Title	A Study on Ensemble Method for Multi-Relational Link Prediction for Commonsense Knowledge
Author(s)	Khine, Myat Thwe
Citation	
Issue Date	2022-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17630
Rights	
Description	Supervisor:Nguyen Minh Le, 先端科学技術研究科, 修士（情報科学）

Master's Thesis

A Study on Ensemble Method for Multi-Relational Link Prediction for Commonsense Knowledge

Khine Myat Thwe

Supervisor: Nguyen Le Minh

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(Master Degreee of Information Science)

March, 2022

Abstract

Commonsense knowledge is crucial in today's artificial intelligence research area. It contains truths and information of our daily situations, for example "Sugar is sweet" or "Breathing makes a person alive", that most of the humans are expected to own by the time they grow up. It is an unsolved artificial intelligence problem till this era. Many working commonsense knowledge programs are needed to assist decision making in AI expert systems. Thus, commonsense knowledge becomes an essential joint for such AI expert systems. The importance of commonsense reasoning was recognized since many of these systems were able to reason, but they were also vulnerable because they frequently offered nonsensical responses when faced with unexpected problematic data.

This becomes our core motivation to dive into this area of research. Speaking of diving into this field, there are many approaches to make machines gain commonsense knowledge by means of natural language processing, computer vision, etc. However, from all the efforts made over years, it is obvious that teaching machines to have commonsense knowledge is a time-consuming and expensive process. Two main reasons why natural language processing is chosen to proceed for this thesis are:

1. Natural language processing can be treated as a base brick for further applications such as audio recognition or computer vision applications that need commonsense knowledge,

2. It is less costly compared to other approaches.

In recent years, knowledge graph embedding algorithms become popular for knowledge base completion tasks. Those prior studies enlightened in a way that if we can use knowledge graph models for knowledge base completion tasks, it can also be helpful in commonsense knowledge mining task along with certain enhancements. Each embedding model has its unique performance when they are dealt with different parts of a dataset. Among various well-known embedding methods, state-of-the-art models [32] were chosen to use in our experiments.

From the aforementioned points, we conducted our thesis with following main contributions:

1. We presented the proposal of using the CKBC dataset to testify the ability of Knowledge Graph Embedding Models,

2. We reproduced the results for 24 knowledge graph embedding models which are built-in models from PyKeen,
3. We found out the different results on each model, select the best ones and implemented ensemble from the results from step 2,
4. Our ensemble method shows that combining the results can produce better performace.

Keywords: Natural Language Processing, Knowledge Graph Embedding Models.

Acknowledgement

First and foremost, I would like to express my over indebtedness and intense sense of gratitude to my grateful supervisor and advisor Professor Nguyen Le Minh, Graduate School of Advanced Science and Technology,Japan Advanced Institute of Science and Technology, for considering and accepting me as his candidate for this worthy Master's Degree in JAIST. I am truly grateful for his encouragement and exhaustive suggestion. During my term from beginning of my thesis work, he directed my attention towards the trends of my research area by giving me intellectual freedom in my work. He gave his all-out of wide knowledge to me and engaged new ideas to utilize in my thesis. I am really indebted to my supervisor Professor Nguyen Le Minh for introducing me to a prosperous and very interesting field of research in natural language processing.

I would also like to acknowledge my respectful thesis advisor Assistant Professor Teeradaj Racharak , for his understanding and kindness to me. The entryway to X sensei's work table was always welcoming whenever I encountered a trouble blot or had some problematic questions about my research or analysis. Without him, this thesis wouldn't have been presented as it is. As an admiring advisor, he is a lodestar for me, and he guided me in practices and showed me directions I will use in my lasting the whole length of the life.

Additionally, I would like to thank to my admirable Professor Kazuhiro Ogata for guiding me to do my minor research. It gave me a lot of experiences and motivation to proceed to finish my master thesis.

I wish to express my arrears to all the present and the former members in my Laboratory for their welcoming, warmness and hospitality. They were always there to help me whenever I am in need. I would like to express my immense sense of gratitude to the member of thesis committee, useful comments and suggestions in my work.

I would like profoundly to express my indebtedness to Daw Saw Sanda Aye, UIT(Myanmar)who greatly support me in various aspects to receive MEXT scholarship and to have a pleasurable student life in Japan.

Finally, I would like to express my special thanks to my parents for their constant love and support. I am beholden to them for their genuine backing and continuous encouragement throughout the period of my study. I am tremendously thankful to their compassionate benevolence and certainty eager attitude to help me to grasp my goals. I hope I will do them to be the proud parents of my achievements and I will delight to my proud family.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Commonsense knowledge is the fundamental level of practical information and judgment that we all require to live in a reasonable and safe manner in our human environment. As people grow older, we become to have more commonsense knowledge by learning around our atmosphere specifically. For example, assume you and your friend are playing basketball and you ask him to give you the ball. Then, there will nearly be a zero chance that he is going to give a soccer ball or a hydrogen balloon to you. That is commonsense. You will have low chances of having errors or mistakes while making an instant ramen, using your vacuum cleaner or riding a bicycle because you already know what to do, what not to do and how to do these actions. Such knowledge is not taught to you explicitly. It is self evident for you. However, in order for robots to complete the same activity, the request "make an instant ramen" will not provide sufficient information to identify task components such as heating water, flavoring the noodle, and so on. The same previous information that you would utilize in the same circumstance must be required by a robot.

The target of artificial intelligence is to construct machines that can mimic people behavior and decision-making. As a result, researchers are working for the creation of computers that can handle problems and achieve objectives on a human level. That is the core reason why machines are needed to be provided commonsense knowledge that we naturally own. To make this goal achieve, AI researchers have put in a lot of work to retrieve commonsense knowledge that they have stored in humans' heads and build it into knowledge base (KBs). However, commonsense knowledge is implicit and it depends on the culture or context, the researchers have faced many challenges and

