Title	クラスタリングに基づく弱教師あり学習によるレビューの暗黙的属性の分類
Author(s)	AYE AYE MAR
Citation	
Issue Date	2024-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/19067
Rights	
Description	Supervisor: 白井 清昭, 先端科学技術研究科, 博士

Doctoral Dissertation

Implicit Aspect Classification of Online Reviews by Clustering-based Weak Supervision

Aye Aye Mar

Supervisor: Kiyoaki Shirai

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
[Information Science]

March, 2024

Supervised by:

Professor SHIRAI Kiyoaki

Reviewed by:

Professor SHINNOU Hiroyuki

Professor NGUYEN Minh Le

Professor INOUE Naoya

Professor OKADA Shogo

Abstract

Aspect Term Extraction (ATE) which is a process of extracting an aspect (also known as opinion target) from a customer review sentence plays a vital role in Aspect-Based Sentiment Analysis (ABSA). Many previous work of ATE focused on explicit aspect but only a few work considered to extract implicit aspects. However, customer reviews containing implicit aspects are widespread on the Web (such as Amazon.com) and these sentences are also important to fully understand the opinions and sentiments of the customers.

One of the bottleneck problem of implicit aspect extraction is lack of a large dataset of reviews annotated with implicit aspects. Although the corpus annotated with implicit aspects is required for every domain due to different types of aspects in different domains, constructing a corpus is labour intensive and time consuming. Therefore, a system to automatically construct a dataset annotated with implicit aspects is required. This study proposed a novel approach that automatically constructs a dataset annotated with implicit aspects using unlabelled Amazon reviews to address the challenge of implicit aspect extraction. To the best of our knowledge, no prior work has been performed on the automatic construction of such a dataset.

The goal of this study is to develop a system of ATE for implicit aspects. A dataset labeled with implicit aspects is automatically constructed by guessing implicit aspects in unlabeled review sentences. The proposed method involves clustering review sentences labeled with explicit aspects (which were extracted by CRF model trained on golden explicit review sentences) and unlabeled review sentences. In this study, using a K-means clustering approach with a relatively large number of clusters (10% of total review sentences) aims to generate many small but accurate clusters.

Cluster labels, considered as implicit aspects, are automatically assigned based on the assumption that sentences with similar context share a common aspect. When selecting the most relevant cluster label among the explicit aspects in the cluster, the frequency of the aspect in the list of aspects extracted by CRF and its occurrence in the review sentences within the cluster are considered to determine

the relevance of the chosen cluster label. When there is more than one aspect that can be the cluster label, we did not consider such kind of cluster since the cluster label is not unique. Moreover, the reliability of the cluster label to be chosen was determined by the threshold value (T_r). Unlabeled sentences in clusters matching pre-defined implicit aspect categories are then obtained as implicit-aspect-labeled sentences. To increase the number of clusters related with the implicit aspects, the aspect synonym list was identified.

The accuracy of the constructed corpus was evaluated by a human annotator by checking manually on 50 random sentences for each implicit aspect. The results showed that accuracy of the sentences in the constructed corpus was reasonably high, i.e., from 0.58 to 0.82. The study presents findings and observations regarding with constructing the corpus annotated with implicit aspects.

In this study, implicit aspect extraction problem is formulated as classification problem. Then, BERT model is fine-tuned for implicit aspect classification using the constructed dataset by investigating the best values of hyper-parameters. Experiments results of implicit aspect classification show that our method achieves 82% and 84% accuracy for the mobile phone and PC reviews respectively, which are 20 and 21 percentage points higher than the baseline.

Furthermore, the study explores the impact of explicit review sentences for implicit aspect classification by combining the explicit sentences and implicit sentences and then by training classification model on the combined dataset. The experimental results showed that it further boosts the performance of implicit aspect classification in both phone and PC domain.

Keywords: Aspect-based Sentiment Analysis, Aspect Extraction, Implicit Aspect, Weakly-supervised Learning, Online Review

Acknowledgements

My studies at Japan Advanced Institute of Science and Technology (JAIST) have been truly incredible. None of this would have been possible without the generous support of many individuals.

Firstly, I would like to express my deepest gratitude to my esteemed supervisor, Professor **SHIRAI Kiyoaki**, for his kind and invaluable supervision, unwavering support, continuous patience and encouragement throughout my doctoral degree program. His expertise and dedication have been instrumental in achieving my research goal and have been a source of inspiration to me. Without his mentorship, insights and counseling, I could not have completed my Ph.D. study. I am particularly grateful for the peaceful and inspiring environment he created, fostering growth both personally and academically. Despite the hardship, my Ph.D. student life under his excellent mentorship has been rewarding.

I would like to extend my deep appreciation to my second supervisor, Professor **NGUYEN Minh Le** who also served as one of the committee members of my Ph.D. defense from JAIST, for his valuable comments and suggestions. I wish to send my sincere thanks to my minor research supervisors Professor **TOJO Satoshi** and Professor **TANAKA Kiyofumi** for supporting me kindly.

Words cannot express my heartfelt gratitude to the committee members of my Ph.D. defense, Professor **SHINNOU Hiroyuki** from Ibaraki University, Professor **INOUE Naoya** and Professor **OKADA Shogo** from JAIST for their fruitful feedback, comments, and suggestions.

I am grateful to Senior Lecturer **KERTKEIDKACHORN Natthawut** from our Laboratory for his counseling, insightful suggestions, helpful comments, and kind support.

I genuinely would like to express my profound gratitude to the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) for graciously awarding me a scholarship. Their generous support has played a pivotal role in enabling my academic pursuits, and I am truly thankful for the opportunities this scholarship has provided me.

I sincerely thank professors and lecturers for teaching interesting courses. Furthermore, I would like to extend my thanks to staff members for consistently creating a safe, convenient, and wonderful environment at JAIST.

I am thankful to my friends for their kind support and for being there to share both my joys and challenges.

I would like to express my appreciation to the unknown significant others (USOs) who have contributed to my academic journey, whether through direct or indirect support, in completing my studies.

Lastly, I am profoundly grateful to my family for their unwavering love, support, and understanding.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research Background

Nowadays, people are more interested in online shopping and their reviews about the products has become the information that draw attention not only to consumers but also manufactures and researchers. While some customers write their opinions and feelings online about the products they bought, some customers are seeking out the the opinions of the other customers about the product they are interested in. Besides the customers, manufacturers are also listening to the customer voices about their products to infer how their products and services are perceived [46]. The exponential growth of digital contents, particularly a substantial volume of online reviews, coupled with the increasing interest of both customers and manufacturers, has significantly amplified the importance of sentiment analysis. It has now become a prominent research area among researchers.

Sentiment analysis, also known as opinion mining, is a subfield of natural language processing (NLP) to determine and analyze the sentiments, attitudes, opinions, or emotions expressed in written text. The primary goal of sentiment analysis is to identify the polarity (positive or negative) of opinions written in customer reviews [46]. According to Hu and Liu [25], there are three kinds of review formats which are are available on the Web.

- Format 1 - Pros and Cons: The format in which the reviewers describe Pros and Cons separately about the product. The reviews from C|Net.com are written in this format.

- Format 2 - Pors, Cons and detailed review: The format in which the reviewers write the Pros and Cons separately as well as description in detials about the product. The reviews from epinion.com and MSN are written in this format.

- Format 3 - free format: the format in which reviewers describe freely about a product without separating Pros and Cons. The reviews from Amazon.com are in this format.

Many studies have been done on sentiment analysis on reviews [4, 5, 63, 23, 34, 37, 76, 65, 64, 1, 68, 40, 67, 74, 14, 2, 29, 20]. There are three levels of sentiment analysis which are document-level, sentence-level and aspect-level (also know as feature-level) [36, 39]. In the document-level, the overall polarity of a document is decided as a positive opinion or negative opinion based on the opinion words in the document by considering the whole document as one topic [47, 72]. In the sentence-level, the polarity of a whole sentence is decided without computing the sentiment for each aspect described in the sentence [75]. The fine-grained level sentiment analysis is aspect-level sentiment analysis which is also called Aspect-Based Sentiment Analysis (ABSA) [70, 42, 73, 56]. It aims at inferring the sentiment of a customer at a fine-grained level by determining the polarity on each aspect of a product, such as "price" and "battery" of a mobile phone. ABSA plays a vital role not only for customers but also for manufacturers, because it allows customers to find the strong and weak points of a product in which they are interested, while manufacturers can identify the customers' needs and expectations accurately.

In general, ABSA consists of two subtasks: Aspect Term Extraction (ATE) and Aspect Polarity Classification (APC). The former involves extracting aspects of a product from a sentence in a review, while the latter aims at classifying whether a customer expresses a positive, neutral, or negative opinion about each extracted aspect. Two kinds of aspects should be considered in ATE: explicit and implicit aspects. Explicit aspects are those that appear as explicit words or phrases in the review sentences, while implicit aspects are expressed implicitly, without directly mentioning the name of the aspect [23]. Explicit aspect expressions are nouns and noun phrases. However, implicit aspect expressions are not just adjectives, adverbs, verbs and verb phrases; they can also be very complex [36]. Some examples of these are as follows.

S1 The battery of the phone lasts many hours, so it does not need to charge frequently.

S2 I don't use it any more, as I get tired of always recharging after using just for a few hours.

Both the sentences S1 and S2 mention the same aspect of the mobile phone, the "battery". S1 contains the explicit aspect "battery" and directly expresses an opinion about it, whereas S2 implicitly expresses an opinion about the battery without using the word "battery" itself. In the second sentence, the battery is an implicit aspect.

Many research has been done explicit aspect extraction from review sentences [79, 25, 24, 43, 7, 28, 45, 50, 9]. Although there has been a great deal of work on extracting explicit aspects, the identification of implicit aspects has not been vigorously studied [71, 53, 48, 59, 41, 13]. Customer reviews typically offer straightforward information (explicit), but they can also include subtle or undisclosed details (implicit) that are important to take into account since neglecting these implicit aspects may lead to inaccurate decisions or recommendations for individual users [69]. Implicit aspects are also important in order to fully understand the opinions and sentiments of customers, since customer reviews containing implicit aspects are widespread on the internet. Zhang and Zhu showed that 30% of the reviews in their corpus contained implicit aspects [80]. Similarly, Cai et al. showed that 44% of the review sentences about laptop PCs contain implicit aspects or implicit opinions [6]. Implicit review sentences are more complex than explicit ones [36]. In addition, capturing the most significant implicit aspects heavily relies on the accessibility of domain-specific knowledge [52]. On the other hand, different people implicitly describe their sentiments about products using different kinds of linguistic expressions and writing styles, meaning that implicit aspects are more difficult to handle in ABSA than explicit ones.

The lack of a large dataset of reviews annotated with implicit aspects is one of the bottlenecks for implicit aspect extraction. Most current methods of ATE rely on supervised learning, in which an aspect extraction model is trained on a labeled dataset [38]. Although explicit aspect extraction were performed with Support Vecotr Machine (SVM) [31] and deep learning model such as Convolutional Neural Network (CNN) [30] and Bidirectional Encoder Representations from Transformers (BERT) [9] by using standard SemEval-2016 dataset that was annotated with explicit aspects, the extraction of implicit aspects cannot be performed in the same way when there is no dataset labeled with implicit aspects.

1.2 Research Objectives

The goal of this study is to develop a system of ATE for implicit aspects. A dataset labeled with implicit aspects is automatically constructed by guessing implicit aspects in unlabeled review sentences. Then, a model for implicit aspect extraction is obtained using the pre-trained language model. We demonstrate the effectiveness of our proposed method when it is applied to customer reviews in two domains: mobile phones and PCs, and present our findings on the complex nature of the implicit aspects and problems in the construction of the dataset through an error analysis. Our contributions can be summarized as follows.

- We propose a novel weakly supervised method to construct a dataset automatically labeled with implicit aspects. To the best of our knowledge, no prior work has been performed on the automatic construction of such a dataset.
- We train a model for implicit aspect extraction by fine-tuning a pre-trained language model using the above dataset coupled with existing review sentences including explicit aspects.
- We empirically evaluate the constructed dataset as well as the performance of implicit aspect extraction achieved by our proposed model.

The most notable contribution is the first one. This study is the first attempt to automatically construct a corpus annotated with implicit aspects.

1.3 Research Questions

Our major research question is as follows.

> *How to develop a model for implicit aspect extraction without using a labeled dataset?*

To answer this research question, we consider a weakly supervised approach. We construct a dataset labeled with implicit aspects automatically, then we train an implicit aspect extraction model from this dataset. In addition, we investigate the additional use of an existing dataset labeled with explicit aspects for implicit aspect extraction.

The above major research questions can be decomposed into three sub-research questions as follows.

RQ1 How to automatically construct a corpus annotated with implicit aspects, which is sufficiently large for training a model?

RQ2 How is the performance of implicit aspect classification by a model trained from an automatically constructed corpus?

RQ3 Is it possible to improve the performance of implicit aspect extraction using both datasets of implicit and explicit aspects?

1.4 Chapter Organization

This dissertation is structured as follows. Chapter 2 gives the literature reviews by divining into two categories. Firstly, we describe related work about implicit aspect extraction task. Then, we introduce previous work regarding with dataset of implicit aspects. We present the most closely related work with our research and clarify the difference between their work and our work.

Chapter 3 explains our proposed system by firstly introducing the overview architecture and explanation. Next, step by step processes of automatically constructing the corpus annotated with implicit aspects are presented with the illustration of flowchart diagram. Then, we fine-tune the BERT model with our constructed corpus and realize implicit aspect classification by the trained BERT model.

