distribution P from distribution Q:

$$\mathrm{KL}\,(P \parallel Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}. \tag{5.8}$$

To effectively display similarities among high-dimensional inputs on a map, utilizing gradient descent to minimize the KL divergence is crucial. Keep in mind, as the KL divergence is not a convex function, starting from different points may result in local minima. It is therefore necessary to try various seeds and select the one with the lowest KL divergence to obtain the optimal embedding.

5.2.3 *K*-nearest neighbor classifier

We constructed an embedding map in the previous subsection that preserves the obtained dissimilarities between materials, as defined in subsection 5.2.1. The proposed distance metric evaluates pairs of materials based on their differences and commonalities in constituent elements with respect to the property of interest. As a result, materials that are close to each other in the embedding space share more common elements and have similar properties of interest. It is worth noting that unknown materials are also positioned near each other, forming regions of materials with high epistemic uncertainty. Therefore, we can use neighbor-based models, such as the k-nearest neighbor classifier (Cover and Hart, 1967), to make inferences or decisions from the map. The k-nearest neighbor classifier is a simple yet powerful algorithm that can be used for classification and regression tasks. It works by identifying the k data points in the embedding space that are closest to a given query point and then predicting the property of interest based on the majority vote or the average of the properties of these neighbors. This approach is particularly suitable for the material embedding map, as it takes advantage of the spatial organization of the materials in the embedding space and the assumption that materials with similar properties are located close to each other. By employing the k-nearest neighbor classifier or other neighbor-based models, researchers can make informed decisions regarding material synthesis, identify promising candidate materials, and gain insights into the relationships between material properties and their constituent elements. This ultimately contributes to more efficient materials discovery and a better understanding of the underlying mechanisms governing material properties.

5.3 Case study 1: Visualization of cocktail effects between transition metals in HEA phase formation

5.3.1 Experimental design

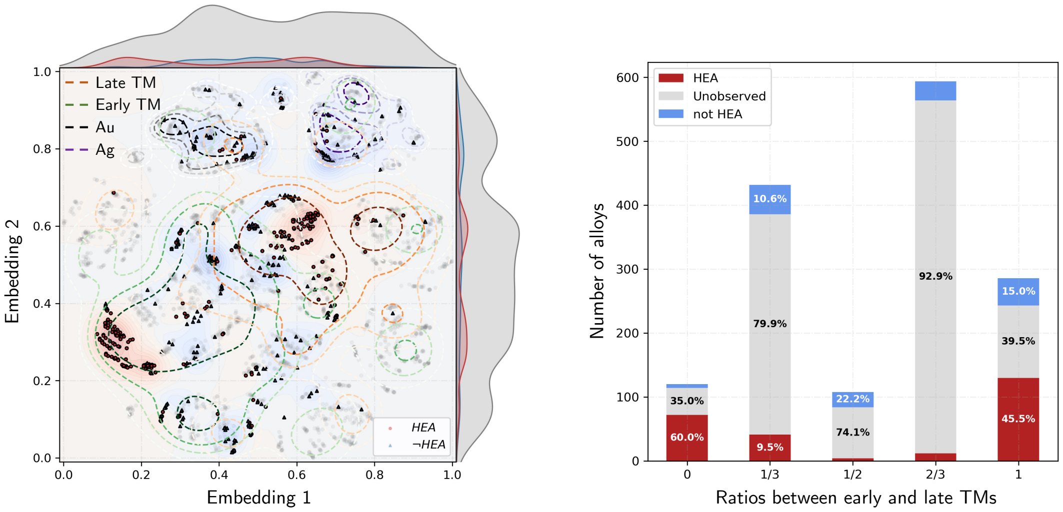
We apply the proposed method to construct an embedding map for binary and ternary alloys, which are comprised of $\mathcal{E} = \{$ Fe, Co, Ir, Cu, Ni, Pt, Pd, Rh, Au, Ag, Ru, Os, Si, As, Al, Tc, Re, Mn, Ta, Ti, W, Mo, Cr, V, Hf, Nb, and Zr}. We denote $\mathcal{D}^{pool}_{binary,ternary}$ as the dataset of 351 (combinations of 2 in 27 elements) binary and 2925 (combinations of 3 in 27) ternary alloys that are comprised of the elements in $\mathcal{E}$. Among the alloys, there are 558 alloys, in which there are 117 binary and 441 ternary alloys, and their phases (either HEA or $\neg HEA$) are estimated by using the AFLOW calculation (Zhang et al., 2014). Details of the 558 alloys are introduced in the previous section 3.3.1 of the chapter 3. We denote the $\mathcal{D}^{observed}_{binary,ternary}$ as the dataset of the 558 calculated alloys.

In this section, we perform two experiments to demonstrate the embedding map's effectiveness in predicting the alloys' HEA phase and providing valuable information for decision support under uncertainty. For the first experiment, we apply the k-nearest neighbor classifier on the embedding map to learn a classification model for predicting the HEA phase of the alloys, in which the alloys are represented by their coordinations on the embedding map. We compare the learned KNN model with the ERS model, which is another classification approach for predicting the HEA phase and is introduced in chapter 3. To evaluate the performance of the two models, we perform a 9-fold cross-validation on $\mathcal{D}^{observed}_{binary,ternary}$ 10 times. This cross-validation process involves dividing the dataset into 9 equal parts, training the models on 8 parts, and testing their performance on the remaining part. This process is repeated 10 times to ensure the stability and reliability of the results. For the second experiment, we investigate the distributions of alloys corresponding to their constituents to extract valuable information about the underlying HEA formation mechanisms with considering the uncertainty.

5.3.2 Results and discussions

The prediction accuracies of the KNN model on the embedding map and the ERS model are 0.81 ± 0.02 and 0.80 ± 0.01, respectively. These results indicate that the embedding map provides descriptive information for representing the alloys concerning HEA phase formation, enabling the KNN model to achieve a similar level of prediction accuracy as the ERS model. The comparable performance of the KNN model on the embedding map and the ERS model highlights the potential of using embedding maps as a valuable tool for material discovery. The embedding map not only offers a visually intuitive representation of the material space but also captures meaningful relationships between materials based on their composition, structure, and properties. This, in turn, enables the development of machine learning models, such as the KNN classifier, that can make accurate and interpretable predictions of material properties. In conclusion, our experiments demonstrate the effectiveness of using embedding maps for predicting the HEA phase of alloys. The results suggest that embedding maps can be a powerful tool for material discovery, providing both a visually intuitive representation of the material space and a foundation for building accurate and interpretable machine learning models.

Figure 5.1 a displays the embedding map of alloys in $\mathcal{D}^{pool}_{binary,ternary}$, constructed from the $\mathcal{D}^{observed}_{binary,ternary}$, providing a visual representation of the material space. We observe that the regions of Au-based alloys (black contours) and Ag-based alloys (purple contours) are dominated by multiphase alloys (blue points). This result indicates that alloys containing either gold or silver are likely not to form the HEA phase. Consequently, in terms of the substitution method, substituting gold with silver and vice versa will not increase the probability of forming the HEA phase. This observation is consistent with the results obtained from the ERS model in section



(a) (b)

FIGURE 5.1: (a) The embedding map of the alloys in $\mathcal{D}^{pool}_{binary,ternary}$, constructed from the $\mathcal{D}^{observed}_{binary,ternary}$, provides a visual representation of the material space. In this map, orange, green, black, and purple contours indicate the regions of alloys containing late transition metals, early transition metals, gold (Au), and silver (Ag), respectively. The red and blue points represent alloys in $\mathcal{D}^{observed}_{binary,ternary}$ that form and do not form the HEA phase, respectively, while the plus points denote unobserved alloys. (b) The bar chart shows the number of alloys concerning the ratios of early and late transition metals in their composition, as well as the proportions of HEA, multiphase alloys (not HEA), and unknown alloys.

3.3.3, which shows that gold and silver are similar to each other in terms of substitutability. Additionally, the regions where alloys form the HEA phase are dominated by alloys consisting of either early transition metals (TM) or late transition metals, respectively. In contrast, the region containing alloys with both early and late TMs is predominantly highlighted by multiphase alloys or "unobserved" compounds. This result is consistent with the distinction between the two groups obtained from the ERS model in section 3.3.3. In other words, alloys containing either only early TMs or late TMs have a high chance of forming the HEA phase, while mixing the elements from both groups may cause a reverse effect, leading to the formation of multiphase alloys.

Further investigations show that 60% of alloys, which comprise only late transition metals, form the HEA phase, and 5% of these alloys do not form the HEA phase, with the remainder being unknown (Fig. 5.1 b). On the other hand, alloys composed solely of early transition metals tend to form the HEA phase more frequently, with a proportion of 45% compared to multiphase alloys at 15%. However, combining these two types of alloys leads to high epistemic uncertainty, and what happens currently needs to be clarified. Therefore, conducting more experiments is imperative to gather additional information and clarify the situation. The embedding map is a crucial tool for understanding the complex interactions between elements that contribute to creating the HEA phase. It is an essential support for the ERS model, providing valuable insights for developing new alloys. By visualizing the material space, this tool allows researchers to identify patterns and connections that can inform future studies and experiments, ultimately leading to the discovery of new high-performance alloys.

5.4 Case study 2: Experiments simulating the discovery of quaternary alloys

5.4.1 Experimental design

We consider the dataset $\mathcal{D}^{pool}_{quaternary}$, containing 990 equiatomic quaternary high-entropy alloys, composed of 14 transition metals: $\{Ag, Cd, Co, Cr, Cu, Fe, Mn, Mo, Ni, Pd, Rh, Ru, Tc, Zn\}$, and their corresponding labels *mag* and T_C. The label y^{mag}_A of a quaternary alloy A in $\mathcal{D}^{pool}_{quaternary}$ is denoted as either *nonzero* or *zero* if its calculated magnetization is nonzero or zero, respectively. Similarly, the label $y^{T_C}_A$ of the quaternary alloy A is denoted as *nonzero* or *zero* if its calculated Curie temperature in the BCC phase is nonzero or zero, respectively. The dataset was collected from an original dataset of 147,630 equiatomic quaternary high-entropy alloys calculated using the Korringa-Kohn-Rostoker coherent approximation method (Fukushima et al., 2022). Each alloy in $\mathcal{D}^{pool}_{quaternary}$ is represented by a set of its constituent elements.

In this section, we perform two experiments to demonstrate the embedding map's effectiveness in predicting the alloys' properties (magnetization and Curie temperature) and providing valuable information for decision support under uncertainty. For the first experiment, we apply the k-nearest neighbor (KNN) classifier on the embedding map to learn a classification model for predicting the y^{mag} and y^{T_C} labels of quaternary alloys, where the alloys are represented by their coordinates on the embedding map. We compare the learned KNN model with the ERS model, another classification approach for predicting whether particular alloys are magnetic or nonmagnetic, introduced in Chapter 3. To evaluate the performance of the two models, we perform a train-test-split cross-validation on $\mathcal{D}^{pool}_{quaternary}$, with the size of the training data ranging from 10% to 90%, in 10 repetitions. This validation process involves training the models on l percent of the dataset ($l \in [10, 90]$) and testing their performance on the remaining part. This process is repeated 10 times to ensure the stability and reliability of the results.

For the second experiment, we design a simulation of the exploration process for the quaternary alloys in $\mathcal{D}^{pool}_{quaternary}$. We randomly select 90 alloys from $\mathcal{D}^{pool}_{quaternary}$ as the initial training data, while the remaining alloys in $\mathcal{D}^{pool}_{quaternary}$ are considered as not-yet-calculated data. For a query time t, we select 20 alloys from the not-yet-calculated data according to criteria of exploration or exploitation. To rank the alloys at the query time, we apply two different models: 1) KNN models learned on the embedding map constructed from calculated data, and 2) the ERS model learned from calculated data. This experiment aims to compare the performance of the two models in the context of an exploration-exploitation scenario and provide insights into the usefulness of the embedding map for decision-making under uncertainty. We perform the validation in 20 times.

In detail, at each query time, we learn the two models based on the available data on $\mathcal{D}^{observed}_{quaternary}$ and use these models to predict the alloys on the not-yet-calculated data. In order to determine how an unobserved alloy called B affects the accuracy of prediction models and the exploration process, we have created two acquisition functions called $\tau^{map}_{exp}(B)$ and $\tau^{ers}_{exp}(B)$. These functions use the KNN model on the embedding map and the ERS model to evaluate the performance of alloy B. The functions are defined as follows:

$$\tau_{exr}^{ers}(B) = m^{B}_{\mathcal{D}^{observed}_{quaternary}}(\{\text{nonzero, zero}\}) \tag{5.9}$$

$$\tau_{exr}^{map}(B) = p(y_B = "unk"|\mathbf{x}_B), \tag{5.10}$$

where the $\mathbf{x}_B$ is the coordinations of the alloy on the embedding map. The probability p is estimated by using the KNN model on the embedding map with considering three variables of the property of interest ("zero," "nonzero," and "unk"). The label *unk* indicates that the alloy is a not-yet-calculated data.

Regarding the exploitation strategy, we define acquisition functions that select not-yet-calculated structures with the high belief that the alloys have nonzero magnetization or nonzero Curie temperature. The functions are defined as follows:

$$\tau_{exp}^{ers}(B) = m^{B}_{\mathcal{D}^{observed}_{quaternary}}(\{\text{nonzero}\}) \tag{5.11}$$

$$\tau_{exp}^{map}(B) = p(y_B = "nonzero"|\mathbf{x}_B), \tag{5.12}$$

where the $\mathbf{x}_B$ is the coordination of the alloy on the embedding map. The probability p is estimated by using the KNN model on the embedding map with considering three variables of the property of interest ("zero," "nonzero," and "unk"). The label *unk* indicates that the alloy is a not-yet-calculated data.

5.4.2 Results and discussions

Figure 5.2 demonstrates the performance comparison between different classification models for predicting magnetic alloys (a) and nonzero T_C alloys (b) in the context of quaternary high-entropy alloys. When using the MDS method to generate the embedding map, the resulting classification models exhibit significantly lower prediction accuracies compared to the ERS models. This indicates that the MDS-generated embedding map is not as effective in capturing the essential features required for accurate predictions of the alloy properties. In contrast, the embedding map generated using the t-SNE method exhibits a similar trend in performance for classifying magnetic alloys and alloys with nonzero Curie temperatures compared to the ERS models, suggesting that the t-SNE-generated map is more effective in representing the data.

As for the classification of magnetic alloys, we observe a distinct improvement in the accuracy of both the KNN (on a t-SNE-generated map) and ERS models as the training size increases from 10% to 30%. Subsequently, the performance stabilizes, hovering at around 90% for the ERS models and 87% for the KNN models on the t-SNE-generated embedding map. Intriguingly, this level of stability persists even as the training set expands. A similar trend is observed in the models' ability to determine alloys with nonzero Curie temperatures (Fig. 5.2 b). The results of the KNN models based on the t-SNE-generated map exhibit a similar trend and performance in prediction accuracies to the ERS models for both properties. In stark contrast, the accuracies of the KNN models based on the MDS-generated map are significantly lower than those of the other two models. These findings provide compelling evidence that the t-SNE-generated embedding map is more adept at predicting the properties of interest than its MDS-generated counterpart. This observation underscores the critical role of choosing the right dimensionality reduction technique for

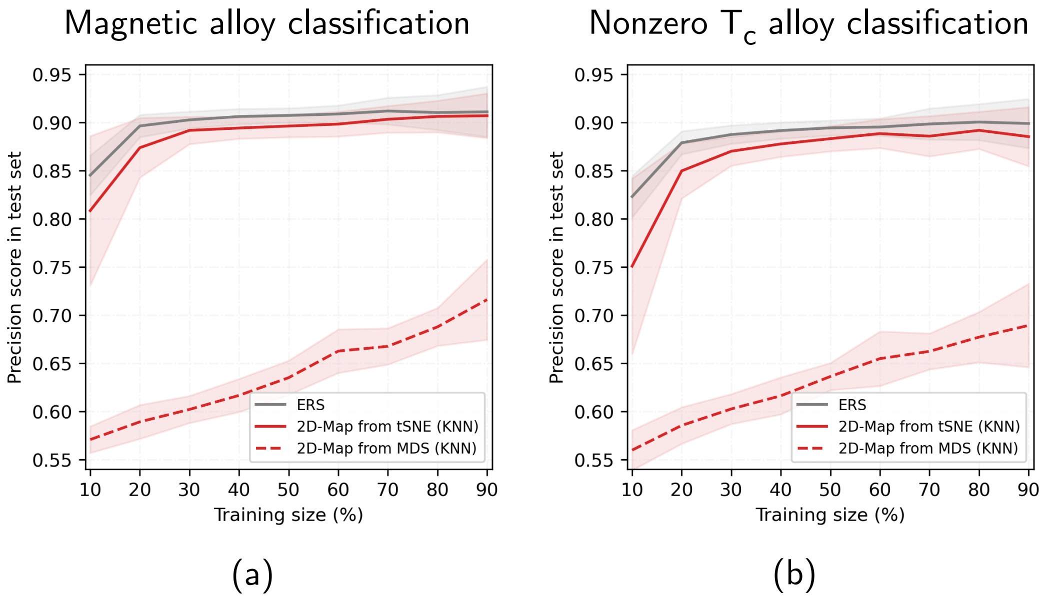


FIGURE 5.2: Comparison of prediction accuracies between classification models for magnetic alloys (a) and nonzero T_C alloys (b). In each plot, the gray line indicates the precision score of the ERS model on the test set, while the red lines show the results of the KNN models on the embedding maps generated by using the t-SNE (solid line) or MDS (dashed line) methods.

creating embedding maps that can effectively capture the intricate relationships embedded within the data. Researchers are better equipped to develop more accurate classification models by leveraging a suitable embedding map. In turn, these models can furnish valuable insights that can be used to inform future efforts in alloy design, discovery, and property prediction, thereby pushing the boundaries of what we currently understand about these complex materials.