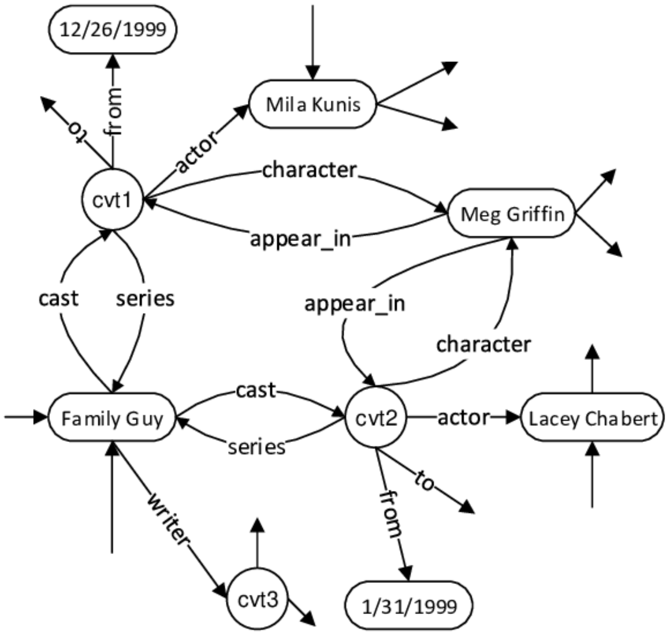

Figure 1.1: Snapshot of Freebase subgraph

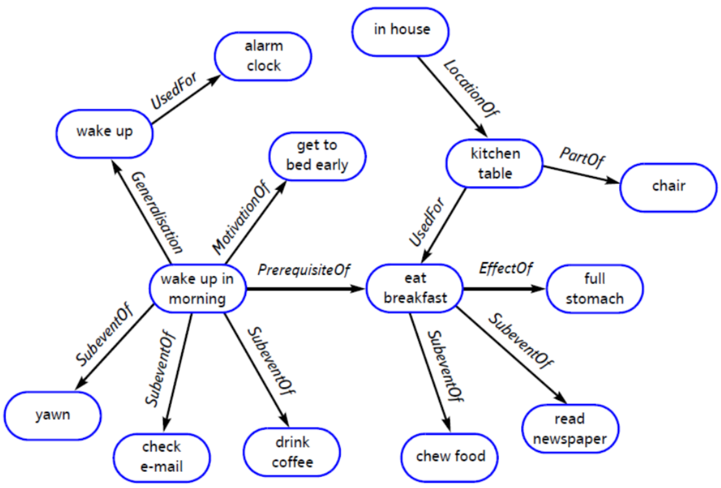

Figure 1.2: Snapshot of ConceptNet semantic network [22]

difficulties.

Cyc [22], SUMO ontology [29], HowNet [15], and Open Mind Common Sense (OMCS) [34] are the early stage milestones for commonsense knowledge development and reasoning. AI researchers had to focus on manual annotation by system experts in that era with the help of manual knowledge gathering by using pubilc platforms[40]. Although the quality of the manual work might be good or acceptable, such manual efforts showed a lot of drawbacks whatsoever such as limited size or diverse nature of the knowledge being collected. To overcome the drawbacks of manual assertions, large scale commonsense knowledge mining were introduced.

Previously, approaches depended on logical models that matched mathematical models to previously acquired knowledge[17]. However, the mathematical complexity of logical reasoning may not fit to present knowledge base (KBs) [9]. Then, knowledge graph completion (KBC) task becomes popular in recent years.With the help of knowledge graph representation, knowledge graph completion is the act of inferring new edges, called facts, in a knowledge graph based on the already existing relational data. There are various knowledge graph embedding (KGE) models that convert the entities and relations of that knowledge graph into vector representations. These approaches excel at adding missing facts to encyclopaedic knowledge bases like DBpedia [21] and Freebase [3]. However, when KGE models are deployed to the commonsense knowledge base completion task (CKBC), a curated repository of commonsense knowledge, similar performance is not found when the problem is formulated as one of knowledge base completion (KBC).

In this thesis, we investigate and study diverse KGE models with CKBC dataset and developed ensemble methods on knowledge graph embedding models aiming to obtain the results on CKBC dataset. Our ultimate goal is to explore CKBC through KGE models. As a result, the improved knowledge graph embedding models are specifically designed to promote commonsense reasoning.

1.2 Contributions

1. Reproducing results for KGE models on CKBC

We propose reproducing the results of the knowledge graph embedding models (KGEs) for CKBC. Then, we propose to make enhancements with the help of hyper parameters. This draws on the idea that each model performs differently on the CKBC dataset.

2. Ensemble KGE Models for CKBC

Here, we propose an ensemble model to combine the best results from our selection knowledge graph embedding models (KGEs) to prove that combination of the models can perform better.

1.3 Issues

Commonsense knowledge mining is a challenging process with distinct obstacles arising from the knowledge's features. As a result, we'll take a look at a few of challenging issues.

1. Implicitness: Humans perceive commonsense knowledge in a way of suppositions about daily situations which should be understood by all. As a consequence, they commonly cease bothering and ignore it in communications. As a result, it is hard to consider and convey which they accept for granted, and commonsense knowledge mining methods that depend on obtaining superficial linguistic resources would struggle to come up with this implicitness. More complicated algorithms capable of reasoning and making inferences are required to do this.

2. Diversity: Commonsense knowledge comprises a large variety of human understanding and touches all element of our everyday lives. It may be characterized as type and domain agnostic. It is difficult to identify all of the concepts, words, or relationships that are involved.

3. Automation: Codifying the generality and universal reach of commonsense knowledge is a massive effort that humans are incapable of doing. As a consequence, manual to automatic and semi-automated operations had to be transitioned. The reasoning approach, in particular, aims to infer new knowledge automatically using similarities and parallels based upon what is understood.

4.Efficiency: One might imagine that as computational performance rises, so will the rate of commonsense knowledge quarrying, however this is not the case. As the size of existing knowledge for reasoning operations expanded, the efficiency of constructing projected lacking commonsense assertions would rise.

1.4 Scope of Research

The goal of this study is to see how useful KGE models are for CKBC when it comes to anticipating missing links between existent entities. To do this, we re-implemented over 20 KGE models using the CKBC dataset. We suggest that the commonsense knowledge mining issue be treated as a commonsense knowledge base completion task. We introduce
(1) reproducing KGE models for CKBC , and
(2) incorporating KGE models together to get better results. Our study used mainly CKBC dataset and a lot of KGE models. For the conduction of our experiments, we mainly use PyKeen [1] and LibKGE [5] for the framework of KGE models that we are going to use in our research.

1.5 Thesis Outline

The following is a breakdown of the report's structure: Chapter 2 locates this research in the context of previous studies. It begins with defining commonsense knowledge and then goes over several KGE models. Following that, it goes through several commonsense knowledge mining strategies. The suggested models are presented in Chapter 3, while experimental settings are described in Chapter 4. We analyze our technique and discuss the findings in Chapter 5. In Chapter 6, we wrap things up by reviewing what we've learned and provide recommendations for future research.

Chapter 2

Related Work

2.1 Commonsense Knowledge

Because there is no formal statement of commonsense knowledge, it might be roughly characterized as a huge stack of agreed-upon facts learned as a person grows up via ordinary occurrences. It encompasses a number of specific knowledge, such as object activities, characteristics, event place and time, human motivations and moods, and so on. It refers to implicit information that is commonly disseminated and well recognized to the point that, regardless of the fact that it is essential to carry out everyday responsibilities, it is commonly left out of discourse. According to [47], commonsense knowledge has the following characteristics: it is shared by practically everyone, it is fundamental and widely understood, it is implicit, it is massive in terms of both volume and quality, open-domain, and contains basic assertions that are accessible to deviations in appropriate settings.