In Chapter 4, we evaluate our proposed system. Firstly, we described the experimental setup of the construction of the dataset. Then, the accuracy of the constructed corpus is checked and computed by a human evaluator and results are presented. Next, we make error analysis of constructed corpus and presented our findings and observations. We also conduct an experiment to evalute the performance of implicit aspect classification. First, we describe the experimental setup of implicit aspect classification and optimize hyper-parameters. Next, the performance of implicit aspect classification model which is trained on three different types of corpora is investigated.

Finally, Chapter 5 summarizes the dissertation with our main contributions as well as the remaining problems. We also describe the future work which can improve some parts of our system.

Chapter 2

Literature Review

2.1 Implicit Aspect Extraction

As discussed in the previous chapter, current methods of ABSA mainly focus on explicit aspects. However, there have been a few attempts to handle implicit aspects [71, 53, 48, 59, 41, 13].

Fei et al. proposed dictionary-based approach for identifying aspects implied by adjectives by formulating the problem as collective classification [12]. They compared their work with WordNet that can attributes given an adjectives and their work performed better than WorkNet.

Zainuddin et al. used the dependency relation rules which can be extracted using dependency parser in order to extract the implicit aspect from hate crime reviews in Twitter [77]. They applied hybrid approach between association rule mining, dependency parsing and SentiSordNet to perform the aspect-based sentiment analysis.

Qui applied semantic ontology as a method to extract implicit aspects [35]. Within the ontology framework, Qui established semantic relations linking various aspects to each entity. The identification and extraction of opinion words lacking explicit features were integral to their approach. The researcher then formulated equations to measure the semantic similarity between features and opinion words, leveraging the underlying ontology. Finally, the extraction of implicit features was conducted based on the associations calculated through this semantic analysis.

Hajar et al. introduced a hybrid methodology that incorporated a corpus, WordNet, and a Naïve Bayes classifier [18]. Initially, the approach involved the extraction of all adjectives from a designated corpus. Subsequently, WordNet was employed to identify words with lexical connections to these adjectives, encompassing synonyms and antonyms. Finally, the acquired dataset was utilized to train a Naïve Bayes classifier, facilitating the implicit extraction of aspects.

Many previous studies considered correlations between sentiment words (such as "excellent" and "bad") and aspect words. Su et al. [60] applied Point-wise Mutual Information (PMI) based sentiment association analysis to identify the implicit aspects. They only presented the PMI scores for some word examples and few mapping results of some example words. The quantitative experimental results such precision, recall have not been described in their work.

Hai et al. proposed a co-occurrence association rule mining approach for identifying implicit aspects [17]. The association rule was in the form (sentiment word $\rightarrow$ explicit aspect), indicating that the sentiment word and explicit aspect frequently co-occurred in a sentence. The rules were generated from a review corpus and converted to more general rules that mapped each sentiment word to a cluster of aspects. The obtained rules were then applied to identify the implicit aspect of sentences that included not an explicit aspect but a sentiment word. The results of an experiment using Chinese review data showed that the F1- score for implicit aspect identification was 74%.

Zang and Li highlighted a limitation of the above association rule-based method in that only a single aspect can be associated with a sentiment word by a rule, but two or more aspects can be related to one sentiment word [78]. For example, since the sentiment word "good" is a general one, it can express opinions towards many aspects, such as "battery", "screen" or "quality". The contextual information of the sentiment word is necessary to identify the exact aspect. Based on this finding, they proposed a classification-based method for the identification of an implicit aspect, where the task was formulated as a classification problem. First, pairs containing an explicit aspect and a sentiment word were obtained by a rule-based method, where the rules were used to extract the pairs from the results of a dependency parsing of the review sentences. Then, sentences including an explicit aspect and a sentiment word were excerpted as a document collection, and were labeled with the aspect. Using this document collection as training dataset, a topic-feature-centroid classifier was trained using bag-of-words features. They evaluated their method on Chinese reviews of cell phones and cameras on Amazon, and found that the F1-scores were 74.66% and 78.76% respectively, i.e., better than the association rule-based method of Hai et al.[17]. The reason for this was that their classification-based approach was able to capture the context of the sentiment words and infrequent dependencies between aspects and sentiment words that were not considered in the association rules.

Sun et al. proposed a method which takes into account not only opinion words but also the context information for extracting implicit features from Chinese mobile phone reviews and computer reviews [61]. In their method, there are three steps to extract implicit features. First, they built a co-occurrence matrix to show the relationship between opinion words and product features. Second, they determined whether there is an implicit feature, and if so, they found a candidate set of implicit features. Finally, they found the correct implicit features according to the candidate implicit features scores calculated based on the opinion words and the context information of the implicit features. They used the data set containing 3656 reviews from mobile phone users and 2446 reviews from computer users. They computed the recall and precision for identifying the implicit features by manual evaluation. For the extraction of features and opining words, their method outperformed Double Propagation method [51] which propagated the information through both sentiment words and features.

Bagheri et al. proposed a graph-based method for implicit aspect extraction [3]. The vertices in the graph were either explicit aspects or sentiment words, while the edges between them were weighted based on the number of their co-occurrences and the degree of the vertices in the graph. To construct the graph, explicit aspects were extracted using an iterative bootstrapping algorithm, starting with the initial seed aspects. For a given review sentence, an aspect connected to sentiment words in the sentence with highly weighted edges was extracted as an implicit aspect. Their method was evaluated using a dataset of reviews of five products, constructed by Hu and Liu [23], and showed that the F1-scores for the implicit aspect extraction method were between 57% and 71%.

Several studies have trained models for implicit aspect extraction using a dataset labeled with implicit aspects. Hendriyana et al. proposed the sentence-level topic model (SLTM) [22]. Their work extracted explicit and implicit features from dummy reviews and amazon customer reviews from the amazon e-commerce website, specifically focusing on the Xperia Z smartphone. They described that amazon customer reviews data they used contain 1407 sentences containing eight features (aspects); camera, battery, endurance, screens, operating system, networks, audio and price. The feature extraction method utilized in their work is the SLTM (Sentence Level Topic Model). Preprocessing steps, such as lowercasing, tokenizing, stop-word removal, lemmatization, and part-of-speech tagging, were applied to structure the unstructured text data from amazon reviews. When applied to the dummy dataset, the system exhibited a performance of 76% in extracting explicit features and 92.59% in extracting implicit features. Nevertheless, when tested on the amazon customer review dataset, the system's performance yielded 88.24% for explicit features and 60% for implicit features. Notably, the system can only detect one product feature in a single sentence.

Schouten and Frasincar aimed to find implicit aspects from sentences that could contain zero or more implicit aspects [57]. A training dataset was first constructed, and the co-occurrence matrix C of implicit aspects and words was created from the training data. For a given sentence, scores for the implicit aspects were calculated using the matrix C, and the implicit aspect with the highest score was chosen. However, if the maximum score was less than a given threshold, the system judged that the sentence contained no implicit aspect. They demonstrated the effectiveness of their method using consumer reviews of products and restaurants. They also reported that a large proportion of the sentences in the product reviews had no implicit aspect, meaning that in such cases, a classification-based approach might not be feasible.

Soni and Rambola proposed a hybrid method incorporating a Recurrent Neural Network (RNN) that is trained on a dataset prepared by themselves and similarity calculations based on WordNet and similarity function from spaCy to detect implicit aspects. Their dataset consisted of 700 records obtained from 160 Samsung M21 phone reviews in Amazon. Precision, recall, and F-measure(weighted average) achieved by their proposed method were 48.7, 42.4 and 41.6, respectively. They described that their method can be improved through dataset enhancement, adoption of alternative text vector representations, utilization of additional lexical resources, exploration of alternative models, or optimization of parameters. They supposed that the dataset should contain more records since 700 records are inappropriate for training and testing the model in their work. Moreover, they mentioned that the dataset should be balanced.

Unlike the previous studies, our proposed method relies on supervised learning using a pre-trained language model that worked well for various natural language processing (NLP) tasks. In addition, instead of using a manually labeled dataset, we use a labeled dataset constructed by a weakly supervised method that requires no human effort. This enables us to develop a large dataset annotated with implicit aspects automatically.

2.2 Dataset of Implicit Aspects

Only explicit aspects are annotated in the most commonly used datasets for ABSA, such as Sentihood [55] and SemEval-2014 Task 4 Aspect Based Sentiment Analysis (we call it "SemEval-2014 dataset" hereafter) [49]. However, a small or pilot dataset with implicit aspects has been constructed. Hu and Liu developed a dataset for ABSA that consisted of corpora based on five product reviews: two digital cameras, a cellular phone, an MP3 player, and a DVD player [23]. Both the explicit and implicit aspects were manually annotated. However, Hu and Liu's dataset was relatively small, and the number of sentences containing implicit aspects for each of the five products was between 14 and 55. Cruz et al. extended this dataset by adding annotations of implicit aspect indicators (IAIs), which were sentiment words indicating a certain implicit aspect [10]. They selected sentences labeled with at least one implicit aspect from Hu and Liu's dataset, and then manually annotated the sentences with the IAIs. They then used the extended dataset to train a conditional random field (CRF) to extract IAIs from the review sentences.

Most methods for ABSA are based on supervised learning, which requires a labeled dataset [66]. In addition, the aspects mentioned in each review are very different for different product types or domains. To perform ABSA for various types of products, it is necessary to individually construct a labeled dataset for each domain. This is our primary motivation for the automatic construction of a large review dataset annotated with implicit aspects.

There have been a few attempts to automatically construct a dataset with explicit aspects. Giannakopoulos et al. constructed a new dataset from Amazon computer reviews [15]. They also proposed a method which used attention mechanism to select sentences from amazon reviews for constructing a dataset with explicit aspects. In their work, they used attention mechanism to remove some noisy sentences and predict review ratings. They kept only the reviews that have at least 3 sentences and at most 10 sentences to leverage attention mechanism. They assigned sentiment scores to each sentence in the reviews based on review ratings and the sentences with high sentiment scores were selected to construct a dataset. For labeling aspect terms in sentences, they used nouns and noun phrases which appeared less than 30 times in the dataset. Then, they selected the aspects from noun and nouns phrases candidates by using Senticnet sentiment lexicon and syntactic rules. The resulting automatically labeled dataset was used to train a model for aspect term extraction with distant supervision. They trained multiple classifiers by using their dataset. They used the SemEval-2014 train data as the validation data and used SemEval-2014 test data as the test data. They presented that B-LSTM followed by CRF classifier trained on their constructed dataset achieved 50.33, 40.49 and 44.87 for precision, recall and F1-score, respectively and outperformed the supervised baseline method of the SemEval-2014 ABSA contest.

Hadano et al. acquired new training data from non-tagged data for aspects identification of sentiment sentences by using clustering approach which was based on cosine similarity measure between the vector of the sentence and vector of the cluster centroid [16]. They assumed that sentences sharing similarities also share common aspects. Consequently, they employed a clustering technique to group these akin sentences together. For clustering implementation, they employed Bayon, a straightforward and expeditious hard-clustering tool. Bayon's clustering methodology relies on the iterative process of repeated bisection. Specifically, the data is divided into two distinct clusters as the first step. As the second step, the cluster displaying the lowest similarity between its centroid and each individual element within it is identified. As the third step, two elements from that cluster are selected randomly. As the fourth step, based on the similarity between them, all elements in the cluster identified in second step are subdivided into two separate clusters. As the fifth step, if swapping elements between these clusters enhances their similarity, swapping is done. Finally, iteration is repeated from the second step to the fifth step. In their work, the aspect of sentences in clusters are determined by an annotator as the representative sentences. At first, the number of annotated sentences was the number of cluster. Based on the assumption that the aspects of sentences belonging to each cluster are equal to the aspects of the representative sentences of each cluster, their system acquired sentences with high similarity as new training by applying different similarity range to increases the number of annotated sentences. They used Support Vector Machine (SVM) as the classifier and they evaluated their system by using game review and the data set for evaluation was constructed by one annotator by using annotation tool. They presented that their proposed method with similarity range "0.7-0.9" which produced 219 sentences achieved the best performance with the accuracy of 73.97%.

In [16], Hadano et al. acquired new training data for aspect extraction based on clustering of sentences, as they assumed that similar sentences shared common aspects. Our method shares the same idea as this method, that is, both methods construct a dataset by clustering of review sentences. However, while Hadano's method is semi-supervised, where a human annotator determines the aspects of the sentences, our method is unsupervised, in the sense that it requires no human intervention.

2.3 Weakly Supervised Learning

This section firstly describes the types of weak supervision, then explain the difference of our method with other weakly-supervised learning methods.

Typically, there are three types of weak supervision; incomplete supervision, inexact supervision and inaccurate supervision [81].

- Incomplete supervision is a kind of supervision where a small amount of labeled data, which is insufficient to train a good learner, are given, while there are abundant unlabelled data. There are two main approaches for incomplete supervision; active learning [58] and semi-supervised learning [8, 82]. Active learning assumes that the ground-truth labels of unlabeled instances can be queried from an oracle such as a human expert. In semi-supervised learning, two fundamental assumptions guide the approach to utilizing labeled and unlabeled data: the cluster assumption and the manifold assumption. The cluster assumption posits that data naturally form clusters, where instances within the same cluster share the same class label. On the other hand, the manifold assumption suggests that data points reside on a manifold, implying that nearby instances should yield similar predictions. Both assumptions revolve around the idea that similar data points should produce similar outputs, and the inclusion of unlabeled data aids in uncovering relationships between similar data points.

- Inexact supervision in machine learning refers to situations where the labeled data provided for training a model is not as precise or accurate as desired. This can result from noisy labels, partial labels, ambiguous labels, or uncertain annotations.

- Inaccurate supervision pertains to scenarios where the provided label information for training a model may not consistently represent the ground truth; in essence, some labels in the training dataset may be erroneous. Several studies focused on how to correct errors in a noisy dataset.

