Exploiting Combined Multi-level Model for Document Sentiment Analysis

Si Li, Hao Zhang, Weiran Xu, Guang Chen and Jun Guo
School of Information and Communication Engineering
Beijing University of Posts and Telecommunications
Beijing, China
E-mail: ls198cf@gmail.com

Abstract—This paper focuses on the task of text sentiment analysis in hybrid online articles and web pages. Traditional approaches of text sentiment analysis typically work at a particular level, such as phrase, sentence or document level, which might not be suitable for the documents with too few or too many words. Considering every level analysis has its own advantages, we expect that a combination model may achieve better performance. In this paper, a novel combined model based on phrase and sentence level’s analyses and a discussion on the complementation of different levels’ analyses are presented. For the phrase-level sentiment analysis, a newly defined Left-Middle-Right template and the Conditional Random Fields are used to extract the sentiment words. The Maximum Entropy model is used in the sentence-level sentiment analysis. The experiment results verify that the combination model with specific combination of features is better than single level model.

Keywords—combined multi-level model; sentiment analysis; phrase-level; sentence-level; document-level

I. INTRODUCTION

As more and more blogs, shopping sites and related reviews are available in the Web, people want to know more details about what they are interested in together with the corresponding comments. This requirement motivates the research on techniques of sentiment analysis. Sentiment analysis is a type of subjectivity analysis that focuses on identifying positive or negative opinions, emotions, and evaluations expressed in natural language [1].

The traditional text sentiment analysis is typically based on one single level, such as phrase-level, sentence-level and document-level. For documents in the websites, some may have only one or two sentences, or even several simple words, which are too short to be viewed as articles, while some may contain too many sentences. Therefore single level approaches can’t deal flexibly with these situations, and the combination model of different levels is promising to achieve an improvement. In this paper, we focus on using the combined multi-level model (also called combination model for short throughout other parts of this paper) of the phrase-level and sentence-level sentiment analysis to get the polarity of the documents.

We follow the hypothesis that text sentiment expression owns its internal pattern, which can be expressed through the words arranged orderly. Base on this hypothesis, in phrase-level sentiment analysis, we design a Left-Middle-Right (LMR) template to simulate this ordered arrangement which obtains sentiment expressions. The Conditional Random Fields (CRFs) is employed to extract sentiment words and the polarity of each word is given automatically [2]. Moreover, the Maximum Entropy (ME) model is used in the sentence-level sentiment analysis [3]. Morphology and grammar features are selected when using CRFs and ME. After the sentiment words and sentences are recognized, the combination model is proposed to improve the F-measure of document sentiment analysis.

In this paper, our definition of the polarities include positive, negative and neutral, which contains objective texts and hybrid texts. The hybrid text includes praise and criticism, the praise and criticism, which have the same weight, and they should be objective in the whole text.

The rest of the paper is organized as follows. We review related works in Section II. Section III describes the document sentiment analysis with different levels and the combination model. In Section IV, the evaluation of document sentiment analysis based on different-level systems are presented. Finally, in Section V, conclusions and comments on the future work are given.

II. RELATED WORKS

The text sentiment analysis is usually performed based on words, sentences and documents. Kaji [4] explored how to build lexicon for sentiment analysis. Wilson [5] focused on the phrase-level sentiment analysis work to automatically distinguish between prior and contextual polarity and find which features are important for this task. Pong [6] studied the text sentiment analysis as a sentiment classification problem via classifying documents not by topic, but by overall sentiment, and the work was improved in [7] for sentence-level sentiment analysis. Abbasi [8] proposed a kind of feature extraction algorithm for sentiment classification. In [9], a survey covered techniques and approaches that promised to directly enable opinion-oriented information-seeking systems.
III. TEXT SENTIMENT ANALYSIS

A. Language Features

In this section, we focus on showing the features that are used in this paper, which are listed as the following:

