Extracting Pseudo-labeled Samples for Sentiment Classification using Emotion Keywords

Abstract—Sentiment and emotion analysis have been traditionally established as independent research topics in NLP. Although they are two important aspects of subjective information and are closely related, there have been few attempts to combine the two analyses. As a preliminary attempt, we integrate emotion information into sentiment analysis by employing emotion keywords to help automatically extract pseudo-labeled samples. The extracted pseudo-labeled samples are then used as the initial training data to perform semi-supervised learning for sentiment classification. Experimental results across four domains show that our approach using emotion keywords is capable of extracting pseudo-labeled samples with high precision (about 90%). Moreover, the pseudo-labeled samples along with the semi-supervised learning approach further improve the classification performance.

Keywords: sentiment classification; emotion; semi-supervised learning

I. INTRODUCTION

Sentiment analysis has mainly adopted the positive and negative dichotomy in categorizing the semantic orientation of a given text [11][12], while emotion analysis has often investigated human affective information toward a particular text, such as happiness, sadness, fear, and so on [9][10]. The two analyses are closely related, they are yet extensively studied as independent topics of research.

Currently, the mainstream of sentiment classification is based on machine learning classification methods. One of the drawbacks of this approach is the need for a large number of labeled samples which involves extensive manual effort. What is more, the domain-specific nature of such approach requires a certain amount of training samples for different domains in question, which further complicates the problem. Another drawback is that machine learning approach relies heavily on the set of content words which can be ambiguous at times, for examples, the word awful could mean either ‘bad’ or ‘very great’ depending on the context.

Unlike sentiment polarities, emotion expressions are generally domain-independent, i.e. descriptions of human emotions are usually found in different texts regardless of domains or genres. Description of these emotions often reflects the polarity of a text. For instance, the emotion keyword happy in a product review very often implies a positive evaluation toward the product, while angry tends to imply a negative evaluation. Furthermore, as compared to sentiment content words, the number of emotion keywords is relatively small and they are easy to obtain.

Also, emotions keywords are generally less ambiguous, which allows higher domain adaptability. Taking into account that emotion keywords are domain-independent, relatively smaller in size, and less ambiguous, which can greatly alleviate the problem of collecting labeled samples for sentiment classification, we believe that the use of emotion keywords can be an effective mean to automatically collect data with positive and negative evaluations.

This paper aims to extract pseudo-labeled samples for sentiment classification using emotion keywords. In order to ensure the high precision of sample data, two measures are adopted: (1) to remove ambiguous emotion keywords and (2) to remove samples containing ambiguous expressions. Based on the collected pseudo-labeled samples and semi-supervised learning methods, we construct a classifier for sentiment classification. Experimental results show that the proposed approach is effective in obtaining reliable ‘labeled’ samples for sentiment classification.

The remainder of this paper is organized as follows. Section II introduces the related work. Section III introduces the approach of the present study. Section IV discusses our experiment and presents the empirical results of our approach. Finally, Section V gives the conclusion and introduces the future work.

II. RELATED WORK

Sentiment analysis has been extensively studied on different levels of a text in NLP, including word level [4], sentence level [5][8], and document level [12][15]. Different classification methods have also been adopted such as supervised learning [12], semi-supervised learning [7] and unsupervised learning [15].

The positive and negative dichotomy adopted in sentiment analysis has been criticized as too general for real world applications; finer-grained approaches to sentiment analysis, such as emotion analysis, were thus recently proposed. Most research on emotion center on emotion classification by identifying the emotion types, for instance, happiness and sadness, of a given sentence or document [1][10][17]. These studies adopt either rule-based approaches [3][9] or statistical approaches [6][10].

Although sentiment analysis and emotion analysis are closely related research topics, there have been few attempts to integrate the two. Tokuhisa et al. [14] are the first to explore such possibility by decomposing emotion classification task into two steps: (1) sentiment polarity classification and (2) emotion classification. They find that such an integrated approach significantly alleviates
sentiment polarity errors. Different from [14], the present work attempts to enhance sentiment classification by making use of emotion information.

