An Empirical Study on Sentiment Analysis for Vietnamese

Nguyen Thi Duyen
Department of Computer Science,
Posts & Telecommunications
Institute of Technology,
Hanoi, Vietnam
nguyenthiduyen.ptit@gmail.com

Ngo Xuan Bach
Department of Computer Science,
Posts & Telecommunications
Institute of Technology,
Hanoi, Vietnam
bachnx@ptit.edu.vn

Tu Minh Phuong
Department of Computer Science,
Posts & Telecommunications
Institute of Technology,
Hanoi, Vietnam
phuongtm@ptit.edu.vn

***Abstract*—This paper presents an empirical study on machine learning based sentiment analysis for Vietnamese, in which we focus on the task of sentiment classification. We investigate the task regarding both learning model and linguistics feature aspects. We also introduce an annotated corpus for sentiment classification extracted from hotel reviews in Vietnamese and conduct a series of experiments and analyses on that corpus. The paper provides useful information for further research as well as for building a real sentiment analysis system for Vietnamese.**

Keywords*—Sentiment Analysis, Naive Bayes, Maximum Entropy Models, Support Vector Machines.

I. Introduction

Sentiment analysis has become an important and hot topic in natural language processing community in recent years. A sentiment analysis system aims to analyze opinionated texts, such as opinions, emotions, sentiments, evaluations, beliefs, and speculations [18]. Based on such analysis, the system can provide helpful information for people, who are finding a restaurant to eat, a movie to see, a book to read, a place to go, or a hotel to stay.

Two major tasks in sentiment analysis include subjectivity classification, the task of labeling a given text as either subjective or objective, and sentiment classification, the task of classifying a subjective text as either positive, negative or neutral [18]. For example, the sentence "*The hotel provides excellent service.*" can be labeled as positive, while the sentence "*I don't want to come back that hotel.*" can be labeled as negative.

Sentiment classification using machine learning has been studied extensively for documents written in English. Various types of models have been proposed to deal with the task, including supervised models [3], [16], [17], [22], unsupervised models [20], semi-supervised models [12], [21], [23], [26], and active learning models [13]. In other languages, such as Japanese, Chinese, German, Spanish, Romanian, research on this task is growing [15]. In Vietnamese, however, rather less attention has been paid to this interesting task. One reason may be the lack of published corpus for the task.

In this paper, we present an empirical study on Vietnamese sentiment analysis, in which we focus on the task of sentiment classification. We investigate the task regarding both learning model and linguistics feature aspects. The contribution of this paper is twofold: 1) we construct a corpus annotated with sentiment labels by extracting hotel reviews followed by manual annotation and statistical verification; 2) we investigate the impact of different features, including non linguistics ones such as the overall score of a review on the classification performance. To the best of our knowledge, this is the first work that conducts such investigation of learning models with rich features for Vietnamese sentiment analysis.

The rest of this paper is organized as follows. Section II presents related work. Section III introduces a system for Vietnamese sentiment analysis. Section IV describes our corpus and experimental setup. Experimental results are presented in Section V. Finally, Section VI concludes the paper and discusses future work.

II. Related Work

Various studies on sentiment analysis have been conducted in English. Pang et al. [17] compare multiple supervised learning methods, including Naive Bayes [19], Maximum Entropy Models [2], and Support Vector Machines [7], [24], for the task of sentiment classification. They conducted experiments with various types of features and obtained 82.9% accuracy on a corpus of movie reviews. Pang and Lee [18] have written an excellent survey on techniques and approaches applied to sentiment classification. A discussion of available resources, benchmark datasets, and evaluation campaigns is also provided. Nakagawa et al. [16] describe a dependency tree-based method for sentiment classification using Conditional Random Fields [11] with hidden variables, and show that the method performs better than other methods based on bag-of-features. Bespalov et al. [3] present a method for classifying the sentiment of English documents using a multi-layered deep neural network [6]. Recently, Socher et al. [22] introduce the Stanford Sentiment Treebank, the first corpus with fully labeled parse trees. They also present a novel method, Recursive Neural Tensor Networks, which pushes state of the art by 5.4% for sentiment classification. Unsupervised learning, semi-supervised learning, and active learning have also been exploited to solve the sentiment classification task [12], [13], [14], [20], [21], [23], [26].

