Sentiment Analysis for Psychological Questionnaires with Semi-supervised Learning

Chen Yaodong¹, Peng Diefei², Liu Qin³
¹ Department of Information and Engineering, Changsha Normal University, Changsha, China.
² Department of Research & Discipline Development, Changsha Normal University, Changsha, China.
³ Department of Vocational & Adult Education Research, Hunan Provincial Research Institute of Education, Changsha, China.

* Corresponding author. Tel.: 86-0731-84036170; email: yaodong.chen@gmail.com
Manuscript submitted April 19, 2018; accepted July 14, 2018.
doi: 10.17706/ijcee.2018.10.3.221-232

Abstract: Sentiment analysis is a more difficult task than topic classification, spam etc. The challenges include how to describe the topic of a given review and how to find the topic-specific sentiment features. This paper applies sentiment analysis to psychological questionnaires. A role-based sentiment analyzer is proposed to detect sentiment features in thousands of questionnaires coming from sandplay therapy and to determine the polarity of them automatically, which help guiding and evaluating the therapeutic process. We decompose the method into 3 steps: detecting topic terms in each review according metadata words, labeling semantic roles on all topic sentences and identifying sentiment features related to topic terms through the labeled roles. A key of our method is that it can find topic-specific sentiment words depending on semantic orientation of roles. Addressing the shortage of manually tagged questionnaires, one semi-supervised learning model -- Spectral Graph Transducer (SGT) is applied to sentiment analysis based on a new feature representation on reviews, i.e. Positive-Related Vector Percentage. The experiments on sandplay questionnaires showed our sentiment analyzer performed well especially on long reviews, and when using SGT it improved rapidly in a small amount of tagged training data.

Key words: Sentiment analysis, sandplay questionnaires, topic-specific sentiment features, topic sentences, spectral graph transducer, semantic role labelling.

1. Introduction

Recently, the bulk of document classification work has focused on Sentiment Analysis (also called Polarity Analysis) with the rapid growth of reviews in discuss websites and questionnaires coming from e-business or other domains. Sandplay therapy (also called sandspiel) is a therapeutic intervention that makes use of a sandbox, toy figures, and sometimes water, to create scenes of miniature worlds that reflect a person's (especially a child) inner thoughts, struggles, and concerns [1]. It is often used to those children who have suffered some form of trauma, neglect, or abuse. Some kinds of questionnaires are designed before or after the therapeutic process for guiding the process or evaluating the results of therapy. In this paper we apply sentiment analysis to medical research, specifically, developing a sentiment analyzer to automatically determine the polarities of thousands of psychological questionnaires.

Previous work on sentiment analysis reported that it is more difficult than other tasks in computational linguistics such as topic classification, spam [2], [3]. There are two challenges: one is how to determine the
topics for a given review, especially several candidates of topics exist in the review, and the other is how to
distinguish related sentiment features corresponding to a specified topic from large noise. Given the topic sandbox in example A, the traditional sentiment analyzer based on topic-based methods [4], [5] or lexical-based methods [6] could classify A into a negative polarity. In fact, the sentiments of enduring and discomfort are not associated with the topic. The sentiment of example A should be positive because the sentiment word likes is positive and it is semantically relevant to the topic sandbox.

A. My boy likes the sandbox game in spite of enduring a lot of discomfort in the closed little room.

We viewed the likes as related sentiment feature since the sentiment of it is relevant to the topic directly. On the other hand, enduring and discomfort are noise because they are not associated with the topic semantically. Addressing this problem, a fine-grained parse is necessary to identify the semantic relationship between the sentiment unit and the specified topic in reviews [7].

One contribution of this paper is introducing a role-based sentiment analyzer (SA for short), which is depending on the identification of related sentiment features according labeled semantic roles. Fig. 1 shows the architecture of the SA. The module of Topic Extraction is designed to detect topic terms based on the survey metadata, where the topic is represented by a collection of words (called topic terms) equivalently. Different from [8] where simple noun and verb groups are scanned for searching the candidate of topics, we use the metadata words that are labeled when designing questionnaires. The metadata words can be viewed as indicator to locate the topic terms in reviews. The Feature Identification is the core module of SA. In one review, we detect all sentences that contain at least one topic term. These sentences are viewed as on-topic sentences. A semantic role labeler is then used to scan roles in these sentences, and then a heuristic algorithm is developed to distinguish the sentiment features relevant to topic terms. Addressing a shortage of manually tagged data about our questionnaires, another contribution of this paper is applying a semi-supervised model -- Spectral Graph Transducer (SGT) [9] to our model. The module of SGT Classification in Fig. 1 exploits the structure and the distribution of untagged data when training and testing.

