Sentence Selection for Language-gap Reduction in Cross-lingual Sentiment Classification

Xinliang Zhao[1,a) Atsushi Fujii[1]

Abstract: Performance of sentiment classification is usually limited by the lack of sentiment resources. Thus cross-lingual sentiment classification techniques have become important to leverage sentiment resources in one (source) language to another (target) language for improving the sentiment classification performance on target language. By far, instance level transfer learning based methods and co-training method are often referred as the most effective mechanism for cross-lingual sentiment classification. However, none of the approaches modeled the language gap or reduced it directly for performance improvement. In this paper, we explicitly model the language gap as the difference between term-frequency distributions of two languages. Then we propose a sentence selection approach to reducing the modeled language gap directly. The evaluations on the NLP&CC 2013 CLSC dataset show the effectiveness of our proposed approach, which can outperform the widely used standard inductive baseline and state-of-the-art systems.

Keywords: Cross-lingual Sentiment Classification, Sentence Selection, Language-gap Reduction, Task-Based Summarization

1. Introduction

Due to the development of social networking services and E-commerce, a huge number of texts like spontaneous submissions to twitter and product reviews has been published onto the Internet. Those texts usually contain valuable subjective opinions that can be used for providing better services and products, etc. Thus, there comes a need to automatically analyze the sentiment behind the texts. In recent years, much research has been focusing on sentiment classification in the natural language processing field. Sentiment classification is the task of judging the sentiment polarity of a given text. It can be applied to many useful aspects, such as opinion mining and summarization [7], [8], [10]. Thus far, supervised learning methods have been quite successful for sentiment classification. However, supervised methods rely heavily on labeled training corpus and are usually limited by the lack of sentiment resources. This problem is more serious in some resource-poor languages due to the imbalanced development of NLP in different languages. To solve this problem, cross-lingual sentiment classification (CLSC) methods are investigated. The main idea of those methods is to transfer the sentiment resources in one (source) language (SL) to another (target) language (TL) so that the TL could get richer resources, thus improving the sentiment classification performance on TL.

Usually, machine translation services are employed when leveraging labeled corpora and sentiment resources across languages. For example, there exist pilot studies directly making use of machine-translated examples to train an inductive classifier, such as SVM (support vector machine) and Naive Bayes classifier, for sentiment classification in other languages [2], [9]. However, due to the insufficient quality of machine translation systems, a language gap[*1] is caused between the original language and the translated language, making the classification performance not satisfactory. Realizing the gap, many existing systems which transfer high quality examples from unlabeled TL dataset to help training have been proposed to improve the performance [3], [12], [15] (See Section 2). However, none of those approaches modeled the language gap or reduced it directly for performance improvement.

Target to this aspect, this paper explicitly models the language gap as the difference between term-frequency distribution of SC and that of TL. Based on this modeled language gap, we then propose a sentence selection approach to reducing it directly for improving accuracy of CLSC. Besides, different from those existing systems, our approach does not iteratively transfer samples for training classifiers many times, which in some sense is more efficient and time-saving.

In this paper, we use English and Chinese for explanation purpose. Machine Translation is used and unlabeled Chinese samples are translated into English side to generate a translated Chinese term-frequency distribution in the proposed approach. After that, the generated distribution is used for adjusting the sentiment scores calculated from Whissell's Dictionary of Affect in Language [13] to boost the score of the word with high frequency in translated Chinese distribution. Then sentence selection approach can fully leverage the generated distribution and adjusted sentiment scores for selecting sentences to reduce the language gap modeled as difference between term-frequency distributions in this research. The SVM classifier is adopted as the basic clas-

[1] Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan
[a)] zhao.x.ac@m.titech.ac.jp

[*1] By the language gap we meant the difference between two languages.

sifier. The evaluations on NLP&CC 2013 CLSC dataset show the effectiveness of our proposed approach, which can outperform the widely used standard inductive baseline and state-of-the-art systems. Note that even if the evaluations are carried out on English-Chinese dataset, the proposed approach can be generalized to any combination of two languages.