Figure 5.3 showcases the performance comparison between ERS learners (gray lines) and KNN learners on the embedding map (red and orange lines) using different selection strategies. The recall rate for querying magnetic alloys (a) and alloys with nonzero Curie temperature (b) is plotted as a function of query time t. Similarly, the prediction accuracies for classifying magnetic alloys (c) and nonzero T_C alloys (d) are also presented as a function of query time t. This figure illustrates that all settings (learner and criteria) manage to recall all the positive alloys (magnetic alloys or non-zero T_C alloys) without querying all the not-yet-calculated alloys (Fig. 5.3 a-b). The designs employing an exploration querying strategy with both KNN learner and ERS learner (solid red and gray lines) exhibit the worst recall performance. By contrast, KNN learners applying the exploitation strategy display the best performances, requiring approximately 30/45 query steps to recall all the magnetic alloys or alloys with non-zero Curie temperature. In comparison, ERS learners require more than 2 or 3 steps to recall all these positive alloys, with higher variance than those of KNN learners. Regarding the prediction accuracies of different settings, Fig. 5.3 c and d show that the settings with exploration querying strategy require less than 10 queries (approximately 30% of alloys in the $\mathcal{D}^{pool}_{quaternary}$) to reach stable, high accuracies, consistent with the results from the first experiment. The results of the settings of ERS or KNN learners exhibit similar performance. In

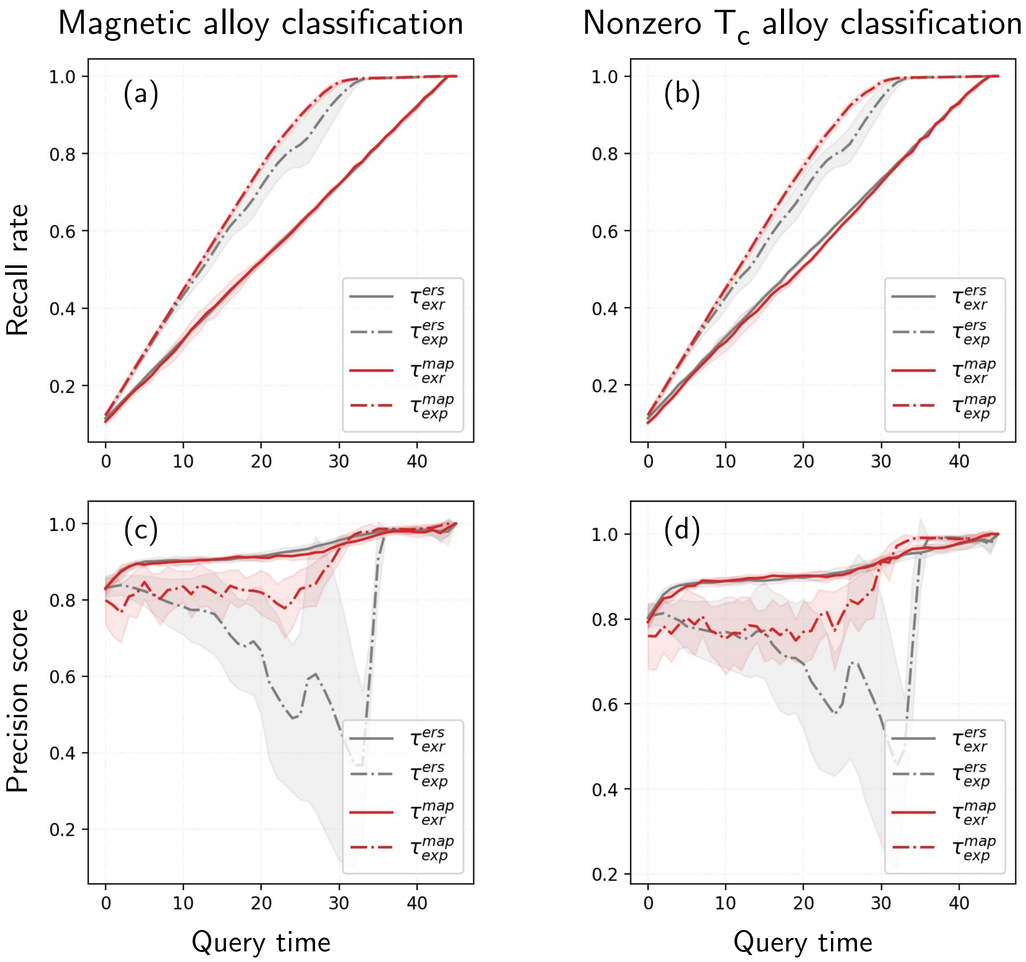


FIGURE 5.3: Comparison of ERS learners (gray lines) and KNN learners on embedding map (red lines) with different criteria. (a-b) Recall rate in querying magnetic alloys and alloys with nonzero Curie temperature depending on the query time t. (c-d) Dependence of prediction accuracies in classifying magnetic alloys (c) and nonzero T_C alloys (d) on querying time t.

summary, the comparison between ERS learners and KNN learners on the embedding map demonstrates the potential of using low-dimensional embedding maps to support scientists in making decisions.

Moreover, using the embedding map provides visual explanations of the learners during the querying process, allowing researchers to observe the evolution of the model's understanding of the material space as more data is incorporated. Figure 5.4 displays the embedding maps of the alloys in the $\mathcal{D}^{pool}_{quaternary}$ at four different query times (initial step, fifth step, tenth step, and ending step) with respect to their magnetization properties. On the maps at steps 0 and 5, the distributions of magnetic and non-magnetic alloys overlap, suggesting that the model has not yet clearly distinguished between the two groups. In contrast, from step 10 onwards, the distributions separate from each other, revealing a more refined understanding of the material space and the relationships between the constituent elements and their magnetic properties. Magnetic alloys are dominated by alloys composed of iron, cobalt, or manganese, indicating that these elements play crucial roles in determining the magnetic behavior of the alloys. However, it is essential to note that some Mn-based alloys do not contain cobalt or iron and exhibit non-magnetic properties, highlighting the complex nature of the relationships between elements and their resulting properties in quaternary alloys.

5.5 Contributions and limitations

In material science, understanding and visualizing the material space is crucial for discovering and designing new materials with desired properties. Recently, a novel evidence-based method for visualizing the material space with respect to properties of interest has been proposed. The effectiveness of this method, which involves constructing embedding maps using manifold learning techniques, has been demonstrated in predicting properties of interest and providing valuable insights into the underlying mechanisms of the data.

1. Visual Representation of Complex Material Space: The evidence-based visualization method offers an intuitive and visually appealing representation of the complex relationships between materials and their properties. This allows scientists and researchers to understand the material space better and make informed decisions during the design and discovery process.

2. Predictive Capabilities: The embedding maps generated using the proposed method have effectively predicted properties of interest, such as magnetic properties and Curie temperatures. This predictive capability can be invaluable in guiding the selection of candidate materials for further experimentation and analysis.

3. Understanding Learned Models: The visualization method provides valuable insights into the learned models from data, such as ERS and KNN learners. This understanding can help researchers better interpret the results of these models and refine them for improved performance in future studies.

However, the method has certain limitations, such as its reliance on the manifold learning method and the need for further research on effective querying strategies.

1. Dependence on Manifold Learning Methods: The proposed method relies on manifold learning techniques, such as t-SNE and MDS, to generate the embedding maps. These techniques can sometimes be sensitive to their hyperparameters and may not always produce optimal visualizations. Furthermore, the choice of manifold learning method can significantly impact the resulting visualization, as seen in the comparison between t-SNE and MDS.

2. Trade-off Between Descriptiveness and Transparency: When constructing visual maps for human interpretation, there is an inherent trade-off between descriptiveness and transparency. To make the visualization more interpretable, some level of detail may be sacrificed, which could limit the ability of the map to represent the material space accurately.

3. Designing Effective Querying Strategies: Using the embedding maps, the current evidence-based visualization method requires further research to develop effective strategies for querying candidate materials. Efficiently exploring the material space and selecting the most promising candidates remains a challenge that needs to be addressed to maximize the potential of the proposed method.

The evidence-based method for visualizing the material space offers several significant contributions, including intuitive visual representation, predictive capabilities, and insights into learned models. However, there are also limitations, such as the dependence on manifold learning methods, the trade-off between descriptiveness and transparency, and the need for further research on effective querying strategies. Despite these limitations, the proposed method holds promise as a valuable tool for scientists and researchers working on material design and discovery. Addressing these limitations and refining the method can advance our understanding and exploration of the complex material space.

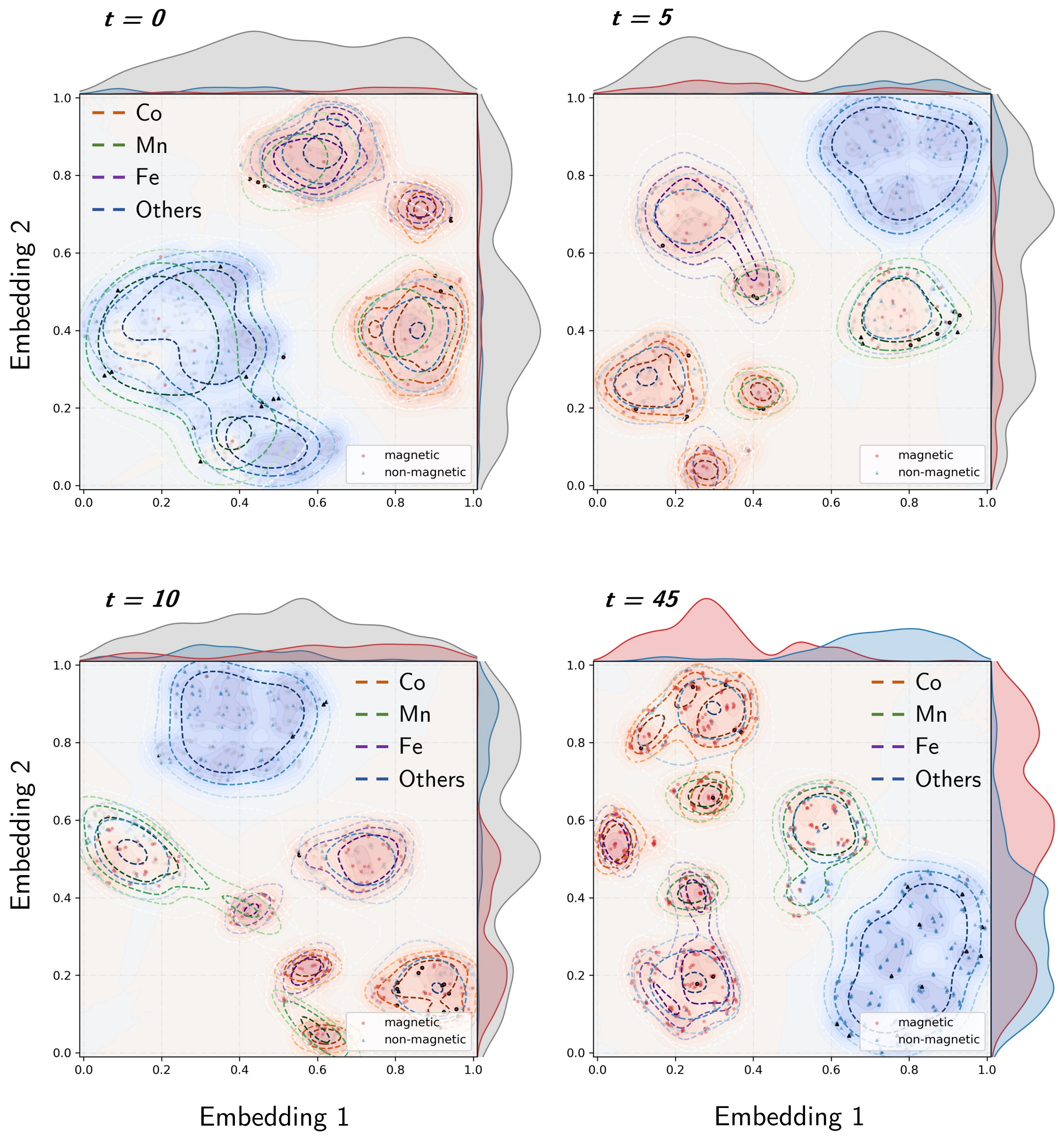


FIGURE 5.4: Embedding maps of the alloys in $\mathcal{D}_4^{pool}$ at four different query times t (0-Initial, 5, 10, and 45-End). Orange, green, purple and blue contours indicate the regions of alloys comprise of cobalt (Co), manganese (Mn), iron (Fe), and other metals, respectively.

Chapter 6

Publication list

List of published/accepted publications

1. Tien-Lam Pham, Duong-Nguyen Nguyen, **Minh-Quyet Ha**, Hiori Kino, Takashi Miyake, and Hieu-Chi Dam, “Explainable Machine Learning for Materials Discovery: Predicting the Potentially Formable Nd-Fe-B Crystal Structures and Extracting Structure-Stability Relationship”, IUCrJ, volume 7, number 6, pages 1036–1047, 2020.

2. Nhan Nu Thanh Ton, **Minh-Quyet Ha**, Takuma Ikenaga, Ashutosh Thakur, Hieu-Chi Dam and Toshiaki Taniike, “Solvent screening for efficient chemical exfoliation of graphite”, 2D Materials, volume 8, number 1, 2020. (Nhan Nu Thanh Ton and Minh-Quyet Ha contributed equally to this work)

3. **Minh-Quyet Ha**, Duong-Nguyen Nguyen, Viet-Cuong Nguyen, Takahiro Nagata, Toyohiro Chikyow, Hiori Kino, Takashi Miyake, Thierry Denœux, Van-Nam Huynh and Hieu-Chi Dam, “Evidence-based recommender system for high-entropy alloys”, Nature Computational Science, volume 1, pages 470–478, 2021.

4. **Minh-Quyet Ha**, Duong-Nguyen Nguyen, Viet-Cuong Nguyen, Hiori Kino, Yasunobu Ando, Takashi Miyake, Thierry Denœux, Van-Nam Huynh, and Hieu-Chi Dam, “Evidence-Based Data Mining Method to Reveal Similarities between Materials Based on Physical Mechanisms”, Journal of Applied Physics, volume 133, issue 5, 053904, 2023.

List of submitted articles

1. Hirosuke Matsui; Yuta Muramoto; Takashi Kakubo; Naoya Amino; Tomoya Uruga; **Minh-Quyet Ha**; Duy-Tai Dinh; Hieu Dam; Mizuki Tada, “Machine-learning-revealed reaction statistics via 3D spectroimaging for copper sulfidation of adhesive layers in rubber/brass composite”, Communications Materials. (submitted)

2. Tien-Sinh Vu, Duy-Tai Dinh, **Minh-Quyet Ha**, Duong-Nguyen Nguyen, Viet-Cuong Nguyen, Yukihiro Abe, Truyen Tran, Huan Tran, Hiori Kino, Takashi Miyake, Hieu-Chi Dam, “Deep learning reveals key aspects to help interpret the structure–property relationships of materials”, npj Computational Materials. (submitted)

List of presentations

1. **Minh-Quyet Ha**, Nguyen-Duong NGUYEN, Van-Doan NGUYEN, Dinh-Khiet LE, Hieu-Chi DAM, "The application of theory of evidence in prediction of high-entropy alloys", '60th Sanibel Symposium: The Theory Meeting for Theoreticians', Georgia, USA, (February 2020)

2. **Minh-Quyet Ha**, Duong-Nguyen NGUYEN, Hiori Kino, Viet-Cuong NGUYEN and Hieu-Chi DAM, "Descriptor free recommender system for new high-entropy alloys: An application of the evidence theory", 'Virtual Conference on a FAIR Data Infrastructure for Materials Genomics', Berlin, Germany, (June 2020)

3. **Minh-Quyet Ha**, Nhan Nu Thanh Ton, Toshiaki Taniike, Hieu-Chi Dam, "Application of evidence theory to recommend solvent mixtures for chemical exfoliation of graphite", August 2021, XXXII IUPAP Conference on Computational Physics (CCP2021). (Poster presentation)

4. **Minh-Quyet Ha**, Duong-Nguyen Nguyen, and Hieu-Chi Dam, "Application of evidence theory to recommend new high entropy alloys", December 2021, Pacifichem 2021, Hawaii, USA. (Poster presentation)

5. **Minh-Quyet Ha**, Duong-Nguyen Nguyen, and Hieu-Chi Dam, "Exploration of high-entropy alloys: An application of the evidence theory", December 2021, MRM2021, Yokohama, Japan. (Poster presentation)

Appendix A

Monitoring HEA recall ratios in test set

A.1 Evaluation of HEA-recommendation capability by cross-validation

In the experiment with $\mathcal{D}_{\text{ASMI16}}$, the result shows that the ERS can significantly reduce the number of trials required to recall all the HEAs in the test set compared to the competitor systems (Figure A.1 a). The proposed ERS requires less than 12, 25, and 80% of all the possible trials to recall one-half, three-quarters, and all the HEAs in the test set, respectively (Table A.1). In the $\mathcal{D}_{\text{CALPHAD}}$ experiment, the ERS requires less than 2 and 5% of all the possible trials to recall one-half and three-quarters of the HEAs in the test set, respectively, which are the fewest trials required among all the recommender systems (Figure A.1 b and Table A.1). Interestingly, in the $\mathcal{D}_{\text{ASMI16}}$ and $\mathcal{D}_{\text{CALPHAD}}$ experiments, the supervised-method-based recommender systems either approximately randomly selected possible HEAs (Naïve Bayes and decision tree) or could not rank any (logistic regression and SVM) at all because these data sets contain only positively labeled HEAs.