Our weakly-supervised learning method falls under the category of Incomplete Supervision. Specifically, labeled and unlabeled data are utilized based on the cluster assumption. To the best of our knowledge, our method is the first attempt of weakly supervised method for implicit aspect classification task.

One of the common approaches to tackle the data sparseness problem is semi-supervised learning. In the survey by Zhou [81], semi-supervised learning is categorized as Incomplete supervision, i.e., a kind of weakly-supervised learning.

In contrast to the conventional semi-supervised approach of combining a small amount of labeled data with a larger unlabeled set, our research adopts another weakly supervised strategy. This involves relying exclusively on a substantial unlabeled dataset without any labeled examples. Unlike traditional methods that strategically blend labeled and unlabeled data, our emphasis is on addressing the challenges posed by the lack of labeled data by fully leveraging the extensive unlabeled dataset. Through our weakly supervised approach, we aim to potentially reduce the resources required for obtaining and annotating labeled samples, presenting a distinctive perspective on mitigating dataset scarcity for implicit aspect extraction.

Chapter 3

Proposed Method

3.1 Overview

In this study, the task of implicit aspect extraction is defined as a classification problem. For a given review sentence, our system chooses a category of the implicit aspect of which the sentence implicitly expresses the reviewer's opinion. It is supposed that those categories of the implicit aspects are pre-defined. For example, the sentence "I get tired of often recharging it." is classified into the implicit aspect category "battery."

In our work, we focus on the identification of implicit aspects for two types of products: mobile phones (or "phones" in short) and personal computers (PCs). For each, six or five categories of implicit aspects are defined, as in Table 3.1. The category "interface" for PCs includes any devices for the human-machine interface, such as a keyboard, track pad, mouse, and so on. Table 3.2 and Table 3.3 shows the example sample sentences for each implicit aspect for mobile phone and PC, respectively.

Table 3.1: Definition of the categories of implicit aspects

Product type	**Category of implicit aspect**
mobile phone	Battery, Case, Look, Price, Screen, Size
PC	Interface, OS, Price, Screen, Software

Table 3.2: Examples of implicit sentences for mobile phone

Implicit Sentence	**Aspect**
it get full charge but lose the power in minutes .	battery
it 's durable and has a strong grip on the phone .	case
aesthetically , he loved it so much we have purchased several .	look
it is so inexpensive and very good value for money .	price
it did n't feel super secure but has n't fallen off my daughter 's phone yet .	screen
this adds a little bulkiness to the phone for people who carry it in their pocket like i do .	size

Table 3.3: Examples of implicit sentences for PC

Implicit Sentence	**Aspect**
:) the trackpad feels good under the finger , and there 's a scroll lever situated between the two buttons . interface	interface
i had not used xp before , and find the improvement over windows98 to be analogous to the improvements of windows98 over windows95 .	OS
would cost more to send it back than i paid , so i suppose we 'll all just have to deal with it until i am in the market for something more reliable .	price
my lcd display went bad after a month of use , which makes me wonder how long this notebook will last past its warranty period .	screen
because of the way the recovery media is designed i could not install any of the sony provided applications onto the newly loaded system .	software

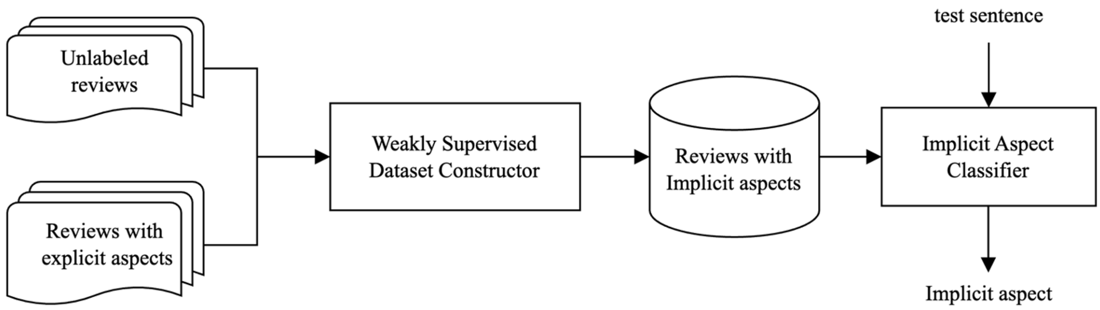

Figure 3.1: Overview of proposed method

Figure 3.1 shows an overview of our proposed method. Since no large-scale dataset labeled with implicit aspects is publically available, it is automatically constructed. Specifically, from a large number of unlabeled reviews and a public dataset of reviews labeled with explicit aspects, the "weakly supervised dataset constructor" module automatically extracts review sentences with implicit aspects to form a dataset labeled with the implicit aspects in a weakly supervised way. This module is essential in this study; the details are described in section 3.2. Next, a classifier of implicit aspect is trained from the obtained dataset. The details of these procedures are described in section 3.3. In the inference, an implicit aspect of a test sentence is identified by the trained classifier.

3.2 Construction of Dataset Annotated with Implicit Aspects

Figure 3.2 shows how the dataset annotated with implicit aspects is constructed. Amazon reviews are used as unlabeled reviews, while the SemEval-2014 dataset is used as a set of reviews with explicit aspects. The dataset is constructed in four steps. First, explicit aspects are extracted from the Amazon reviews using an aspect extraction model trained from the SemEval-2014 dataset [49]. Second, a clustering of the sentences of the reviews is performed, where sentences that mention the same aspect, regardless of whether it is implicit or explicit, are intended to be merged into a cluster. Third, a label is determined for each cluster: it is the aspect of the reviews in that cluster. Finally, the sentences labeled with implicit aspects are retrieved to form the dataset. The details of these steps are presented in the following subsections.

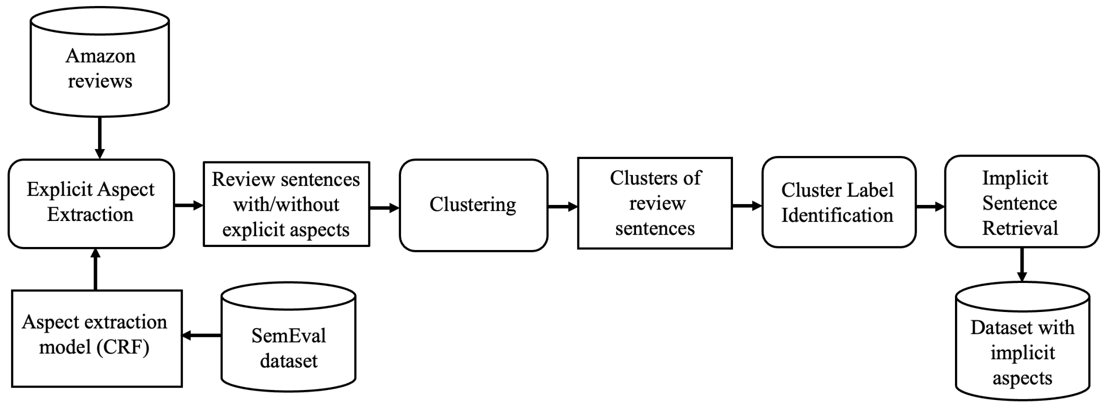

Figure 3.2: Flowchart of constructing corpus annotated with implicit aspects

3.2.1 Explicit Aspect Extraction

The goal of explicit aspect extraction is to extract, from unlabeled Amazon reviews, words and phrases that explicitly represent an aspect of a product. To achieve this, a model to extract explicit aspects is obtained by supervised learning. The dataset of SemEval-2014 Task 4 ABSA is used as the training data. Although there are four subtasks in Task 4 ABSA, subtask 1, "aspect term extraction," is the most appropriate for this study. All aspect terms in the review sentences are marked up in the dataset. The task organizers provide two domain-specific datasets: one for laptops, and the other for restaurants. Each consists of around 3,000 reviews. This is one of the largest publicly available datasets for ABSA. In the present study, the laptop dataset is used for extracting explicit aspects for the PC domain. It is also used for the mobile phone domain, since there is no available dataset of ABSA for the phone domain. Note that the disagreement of the domains, between the training and test data, may decrease the performance of the explicit aspect extraction. A possible solution is to apply a domain adaptation technique that enables us to train an accurate classification model from training data of a different domain.

Conditional Random Field

Explicit aspects are extracted by applying Conditional Random Field (CRF) [33, 32, 26, 62]. It is known that CRF performs relatively well even when it is trained on a small training data and it showed comparatively good results for the task of aspect extraction from reviews in SemEval shared task related to ABSA, where two best results were based on CRF [49]. Moreover, CRF has two main advantages. The first one is independence of assumptions of the observed varibles. The second one is that CRF avoids the label bias problem. CRF has a single exponential model for the joint probability of the entire sequence of labels given the observation sequence. Therefore, even if some data is missing, the observation sequence can still be labeled with less number of features [54].

CRF is a type of discriminative undirected probalistic graphical model. It is used to encode known relationship between observations and construct consistent interpretations. Let $X = (x_1, ...x_n)$ be the sequence of observed data. Let $Y = (y_1, ..., y_n)$ be the sequence of random variables associated with the vertices of the graph. CRF models a conditional probablility $P(Y|X)$ over hidden sequence Y given observation sequence X. That is the conditional model is trained to label an unknown observation sequence X by selecting the hidden sequence Y which maximizes $p(Y|X)$. Then the conditional distribution $p(Y|X)$ can be formalized as follows:

$$P(Y|X) = \frac{1}{Z(X)} \exp\left(\sum_{c \epsilon C} \lambda_c f_c(y_c, X)\right), \tag{3.1}$$

where C is all set of all graphs cliques, f_c is set of all features, λ_c is its corresponding weight. $Z(x)$ is a normalization function as follows:

$$Z(X) = \sum_{y} \exp\left(\sum_{c \epsilon C} \lambda_c f_c(y_c, X)\right). \tag{3.2}$$

Labelling and Features

CRF accepts a review sentence (sequence of words) as input and identifies a label for each word as output. Job and Gurevych [27] represented the possible labels following the Inside-Outside-Begin (IOB) labelling schema, where B, I, and O stand for beginning of an aspect word or phrase, continuation (inside) of an aspect word or phrase and other words which are not aspect tokens, respectively. As we applied sequential labelling, each word in the sentences in SemEval-2014 laptop dataset is assigned with one of the IOB labels. Table 3.4 shows IOB format of an example sentence "The phone would keep making the charging notification sound which caused the screen to also turn on thus killing the battery ." The explicit aspects in this sentence are "charging notification sound", "screen" and "battery".

Table 3.4: Example of an extracted sentence labeled with IOB format

Word	**POS tag**	**IOB tag**
The	DT	O
phone	NN	O
would	MD	O
keep	VB	O
making	VBG	O
the	DT	O
charging	VBG	B
notification	NN	I
sound	NN	I
which	WDT	O
caused	VBD	O
the	DT	O
screen	NN	B
to	TO	O
also	RB	O
trun	VB	O
on	IN	O
thus	RB	O
killing	VBG	O
the	DT	O
battery	NN	B
.	.	O

We used the sklearn-crfsuite library with the default settings to train CRF model. The features used for training the CRF model are in the follwoing list. Those features are extracted from the previous, current, and succeeding words. POS tagging of each word in the review sentences was done by Natural Language Toolkit (NLTK)[1].

- word in lower case
- part-of-speech (POS) of a word
- last two characters of POS of a word
- flag whether all characters are upper case
- flag whether the initial character is upper case
- flag whether it is a digit
- last two characters (of the current word only)
- last three characters (of the current word only)

Extracting Amazon Review Sentences with Aspect Information

By applying the trained CRF model, review sentences annotated with explicit aspects are extracted from Amazon unlabeled review sentences. In addition, review sentences from which no explicit aspect is extracted are also retrieved. These sentences might include no aspect, but sometimes could include an implicit aspect. In other words, the sentences without the explicit aspects can be the potential sentences including the implicit aspects. As a result, a set of review sentences with and without explicit aspects is obtained at the explicit aspect extraction step. Table 3.5 shows examples of extracted sentences.[2] The second and third sentence contain the explicit aspect of “screen” and “price” respectively, while other sentences include no explicit aspect.

[1]https://www.nltk.org/

[2]In this thesis, outputs of the tokenizer are shown as example sentences. For example, “does’t” is split into two tokens “does” and “n’t”.

Table 3.5: Example of sentences obtained by explicit aspect extraction

Review sentence	**Aspect**
It does n't click with the white piece at all, and it easily slides off	*none*
I wanted so much to keep this case on, but I also did n't wan na risk my phone having a giant crack on the screen due to a case that does n't stay on	screen
Or the price, it is neat, but I really doubt I 'm going to keep it on my phone	price
What a bummer	*none*
It is cute and light weight	*none*

3.2.2 Clustering of Sentences

The review sentences, either labeled with explicit aspects or unlabeled, were then merged into clusters. The goal of this clustering is to make clusters of sentences that express opinions about the same aspect.

Each review sentence is converted to Sparse Composite Document Vectors (SCDVs) [44]. Sparse Composite Document Vectors (SCDVs) emerge as a sophisticated and highly effective method for encapsulating the rich semantics embedded in review sentences, surpassing the simplicity of traditional word embedding averages. Notable for their excellence in document representation, SCDVs achieve a sparse yet information-rich encoding through a meticulous fusion of global and local TF-IDF weighting. This unique characteristic not only enhances computational efficiency but also captures the intricate nuances of language composition. Significantly, SCDVs excel in capturing semantic similarity, particularly in discerning implicit sentiments within review sentences. This capability positions SCDVs as a potent tool for extracting the subtle meanings embedded in textual data, contributing to a deeper understanding of reviews. In the broader context of document analysis, SCDVs offer a versatile and robust approach, promising to advance the precision and depth of document representation across various applications within the field of natural language processing, thereby enriching the landscape of computational linguistics and information retrieval.