In contrary to true knowledge, commonsense is an ontological knowledge that is focused with the connections and qualities of vague concepts and categories rather than real entities or instances of these classes. Commonsense knowledge includes ideas and link hierarchy, which are facilitators for commonsense reasoning and inference. ConceptNet is a public commonsense knowledge base and natural language processing tool set designed to help computers understand the definitions of words humans use. Linguistic networks, which are a sort of knowledge representation, are used to do this.These use graph ways to demonstrate linkages between ideas and events to demonstrate common sense tasks. This inferential data knowledge-base contains about 1.6 million asserts of common sense information, such as the spatial, physiological, cultural, and mental aspects of daily life. It originated as a system named Open Mind Common Sense [34]. It's been utilized in Chatbots and some natural language assistance with great success.

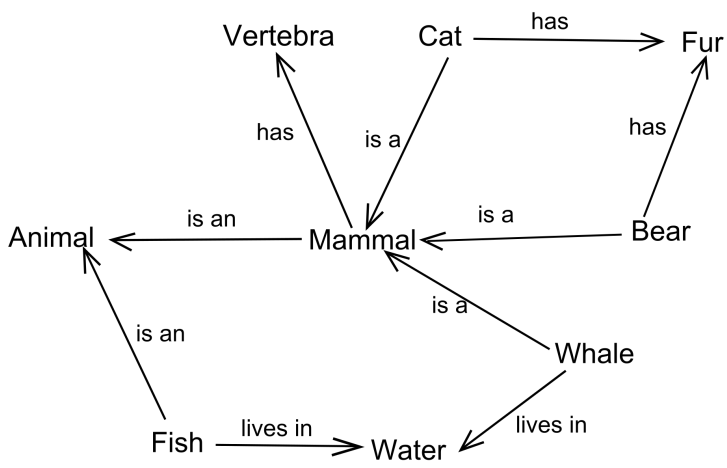

Figure 2.1: Example of Semantic network (Source: Wikipedia)

2.2 Commonsense Knowledge Bases

A knowledge base is a bundle of statements expressed as triples of the type (head , relation, tail), signifying the presence of a marked link between two terms. In commonsense knowledge bases [14], rather than actual examples of these notions, phrases relate to vague concepts (formalisms). In the past 33 years, a lot of commonsense Knowledge sources have been built.Among the most well-known are Cyc [22], [25] and ConceptNet [35] and WebChild [38].

2.2.1 Cyc

Cyc, which started in the nineties and spanned two decades, is for building entire commonsense knowledge bases. It was first manually developed by a team of dedicated system professionals (CycL) in a formal logic mathematics syntax language. Cyc's commonsense knowledge is made up of observations, rules of wisdom, and strategies for reasoning about everyday items and events. By necessity, CycL assertions are true only in certain contexts. As a consequence, Cyc's assertions have been broken down into 20,000 microtheories sharing of common assumptions. 7,000,000 claims, 500,000 words, 17,000 relations, and 17,000 connections make up Cyc. In supplement to the knowledge base, Cyc has a number of reasoning engines for reason on its knowledge [18].

2.2.2 HowNet

[15]HowNet is a multilingual commonsense knowledge base that identifies linkages among concepts or attributes of concepts. Around 1.9 million records on HowNet are expressed using Knowledge Database Markup Language (KDML). It expresses its concepts using Chinese and English phrases and idioms. These concepts are based on the foundation of sememes, the tiniest units of meaning.

2.2.3 ThoughtTreasure

ThoughtTreasure (Mueller, 1998) [27] is a commonsense knowledge-based natural language comprehension framework. The ideas in ThoughtTreasure are grouped into a top ontology and a number of domain-specific bottom ontologies. Furthermore, each concept is associated with one or more lexical components (words and phrases). ThoughtTreasure comprises 27,000 ideas with 51,000 assertions connecting them.

2.2.4 WordNet

[6]WordNet is a created lexicon of English vocabulary that includes nouns, adjectives, verbs, and adverbs and is intended for lexical classification and language modeling identification. WordNet differentiates among several meanings of a word, each one is a different interpretation that a word might have, and organizes words with the same notion into'synsets,' which are clusters of cognitive synonyms. A score is also assigned to each synset, indicating how commonly it occurs in textual. WordNet also keeps a record of the occurrence of interactions between synsets or individual words, as well as providing short description and word samples. WordNet 3.1 has 155,327 words arranged into 175,979 synsets for a total of 207,016 word-sense pairs. WordNet's semantic relations are either linguistic or commonsense, and they are between synsets rather than words. [29].

2.2.5 Open Mind Common Sense

The Common Sense Computing Initiative launched the Open Mind Common Sense (OMCS) [34]initiative with the purpose of directly collecting commonsense information on a wide level. It depended on contributors from the general populace to gather commonsense information in the form of natural language assertions, which were then analyzed to create claims. OMCS has amassed over a million bits of common sense knowledge in English from over

Reference	Year	Source	Concepts	Relations	Assertions
Cyc	1984	Curated	500,000	17,000	7,000,000
ThoughtTreasure	1994	Curated	27,000	N.A	51,000
WordNet	1995	Curated	155,327	$\approx$ 10	207,016
ConceptNet5.5	2016	Semi Automated	1,803,873	38	28,000,000

Table 2.1: Commonsense Knowledge Base Stats

15,000 contributors since its inception , with modifications to numerous other language families.

2.2.6 ConceptNet

ConceptNet (Lenat, 1995) is a vast semi-automated and heterogeneous commonsense knowledge resource obtained primarily from OMCS as well as other alternative data and modeled after the WordNet semantic network. Its vertices are natural language terms, and its associations are based on the linguistic relations paradigm of WordNet. It has been improved and published in different variants, beginning with ConceptNet 2 and finishing with ConceptNet 5.5. The most current version of ConceptNet,[35], was built from 7 formal and informal knowledge - based resources . It comprises about 8 million nodes and 21 million edges from a multilingual lexicon, ultimately connected by 38 connections. In English, there are 1,803,873 ideas and around 28 million statements.

2.3 Knowledge Graph

The term "knowledge graph" is becoming widely used in current history to describe to graph-based expert systems, and the phrases "knowledge base" and "knowledge graph" are frequently interchanged. It became famous after Google's Knowledge Graph copied it. It has been characterized in a fuzzy sense ever since due to a complete lack of consensus on an exact meaning. Before presenting their own definition, Ehrlinger and Woß [16] endeavored to assemble a list of existing definitions used in the literature. Paulheim [31] proposes a fascinating term to separate knowledge graphs from just graph-formatted data collection:

A knowledge graph
(i) characterizes regular items and associated interrelationships

in a graph,
(ii) specifies alternative categories and relationships of things in a framework,
(iii) provides for possibly interlinking unconstrained entities with one another, and
(iv) encompasses a broad range of thematic topics.
A knowledge graph is a multi-relational graph that has nodes for entities and typed-edges for entities. Every edge represents a formal truth (head entity, predicate, tail entity).

