Automatic Construction of an Annotated Corpus with Implicit Aspects

Aye Aye Mar, Kiyoaki Shirai

Graduate School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292, Japan
{ s1920405, kshirai }@jaist.ac.jp

Abstract

Aspect-based sentiment analysis (ABSA) is a task that involves classifying the polarity of aspects of the products or services described in users' reviews. Most previous work on ABSA has focused on explicit aspects, which appear as explicit words or phrases in the sentences of the review. However, users often express their opinions toward the aspects indirectly or implicitly, in which case the specific name of an aspect does not appear in the review. The current datasets used for ABSA are mainly annotated with explicit aspects. This paper proposes a novel method for constructing a corpus that is automatically annotated with implicit aspects. The main idea is that sentences containing explicit and implicit aspects share a similar context. First, labeled sentences with explicit aspects and unlabeled sentences that include implicit aspects are collected. Next, clustering is performed on these sentences so that similar sentences are merged into the same cluster. Finally, the explicit aspects are propagated to the unlabeled sentences in the same cluster, in order to construct a labeled dataset containing implicit aspects. The results of our experiments on mobile phone reviews show that our method of identifying the labels of implicit aspects achieves a maximum accuracy of 82%.

Keywords: aspect-based sentiment analysis, implicit aspect, opinion mining

1. Introduction

With the increase in customer reviews of products on the Web, opinion mining from those reviews has become a prominent research area. In particular, a fundamental technique for fine-grained opinion mining called aspect-based sentiment analysis (ABSA), in which sentiments are analyzed in relation to a specific aspect of a product, plays a vital role not only for customers but also for manufacturers. It allows customers to find the strong and weak points of the products in which they are interested, while manufacturers can identify the customers' needs more accurately. In the product domain, the term "aspect" (also called "feature" in some work) means a component or attribute of the product. For example, when carrying out opinion mining from phone reviews, "battery" and "price" are examples of the aspects of a phone.

In general, ABSA consists of two tasks: aspect term extraction and aspect polarity classification. The former involves extracting all aspects from the sentences in the review, while the aim of the latter is to classify whether the customer is expressing a positive, neutral, or negative opinion of each extracted aspect. Two kinds of aspects are considered in the task of aspect term extraction, namely 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 names of the aspects. Some examples of these are as follows:

- [review sentence 1] The battery of the phone lasts many hours, so it does not need to charge frequently.
- [review sentence 2] I don't use it any more, as I get tired of always recharging after using just for a few hours.

Both reviews mention the same aspect of the phone, the "battery". Review sentence 1 contains the explicit aspect "battery" and directly expresses an opinion about it, whereas review sentence 2 implicitly expresses an opinion about the battery but without using the word "battery" itself. In the second sentence, the battery is an implicit aspect.

Most of the current studies on aspect term extraction focus only on the explicit aspects, and ignore implicit ones. However, 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 Web. In addition, implicit review sentences are more complex than explicit ones (Liu, 2012). 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.

A lack of a large review corpus annotated with implicit aspects is one of the bottleneck problems for implicit aspect extraction. Most current methods of aspect term extraction rely on supervised learning, in which an aspect extraction model is trained on labeled dataset. The extraction of implicit aspects cannot be performed in the same way when there is no dataset labeled with implicit aspects. To the best of our knowledge, no prior work has been done to automatically construct a large review corpus annotated with implicit aspects from raw review texts.

The final goal of this study is to develop a system of ABSA for implicit aspects, which can extract implicit aspects from customer reviews and classify their polarity. As an initial step, this paper proposes a novel method of constructing an annotated corpus of product reviews that is automatically labeled with the implicit aspects. This corpus can be used to develop our ABSA system using a sophisticated machine learning method. We will also demonstrate the effectiveness of our proposed method using unlabeled reviews of mobile phones. Furthermore, we will present our findings on the complex nature of the implicit aspects and the problems with the construction of the corpus through an error analysis.

The rest of the paper is organized as follows. Section 2 describes related work on the extraction of implicit aspects and existing labeled corpora. Section 3 introduces the framework of the proposed method and explains each component of it in detail. The results of our experiments and a discussion are presented in Section 4. Finally, Section 5 concludes the paper.

2. Related Work

As discussed above, current methods of ABSA mainly focus on explicit aspects, and there have been few attempts to study implicit aspects. Hai et al. (2011) proposed a co-occurrence association rule mining approach for identifying implicit aspects. The association rule was in the form (sentiment word → 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 (2013) 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. 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 (Hai et al., 2011). 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.