The rest of this paper is organized as follows: Section 2 surveys past research on CLSC. Section 3 describes in detail our proposed sentence selection approach for cross-lingual sentiment classification. Section 4 gives the experimental results and discussions. Finally, we conclude and give some insights for future work in Section 5.

2. Related Work

By far, instance level transfer learning based methods and co-training method are often referred as the most effective mechanism for CLSC. Instance level transfer learning based methods try to transfer high quality training examples for improving the TL sentiment classification rather than use the whole translated training data. Dai et al. [3] proposed transfer AdaBoost algorithm that uses boosting-like strategy to enhance the weights of TL samples and down-weight the incorrectly classified translated examples iteratively and improved the system performance. Xu et al. [15] extended the transfer AdaBoost algorithm to handle multi-category problem and proposed a new Transfer Self-training algorithm to iteratively select high quality translated examples to enrich the training data set for improving the CLSC performance. The typical co-training approach for CLSC was introduced by Wan et al [12]. The co-training approach considers English and Chinese features as two independent views of the classification problem and makes use of unlabeled Chinese data for improving the sentiment classification performance on both languages.

3. The Sentence Selection Approach

3.1 Overview

The purpose of our approach is to make use of labeled training examples in SL and unlabeled raw samples in TL for improving the performance of CLSC, without using any labeled samples in TL. Given the labeled training examples in SL, one straightforward model [2], [9] which is also referred as unigram model can be used to handle this problem. This model first learns a classifier based on unigram term-frequency vectors generated from labeled training examples in SL. Then samples that need to be classified in TL are translated into SL. Lastly, the model classifies the translated samples also based on unigram vectors. Although the straightforward model can somehow solve the cross-lingual problem, the result is not so promising due to the language gap modeled as the difference of term-frequency distributions between languages in this study.

To deal with the modeled language gap, our proposed sentence selection approach chooses appropriate sentences from each training example so that the gap can be reduced, thus improving the classification accuracy. The choices made are based on the term-frequency distribution generated from machine-translated unlabeled raw TL data and sentiment scores adjusted by the distribution. The framework of the proposed approach is illustrated in **Fig. 1**. Details of each component will be discussed in the following subsections.

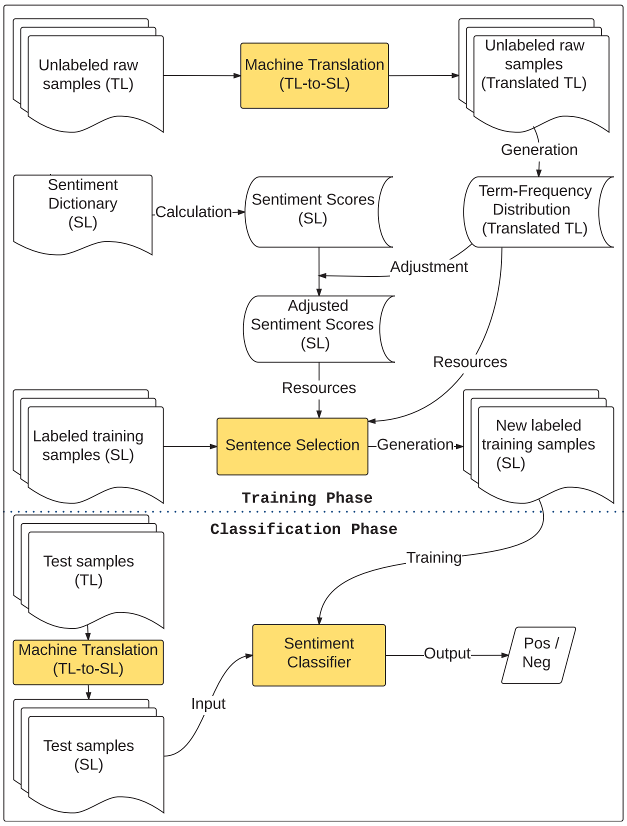
Unlabeled raw samples (TL)
Machine Translation (TL-to-SL)
Unlabeled raw samples (Translated TL)
Generation
Sentiment Dictionary (SL)
Calculation
Sentiment Scores (SL)
Term-Frequency Distribution (Translated TL)
Adjustment
Adjusted Sentiment Scores (SL)
Resources
Resources
Labeled training samples (SL)
Sentence Selection
Generation
New labeled training samples (SL)
Training Phase
Classification Phase
Test samples (TL)
Training
Machine Translation (TL-to-SL)
Sentiment Classifier
Output
Pos / Neg
Input
Test samples (SL)