In other languages, research on sentiment analysis is growing. Cheng and Zhulyn [4] present an empirical study on

This work was done when the second author was at the School of Information Science, Japan Advanced Institute of Technology, Japan.

978-1-4799-6956-2/14/$31.00 ©2014 IEEE

sentiment classification in nine languages, including Japanese, English, German, Chinese, French, Italian, Spanish, Dutch, and Portuguese. They describe experiment results with two learning algorithms, Naive Bayes and Boosting with a logistic regression, on large datasets. Mihalcea et al. [15] present a tutorial on multilingual subjectivity and sentiment analysis at the ACL 2012 conference. They focus on the subjectivity and sentiment research carried out on languages other than English. In languages such as Chinese, character-based sentiment analysis has been investigated [25]. In Romanian, Banea et al. [1] show that subjectivity detection may be easier to achieve due to markers of politeness and additional verbal modes embedded in the language.

In Vietnamese, Kieu and Pham [10] introduce a rule-based system for Vietnamse sentiment classification using the Gate framework [8] and describe experiments on a corpus of computer product reviews. Unlike the work of Kieu and Pham [10], we take machine learning approaches to examine the task. We also investigate the impact of the overall score of a review on classifying the sentiment of a sentence. Machine learning has been shown to have several advantages over a rule-based approach. First, we do not need to design rules, which is a difficult and time-consuming task. Second, machine learning has better generalization capability. Third, machine learning can deal with ambiguities, which are a problem in a rule-based system. Last, a system using machine learning can adapt easily from one domain to anther domain.

III. A System for Vietnamese Sentiment Analysis

In this section, we present our system for Vietnamese sentiment analysis, which has been built to analyze hotel reviews. Hotel reviews are used for the illustration purpose. A system which analyzes other kinds of texts should have the same architecture with our system.

The system receives reviews of a hotel, which were commented by customers stayed in that hotel in the past, as the input, and outputs useful information for prospective customers. The output information can be represented in several formats, such as a number ranging from 1 (bad) to 5 (excellent) or a suggestion which is either positive (should book that hotel) or negative (should not book that hotel).

Figure 1 illustrates the framework of our sentiment analysis system. The system consists of a preprocessing module and two main modules, subjectivity identification and sentiment classification.

- **Preprocessing module**: this module conducts some preprocessing steps, including sentence detection, word segmentation, and part-of-speech tagging.
- **Subjectivity identification module**: this module receives a preprocessed sentence and labels it as either objective or subjective (opinionated sentence).
- **Sentiment classification module**: this module receives an opinionated sentence and classifies it as either positive, negative or neutral.

Among these three modules, in this paper we focus on sentiment classification, the most important step, which provides beneficial output for users. In the next sections, we will present our dataset on the hotel domain and experiments on Vietnamese sentiment classification.

IV. Data & Experimental Setup

A. Data

Our dataset was retrieved from Agoda[1], a website for booking hotels online. We extracted reviews in Vietnamese for 50 hotels, which are located in Vietnam (mainly in Hanoi, Ho Chi Minh city, Danang, and Nhatrang). We then conducted some preprocessing steps, including sentence detection[2], word segmentation[3], and part-of-speech tagging[4]. We also removed sentences which are not standard Vietnamese, i.e. sentences without tone marks[5].

Figure 2 shows an example of a preprocessed review in our corpus and its translation. Reviews are stored in the XML format, in which the *Score* tag indicates the overall score of the review, and the *Id* tag indicates the index of a review or the index of a sentence. In each sentence, words are separated by white spaces and part-of-speech tags follow words by slashes "/". For example, P means a pronoun, V means a verb, A means an adjective, and N means a common noun. In a Vietnamese sentence, each word may contain several syllables, which are separated by white spaces. At the word segmentation step, we connected syllables in the same word by underlines "_".