The result in $\mathcal{D}_{\text{AFLOW}}$ experiment demonstrates that the ERS also outperforms the competitor systems in recalling one-half of the HEAs in the test set. However, the ERS cannot reliably recall the one-quarter of the HEAs remaining in the test set because not enough evidence is available in the training data to make inferences about the remaining HEAs (Figure A.1 c and Table A.1). The $\mathcal{D}_{\text{LTVC}}$ and $\mathcal{D}_{\text{AFLOW}}$ experimental results are identical(Figure A.1 d). Although the ERS performs better than the other recommendation systems in recovering one-half of the test HEAs in the $\mathcal{D}_{\text{LTVC}}$ data set (requiring only less than 3% of the number of possible trials), it cannot reliably recover the remaining one-quarter of the test HEAs owing to the lack of evidence in the training data (Table A.1).

A.2 Evaluation of HEA-recommendation capability by extrapolation

In the $\mathcal{D}_{\text{AFLOW}}^{\text{quaternary}}$ experiment, the ERS performs significantly better than the NMF-based recommender system, requiring less than 5 and 19% of the total number of possible HEA candidates to recall 50 and 75% of the HEAs in the test set, respectively (Table A.2). In the $\mathcal{D}_{\text{LTVC}}^{\text{quaternary}}$ experiment, the ERS and competitor matrix-based system developed using the first type of matrix representation require 13 and 32% and 14 and 41% of the total number of possible HEA candidates to recall 50 and 75% of the HEAs in the test set, respectively (Table A.2). Further investigation indicates

Table A.1: Ratio of number of trials (out of total number of possible trials) required to recall 50, 75, and 100% of HEAs in test set.

Data set	Model	Recall rates Half	Recall rates Three-quarters	Recall rates Full
$\mathcal{D}_{\text{ASMI16}}$	ERS	**12%**	**25%**	**80%**
	NMF (type 1)	16%	25%	92%
	NMF (type 2)	13%	26%	98%
	SVD (type 1)	31%	68%	99%
	SVD (type 2)	23%	64%	99%
	Decision Tree	77%	90%	99%
	Naïve Bayes	77%	90%	99%
	Logistic Regression	-	-	-
	SVM	-	-	-
$\mathcal{D}_{\text{CALPHAD}}$	ERS	**2%**	**5%**	92%
	NMF (type 1)	3%	7%	**89%**
	NMF (type 2)	3%	8%	93%
	SVD (type 1)	14%	28%	94%
	SVD (type 2)	17	37%	93%
	Decision Tree	39%	52%	94%
	Naïve Bayes	39%	52%	94%
	Logistic Regression	-	-	-
	SVM	-	-	-
$\mathcal{D}_{\text{AFLOW}}$	ERS	**2%**	8%	97%
	NMF (type 1)	3%	**6%**	96%
	NMF (type 2)	3%	6%	**85%**
	SVD (type 1)	16%	35%	99%
	SVD (type 2)	20%	50%	99%
	Decision Tree	31%	51%	99%
	Naïve Bayes	33%	53%	99%
	Logistic Regression	20%	29%	93%
	SVM	15%	26%	99%
$\mathcal{D}_{\text{LTVC}}$	ERS	**3%**	23%	97 %
	NMF (type 1)	4%	**6%**	96%
	NMF (type 2)	4%	7%	**86%**
	SVD (type 1)	14%	33%	99%
	SVD (type 2)	19%	52%	99%
	Decision Tree	32%	48%	99%
	Naïve Bayes	26%	42%	99%
	Logistic Regression	17%	26%	89%
	SVM	12%	26%	99%

that the ERS hardly recommends any quaternary alloys in $\mathcal{D}_{\text{LTVC}}^{\text{quaternary}}$ because these alloys cannot be generated by substituting elements in any of the ternary alloys in $\mathcal{D}_{\text{LTVC}}$ (Table A.3). Therefore, the properties of these alloys cannot be inferred from the evidence collected from $\mathcal{D}_{\text{LTVC}}$. As a result, the rankings obtained for these alloys are significantly low; therefore, the HEA recall rate is even lower than those obtained for randomly recommended HEAs. The results obtained for $\mathcal{D}_{\text{LTVC}}^{\text{quinary}}$ and

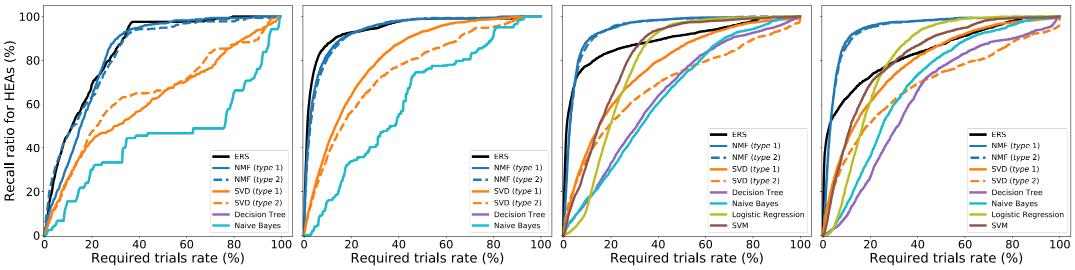
TABLE A.2: Ratio of number of trials (out of total number of possible trials) required to recall 50, 75, and 100% of HEAs in test set by extrapolating HEA-recommendation capability.

Data set	Model	Recall rates Half	 Three-quarters	 Full
$\mathcal{D}_{\text{AFLOW}}^{\text{quaternary}}$	ERS	**5%**	**19%**	**99%**
	NMF (type 1)	10%	24%	99%
	NMF (type 2)	50%	67%	99%
	SVD (type 1)	13%	32%	99%
	SVD (type 2)	53%	67%	99%
$\mathcal{D}_{\text{AFLOW}}^{\text{quinary}}$	ERS	**0.4%**	**1%**	**3%**
	NMF (type 1)	10%	56%	98%
	NMF (type 2)	9%	14%	47%
	SVD (type 1)	15%	27%	99%
	SVD (type 2)	8%	57%	99%
$\mathcal{D}_{\text{LTVC}}^{\text{quaternary}}$	ERS	**13%**	**32%**	**99%**
	NMF (type 1)	14%	41%	99%
	NMF (type 2)	50%	71%	99%
	SVD (type 1)	15%	39%	99%
	SVD (type 2)	53%	71%	99%
$\mathcal{D}_{\text{LTVC}}^{\text{quinary}}$	ERS	**0.07%**	**0.2%**	**2%**
	NMF (type 1)	11%	16%	47%
	NMF (type 2)	10%	53%	93%
	SVD (type 1)	15%	27%	99%
	SVD (type 2)	7%	54%	93%

$\mathcal{D}_{\text{AFLOW}}^{\text{quinary}}$ both show that the ERS drastically outperforms the capability of the competitor systems for recommending quinary HEAs. To recall 50, 75, and 100% of the HEAs from these data sets, 10–100 times fewer trials are required using the ERS than are required using the matrix-based recommender systems (Table A.2).

Table A.3: List of 63 quaternary HEAs in $\mathcal{D}^{\text{quaternary}}_{\text{LTVC}}$ that no evidence about their properties is found.

FeAuRePd
FeNiRePd
FeMoOsAg
FeRhPdOs
FeRePdRu
FeRePdCu
FeRePdOs
FeReRuAg
FeReOsAg
FePdRuOs
FePdCuOs
FePdOsCr
FeRuOsAg
AuNiPdOs
AuRhRePd
AuRhPdOs
AuRePdRu
AuRePdCu
AuRePdOs
AuRePdAg
NiMoPdOs
NiRhRePd
NiRhPdOs
NiCoRePd
NiRePdRu
NiRePdCu
NiRePtOs
NiPdRuOs
NiPdCuOs
NiPdOsCr
NiPdOsAg
MoRhPdOs
MoRePdOs
MoRePtOs
MoReOsAg
MoPdRuOs
MoRuPtOs
RhCoPdOs
RhRePdCu
RhRePtOs
RhPdOsAg
CoRePdRu
CoRePdCu
CoRePdOs
CoRePdAg
CoReRuPt
CoRePtOs
CoPdRuOs
CoRuPtOs
RePdRuOs
RePdPtOs
RePdOsAg
RePtOsAg
PdRuOsAg
PdCuOsCr
ReCuPtOs
RhRePdAg
NiRePdAg
AuNiReAg
ReRuPtOs
RhRePdOs
NiRePdOs
FeCuOsAg



(a) (b) (c) (d)

Figure A.1: Dependence of HEA recall ratio in the test sets on the number of trial required using *k*-fold cross-validation on (a) $\mathcal{D}_{\text{ASMI16}}$, (b) $\mathcal{D}_{\text{CALPHAD}}$, (c) $\mathcal{D}_{\text{AFLOW}}$, and (d) $\mathcal{D}_{\text{LTVC}}$ data sets.

Appendix B

Synthesis of FeMnCoNi high entropy alloy thin film

As a case study, we fabricated a high entropy alloy film of $Fe_{0.25}Co_{0.25}Mn_{0.25}Ni_{0.25}$. A 100 nm thick-thermal oxidized SiO_2/Si (100) substrate was used. After the organic solvent and deionized water cleaning, the substrate was loaded in a combinatorial multi target RF-sputtering system (COMET inc., CMS-6400). To identify the stable crystal structure and its composition dependence, a composition spread film was fabricated by combinatorial method (Koinuma and Takeuchi, 2004). For the composition spread film, we used two targets of FeCoMn (1:1:1) and Ni (3N grade). The base pressure was below 1×10^{-5} Pa, and Ar gas pressure was set as 0.3 Pa. To adjust the deposition rate as 0.23 ± 0.01 nm/s, RF-sputtering powers of FeCoMn and Ni targets were set at 100 and 120 W, respectively. To enhance the crystallinity, the sample was annealed at 400°C for 30 min under a vacuum condition below 6×10^{-3} Pa (Advanced RIKO, MILA-3000).

Figure 5(b) shows the sample structure. The composition film layer consists of three layers. One is a single FeCoMn layer with a thickness of 0.25 nm. The other layers are composition spread film formed by FeCoMn and Ni layers. For the composition-spread film deposition, during the FeCoMn layer deposition, a mask moved 18.5 mm at constant speed from a point 1.5 mm from the edge of the substrate to another end where the film thickness gradually changed. After that, the targets were changed to Ni. The mask moved to the opposite direction during the Ni film deposition. The total thickness of one unit of the 1-x(FeCoMn)-xNi composition spread layer/ FeCoMn stack structure is 0.5 nm. Alternating between the three deposition steps created composition-spread region with a width of 18.5 mm. The total film thickness in the composition-spread region was set to 100 nm. The composition spread was confirmed by an X-ray fluorescence spectrometer (XRF: Shimadzu, μEDX-1400) with a measuring spot diameter of 50 μm, as shown in Figure B.1.

The crystal structure was identified by X-ray diffraction (XRD). An XRD system with a 5-kW rotating anode Cu target x-ray source and a high-resolution 2D-detector (BRUKER AXS, D8 Discover Super Speed with GADDS) was used to determine the crystal structure. The 2D-detector system can detect part of the Debye–Scherrer ring rapidly and two-dimensionally (He, 2018).

In the evaluation of the phase separation temperature and magnetization properties of the other FeCoMn-X compositions, we found that the phase separation and inflection point were observed near 400°C. Therefore, we set the annealing temperature as 400°C. In the reported experiment, the annealing was performed at only 400°C; however, for FeCoMnNi, structural changes at higher temperatures are expected and are currently under investigation. Figure B.3 shows the XRD patterns of the sample as deposited and annealed. The BCC phase was confirmed for the annealed thin film sample at the equiatomical composition of FeCoMnNi (x=0.25).

Even at room temperature, a weak peak of the BCC can be observed for the FeCoM-rich composition.

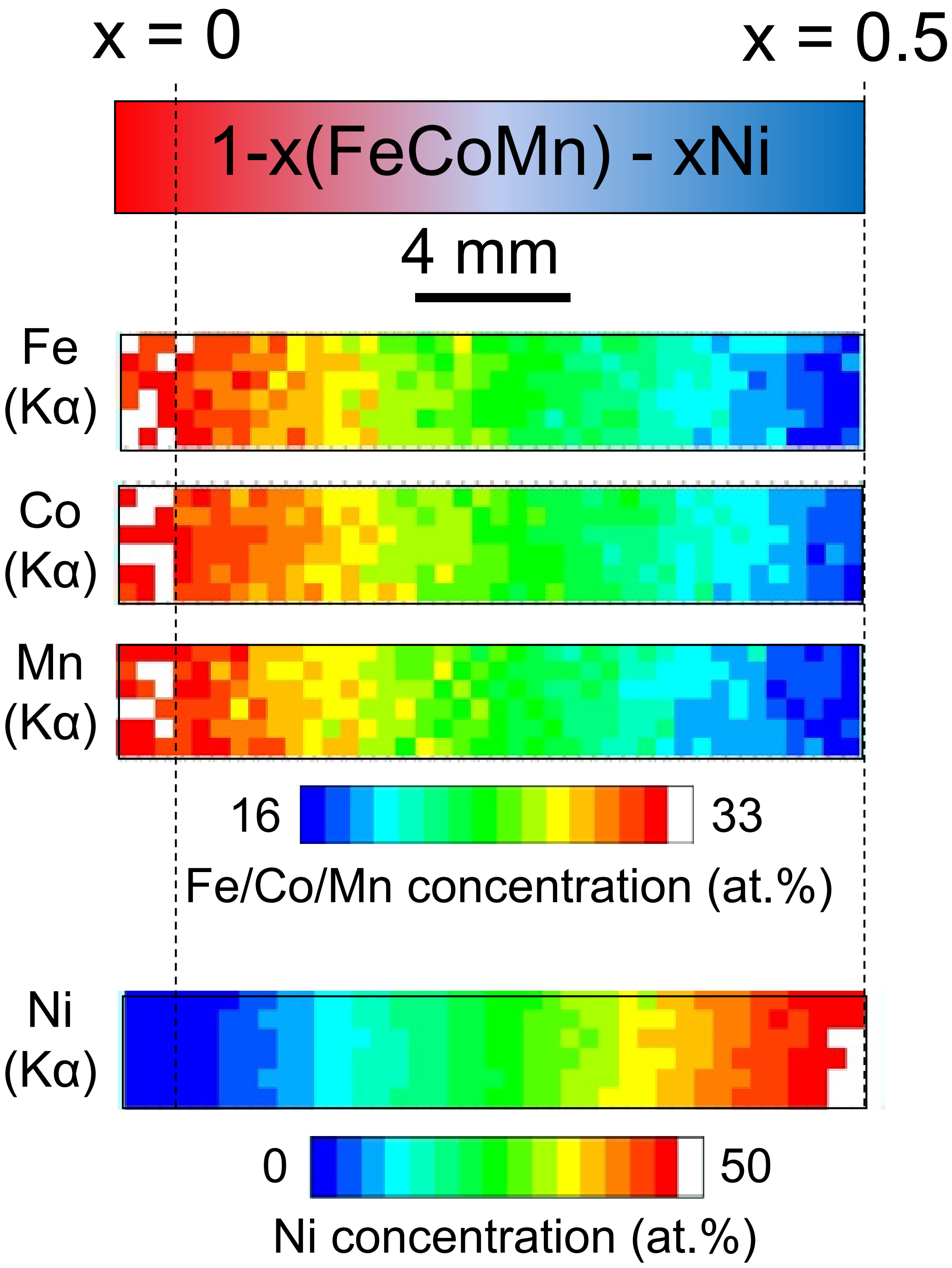


FIGURE B.1: Heatmapping image of Fe, Co, Mn, and Ni concentration estimated by EDX analysis. Composition was estimated from the XRF intensity of balk target materials and single-phase films of FeCoMn and Ni.