<table>
<thead>
<tr>
<th>Features for sentiment analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word N-Grams:</strong> The bag-of-words, also the standard word vector, is the basic and popular used representation of documents. While being put into the n-gram framework, the bag-of-words model is a context free document representation model and can be viewed as a basic unigram form. In this paper, unigram and bigram language features are adopted.</td>
</tr>
<tr>
<td><strong>Part of Speech (POS) tag:</strong> This tag can be considered to be a crude form of word sense disambiguation.</td>
</tr>
<tr>
<td><strong>Syntax tag:</strong> It’s a kind of deeper linguistic analysis.</td>
</tr>
<tr>
<td><strong>Negation terms:</strong> These words can change the sentence polarity obviously, such as “no” and “not”.</td>
</tr>
<tr>
<td><strong>Degree modifiers:</strong> When these words appear, the degree of the sentence polarity may be enhanced or reduced, such as “very” and “much”.</td>
</tr>
<tr>
<td><strong>Transitional words:</strong> These words may change the sentences’ orientation, such as “but”.</td>
</tr>
<tr>
<td><strong>The dependency relationship:</strong> The relationship is an asymmetric binary relationship and obtained according to the dependency parsing tree.</td>
</tr>
</tbody>
</table>

Because original features are usually too many to effectively be used, there are two procedures needed to be performed on the original language-grams, one is feature extraction and the other is feature weighting. A simple feature elimination method is used to abandon the features they appear fewer times than a threshold. Boolean value and absolute term frequency (ABS) are introduced for feature weighting.

B. Phrase-level

1) The LMR sentiment word template: A word appears in the document may be quite different from the prior polarity in the lexicon as shown in [4]. For example, “What a blue day!” and “This is a blue cup.”. In the first sentence, the word “blue” means melancholy, it has a negative polarity, while the word just means one kind of color without any polarity in the second sentence. Therefore, in order to judge the polarity of a word, we put the word into the context. The LMR template takes such condition into account.

We use the word n-grams feature to explain the LMR template. Letters in the “LMR” have different meanings, “M” represents the word which is required to be judged, “L” is the word that on the left of the M word in the ordered arrangement context, and “R” means the word on the right side of the M word. So the words sequence which contains \(2n+1\) words can be denoted as \(L_0L_1…L_nMR_1…R_n\). Some information which is got from “L” and “R” can help to judge the polarity of the “M” word.

Word n-grams, POS tag, syntax tag, negation terms, the dependency relationship, transitional words and degree modifiers are selected as features to extract sentiment word based on CRFs.

Considering the different importance of each feature, we adopted different LMR templates for different features:

- **Strategy A:** Unigram and Bigram + Boolean;
- **Strategy B:** Unigram and Bigram + ABS;
- **Strategy C:** Unigram and Bigram + ABS, negation
words + ABS;

Strategy D: Unigram and Bigram + ABS, degree modifiers + Boolean

Strategy E: Unigram and Bigram + ABS, negation words + ABS, degree modifiers + Boolean.

The ME classification is employed to train a classifier. Then, the following information can be got in each document: The number of the positive sentences is denoted as SenPosNum; the number of the negative sentences is marked as SenNegNum; while the sum of the positive sentences and the negative sentences is denoted as SenSubNum.

SenScale is given as (3):

\[ SenScale = \frac{SenPosNum}{SenSubNum} \]  

(3)

And the document polarity is judged based on sentence-level as (4).

\[ Textpol = \begin{cases} 
-1 & \text{SenScale} < SL \\
1 & \text{SenScale} > SH \\
0 & \text{otherwise}
\end{cases} \]  

(4)

In the sentence-level sentiment analysis, different strategies have their own characteristics. Different thresholds are set for different strategies. SL and SH are set as 0.6 and 0.8 in Strategy A, while in Strategy B, C, D and E they are set as 0.4 and 0.6 respectively.