We chose k-means as the clustering algorithm, since it is an efficient method [19]. The algorithm follows an iterative process that converges towards an optimal solution, aiming to minimize the intra-cluster variance and maximize inter-cluster separation. The fundamental idea behind k-means is to assign each data point to the cluster whose centroid is the nearest, based on a chosen distance metric, typically Euclidean distance.

The algorithm initiates by randomly selecting k initial centroids, where k represents the predetermined number of clusters. Subsequently, data points are assigned to the cluster whose centroid is closest, and the centroids are recalculated based on the mean of the assigned data points. This assignment and centroid update process iterates until convergence, signifying stability in cluster assignments.

In our application of k-means to sentence clustering, the vectors representing sentences in a high-dimensional space are analogous to data points. The Euclidean distance, as the chosen metric, quantifies the dissimilarity between sentence vectors. This distance-based approach is conducive to grouping similar sentences into clusters, making k-means particularly suitable for tasks such as sentiment analysis and review clustering.

One critical consideration in implementing k-means is the determination of the optimal number of clusters, denoted by k. In this study, it is not preferable to merge the sentences referring to different aspects into one cluster. In other words, the purity of the clusters should be high. Therefore, we set the number of clusters to a relatively large number so that we could create many small but accurate clusters of review sentences. Specifically, the parameter k is set to 10 % of the total number of review sentences.

In summary, the application of k-means in our proposed methodology is rooted in the foundational principles of unsupervised learning, iterative optimization, and distance-based clustering. This theoretical foundation serves as the basis for the subsequent application of the algorithm to review sentence clustering, aligning with our goal of extracting meaningful insights from diverse aspects of the review data.

1) For the prices, was n't worth sending back & is really for those few times away from home or do n't have outlet handy & the battery gets really low anyway . (price, battery)

2) Like it but it causes the battery to get really hot and lock the phone . (battery)

3) I really like the design, but however the casing did not snap nicely with my phone in place . (design)

4) took a while to get to me its really cute just hard to come off which is good and bad i guess good because its secure if you drop the phone and bad because you may have to use something to get it open to clean or switch cases in any event i like its hard rubber and design . (hard rubber, design)

5) I would have given it one star since it really does n't hold a charge or even charge for that matter , but I decided to add another for the design of the case although the kickstand is extremely flimsy and half of the time wo n't even hold up my phone . (*none*)

Figure 3.3: Example of cluster

Figure 3.3 shows an example of a constructed cluster. It consists of five review sentences. In general, a cluster contains two kinds of sentences. One is a sentence containing one or more explicit aspects, such as the sentences 1)–4) in Figure 3.3, where the explicit aspects extracted by the CRF model are indicated by being in parentheses. The other is a sentence that does not contain explicit aspects, such as the sentence 5). The explicit aspects that do not correspond to any pre-defined aspect categories (and their synonyms that will be explained in subsection 3.2.4), such as "hard rubber", are ignored.

3.2.3 Cluster Label Identification

The task of cluster label identification involves choosing the most significant aspect for a sentence cluster. This is not always obvious, since there are two or more explicit aspects in the cluster as shown in Figure 3.3. Algorithm 1 shows pseudocode for this process.

Algorithm 1: Algorithm for cluster label identification

Input : Cluster of review sentences

Output: Cluster label

Let s_i be a sentence in the cluster, and a_i be the explicit aspect of s_i;
Let $Fre(a_i)$ be the frequency of a_i in the cluster;
Let $Oc(a_i)$ be the number of occurrence of a_i in the set of sentences $\{s_i\}$;
label $\leftarrow$ aspect with the maximum $Fre(a_i)$;
if *label is unique* **then**
    return `ReliabilityCheck(`*label*`)`
else
    Let $\{a'_i\}$ be the set of aspects with maximum $Fre(a_i)$;
    label $\leftarrow$ aspect with the maximum $Oc(a'_i)$;
    if *label is unique* **then**
        return `ReliabilityCheck(`*label*`)`
    else
        return *INDETERMINABLE*
    end
end
def `ReliabilityCheck(`*label*`)`
    if $Rel(label) \geq T_r$ **then**
        return *label*
    else
        return *INDETERMINABLE*
    end
end

The basic idea of Algorithm 1 is that the most frequent aspect is chosen as the cluster label. Two kinds of frequency, $Fre(a_i)$ and $Oc(a_i)$, are considered. $Fre(a_i)$ is the number of the times an aspect a_i is extracted, while $Oc(a_i)$ is the number of occurrences of the aspect in the review sentences. The explicit aspects in a cluster are compared with respect to Fre and Oc in this order, then the most frequent one is chosen as the cluster label (lines 4–11). If two or more aspects have the same Fre and Oc, the label is defined as *INDETERMINABLE* (line 13), which indicates that the cluster may be wrongly made up of sentences about different aspects.

Next, we measure the reliability of the label, which is defined as the proportion of the sentences with the explicit aspect to all sentences in the cluster, as shown in Equation (3.3).

$$Rel(aspect) = \frac{Fre(aspect)}{\text{number of sentences in the cluster}} \tag{3.3}$$

If the chosen aspect is reliable enough (the reliability is greater than or equal to a threshold T_r), it is chosen as the cluster label; otherwise, the label is defined as *INDETERMINABLE* (lines 16–22). The threshold T_r was empirically determined for each aspect category in the experiment.

Let us explain how the label of the example cluster in Figure 3.3 is identified. Fre(design)=2 since the explicit aspect "design" is extracted twice, and Oc(design)=3 since the word "design" appears three times in the sentences. Similarly, Fre(battery)=2 and Oc(battery)=2. Since Fre(design) and Fre(battery) are the same, Oc is compared. Then "design" is chosen because Oc(design) > Oc(battery). Finally, the reliability is assigned the measure Rel(design)=2/5. If it is higher than T_r, "design" is chosen as the label of the cluster.

3.2.4 Implicit Sentence Retrieval

The last step is to collect the sentences containing implicit aspects. As explained in section 3.1, six implicit aspects for the phone domain and five implicit aspects for the PC domain have been defined. For each implicit aspect, the cluster whose label coincides with it is chosen. To get more clusters, a list of synonyms of the implicit aspects is created manually, and clusters for which the label is a synonym are also chosen. For example, synonyms of the aspect "battery" are "battery case," "battery life," "power," and so on. Table 3.6 shows examples of the synonyms. The full list of the synonyms are shown in the second column of Tables A.1 and A.2 in Appendix A. No synonym is used for "price" and "size" of the phone domain and "price" of the PC domain. Note that the cost of constructing the list of the synonyms is much less than the manual annotation of many sentences with implicit aspect labels.

Table 3.6: Examples of synonyms.

(**a**) Phone domain

Aspect	**Synonym**
Battery	battery case, battery life, power
Case	case quality, case cover
Look	design, color
Price	—
Screen	screen protector, screen cover
Size	—

(**b**) PC domain

Aspect	**Synonym**
Interface	keyboard, touchpad
OS	windows, windows xp
Price	—
Screen	monitor, screen size
Software	program, applications

Sentences for which no explicit aspect has been extracted are then retrieved from the chosen clusters. A cluster label is attached to these retrieved sentences as their implicit aspects. In the example in Figure 3.3, sentence (5) is retrieved with the label "look" as its implicit aspect, since the cluster label is "design", which is a synonym of "look".

Recall that the number of the clusters in k-means is set to a large value (10%

of total sentences). Our motivation for this is to avoid making clusters containing multiple aspects, since they cause errors in the process of the cluster label identification and retrieval of implicit sentences. Although sentences with the same aspect may be scattered to different clusters, it might not be a problem because we can retrieve sentences with the implicit aspect from each cluster.

3.3 Implicit Aspects Classification by BERT

A classifier for implicit aspect identification is trained using the constructed dataset. Bidirectional Encoder Representations from Transformers (BERT) [11] is chosen as our classification model, since it has achieved outstanding performance for many NLP tasks. The bert-base-uncased[3] is chosen as the pre-trained BERT model. Then it is fine-tuned using the dataset of implicit aspect sentences we constructed.

In addition, the SemEval-2014 dataset is also used for fine-tuning. Although it contains not implicit but explicit aspects, the linguistic expressions in sentences with explicit aspects may be similar to those of the implicit aspects. Thus the explicit aspect sentences can also be used for fine-tuning BERT, resulting in an increase in the number of the training samples. To form the training data for the implicit aspect classifier, the review sentences with an explicit aspect that agrees with one of the pre-defined implicit aspect categories are excerpted. Similar to the retrieval of implicit aspect sentences described in subsection 3.2.4, soft-matching using a list of synonyms of the implicit aspect is applied to retrieve more sentences. The synonyms for each implicit aspect category are manually and exhaustively excerpted from the SemEval-2014 dataset. The complete list of the synonyms is shown in the third column in Tables A.1 and A.2 in Appendix A. In addition, a review sentence is not included in the training data when it contains two or more implicit aspect categories. For example, let us consider the sentence "this laptop is a great price and has a sleek look." Since it contains two aspects, "price" and "look," it is not included in the training data for the phone domain. Note that it is included in the training data for the PC domain, since only "price" is an

[3] `https://huggingface.co/bert-base-uncased`

implicit aspect category for this domain.

In summary, the BERT models of implicit aspect classification are trained from two different datasets: one is the dataset of the implicit aspects created by our proposed method, the other is the union of this dataset and another corpus extracted from the SemEval-2014 dataset.

3.4 Summary

In this chapter, we presented the overview of our proposed method, including the overview architecture of our method. Then, the flowchart describing the main steps in our approach was described and each step was explained in detail with some examples. We presented our proposed approach by dividing into two main parts; construction of the dataset annotated with implicit aspects and classification of the implicit aspects for unknown reviews. The first part consisted of four steps; explicit aspect extraction, clustering of review sentences, cluster label identification and implicit sentence retrieval. The second part was the proposal of training implicit aspect classification using BERT model, where both the proposed dataset labeled with implicit aspects and the existing dataset labeled with explicit aspects.

Chapter 4

Evaluation

This chapter reports the results of experiments to evaluate our proposed method. First, in section 4.1, the quality and quantity of the dataset annotated with implicit aspects described in section 3.2 is assessed. Next, in section 4.2, the performance of the classification of the implicit aspect by the method described in section 3.3 is evaluated.

4.1 Evaluation of Dataset Annotated with Implicit Aspects

4.1.1 Experimental Setup of the Construction of the Dataset

Amazon product data [21] was used to construct the dataset. We excerpted 30,000 review sentences from the category entitled "Cell Phones" and "Accessories" for the phone domain, and 10,000 review sentences from the category entitled "computers" for the PC domain.

The CRF model for explicit aspect extraction was trained first. The laptop reviews from the SemEval-2014 dataset were used for training the CRF model for both the phone domain and the PC domain. A preliminary evaluation of the performance of the CRF model was carried out on the SemEval-2014 dataset, where 90% of the datasets was used for training and 10% for testing. The precision, recall, and F1-score for explicit aspect extraction by CRF were shown in Table 4.1. They were sufficiently high for the subsequent procedures.

Table 4.1: Results of explicit aspect extraction by CRF

Precision	**Recall**	**F1-score**
0.77	0.64	0.70

For each aspect category, we randomly chose 50 sentences associated with the target implicit aspect (or all of them when the number of such sentences was less than 50). The chosen sentences were then manually judged so as to determine whether they expressed users' opinions about the implicit aspect. In our experiment, a human evaluator judged the sentence without referring its surrounding context. As an evaluation criterion, we used the accuracy, defined as the ratio of the correct review sentences containing implicit aspects to the total number of manually checked sentences.

T_r is the parameter used in Algorithm 1, where the cluster label is judged as *INDETERMINABLE* when the ratio of the majority aspect in the cluster is lower than T_r. If we set T_r higher, the number of the retrieved implicit sentences will be reduced, but the accuracy will be improved. In this experiment, T_r was initially set to 0.1. For some implicit aspect categories, we set T_r higher when the accuracy was relatively low. We argue that it is not necessary to optimize T_r using validation data. Once T_r is set so that the accuracy is high, we can easily increase the number of sentences with implicit aspects by using more unlabeled review sentences. However, empirical investigation of how the increase in unlabeled sentences can contribute to enlarge the dataset and to improve the performance of the implicit aspect classification should be carried out in the future.

4.1.2 Results of Constructed Dataset

Tables 4.2 and 4.3 show the statistics of the constructed dataset as well as the accuracy and T_r for the phone and PC domains, respectively. The third column shows the average size of the clusters and the standard deviation in the form of ave±sd. As for the phone domain, we obtained 290 clusters for six aspects, and the average size of the clusters (the number of sentences per cluster) was between 8 and 12. Recall that each cluster consists of sentences with both explicit and implicit aspects; the numbers of each are shown in the fourth and fifth columns, respectively. As a result, from 90 to 393 implicit sentences were obtained for six aspect categories. As for the PC domain, 149 clusters were obtained in total. The number of the obtained implicit sentences was between 45 and 261 for five aspect categories.

Next, we discuss the accuracy of the obtained sentences including implicit aspects. The accuracy was 0.58 or more for all aspect categories in the phone domain. The threshold T_r was set higher for the "screen" and "price" categories to improve the accuracy. As for the PC domain, the accuracy of the five aspect categories was between 0.56 and 0.72. T_r was set to 0.3 for the "price" category, but 0.1 (the default value) for the others.