A knowledge graph (KG) is a directed heterogeneous multigraph with domain-specific node and relation types. KGs permit users to encapsulate information in a human-readable format that can be analyzed and inferred automatically. KGs are a common way of representing many sorts of collected in the form of various types of objects linked by various kinds of relationships. [20] When engaging with knowledge graphs, we employ a different vocabulary than the standard vertices and edges used in graphs. Entities and triplets are the vertices and directed edges of the knowledge graph, respectively, and are represented as a (h, r,t) tuple, where h is the head entity, t is the tail entity, and r is the relation between the head and tail entities. The term "relationship" refers to the sort of relationship in question (e.g., isA, ComprisesOf, and LocatedIn). A knowledge graph is self-descriptive in that it gives a single location to locate facts and comprehend its meaning. The term semantics is connected with the knowledge graph since the meaning of the data is encoded alongside the data in the network itself. Knowledge graphs provide value by allowing us to:

- Context: Knowledge graphs provide algorithms context by integrating multiple forms of data into an ontology and allowing for the addition of additional derived knowledge on the fly. Most classic knowledge graphs may employ many forms of raw data at the same time.

- Efficiency: Knowledge graphs enable computational efficiency for querying stored data, resulting in effective data utilization for generating insights, once necessary entities and relations are accessible.

- Explainability: By integrating the meaning of entities available inside the graph itself, large networks of entities and interactions give answers to the challenge of understandability. As a result, knowledge graphs are inherently explainable.

2.4 Knowledge Graph Embeddings

It is described as the challenge of discovering vector space representations for knowledge base entities and relations. Assume we have head and tail entities, as well as relationships between them. A scoring function that may be specified by the connection linking the two entities, a function f can be used to calculate the likelihood of a relation that can connect a certain head and tail entity (h, r, t). To put it another way, the core concept behind these models is that relationships between things may be depicted as interactions between their vector representations, and that these interactions can take a variety of forms.

These representations can help with a variety of useful applications. In past few years, several relation modeling approaches have been presented, with the majority of differences being in the construction of the score function, which is specified by how relation conversion operates. Moreover, many of these approaches seek to achieve the optimal balance of model expressive power and complexities in place to ensure predictability across large knowledge networks.

The embedding models acquire entities and relations representations that encapsulate local connection patterns by optimizing a global loss function, facilitating in the understanding of actual revelations by making generalisations over existing ones. According to [12], the following will be a brief summary of state of the art knowledge graph embeddings [26].

The translational and bilinear models are the two primary groups of methodologies. Translational models (e.g., TransE) frequently use a multiplicative method and represent relationships as matrices in the vector space, whereas bilinear models frequently use a multiplicative approach and express relationships as matrices in the vector space. Bilinear models perform well in link prediction tasks in general [41]. RESCAL, DistMult, and ComplEx are the most popular models in this area.

2.4.1 TransE

TransE is a knowledge base embedding model that is based on energy. It interprets relationships as translations on the entities' low-dimensional embeddings in order to model them. Relationships are represented in the embedding space as translations: if holds, the tail entity's embedding should be near to the head entity's embedding plus some vector that relies on the relationship. Bordes [4] proposed the TransE model for the first time in 2013. Consider the relationship $r$in each triple instance (head, relation, tail) as a translation from the head entity h to the tail entity t, and make $(h + r)$ as equal to t as possible by continuously adjusting h, r, and t to achieve the learning goal based on the distributed vector representation of entities and relationships.

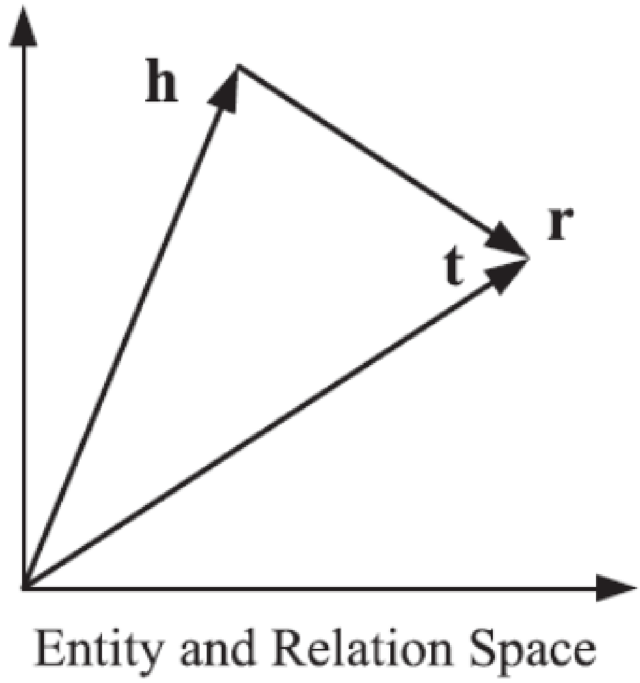

Figure 2.2: Simple illustrations of TransE model

2.4.2 Rescal

RESCAL [28]is a program that computes a three-way factorization of the knowledge graph's adjacency tensor. It may also be understood as a compositional model, in which the tensor product of two things' embeddings is used to represent them. RESCAL is an effective model for capturing complicated relationship patterns over several graph hops.This model models interactions between latent characteristics by representing relations as matrices. It is a bilinear model in which entities are represented as vectors and relationships are represented as matrices. Therefore, the score will be:

$$f(h, r, t) = \mathbf{e}_h^T \mathbf{W}_r \mathbf{e}_t = \sum_{i=1}^{d} \sum_{j=1}^{d} w_{ij}^{(r)} \left(\mathbf{e}_h\right)_i \left(\mathbf{e}_t\right)_j$$

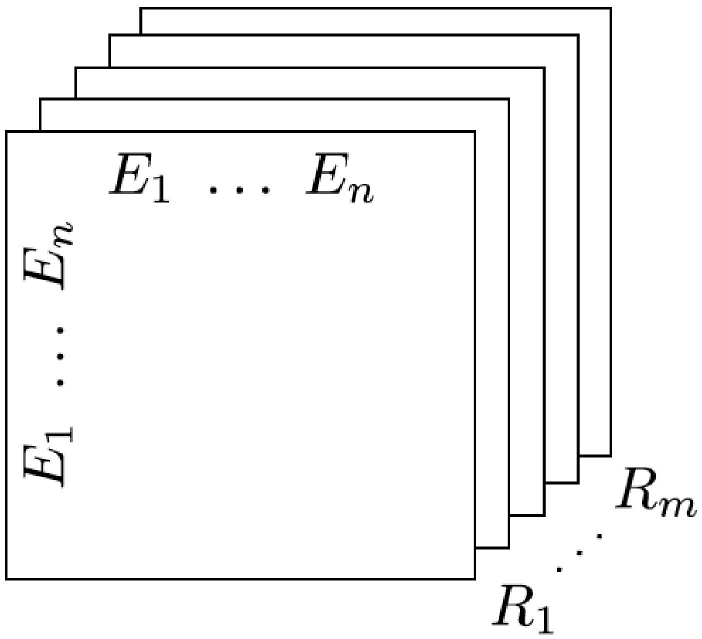

Figure 2.3: Simple Illustration of Rescal Model

2.4.3 ComplEx

The entities and relations in the CompleX model [19]are represented using complex valued representations. The plausibility score is computed using for the and entities and relations. Entities and relations are represented as vectors and the score function is :

$$f(h,r,t) = \operatorname{Re}\left(\mathbf{e}_h \odot \mathbf{r}_r \odot \overline{\mathbf{e}}_t\right)$$