Schoten and Frasincar (2014) proposed a method that was applicable to more realistic scenarios. They aimed to find implicit aspects from sentences that could contain zero or more implicit aspects. A training dataset, a corpus annotated with implicit aspects, 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. These authors 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.

Bagheri et al. (2013) proposed a graph-based method for implicit aspect extraction. The vertices in the graph were either explicit aspects or sentiment words, while the edges between them were weighted based on the number of co-occurrences of the aspect and sentiment words 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, aspects connected to sentiment words in a sentence with highly weighted edges were extracted as implicit aspects. Their method was evaluated using a dataset of reviews of five products, constructed by Hu and Liu (2004), and showed that the F1-scores for the implicit aspect extraction method were between 57% and 71%.

Only explicit aspects are annotated in the most commonly used datasets for ABSA, such as Sentihood (Saeidi et al., 2016) and SemEval-2014 Task 4 (Pontiki et al., 2014). However, a small or pilot dataset with implicit aspects has been constructed. Hu and Liu (2004) 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. Both the explicit and implicit aspects were manually annotated. Cruz et al. (2014) extended this dataset by adding annotations of implicit aspect indicators (IAIs), which were sentiment words indicating a certain implicit aspect. They selected sentences labeled with at least one implicit aspect from Hu and Liu's dataset, and then manually annotated the IAIs. They then used the extended dataset to train a conditional random field (CRF) to extract IAIs from the review sentences. However, Hu and Liu's dataset was relatively small, and the numbers of sentences containing implicit aspects for each of the five products were between 14 and 55. Current state-of-the-art methods for ABSA are based on deep learning, which requires huge labeled datasets. 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 corpus for each domain. This is our primary motivation for the automatic construction of a large review corpus annotated with implicit aspects.

3. Proposed Method

3.1 Overview

The idea behind our method of corpus construction is that explicit and implicit review sentences tend to be similar (i.e., contain similar words or phrases) when they mention the same aspect. By grouping similar review sentences together and identifying the relevant aspect mentioned in the sentences in the cluster, review sentences without explicit aspects can be labeled with their implicit aspects.

Figure 1 shows an overview of our proposed method. It consists of four steps. First, explicit aspects are extracted from review sentences by supervised machine learning. Since no aspect will be extracted from a large number of the sentences, review sentences with and without explicit aspects are obtained. Next, we perform clustering to merge similar review sentences. As a result, clusters containing sentences with both explicit and implicit aspects are created. A label is identified for each cluster, where the label indicates an aspect mentioned by the sentences in the cluster. Finally, unlabeled sentences in
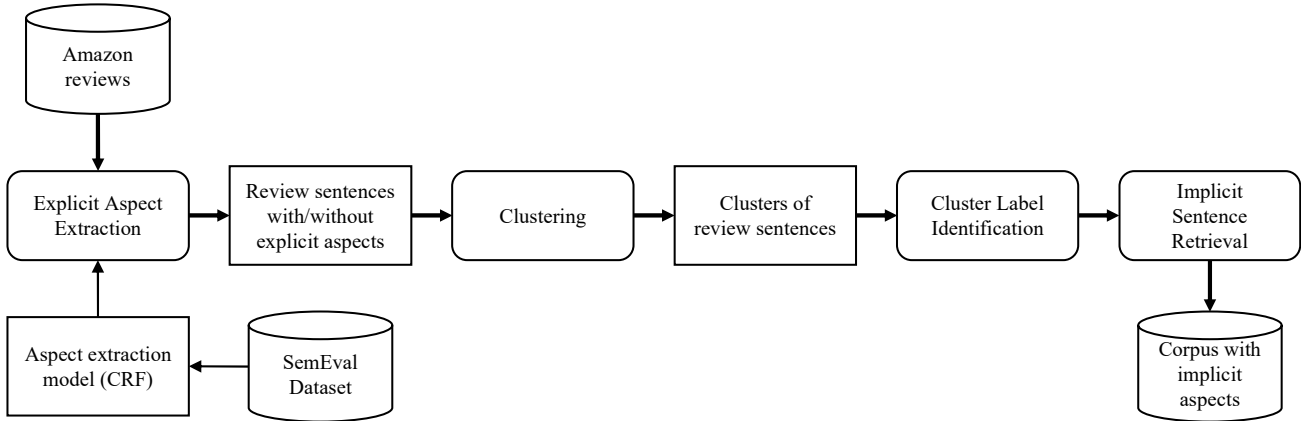
Figure 1: Overview of the proposed method

the cluster are annotated with the identified cluster label, i.e., the implicit aspect. The following subsections describe each step in more detail.