Figure 4.1 shows examples of sentences with implicit aspects. The labels for the clusters are (price) and (look), and the check marks indicate the obtained implicit sentences, in which the cluster label (price or look) is annotated as the implicit aspect. Sentence (4) in cluster 1 was successfully annotated with the aspect "price", although the word "price" was not explicitly used. The sentences with check marks in cluster 2 are other good examples of the implicit aspect of "look". Note that the cluster label was identified as "look" since the majority of the explicit aspects in this cluster were identified as "design", which is a synonym for the aspect category of "look". In summary, the results indicate that our proposed method is promising in terms of automatically constructing a dataset annotated with implicit aspects.

Table 4.2: Details of constructed dataset with implicit aspects (phone domain)

Aspect	**# of clusters**	**Average size of cluster**	**# of Explicit Sentence**	**# of Implicit Sentence**	**Accuracy**	**T$_r$**
Battery	58	12±16	274	393	0.82	0.1
Case	23	8.6±5.5	104	94	0.74	0.1
Look	62	11±9.2	303	353	0.58	0.1
Price	89	11±10	751	252	0.78	0.4
Screen	31	8.7±11	179	90	0.76	0.2
Size	27	8.4±6.2	121	106	0.70	0.1

"#" means the number of clusters or sentences.

Table 4.3: Details of constructed dataset with implicit aspects (PC domain)

Aspect	**# of clusters**	**Average size of cluster**	**# of Explicit Sentence**	**# of Implicit Sentence**	**Accuracy**	**T$_r$**
Interface	24	9.0±5.4	117	100	0.62	0.1
OS	25	12±9.0	145	163	0.72	0.1
Price	15	8.6±5.8	84	45	0.56	0.3
Screen	44	11±7.5	228	261	0.70	0.1
Software	41	10±7.0	147	250	0.64	0.1

"#" means the number of clusters or sentences.

Cluster 1: (price)

1) Great price, great service from the vendor . (price, service)
2) Cheap price for a good quality made item . (price, quality)
3) Very pleased with this item and it was an exellent price ! (price)
4) This was such a nice small and cheap item , I had to order 2 of them , just to have one in each car . (*none*) ✓
5) for its price, it's not too bad, with a beautiful design (price, design)
6) good item for great price . (price)

Cluster 2: (look)

1) The color options are awesome and its very portable . (color options)
2) Very vivid colors and the car charger is an awesome bonus . (car charger)
3) The design is amazing and the lettering is a little light but that does n't matter as long as it fit and you are satisfied with your purchase , because I was ! (design)
4) The design was ok for a cheap case , but it was not the color it should have been ! ! ! ! (design)
5) This case is beautiful and vibrant in color , it has somewhat of a grip so it does n't slip out of your hands easily . (*none*) ✓
6) I 've always had plain solid colors , but when I saw this I thought it would look nice . (*none*) ✓
7) ONLY THING NICE ABOUT THIS ITEM IS THE ARRAY OF COLORS . (*none*) ✓
8) A great buy as it does not slip out of your hand and has an awesome vivid design . (*none*) ✓
9) Nice design and color . (Nice design)
10) like the design and color . (design)
11) I love the leopard design and colors defiantly makes my phone unique ! (leopard design)
12) I love the design and colors . (design)
13) The colors are vibrant , the design is unique , and the case snaps together easily and is actually hard to pry back off (I tried ! (design)
14) I do like this Owl & case and the colors and the design is great also . (design)

Figure 4.1: Example of sentences labeled with implicit aspects

4.1.3 Error Analysis

We found some major causes of error in the process of implicit aspect identification for the phone domain. When we initially set the threshold T_r to 0.1, numerous errors were found in the extraction of the implicit aspect "price." This was because "price" is a rather general concept, and frequently occurred with other aspects, such as "service," "battery," "case," or "look." For example, in Cluster 1 in Figure 4.1, sentences 1), 2) and 5) include "price" with other aspects. In this example of a cluster, sentence 4) was correctly extracted as a sentence with this implicit aspect, but many sentences in other clusters were wrongly extracted. However, by setting T_r to 0.4, the accuracy was improved to 0.78, although this was offset by a decrease in the number of extracted sentences.

"Screen" was another implicit aspect for which we found many errors. Even when sentences contained the explicit aspect "screen," they often mentioned not the screen itself but other related concepts, such as notifications or information shown on the phone screen. However, by changing T_r to 0.2, the accuracy was improved to 0.76. In addition, errors were caused by ambiguity in the meanings of words. For example, the word "look" was used both to represent the design of the mobile phone and as a verb that was almost equivalent to "seem" (as in the phrase "looks like ..."). Another problem was ambiguity in the aspect itself; for example, the word "cover" was ambiguous, and could have meant "phone cover" or "screen cover."

We also found some major causes of errors in the process of implicit aspect identification for the PC domain. Five percent of the manually assessed sentences labeled with the implicit aspect "software" were written about RAM. RAM might be related to the software since it enables some software and applications to run quickly. However, RAM itself is not software but hardware. Besides, 79% erroneous sentences for the category "interface" mentioned a port such as "USB port" and "serial port".[1] They originated from one cluster that consisted of many sentences including the word "port." The label of this cluster was identified as "interface" since some sentences included both "keyboard" (a synonym of "interface") and "port," e.g., "Because your keyboard itself has 2 USB ports, you can plug your mouse and printer into your keyboard." Accidental co-occurrence of

[1] Note that the implicit aspect "interface" is defined as a man–machine interface such as a keyboard, mouse, and trackpad in this study. The interfaces to connect other devices (e.g. USB port, display port) are not included.

an aspect and another word (such as "keyboard" and "port") could be a cause of the incorrect assignment of the implicit aspect.

4.2 Evaluation of Implicit Aspect Classification

4.2.1 Experimental Setup of Implicit Aspect Classification

The performance of the proposed model for classifying of implicit aspects was evaluated. First, the test data was constructed by the following procedures. As described in subsection 4.1.1, a few sentences in the constructed dataset labeled with implicit aspects were manually evaluated. For each aspect category, 30 (or 10 when the number of extracted implicit sentences is small) sentences were randomly chosen from the sentences that were judged as correct. Thus the test data consisted of genuine sentences with implicit aspects where the distribution of the aspect categories was relatively balanced.

Next, three datasets were constructed as follows.

$\boldsymbol{D_e}$ A set of review sentences with explicit aspects. It was made from the SemEval-2014 dataset as described in section 3.3.

$\boldsymbol{D_i}$ A set of review sentences with implicit aspects. It was constructed by our proposed method as described in section 3.2.

$\boldsymbol{D_{e+i}}$ A set of both sentences with explicit and implicit aspects.

Table 4.4 (a) and (b) show the number of sentences in D_e, D_i, and D_{e+i} as well as the test data for the phone and PC domain, respectively.

Three classifiers were obtained by fine-tuning BERT using these datasets. Hereafter, the models trained from D_e, D_i, and D_{e+i} are denoted by M_e, M_i, and M_{e+i}, respectively. These three models were compared in this experiment, where M_e is the baseline model.

Table 4.4: Statistics of the training and test data.

(**a**) Phone domain

Aspect	**Dataset**			**Test Data**
	$\boldsymbol{D_e}$	$\boldsymbol{D_i}$	$\boldsymbol{D_{e+i}}$	
Battery	106	363	469	30
Case	4	64	68	30
Look	21	323	344	30
Screen	93	80	173	10
Size	21	96	117	10
Price	82	222	304	30
Total	327	1148	1475	140

(**b**) PC domain

Aspect	**Dataset**			**Test Data**
	$\boldsymbol{D_e}$	$\boldsymbol{D_i}$	$\boldsymbol{D_{e+i}}$	
Interface	83	70	153	30
OS	45	133	178	30
Price	83	35	118	10
Screen	88	231	319	30
Software	104	220	324	30
Total	403	689	1092	130

When we fine-tuned BERT, the hyperparameters were optimized on the validation data. Specifically, the dataset in Table 4.4 was randomly split into 90% for the training data and 10% for the validation data. The optimized hyperparameters and their possible values are as follows:

- batch size: {8, 16, 32}
- learning rate: $\{2e^{-5}, 3e^{-5}, 4e^{-5}, 5e^{-5}\}$
- number of epochs: {2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50}

The best set of the hyperparameters was chosen by several criteria on the validation data. More concretely, the criteria were checked in the following order: (1) the highest accuracy, (2) the highest macro average of F1-score for all aspect categories, and (3) the lowest validation loss. The best hyperparameters for each dataset are presented in Table 4.5. The final BERT model was fined-tuned using the overall dataset (both the training and validation data) with the optimized hyperparameters.

Table 4.5: Optimized hyperparameters.

Domain	**Phone**			**PC**		
Dataset	$\boldsymbol{D_e}$	$\boldsymbol{D_i}$	$\boldsymbol{D_{e+i}}$	$\boldsymbol{D_e}$	$\boldsymbol{D_i}$	$\boldsymbol{D_{e+i}}$
Batch size	8	8	8	8	8	8
Learning rate	$3e^{-5}$	$5e^{-5}$	$4e^{-5}$	$4e^{-5}$	$5e^{-5}$	$5e^{-5}$
Number of epochs	20	30	35	5	35	50

4.2.2 Results of Implicit Aspect Classification

Tables 4.6 and 4.7 show the results of the classification of implicit aspects. These tables present the precision (P), recall (R), and F1-score (F) for each aspect category, their macro average, and the accuracy (micro average). The best score among the three models is shown in bold. Graphical representations of precision, recall and F-score of Table 4.6 are shown in Figure 4.2, Figure 4.3 and Figure 4.4, respectively. Similarly, Figure 4.5, Figure 4.6 and Figure 4.7 illustrates the precision, recall and F-score of Table 4.7.

Table 4.6: Results of implicit aspect classification (phone domain).

Aspect		M_e			M_i			M_{e+i}	
	P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
Battery	0.68	0.90	0.77	**0.96**	0.87	0.91	0.91	**0.97**	**0.94**
Case	0.38	0.30	0.33	0.79	**0.50**	**0.61**	**0.82**	0.47	0.60
Look	0.68	0.57	0.62	0.65	0.80	0.72	**0.74**	**0.87**	**0.80**
Price	0.92	0.80	0.86	**0.94**	**0.97**	**0.95**	**0.94**	**0.97**	**0.95**
Screen	0.39	0.70	0.50	**0.89**	**0.80**	**0.84**	0.62	**0.80**	0.70
Size	0.43	0.30	0.35	0.53	**0.90**	0.67	**0.75**	**0.90**	**0.82**
Macro avg.	0.58	0.59	0.57	0.79	0.81	0.78	**0.80**	**0.83**	**0.80**
Accuracy		0.62			0.79			**0.82**	

Table 4.7: Results of implicit aspect classification (PC domain).

Aspect		M_e			M_i			M_{e+i}	
	P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
Interface	0.95	0.60	0.73	**1.00**	0.67	0.80	0.83	**0.83**	**0.83**
OS	0.82	0.30	0.44	0.89	**0.83**	0.86	**0.93**	**0.83**	**0.88**
Price	0.67	0.40	0.50	**0.86**	**0.60**	**0.71**	**0.86**	**0.60**	**0.71**
Screen	0.73	0.80	0.76	0.89	0.80	0.84	**0.90**	**0.90**	**0.90**
Software	0.39	0.80	0.53	0.52	0.83	0.64	**0.72**	**0.87**	**0.79**
Macro avg.	0.71	0.58	0.59	0.83	0.75	0.77	**0.85**	**0.81**	**0.82**
Accuracy		0.61			0.77			**0.84**	