Thousands of questionnaires are collected by our Child Behavior Laboratory and are used to evaluate the role-based sentiment analyzer in experiments. We propose a new kind of feature description of reviews, called Positive-Related Vector Percentage. This description can capture the sentiment characteristics of reviews more effectively compared to Positive-Sentence Percentage [10]. The results showed that the role-based sentiment analyzer performs well, especially on long reviews, and when using SGT, it improves rapidly in a small amount of tagged training data.

The rest of this paper is organized as follows. Section 2 gives the related work and introduces Semantic Role Labeling. Section 3 describes our role-based sentiment analyzer. The spectral graph transducer is introduced in Section 4. Section 5 shows the experiment results of the proposed sentiment analyzer and further compares SGT with SVM. Finally, we conclude our work in section 6.

2. Related Work

Sentiment Analysis Most prior work on sentiment analysis focused on lexical-based methods [6], [11],
n-gram based methods [12] and syntax-based methods [13], [14]. Usually, the polarity of a review was classified depending on the average semantic orientation of pre-selected words or phrases that are mainly composed of sentiment adjectives, verbs and nouns [15]-[18] reported an inherent problem of sentiment analysis in movie domain, i.e. “the whole is not necessarily the sum of the parts”. Another problem in sentiment analysis, reported in [19], was called ‘thwarted expectations’ narrative. In that paper, three machine learning methods, i.e. Naïve Bayes, Maximum Entropy, and Support Vector Machine, were compared in the form of two features--unigram and bigram. [12] had concerns about the detection on focus of sentences and identification of features indicating whether words or sentences are on-topic. Recently, many papers had discussed semi-supervised models for sentiment analysis. [20] firstly adopted an unsupervised bootstrapping method for separating one review into personal and impersonal views, and then performed a co-training algorithm incorporating unlabeled data for semi-supervised sentiment classification. [21] mining unambiguous reviews through a spectral technique and classifying ambiguous reviews via a combination of active learning, transductive learning, and ensemble learning. Rather than these two-step semi-supervised methods, the SGT in this paper trains the model using labeled data and unlabeled data simultaneously.

**Semantic Role Labeling** Semantic Role Labeling (SRL for short) is a shallow semantic parsing technology, which is defined as a shared task in CoNLL-04 [22]. SRL aims at identifying semantic roles for each target verb (also called predicate) in sentences using machine learning. There are two kinds of roles: one is core role, i.e. Arg1~Arg5, and the other is additional role, i.e. ArgM-Tmp, ArgM-Loc, ArgM-Tpc etc. The core role is basic semantic components of a sentence, such as Arg0 stands for agent, Arg1 stands for theme or patient. Additional role, related to the target verb, is optional. In example B, for instance, He was labeled Arg0 as the subject of the predicate talk about, and the sentiment word cheerfully, as a modifier to the predicate, was labeled ArgM-Adv.

B. [Arg0 He] [ArgM-Adv cheerfully] [Predicate talked out] [Arg1 the arranged little train].

Some adjectives or nouns could be recognized as predicate when the main verb is copulative verb. SRL can be typically divided into two steps: the first is identification and the second is classification. Identification finds out the argument candidates in the constituents in a sentence, and classification assigns the argument to those candidates. There has some works on Semantic Role Labeling in Chinese. [23] utilized transductive SVM and some semantic heuristics to achieve the improvement on Chinese SRL. [24] researched a relationship between opinion holders and topics through SRL. The method is divided into three steps: identifying an opinion-bearing word, labeling semantic roles related to the word and finding the holder and the topic of the opinion word among the labeled semantic roles. The principle of our method is similar with [24] but we uses SRL to search sentiment words relevant to topics. [25] integrated shallow syntactic parsing features and heuristic position information for extracting opinion targets. However, we develop heuristics for extracting topic terms based on simple syntactic information and identification of related sentiment features depending on semantic roles.