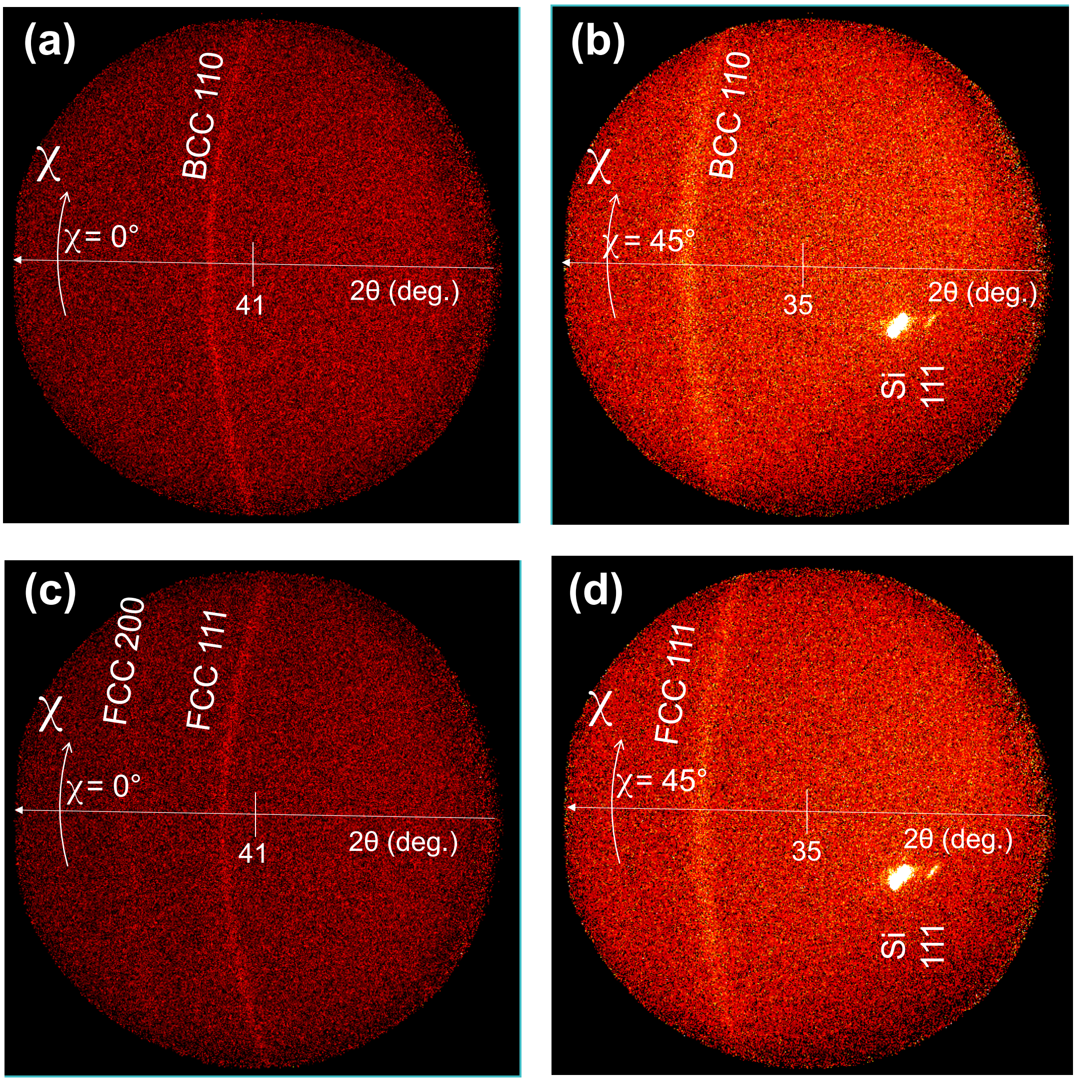


FIGURE B.2: 2D-XRD images at center γ angles of 0 and 45 $\circ$of FeCoMn-Ni films with low- (a,b) and high- (c,d) Ni concentrations. According to the powder diffraction pattern data baseGates-Rector and Blanton, 2019 (PDF 03-065-7519 and PDF 03-065-5131), for BCC, except for the reflection from (110), the signal intensities from other plane are not enough high to detect them in film form. So, the reflection from (110) is only detected. For FCC, in addition to the reflection from (111), the second strongest signal from (200) can barely be detected. The signals do not show no γ angle dependence, meaning the films are polycrystals in disordered crystal orientation.

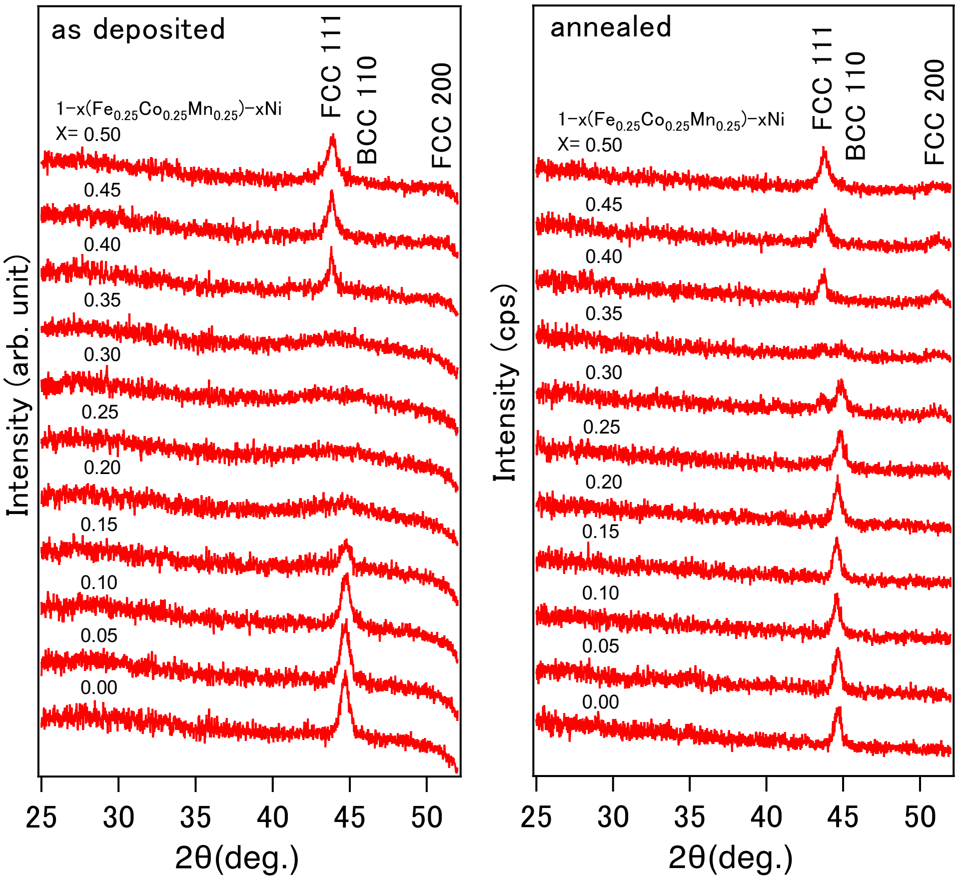


FIGURE B.3: XRD patterns of the as deposited and annealed at 400°C of FeCoMn-Ni film using an XRD system with a 5-kW rotating anode Cu target x-ray source. The BCC phase was confirmed for the annealed thin film sample at the equiatomical composition of FeCoMnNi (x=0.25)

Appendix C

Evaluation the eRSM using toy datasets

C.1 Datasets

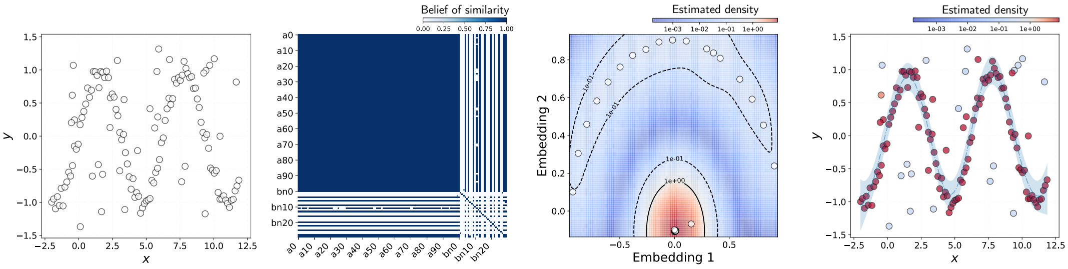
The details of the datasets investigated in this study are as follows.

- Noisy datasets $\mathcal{D}^{k}_{noisy}$: Two-dimensional synthesized signals with background noisy datasets consisting of 100 data instances generated from a nonlinear function $f(x) = \sin(x) + \mathcal{N}(0, 0.05^2)$ and several random background noise with ratio k. The data instances generated from $f(x)$ are denoted by the prefix a, whereas the random background noises are indicated by the prefix bn. The ratio k of random background noise bn varies from 20% to 90% of the total number of data instances generated from $f(x)$. For example, Figure C.1 a shows the synthesized dataset $\mathcal{D}^{30}_{noisy}$ with a ratio of the background noise bn of 30%.

- Bifurcate dataset $\mathcal{D}_{bifurcate}$: A two-dimensional synthesized dataset containing 150 data instances with a descriptive variable x ranging from 0.0 to 1.0 and a target variable y ranging from -1.0 to 1.0. $\mathcal{D}_{bifurcate}$ is generated from a mixture of three mechanisms. Regarding $x \leq 0.4$, y oscillates around 0.0 (A). Considering $x > 0.4$, y is bifurcated, with one branch oscillating and increasing from 0.0 to 1.0 (B), and the other oscillates and decreases from 0.0 to -1.0 (C) (Fig. C.4 a). The data instances generated from mechanisms A, B, and C are denoted by the prefixes a, b, or c, respectively. Regarding each mechanism, each data instance is represented by an index suffix that corresponds to an ascending value of x.

C.2 Evaluation of outlier detection capability

We apply the eRSM to detect outliers unrelated to the main mechanism in the synthesized noisy datasets $\mathcal{D}^{k}_{noisy}$. Figure C.1 b shows the similarity matrix M^{30}_{noisy} constructed by combining pieces of evidence using the Dempster–Shafer theory applied to the synthesized dataset $\mathcal{D}^{30}_{noisy}$ with a noise level bn of 30%. The background noise instances denoted by the prefix bn exhibit low similarity to all the other data instances. By contrast, data instances without the prefix bn are all similar in matrix M^{30}_{noisy}. To obtain a better visual interpretation of the obtained results, we apply the multidimensional scaling (MDS) algorithm (Kruskal, 1964; Borg, 1997; Mead, 1992) to represent the obtained similarities/dissimilarities between data instances as distances between points (representing the corresponding data instances) in a



(a) (b) (c) (d)

FIGURE C.1: (a) Sample dataset $\mathcal{D}^{30}_{noisy}$ with the noise level $bn - 30\%$. (b) Similarity matrix M^{30}_{noisy} constructed by applying the eRSM in dataset $\mathcal{D}^{30}_{noisy}$. The pattern and background instances are labelled with prefixes a and bn, respectively. (c) Projection of the data instances in $\mathcal{D}^{30}_{noisy}$ into a two-dimensional embedding space in which the distances correspond to the obtained similarity in M^{30}_{noisy}. The background color and contour lines indicate the densities of these data instances which are estimated using kernel density estimation (Silverman, 1998) in the projection space. (d) Visualization of data instances $\mathcal{D}^{30}_{noisy}$ colored according to the estimated density in the projection space. The blue dashed line and blue region show the mean and 2σ confidence interval of the regression models derived using the Gaussian process for inliers predicted using the local outlier factor (LOF) algorithm with eRSM.

low-dimensional space. This makes the distances correspond as closely as possible to the similarities/dissimilarities. Figure C.1 c shows a visual representation of the similarity matrix M^{30}_{noisy} by applying the MDS algorithm. Approximately, all the data instances are distributed primarily in the dense region (red), whereas the rest are scattered in the sparse region (blue). These data instances with high estimated density form a sine pattern in the original space; nevertheless, the remaining points are randomly located in the original space (Fig. C.1 d). We observed similar results for the datasets with different noise levels in this experiment (Fig. C.2).

To quantitatively assess the capability of the eRSM for outlier detection, we use the eRSM with the local outlier factor (LOF) algorithm (Breunig et al., 2000), which identifies outlier based on the density of its local neighborhood, and measure the precision and recall of the inlier and outlier labels in noisy datasets. Furthermore, we compare the results of the eRSM with those using the Euclidean distance in the descriptive space of the datasets (including the target y). The result of the comparison between the eRSM and Euclidean distance shows that the application of LOF with eRSM effectively removes noise (Fig. C.3), especially the datasets with high noise levels. These results indicate that the obtained similarities facilitate the detection of outliers unrelated to the dominant mechanism that generates the datasets.

C.3 Evaluation of capability in detecting the mixture mechanism

We now consider an application of the similarity measure to determine multiple mechanisms in a heterogeneous dataset. Figure C.4 b shows the similarity matrix M_{bif} obtained by the eRSM for all the data instances in $\mathcal{D}_{bif}$. This uses a mixture of

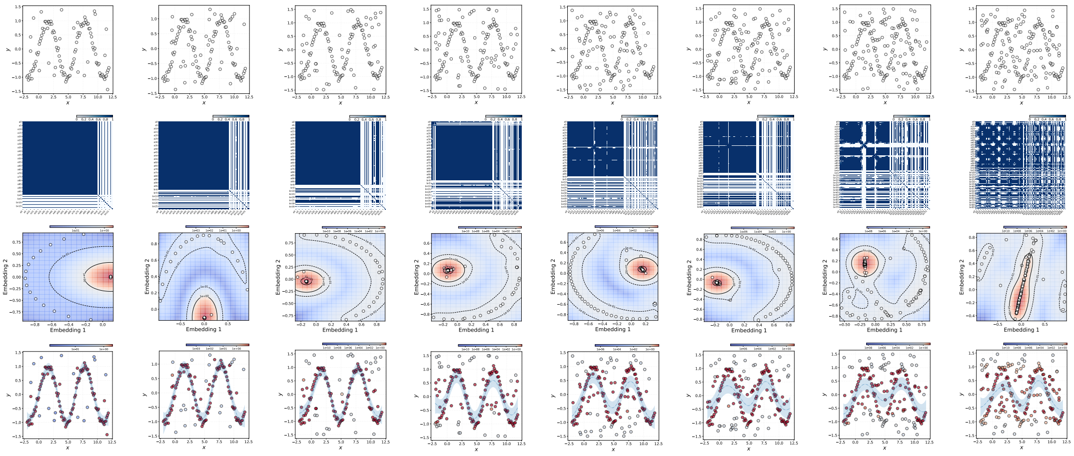


FIGURE C.2: Results of applying the eRSM to noisy datasets $\mathcal{D}^k_{noisy}$ with the noise level k from 20% to 90% (left to right). Plots in the first row visualize the noisy datasets. Similarity matrices constructed by applying the eRSM in these datasets are shown in the second row. The pattern and background instances are labelled with prefixes a and bn. In the next row, we exhibit projections of the data instances in each dataset into a two-dimensional embedding space in which the distances correspond to the obtained similarity. The background color and contour lines indicate the densities of these data instances, which are estimated using kernel density estimation (Silverman, 1998) in the projection space. The last row consists of visualizations of data instances colored according to the estimated density in the projection space. The blue dashed line and shaded region show the mean and 2σ confidence interval of the regression models derived using the Gaussian process (Williams and Rasmussen, 1996) for inliers predicted using the local outlier factor (LOF) algorithm (Breunig et al., 2000) with eRSM.

two Gaussian processes to collect the reference functions f_r. The dark blue cells in the similarity matrix indicate pairs of data instances associated with high similarity values and vice versa. Regarding the matrix, the similarity values among data instances with the same prefixes of indexes are homogeneous. These indicate that the data instances with the same prefix can be modeled using the same smooth function. By contrast, approximately all data instances with different prefixes show low similarity values, especially between prefixes b and c.

We apply the MDS algorithm to the similarity matrix M_{bif} to visualize the similarities between all the data instances in $\mathcal{D}_{bif}$. Figure C.4 c shows the projection of these data instances in a two-dimensional embedding space in which the distances are consistent with the obtained similarities. Regarding the embedding space, the data instances in $\mathcal{D}_{bif}$ are distributed in three well-separated clusters. We apply k-means clustering (MacQueen, 1967; Lloyd, 1982) to extract three groups of data instances in the embedding space. The clustering results show that the data instances belonging to each of the three clusters form different nonlinear smooth functions. This is consistent with the three mechanisms that generated the data in the original space (Fig. C.4 d). We compare the obtained result extracted by eRSM and k-means clustering with the result extracted by a mixture of three Gaussian processes (Lázaro-Gredilla, Van Vaerenbergh, and Lawrence, 2012). The mixture of GPs can distinguish mechanisms B and C; nonetheless, it cannot detect the difference between mechanisms A and B and between mechanisms A and C (Fig. C.5 d). It

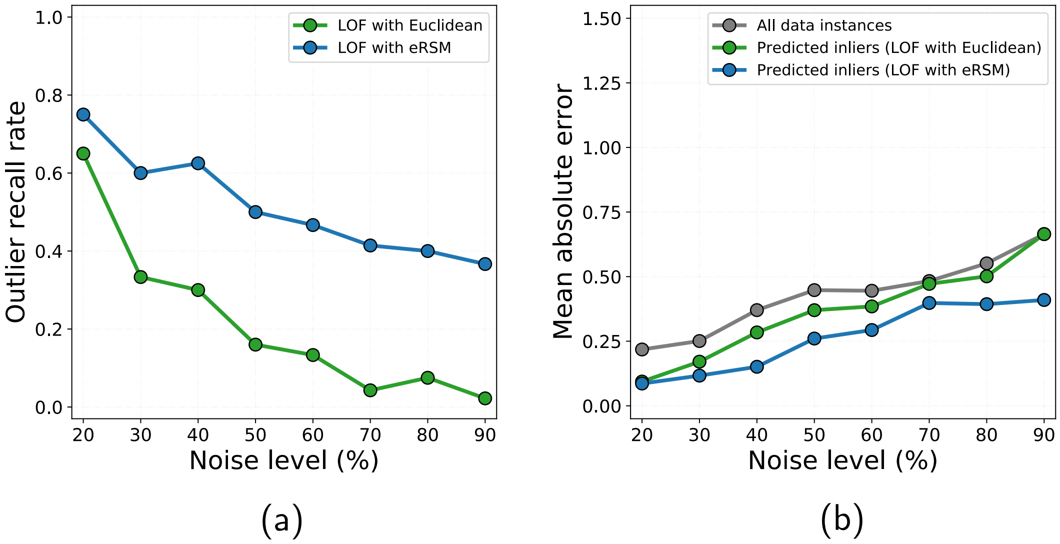
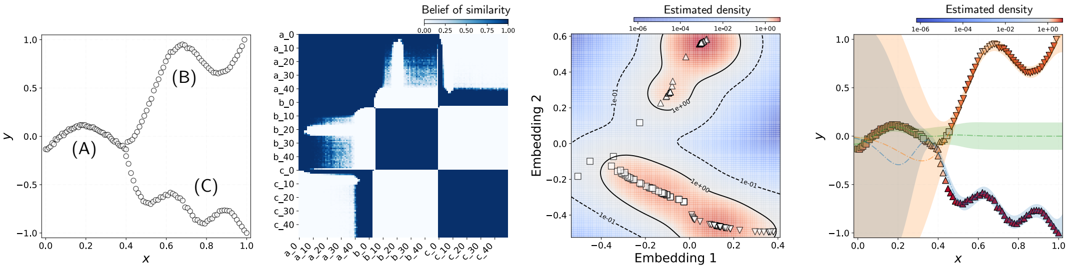