3.2 Explicit Aspect Extraction

Extraction of the explicit aspects from review sentences is an important step in detecting implicit aspects, since we assume that clues for identifying implicit review sentences can be derived from the common words in the explicit reviews. In this study, CRF (Rubtsova and Koshelnikov, 2015) is applied for aspect extraction, since it performs relatively well even on a small set of features. CRF models a conditional probability $P(Y|X)$ over a hidden sequence Y based on an observation X. The model labels an unknown observation sequence X by choosing the hidden sequence Y which maximizes the value of $p(Y|X)$, defined as follows:

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

where C is the set of all graph cliques, f_c is the set of all features, and λ_c are their corresponding weights. $Z(X)$ is a normalized function as described in Equation (2):

$$Z(X) = \sum_y \exp\left(\sum_{c \in C} \lambda_c f_c(y_c, X) \right). \quad (2)$$

A standard model of aspect extraction is trained using CRF. Aspect extraction is defined as a sequential labeling problem with IOB encoding, where B, I, and O stand for the beginning, inside, and outside of the aspect, respectively. The features used to train the CRF are the surface form and the POSs of the words in the context as well as the labels of the previous words. The size of the context window is set to three. The CRF was trained using a public dataset annotated with explicit aspects.

After training the CRF for aspect extraction, it was applied to a large number of unlabeled review sentences. The explicit aspects in each sentence were expected to be successfully identified by the CRF model. Note that explicit aspects cannot be extracted from all sentences. No aspect will be extracted from sentences that include only implicit aspects or no aspect. As a result, a set of review sentences with and without explicit aspects is obtained.

3.3 Clustering of Review Sentences

The review sentences, either labeled with the explicit aspects or unlabeled, were then merged into clusters, where each cluster contained sentences that expressed opinions on the same aspect. First, each review sentence was converted to sparse composite document vectors (SCDVs) (Mekala et al., 2017). SCDV is well-known as an excellent vector representation of a document, and is usually better than the average of the word embeddings. SCDV performs well on heterogeneous tasks such as Topic Coherence, Context-sensitive Learning, and Information Retrieval. We chose SCDV as a representation of the review sentences because it is suitable for capturing the semantic similarity between explicit and implicit sentences.

ID	Explicit Aspect	Review Sentence
1	price, battery	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 .
2	battery	Like it but it causes the battery to get really hot and lock the phone .
3	design	I really like the design, but however the casing did not snap nicely with my phone in place .
4	hard rubber, design	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.
5	*none*	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 .

Table 1: Example of a cluster of review sentences

Next, clustering was performed by k-means, which is an efficient and precise clustering algorithm. In this approach, it is necessary to determine the number of clusters in advance. We set this to 10% of the total number of review sentences, to obtain good-quality clusters. In this step, we do not aim to merge all the sentences expressing opinions on the same aspect into a single cluster; they may be divided into several clusters. However, it is not preferable to merge sentences referring to different aspects into one cluster. In other words, the purity of the clusters should be high. We therefore chose a relatively large number of clusters, so that we could create many small but accurate clusters.

Table 1 shows an example of a cluster. The column entitled "Explicit Aspect" shows the explicit aspects extracted from the sentence by the CRF model. In general, a cluster contains two kinds of review sentences:

1. Review sentences containing one or more explicit aspects, such as the first four sentences in Table 1.
2. Review sentences that do not contain explicit aspects, such as sentence #5, indicated by "*none*" in the "Explicit Aspect" column.