### 3. Role-Based Sentiment Analyzer

#### 3.1. Topic Term Detection

In this section, we describe the implement of the role-based sentiment analyzer. A topic term is a word that represents an attribute or a component of a specified topic. For one sandplay test, some basic information, such tester’s name, toys’ names, family etc. are listed in the front of questionnaires. We called them metadata words. Metadata words are mostly talking points of the answers in questionnaires. Intuitively metadata words are important indicators to represent or locate the topics for a given review in questionnaires. For instance, in example C, supposed Tom, a tester's name, is a metadata word,
**performance** could be identified as topic term depending on the syntactic relation between them. We consider two kinds of syntactic relations between candidates of topic terms and metadata words. One is adjunction. If a metadata word is an adjunct to some noun, the noun is identified as the candidate of topic term, seen in example C.

C. **Tom's performance** is wonderful.
D. The **decoration** and the **style** of this room are comfortable to us.

The other is co-ordination. In accordance with a metadata word, this noun was considered to the candidate of feature term. In example D, supposed **style** is a metadata words, **decoration** is a candidate of topic term. After identifying all candidates, we choose top $k$ candidates as topic terms about our questionnaires. In experiments we empirically assigned 200 to $k$ based on cross-validation. The Algorithm 1 details the heuristics of topic term detection.

```plaintext
Algorithm 1 --- Topic Term Detection
Input review set: dataset; nature number: $k$; temporary metadata word set: tempset
Output the feature terms set: termset;
1. termset ← null;
2. Loop: for each review in dataset, do next steps:
   2.1 tempset ← null;
   2.2 obtain the metadata words from this review, and store them into tempset;
   2.3 add termset into tempset;
   2.4 scan the nouns which are satisfied with two syntactic relations with any word in tempset and identify the candidates of feature term.
   2.5 if termset does not contain candidates, add candidates into termset;
3. if termset enlarge, return step 2; else do step 4;
4. Loop: for each candidate in termset, do figure out the frequency of the candidate
6. choose top $k$ candidates and refill them in termset;
7. return termset;
```

After detection of topic terms, a simple filter is developed to collect all on-topic sentences where each one contains at least one term. We use a semantic role labeler [23] to detect all roles and predicates in each on-topic sentence, and store all predicate-role vectors as a semantic representation.

### 3.2. Identification of Related Sentiment Features

We use the Sentiment Lexicon in HowNet [26] to provide basic sentiment words for feature description. HowNet is an on-line Chinese-English bilingual base providing common-sense knowledge. The Sentiment Lexicon consists of sentiment words and adverbs/adjectives of degree in Chinese and English respectively. The detailed information was showed in Table 1.

<table>
<thead>
<tr>
<th>Word Type</th>
<th>Chinese</th>
<th>English</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentiment words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pos</td>
<td>4370</td>
<td>4574</td>
<td>happy, favor</td>
</tr>
<tr>
<td>Neg</td>
<td>4566</td>
<td>4373</td>
<td>aimless, alone</td>
</tr>
<tr>
<td>Adverbs/Adjectives of degree</td>
<td>219</td>
<td>170</td>
<td>Too, very, a little</td>
</tr>
</tbody>
</table>

#### 3.2.1 Related sentiment features

Related sentiment features (**ReSF** for short) mean sentiment words among roles that have a semantic relationship with the roles including topic terms. The polarity of a semantic role was determined through
counting the polarity of included sentiment words defined in HowNet. For example, good music (JJ NN) is a positive role because good (JJ) is a positive sentiment word. The polarity of a predicate-role vector can be determined in the same way. In example A, for instance, the sentiment of the predicate-role <Arg0My boy>-<predicate>likes>-<Arg1sandbox> is positive because the predicate likes is the role relevant to the topic term sandbox and it is positive. There are two noticeable situations: the first one is, when the role is modified by a negative meaning, such as not, never etc. the polarity should be reversed;

E. [Arg0 What I favor in the game] [Predicate is] [Arg1 his concentration].

The second one is, when a role corresponds to a clause, such as IP, CP, the polarity of role should be recursively computed according to sentiment of the sub predicate-role vector. For example, in E, Arg0 is a CP, the polarity of Arg0 is positive because the favor in CP is positive sentiment word. Now we consider the sentiment word distribution and the sentiment-directed role distribution for implementing ReSF identification.