Preprocessed sentences were then annotated by two people in two steps. At the first step, each annotator assigned to each sentence a sentiment label, POSITIVE, NEGATIVE or NEUTRAL. At the second step, sentences with two different labels, or contradiction sentences, were reconsidered. Two annotators discussed and assigned a unique label to each contradiction sentence. To measure the inter-annotator agreement, we used the Cohen's kappa coefficient as follows:

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

where $Pr(a)$ is the relative observed agreement between two annotators, and $Pr(e)$ is the hypothetical probability of chance agreement. The Cohen's kappa coefficient of our corpus was 0.89, which can be interpreted as almost perfect agreement.

Figure 3 shows an example of a review with annotated sentences, in which the *Class* tag indicates the label of a sentence. In this example, the first and the second sentences have POSITIVE tags, while the third and the fourth sentences have NEGATIVE tags.

Table I shows statistical information of our corpus. The corpus contains 3304 sentences, including 1980 positive sentences, 777 negative sentences, and 547 neutral sentences. The corpus is available from the second author upon request.

[1]http://www.agoda.com
[2]http://mim.hus.vnu.edu.vn/phuonglh/softwares/vnSentDetector
[3]http://mim.hus.vnu.edu.vn/phuonglh/softwares/vnTokenizer
[4]http://mim.hus.vnu.edu.vn/phuonglh/softwares/vnTagger
[5]Vietnamese language consists of several tone marks. Some people, however, write sentences without using them to save time.

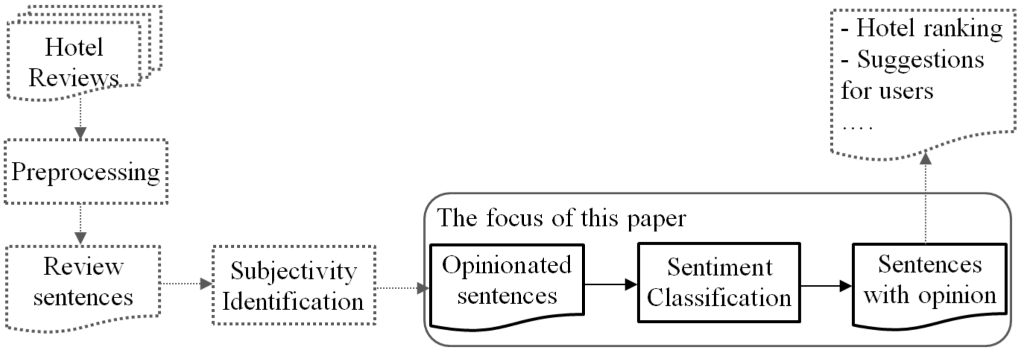

Fig. 1. The overall framework of a Vietnamese sentiment analysis.

```
<review Score="8" Id="4">
    <sentence Id="1">Chúng_tôi/P hài_lòng/V với/E kỳ/N nghỉ/V tại/E resort/N ./.
    </sentence>
    <sentence Id="2">Giá/N phòng/N rất/R tốt/A so/V với/E các/L resort/N khác/A
    chất_lượng/N tương_đương/A ./. </sentence>
    <sentence Id="3">Tuy_nhiên/C nhân_viên/N cần/V được/V đào_tạo/V để/E phục_vụ/V chuyên_nghiệp/A
    hơn/R ./. </sentence>
    <sentence Id="4">Chúng_tôi/P phải/V chờ/V đến/E hơn/A 3h/M chiều/N mới/R được/V nhận/V phòng/N ./.
    </sentence>
</review>
```

Translation

```
<sentence Id="1">We are satisfied with the vacation at the resort.</sentence>
<sentence Id="2">The price is very good in comparison with other resorts with the same quality.</sentence>
<sentence Id="3">However, staffs need to be trained to serve more professionally.</sentence>
<sentence Id="4">We had to wait for rooms till 3pm.</sentence>
```