2.4.4 DistMult

By constraining matrices describing relations to diagonal matrices, DistMult[43] simplifies RESCAL. The relation matrices are restricted to diagonal matrices:

$$f(h,r,t) = \mathbf{e}_h^T \mathbf{W}_r \mathbf{e}_t = \sum_{i=1}^{d} \left(\mathbf{e}_h\right)_i \cdot \operatorname{diag}\left(\mathbf{W}_r\right)_i \cdot \left(\mathbf{e}_t\right)_i$$

DistMult is more computationally efficient than RESCAL due to its restriction to diagonal matrices, but it is also less expressive. It cannot, for example, model anti-symmetric relationships, since $f(h,r,t) = f(t,r,h)$. This can alternatively be formulated with relation vectors and the Hadamard operator and the l_1 norm.

$$f(h,r,t) = \left\|\mathbf{e}_h \odot \mathbf{r}_r \odot \mathbf{e}_t\right\|_1$$

2.4.5 RotatE

RotatE [36] was introduced to provide a method to effectively represent symmetric properties in knowledge graph embeddings. The authors of this paper propose to use rotation in a complex space to support symmetry and other properties. RotatE models relations as rotations from head to tail entities in complex space:

$$\mathbf{e}_t = \mathbf{e}_h \odot \mathbf{r}_r$$

The interaction model is then defined as:

$$f(h,r,t) = -\left\|\mathbf{e}_h \odot \mathbf{r}_r - \mathbf{e}_t\right\|$$

which allows to model symmetry, anti-symmetry, inversion, and composition.

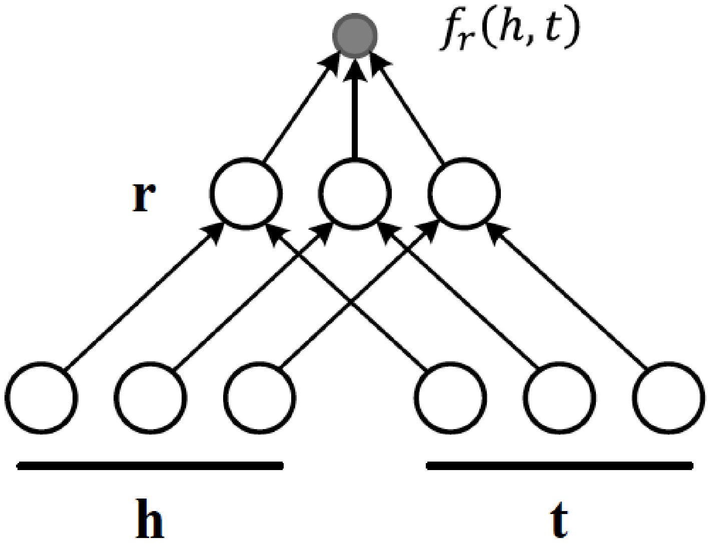

Figure 2.4: Simple Illustration of Rescal Model

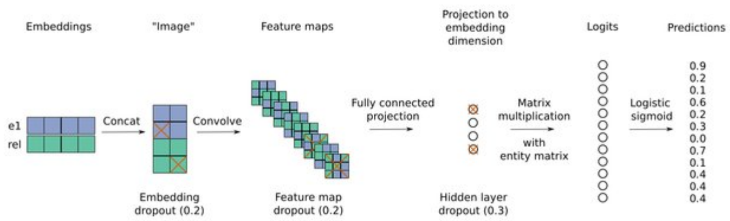

Figure 2.5: ConvE Model Architecture

“

2.5 Knowledge Base Completion

Knowledge base completion is a job that uses reasoning to infer missing facts from the information currently in the knowledge base. A knowledge base is a set of related information that are commonly expressed as "subject", "relation", and "object"-triples.

2.6 Commonsense Knowledge Base Completion (CKBC)

The majority of KBC research focuses on knowledge bases, such as Freebase, that connect things selected from a defined set. ConceptNet uses tuples to create relationships between an infinite number of words. They create neural network models for scoring tuples on arbitrary phrases and assess their ability to differentiate true from false held-out tuples. By framing the problem as one of knowledge base completeness, the coverage of commonsense resources is enhanced. We concentrate on ConceptNet, a curated commonsense resource . Tuples consisting of a left term, a relation, and a right term are found in ConceptNet. The relationships are drawn from a predetermined list. ConceptNet words can be any sentences, but terms in Freebase tuples are entities.The baseline method was based on

	ConceptNet
train	100,000
validation1	1,200
validation2	1,200
test	2,400
size of relation	34
size of vocabulary	21,471
average word length	2.02

Figure 2.6: Snapshot of CKBC Statistics

2.6.1 Task Design of CKBC

The tuples are obtained from the Open Mind Common Sense (OMCS) entries in the ConceptNet5 dataset. They are sorted by a confidence score. The most confident 1200 tuples were reserved for creating the test set (TEST). To generate a development set (DEV1), the next most confident 600 tuples were utilized, and the next most confident 600 tuples were used for (DEV2). For each set , for each tuple S, a negative example was created and added it to S. So each set doubled in size.Each of DEV1 and DEV2 has 1200 tuples.

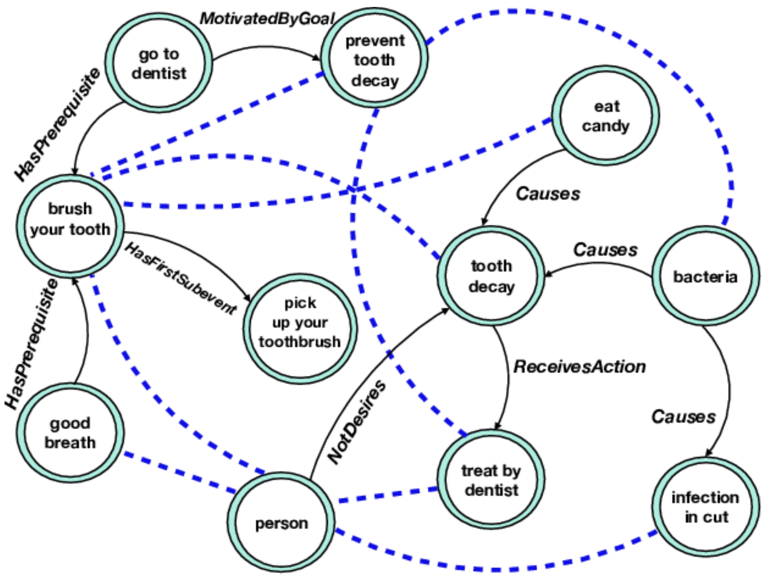

Figure 2.7: CKBC Knowledge Graph

2.7 Ensemble Technique

Ensemble techniques are a type of machine learning methodology that integrates numerous base models to create a single best-fit prediction model. Because of their extensive use, there are three "standard" ensemble learning procedures: bagging, stacking, and boosting. Each strategy has an algorithm that specifies it, but the success of each approach has produced a plethora of extensions and corresponding approaches.