Sentiment Word Distribution. According to our dataset, we summed up all presences of sentiment words in roles and pictured probability distribution of role type in Fig. 2 where verb/predicate means the predicate is regular verb (excludes copulative verb), Adj/nn/predicate represents the predicate is adjective or noun, and no/predicate represents the irregular sentence that has no predicate, such as ‘so cool!’, ‘up!’.

There are nearly 36,000 presences of sentiment words. It is obvious that the whole presence ratio of them in Arg0, verb/predicate, adj/nn/predicate and no/predicate was 80%, and 9.37% in ArgMs (ArgM-Tpc, ArgM-Mnr, ArgM-Adv).

Sentiment-Orientated Role Distribution. Based on sentiment word distribution, we further calibrated the characteristics of sentiment-orientated role distribution. In Table 2, we noted that when sentiment roles, which means sentiment words are included, is core role, most possible role orientated by the sentiment roles is the core role itself; And When sentiment roles are verb/predicate, Adj/nn/predicate and additional roles ArgMs, beyond 80% possibility of roles orientated are Arg1, Arg0 and Arg1+Arg0 respectively.

<table>
<thead>
<tr>
<th>Sentiment roles</th>
<th>roles semantically orientated by sentiment roles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg0</td>
<td>Arg0 (85.2%) Arg1 (15.5%) ...</td>
</tr>
<tr>
<td>Arg1</td>
<td>Arg1 (89.3%) ArgM-Tpc (6.5%) ...</td>
</tr>
<tr>
<td>Arg2</td>
<td>Arg2 (80.4%)</td>
</tr>
<tr>
<td>Verb_predicate</td>
<td>Arg1 (87%) ArgM-Tpc (7.6%) ...</td>
</tr>
<tr>
<td>Adj/nn_predicate</td>
<td>Arg0 (92.2%) ArgM-Tpc (3.3%) ...</td>
</tr>
<tr>
<td>ArgMs</td>
<td>Arg1 (45.9%) Arg0 (38.1%) ArgM-Tpc (5.6%) ...</td>
</tr>
</tbody>
</table>

3.2.2 Identifying related sentiment features

Based on Fig. 2 and Table 2 we implemented an algorithm to identify the ReSF. For each sentiment role...
and predicate (collectively called source role) in a sentence, the algorithm is divided into two steps. The first step is locating the target role which the sentiment of source role is directed to. The second step is determining whether the source role is relevant or not based on semantic relationship between target role and topic terms.

Algorithm 2 --- Identification of Related Sentiment Features

**Input** role: source role, target role; sentiment predicate-roles vector contains source role; vector;  
**Output** Boolean yes or not  
1. target role ← null;  
2. if type of source role is predicate  
   2.1 if source role is adj/nnPredicate  
      2.1.1 if vector’s Arg0 existed,  
           target role ← Arg0;  
      2.1.2 else if vector’s ArgM-Tpc existed, target role ← ArgM-Tpc;  
   2.2 else if source role is verbPredicate  
      2.2.1 if vector’s Arg1 existed,  
           target role ← Arg1;  
      2.2.2 else if vector’s ArgM-Tpc existed, target role ← ArgM-Tpc;  
3. if type of source role is role  
   3.1 if source role is Arg0, Arg1 or Arg2,  
        target role ← source role  
   3.2. else if source role is ArgMs  
      3.2.1 recall the algorithm 2 using the vector’s predicate as input  
      3.2.2 target role ← the result of 3.2.1;  
4. obtain the headword of constituent corresponding to the target role;  
5. if topic terms contain headword, return yes, else return no;

Compared to an analogy work in [7], algorithm 2 has two advantages: Firstly, it exploits the inherent characteristics of sentiment words through semantic roles. In example F and G, it is easy to recognize dress and plots as the target which loves and far-fetched is directed to according to the role types Theme and ArgM-Tpc. However, [7] used lexical-based sentiment information, which cannot represent directly the relationship between a sentiment word and its target.

F. [Arg0 She] [Predicate loves] [Theme the beautiful dress of cartoon character] [ArgM-Adv very much].

G. As to [ArgM-Tpc the plots of those stories narrated], [Arg0 Most persons] would [Predicate feel] [Arg1 a little far-fetched].