D. The Combined Multi-level Model

Previous research in information retrieval [11] showed that combining passage retrieval with document retrieval often yields better results than using any single retrieval model alone. This idea can also be considered for sentiment analysis. In this paper, the phrase-level sentiment analysis is the basis of the combination model and a trade-off between phrase-level and sentence-level is involved in the combination model as (5).

\[ TextScale = \mu \times WordScale + (1 - \mu) \times \frac{2 \times WordScale \times SenScale}{WordScale + SenScale} \]  

(5)

TextScale is the parameter for judging the document’s polarity based on the combination model. \( \mu \) is a weighting parameter distributing in the interval \([0, 1]\) and balances the scores of phrase-level and sentiment-level sentiment analysis. In this paper, \( \mu \) is set as 0.8. Then the document polarity is judged as (6) by TextScale.

\[ Textpol = \begin{cases} 
-1 & \text{TextScale} < TL \\
1 & \text{TextScale} > TH \\
0 & \text{otherwise}
\end{cases} \]  

(6)

where TL and TH are set as 0.6 and 0.8 respectively. This trade-off method, which employs the sentence-level result to modify the phrase-level result, improves the term windows limitation in phrase-level.

In this section, we present experimental evaluation results to assess the effectiveness of our sentiment analysis technique on different levels. In particular, we conducted experiments on the 829 website texts which were downloaded from [10]. This collection, including 164 positive texts, 108 negative texts and 557 neutral texts, contains news, expository essay, movie reviews, product reviews and so on. The training data includes 1741 tagged sentiment words, 10147 sentiment sentences and 10000 sentences without emotion. We employed three performance metrics: Precision, Recall and F-measure. Then the Average performance metrics of three kinds of orientations showed the performance of overall classification.

To achieve better results in our experiments, we used the thresholds mentioned in the previous sections for different-level sentiment analysis and the combination model.

The results of the 5 strategies for sentence-level sentiment analysis are illustrated in Table I. From this table, we draw the following observations for text sentiment analysis based on the sentence-level:

1. In positive texts identification, Strategy C reaches the best precision and F-measure values, while Strategy D and Strategy E get the best recall value. For identifying the negative texts, Strategy C and Strategy E reach the best evaluation results respectively. And Strategy A and Strategy E obtain the best results in neutral analysis. These phenomena reveal that Strategy A is suitable for neutral text analysis, while Strategy C is fit for positive and negative sentence-level sentiment analysis. The Strategy C results illustrate that the negation word feature is an important feature for sentiment analysis, especially for the recall of the negative analysis.

2. The Average value is a kind of comprehensive consideration for the three orientations. From the analysis by synthesis, Strategy E brings in improvements over others in precision and F-measure. This is in line with our expectation. Strategy C gets the best average recall value,
which indicates that the negation feature can distinguish three orientations very well. Strategy C performs better than Strategy D. We conjecture that the negation terms feature is much more effective than the degree modifiers feature. When these two features are combined in Strategy E, the best average F-measure is obtained.

Table II shows the evaluation results of each combination model, comparing with the phrase-level method. From Table I and Table II, we can see that document sentiment analysis based on phrase-level significantly outperforms the system based on sentence-level. The improvements over sentence-level from phrase-level are found to be statistically significant for precision, recall and F-measure with large margin. This result implies that the document analysis based on phrase-level is much more effective than the document analysis based on sentence-level. Meanwhile, in the information retrieval system, the smaller the chunk is, the more correct the result is. The result of multi-level sentiment analysis is in accordance with the result of multi-level information retrieval.

The combination model with phrase-level and Strategy A shows the best performances on average precision and average F-measure. Moreover, it is also with significantly improvement over other combination models and the single level document sentiment analysis. It is concluded that the unigram and bigram features with Boolean weighting are complementary with the features which are used in the phrase-level sentiment analysis. This result is in conformity with the previous research mentioned in part D of Section III. And the negation terms feature is also helpful in the combination model for the negative analysis.