3.4   Identification of Cluster Labels

The task of cluster label identification involves choosing the most relevant aspect for a sentence cluster. This is not always obvious, since there are two or more explicit aspects in the cluster, as shown in the example in Table 1. Figure 2 shows pseudocode for this process.

```
Input: Cluster of review sentences
Output: Cluster label
1  Let sᵢ be a sentence in the cluster, and aᵢ
   be the explicit aspect of sᵢ.
2  Let Fre(aᵢ) be the frequency of aᵢ in the
   cluster.
3  Let Oc(aᵢ) be the number of occurrences of
   aᵢ in the set of sentences {sᵢ}
4  label ← aspect with the maximum Fre(aᵢ)
5  IF label is unique THEN
6       GOTO 14
7  ELSE
8       Let {a'ᵢ} be the set of aspects with
        maximum Fre(aᵢ).
9       label ← aspect with the maximum Oc(a'ᵢ)
10  IF label is unique THEN
11      GOTO 14
12  ELSE
13      return INDETERMINABLE
14  IF Rel(label) ≥ Tᵣ THEN
15      return label
16  ELSE
17      return INDETERMINABLE
```

Figure 2: Algorithm for cluster label identification

The most frequent aspect is chosen as the cluster label. Two kinds of frequency, *Fre(aᵢ)* and *Oc(aᵢ)*, are considered. *Fre(aᵢ)* is the number of the times an aspect a_i is extracted, while *Oc(aᵢ)* is the number of occurrences of the aspect in the review sentences. For example, in the cluster in Table 1, *Fre*(design) = 2[1] and *Oc*(design) = 3[2].

First, the aspect with the maximum value of *Fre* is chosen. When two or more aspects have the same maximum value of *Fre* (e.g., *Fre*(battery) = *Fre*(design) = 2 in Table 1), the aspect with the maximum *Oc* is chosen (e.g., "design" is chosen since *Oc*(design)=3 > *Oc*(battery)=2).

In addition, the reliability of the label is also considered. This is defined in Equation (3):

$$Rel(aspect) = \frac{Fre(aspect)}{number\ of\ sentences\ in\ the\ cluster} \quad (3)$$

In the example in Table 1, *Rel*(design)=2/5. If the reliability of the aspect is higher than or equal to a threshold T_r, it is chosen as the cluster label; otherwise, the label is defined as *INDETERMINATE*, which indicates that the cluster may be wrongly made up of the sentences about different aspects. The threshold T_r was empirically determined in the experiment in Section 4.

3.5   Retrieval of Sentences Labeled with Implicit Aspect

The last step is to collect sentences containing implicit aspects. In this paper, we focus on opinion mining in the mobile phone domain, and aim to extract sentences including one of the following implicit aspects: "battery", "case", "look", "size", "screen", and "price". For each implicit aspect, the cluster whose label is coincident 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 the aspect "battery", its synonyms are "battery case", "battery life", "power" and so on. The list of all synonyms is shown in Table 2. Note that no synonyms are used for the categories of "size" and "price".

Aspect	Synonym
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
case	case quality, case cover
look	design, color
screen	screen protector, screen protectors, screen cover, screen look, precut screen protectors

Table 2 : Synonyms of aspects

Sentences for which no explicit aspect was extracted by CRF are then retrieved from the chosen clusters. The cluster label is attached to these retrieved sentences as their implicit aspects. In the example in Table 1, sentence #5 is retrieved with the label "look" as its implicit aspect, since the cluster label is "design," which is a synonym for "look".

---

[1] Since the aspect "design" appears twice in the "Explicit Aspect" column.

[2] Since the word "design" appears three times in the sentences in the "Review Sentence" column.

Aspect	# of clusters	Ave. size of cluster	# of sentences		Accuracy	T_r
			Explicit	Implicit		
Battery	28	24	274	393	0.82	0.1
Case	15	9.3	66	74	0.74	0.1
Look	24	20	234	252	0.58	0.1
Size	2	10	13	7	0.14	0.1
Screen	7	19	115	21	0.76	0.2
Price	20	22	342	100	0.78	0.4

Table 3: Results of implicit aspect extraction

Cluster 1: (price)

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

Cluster 2: (look)

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

Table 4: Examples of sentences labeled with implicit aspects

4. Evaluation

4.1 Experimental Setup

As described above, we focused on reviews of mobile phones. A set of Amazon product data (He and McAuley, 2016) was used to evaluate our proposed method. We excerpted 10,000 review sentences from the category entitled Cell Phones and Accessories, and these sentences were used to mine implicit aspects.

The CRF model for explicit aspect extraction was trained first. A corpus annotated with gold aspects is required to train CRF, but there is no appropriate public corpus for the mobile phone domain. Hence, laptop reviews from the SemEval-2014 dataset (Pontiki et al., 2014) were used for training, since laptops are fairly similar to mobile phones.

A preliminary evaluation of the performance of the CRF model was carried out on the SemEval-2014 dataset. The precision, recall, and F1-score for aspect extraction were 0.77, 0.64, and 0.70 respectively, which were sufficiently high for the subsequent procedures.