Fig. 2. A preprocessed review in our corpus.

```
<review Score="8" Id="4">
    <sentence Class="POSITIVE" Id="1">Chúng_tôi/P hài_lòng/V với/E kỳ/N nghỉ/V tại/E resort/N ./.
    </sentence>
    <sentence Class="POSITIVE" Id="2">Giá/N phòng/N rất/R tốt/A so/V với/E các/L resort/N khác/A
    chất_lượng/N tương_đương/A ./. </sentence>
    <sentence Class="NEGATIVE" Id="3">Tuy_nhiên/C nhân_viên/N cần/V được/V đào_tạo/V để/E
    phục_vụ/V chuyên_nghiệp/A hơn/R ./. </sentence>
    <sentence Class="NEGATIVE" Id="4">Chúng_tôi/P phải/V chờ/V đến/E hơn/A 3h/M chiều/N mới/R
    được/V nhận/V phòng/N ./. </sentence>
</review>
```

Fig. 3. An annotated review in our corpus.

TABLE I. STATISTICAL INFORMATION OF THE CORPUS

Class	Number of Sentences
POSITVE	1980
NEGATIVE	777
NEUTRAL	547
Total	3304

B. Experimental Setup

1) Method for Conducting Experiments: We randomly divided the corpus into 5 folds and conducted 5-fold cross-validation test. The performance of the sentiment classification system was measured using accuracy.

$$accuracy = \frac{\#\ of\ correctly\ classified\ sentences}{\#\ of\ sentences}.$$

2) Learning Algorithms: As the learning methods, we used three different models, including Naive Bayes [19], Maximum Entropy Models (MEMs) [2], and Support Vector Machines (SVMs) [7], [24]. These learning methods have been shown to be effective for sentiment analysis [17]. In the following, we give a brief introduction to Naive Bayes, Maximum Entropy Models, and Support Vector Machines.

- *Naive Bayes*. Naive Bayes is a simple classification method based on Bayes' theorem and the independence assumptions between the features [19]. Given a sample $x = (x_1, x_2, \ldots, x_n)$, the method will find the most probable class label y^* for x as follows:

$$\begin{aligned} y^* &= argmax_{y \in Y} p(y|x) \\ &= argmax_{y \in Y} \frac{p(x|y)p(y)}{p(x)} \\ &= argmax_{y \in Y} p(x|y)p(y). \end{aligned}$$

where Y is the set of all class labels. Using the independence assumptions between the features, we have:

$$p(x|y) \approx \prod_{i=1}^{n} p(x_i|y).$$

The probabilities $p(y)$ and $p(x_i|y)$ can be estimated simply by counting on training samples.

- *Maximum Entropy Models*. Maximum Entropy Models (MEMs) [2] are a method of estimating the conditional probability $p(y|x)$ that a model outputs a label y given a context x,

$$p(y|x) = \frac{1}{Z(x)} exp(\sum_i \lambda_i f_i(x, y))$$

  where $f_i(x, y)$ refers to a feature function; λ_i is a parameter of the model; and $Z(x)$ is a normalization factor. To capture statistic information, this method requires that the model accord with some constraints (properties). Among the models that satisfy these constraints, the MEM method chooses the model with the flattest probability distribution (the model with the highest entropy).

- *Support Vector Machines*. Support Vector Machines (SVMs) are a statistical machine learning technique proposed by Vapnik et al. [7], [24]. To choose a hyperplane separating samples in a classification task, SVMs use a strategy that maximizes the margin between training samples and the hyperplane.
  In the cases where we cannot separate training samples linearly (because of some noises in the training data, for example) we can build the separating hyperplane by allowing some misclassifications. In those cases, we can build an optimal hyperplane by introducing a soft margin parameter, which trades off between the training error and the magnitude of the margin [24].

3) Feature Set: We investigated the task with the following kinds of features and their combinations:

- *Words*: all words in the sentence.
- *Important words*: main words in the sentence, including proper nouns, common nouns, verbs, adjectives, adverbs, and subordinating conjunctions.
- *N-grams of words*: unigrams, bigrams, and trigrams of words.
- *Syllables*: all syllables of all words in the sentence.
- *Important syllables*: all syllables of important words in the sentence.