V. CONCLUSION

In this paper, we propose a novel combination model which is a new perspective for document sentiment analysis based on different levels and discuss the complementation between different levels with different features. Our empirical evaluation results show that the combination model with phrase-level and Strategy A is an effective model to improve the document sentiment analysis results. The features for sentence-level sentiment analysis in Strategy A and the features for phrase-level are complementary.

In summary, we can see that our model is promising for document sentiment analysis. But this combination model still has some weak points:

1. As shown in Table II, the combination model isn’t effective for all sentence-levels with phrase-level.
2. We have not dealt with the sentence’s position in a document, from which some useful information might be extracted.
3. It is also helpful to take the strength of opinion words and the prior word’s polarity into consideration, due to their contribution in improving the precision of sentiment analysis.

In our future work, we plan to further improve and refine the combined multi-level model, and to deal with the outstanding problems identified above.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China No. 60905017.

REFERENCES


<table>
<thead>
<tr>
<th>Phrase-level</th>
<th>Positive</th>
<th>Negative</th>
<th>Neutral</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P%$</td>
<td>66.8142</td>
<td>88.2353</td>
<td>93.8124</td>
<td>82.9539</td>
</tr>
<tr>
<td>$R%$</td>
<td>92.0732</td>
<td>83.3333</td>
<td>84.2806</td>
<td>86.5957</td>
</tr>
<tr>
<td>$F%$</td>
<td>77.4559</td>
<td>85.7143</td>
<td>88.8469</td>
<td>83.9999</td>
</tr>
<tr>
<td>Phrase-level + A</td>
<td>$P%$</td>
<td>69.5122</td>
<td>88.8899</td>
<td>92.2801</td>
</tr>
<tr>
<td>$R%$</td>
<td>75.4967</td>
<td>86.0987</td>
<td>90.7326</td>
<td>84.1093</td>
</tr>
<tr>
<td>$F%$</td>
<td>69.5467</td>
<td>75.4967</td>
<td>82.5743</td>
<td>76.4956</td>
</tr>
<tr>
<td>Phrase-level + B</td>
<td>$P%$</td>
<td>60.0917</td>
<td>61.6532</td>
<td>90.4867</td>
</tr>
<tr>
<td>$R%$</td>
<td>79.878</td>
<td>90.7407</td>
<td>73.4291</td>
<td>73.4291</td>
</tr>
<tr>
<td>$F%$</td>
<td>68.5864</td>
<td>70.3408</td>
<td>81.0704</td>
<td>74.355</td>
</tr>
<tr>
<td>Phrase-level + C</td>
<td>$P%$</td>
<td>60.2804</td>
<td>63.6943</td>
<td>90.6114</td>
</tr>
<tr>
<td>$R%$</td>
<td>78.6585</td>
<td>92.5926</td>
<td>74.5062</td>
<td>81.9191</td>
</tr>
<tr>
<td>$F%$</td>
<td>68.254</td>
<td>75.4717</td>
<td>81.7734</td>
<td>75.1664</td>
</tr>
<tr>
<td>Phrase-level + D</td>
<td>$P%$</td>
<td>60.7306</td>
<td>63.4616</td>
<td>91.1894</td>
</tr>
<tr>
<td>$R%$</td>
<td>81.0976</td>
<td>91.6667</td>
<td>74.3268</td>
<td>82.3637</td>
</tr>
<tr>
<td>$F%$</td>
<td>69.4517</td>
<td>75.0000</td>
<td>81.8991</td>
<td>75.4305</td>
</tr>
<tr>
<td>Phrase-level + E</td>
<td>$P%$</td>
<td>60.9865</td>
<td>65.3595</td>
<td>92.053</td>
</tr>
<tr>
<td>$R%$</td>
<td>82.9268</td>
<td>92.5926</td>
<td>74.8654</td>
<td>83.4616</td>
</tr>
<tr>
<td>$F%$</td>
<td>70.2842</td>
<td>76.6284</td>
<td>82.5743</td>
<td>76.4956</td>
</tr>
</tbody>
</table>