Fig. 1 Framework of the proposed approach

The framework is composed of a training phase and a classification phase. In the training phase, the input are labeled training samples in SL and unlabeled raw samples in TL. Unlabeled samples are first translated into SL by using machine translation services. Then a term-frequency distribution can be generated based on the translated unlabeled samples by counting time of appearance of each word. Usually, if summaries exist, the words contained in them should be counted more since summaries convey more information than normal text. This distribution is normalized in the way that term-frequency of each word is divided by the maximum among them, so that it can be used to adjust the sentiment scores scaling from 0 to 1 and cooperate with it for sentence selection in later steps. Sentiment scores are calculated from sentiment dictionary in SL and adjusted by generated distribution to boost the score of the word with high term-frequency in the distribution. After that, sentence selection approach can fully leverage the generated distribution and adjusted sentiment scores for selecting sentences from each labeled training sample to reduce the language gap modeled as the difference between term-frequency distributions in this research. Selected sentences of each sample are treated as a new sample, and thus an identical number of new samples should be obtained from the original labeled training samples on the SL side. Finally, a classifier (e.g. SVM, NB) can be learned using the newly obtained samples. In the classification phase, each unlabeled test sample in TL is first translated into SL, and then the learned classifier on the SL side is applied to classify the translated test sample into either positive or

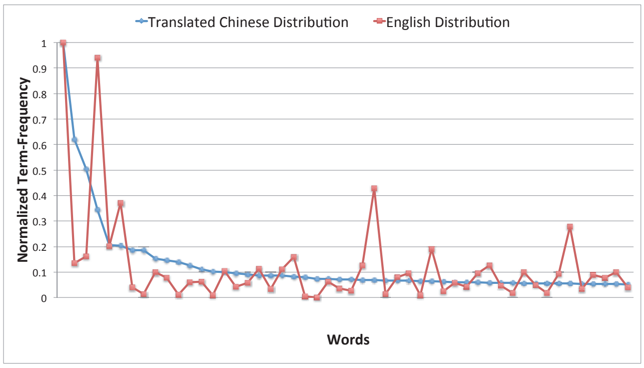

Fig. 2 Difference of term-frequency distribution between *English* and *Translated Chinese*

negative. Although we use a specific classifier in our experiments (see Section 4), in principle, our sentence selection approach is independent of the classification phase, and the newly generated samples by our approach can be used to train any advanced CLSC model.

3.2 Machine Translation

In the context of CLSC, we must translate data samples from one language to another language for further processing. So far, many machine translation techniques have been developed to achieve this goal, even though the translation quality of these techniques is not satisfying. A few machine translation services are publicly available and can be used for scientific research, e.g. *Google Translate*[*2], *Yahoo Babel Fish*[*3] and *Windows Live Translate*[*4]. [12] In this study, we adopt *Google Translate* for translating Chinese samples into English because it is one of the well-known state-of-the-art commercial machine translation systems.

3.3 Language Gap Modeling

Although one language can be machine-translated into another language to make them comparable, a language gap is caused between the original language and the translated language due to the insufficiency of machine translation quality. In this study, we explicitly model the language gap as the difference of term-frequency distributions between languages. Here, an example is given in **Fig. 2** illustrating this difference. The example is generated using the labeled English training data and unlabeled Chinese raw data in the book review category of NLP&CC 2013 CLSC dataset. The calculated term-frequency is normalized in the way that term-frequency of each word is divided by the maximum among them, so that the two distributions are comparable.