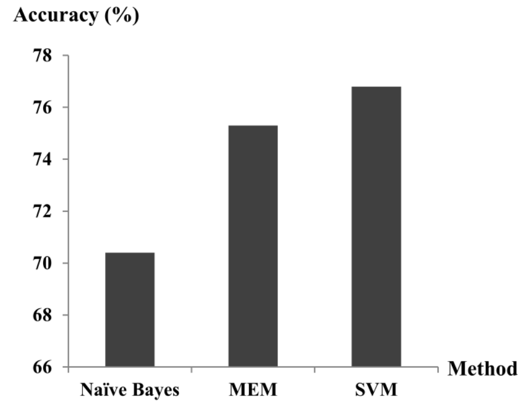

Fig. 4. Comparison between different learning methods.

- *N-grams of syllables*: unigrams, bigrams, and trigrams of syllables.
- *Overall score*: the score of the review that contains the sentence. When writing some kinds of reviews such as hotel reviews, users can explicitly specify overall scores, e.g. using one to ten scale, to express overall evaluation. We used such scores as additional features.

4) Experimental Purposes: The purposes of our experiments are as follows.

- To investigate different machine learning methods.
- To compare using word features with using syllable features.
- To compare using all words with using important words.
- To investigate different combinations of n-grams.
- To investigate the impact of overall scores.

V. Experimental Results

A. Different Learning Methods

We conducted experiments to compare three learning methods, Naive Bayes, MEM[6], and SVM[7]. For each learning method, we conducted experiments with various combinations of features, including word features, syllable features, important word features, important syllable features, n-gram features, and overall score features.

Figure 4 shows the best results of three methods. Among three methods, the SVM method achieved the highest result, 76.8% accuracy, the MEM method achieved 75.3% accuracy, and the Naive Bayes method got the worst result, 70.4% accuracy.

B. Words vs. Syllables

Figure 5 compares the results between two feature extraction methods: word-based method and syllable-based method. We can see that, for all three learning algorithms, extracting

[6]For Naive Bayes and MEM, we used WEKA [9].
[7]We used LIBSVM [5] with RBF kernel.

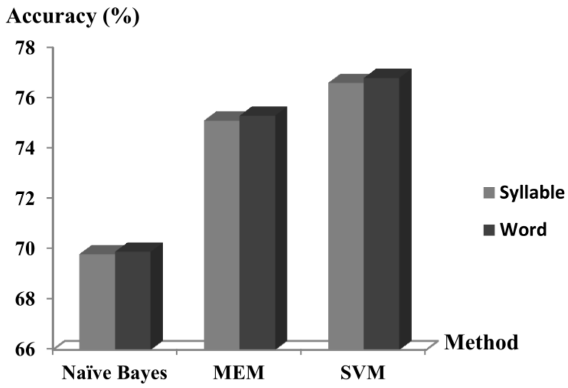

Fig. 5. Comparison between using word features and using syllable features.

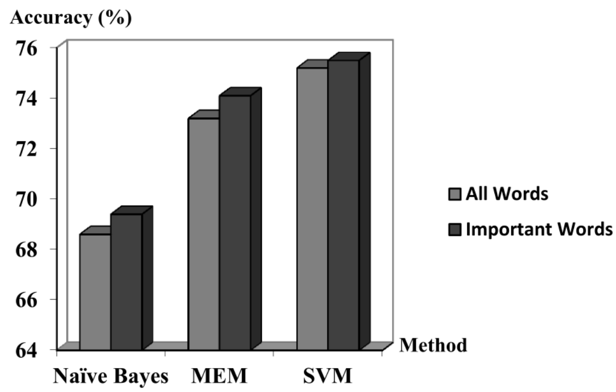

Fig. 6. Comparison between using all words and using important words.

features based on words achieved a little bit higher results than extracting features based on syllables: 76.8% and 76.4% for SVM, 75.3% and 75.1% for MEM, and 70.4% and 69.8% for Naive Bayes.