H. [Arg0 John] [Predicate performs] [ArgM-Adv clumsily].

Secondly, our algorithm is more robust than lexical method practically. In example H. Based on a fact in SRL that all additional roles are viewed as the modifiers of the predicate, our algorithm can automatically transfer the target that the predicate performs is directed to the target of the sentiment additional role clumsily.

4. Spectral Graph Transducer

Spectral Graph Transducer (SGT for short) [9] is transductive learning based on K nearest-neighbor classifier. Different from inductive learning such as SVM, a transductive learner exploits the structure or the distribution of test data when training and decoding this information into the model. Since it is difficult of acquiring a large amount of tagged data in practice, the global models are limited if tagged data are not
enough to search target function in the hypothesis space. Compared to other transductive version, the key advantage of SGT is that it uses spectral methods to solve globally optimization problem, rather than running a greedy search. Here we introduce the algorithm briefly, and the detail is referred to [09]. Typically, a SGT consists of two steps for a classification task given training set and test set. The training set is represented by an undirected graph $L$ stored in matrix form.

**Pre-process.** This step computes the nearest neighbor graph according to similarity-weighted $K$ nearest-neighbor graph and set $d$ eigenvectors of the (normalized) Laplacian for $L$, where $K$ means the number of neighbors considered for a node in $L$ and $d$ means the number of eigenvalues / eigenvectors to extract from the Laplacian. As the optional parameters, $K$ and $L$ could be specified by the user. The model stored in this step could be used to a new classification task.

**Classification.** This step implements a process of transductive learning. The test set and training set (or other labeled dataset) are transformed into vectors. Then, the label value of each test vector is computed according to the obtained model. Note that, a parameter $C$, which trade-offs between training error and complexity, could be specified by a user.

5. **Experiments**

In this section, we evaluate the role-based sentiment analyzer (SA) described above. The experiments on sandplay questionnaires consist of Test 1 and Test 2.

5.1. **Dataset and Semantic Role Labeler**

An application of our sentiment analyzer is to automatically determine the polarity of sandspiel questionnaires. The questionnaires and metadata corresponding were collected from the Child Behavior Laboratory where researches behavior problems of preschool children. We designed the templates and obtained about 3,200 questionnaires coming from children having psychological disorders and normal children during 3 years. We develop a semantic role labeler with 10 features (listed in Table 3). We used a SVM classifier toolkit SVMLight [27] (linear kernel, parameter $C$ is set to 2), which is an implementation of SVM in Program C. The classifier was trained on Chinese PropBank\(^3\) 1.03 using 10 cross-validations. We chose Stanford Chinese Parser for syntactic parse. As a start-of-art parser in Chinese, Stanford Parser integrated three statistical parsers (i.e. a PCFG parser, a dependency parser, and a lexical-factored parser). The performance of parser and semantic role labeler (using automatic parse) were listed in Table 4.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path</td>
<td>The syntactic path from constituent parsed to the predicate</td>
</tr>
<tr>
<td>Phrase Type</td>
<td>The phrase category the argument corresponds to</td>
</tr>
<tr>
<td>Predicate</td>
<td>The verb lemma</td>
</tr>
<tr>
<td>Position</td>
<td>The phrase is before or after the predicate</td>
</tr>
<tr>
<td>Subcat-Frame</td>
<td>The rule that expands the parent of the predicate verb</td>
</tr>
<tr>
<td>Phrase type of the sibling to the left</td>
<td>The phrase category of the sibling to the argument in left</td>
</tr>
<tr>
<td>Head Word and Part of Speech</td>
<td>The syntactic head of the phrase</td>
</tr>
<tr>
<td>First and last word of the constituent parsed</td>
<td>First and last word of phrase corresponding to the argument</td>
</tr>
<tr>
<td>Syntactic Frame</td>
<td>The syntactic frame consists of the NPs surrounding predicate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parser</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stanford Parser</td>
<td>84.74</td>
<td>84.62</td>
<td>84.68</td>
</tr>
<tr>
<td>Syntactic Frame</td>
<td>74.33</td>
<td>75.02</td>
<td>74.68</td>
</tr>
</tbody>
</table>
5.2. Features