3.4 Sentiment Score Adjustment

It is intuitive to see that sentiment resources like sentiment dictionaries are helpful for sentiment classification task since sentiment words usually contain more sentiment information than other words. So far, many sentiment analysis lexicons have been developed in English, e.g. Whissell's Dictionary of Affect in Language [13], SentiWordNet [4], MPQA Subjectivity Lexicon [14] and Bing Liu and Minqing Hu Sentiment Lexicon [6]. Much research has successfully leveraged these sentiment analysis lexicons for improving the performance of sentiment classification. For example, Agarwal et al. used Whissell's Dictionary of Affect in Language to construct a set of new features for improving the performance of sentiment classification on twitter data in English [1]. In this paper, Whissell's Dictionary of Affect in Language which gives scores of pleasantness of English words is employed. The scores that are real numbers in the dictionary range from 1 to 3, where 1 represents unpleasant, 2 means neutral and 3 denotes pleasant. We then take absolute value of each score subtracted by 2 to measure the sentiment strength of each word, ignoring the sentiment polarity of the word. Thus, the sentiment score now scale from 0 to 1, representing the strength of sentiment of each word.

However, without adjustment, sentiment scores do not work well for the proposed sentence selection approach, because the scores are statistically based on English, causing that the selected sentences could not reduce the modeled language gap. Thus, adjustment is carried out using term-frequency distribution of translated TL mentioned above to boost the score of the word with high term-frequency in the distribution. In this research, we simply add together the sentiment score of each word and the normalized term-frequency in the distribution. Then we multiply the result with factor α which indicates the relative weight between sentiment scores and normalized term-frequency in the following sentence selection step. Letting $s(w)$ denote sentiment score and $q(w)$ denote normalized term-frequency of word w, the formula is shown as below:

$$s_{after}(w) = \alpha \times (s_{before}(w) + q(w)) \qquad (1)$$

It is worth noting that since the normalized term-frequency of most strong sentiment words in the translated TL distribution usually lies in the range under 0.01, say, the normalized term-frequency of sentiment words have different magnitudes when being used to adjust the sentiment scores, the factor α usually need to be assigned different values for different magnitudes. This is to give different relative weights for handling the problem brought by different magnitudes. Another solution for this problem is to use only a particular range of normalized term-frequency and adjust only the scores of sentiment words whose term-frequency lie in that range. In this case, the factor α is assigned only one value since we are dealing with only one magnitude. By adjusting sentiment scores of words whose normalized term-frequency lie in the range containing most sentiment words, we can achieve really good performance for CLSC. In the following sections, this solution is taken.

3.5 Sentence Selection

The sentence selection algorithm is motivated by Frequency-driven Approaches for extractive text summarization. One system called SumBasic proposed by Vanderwende et al. [11] utilizes word probability computed from the input for sentence selection. For each sentence, an importance score is calculated by taking average probability of the content words. The system then iteratively picks the best scoring sentences in a greedy manner until the desired summary length is achieved.

[*2] http://translate.google.com/translate_t
[*3] http://babelfish.yahoo.com/translate_txt
[*4] http://www.windowslivetranslator.com/

One important aspect of the Frequency-driven Approaches is that we can define our own sentence scoring function for specific task with continuous word probability. In the context of CLSC, here, we use normalized term-frequency $q(w)$ of word w generated from translated raw TL data instead of word probability. As explained in previous sections, it is calculated as the number of occurrences, $c(w)$ in the corpus, T of a word, w divided by the maximum occurrences among all words:

$$q(w) = \frac{c(w)}{\max_{w_i \in T} c(w_i)} \tag{2}$$

Then based on normalized term-frequency q and adjusted sentiment scores s_{after}, we assign each sentence S_j in SL training data a weight using sentence scoring function defined as below:

$$Weight(S_j) = \sum_{w_i \in S_j} q(w_i) + s_{after}(w_i) \tag{3}$$

Here, we let $q(w_i)$ or $s_{after}(w_i)$ be 0 if the word w_i can not be found in the translated TL or sentiment dictionary. After that, we select a number of sentences with best scores from each labeled training sample in SL. Without losing generality, half of the sentences in each sample are selected in this paper. Selected sentences of each sample are treated as a new sample, and thus an identical number of new samples should be obtained from the original labeled training samples on the SL side. Different from the usual extractive text summarization whose objective is to create a summary that retains the most important points of the original document, the sentence selection algorithm is task-oriented and those newly generated samples will only serve for the CLSC training phase instead of maintaining contents.