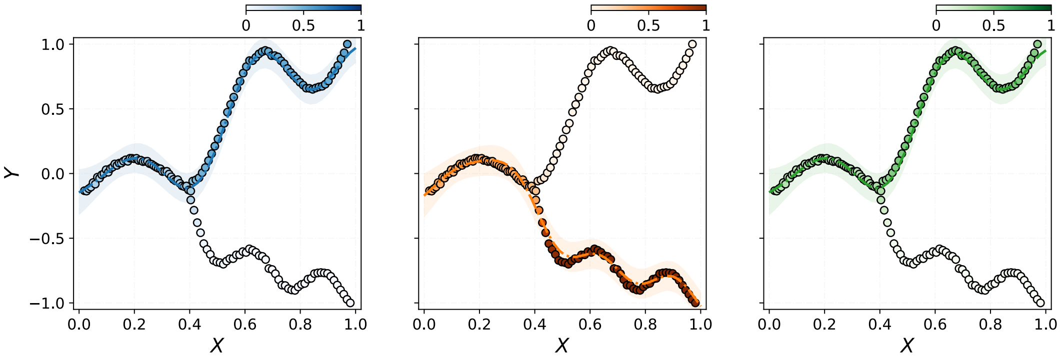
FIGURE C.3: Evaluation of the outlier detection capability of eRSM on datasets with different noise levels. (a) Outlier recall rate via local outlier factor (LOF) algorithm with eRSM (blue) and Euclidean distance (green). (b) Inlier prediction errors of regression models learned from all data instances (black), inliers predicted using the LOF algorithm with eRSM (blue), and Euclidean distance (green).

is difficult to discuss the differences between mechanisms A and B or mechanisms A and C if they are compared pairwise. However, mechanism A cannot be distinguished from mechanisms B or C, which shows a peculiar property of mechanism A. Therefore, the proposed eRSM has shown its capability to identify such complicated three-way relationships.



(a) (b) (c) (d)

FIGURE C.4: (a) Scatter plot of data instances in dataset $\mathcal{D}_{bif}$. (b) Heatmap showing the similarity matrix M_{bif} constructed by applying the eRSM in $\mathcal{D}_{bif}$. (c) Visualization of the data instances in a two-dimensional embedding space corresponding to their similarities M_{bif}. The background color and contour lines indicate the estimated densities of these data instances. The upward triangle, square, and downward triangle points indicate that the data instances are assigned to groups one, two, and three, respectively, using k-means clustering. (d) Visualization dataset $\mathcal{D}_{bif}$ colored according to the estimated density in the projection space. The blue, green, and orange dashed lines show the Gaussian process predictions of data instances from the predicted groups one, two, and three, respectively. The shaded regions indicate the 2σ confidence interval of these predictions.



Gaussian Process 1 Gaussian Process 2 Gaussian Process 3

FIGURE C.5: Mixture of three regression models (Gaussian processes) is regressed from the dataset $\mathcal{D}_{bif}$ using the mixture of Gaussian processes (MGP) (Lázaro-Gredilla, Van Vaerenbergh, and Lawrence, 2012). The dashed lines and shaded regions show the mean and 2σ confidence interval of the regression models. Colors of scatter points indicate the mixture weights of the regression models for each of data instance.

Appendix D

Supplementary materials for the eRSM

D.1 Modeling unreliability of evidence

When the evidence as a whole is itself affected by some uncertainty (unreliability), Shafer has suggested a discounting factor in his research (Shafer, 1976) to consider the new uncertainty. Discounting operator, an application of the concept, was explained in Smets's research (Smets, 1993). According to such studies, we adopt an discounting factor $\gamma_{i,j}$ to describe the unreliability of evidence about the similarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$ collected from a source of evidence f_r. We define the factor as follows:

$$\gamma_{i,j} = (e^{\frac{\bar{\sigma}}{\Delta_y}} + 1) \times (\frac{\sigma_{\boldsymbol{x}_i}}{\bar{\sigma}} + 1) \times (\frac{\sigma_{\boldsymbol{x}_j}}{\bar{\sigma}} + 1), \tag{D.1}$$

where Δ_y is the variation range of the target variable y in the dataset $\mathcal{D}$. $\bar{\sigma}$ is the average of the predictive standard error of all the data instances in $\mathcal{D}_{ref}$, a subset of dataset $\mathcal{D}$ is used to build the function f_r.

Given a source of evidence f_r, the $\gamma_{i,j}$ considers the unreliability of collected evidence about the similarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$ is caused due to uncertainties from:

- Source of evidence f_r: We use a function f_r regressed from $\mathcal{D}_{ref}$ ($\subset \mathcal{D}$) to consider the similarity/dissimilarity between remaining data instances in $\mathcal{D}$. So, function f_r with higher accuracy is expected to provide more reliable evidence about the similarities between these data instances. The first component of the discounting factor $(e^{\frac{\bar{\sigma}}{\Delta_y}} + 1)$ is used to describe the uncertainty from f_r.
- Prediction for x_i (or x_j): Given f_r, we use $\Delta_i = |y_i - f_r(\boldsymbol{x}_i)|$, the deviation from the true to the predicted target values of data instance i, and $\sigma_{\boldsymbol{x}_i}$, standard error of the prediction, to model the probability $p(O_i|f_r)$ that instance i is generated from f_r. However, in the case that the f_r does not make sure about its prediction ($\sigma_{\boldsymbol{x}_i} \rightarrow +\infty$), the probability $p(O_i|f_r)$ approximates to 1 even the prediction error Δ_i is large. Therefore, we use $(\frac{\sigma_{\boldsymbol{x}_i}}{\bar{\sigma}} + 1)$ and $(\frac{\sigma_{\boldsymbol{x}_j}}{\bar{\sigma}} + 1)$ to describe the prediction uncertainty of f_r for x_i and x_j, respectively.

The smaller $\bar{\sigma}$ is relative to Δ_y, the more reliable the learned regression function f_r is. Also, when $\sigma_{\boldsymbol{x}_i}$ and $\sigma_{\boldsymbol{x}_j}$ are smaller than $\bar{\sigma}$, f_r can provide reliable evidence for the relationship between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$. By contrast, when $\sigma_{\boldsymbol{x}_i}$ and $\sigma_{\boldsymbol{x}_j}$ are large compared to $\bar{\sigma}$, f_r cannot provide reliable evidence for the relationship between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$.

D.2 Combination method for the eRSM and Jaccard Index

To identify the subgroup that unobserved alloys belong to, we propose an approach that constructs an embedding space for the observed and unobserved alloys from a distance matrix M, which is an element-wise product of the distance matrix obtained by using our proposed similarity measurement, the eRSM, for the 36 numerical descriptors and the distance matrix obtained by applying the Jaccard index for four categorical features. Considering any pair of two materials a_i and a_j. The combined distance is defined as follows:

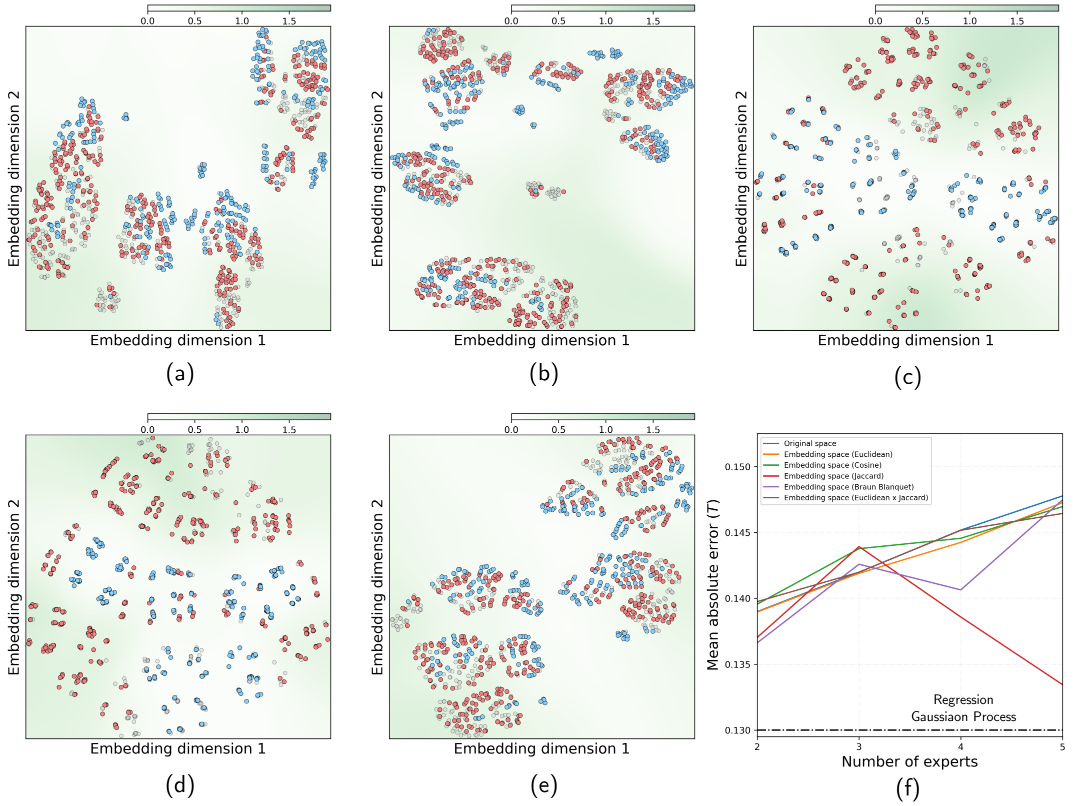
$$M[i,j] = (m^{i,j}(\{ds\}) + \frac{m^{i,j}(\{s,ds\})}{2}) \times (1 - J(a_i,a_j)) \tag{D.2}$$

, where $m^{i,j}$ is the obtained similarity information from the eRSM and $J(a_i,a_j)$ is Jaccard similarity coefficient between the two materials. It should be noted that if one of the two alloys is unobserved in dataset, the eRSM considers no information about the similarity between the alloys ($m^{i,j}(\{s,ds\}) = 1$ and $m^{i,j}(\{ds\}) = 0$). Therefore, the combined distance M[i,j]=$\frac{1-J(a_i,a_j)}{2}$.

D.3 Comparison between the eRSM and traditional similarity measurements

We perform an additional experiment to evaluate the effectiveness of our proposed similarity measure for detecting subgroups of quaternary alloys in the quaternary alloy dataset $\mathcal{D}_{quaternary}$. In this experiment, we construct six mixtures of experts (MoE), in which their gating functions are learned by applying the K-Means clustering on the material features (36 selected compositional descriptors) and five embedding spaces. The four embedding spaces are developed using the tSNE algorithms with four traditional similarity/dissimilarity measures: 1) Euclidean distance, 2) Cosine distance, 3) Jaccard index, and 4) Braun-Blanquet similarity coefficient. The Euclidean and cosine distances are applied to the materials space of 36 selected compositional descriptors. In contrast, the Jaccard index and Braun-Blanquet similarity coefficient are applied to categorical descriptors of the dataset. In addition, the last embedding space is constructed using a distance matrix that is an element-wise product of the distance matrix obtained from the euclidean distance and distance matrix obtained Jaccard index. It should be noted that the expert (regression model) of each detected cluster is learned by applying Gaussian Process to the alloys in the cluster; these alloys are represented by 36 selected compositional descriptors.

Figure D.1 a-e show visualization in the five embedding spaces of quaternary alloys in two subgroups detected using our method, G_1^{Mag} (red points) and G_2^{Mag} (blue points). In the embedding spaces built using the Jaccard Index and Braun-Blanquet similarity coefficient, we can identify the local regions in which the alloys belong to the same subgroup (G_1^{Mag} or G_2^{Mag}). However, such embedding spaces do not maintain a global structure like the embedding space obtained by using the eRSM (Figure 4 c). In contrast, embedding spaces constructed from the euclidean, cosine distances and the combination of euclidean and Jaccard index cannot provide valuable information to distinguish alloys from the two subgroups G_1^{Mag} and G_2^{Mag}, making it challenging to distinguish the alloys in the two subgroups. As a result, the cross-validation mean absolute errors (MAE) of the constructed MoEs based on clusters in the embedding spaces constructed from the Jaccard Index and Braun-Blanquet



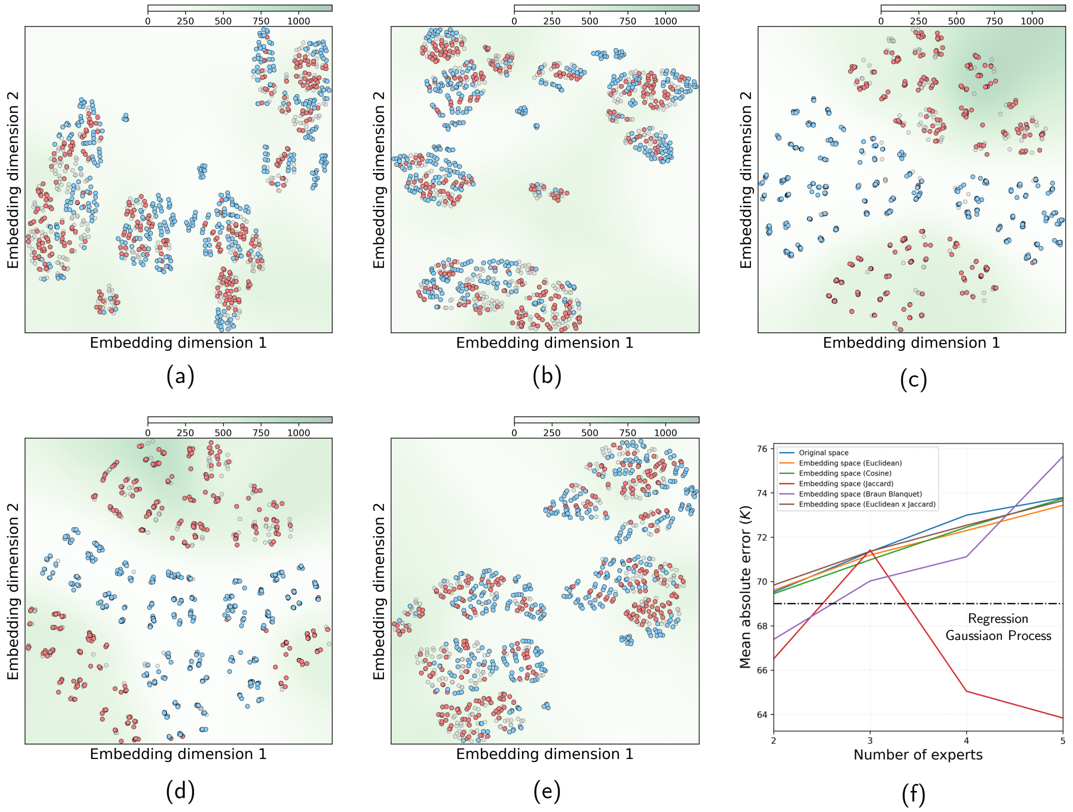
(a) (b) (c)

(d) (e) (f)

FIGURE D.1: (a-e) Visualization in the five embedding spaces of two subgroups of alloys G_1^{Mag} (red points) and G_2^{Mag} (blue points), which are detected by using the similarity information obtained from the eRSM. These spaces are constructed by applying (a) Euclidean distance, (b) Cosine distance, (c) Jaccard index, (d) Braun-Blanquet similarity coefficient, and (e) a combination between Euclidean distance and Jaccard index. (f) Cross-validation prediction error for magnetization of Mixtures of Experts in which their gating functions are learned by using the K-means algorithms on the embedding space and material features space.

similarity coefficient exhibit better performance for predicting magnetization than the MoEs on other embedding spaces and original space (Fig. D.1 f). However, the results are significantly higher than those of a Gaussian Process Regression and the constructed MoE based on detected subgroups G_1^{Mag} and G_2^{Mag}.

Similarly, Figure D.2 a-e show visualization in the five embedding spaces of quaternary alloys in two subgroups detected using our method, $G_1^{T_C}$ (red points) and $G_2^{T_C}$ (blue points). In the embedding spaces built from the Jaccard Index and Braun-Blanquet similarity coefficient, we can identify the local regions in which the alloys belong to the same subgroup ($G_1^{T_C}$ or $G_2^{T_C}$) and construct MoEs exhibiting better performance for predicting the Curie temperature than those on other embedding spaces and original space (Fig. D.2 f). In addition, the prediction errors of the MoEs based on the clusters detected from the embedding space, which are constructed using the Jaccard Index, are lower than the result of a Gaussian Process Regression but are still significantly higher than the result of the MoE based on subgroups that are detected by using the eRSM.