C. All Words vs. Important Words

In our system, features were extracted mainly based on words. An important question is that should we use all words or use only important words[8]? To answer this question, we compared experimental results of two feature extraction methods: using all words and using only important words. Figure 6 shows that for all three learning algorithms, extracting features using important words achieved the better results. It is reasonable because the meaning of a sentence is mainly constituted from nouns, verbs, adjectives, and adverbs.

D. Different Combination of N-grams

Figure 7 compares experimental results when using only unigrams, unigrams and bigrams, and using all unigrams, bigrams, and trigrams. The learning algorithm here was SVM. We can see that using only unigrams gave the best result.

E. Impact of Overall Scores

An important purpose of our experiments is to investigate the impact of overall scores. Figure 8 shows the experimental

[8]Important words are main words in a sentence, including proper nouns, common nouns, verbs, adjectives, adverbs, and subordinating conjunctions.

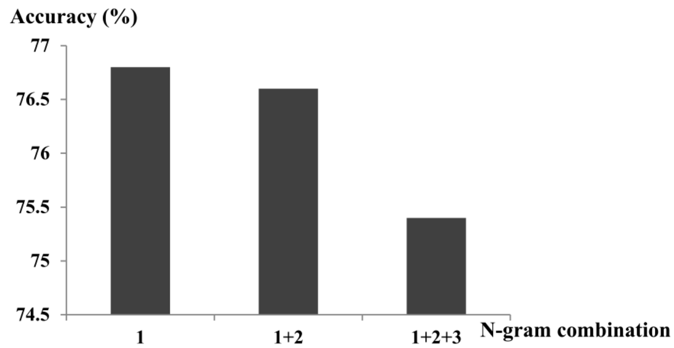

Fig. 7. Comparison between different combinations of N-gram features.

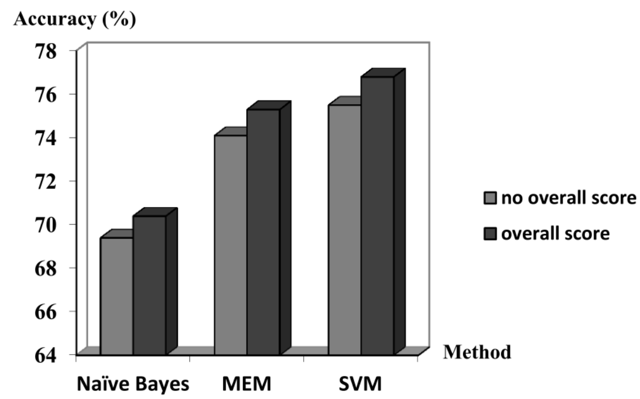

Fig. 8. The impact of overall scores.

results of three learning algorithms with and without overall scores. For all learning algorithms, using overall scores improved the performance of the system. Without overall scores, the system achieved 74.5%, 74.1%, and 69.4% accuracy when using SVM, MEM, and Nave Bayes, respectively. Adding overall score features improved the performance of the system to 76.8%, 75.5%, and 70.4% accuracy, respectively. The results showed that overall score features were effective for Vietnamese sentiment analysis.

F. Positive vs. Negative

We also measured the ability of the system to predict positive and negative classes separately using precision, recall, and the F_1 score. The precision, recall, and the F_1 score of the positive class were calculated as follows (negative class is similar):

$$precision = \frac{|A \cap B|}{|A|},$$

$$recall = \frac{|A \cap B|}{|B|}, and$$

$$F_1 = \frac{2 * precision * recall}{precision + recall},$$

where A is the set of test samples that the system predicted as positive, and B is the set of positive samples in the test dataset.