3.6 A Theoretical Intuition

To get an intuition of why our proposed approach can work by directly reducing the modeled language gap in theory, we may think of the principle behind the classifiers. Here, the inductive classifiers, in particular, the SVM classifier is used for explanation because it is also employed in our following experiments.

The hypothesis of SVM classifier can be written as:

$$h_\theta(x) = \begin{cases} 1 & \texttt{if}\, \theta^T x \geqslant 0 \\ 0 & otherwise \end{cases} \tag{4}$$

Where θ is a parameter vector that need to be trained and x is the vector of a data sample. Through training phase, the parameter vector is fixed and thus a decision boundary should be obtained.

In the context of CLSC, if we simply leverage the original samples in SL for training a classifier, in the classification phase, there will be more uncertainty for the sign of $\theta^T x$, thus more uncertainty for the classification results due to the modeled language gap.

It can also be said like this: Because of the difference of term-frequency distribution between original language and translated language, even if the terms contained in a translated test sample are the same as those contained in training samples, the result of $\theta^T x$ becomes more unexpectable in the classification phase. However, by directly reducing this modeled language gap in the training phase, the result is more expectable and more likely to be similar to those of training samples in the classification phase. Thus our proposed approach can predict the sentiment polarity more correctly and achieve better performance for CLSC.

4. Experiments

4.1 Experiment Settings

The proposed sentence selection approach is evaluated on NLP&CC 2013 CLSC dataset[*5]. This dataset consists of product reviews in three different categories, namely, DVD, Book and Music. In each category, there are 4,000 labeled English reviews whose ratio between positive and negative samples is 1:1. An unlabeled corpus is composed of 17,814 DVD reviews, 47,071 Book reviews and 29,677 Music reviews in Chinese. For the test dataset, each category contains 4,000 labeled Chinese reviews. Although this dataset also provide 40 labeled Chinese reviews in each category for training, in this study, we did not use them because we want our approach to be more general even for target languages without any labeled data. The performance is evaluated by the classification accuracy for each category, and the average accuracy of the three categories, respectively[*6].

The category accuracy is defined as:

$$Accuracy_c = \frac{\#correctly\ classified\ samples\ in\ category\ c}{4000} \tag{5}$$

Where c is one of the three categories, namely, DVD, Book or Music category. The overall accuracy is defined as:

$$Accuracy = \frac{1}{3} \sum_c Accuracy_c \tag{6}$$

In this experiment, only Chinese-to-English translation is needed since all language processing work is done on the English side. It is a merit because lots of language resources in English could be leveraged in this situation. The monolingual sentiment classifier used is SVM^{light}[*7] and only word unigram features are employed in this study.

4.2 Baseline Methods

In the experiments, the proposed sentence selection approach is compared with the following three baseline methods.

Uni: This is the widely used unigram model described in previous sections. In more detail, it directly uses the labeled English data as training samples with summary part of each sample given more weights when generating unigram vectors. The stopwords are eliminated and remaining words are stemmed. Unlabeled Chinese data are not used in this model.

Uni(Neg): This method is basically the same as **Uni** unless it does not exclude all the stopwords. The negation words such as ”no” and ”not” are left to introduce negation features for better negative case classification.

Uni(Neg)+SenSel(Random): This method is similar to the proposed sentence selection approach. But instead of selecting sentences based on the scores computed, it selects sentences randomly. After getting the new set of labeled English training samples, it proceeds the same way as **Uni(Neg)**.