(a) (b) (c)

(d) (e) (f)

FIGURE D.2: (a-e) Visualization in the five embedding spaces of two subgroups of alloys $G_1^{T_C}$ (red points) and $G_2^{T_C}$ (blue points), which are detected by using the similarity information obtained from the eRSM. These spaces are constructed by applying (a) Euclidean distance, (b) Cosine distance, (c) Jaccard index, (d) Braun-Blanquet similarity coefficient, and (e) a combination between Euclidean distance and Jaccard index. (f) Cross-validation prediction error for Curie temperature of Mixtures of Experts in which their gating functions are learned by using the K-means algorithms on the embedding space and material features space.

Bibliography

Aggarwal, Charu C. et al. (Jan. 2014). “Active learning: A survey”. English (US). In: *Data Classification*. Publisher Copyright: © 2015 by Taylor & Francis Group, LLC. CRC Press, pp. 571–605. ISBN: 9781466586741. DOI: 10.1201/b17320.

Agrawal, Ankit and Alok Choudhary (2016). “Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science”. In: *APL Materials* 4.5, p. 053208. DOI: 10.1063/1.4946894.

Akai, Hisazumi et al. (Feb. 1990). “Theory of Hyperfine Interactions in Metals”. In: *Progress of Theoretical Physics Supplement* 101, pp. 11–77. ISSN: 0375-9687. DOI: 10.1143/PTP.101.11. URL: https://doi.org/10.1143/PTP.101.11.

al., Peter Blake et. (June 2008a). “Graphene-Based Liquid Crystal Device”. In: *Nano Letters* 8.6, pp. 1704–1708. DOI: 10.1021/nl080649i. URL: https://doi.org/10.1021/nl080649i.

al., Yenny Hernandez et. (Aug. 2008b). “High-yield production of graphene by liquid-phase exfoliation of graphite”. In: *Nature Nanotechnology* 3.9, pp. 563–568. DOI: 10.1038/nnano.2008.215. URL: https://doi.org/10.1038/nnano.2008.215.

Alman, David (Oct. 2013). “Searching for Next Single-Phase High-Entropy Alloy Compositions”. In: *Entropy* 15, pp. 4504–4519. DOI: 10.3390/e15104504.

Andersson, J-O et al. (2002). “Thermo-Calc DICTRA, computational tools for materials science”. In: *Calphad* 26.2, pp. 273–312. ISSN: 0364-5916. DOI: https://doi.org/10.1016/S0364-5916(02)00037-8. URL: http://www.sciencedirect.com/science/article/pii/S0364591602000378.

Arrow, Kenneth J. (1971). *Essays in the theory of risk-bearing*. Markham economics series. Chicago: Markham Pub. Co. ISBN: 0841020019; 9780841020016.

Bardeen, J., L. N. Cooper, and J. R. Schrieffer (1957). “Theory of Superconductivity”. In: *Phys. Rev.* 108 (5), pp. 1175–1204. DOI: 10.1103/PhysRev.108.1175. URL: https://link.aps.org/doi/10.1103/PhysRev.108.1175.

Barredo Arrieta, Alejandro et al. (2020). “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI”. In: *Information Fusion* 58, pp. 82–115. ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2019.12.012. URL: https://www.sciencedirect.com/science/article/pii/S1566253519308103.

Bissantz, Caterina, Bernd Kuhn, and Martin Stahl (July 2010). “A Medicinal Chemist’s Guide to Molecular Interactions”. In: *Journal of Medicinal Chemistry* 53.14, pp. 5061–5084. DOI: 10.1021/jm100112j. URL: https://doi.org/10.1021/jm100112j.

Borg, I.; Groenen P. (1997). *Modern Multidimensional Scaling - Theory and Applications*. Springer Series in Statistics.

Bourlinos, Athanasios B. et al. (Aug. 2009). “Liquid-Phase Exfoliation of Graphite Towards Solubilized Graphenes”. In: *Small* 5.16, pp. 1841–1845. DOI: 10.1002/smll.200900242. URL: https://doi.org/10.1002/smll.200900242.

Breiman, Leo (2001). “Random Forests”. In: *Machine Learning* 45.1, pp. 5–32. DOI: 10.1023/A:1010933404324. URL: https://doi.org/10.1023/A:1010933404324.

Breunig, Markus et al. (June 2000). “LOF: Identifying Density-Based Local Outliers.” In: *ACM Sigmod Record* 29, pp. 93–104. DOI: 10.1145/342009.335388.

Brighton, Henry and Gerd Gigerenzer (2012). “Are rational actor models “rational” outside small worlds”. In: *Evolution and rationality: Decisions, co-operation, and strategic behavior*, pp. 84–109.

Cai, Minzhen et al. (2012). “Methods of graphite exfoliation”. In: *Journal of Materials Chemistry* 22.48, p. 24992. DOI: `10.1039/c2jm34517j`. URL: `https://doi.org/10.1039/c2jm34517j`.

Cantor, B. et al. (2004). “Microstructural development in equiatomic multicomponent alloys”. In: *Mater. Sci. Eng. A* 375, pp. 213 –218. ISSN: 0921-5093. DOI: `https://doi.org/10.1016/j.msea.2003.10.257`.

Ceriotti, Michele (2019). “Unsupervised machine learning in atomistic simulations, between predictions and understanding”. In: *The Journal of chemical physics* 150.15, p. 150901.

Chen, Chaofan et al. (2019). “This Looks Like That: Deep Learning for Interpretable Image Recognition”. In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc. URL: `https://proceedings.neurips.cc/paper/2019/file/adf7ee2dcf142b0e11888e72b43fcb75-Paper.pdf`.

Cheng, Bingqing et al. (Sept. 2020). “Mapping Materials and Molecules”. In: *Accounts of Chemical Research* 53.9, pp. 1981–1991. DOI: `10.1021/acs.accounts.0c00403`. URL: `https://doi.org/10.1021/acs.accounts.0c00403`.

Chow, C (1970). “On optimum recognition error and reject tradeoff”. In: *IEEE Transactions on information theory* 16.1, pp. 41–46.

Chryssolouris, George, Moshin Lee, and Alvin Ramsey (1996). “Confidence interval prediction for neural network models”. In: *IEEE Transactions on neural networks* 7.1, pp. 229–232.

Cover, T. and P. Hart (1967). “Nearest neighbor pattern classification”. In: *IEEE Transactions on Information Theory* 13.1, pp. 21–27. DOI: `10.1109/TIT.1967.1053964`.

Cui, Peng et al. (2018). “Effect of Ti on microstructures and mechanical properties of high entropy alloys based on CoFeMnNi system”. In: *Mater. Sci. Eng. A* 737, pp. 198–204. ISSN: 0921-5093. DOI: `https://doi.org/10.1016/j.msea.2018.09.050`.

Cui, Xu et al. (2011). “Liquid-phase exfoliation, functionalization and applications of graphene”. In: *Nanoscale* 3.5, 2118—2126. ISSN: 2040-3364. DOI: `10.1039/c1nr10127g`. URL: `https://doi.org/10.1039/c1nr10127g`.

Dam, Hieu Chi et al. (2018). “Important Descriptors and Descriptor Groups of Curie Temperatures of Rare-earth Transition-metal Binary Alloys”. In: *Journal of the Physical Society of Japan* 87.11, p. 113801. DOI: `10.7566/JPSJ.87.113801`.

Dempster, A. P. (Apr. 1967). “Upper and Lower Probabilities Induced by a Multivalued Mapping”. In: *The Annals of Mathematical Statistics* 38.2, pp. 325–339. DOI: `10.1214/aoms/1177698950`. URL: `https://doi.org/10.1214/aoms/1177698950`.

— (1968). “A Generalization of Bayesian Inference”. In: *J R STAT SOC B* 30.2, pp. 205–232. DOI: `10.1111/j.2517-6161.1968.tb00722.x`.

Deng, Yong (Apr. 2014). “Generalized Evidence Theory”. In: *Applied Intelligence* 43. DOI: `10.1007/s10489-015-0661-2`.

Denœux, Thierry, Didier Dubois, and Henri Prade (2020). “Representations of Uncertainty in Artificial Intelligence: Beyond Probability and Possibility”. In: *A Guided Tour of Artificial Intelligence Research*. Ed. by P. Marquis, O. Papini, and H. Prade. Vol. 1. Springer Verlag. Chap. 4, pp. 119–150. DOI: `10.1007/978-3-030-06164-7_4`.

Der Kiureghian, Armen and Ove Ditlevsen (2009). “Aleatory or epistemic? Does it matter?” In: *Structural safety* 31.2, pp. 105–112.

DeVries, Terrance and Graham W. Taylor (2018). *Learning Confidence for Out-of-Distribution Detection in Neural Networks*. arXiv: 1802.04865 [stat.ML].

Dreyer, Daniel R. et al. (2010). "The chemistry of graphene oxide". In: *Chem. Soc. Rev.* 39 (1), pp. 228–240. DOI: 10.1039/B917103G. URL: http://dx.doi.org/10.1039/B917103G.

Drosback, Meredith (2014). "Materials Genome Initiative: Advances and Initiatives". In: *JOM* 66.3, pp. 334–335. DOI: 10.1007/s11837-014-0894-2. URL: https://doi.org/10.1007/s11837-014-0894-2.

Du, Wencheng et al. (May 2013). "Organic salt-assisted liquid-phase exfoliation of graphite to produce high-quality graphene". In: *Chemical Physics Letters* 568-569, pp. 198–201. DOI: 10.1016/j.cplett.2013.03.060. URL: https://doi.org/10.1016/j.cplett.2013.03.060.

Dubois, Didier and Henri Prade (2012). *Possibility theory: an approach to computerized processing of uncertainty*. Springer Science & Business Media.

Dubois, Didier, Henri Prade, and Philippe Smets (1996). "Representing partial ignorance". In: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 26.3, pp. 361–377.

Economopoulos, Solon P. et al. (Nov. 2010). "Exfoliation and Chemical Modification Using Microwave Irradiation Affording Highly Functionalized Graphene". In: *ACS Nano* 4.12, pp. 7499–7507. DOI: 10.1021/nn101735e. URL: https://doi.org/10.1021/nn101735e.

Eigler, Siegfried et al. (2013). "Graphene oxide: efficiency of reducing agents". In: *Chemical Communications* 49.67, p. 7391. DOI: 10.1039/c3cc43612h. URL: https://doi.org/10.1039/c3cc43612h.

Eyke, Natalie S, William H Green, and Klavs F Jensen (2020). "Iterative experimental design based on active machine learning reduces the experimental burden associated with reaction screening". In: *Reaction Chemistry & Engineering* 5.10, pp. 1963–1972.

Feng, Hongbin et al. (2013). "A low-temperature method to produce highly reduced graphene oxide". In: *Nature Communications* 4.1, p. 1539. DOI: 10.1038/ncomms2555. URL: https://doi.org/10.1038/ncomms2555.

Flach, Peter A. (2017). "Classifier Calibration". In: *Encyclopedia of Machine Learning and Data Mining*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, pp. 210–217. ISBN: 978-1-4899-7687-1. DOI: 10.1007/978-1-4899-7687-1_900. URL: https://doi.org/10.1007/978-1-4899-7687-1_900.

Fukushima, T. et al. (2022). "Automatic exhaustive calculations of large material space by Korringa-Kohn-Rostoker coherent potential approximation method applied to equiatomic quaternary high entropy alloys". In: *Phys. Rev. Materials* 6 (2), p. 023802. DOI: 10.1103/PhysRevMaterials.6.023802. URL: https://link.aps.org/doi/10.1103/PhysRevMaterials.6.023802.

Gal, Yarin and Zoubin Ghahramani (2016). "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning". In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by Maria Florina Balcan and Kilian Q. Weinberger. Vol. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, pp. 1050–1059. URL: https://proceedings.mlr.press/v48/gal16.html.

Gates-Rector, Stacy and Thomas Blanton (2019). "The Powder Diffraction File: a quality materials characterization database". In: *Powder Diffr.* 34.4, pp. 352–360. DOI: 10.1017/S0885715619000812.

George, Easo P., Dierk Raabe, and Robert O. Ritchie (2019). "High-entropy alloys". In: *Nat. Rev. Mater.* 4.8, pp. 515–534. DOI: 10.1038/s41578-019-0121-4.

Gigerenzer, Gerd and Wolfgang Gaissmaier (2011). “Heuristic Decision Making”. In: *Annual Review of Psychology* 62.1. PMID: 21126183, pp. 451–482. DOI: `10.1146/annurev-psych-120709-145346`. URL: `https://doi.org/10.1146/annurev-psych-120709-145346`.

Gneiting, Tilmann and Adrian Raftery (Mar. 2007). “Strictly Proper Scoring Rules, Prediction, and Estimation”. In: *Journal of the American Statistical Association* 102, pp. 359–378. DOI: `10.1198/016214506000001437`.

Goldsmith, Bryan R et al. (2017). “Uncovering structure-property relationships of materials by subgroup discovery”. In: *New Journal of Physics* 19.1, p. 013031. DOI: `10.1088/1367-2630/aa57c2`. URL: `https://doi.org/10.1088/1367-2630/aa57c2`.

Goldstone, Robert L., Douglas L. Medin, and Jamin Halberstadt (1997). “Similarity in context”. In: *Memory & Cognition* 25.2, pp. 237–255. DOI: `10.3758/BF03201115`. URL: `https://doi.org/10.3758/BF03201115`.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep learning*. MIT press.

Graves, Alex (2011). “Practical variational inference for neural networks”. In: *Advances in neural information processing systems* 24.

Gu, Xingxing et al. (Oct. 2018). “Highly Reversible Li–Se Batteries with Ultra-Lightweight N,S-Codoped Graphene Blocking Layer”. In: *Nano-Micro Letters* 10. DOI: `10.1007/s40820-018-0213-5`.

Guardia, L. et al. (Apr. 2011). “High-throughput production of pristine graphene in an aqueous dispersion assisted by non-ionic surfactants”. In: *Carbon* 49.5, pp. 1653–1662. DOI: `10.1016/j.carbon.2010.12.049`. URL: `https://doi.org/10.1016/j.carbon.2010.12.049`.

Gubaev, Konstantin, Evgeny V Podryabinkin, and Alexander V Shapeev (2018). “Machine learning of molecular properties: Locality and active learning”. In: *The Journal of chemical physics* 148.24, p. 241727.

GUO, Sheng and C.T. LIU (2011). “Phase stability in high entropy alloys: Formation of solid-solution phase or amorphous phase”. In: *Prog. Nat. Sci.* 21, pp. 433–446. DOI: `https://doi.org/10.1016/S1002-0071(12)60080-X`.

Ha, Minh-Quyet et al. (2021). “Evidence-based recommender system for high-entropy alloys”. In: *Nature Computational Science* 1.7, pp. 470–478. DOI: `10.1038/s43588-021-00097-w`. URL: `https://doi.org/10.1038/s43588-021-00097-w`.

Hamilton, Christopher E. et al. (Oct. 2009). “High-Yield Organic Dispersions of Unfunctionalized Graphene”. In: *Nano Letters* 9.10, pp. 3460–3462. DOI: `10.1021/nl9016623`. URL: `https://doi.org/10.1021/nl9016623`.

Hao, Rui et al. (Nov. 2008). “Aqueous dispersions of TCNQ-anion-stabilized graphene sheets”. In: *Chemical communications (Cambridge, England)* 48, pp. 6576–6578. DOI: `10.1039/b816971c`.

Hasan, Tawfique et al. (Aug. 2007). “Stabilization and Debundling of Single-Wall Carbon Nanotube Dispersions in N-Methyl-2-pyrrolidone (NMP) by Polyvinylpyrrolidone (PVP)”. In: *The Journal of Physical Chemistry C* 111.34, pp. 12594–12602. DOI: `10.1021/jp0723012`. URL: `https://doi.org/10.1021/jp0723012`.

He, B. B. (2018). “Geometry and Fundamentals”. In: *Two-dimensional X-ray Diffraction*. John Wiley & Sons, Ltd. Chap. 2, pp. 29–55. ISBN: 9781119356080. DOI: `https://doi.org/10.1002/9781119356080.ch2`. URL: `https://doi.org/10.1002/9781119356080.ch2`.

Hearst, Marti A. (July 1998). “Support Vector Machines”. In: *IEEE Intell. Syst.* 13.4, pp. 18–28. ISSN: 1541-1672. DOI: `10.1109/5254.708428`.

Hellman, Martin E (1970). “The nearest neighbor classification rule with a reject option”. In: *IEEE Transactions on Systems Science and Cybernetics* 6.3, pp. 179–185.

Hieu-Chi, Dam (Jan. 2023). “Datasets of binary and quaternary alloys with Curie temperature and magnetization for the eRSM”. In: DOI: 10.5281/zenodo.7540840. URL: https://doi.org/10.5281/zenodo.7540840.

Hong, Seungbum et al. (Mar. 2021). “Reducing Time to Discovery: Materials and Molecular Modeling, Imaging, Informatics, and Integration”. In: *ACS Nano* 15.3, pp. 3971–3995. DOI: 10.1021/acsnano.1c00211. URL: https://doi.org/10.1021/acsnano.1c00211.

Hora, Stephen C (1996). “Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management”. In: *Reliability Engineering & System Safety* 54.2-3, pp. 217–223.

Hossein Zadeh Bazargani, Mehran and Brian Mac Namee (Apr. 2020). “The Elliptical Basis Function Data Descriptor (EBFDD) Network: A One-Class Classification Approach to Anomaly Detection”. In: pp. 107–123. ISBN: 978-3-030-46149-2. DOI: 10.1007/978-3-030-46150-8_7.