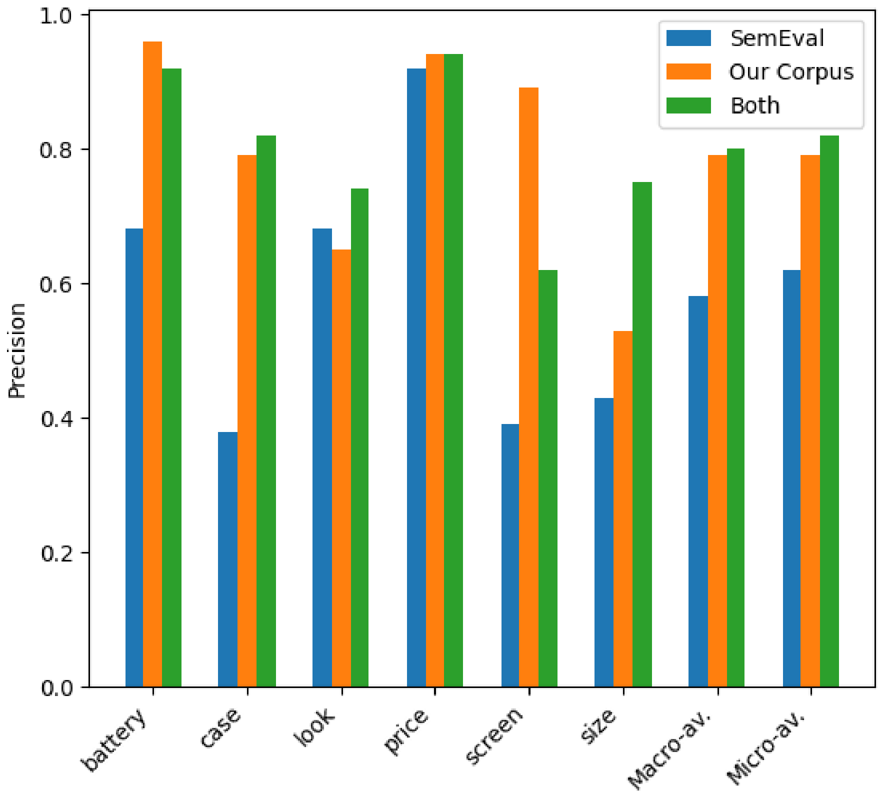

Figure 4.2: Precision of implicit aspect extraction in phone domain

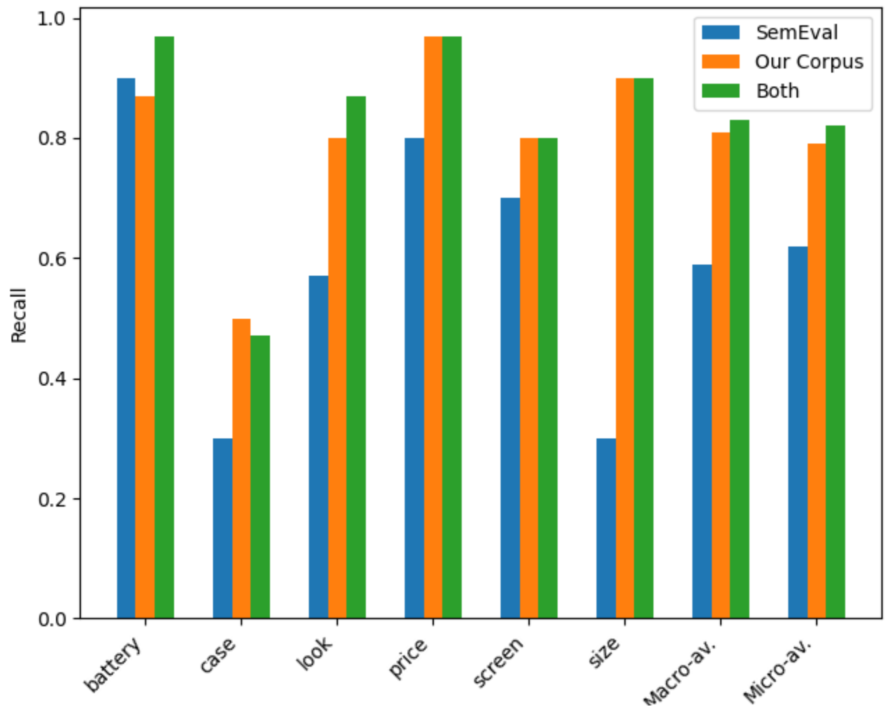

Figure 4.3: Recall of implicit aspect extraction in phone domain

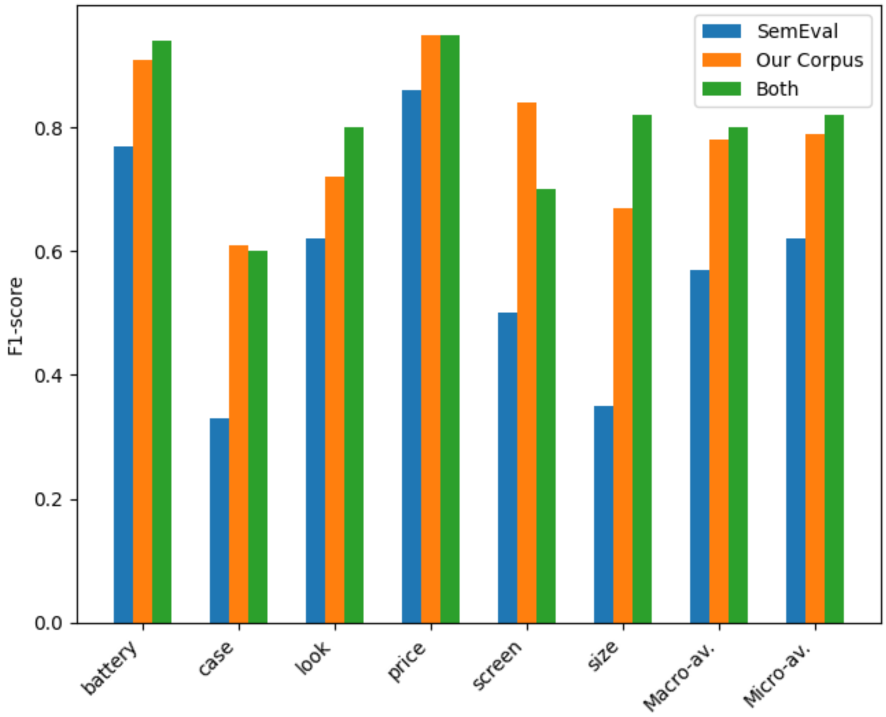

Figure 4.4: F1-score of implicit aspect extraction in phone domain

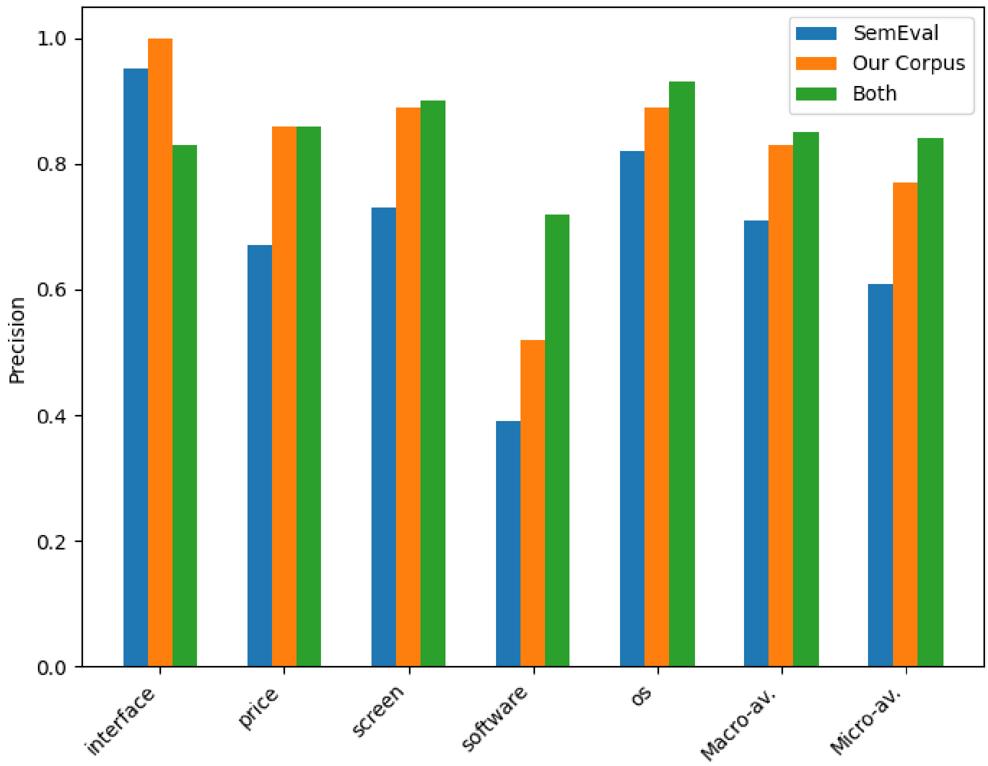

Figure 4.5: Precision of implicit aspect extraction in PC domain

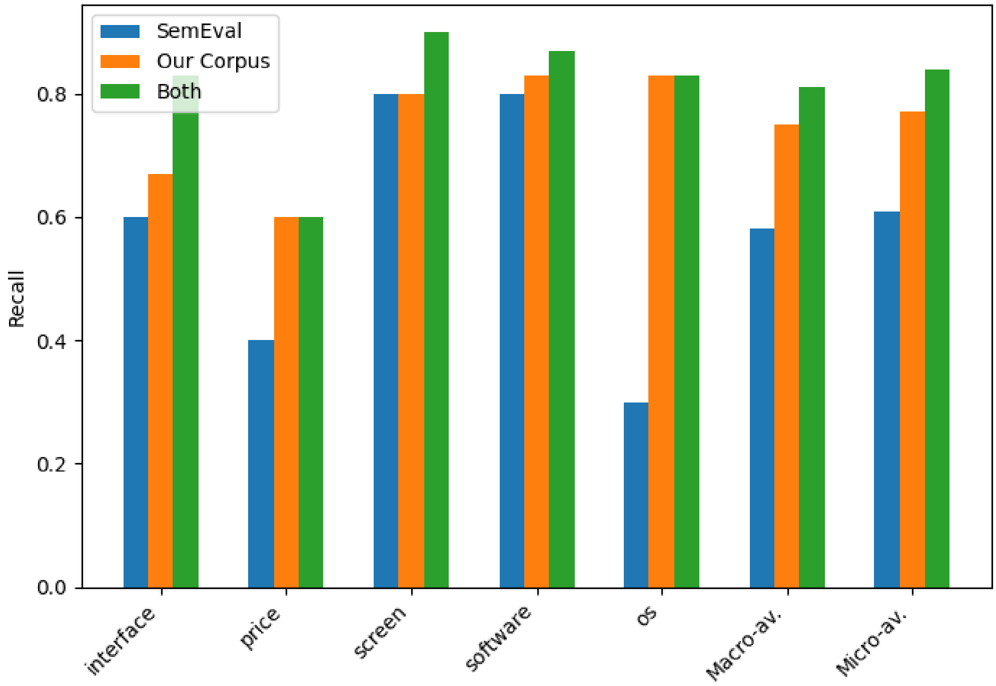

Figure 4.6: Recall of implicit aspect extraction in PC domain

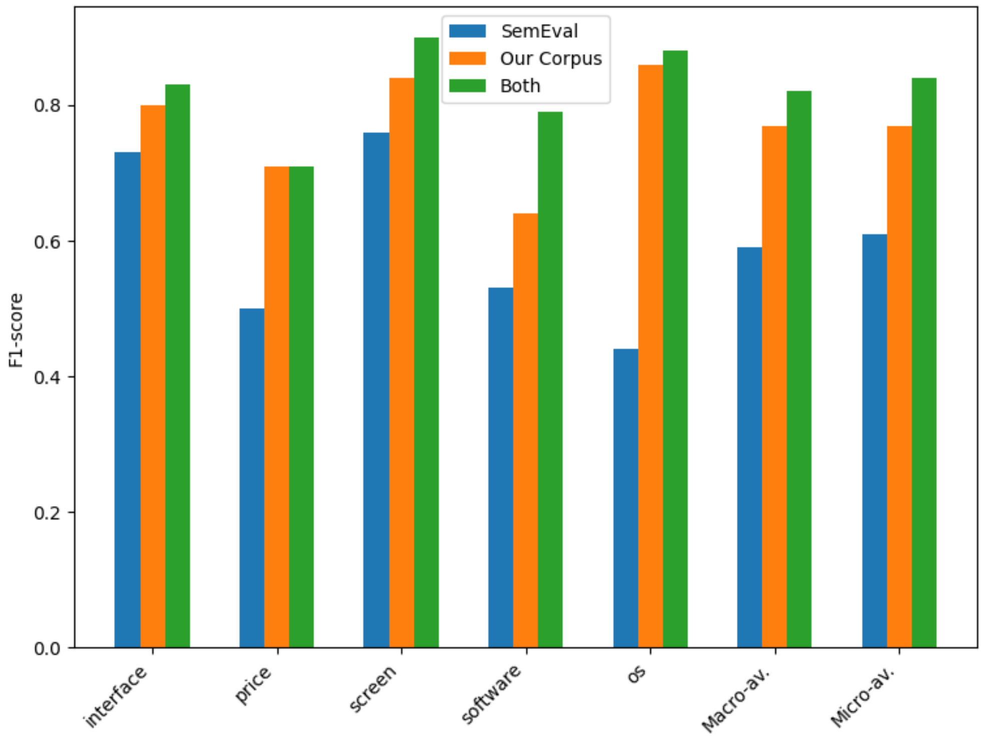

Figure 4.7: F1-score of implicit aspect extraction in PC domain

Our model M_i outperformed the baseline model M_e for all evaluation criteria except for the recall of "battery" for the phone domain, the precision of "look" for the phone domain, and the recall of "screen" for the PC domain. In addition, large differences between M_i and M_e were found. The macro average of the F1-score and the accuracy of M_i for the phone domain was better than those of M_e by 0.21 and 0.17 points, respectively. Similarly, the improvement of 0.18 points of the macro F1 and 0.16 points of the accuracy was found for the PC domain. Therefore, the corpus of sentences with implicit aspects, which was constructed by our proposed method, was an effective training dataset for implicit aspect classification. It seems reasonable, since both the test data and D_i consisted of implicit sentences, while D_e consisted of explicit sentences.

Comparing models M_i and M_{e+i}, it was confirmed that the use of both the sentences with explicit and implicit aspects could further boost the performance of the classification. As for the phone domain, M_{e+i} outperformed M_i by 0.02 points with respect to the macro average of the F1-score and 0.03 points with respect to the accuracy. However, when the McNemar test was performed to check the difference between M_i and M_{e+i}, it was not statistically significant in 95% confidence level. Besides, further improvement was found in the results for the PC domain; M_{e+i} was better than M_i by 0.05 points in the macro F1 and 0.07 points in the accuracy. It was confirmed by the McNemar test that the difference between them was statistically significant in 95% confidence level. Since the reviews in SemEval-2014 dataset are in the laptop domain, adding the explicit sentences to the implicit sentences can remarkably contribute to improve the classification performance in the PC domain, but not in the phone domain. In addition, the F1-score of M_{e+i} was worse than that of M_i in two aspects: "case" and "screen" for the phone domain. Adding the sentences with explicit aspects to the dataset made by the sentences with implicit aspects did not always contribute to improving the performance.

Our best model M_{e+i} was always better than the baseline (M_e) with respect to the F1-score of all aspect categories. As for the macro average of F1-score and the accuracy, improvements by 0.23 and 0.20 points were found. These results prove the effectiveness of our proposed method.