[*5] http://tcci.ccf.org.cn/conference/2013/dldoc/evdata03.zip
[*6] http://tcci.ccf.org.cn/conference/2013/dldoc/evres03.pdf
[*7] http://svmlight.joachims.org/

Table 1 Comparison with Baseline Methods

Method	DVD(%)	Music(%)	Book(%)	Accuracy(%)
Uni	72.65	71.13	72.81	72.20
Uni(Neg)	78.92	74.22	76.91	76.68
Uni(Neg)+SenSel(Random)	79.36	73.79	75.74	76.30
Uni(Neg)+SenSel	**80.02**	**75.30**	**78.26**	**77.86**
Uni(Neg)+SenSel(TD)	79.96	75.25	78.34	77.85

Table 2 Comparison with NLP&CC 2013 CLSC Evaluation Results

Team	DVD(%)	Music(%)	Book(%)	Accuracy(%)
BUAA	48.05	50.30	49.78	49.38
BISTU	64.73	66.05	59.80	63.53
HLT-Hitsz	77.73	75.13	78.50	77.12
THUIR-SENTI	73.90	73.25	74.23	73.79
SJTUGSLIU	77.20	74.53	72.40	74.71
LEO_WHU	78.33	75.95	77.00	77.09
Our Approach	**80.02**	**75.30**	**78.26**	**77.86**

4.3 Compared with Baseline Methods

In the experiments, we first compare the proposed sentence selection approach with the 3 baseline methods. The comparison results are shown in **Table 1**.

It can be seen from the table that our proposed sentence selection approach outperforms all three baselines based on the accuracy metric. By carrying out paired student's t-test on each category between our proposed approach and each of the three baselines, we can see that the difference on accuracy is statistically significant at the 1% level. Especially, our approach **Uni(Neg)+SenSel** can improve the baseline **Uni(Neg)** from 76.68% to 77.86% by directly reducing the modeled language gap. Also, from the comparison between our proposed approach **Uni(Neg)+SenSel** and **Uni(Neg)+SenSel(Random)**, we can conclude that selecting sentences randomly could not help improve the performance because the modeled language gap will probably not decrease in this case.

Besides the baselines, we also add another method **Uni(Neg)+SenSel(TD)** to compare with. This method is similar to the proposed sentence selection approach. But instead of selecting sentences based on the scores computed partially from the term-frequency distribution generated from translated unlabeled Chinese data, it selects sentences based on the scores calculated in part from the distribution built from translated labeled Chinese test data by ignoring their labels. After getting the new set of labeled English training samples, it proceeds the same way as **Uni(Neg)**. By using the translated test data distribution, nearly the same performance is achieved, because distribution generated from translated Chinese test data is similar to that generated from translated unlabeled Chinese data. This result can further prove the effectiveness and correctness of the proposed sentence selection approach.

4.4 Compared with Other Evaluation Results on NLP&CC 2013 CLSC Dataset

In the second set of experiments, we compare our proposed approach with official results in NLP&CC 2013 CLSC evaluation task and the result is shown in **Table 2**.

Note that among the 6 participants, HLT-Hitsz achieved the best performance on accuracy, using the co-training method in transfer learning [5]. However, unlike HLT-Hitsz system, without using any labeled Chinese samples, our proposed approach further improves the overall accuracy performance. This means that our approach is not only better on performance, but also more generic.

4.5 Language-gap Reduction

In order to quantify the language gap, we employ KL divergence which is a non-symmetric measure of the difference between two probability distributions P and Q.

For discrete probability distributions P and Q, the KL divergence of Q from P is defined to be:

$$D_{KL}(P\|Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)} \tag{7}$$

Due to the asymmetry, we take the average of $D_{KL}(P\|Q)$ and $D_{KL}(Q\|P)$. By calculating the KL divergence before and after our proposed sentence selection approach is applied, we find that there is an average decrease of 1.2% in language gap.