Huang, Wenjiang, Pedro Martin, and Houlong L. Zhuang (2019). “Machine-learning phase prediction of high-entropy alloys”. In: *Acta Mater.* 169, pp. 225–236. ISSN: 1359-6454. DOI: https://doi.org/10.1016/j.actamat.2019.03.012.

Huhn, William Paul and Michael Widom (2013). “Prediction of A2 to B2 Phase Transition in the High-Entropy Alloy Mo-Nb-Ta-W”. In: *JOM* 65.12, pp. 1772–1779. DOI: 10.1007/s11837-013-0772-3.

Hullermeier, Eyke and Klaus Brinker (2008). “Learning valued preference structures for solving classification problems”. In: *Fuzzy Sets and Systems* 159.18, pp. 2337–2352.

Hüllermeier, Eyke and Willem Waegeman (2021). “Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods”. In: *Machine Learning* 110.3, pp. 457–506. DOI: 10.1007/s10994-021-05946-3. URL: https://doi.org/10.1007/s10994-021-05946-3.

Hummers, William S. and Richard E. Offeman (Mar. 1958). “Preparation of Graphitic Oxide”. In: *Journal of the American Chemical Society* 80.6, pp. 1339–1339. DOI: 10.1021/ja01539a017. URL: https://doi.org/10.1021/ja01539a017.

Hwang, JT Gene and A Adam Ding (1997). “Prediction intervals for artificial neural networks”. In: *Journal of the American Statistical Association* 92.438, pp. 748–757.

Isayev, Olexandr et al. (2015). “Materials cartography: representing and mining materials space using structural and electronic fingerprints”. In: *Chemistry of Materials* 27.3, pp. 735–743.

Jaeger, Sabrina, Simone Fulle, and Samo Turk (Jan. 2018). “Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition”. In: *Journal of Chemical Information and Modeling* 58.1, pp. 27–35. DOI: 10.1021/acs.jcim.7b00616. URL: https://doi.org/10.1021/acs.jcim.7b00616.

Jain, A.K., Jianchang Mao, and K.M. Mohiuddin (1996). “Artificial neural networks: a tutorial”. In: *Computer* 29.3, pp. 31–44. DOI: 10.1109/2.485891.

Jiang, Guodong et al. (July 2011). “TiO_2 nanoparticles assembled on graphene oxide nanosheets with high photocatalytic activity for removal of pollutants”. In: *Carbon* 49.8, pp. 2693–2701. DOI: 10.1016/j.carbon.2011.02.059. URL: https://doi.org/10.1016/j.carbon.2011.02.059.

Jiménez-Luna, José, Francesca Grisoni, and Gisbert Schneider (2020a). “Drug discovery with explainable artificial intelligence”. In: *Nature Machine Intelligence* 2.10, pp. 573–584. DOI: 10.1038/s42256-020-00236-4. URL: https://doi.org/10.1038/s42256-020-00236-4.

Jiménez-Luna, José, Francesca Grisoni, and Gisbert Schneider (2020b). "Drug discovery with explainable artificial intelligence". In: *Nature Machine Intelligence* 2.10, pp. 573–584.

Johansson, Ulf et al. (2018). "Venn predictors for well-calibrated probability estimation trees". In: *Proceedings of the Seventh Workshop on Conformal and Probabilistic Prediction and Applications*. Ed. by Alex Gammerman et al. Vol. 91. Proceedings of Machine Learning Research. PMLR, pp. 3–14. URL: `https://proceedings.mlr.press/v91/johansson18a.html`.

Kailkhura, Bhavya et al. (2019). "Reliable and explainable machine-learning methods for accelerated material discovery". In: *npj Computational Materials* 5.1, p. 108. DOI: `10.1038/s41524-019-0248-2`. URL: `https://doi.org/10.1038/s41524-019-0248-2`.

Kendall, Alex and Yarin Gal (2017). "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?" In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc. URL: `https://proceedings.neurips.cc/paper_files/paper/2017/file/2650d6089a6d640c5e85b2b88265dc2b-Paper.pdf`.

Kennes, P Smets, R, and P Smets (1994). "The transferable belief model". In: *Artif Intell* 66, pp. 191–234.

Khan, Shehroz S and Michael G Madden (2014). "One-class classification: taxonomy of study and review of techniques". In: *The Knowledge Engineering Review* 29.3, pp. 345–374.

Khan, Umar et al. (Apr. 2010). "High-Concentration Solvent Exfoliation of Graphene". In: *Small* 6.7, pp. 864–871. DOI: `10.1002/smll.200902066`. URL: `https://doi.org/10.1002/smll.200902066`.

Kingma, Diederik P. and Max Welling (2019). "An Introduction to Variational Autoencoders". In: *Foundations and Trends® in Machine Learning* 12.4, pp. 307–392. ISSN: 1935-8237. DOI: `10.1561/2200000056`. URL: `http://dx.doi.org/10.1561/2200000056`.

Knight, Frank H. (1921). *Risk, Uncertainty, and Profit*. online. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.

Kobayashi, Ryo (2021). "nap: A molecular dynamics package with parameter-optimization programs for classical and machine-learning potentials". In: *J. Open Source Softw.* 6.57, p. 2768. DOI: `10.21105/joss.02768`.

Kobayashi, Ryo et al. (2017). "Neural network potential for Al-Mg-Si alloys". In: *Phys. Rev. Materials* 1 (5), p. 053604. DOI: `10.1103/PhysRevMaterials.1.053604`.

Koh, Pang Wei and Percy Liang (2017). "Understanding Black-box Predictions via Influence Functions". In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. PMLR, pp. 1885–1894. URL: `https://proceedings.mlr.press/v70/koh17a.html`.

Koinuma, Hideomi and I. Takeuchi (Aug. 2004). "Combinatorial Solid-State Chemistry of Inorganic Materials". In: *Nat. Mater.* 3, pp. 429–438. DOI: `10.1038/nmat1157`.

Konno, Tomohiko et al. (2021). "Deep learning model for finding new superconductors". In: *Phys. Rev. B* 103 (1), p. 014509. DOI: `10.1103/PhysRevB.103.014509`.

Kruppa, Jochen et al. (2014). "Probability estimation with machine learning methods for dichotomous and multicategory outcome: theory". In: *Biometrical Journal* 56.4, pp. 534–563.

Kruskal, J. B. (1964). "Nonmetric multidimensional scaling: A numerical method". In: *Psychometrika* 29.2, pp. 115–129. ISSN: 1860-0980. DOI: `10.1007/BF02289694`. URL: `https://doi.org/10.1007/BF02289694`.

Kull, Meelis and Peter Flach (Sept. 2014). “Reliability Maps: A Tool to Enhance Probability Estimates and Improve Classification Accuracy”. In: vol. 8725. ISBN: 978-3-662-44850-2. DOI: 10.1007/978-3-662-44851-9_2.

Larciprete, Rosanna et al. (Nov. 2011). “Dual Path Mechanism in the Thermal Reduction of Graphene Oxide”. In: *Journal of the American Chemical Society* 133.43, pp. 17315–17321. DOI: 10.1021/ja205168x. URL: https://doi.org/10.1021/ja205168x.

LaValley, Michael P. (2008). “Logistic Regression”. In: *Circulation* 117.18, pp. 2395–2399. DOI: 10.1161/CIRCULATIONAHA.106.682658.

Lederer, Yoav et al. (2018). “The search for high entropy alloys: A high-throughput ab-initio approach”. In: *Acta Mater.* 159, pp. 364–383. ISSN: 1359-6454. DOI: https://doi.org/10.1016/j.actamat.2018.07.042.

Letham, Benjamin et al. (Sept. 2015). “Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model”. In: *Ann. Appl. Stat.* 9.3, pp. 1350–1371. DOI: 10.1214/15-AOAS848. URL: https://doi.org/10.1214/15-AOAS848.

LEVANDOWSKY, MICHAEL and DAVID WINTER (1971). “Distance between Sets”. In: *Nature* 234.5323, pp. 34–35. DOI: 10.1038/234034a0. URL: https://doi.org/10.1038/234034a0.

Li, Ruixuan et al. (2020). “High-Throughput Calculations for High-Entropy Alloys: A Brief Review”. In: *Frontiers in Materials* 7. ISSN: 2296-8016. DOI: 10.3389/fmats.2020.00290. URL: https://www.frontiersin.org/articles/10.3389/fmats.2020.00290.

Li, Xianye et al. (Sept. 2013). “Preparation of polylactide/graphene composites from liquid-phase exfoliated graphite sheets”. In: *Polymer Composites* 35.2, pp. 396–403. DOI: 10.1002/pc.22673. URL: https://doi.org/10.1002/pc.22673.

Li, Yi-Pei et al. (2019). “Self-evolving machine: A continuously improving model for molecular thermochemistry”. In: *The Journal of Physical Chemistry A* 123.10, pp. 2142–2152.

Liang, Shiyu, Yixuan Li, and R. Srikant (2018). “Enhancing The Reliability of Out-of-distribution Image Detection in Neural Networks”. In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=H1VGkIxRZ.

Liang, Yu Teng and Mark C. Hersam (Dec. 2010). “Highly Concentrated Graphene Solutions via Polymer Enhanced Solvent Exfoliation and Iterative Solvent Exchange”. In: *Journal of the American Chemical Society* 132.50, pp. 17661–17663. DOI: 10.1021/ja107661g. URL: https://doi.org/10.1021/ja107661g.

Lin, Pin-Chun, Jhao-Yi Wu, and Wei-Ren Liu (June 2018). “Green and facile synthesis of few-layer graphene via liquid exfoliation process for Lithium-ion batteries”. In: *Scientific Reports* 8.1. DOI: 10.1038/s41598-018-27922-z. URL: https://doi.org/10.1038/s41598-018-27922-z.

Linardatos, Pantelis, Vasilis Papastefanopoulos, and Sotiris Kotsiantis (2020). “Explainable ai: A review of machine learning interpretability methods”. In: *Entropy* 23.1, p. 18.

Lindsay, Bruce G. (1995). “Mixture Models: Theory, Geometry and Applications”. In: *NSF-CBMS Regional Conference Series in Probability and Statistics* 5, pp. i–163. ISSN: 19355920, 23290978. URL: http://www.jstor.org/stable/4153184 (visited on 11/15/2022).

Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou (2008). “Isolation Forest”. In: *2008 Eighth IEEE International Conference on Data Mining*, pp. 413–422. DOI: 10.1109/ICDM.2008.17.

Liu, Xinjuan et al. (2011). “Microwave-assisted synthesis of TiO_2-reduced graphene oxide composites for the photocatalytic reduction of Cr(vi)”. In: *RSC Advances*

1.7, p. 1245. DOI: 10.1039/c1ra00298h. URL: https://doi.org/10.1039/c1ra00298h.

Lloyd, S. (1982). "Least squares quantization in PCM". In: *IEEE Transactions on Information Theory* 28.2, pp. 129–137. DOI: 10.1109/TIT.1982.1056489.

Lotya, Mustafa et al. (May 2010). "High-Concentration, Surfactant-Stabilized Graphene Dispersions". In: *ACS Nano* 4.6, pp. 3155–3162. DOI: 10.1021/nn1005304. URL: https://doi.org/10.1021/nn1005304.

Lázaro-Gredilla, Miguel, Steven Van Vaerenbergh, and Neil D. Lawrence (2012). "Overlapping Mixtures of Gaussian Processes for the data association problem". In: *Pattern Recognition* 45.4, pp. 1386–1395. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2011.10.004. URL: https://www.sciencedirect.com/science/article/pii/S0031320311004109.

Maaten, Laurens van der and Geoffrey Hinton (2008). "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research* 9.86, pp. 2579–2605. URL: http://jmlr.org/papers/v9/vandermaaten08a.html.

MacQueen, J. B. (1967). "Some Methods for Classification and Analysis of MultiVariate Observations". In: *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*. Ed. by L. M. Le Cam and J. Neyman. Vol. 1. University of California Press, pp. 281–297.

Mcfall, Joseph (Apr. 2015). "Rational, Normative, Descriptive, Prescriptive, or Choice Behavior? The Search for Integrative Metatheory of Decision Making". In: *Behavioral Development Bulletin* 20, p. 45. DOI: 10.1037/h0101039.

McInnes, Leland, John Healy, and James Melville (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. arXiv: 1802.03426 [stat.ML].

Mead, A. (1992). "Review of the Development of Multidimensional Scaling Methods". In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 41.1, pp. 27–39. URL: http://www.jstor.org/stable/2348634.

Meder, Björn, Fabrice Le Lec, and Magda Osman (2013). "Decision making in uncertain times: what can cognitive and decision sciences say about or learn from economic crises?" In: *Trends in Cognitive Sciences* 17.6, pp. 257–260. DOI: 10.1016/j.tics.2013.04.008. URL: https://doi.org/10.1016/j.tics.2013.04.008.

Mikolov, Tomas et al. (2013). *Distributed Representations of Words and Phrases and their Compositionality*. arXiv: 1310.4546 [cs.CL].

Murphy, Allan H. (1996). "The Finley Affair: A Signal Event in the History of Forecast Verification". In: *Weather and Forecasting* 11.1, pp. 3 –20. DOI: 10.1175/1520-0434(1996)011<0003:TFAASE>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/wefo/11/1/1520-0434_1996_011_0003_tfaase_2_0_co_2.xml.

Murtagh, Fionn and Pierre Legendre (2014). "Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?" In: *J. Classif.* 31.3, pp. 274–295. ISSN: 1432-1343. DOI: 10.1007/s00357-014-9161-z.

Nacken, T. J. et al. (2015). "Delamination of graphite in a high pressure homogenizer". In: *RSC Advances* 5.71, pp. 57328–57338. DOI: 10.1039/c5ra08643d. URL: https://doi.org/10.1039/c5ra08643d.

Neumann, John von, Oskar Morgenstern, and Ariel Rubinstein (1944). *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. Princeton University Press. ISBN: 9780691130613. URL: http://www.jstor.org/stable/j.ctt1r2gkx (visited on 04/20/2023).

Nguyen, Duong-Nguyen et al. (Nov. 2018a). "Committee machine that votes for similarity between materials". In: *IUCrJ* 5, pp. 830–840. DOI: 10.1107/S2052252518013519.

Nguyen, Duong-Nguyen et al. (2018b). "Committee machine that votes for similarity between materials". In: *IUCrJ* 5.6, pp. 830–840. DOI: 10.1107/S2052252518013519. URL: https://doi.org/10.1107/S2052252518013519.

Nguyen, Duong-Nguyen et al. (2019). "Ensemble learning reveals dissimilarity between rare-earth transition binary alloys with respect to the Curie temperature". In: *Journal of Physics: Materials*. URL: http://iopscience.iop.org/10.1088/2515-7639/ab1738.

Nguyen, Duong-Nguyen et al. (Sept. 2022). "Explainable active learning in investigating structure–stability of SmFe12 $-\alpha-\beta X\alpha Y\beta$ structures X, Y {Mo, Zn, Co, Cu, Ti, Al, Ga}". In: *MRS Bulletin* 48. DOI: 10.1557/s43577-022-00372-9.

Nicolosi, V. et al. (June 2013). "Liquid Exfoliation of Layered Materials". In: *Science* 340.6139, pp. 1226419–1226419. DOI: 10.1126/science.1226419. URL: https://doi.org/10.1126/science.1226419.

Nielsen, Frank (Feb. 2016). "Hierarchical Clustering". In: pp. 195–211. ISBN: 978-3-319-21902-8. DOI: 10.1007/978-3-319-21903-5_8.

Nix, D.A. and A.S. Weigend (1994). "Estimating the mean and variance of the target probability distribution". In: *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*. Vol. 1, 55–60 vol.1. DOI: 10.1109/ICNN.1994.374138.

Novoselov, K. S. (2004). "Electric Field Effect in Atomically Thin Carbon Films". In: *Science* 306.5696, pp. 666–669. DOI: 10.1126/science.1102896. URL: https://doi.org/10.1126/science.1102896.

Nu Thanh Ton, Nhan et al. (2020). "Solvent screening for efficient chemical exfoliation of graphite". In: *2D Materials* 8.1, p. 015019. DOI: 10.1088/2053-1583/abc08a. URL: http://dx.doi.org/10.1088/2053-1583/abc08a.

Nyshadham, Chandramouli et al. (2017). "A computational high-throughput search for new ternary superalloys". In: *Acta Mater.* 122, pp. 438–447. ISSN: 1359-6454. DOI: https://doi.org/10.1016/j.actamat.2016.09.017.

"Binary Alloy Phase Diagrams" (Apr. 2016). In: *Alloy Phase Diagrams*. Ed. by H. Okamoto, M.E. Schlesinger, and E.M. Mueller. Vol. 3. ASM International. ISBN: 978-1-62708-163-4. DOI: 10.31399/asm.hb.v03.a0006247. URL: https://doi.org/10.31399/asm.hb.v03.a0006247.

Papernot, Nicolas and Patrick McDaniel (2018). *Deep k-Nearest Neighbors: Towards Confident, Interpretable and Robust Deep Learning*. arXiv: 1803.04765 [cs.LG].

Patole, Archana S. et al. (Oct. 2010). "A facile approach to the fabrication of graphene/polystyrene nanocomposite by in situ microemulsion polymerization". In: *Journal of Colloid and Interface Science* 350.2, pp. 530–537. DOI: 10.1016/j.jcis.2010.01.035. URL: https://doi.org/10.1016/j.jcis.2010.01.035.