III. METHODOLOGY

This section describes how emotion keywords are identified, how pseudo-labeled samples are extracted using emotion keywords, and how semi-supervised learning is adopted based on the pseudo-labeled samples.

A. Emotion Keywords

As a classic cognitive emotion classification model, Plutchik [13] follows the division of primary emotions and complex emotions and suggests a list of English emotion keywords. Extending Plutchik’s work, Turner’s [16] emotion taxonomy allows more flexible combinations of primary emotions to form complex emotions. In this paper, we adapt Turner’s taxonomy in which there are four primary emotions and 26 complex emotions, each represented with a number of keywords. Turner’s emotion taxonomy does not include some frequently used emotion key words as can be seen in that of Plutchik’s [13]. Hence, we modify Turner’s emotion taxonomy by adding the emotion keywords in Plutchik’s classification model as well as the relevant emotion terms identified in Harvard Psychological Scale to the corresponding classes and levels.

The modified emotion word list is further revised by (1) manually adding the parts of speech (POS) of each emotion keyword such as sadden, sadders, saddened, sadness and sadness are added to the keyword sad; (2) removing emotional keywords that are ambiguous in terms of semantic orientation, i.e. when an emotion keyword is neither positive nor negative. Figure 1 shows the statistics of the emotion keywords used in this study.

<table>
<thead>
<tr>
<th>Number of original emotion keywords</th>
<th>Positive</th>
<th>Negative</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>37</td>
<td>130</td>
<td>enjoy, happy, dissatisfied, hate</td>
</tr>
<tr>
<td>Number of ambiguous emotion keywords</td>
<td>9</td>
<td>62</td>
<td>obliging, surprise, shy</td>
</tr>
<tr>
<td>Number of used emotion keywords with POS</td>
<td>112</td>
<td>301</td>
<td>enjoy, enjoying, enjoyed, happy, happily, happiness</td>
</tr>
</tbody>
</table>

Table 1: Statistics of emotion keywords

B. Pseudo-labeled Sample Extraction

Our basic idea of extracting pseudo-labeled samples is to extract the samples which contain an emotion keyword or more keywords with the same semantic orientation (positive or negative). Since the presence of emotion keywords does not necessarily mean that the sample document contains emotion information, we perform certain specific modifications to the pseudo-labeled samples with the following three situations:

1. Negation

Negation is a major cause for an emotion keyword expressing a different polarity from the whole sentence. For examples,

I really did not enjoy this book

2. Uncertainty

We thought we would continue to be pleased for quite some time. However, after minimal use, we discovered that the finish just didn’t seem to hold up to everyday use.

If I could replace just the mugs, I’d be a happy camper.

3. Irrelevance

I think David Cross is a brilliant comic - even more amazing is that each show (shut up you..., pride is back, etc.) is different material, not the same jokes again and again.

Jason hates sex.

However, accurately recognizing this situation requires a deep understanding of the text. To minimize the manual effort, emotion keywords appearing in quotation marks [““] or parentheses [( )] are all removed.

In summary, each of the above situations is identified based on lists of cue words. Figure 2 shows the number of cue words identified together with some examples.

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negation</td>
<td>24</td>
<td>not, no, never, hardly...</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>22</td>
<td>would, may, if, perhaps...</td>
</tr>
<tr>
<td>Irrelevance</td>
<td>3</td>
<td>(, “ “, )</td>
</tr>
</tbody>
</table>

Table 2: Cue words for ambiguous situations

C. Semi-supervised Learning with Pseudo-labeled Sample Extraction

The extracted pseudo-labeled samples can be used as the training data for the sentiment classifier. However, since the size of the pseudo-labeled samples might be relatively small, the performance of the resulted classifier will be relatively poor. Another way to take advantage of the pseudo-labeled samples is to construct a sentiment
classification system using semi-supervised learning methods.