After the explicit aspects had been extracted, the review sentences were converted to SCDVs and clustered using k-means, where the number of clusters was set to 1,000. Finally, the label for each cluster was identified to extract review sentences containing implicit aspects.

In the experiment, we extracted six major implicit aspects for the mobile phone domain, as shown in Subsection 3.5. 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 to

determine whether they expressed users' opinions about the implicit aspect. 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.

4.2 Results

Table 3 summarizes the results of the experiment. We obtained 96 clusters for six aspects, and the average sizes of the clusters (the number of sentences per cluster) were between 10 and 24. 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. T_r shows the threshold used in the algorithm in Figure 2, and this was set to 0.1 in most cases. Since the accuracy was less high for the aspects of "screen" and "price", we increase the threshold for these. Note that when T_r is set to a high value, the number of extracted implicit sentences is reduced, but the accuracy is improved.

It was found that the number of the implicit sentences was small and the accuracy was low for the aspect of "size". It seems that by chance, there were few sentences that mentioned the size of the mobile phone in the reviews used in this experiment. However, the accuracy was relatively high for the other aspects.

Table 4 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 was a synonym for the aspect category of "look".

The number of implicit sentences extracted in this experiment was not extremely large, but they were extracted from only 10,000 review sentences. We could easily increase this number by mining more review sentences. In summary, the results indicate that our proposed method is promising in terms of automatically constructing a dataset annotated with implicit aspects.

4.3 Discussion

This subsection discusses the major causes of error in the process of implicit aspect extraction. 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 the 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 Table 4, sentences #1, #2 and #5 include "price" with other aspects. In this example 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 term "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."

5. Conclusion

This paper has proposed a novel method of automatically constructing an annotated corpus with implicit aspects. A CRF model was used to extract the explicit aspects, clustering of the sentences with both explicit and implicit aspects was carried out by k-means, the cluster labels were identified by some heuristics, and the sentences were automatically annotated with the implicit aspects. The results of our experiments showed that the accuracy of annotation was relatively high for most implicit aspects. Thus, our method was promising, although the automatic annotation of implicit aspects was a difficult task. We have also discussed the current problems in identifying implicit aspects from review sentences.

The future work of this study can be enumerated as follows.

- By handling more unlabeled review sentences, we will enlarge the size of the corpus labeled with implicit aspects.
- A more comprehensive evaluation of the method using additional domains other than mobile phones is required.
- 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, and can also be used as initial clusters. Such an approach may improve the performance of the clustering.
- Manual construction of the synonym lists shown in Table 2 can be replaced with an automatic synonym expansion method.
- Finally, we will investigate the use of sophisticated deep learning methods to automatically identify implicit aspects, using the constructed corpus as training data.

6. Bibliographical References

Bagheri, A., Saraee, M., and De Jong, F. (2013). Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews. *Knowledge-Based Systems*, 52:201-213.

Cruz, I., Gelbukh, A. F., and Sidorov, G. (2014). Implicit aspect indicator extraction for aspect based opinion mining. *International Journal of Computational Linguistics and Applications*, 5(2):135-152.

Hai, Z., Chang, K., and Kim, J. J. (2011). Implicit feature identification via co-occurrence association rule mining. In *Proceedings of the International Conference on*

Intelligent Text Processing and Computational Linguistics, pp. 393-404.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 168-177.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1-167.

Mekala, D., Gupta, V., Paranjape, B., and Karnick, H. (2017) SCDV: Sparse composite document vectors using soft clustering over distributional representations, In *Proceedings of 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 659-669.

Schouten, K. and Frasincar, F. (2014). Finding implicit features in consumer reviews for sentiment analysis. In *Proceedings of the International Conference on Web Engineering*, pp. 130-144.

Rubtsova, Y. and Koshelnikov, S. (2015). Aspect extraction from reviews using conditional random fields. In *Proceedings of the International Conference on Knowledge Engineering and the Semantic Web*, pp. 158-167.

Zeng, L. and Li, F. (2013). A classification-based approach for implicit feature identification. In *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, pp. 190-202.

7.   Language Resource References

He, R. and McAuley, J. (2016). Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *Proceedings of the 25th International Conference on World Wide Web*, pp. 507-517.

Pontiki, M., Galanis D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., and Manandhar, S. (2014). SemEval-2014: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 27-35.