We discuss the reason why M_e, the model trained from the sentences with explicit aspects, performed poorly. One obvious reason is the lack of training samples. For example, there were only four sentences for the aspect "case" in D_e for the phone domain, thus the F1-score of the model M_e was low, 0.33. Another reason may be the disagreement of the domains. Recall that the dataset D_e was extracted from the SemEval 2014 ABSA laptop dataset. When it was applied to the phone domain, the domains of the training and test data were different. We observed that the aspect "screen" in the phone and PC domains referred to different concepts, although it was the same aspect for electronic devices. Customers pay attention to a protector, cover, fingerprint or swipe on the screen in the phone domain, while they focus on resolution or size of a screen in the PC domain. This might be the reason why the F1-score for "screen" in the phone domain was low, viz., 0.50. On the other hand, in our approach, the sentences with an implicit aspect were extracted from unlabeled data of the same domain. Therefore, the problem of the domain-shift can be alleviated. In addition, the increase in the number of the samples in the training data can obviously contribute to improve the performance of the classifier.

4.3 Summary

In this chapter, we have presented the experimental setup and the results of two main tasks of our approach; constructing the corpus annotated with implicit aspects and classification of implicit aspects. We did qualitative and quantitative analysis for our constructed corpus and error analysis was also presented. We presented how we chose the best hyperparameters for fine-tuning BERT classifier for the implicit aspect classification task. We evaluated the performance of implicit classification by our proposed method and baseline method and discussed the reasons of results of our method and baseline. We experimented our proposed method in two domains: mobile phone domain and PC domains. The results have shown that our method was better than the baseline method significantly.

Chapter 5

Conclusion

5.1 Summary of This Study

This paper has proposed a weakly-supervised leaning method to classify a sentence in a customer review into pre-defined categories of implicit aspects. First, a dataset annotated with implicit aspects was automatically constructed. For a given unlabeled dataset consisting of sentences with explicit and implicit aspects, clustering was performed to merge sentences having the same (explicit or implicit) aspect. Then, the explicit aspect was transferred to a sentence within the same cluster as a label of an implicit aspect. Next, the constructed dataset and the existing dataset with explicit aspects were used to fine-tune the BERT model to identify an implicit aspect of a sentence. The results of an experiment on two domains (mobile phones and PCs) showed that our proposed model, trained from the weakly labeled dataset, outperformed the baseline, trained from the sentences with explicit aspects only, by a large margin. An error analysis was also carried out to reveal the major problems in the construction of the implicit-aspect-labeled dataset.

We would like to discuss the difference between our proposed method and other weakly-supervised methods. As discussed in section 2.3, our method was categorized as Incomplete Supervision using a small amount of labeled data and a large amount of unlabeled data, where the gold labels of the unlabeled data were guessed by the cluster assumption. In conventional weakly-supervised approach, a small amount of data annotated with implicit aspects might be used. In our weakly supervised method, any labeled data annotated with implicit aspects was not used. Instead, the sentences with the explicit aspects were used as the initial labeled data, then the explicit aspects were transferred to the labels of implicit aspect sentences. No use of labeled data was our important challenge.

5.2 Answer for Research Question

As described in section 1.3, our major research question is "*How to develop a model for implicit aspect extraction without using a labeled dataset?*". It was decomposed to three sub-research questions RQ1, RQ2, and RQ3. We answer these research questions as follows.

RQ1 *How to automatically construct a corpus annotated with implicit aspects, which is sufficiently large for training a model?*

An automated approach to construct a corpus annotated with implicit aspects was developed by the following procedures. We clustered the labelled review sentences (which were extracted with explicit aspects by CRF model trained with the dataset annotated with explicit aspects) and unlabeled review sentences. Then, a cluster label for each cluster (which was also regarded as an implicit aspect) was automatically identified based on the assumption that the sentences with similar context shared a common aspect. The unlabled sentences in a cluster whose label was coincident with the predefined implicit aspect category were obtained as the implicit-aspect-labeled sentences. The manually designed synonyms were also used to increase the number of labeled sentences. In our clustering approach by K-means, the number of clusters (parameter k) is set to a relatively large number (10 % of the total number of review sentences) to obtain many small but accurate clusters of review sentences. The accuracy of the the constructed corpus was evaluated by manually checking 50 review sentences for each implicit aspect category. The accuracy was reasonably high, i.e., from 0.58 to 0.82.

RQ2 *How is the performance of implicit aspect classification by a model trained from an automatically constructed corpus?*

We presented the method to fine-tune the BERT model for implicit aspect extraction (called "implicit aspect classification" in this thesis because it was defined as the classification task). The constructed corpus annotated with implicit aspects was used as the labeled dataset for fine-tuning of BERT. The results of the experiments showed that the accuracy of the BERT model trained with implicit sentences was 0.17 or 0.16 points better than that with explicit sentences. It indicated that our corpus was obviously better than the existing dataset labeled with explicit aspect.

RQ3 *Is it possible to improve the performance of implicit aspect extraction using both datasets of implicit and explicit aspects?*

To answer this research question, we compared the implicit aspect classification performance of M_i (which is the BERT model trained with the dataset conatining implicit aspects only) and M_{e+i} (which is the BERT model trained with both implicit aspects and explicit aspects). The results of the experiments showed that the integration of implicit sentences and explicit sentences could further boost the performance of implicit aspect classification. Specifically, M_{e+i} is better than M_i in accuracy by 0.03 and 0.07 points for the phone and PC domains, respectively.

5.3 Future Work

Future avenues of research continuing those of this study include the following:

- It is necessary to extend our method of constructing the dataset as well as implicit aspect classification to handle sentences including multiple implicit aspects. One of the possible solutions is as follows. Instead of a hard clustering, a soft clustering is applied to allow a sentence to belong to multiple clusters. It enables us to add multiple implicit aspects for one sentence in the dataset. Then, we train the model for multi-class classification that can identify multiple implicit aspects in one sentence.

- The explicit aspects were automatically extracted, but some of them may be incorrect. On the other hand, the sentences including the explicit aspects can be obtained from the existing dataset for ABSA. These sentences can be mixed with unlabeled sentences for the clustering. Such an approach may improve the performance of the clustering.

- Manual construction of the synonym lists shown in Table 3.6 (also Table A.1 and A.2) can be replaced with an automatic synonym expansion method.

- More appropriate clustering algorithms other than k-means should be investigated.

- The performance of the the implicit aspect classification model, which is trained on the corpus containing the review sentences with explicit aspects exactly from the same domain, can be further investigated.

Publications

Journal

- Aye Aye Mar, Kiyoaki Shirai, and Natthawut Kertkeidkachorn. 2023. Weakly Supervised Learning Approach for Implicit Aspect Extraction. Information 14, no. 11: 612. https://doi.org/10.3390/info14110612

International Conference

- Aye Aye Mar and Kiyoaki Shirai. 2022. Automatic Construction of an Annotated Corpus with Implicit Aspects. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 6985–6991, Marseille, France. European Language Resources Association.

Bibliography

[1] Barakat AlBadani, Ronghua Shi, Jian Dong, Raeed Al-Sabri, and Oloulade Babatounde Moctard. Transformer-based graph convolutional network for sentiment analysis. *Applied Sciences*, 12(3), 2022.

[2] Cem Rıfkı Aydın and Tunga Güngör. Sentiment analysis in Turkish: Supervised, semi-supervised, and unsupervised techniques. *Natural Language Engineering*, 27(4):455–483, 2021.

[3] Ayoub Bagheri, Mohamad Saraee, and Franciska de Jong. Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews. *Knowledge-Based Systems*, 52:201–213, 2013.

[4] Alexandra Balahur, Jesús Hermida, and Andrés Montoyo. Detecting implicit expressions of emotion in text: A comparative analysis. *Decision Support Systems*, 53:742–753, 11 2012.

[5] Samuel Brody and Noemie Elhadad. An unsupervised aspect-sentiment model for online reviews. In Ron Kaplan, Jill Burstein, Mary Harper, and Gerald Penn, editors, *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 804–812, Los Angeles, California, June 2010. Association for Computational Linguistics.

[6] Hongjie Cai, Rui Xia, and Jianfei Yu. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In *59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 340–350, 2021.

[7] Keke Cai, Scott Spangler, Ying Chen, and Li Zhang. Leveraging sentiment analysis for topic detection. In *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, volume 1, pages 265–271, 2008.

[8] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien, editors. *Semi-Supervised Learning*. The MIT Press, 2006.

[9] Ganpat Singh Chauhan, Yogesh Kumar Meena, Dinesh Gopalani, and Ravi Nahta. A mixed unsupervised method for aspect extraction using bert. *Multimedia Tools Appl.*, 81(22):31881–31906, sep 2022.

[10] Ivan Cruz, Alexander Gelbukh, and Grigori Sidorov. Implicit aspect indicator extraction for aspect based opinion mining. *International Journal of Computational Linguistics and Applications*, 5:135–152, 2014.

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186, 2019.

[12] Geli Fei, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. A dictionary-based approach to identifying aspects implied by adjectives for opinion mining. In Martin Kay and Christian Boitet, editors, *Proceedings of COLING 2012: Posters*, pages 309–318, Mumbai, India, December 2012. The COLING 2012 Organizing Committee.

[13] Vaishali Ganganwar and R Rajalakshmi. Implicit aspect extraction for sentiment analysis: A survey of recent approaches. *Procedia Computer Science*, 165:485–491, 2019.

[14] M. Ghiassi and S. Lee. A domain transferable lexicon set for twitter sentiment analysis using a supervised machine learning approach. *Expert Systems with Applications*, 106:197–216, 2018.

[15] Athanasios Giannakopoulos, Diego Antognini, Claudiu Musat, Andreea Hossmann, and Michael Baeriswyl. Dataset construction via attention for aspect

term extraction with distant supervision. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 373–380. IEEE, 2017.

[16] Masashi Hadano, Kazutaka Shimada, and Tsutomu Endo. Aspect identification of sentiment sentences using a clustering algorithm. *Procedia-Social and Behavioral Sciences*, 27:22–31, 2011.

[17] Zhen Hai, Kuiyu Chang, and Jung-jae Kim. Implicit feature identification via co-occurrence association rule mining. In Alexander F. Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, pages 393–404, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[18] El Hannach Hajar and Benkhalifa Mohammed. Hybrid approach to extract adjectives for implicit aspect identification in opinion mining. In *2016 11th International Conference on Intelligent Systems: Theories and Applications (SITA)*, pages 1–5, Oct 2016.

[19] J. A. Hartigan and M. A. Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.

[20] Qinze He. Recent works for sentiment analysis using machine learning and lexicon based approaches. In *2022 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, pages 422–426, April 2022.

[21] Ruining He and Julian McAuley. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *25th International Conference on World Wide Web*, pages 507–517, 2016.

[22] Hendriyana, Arief Fatchul Huda, and ZK Abdurahman Baizal. Feature extraction amazon customer review to determine topic on smartphone domain. In *13th International Conference on Information & Communication Technology and System (ICTS)*, pages 342–347, Oct 2021.

[23] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004.

[24] Minqing Hu and Bing Liu. Mining opinion features in customer reviews. In *Proceedings of the 19th National Conference on Artifical Intelligence*, AAAI'04, page 755–760. AAAI Press, 2004.

[25] Minqing Hu and Bing Liu. Opinion feature extraction using class sequential rules. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, pages 61–66. Stanford, CA, USA, 2006.

[26] Niklas Jakob and Iryna Gurevych. Extracting opinion targets in a single and cross-domain setting with conditional random fields. In Hang Li and Lluís Màrquez, editors, *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1035–1045, Cambridge, MA, October 2010. Association for Computational Linguistics.

[27] Niklas Jakob and Iryna Gurevych. Extracting opinion targets in a single and cross-domain setting with conditional random fields. In *Proceedings of the 2010 conference on empirical methods in natural language processing*, pages 1035–1045, 2010.

[28] S Jayanthi and S S Arumugam. Exhibiting the explicit aspects in twitter sentimental data. In *2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*, pages 1–6, 2023.

[29] Fatma Jemai, Mohamed Hayouni, and Sahbi Baccar. Sentiment analysis using machine learning algorithms. In *2021 International Wireless Communications and Mobile Computing (IWCMC)*, pages 775–779, June 2021.

[30] Nadheesh Jihan, Yasas Senarath, and Surangika Ranathunga. Aspect extraction from customer reviews using convolutional neural networks. In *2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer)*, pages 215–220, 2018.

[31] Nadheesh Jihan, Yasas Senarath, Dulanjaya Tennekoon, Mithila Wickramarathne, and Surangika Ranathunga. Multi-domain aspect extraction using support vector machines. In Lun-Wei Ku and Yu Tsao, editors, *Proceedings of the 29th Conference on Computational Linguistics and Speech Processing (ROCLING 2017)*, pages 308–322, Taipei, Taiwan, November 2017. The Association for Computational Linguistics and Chinese Language Processing (ACLCLP).

[32] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

[33] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, page 282–289, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[34] Chenghua Lin, Yulan He, Richard Everson, and Stefan Ruger. Weakly supervised joint sentiment-topic detection from text. *IEEE Transactions on Knowledge and Data Engineering*, 24(6):1134–1145, 2012.

[35] Qiu Lirong. An opinion analysis model for implicit aspect expressions based on semantic ontology. *International Journal of Grid and Distributed Computing*, 8:165–172, 10 2015.

[36] Bing Liu. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, May 2012.

[37] Bing Liu, Minqing Hu, and Junsheng Cheng. Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, page 342–351, New York, NY, USA, 2005. Association for Computing Machinery.