5. Conclusion and Future Work

This paper proposes a sentence selection approach for cross-lingual sentiment classification, which can directly reduce the explicitly modeled language gap, without using any labeled samples in target language. Different from the mainstream instance level transfer learning based methods and co-training method, our approach does not iteratively transfer samples for training classifiers many times, which in some sense is more efficient and time-saving. Experiments on NLP&CC 2013 CLSC dataset show effectiveness of our proposed approach. What is more, the proposed approach can serve as a preprocessing step, which could be combined with other developed methods for cross-lingual sentiment classification. The newly generated samples from our approach might be more effective for training since the language gap is reduced, thus improving the performance of developed methods. This aspect is worth exploring in the future research.

In this paper, in order to prove the effectiveness of our approach without bias, for each sample in labeled source language training data, we select half of its sentences. However, some selected sentences can not actually reduce the language gap, because they are selected just to make up the number. Therefore, the strategy for accurately filtering out those sentences and selecting exactly the sentences needed is the important problem to be solved in our future study.

References

[1] Agarwal, A., Xie, B., Vovsha, I., Rambow, O. and Passonneau, R.: Sentiment analysis of twitter data, *Proceedings of the Workshop on Languages in Social Media*, Association for Computational Linguistics, pp. 30–38 (2011).

[2] Banea, C., Mihalcea, R., Wiebe, J. and Hassan, S.: Multilingual subjectivity analysis using machine translation, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 127–135 (2008).

[3] Dai, W., Yang, Q., Xue, G.-R. and Yu, Y.: Boosting for transfer learning, *Proceedings of the 24th international conference on Machine learning*, ACM, pp. 193–200 (2007).

[4] Esuli, A. and Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining, *Proceedings of LREC*, Vol. 6, Citeseer, pp. 417–422 (2006).

[5] Gui, L., Xu, R., Xu, J., Yuan, L., Yao, Y., Zhou, J., Qiu, Q., Wang, S., Wong, K.-F. and Cheung, R.: A mixed model for cross lingual opinion analysis, *Natural Language Processing and Chinese Computing*,

Springer, pp. 93–104 (2013).

[6] Hu, M. and Liu, B.: Mining and summarizing customer reviews, *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 168–177 (2004).

[7] Ku, L.-W., Liang, Y.-T. and Chen, H.-H.: Opinion Extraction, Summarization and Tracking in News and Blog Corpora., *AAAI spring symposium: Computational approaches to analyzing weblogs*, Vol. 100107 (2006).

[8] Liu, B., Hu, M. and Cheng, J.: Opinion observer: analyzing and comparing opinions on the web, *Proceedings of the 14th international conference on World Wide Web*, ACM, pp. 342–351 (2005).

[9] Mihalcea, R., Banea, C. and Wiebe, J. M.: Learning multilingual subjective language via cross-lingual projections (2007).

[10] Titov, I. and McDonald, R. T.: A Joint Model of Text and Aspect Ratings for Sentiment Summarization., *ACL*, Vol. 8, Citeseer, pp. 308–316 (2008).

[11] Vanderwende, L., Suzuki, H., Brockett, C. and Nenkova, A.: Beyond SumBasic: Task-focused summarization with sentence simplification and lexical expansion, *Information Processing & Management*, Vol. 43, No. 6, pp. 1606–1618 (2007).

[12] Wan, X.: Co-training for cross-lingual sentiment classification, *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, Association for Computational Linguistics, pp. 235–243 (2009).

[13] Whissell, C.: The dictionary of affect in language, *Emotion: Theory, research, and experience*, Vol. 4, No. 113-131, p. 94 (1989).

[14] Wilson, T., Wiebe, J. and Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis, *Proceedings of the conference on human language technology and empirical methods in natural language processing*, Association for Computational Linguistics, pp. 347–354 (2005).

[15] Xu, J., Xu, R., Ding, Y., Wang, X. and Kit, C.: Cross lingual opinion analysis via transfer learning, *Australian Journal of Intelligent Information Processing Systems*, Vol. 11, No. 2, pp. 28–34 (2010).