2.7.1 Bagging Learning in a group

The diversity of the ensemble is guaranteed by changes in the bootstrap sample versions on which every classification is trained, and the employment of a particularly poor classifier whose classification algorithm move significantly with respect to comparatively small modifications in the training data. If a row is picked, it is returned to the training dataset for possible re-selection within the same training dataset.A better overall estimate of the desired quantity can be accomplished by producing numerous separate bootstrap samples, estimating a statistical quantity, and finding the mean of the estimates rather than estimating directly from the dataset. More modifications to the training dataset, for example, could be made, the algorithm that fits the training data could be altered, and the mechanism that combines predictions might be changed in ways of voting and so forth.

2.7.2 Stacking Ensemble Learning

Stacking has its own terminology, with level-0 models referring to ensemble members and level-1 models referring to the model that is used to integrate the forecasts.

Instead of a single level-1 model, we may have three or five level-1 models and a single level-2 model that integrates level-1 model predictions to generate a forecast.

2.7.3 Boosting Ensemble Learning

Boosting is an ensemble strategy that attempts to alter the training data in order to focus attention on cases that prior fit models on the training dataset have incorrectly identified. The models are fitted and added to the ensemble in order, with the second model attempting to correct the first model's predictions, the third model correcting the second model, and so on.

2.8 Limitations

Prior approaches on commonsense knowledge base completion task with CKBC dataset includes using deep neural network models, using pre-trained language models and similar alternatives . Although these models performed well, there is still a question that whether they can handle the missing link prediction when it is necessary or not. Moreover, according to our findings and surveys as much as we could, there is still not any implementation on commonsense knowledge base completion task on CKBC dataset with knowledge graph embedding models. Hence, our approach mainly focuses on predicting the connections among entities in a commonsense knowledge base completion dataset.

2.9 Application Area

The followings are the applications that can be more useful with commonsense knowledge.

- Expert systems are for mimicking the judgment and behavior of people in a certain industry, such as banking system, communications, nursing, service and support, logistics, so on and so forth. An expert system is often made up of a specified knowledge base based on cumulative social encounters and a collection of pre-defined criteria issues and scenarios. When confronted with novel scenarios, these systems fail. To transcend over their initial focus and adequately simulate human judgment in new contexts, ESs must have commonsense information as well as the ability to learn from it. [24] [42]

- NLP:Commonsense knowledge is critical for natural language processing tasks like disambiguation and machine translation. Commonsense knowledge is especially important in situations where mere default rules are insufficient and a true comprehension of real-world information is required. Machine translation, for example, is one of the most difficult and unsolved tasks in NLP. Other examples include [13]; [11]; [8], [48], sentiment analysis [7], [33],[30],[45], and [44].

- Computer vision: Commonsense, like NLP, plays a critical role in progressing various critical computer vision tasks, like as [46], [37], and [10].

- Robotics: For autonomous robots functioning in an uncontrolled environment, commonsense thinking is a must. Bots ought to be able to fit in with their surroundings and interrupt scenarios. For example, a droid that is required to comprehend a scenario of a human rock climbing should be aware with the scene's semantics. Based on its existing beliefs and directives, a domestic robot is anticipated to estimate a user's wants. [39].

- To enhance search terms with structured information, browsers or question answering tools such as personal assistants or visual question answering can translate a query into some type of request against an existing knowledge. Furthermore, voice recognition-powered personal secretary software such as Siri has reduced mistake percentages. [2].

Chapter 3

Proposed Method

3.1 Problem Formulation

Reasoning-based approaches for commonsense knowledge mining make educated estimates about what constitutes a legitimate commonsense statement based on analogies and trends derived from established commonsense knowledge regularities. Knowledge graph embedding models learn encoding of graph connections in low-dimensional continuous vector spaces that maintain graph characteristics and functional consistencies by portraying a base of knowledge as a graph with nodes (entities) connected by edges (relations) . After that, these embeddings may be employed for tasks like entity categorization, relation retrieval, and link prediction. The completeness of commonsense knowledge bases is one job that we're interested in and might benefit from these embeddings.To reach our aim, we need to be able to forecast new asserts that may not be already in a knowledge base by bringing in lacking entries of missing triples. We need to guess the absent link for the incoming unidentified triples using knowledge expressed in triples, i.e. (h, r, t), and a scoring function $f(h, r, t)$ that scores right triples greater than wrong triples.

In order to conduct our research, we created a simple ensemble wherein we combined the link projections of knowledge graph embedding methods reported in previous papers. Extending the previous projects, the ultimate probability of a triple is calculated from the ensemble of such assumptions.

$$P\left(x_{h,r,t} = 1 \mid \Theta\right) = \frac{1}{n} \sum_{\theta^m \in \Theta}^{n} P\left(x_{h,r,t} \mid \theta^m_{h,r,t}\right)$$

where

$$P\left(x_{h,r,t} = 1 \mid \theta^m_{h,r,t}\right) = \frac{1}{1 + e^{-\left(\omega^m_1 \theta^m_{h,r,t} + w^m_0\right)}}$$

The pool of model parameters is termed $Theta$, where xh, r, t is the goal factor that indicates when a triple (h, r, t), comprising of the two end objects h and t and the relation r, is correct. We are assessing all potential model combos, so $Theta$ might be a subgroup of the KGE models we addressed in this paper. The predicted confidence scores for triples differ dramatically throughout all models in the ensembles; RESCAL can yield any value, whereas TransE gives negative distances. To adjust the tree methodologies' outputs, we thought to use a straightforward meta-learner however and we expected that this approach will indeed bumble the significance of every single classifier in link prediction tasks. To acquire suitably scaled probability, we employed a PlattScaler one per model based on a small subset of the training data. A Platt-Scaler is a logistic regression model that maps into the interval $[0, 1]$ with precisely one input (the output $thetah, r, tm$ of the model m). $omega1m$ and $omega0m$ reflect the learnt weights and biases of the Platt-Scaler for the model m, respectively.We implement the scalers to the confidence score $thetah, r, tm$ of every model m to produce the likelihood $Pleft(xh, r, tmidthetah, r, tmright)$, which is the likelihood of the triple (h, r, t) provided the models m. Afterwards when, we start adding every one of these probabilities together by finding the arithmetic average.

3.2 Preliminary Experiments

As our goal is to explore the performance of KGE models, we implemented each available model in KGE using PyKeen. PyKeen provides many existing knowledge graph embedding models which makes our experiments more complete. We will discuss about our preliminary experiments in the section 5.2.

Chapter 4

Experimental Setup

4.1 Dataset

Our experimental procedure is based on the [23],CKBC dataset, which uses its test set with an adequate proportion of correct and incorrect triples to evaluate our model.