Pennington, Jeffrey, Richard Socher, and Christopher Manning (Oct. 2014). "GloVe: Global Vectors for Word Representation". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, pp. 1532–1543. DOI: 10.3115/v1/D14-1162. URL: https://aclanthology.org/D14-1162.

Perello Nieto, Miquel et al. (Dec. 2016). "Background Check: A General Technique to Build More Reliable and Versatile Classifiers". In: pp. 1143–1148. DOI: 10.1109/ICDM.2016.0150.

Pham, Tien Lam et al. (2016a). "Novel mixture model for the representation of potential energy surfaces". In: *J. Chem. Phys.* 145.15, p. 154103. DOI: 10.1063/1.4964318.

— (2016b). "Novel mixture model for the representation of potential energy surfaces". In: *The Journal of Chemical Physics* 145.15, p. 154103. DOI: 10.1063/1.4964318.

Pham, Tien Lam et al. (2017). “Machine learning reveals orbital interaction in materials”. In: *Sci. Technol. Adv. Mater.* 18.1, pp. 756–765. DOI: 10.1080/14686996.2017.1378060.

Platt, John et al. (1999). “Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods”. In: *Advances in large margin classifiers* 10.3, pp. 61–74.

Potyrailo, Radislav et al. (Nov. 2011). “Combinatorial and High-Throughput Screening of Materials Libraries: Review of State of the Art”. In: *ACS Combinatorial Science* 13.6, pp. 579–633. DOI: 10.1021/co200007w. URL: https://doi.org/10.1021/co200007w.

Qi, Xin et al. (Mar. 2017). “Highly Efficient High-Pressure Homogenization Approach for Scalable Production of High-Quality Graphene Sheets and Sandwich-Structured - Fe_2O_3/Graphene Hybrids for High-Performance Lithium-Ion Batteries”. In: *ACS Applied Materials Interfaces* 9.12, pp. 11025–11034. DOI: 10.1021/acsami.7b00808. URL: https://doi.org/10.1021/acsami.7b00808.

Qian, Wen et al. (Sept. 2009). “Solvothermal-assisted exfoliation process to produce graphene with high yield and high quality”. In: *Nano Research* 2, pp. 706–712. DOI: 10.1007/s12274-009-9074-z.

Quinlan, J. R. (Mar. 1986). “Induction of Decision Trees”. In: *Mach. Learn.* 1.1, 81–106. ISSN: 0885-6125. DOI: 10.1023/A:1022643204877.

Ramprasad, Rampi et al. (2017). “Machine learning in materials informatics: recent applications and prospects”. In: *npj Computational Materials* 3.1, p. 54. DOI: 10.1038/s41524-017-0056-5. URL: https://doi.org/10.1038/s41524-017-0056-5.

Rao, Jiahua et al. (2022). “Quantitative evaluation of explainable graph neural networks for molecular property prediction”. In: *Patterns* 3.12, p. 100628.

Rickman, J. M. et al. (2019). “Materials informatics for the screening of multi-principal elements and high-entropy alloys”. In: *Nat. Commun.* 10, p. 2618. DOI: 10.1038/s41467-019-10533-1.

Rodríguez-Pérez, Raquel and Jurgen Bajorath (2021). “Explainable Machine Learning for Property Predictions in Compound Optimization: Miniperspective”. In: *Journal of medicinal chemistry* 64.24, pp. 17744–17752.

Rudin, Cynthia (2019). “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead”. In: *Nature Machine Intelligence* 1.5, pp. 206–215. URL: https://doi.org/10.1038/s42256-019-0048-x.

Saal, James E. et al. (2013). “Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD)”. In: *JOM* 65.11, pp. 1501–1509. DOI: 10.1007/s11837-013-0755-4. URL: https://doi.org/10.1007/s11837-013-0755-4.

Sahoo, Sumanta et al. (2013). “One Pot Synthesis of Graphene by Exfoliation of Graphite in ODCB”. In: *Graphene* 02.01, pp. 42–48. DOI: 10.4236/graphene.2013.21006. URL: https://doi.org/10.4236/graphene.2013.21006.

Seko, Atsuto, Atsushi Togo, and Isao Tanaka (2018). “Descriptors for Machine Learning of Materials Data”. In: *Nanoinformatics*. Ed. by Isao Tanaka. Singapore: Springer Singapore, pp. 3–23. ISBN: 978-981-10-7617-6. DOI: 10.1007/978-981-10-7617-6_1. URL: https://doi.org/10.1007/978-981-10-7617-6_1.

Seko, Atsuto et al. (2017). “Representation of compounds for machine-learning prediction of physical properties”. In: *Phys. Rev. B* 95 (14), p. 144110. DOI: 10.1103/PhysRevB.95.144110.

Seko, Atsuto et al. (2018). "Matrix- and tensor-based recommender systems for the discovery of currently unknown inorganic compounds". In: *Phys. Rev. Mater.* 2, p. 013805.

Selten, Reinhard, Sabine Pittnauer, and Martin Hohnisch (2012). "Dealing with dynamic decision problems when knowledge of the environment is limited: an approach based on goal systems". In: *Journal of Behavioral Decision Making* 25.5, pp. 443–457.

Senge, Robin et al. (2014). "Reliable classification: Learning classifiers that distinguish aleatoric and epistemic uncertainty". In: *Information Sciences* 255, pp. 16–29.

Senkov, O. N. et al. (2015). "Accelerated exploration of multi-principal element alloys with solid solution phases". In: *Nat. Commun.* 6.1, p. 6529. DOI: `10.1038/ncomms7529`.

Shafer, Glenn (1976). *A Mathematical Theory of Evidence*. Princeton University Press. DOI: `10.1515/9780691214696`.

Shen, Jianfeng et al. (July 2015). "Liquid Phase Exfoliation of Two-Dimensional Materials by Directly Probing and Matching Surface Tension Components". In: *Nano Letters* 15.8, pp. 5449–5454. DOI: `10.1021/acs.nanolett.5b01842`. URL: `https://doi.org/10.1021/acs.nanolett.5b01842`.

Shen, Zhigang et al. (Aug. 2011). "Preparation of graphene by jet cavitation". In: *Nanotechnology* 22.36, p. 365306. DOI: `10.1088/0957-4484/22/36/365306`. URL: `https://doi.org/10.1088/0957-4484/22/36/365306`.

Silverman, B.W. (1998). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, pp. 1–175. ISBN: 9781315140919. DOI: `10.1201/9781315140919`.

Silveyra, Josefina M. et al. (2018). "Soft magnetic materials for a sustainable and electrified world". In: *Science* 362.6413. ISSN: 0036-8075. DOI: `10.1126/science.aao0195`.

Simon, Herbert A (1990). "Invariants of human behavior". In: *Annual review of psychology* 41.1, pp. 1–20.

Smets, Philippe (1993). "Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem". In: *International Journal of Approximate Reasoning* 9.1, pp. 1–35. ISSN: 0888-613X. DOI: `https://doi.org/10.1016/0888-613X(93)90005-X`. URL: `https://www.sciencedirect.com/science/article/pii/0888613X9390005X`.

Snow, R. J. et al. (2018). "Large moments in bcc $Fe_xCo_yMn_z$ ternary alloy thin films". In: *Appl. Phys. Lett.* 112.7, p. 072403. DOI: `10.1063/1.5006347`.

Soleimany, Ava P et al. (2021). "Evidential deep learning for guided molecular property prediction and discovery". In: *ACS central science* 7.8, pp. 1356–1367.

Tamura, Tomoyuki et al. (2017). "Fast and scalable prediction of local energy at grain boundaries: machine-learning based modeling of first-principles calculations". In: *Model. Simul. Mater. Sci. Eng.* 25.7, p. 075003. DOI: `10.1088/1361-651x/aa8276`.

Tax, David MJ and Robert PW Duin (2004). "Support vector data description". In: *Machine learning* 54, pp. 45–66.

Tenenbaum, Joshua, Vin Silva, and John Langford (Jan. 2000). "A Global Geometric Framework for Nonlinear Dimensionality Reduction". In: *Science* 290, pp. 2319–2323. DOI: `10.1126/science.290.5500.2319`.

Tenenbaum, Joshua B. (1996). "Learning the Structure of Similarity". In: *Advances in Neural Information Processing Systems 8*. Ed. by D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, pp. 3–9. URL: `http://papers.nips.cc/paper/1052-learning-the-structure-of-similarity.pdf`.

Teye, Mattias, Hossein Azizpour, and Kevin Smith (2018). "Bayesian Uncertainty Estimation for Batch Normalized Deep Networks". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, pp. 4907–4916. URL: `https://proceedings.mlr.press/v80/teye18a.html`.

Thaler, Richard and C. Sunstein (June 2009). *NUDGE: Improving Decisions About Health, Wealth, and Happiness*. Vol. 47. ISBN: 9780141040011.

Toh, Shaw Yong et al. (2014). "Graphene production via electrochemical reduction of graphene oxide: Synthesis and characterisation". In: *Chemical Engineering Journal* 251, pp. 422 –434. ISSN: 1385-8947. DOI: `https://doi.org/10.1016/j.cej.2014.04.004`. URL: `http://www.sciencedirect.com/science/article/pii/S138589471400432X`.

Ton, Nhan Nu Thanh et al. (July 2018). "One-pot synthesis of TiO_2/graphene nanocomposites for excellent visible light photocatalysis based on chemical exfoliation method". In: *Carbon* 133, pp. 109–117. DOI: `10.1016/j.carbon.2018.03.025`. URL: `https://doi.org/10.1016/j.carbon.2018.03.025`.

Tsai, Ming-Hung (2016). "Three Strategies for the Design of Advanced High-Entropy Alloys". In: *Entropy* 18.7, p. 252. ISSN: 1099-4300. DOI: `10.3390/e18070252`.

Tsai, Ming-Hung and Jien-Wei Yeh (2014). "High-Entropy Alloys: A Critical Review". In: *Mater. Res. Lett.* 2, pp. 107–123. DOI: `10.1080/21663831.2014.912690`.

Tsai, Ming-Hung et al. (2019). "Intermetallic Phases in High-Entropy Alloys: Statistical Analysis of their Prevalence and Structural Inheritance". In: *Metals* 9.2, p. 247. ISSN: 2075-4701. DOI: `10.3390/met9020247`.

Tshitoyan, Vahe et al. (2019). "Unsupervised word embeddings capture latent knowledge from materials science literature". In: *Nature* 571.7763, pp. 95–98. DOI: `10.1038/s41586-019-1335-8`. URL: `https://doi.org/10.1038/s41586-019-1335-8`.

Tversky, Amos (1977). "Features of similarity." In: *Psychol. Rev.* 84.4, pp. 327–352. DOI: `10.1037/0033-295X.84.4.327`.

Tversky, Amos and Daniel Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases". In: *Science* 185.4157, pp. 1124–1131. DOI: `10.1126/science.185.4157.1124`. eprint: `https://www.science.org/doi/pdf/10.1126/science.185.4157.1124`. URL: `https://www.science.org/doi/abs/10.1126/science.185.4157.1124`.

Varshney, Kush R. and Homa Alemzadeh (2017). *On the Safety of Machine Learning: Cyber-Physical Systems, Decision Sciences, and Data Products*. arXiv: `1610.01256 [cs.CY]`.

Villars, P. et al. (2004). "The Pauling File, Binaries Edition". In: *Journal of Alloys and Compounds* 367. DOI: `10.1016/j.jallcom.2003.08.058`.

Vovk, Vladimir (2013). "Kernel ridge regression". In: *Empirical inference*. Springer, pp. 105–116. DOI: `10.1007/978-3-642-41136-6_11`.

Wajid, Ahmed S. et al. (Feb. 2012). "Polymer-stabilized graphene dispersions at high concentrations in organic solvents for composite production". In: *Carbon* 50.2, pp. 526–534. DOI: `10.1016/j.carbon.2011.09.008`. URL: `https://doi.org/10.1016/j.carbon.2011.09.008`.

Walley, Peter (1991). *Statistical reasoning with imprecise probabilities*. Vol. 42. Springer.

Wang, Jingshi, Zhigang Shen, and Min Yi (Nov. 2019). "Liquid-exfoliated graphene as highly efficient conductive additives for cathodes in lithium ion batteries". In: *Carbon* 153, pp. 156–163. DOI: `10.1016/j.carbon.2019.07.008`. URL: `https://doi.org/10.1016/j.carbon.2019.07.008`.

Ward, Logan et al. (2016). "A general-purpose machine learning framework for predicting properties of inorganic materials". In: *npj Computational Materials* 2.1, p. 16028. DOI: 10.1038/npjcompumats.2016.28. URL: https://doi.org/10.1038/npjcompumats.2016.28.

White, Ashley (2012). "The Materials Genome Initiative: One year on". In: *MRS Bulletin* 37.8, pp. 715–716. DOI: 10.1557/mrs.2012.194. URL: https://doi.org/10.1557/mrs.2012.194.

Williams, CKI. and CE. Rasmussen (June 1996). "Gaussian Processes for Regression". In: *Advances in neural information processing systems 8*. Max-Planck-Gesellschaft. Cambridge, MA, USA: MIT Press, pp. 514–520. URL: https://proceedings.neurips.cc/paper/1995/file/7cce53cf90577442771720a370c3c723-Paper.pdf.

Wu, Z. et al. (2014). "Recovery, recrystallization, grain growth and phase stability of a family of FCC-structured multi-component equiatomic solid solution alloys". In: *Intermetallics* 46, pp. 131–140. ISSN: 0966-9795. DOI: https://doi.org/10.1016/j.intermet.2013.10.024.

Xu, Tongguang et al. (Jan. 2011). "Significantly enhanced photocatalytic performance of ZnO via graphene hybridization and the mechanism study". In: *Applied Catalysis B: Environmental* 101.3-4, pp. 382–387. DOI: 10.1016/j.apcatb.2010.10.007. URL: https://doi.org/10.1016/j.apcatb.2010.10.007.

Xu, Yibin, Masayoshi Yamazaki, and Pierre Villars (2011). "Inorganic Materials Database for Exploring the Nature of Material". In: *Jpn. J. Appl. Phys.* 50.11RH02. DOI: 10.1143/JJAP.50.11RH02.

Yager, Ronald R (2006). "An extension of the naive Bayesian classifier". In: *Inf. Sci.* 176.5, pp. 577–588.

Yang, Chu-I and Yi-Pei Li (2023). "Explainable uncertainty quantifications for deep learning-based molecular property prediction". In: *Journal of Cheminformatics* 15.1, p. 13. DOI: 10.1186/s13321-023-00682-3. URL: https://doi.org/10.1186/s13321-023-00682-3.

Yang, Yingzhen et al. (2014). "On a Theory of Nonparametric Pairwise Similarity for Clustering: Connecting Clustering to Classification". In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani et al., pp. 145–153. URL: http://papers.nips.cc/paper/5304-on-a-theory-of-nonparametric-pairwise-similarity-for-clustering-connecting-clustering-to-classification.pdf.

Ye, Y.F. et al. (2015). "Design of high entropy alloys: A single-parameter thermodynamic rule". In: *Scr. Mater.* 104, pp. 53–55. ISSN: 1359-6462. DOI: https://doi.org/10.1016/j.scriptamat.2015.03.023.

Yeh, J.-W. et al. (2004). "Nanostructured High-Entropy Alloys with Multiple Principal Elements: Novel Alloy Design Concepts and Outcomes". In: *Adv. Eng. Mater.* 6, pp. 299–303. DOI: 10.1002/adem.200300567.

Yi, Min et al. (Sept. 2011). "Morphology and structure of mono- and few-layer graphene produced by jet cavitation". In: *Applied Physics Letters* 99.12, p. 123112. DOI: 10.1063/1.3641863. URL: https://doi.org/10.1063/1.3641863.

Zadeh, Lotfi A. (1984). "Review of A Mathematical Theory of Evidence". In: *AI Magazine* 5.3, p. 81. DOI: 10.1609/aimag.v5i3.452. URL: https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/452.

Zadrozny, Bianca and Charles Elkan (May 2001). "Obtaining Calibrated Probability Estimates from Decision Trees and Naive Bayesian Classifiers". In: *ICML* 1.

— (Aug. 2002). “Transforming Classifier Scores into Accurate Multiclass Probability Estimates”. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. DOI: `10.1145/775047.775151`.

Zhang, F. et al. (2014). “An understanding of high entropy alloys from phase diagram calculations”. In: *CALPHAD* 45, pp. 1–10. ISSN: 0364-5916. DOI: `https://doi.org/10.1016/j.calphad.2013.10.006`.

Zhang, Ruqi et al. (2020). *Cyclical Stochastic Gradient MCMC for Bayesian Deep Learning*. arXiv: `1902.03932 [cs.LG]`.

Zhang, Y. et al. (2008). “Solid-Solution Phase Formation Rules for Multi-component Alloys”. In: *Adv. Eng. Mater.* 10.6, pp. 534–538. DOI: `10.1002/adem.200700240`.

Zhang, Yong et al. (2016). “Phase Formation Rules”. In: *High-Entropy Alloys: Fundamentals and Applications*. Cham: Springer International Publishing, pp. 21–49. DOI: `10.1007/978-3-319-27013-5_2`. URL: `https://doi.org/10.1007/978-3-319-27013-5_2`.

Zhu, Yanwu et al. (June 2010). “Graphene and Graphene Oxide: Synthesis, Properties, and Applications”. In: *Advanced Materials* 22.35, pp. 3906–3924. DOI: `10.1002/adma.201001068`. URL: `https://doi.org/10.1002/adma.201001068`.