We use the following feature framework in the machine learning algorithms, i.e. SVM and SGT: (1) **Bag-of-Words framework.** Let \( \{ f_1, \ldots, f_m \} \) be a predefined set of features that appear in a review. Each review is represented by the vector \( v = (n_1, n_2, \ldots, n_m) \) where \( n_i \) is assigned to 1 if and only if \( f_i \) appears in the review, or 0 if it does not appear. There are three feature representations according to the sources: sentiment lexicon in HowNet, sentiment word extracted in on-topic sentences based on Algorithm 1, related sentiment words extracted by Algorithm 1+2. (2) **Positive X Percentage framework.** The principle of this feature framework comes from [10] where the Positive Sentence Percentage is used as the effective review feature. Here, the \( X \) is designed to represent on-topic sentences extracted by Algorithm 1 or related sentiment vectors extracted by Algorithm 1+2. The all features used for experiments are listed as follows:

- Basic Features with Sentiment Words (**BFSW**). Bag-of-words using sentiment words come from Sentiment Lexicon.
- Topic Features with Sentiment Words (**TFSW**). Bag-of-words using sentiment words in all on-topic sentences based on Algorithm 1.
- Related Features with Sentiment Words (**RFSW**). Bag-of-words using related sentiment words extracted by Algorithm 1+2.
- Positive Sentence Percentage (**PSP**). defined in [Pang 2005].
- Positive on-Topic Sentence Percentage (**PTSP**). The percentage of number of positive on-topic sentences divided by number of on-topic sentences in a review.
- Positive Related Vector Percentage (**PRVP**). The percentage of number of positive related vectors divided by number of relevant vectors.

5.3. Results and Discusses

Table 5. Results of the Evaluation on BFSW, LFSW and RFSW (5 Average Sentences per Review)

<table>
<thead>
<tr>
<th># training data</th>
<th>BFSW</th>
<th>TFSW</th>
<th>RFSW</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>70.11</td>
<td>67.55</td>
<td>64.13</td>
</tr>
<tr>
<td>200</td>
<td>72.05</td>
<td>70.12</td>
<td>68.44</td>
</tr>
<tr>
<td>500</td>
<td>75.50</td>
<td>75.10</td>
<td>74.12</td>
</tr>
<tr>
<td>1000</td>
<td>77.42</td>
<td>78.76</td>
<td>77.23</td>
</tr>
<tr>
<td>5000</td>
<td>80.19</td>
<td>83.31</td>
<td>81.33</td>
</tr>
<tr>
<td>10000</td>
<td>84.50</td>
<td>86.23</td>
<td>84.55</td>
</tr>
</tbody>
</table>

**Test 1.** In order to investigate whether the ReSF identified have influence on the performance of sentiment analysis in psychological questionnaires, we evaluate TFSW with BFSW and RFSW. Different number of reviews was extracted randomly from our dataset as tagged training data and other 1000 reviews are held for test data. Table 5 showed the results (F1 value). It could be found that the BFSW bettered than TFSW and RFSW when number of tagged reviews is less than 500. The reason we considered was that the features of RFSW and TFSW can be sparse in small training data. (BFSW has a fixed number of features). As the number beyond 500, the performance of TFSW and RFSW improved rapidly. To our surprise, the performances of RFSW lower than TFSW in whole test. Observing the reviews which were classified incorrectly, we included two characteristics. One is that most of these reviews contained small sentences (more than half of them just contained one or two sentences). The other characteristic is that large spoken languages and incomplete sentences exist, which may result in a low accuracy on syntactic parse, labeling roles and the identification of relevant feature words. We then chose the different number of reviews that contain more than 10 sentences and retrain model. As we can see in Table 6, the RFSW
outperforms BFSW and TFSW even though the number of tagged training data was small. Note that BFSW in Table 5 and 6, where the performance in long reviews lower than it in small reviews (column 1 in Table 5 and 6). It indicates that the noise (i.e. the sentiment information has not relevant to the topic) increased along with enlargement of review size. Test 1 suggested that it is necessary to research on parsing spoken languages and non-complete sentences for improving sentiment analysis.