4.2 PyKeen

The purpose of the PykEnn is to implement knowledge graph embedding models using a variety of modelling techniques, training methodologies, and loss functions. It is also easily to get impact of each constituent on the model's effectiveness independently in PyKeen. Furthermore, an automated memory reduction has been implemented in PyKeen for maximizing the use of the available resources. For our tests, we use GoogleColab. There have already been 27 installed datasets in PyKeen. However, because CKBC is not a built-in dataset, we must import it manually.

4.3 LibKGE

LibKGE (https://github.com/uma-pi1/kge) is a fully accessible PyTorch-based platform for training, hyperparameter adjustment, and evaluation of knowledge base embedding models for link prediction. The major aims of LibKGE are to improve the reproducibility of research, serve as a baseline for large-scale experimentation, and make it simpler to assess the effects of different elements of training techniques, model designs, and assessment processes. Each trial may be successfully duplicated only with a configuration file because to LibKGE's flexibility. To the maximum degree possible, constituents are separated, permitting them to be combined and compared. By a systematic review, LibKGE provides close to state-of-the-art efficiency for several models with a little amount of automatic hyperparameter adjustment.

Chapter 5

Evaluation

5.1 Evaluation Metrics

Mean Reciprocal Rank (MRR)

The mean reciprocal rank is a metric that may be used to evaluate any activities that generate a collection of probable replies to a set of inquiries, sorted by likelihood of accuracy. The multiplicative inverse of the score of the first correct solution is the reciprocal rank of a requests.

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}.$$

where rank_i refers to the rank position of the first relevant document for the i-th query.

Mean Rank

To compute the average rank of outputs from the document, we need mean rank.

$$\text{score } = \frac{1}{|\mathcal{I}|} \sum_{r \in \mathcal{I}} r$$

It has the benefit over hits @ k in that it is attentive to any variations in prediction accuracy, including those that happen below certain threshold, and hence captures average rating. The mean rank in PyKEEN's conventional 1-based indexing is on the range $[1, infty)$, with lesser being preferable.

Hits at k (Hits@k)

This metric is calculated by the percentage of correct triples resulted in the top elements in which the number is k. The overall k-truncated list may consists of many correct assertions, this value may surpass 1.00. The proportion of genuine entities that occur in the first entities of the sorted rank list is described by hits @ k. It is written as follows: For instance, if a search engine returns 10 results on the first page, the hits @ 10 is the proportion of relevant results.

5.2 Results

First, the built-in models from PyKeen were tested and evaluated on CKBC dataset to see the results of each model so that we can see how each model performs individually for the dataset. We did not use any tuned hyperparameters because we wanted to know the original performace on each model. That is why we used the default pipeline model training from PyKeen.

Next, based from the results of table 2, we chose top 5 models. From these results, we can see that DistMult significantly worse than other models and ComplEx and ConvE show best results overall. In filtered hits@K results, most model performs slightly better and ConvE improves the most. After applying ensemble method to our

Models	Mean rank	Mean reciprocal rank
TransE	4112.73	0.0279
TransD	9877.45	0.0125
TransF	23081.42	4.03E-03
TransH	33950.49	2.00E-03
TransR	30769.40	2.19E-02
RotatE	13502.05	1.30E-02
TorusE	5086.65	0.0064
StructuredEmbedding	38003.16	9.95E+00
SimplE	16781.44	1.80E-03
RGCN	8256.60	4.44E-05
RESCAL	4081.01	0.078
QuatE	5751.13	1.00E-04
ProjE	8419.89	0.00122
PairRE	5278.03	1.00E-03
MuRE	38645.65	6.06E-05
HolE	20254.22	1.00E-04
ERMLPE	28110.41	3.70E-03
ERMLP	41002.39	2.00E-04
DistMult	3873.39	0.051
DistMA	9037.09	3.00E-04
CrossE	8963.77	6.74E-05
ConvE	3711.98	0.0938
ComplEx	3915.91	0.0748

Table 5.1: Preliminary Results obtained from default training

Models	Mean rank	Mean rank (F)	Mean reciprocal rank	Mean reciprocal rank (F)
TransE	4112.7341	4105.6158	0.0279	0.0292
Rescal	4081.0158	4069.3066	0.0787	0.1422
ComplEx	3915.9141	3904.2066	0.0748	0.1496
DistMult	3873.3972	3861.6883	0.0518	0.0867
ConvE	3711.9875	3700.6108	0.0938	0.1676

Table 5.2: Selection models for mean rank and mean reciprocal rank metrics

Models	Hits@1	Hits@10	Hits@100	Hits@1000
TransE	0.088	0.202	0.303	0.400
Rescal	0.076	0.138	0.198	0.281
ComplEx	0.109	0.177	0.240	0.313
DistMult	0.003	0.006	0.015	0.038
ConvE	0.080	0.195	0.205	0.381

Table 5.3: Preliminary Results for unfiltered hits@k results

Models	Hits@1	Hits@10	Hits@100	Hits@1000
TransE	0.091	0.207	0.317	0.400
Rescal	0.094	0.148	0.213	0.281
ComplEx	0.128	0.233	0.308	0.382
DistMult	0.003	0.006	0.015	0.038
ConvE	0.185	0.366	0.326	0.425

Table 5.4: Preliminary Results for filtered hits@k results

Models	Hits@1	Hits@10	Hits@100	Hits@1000
TransE	0.088	0.202	0.303	0.400
Rescal	0.076	0.138	0.198	0.281
ComplEx	0.109	0.177	0.240	0.313
DistMult	0.003	0.006	0.015	0.038
ConvE	0.080	0.195	0.205	0.381
Rescal + TransE	0.092	0.142	0.247	0.314
ComplEx + TransE	0.138	0.251	0.332	0.428
ConvE + TransE	0.190	0.281	0.310	0.501
ComplEx+ ConvE + TransE	0.162	0.228	0.285	0.415

Table 5.5: Ensemble results

Chapter 6

Conclusion and Future Work

Using knowledge graph embedding models, we suggested a strategy for completing a commonsense knowledge base with regard to link prediction. We demonstrated how we can improve performance above the baseline by ensembling the current KGE models. In the future, we'll look at how we can utilize our model to better downstream NLP jobs, as well as how we can adapt our approaches to different commonsense knowledge bases.

Bibliography

[1] Mehdi Ali et al. "PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings". In: *Journal of Machine Learning Research* 22.82 (2021), pp. 1–6.

[2] Stanislaw Antol et al. "Vqa: Visual question answering". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 2425–2433.

[3] Kurt Bollacker et al. "Freebase: a collaboratively created graph database for structuring human knowledge". In: *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. 2008, pp. 1247–1250.

[4] Antoine Bordes et al. "Translating embeddings for modeling multi-relational data". In: *Advances in neural information processing systems* 26 (2013).

[5] Samuel Broscheit et al. "LibKGE-A knowledge graph embedding library for reproducible research". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. 2020, pp. 165–174.

[6] Erik Cambria, Daniel Olsher, and Dheeraj Rajagopal. "SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis". In: *Twenty-eighth AAAI conference on artificial intelligence*. 2014.

[7] Erik Cambria et al. "Affective computing and sentiment analysis". In: *A practical guide to sentiment analysis*. Springer, 2017, pp. 1–10.