In this study, the collected pseudo-labeled samples are taken as the initial training data in a semi-supervised learning way. Specifically, we adopt the two-view learning, i.e., personal/impersonal views [7], the state-of-the-art method for semi-supervised sentiment classification, to perform sentiment classification. Strictly speaking, our approach can be considered an unsupervised sentiment classification since it does not require any labeled samples.

Input:
The unlabeled data \( U \)
The emotion keyword list
The labeled data \( L = \phi \)

Output:
Classifier for sentiment classification

Procedure:
1. Employ the personal/impersonal mining approach to get personal sentence set \( S_{\text{personal}} \) and impersonal sentence set \( S_{\text{impersonal}} \) of \( U \).
2. Use emotion keywords to extract pseudo-labeled samples with their personal sentence set \( S_{\text{pl-personal}} \) and impersonal sentence set \( S_{\text{pl-impersonal}} \).
3. Loop for \( N \) iterations until \( U = \phi \)
   3.1) Learn the first classifier \( f_1 \) with \( S_{\text{pl-personal}} \).
   3.2) Use \( f_1 \) to label samples from \( U \) with \( S_{\text{U-personal}} \).
   3.3) Choose \( n_1 \) positive and \( n_1 \) negative most confidently predicted samples \( A_1 \).
   3.4) Learn the second classifier \( f_2 \) with \( S_{\text{pl-impersonal}} \).
   3.5) Use \( f_2 \) to label samples from \( U \) with \( S_{\text{U-impersonal}} \).
   3.6) Choose \( n_2 \) positive and \( n_2 \) negative most confidently predicted samples \( A_2 \).
   3.7) Add samples \( A_1 \cup A_2 \) with the corresponding labels into \( L \).
   3.8) Update \( S_{\text{pl-personal}} \) and \( S_{\text{pl-impersonal}} \).
4. Train a classifier with the new labeled data \( L \).

Figure 1: Our unsupervised learning approach that integrating emotion keywords and personal/impersonal semi-supervised learning

Figure 1 illustrates our algorithm for training the sentiment classifier. The mining approach in Step (1) follows the one suggested in [7]. The difference is that the initial samples used in [7] are manually-labeled, while ours are automatically obtained based on emotion keywords.

IV. EXPERIMENTAL STUDIES

A. Experiment Setup

Our experiments are conducted on the product reviews of four domains: book, DVD, electronics, and kitchen appliances, taken from the multi-domain sentiment classification corpus collected by [2]. In each of the four domains, we recollect 3,000 positive and 3,000 negative reviews from the data.

The classification algorithm is Maximum Entropy (ME) algorithm implementing with the help of the Mallet tool. All parameters are set to their default values. As far as the features are concerned, Boolean word unigram features are employed, representing the presence or absence of a word in a document.

B. Results on Pseudo-labeled Sample Extraction

Table 3 shows the numbers of the extracted pseudo-labeled samples with different disambiguation strategies. We can see that 1,000 articles (without any disambiguation strategy) can be extracted using emotion keywords, which indicates that emotion keywords frequently occur in product reviews.

With the use of disambiguation strategies, the size of the extracted pseudo-labeled samples becomes smaller. However, we can find in Table 4 that the precision of the extracted pseudo-labeled samples greatly increases after employing the disambiguation strategies. The average precision reaches 90% when applying all the strategies at the same time.