Table 6. Results of the Evaluation on BFSW, TFSW and RFSW (15 Average Sentences per Review)

<table>
<thead>
<tr>
<th># training data</th>
<th>BFSW</th>
<th>TFSW</th>
<th>RFSW</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>67.54</td>
<td>70.11</td>
<td>68.30</td>
</tr>
<tr>
<td>200</td>
<td>69.66</td>
<td>73.32</td>
<td>74.10</td>
</tr>
<tr>
<td>500</td>
<td>71.56</td>
<td>77.15</td>
<td>79.80</td>
</tr>
<tr>
<td>1000</td>
<td>74.11</td>
<td>79.79</td>
<td>83.80</td>
</tr>
<tr>
<td>2000</td>
<td>77.77</td>
<td>83.01</td>
<td>86.61</td>
</tr>
<tr>
<td>4000</td>
<td>80.57</td>
<td>86.22</td>
<td>88.39</td>
</tr>
</tbody>
</table>

Test 2. In order to investigate the learning performance of SGT and SVM in small tagged data and the effectivity of PRVP, we evaluated PRVP with PTSP and PSP in different number of tagged data. About 5000 reviews were held for test data. Table 7 showed the results. The SGT performed better than SVM (raising about 3%~5%) when tagged training data is small (less than 200) in PSP, PTSP and PRVP. It is because that SGT made effectively use of the distribution information of test data. Unfortunately, the SGT did not do as well with large tagged data comparing to SVM. It maybe resulted from the features presented in this paper is not accurate enough to capture the characteristics of questionnaires. In addition, PRVP outperforms PTSP about 1%. It proved that PRVP is more effective than PTSP on feature representation of reviews, which encourages us to develop more characteristics of reviews in the future.

Table 7. Results of Evaluation on PSP, PLSP and PRVP Using SGT and SVM (Holding 5000 Test Data)

<table>
<thead>
<tr>
<th># training data</th>
<th>SVM</th>
<th>SGT</th>
<th>SVM</th>
<th>SGT</th>
<th>SVM</th>
<th>SGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>62.90</td>
<td>65.09</td>
<td>62.19</td>
<td>64.13</td>
<td>61.06</td>
<td>61.39</td>
</tr>
<tr>
<td>200</td>
<td>63.65</td>
<td>68.90</td>
<td>64.00</td>
<td>69.28</td>
<td>64.90</td>
<td>70.34</td>
</tr>
<tr>
<td>500</td>
<td>67.01</td>
<td>70.99</td>
<td>67.51</td>
<td>71.54</td>
<td>69.55</td>
<td>71.97</td>
</tr>
<tr>
<td>1000</td>
<td>70.06</td>
<td>71.00</td>
<td>71.30</td>
<td>71.32</td>
<td>72.87</td>
<td>72.54</td>
</tr>
<tr>
<td>2000</td>
<td>75.56</td>
<td>71.12</td>
<td>75.50</td>
<td>72.78</td>
<td>76.31</td>
<td>73.91</td>
</tr>
<tr>
<td>4000</td>
<td>77.50</td>
<td>72.10</td>
<td>77.98</td>
<td>73.41</td>
<td>79.43</td>
<td>74.10</td>
</tr>
</tbody>
</table>

6. Conclusion and Future Work

Identifying the topic-specific features has influences on the precision and performance of sentiment analysis. This paper introduced a role-based sentiment analyzer to identify automatically relevant sentiment roles. We use Spectral Graph Transducer to improve the performance of sentiment analyzer in small tagged training data. The experiments demonstrated our analyzer performed well in bag-of-word framework comparing to the methods based on pre-selected sentiment words and on on-topic sentiment words. In addition, the evaluation in Positive X Percent-age framework proves that SGT achieves the better performance than SVM in small tagged data. It could be concluded that role-based analyzer deal more properly with the long reviews. It is necessary to research on spoken languages and non-complete sentences for sentiment analysis. In the future, we will develop more effective features of reviews for improve the performance of our system.
Acknowledgment

This work is supported by The Ministry of Education of Humanities and Social Science Project (No. 17YJCZH028).

References


Madrid, Spain.


Chen Yaodong was born in 1978. He received a Ph.D. degree from Hunan University in 2014. He is currently an assistant professor at Changsha Normal University, Changsha, China. His current research interests include natural language processing, pattern recognitions, and computer vision. He is a member of Chinese Computer Society.

Peng Diefei was born 1967. She received a Ph.D. degree from Hunan Agricultural University in 2010. She is currently a professor at Changsha Normal University, Changsha, China. Her current research interests include smart tourism, tourism information management.
Liu Qin was born in 1976. She received a bachelor degree from Hunan Normal University in 1998. She is currently a vice professor at Hunan Provincial Research Institute of Education, Changsha, China. Her current research interests include educational information, educational technology.