[38] Jakub Macháček. BUTknot at SemEval-2016 task 5: Supervised machine learning with term substitution approach in aspect category detection. In

Steven Bethard, Marine Carpuat, Daniel Cer, David Jurgens, Preslav Nakov, and Torsten Zesch, editors, *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 301–305, San Diego, California, June 2016. Association for Computational Linguistics.

[39] Zohreh Madhoushi, Abdul Razak, and Suhaila Zainudin. Aspect-based sentiment analysis methods in recent years. *Asia-Pacific Journal of Information Technology & Multimedia*, 08:79–96, 06 2019.

[40] Arash Maghsoudi, Sara Nowakowski, Ritwick Agrawal, Amir Sharafkhaneh, Mark E Kunik, Aanand D Naik, Hua Xu, and Javad Razjouyan. Sentiment analysis of insomnia-related tweets via a combination of transformers using dempster-shafer theory: Pre– and peri–covid-19 pandemic retrospective study. *J Med Internet Res*, 24(12):e41517, Dec 2022.

[41] Jaafar Zubairu Maitama, Norisma Idris, Asad Abdi, Liyana Shuib, and Rosmadi Fauzi. A systematic review on implicit and explicit aspect extraction in sentiment analysis. *IEEE Access*, 8:194166–194191, 2020.

[42] Edison Marrese-Taylor, Juan D. Velásquez, and Felipe Bravo-Marquez. A novel deterministic approach for aspect-based opinion mining in tourism products reviews. *Expert Systems with Applications*, 41(17):7764–7775, 2014.

[43] Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. Topic sentiment mixture: Modeling facets and opinions in weblogs. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, page 171–180, New York, NY, USA, 2007. Association for Computing Machinery.

[44] Dheeraj Mekala, Vivek Gupta, Bhargavi Paranjape, and Harish Karnick. SCDV : Sparse composite document vectors using soft clustering over distributional representations. In *2017 Conference on Empirical Methods in Natural Language Processing*, pages 659–669, Copenhagen, Denmark, Sep 2017. Association for Computational Linguistics.

[45] Pallavi Mishra and Sandeep Kumar Panda. Dependency structure-based rules using root node technique for explicit aspect extraction from online reviews. *IEEE Access*, 11:65117–65137, 2023.

[46] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1–2):1–135, 2008.

[47] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pages 79–86. Association for Computational Linguistics, July 2002.

[48] Ameya Parkar and Rajni Bhalla. A survey paper on the latest techniques for implicit feature extraction using ccc method. In *2022 Algorithms, Computing and Mathematics Conference (ACM)*, pages 20–29, Aug 2022.

[49] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. SemEval-2014 task 4: Aspect based sentiment analysis. In *8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland, Aug 2014. Association for Computational Linguistics.

[50] Soujanya Poria, Erik Cambria, and Alexander Gelbukh. Aspect extraction for opinion mining with a deep convolutional neural network. *Know.-Based Syst.*, 108(C):42–49, sep 2016.

[51] Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. Expanding domain sentiment lexicon through double propagation. In *Twenty-First International Joint Conference on Artificial Intelligence*. Citeseer, 2009.

[52] Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1):9–27, 2011.

[53] Toqir A Rana and Yu-N Cheah. Aspect extraction in sentiment analysis: Comparative analysis and survey. *Artificial Intelligence Review*, 46:459–483, 2016.

[54] Yuliya Rubtsova and Sergey Koshelnikov. Aspect extraction from reviews using conditional random fields. In Pavel Klinov and Dmitry Mouromtsev, editors, *Knowledge Engineering and Semantic Web*, pages 158–167, Cham, 2015. Springer International Publishing.

[55] Marzieh Saeidi, Guillaume Bouchard, Maria Liakata, and Sebastian Riedel. SentiHood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods. In *26th International Conference on Computational Linguistics*, pages 1546–1556, Osaka, Japan, Dec 2016. The COLING 2016 Organizing Committee.

[56] Amani Khalaf Samha, Yuefeng Li, and Jinglan Zhang. Aspect-based opinion mining from product reviews using conditional random fields. In *Data Mining and Analytics: Proceedings of the 13th Australasian Data Mining Conference [Conferences in Research and Practice in Information Technology, Volume 168]*, pages 119–128. Australian Computer Society, 2015.

[57] Kim Schouten and Flavius Frasincar. Finding implicit features in consumer reviews for sentiment analysis. In *14th International Conference on Web Engineering*, pages 130–144. Springer, 2014.

[58] Burr Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

[59] Piyush Kumar Soni and Radhakrishna Rambola. A survey on implicit aspect detection for sentiment analysis: Terminology, issues, and scope. *IEEE Access*, 10:63932–63957, 2022.

[60] Qi Su, Kun Xiang, Houfeng Wang, Bin Sun, and Shiwen Yu. Using pointwise mutual information to identify implicit features in customer reviews. In Yuji Matsumoto, Richard W. Sproat, Kam-Fai Wong, and Min Zhang, editors, *Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead*, pages 22–30, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[61] Li Sun, Sheng Li, JiYun Li, and JuTao Lv. A novel context-based implicit feature extracting method. In *2014 International Conference on Data Science and Advanced Analytics (DSAA)*, pages 420–424, Oct 2014.

[62] Charles Sutton and Andrew McCallum. An introduction to conditional random fields. *Found. Trends Mach. Learn.*, 4(4):267–373, apr 2012.

[63] S. Santhoshkumar T. Hemakala. Advanced classification method of twitter data using sentiment analysis for airline service. *International Journal of Computer Sciences and Engineering*, 6:331–335, 7 2018.

[64] Sayyida Tabinda Kokab, Sohail Asghar, and Shehneela Naz. Transformer-based deep learning models for the sentiment analysis of social media data. *Array*, 14:100157, 2022.

[65] Kian Long Tan, Chin Poo Lee, Kalaiarasi Sonai Muthu Anbananthen, and Kian Ming Lim. Roberta-lstm: A hybrid model for sentiment analysis with transformer and recurrent neural network. *IEEE Access*, 10:21517–21525, 2022.

[66] Kian Long Tan, Chin Poo Lee, and Kian Ming Lim. A survey of sentiment analysis: Approaches, datasets, and future research. *Applied Sciences*, 13(7), 2023.

[67] Kian Long Tan, Chin Poo Lee, Kian Ming Lim, and Kalaiarasi Sonai Muthu Anbananthen. Sentiment analysis with ensemble hybrid deep learning model. *IEEE Access*, 10:103694–103704, 2022.

[68] Senait Gebremichael Tesfagergish, Jurgita Kapočiūtė-Dzikienė, and Robertas Damaševičius. Zero-shot emotion detection for semi-supervised sentiment analysis using sentence transformers and ensemble learning. *Applied Sciences*, 12(17), 2022.

[69] Sameer Yadavrao Thakur, K. H. Walse, and V. M. Thakare. A systematic review on explicit and implicit aspect based sentiment analysis. In Mukesh Saraswat, Chandreyee Chowdhury, Chintan Kumar Mandal, and Amir H.

Gandomi, editors, *Proceedings of International Conference on Data Science and Applications*, pages 329–356, Singapore, 2023. Springer Nature Singapore.

[70] Tun Thura Thet, Jin-Cheon Na, and Christopher SG Khoo. Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of information science*, 36(6):823–848, 2010.

[71] Mohammad Tubishat, Norisma Idris, and Mohammad AM Abushariah. Implicit aspect extraction in sentiment analysis: Review, txonomy, oppportunities, and open challenges. *Information Processing & Management*, 54(4):545–563, 2018.

[72] Peter Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In Pierre Isabelle, Eugene Charniak, and Dekang Lin, editors, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.

[73] S. Josephine Vanaja and Meena Belwal. Aspect-level sentiment analysis on e-commerce data. *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pages 1275–1279, 2018.

[74] Asma Al Wazrah and Sarah Alhumoud. Sentiment analysis using stacked gated recurrent unit for arabic tweets. *IEEE Access*, 9:137176–137187, 2021.

[75] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In Raymond Mooney, Chris Brew, Lee-Feng Chien, and Katrin Kirchhoff, editors, *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 347–354, Vancouver, British Columbia, Canada, October 2005. Association for Computational Linguistics.

[76] Fu Xianghua, Liu Guo, Guo Yanyan, and Wang Zhiqiang. Multi-aspect sentiment analysis for chinese online social reviews based on topic modeling and hownet lexicon. *Knowledge-Based Systems*, 37:186–195, 01 2013.

[77] Nurulhuda Zainuddin, Ali Selamat, and Roliana Ibrahim. Improving twitter aspect-based sentiment analysis using hybrid approach. In Ngoc Thanh Nguyen, Bogdan Trawiński, Hamido Fujita, and Tzung-Pei Hong, editors, *Intelligent Information and Database Systems*, pages 151–160, Berlin, Heidelberg, 2016. Springer Berlin Heidelberg.

[78] Lingwei Zeng and Fang Li. A classification-based approach for implicit feature identification. In Maosong Sun, Min Zhang, Dekang Lin, and Haifeng Wang, editors, *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, pages 190–202, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[79] Lei Zhang and Bing Liu. *Aspect and Entity Extraction for Opinion Mining*, pages 1–40. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[80] Yu Zhang and Weixiang Zhu. Extracting implicit features in online customer reviews for opinion mining. In *22nd International Conference on World Wide Web*, pages 103–104, 2013.

[81] Zhi-Hua Zhou. A brief introduction to weakly supervised learning. *National Science Review*, 5:44–53, 2018.

[82] Xiaojin Zhu. Semi-supervised learning literature survey. Technical Report 1530, Computer Sciences, University of Wisconsin-Madison, 2005.

Appendix A

Full list of synonyms

Table A.1 and A.2 show all synonyms of the implicit aspects for mobile phone and PC domains, respectively. These tables include two lists of the synonyms used for different purposes. The column "Synonyms for soft-matching with cluster label" shows the synonyms to choose the clusters of review sentences including the target implicit aspects (as described in Subsubsection 3.2.4), while the column "Synonyms for soft-matching with explicit aspect" shows the synonyms to extract sentences including the target explicit aspects from SemEval 2014 Task 4 ABSA dataset (as described in Subsection 3.3).

Table A.1: All synonyms of aspects for mobile phone domain

Aspect	**Synonyms for soft-matching with cluster label**	**Synonyms for soft-matching with explicit aspect**
Battery	battery case, battery life, battery percentages, battery access, battery pack, battery charge, battery charger, charger, blackberry charger brand, blackberry charger, USB charger, USB adapter, cord, USB cord, USB port, USB ports, USB plugs, car charger, USB cable, USB cables, Samsung car charger, quality charger, power, power port, power loss, power light	battery life, charger, cord, usb port, usb ports, power,power light
Case	case quality, case cover	case design
Look	design, color	design ,designed, color
Price	—	price tag, price range, cost, costing, priced, costed, shipping, budget, value
Screen	screen protector, screen protectors, screen cover, screen look, precut screen protectors	screen resolutions, screen resolution, 18.4” screen, screen graphics, looking, service center, seventeen inch screen, 17” inch screen, 17-inch screen, 17 ince screen, 17 inch screen, resolution of the screen, screen brightness, display, monitor, surface, stock screen, screen size, acer screen, lcd, lcd screen, screen hinges,picture quality, resolution on the screen
Size	—	size, sized

Table A.2: All synonyms of aspects for PC domain

Aspect	Synonyms for SM[1] with cluster label	Synonyms for soft-matching with explicit aspect
Interface	keyboard, touchpad, touch pad, keyboard flex	keyboard, touchpad, touch pad, keyboard flex,Keyboard , KEYBOARD, touch pad, keys, mouse, trackpad,left mouse key, key bindings,10-key,regular layout keyboard, right click key, touch-pad , mouse keys, island backlit keyboard,multi-touch mouse,multi-touch track pad,Apple keyboard,mouse on the pad, left button,shift key, mouse pointer, flatline keyboard
OS	windows xp home edition,windows media player,windows xp,windows xp pro, windows convert, operating system, os	windows xp home edition , windows media player, windows xp , windows xp pro , windows convert , operating system , os, Windows 7, operating system , operating systems, XP , Vista , Windows applications , Windows Vista , key pad , Mac OS , antivirus software, Windows XP SP2, Windows 7 Ultimate , OSX 16 , Windows 7 Starter, Windows 7 Home Premium, Windows 7 Professional, Windows operating system, Windows operating systems, Windows update, Windows XP drivers, Window update, Windows, Windows Vista Home Premium, Win 7
Price	—	price tag, price range, cost, costing, priced, costed, shipping, budget, value
Screen	monitor, screen size,screen real estate, screen flickers, screen distortion	monitor, screen size, screen real estate, screen flickers, screen distortion, screen resolutions, screen resolution, screen dispaly, 18.4” screen, screen graphics, looking, service center, seventeen inch screen, 17” inch screen, 17-inch screen, 17 ince screen, Screen size, resolution of the screen, screen brightness,30” HD Monitor, display, Resolution, display, surface, stock screen, Acer screen,17 inch screen, LCD, screen hinges, picture quality, resolution on the screen
Software	programs, program, isoftware, applications, software kit, software problem, itools software	programs, program,isoftware, applications, software kit, software problem, itools software, MS Applications, suite of software, system, Microsoft office for the mac, preloaded software, Microsoft office, software packages, trackpad, Software, antivirus software, Microsoft Word for Mac, MS Office, MS Office apps, Dreamweaver, Final Cut Pro 7, Photoshop, Safari, Firefox , MSN Messenger, Apple applications, music software, Office Mac applications, Word, Excel, software options, Sony Sonic Stage software, Garmin GPS software, Microsoft Office 2003, powerpoint, iMovie, iWork, Internet Explorer

[1] soft-matching.