<table>
<thead>
<tr>
<th>Positive/Negative</th>
<th>Book</th>
<th>Kitchen</th>
<th>Electronics</th>
<th>DVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without Disambiguation</td>
<td>567/470</td>
<td>429/354</td>
<td>333/295</td>
<td>532/493</td>
</tr>
<tr>
<td>Negation</td>
<td>559/472</td>
<td>429/346</td>
<td>326/304</td>
<td>532/500</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>313/312</td>
<td>313/254</td>
<td>236/218</td>
<td>311/308</td>
</tr>
<tr>
<td>Irrelevance</td>
<td>399/305</td>
<td>349/269</td>
<td>272/209</td>
<td>341/308</td>
</tr>
<tr>
<td>All strategies</td>
<td>232/222</td>
<td>265/195</td>
<td>199/163</td>
<td>199/205</td>
</tr>
</tbody>
</table>

Table 3: The number of the extracted pseudo-labeled samples with different disambiguation strategies

<table>
<thead>
<tr>
<th>Precision (%)</th>
<th>Book</th>
<th>Kitchen</th>
<th>Electronics</th>
<th>DVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without Disambiguation</td>
<td>74.18</td>
<td>83.06</td>
<td>80.41</td>
<td>73.64</td>
</tr>
<tr>
<td>Negation</td>
<td>75.07</td>
<td>87.63</td>
<td>85.24</td>
<td>75.39</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>81.28</td>
<td>87.65</td>
<td>86.12</td>
<td>79.97</td>
</tr>
<tr>
<td>Irrelevance</td>
<td>75.99</td>
<td>84.14</td>
<td>79.88</td>
<td>77.50</td>
</tr>
<tr>
<td>All strategies</td>
<td>85.02</td>
<td>92.39</td>
<td>91.99</td>
<td>86.14</td>
</tr>
</tbody>
</table>

Table 4: The precision of the extracted pseudo-labeled samples with different disambiguation strategies

Although the pseudo-labeled samples achieve a high precision rate, there are still some errors. Through manual checking of the extracted samples, errors are analyzed as follows.

One type of common errors is the more specific kind of polarity shifting phenomenon which requires a deep understanding of contextual information. For example,

Initially we were really pleased, a month later two out of three of them have died.

When I first purchased these towels I was very happy. They were soft and luxurious. After a couple of months of use, I am very sorry I ever purchased them.

Another type of common errors relates to the irrelevant description of the reviewed object. Since we adopt a simple approach in that emotion keywords in quotation marks or parentheses are removed to avoid irrelevant descriptions, cases of irrelevant description without quotation marks or parentheses are thus not recognizable. For instance,

The actors make you feel what they are feeling, the love, hate, indifference, and fear.

Hugh Grant is not too offensive in this role.

C. Results on Semi-supervised Learning with the Pseudo-labeled Sample Extraction

In this experiment, we select 1,000 samples in each domain as testing data and the remaining samples (not including the extracted pseudo-labeled samples) as the unlabeled data. Figure 2 demonstrates the performances of the classifier by the algorithm described in Figure 1.

![Figure 2: The results of running personal/impersonal semi-supervised learning approach with the pseudo-labeled sample as the initial training data](image)

In Figure 2, we can see that the extracted pseudo-labeled samples are consistently effective for semi-supervised learning with personal/impersonal views. The average improvement over four domains is 2.82%, which is similar (or better, e.g., in the DVD domain) to the results reported in [7] in which manual-labeled data are used as the initial training data. This suggests that the pseudo-labeled samples perform as well as manual-labeled data.

V. CONCLUSION AND FUTURE WORK

In this paper, we employ emotion keywords to extract high-precision pseudo-labeled samples which are then used as the initial training data for semi-supervised sentiment classification. Specifically, we propose three different disambiguation strategies to ensure the high precision of the extracted samples. Experimental results show the effectiveness of the proposed strategies. They also demonstrate that our unsupervised learning approach, integrating the pseudo-labeled samples with the personal/impersonal semi-supervised approach, achieves a decent performance for sentiment classification.

As shown in the error analysis, there is still room to improve on the precision. Additional useful strategies will be identified to enhance the precision, especially to cope with the problem of long-distance polarity shifting. Furthermore, attempts will be made to apply our extraction model to sources where more unlabeled data can be obtained such as Twitter. By doing so, the scale of the pseudo-labeled samples will be substantial enough to train a better-performed sentiment classifier.

REFERENCES