[8] Erik Cambria et al. "Senticnet: A publicly available semantic resource for opinion mining". In: *2010 AAAI fall symposium series*. 2010.

[9] Timothy Anatolievich Chklovski. "Using analogy to acquire commonsense knowledge from human contributors". PhD thesis. Massachusetts Institute of Technology, 2003.

[10] Bob Coyne and Richard Sproat. "WordsEye: An automatic text-to-scene conversion system". In: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. 2001, pp. 487–496.

[11] Jon Curtis, John Cabral, and David Baxter. "On the Application of the Cyc Ontology to Word Sense Disambiguation." In: *FLAIRS Conference*. 2006, pp. 652–657.

[12] Vincenzo Cutrona et al. "Tough tables: Carefully evaluating entity linking for tabular data". In: *International Semantic Web Conference*. Springer. 2020, pp. 328–343.

[13] Kathleen Dahlgren and Joyce P McDowell. "Using Commonsense Knowledge to Disambiguate Prepositional Phrase Modifiers." In: *AAAI*. 1986, pp. 589–593.

[14] Ernest Davis. *Representations of commonsense knowledge*. Morgan Kaufmann, 2014.

[15] Zhendong Dong and Qiang Dong. "HowNet-a hybrid language and knowledge resource". In: *International Conference on Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003*. IEEE. 2003, pp. 820–824.

[16] Lisa Ehrlinger and Wolfram Wöß. "Towards a Definition of Knowledge Graphs." In: *SEMANTiCS (Posters, Demos, SuCCESS)* 48.1-4 (2016), p. 2.

[17] Vladimir Fomichov. *Semantics-Oriented Natural Language Processing: Mathematical Models and Algorithms*. Vol. 27. Springer Science & Business Media, 2009.

[18] Andrew Galbraith et al. "The formation of cycl [3.2. 2] azine derivatives via the reaction of pyrrocoline with dimethyl acetylenedicarboxylate". In: *Journal of the American Chemical Society* 83.2 (1961), pp. 453–458.

[19] Robert Grone et al. "Positive definite completions of partial Hermitian matrices". In: *Linear algebra and its applications* 58 (1984), pp. 109–124.

[20] Nicola Guarino and Pierdaniele Giaretta. "Ontologies and knowledge bases". In: *Towards very large knowledge bases* (1995), pp. 1–2.

[21] Jens Lehmann et al. "Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia". In: *Semantic web* 6.2 (2015), pp. 167–195.

[22] Douglas B Lenat. "CYC: A large-scale investment in knowledge infrastructure". In: *Communications of the ACM* 38.11 (1995), pp. 33–38.

[23] Xiang Li et al. "Commonsense knowledge base completion". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2016, pp. 1445–1455.

[24] John McCarthy. "Some expert systems need common sense". In: (1984).

[25] George A Miller. "WordNet: a lexical database for English". In: *Communications of the ACM* 38.11 (1995), pp. 39–41.

[26] Sameh K Mohamed et al. "Loss Functions in Knowledge Graph Embedding Models." In: *DL4KG@ ESWC* 2377 (2019), pp. 1–10.

[27] ET Mueller. "Natural Language Processing with ThoughtTreasure, Signiform". In: *New York* (1998).

[28] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. "A three-way model for collective learning on multi-relational data". In: *Icml*. 2011.

[29] Ian Niles and Adam Pease. "Mapping WordNet to the SUMO ontology". In: *Proceedings of the ieee international knowledge engineering conference*. 2003, pp. 23–26.

[30] Ethel ChuaJoy Ong. "A commonsense knowledge base for generating children's stories". In: *2010 AAAI Fall Symposium Series*. 2010.

[31] Heiko Paulheim. "Knowledge graph refinement: A survey of approaches and evaluation methods". In: *Semantic web* 8.3 (2017), pp. 489–508.

[32] Daniel Ruffinelli, Samuel Broscheit, and Rainer Gemulla. "You can teach an old dog new tricks! on training knowledge graph embeddings". In: *International Conference on Learning Representations*. 2019.

[33] Push Singh, Barbara Barry, and Hugo Liu. "Teaching machines about everyday life". In: *BT Technology Journal* 22.4 (2004), pp. 227–240.

[34] Push Singh et al. "Open mind common sense: Knowledge acquisition from the general public". In: *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer. 2002, pp. 1223–1237.

[35] Robyn Speer, Joshua Chin, and Catherine Havasi. "Conceptnet 5.5: An open multilingual graph of general knowledge". In: *Thirty-first AAAI conference on artificial intelligence*. 2017.

[36] Zhiqing Sun et al. "Rotate: Knowledge graph embedding by relational rotation in complex space". In: *arXiv preprint arXiv:1902.10197* (2019).

[37] Wandi Susanto, Marcus Rohrbach, and Bernt Schiele. "3d object detection with multiple kinects". In: *European Conference on Computer Vision*. Springer. 2012, pp. 93–102.

[38] Niket Tandon et al. "Webchild: Harvesting and organizing commonsense knowledge from the web". In: *Proceedings of the 7th ACM international conference on Web search and data mining*. 2014, pp. 523–532.

[39] Moritz Tenorth et al. "Knowrob-map-knowledge-linked semantic object maps". In: *2010 10th IEEE-RAS International Conference on Humanoid Robots*. IEEE. 2010, pp. 430–435.

[40] Luis Von Ahn and Laura Dabbish. "Designing games with a purpose". In: *Communications of the ACM* 51.8 (2008), pp. 58–67.

[41] Yanjie Wang, Rainer Gemulla, and Hui Li. "On multi-relational link prediction with bilinear models". In: *Thirty-Second AAAI Conference on Artificial Intelligence*. 2018.

[42] Donald A Waterman. *A guide to expert systems*. Addison-Wesley Longman Publishing Co., Inc., 1985.

[43] Bishan Yang et al. "Embedding entities and relations for learning and inference in knowledge bases". In: *arXiv preprint arXiv:1412.6575* (2014).

[44] Fei Yin et al. "ICDAR 2013 Chinese handwriting recognition competition". In: *2013 12th international conference on document analysis and recognition*. IEEE. 2013, pp. 1464–1470.

[45] Mary Young et al. "Becoming 'real'Aboriginal teachers: Attending to intergenerational narrative reverberations and responsibilities". In: *Teachers and Teaching: theory and practice* 16.3 (2010), pp. 285–305.

[46] Zheng Yuan et al. "Role of magnetic resonance diffusion-weighted imaging in evaluating response after chemoembolization of hepatocellular carcinoma". In: *European journal of radiology* 75.1 (2010), e9–e14.

[47] Liang-Jun Zang et al. "A survey of commonsense knowledge acquisition". In: *Journal of Computer Science and Technology* 28.4 (2013), pp. 689–719.

[48] Yaoyun Zhang et al. "ICRC_HITSZ at RITE: Leveraging Multiple Classifiers Voting for Textual Entailment Recognition." In: *NTCIR*. Citeseer. 2011.