Table II shows the experimental results on positive and negative classes with three learning methods. For all learning

TABLE II. PERFORMANCE OF THE SYSTEM ON POSITIVE AND NEGATIVE CLASSES

Method	Class	Precision	Recall	F_1
Naive Bayes	Positive	85.4	80.8	83.0
	Negative	66.3	62.4	64.3
MEM	Positive	88.7	90.7	89.7
	Negative	65.5	89.2	75.5
SVM	Positive	82.2	94.0	87.7
	Negative	73.7	72.6	73.1

methods, the F_1 score of the positive class was much higher than the F_1 score of the negative class. For example, when we used SVM, the F_1 scores of two classes were 87.7% and 73.1%, respectively. This can be explained that the number of positive samples (1980) is larger than the number of negative samples (777) in the corpus. Another reason may be positive sentences are usually stated clearly, while negative sentences are often stated implicitly. Let us examine four sentences in the Figure 2. Two positive sentences, "*We are satisfied with the vacation at the resort.*" and "*The price is very good in comparison with other resorts with the same quality.*", contain strong indication words, "*satisfied*" and "*very good*". However, two negative sentences, "*However, staffs need to be trained to serve more professionally.*" and "*We had to wait for rooms till 3pm.*", seem to be stated implicitly.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have presented an empirical study on Vietnamese sentiment analysis based on machine learning. Experimental results on a corpus extracted from hotel reviews showed that: 1) SVM was effective for Vietnamese sentiment classification; 2) Using word-based features achieved the better results than using syllable-based features; 3) Extracting features based on important words, including proper nouns, common nouns, verbs, adjectives, adverbs, and subordinating conjunctions outperformed extracting features using all words; 4) Using only unigrams was effective for the task; and 5) The overall scores were very important for accurately predicting sentiment of sentences in Vietnamese.

For future work, we would like to study semi-supervised learning methods for Vietnamese sentiment analysis. Exploiting annotated data in other languages, for example in English, to solve the task for Vietnamese language may also be an interesting direction.

ACKNOWLEDGMENT

This work was supported by Posts and Telecommunications Institute of Technology, Hanoi, Vietnam.

REFERENCES

[1] C. Banea, R. Mihalcea, and J. Wiebe. 2008. A Bootstrapping method for building subjectivity lexicons for languages with scarce resources. In *Proceedings of LREC*, pp. 2764–2767, 2008.

[2] A.L. Berger, V.J.D. Pietra, and S.A.D. Pietra. A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics*, 22(1), 1996.

[3] D. Bespalov, B. Bai, Y. Qi, and A. Shokoufandeh. Sentiment Classification Based on Supervised Latent N-gram Analysis. In *Proceedings of CIKM*, pp. 375–382, 2011.

[4] A. Cheng and O. Zhulyn. A System for Multilingual Sentiment Learning On Large Data Sets. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, pp. 577–592, 2012.

[5] C. Chih-Chung and L. Chih-Jen. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (ACM TIST)*. 2(3):1–27.

[6] R. Collobert and J. Weston. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning. In *Proceedings of ICML*, pp. 160–167, 2008.

[7] C. Cortes and V. Vapnik. Support-Vector Networks. *Machine Learning*, 20(3):273–297, 1995.

[8] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. GATE, A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 168–175, 2002.

[9] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11(1):10–18, 2009.

[10] B.T. Kieu and S.B. Pham. Sentiment Analysis for Vietnamese. In *Proceedings of Second International Conference on Knowledge and Systems Engineering (KSE)*, pp. 152–157, 2010.

[11] J. Lafferty, A. McCallum, and F. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 282–289, 2001.

[12] S. Li, Z. Wang, G. Zhou, and S.Y.M. Lee. Semi-Supervised Learning for Imbalanced Sentiment Classification. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1826–1831, 2011.

[13] S. Li, S. Ju, G. Zhou, and X. Li. Active Learning for Imbalanced Sentiment Classification. In *Proceedings of EMNLP-CoNLL*, pp. 139–148, 2012.

[14] A.L. Maas, R.E. Daly, P.T. Pham, D. Huang, A. Ng, and C. Potts. Learning Word Vectors for Sentiment Analysis. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 142–150, 2011.

[15] R. Mihalcea, C. Banea, and J. Wiebe. Multilingual Subjectivity and Sentiment Analysis. *Tutorial at ACL 2012 Conference*, 2012.

[16] T. Nakagawa, K. Inui, and S. Kurohashi. Dependency Tree-based Sentiment Classification using CRFs with Hidden Variables. In *Proceedings of NAACL*, pp. 786–794, 2010.

[17] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 79–86, 2002.

[18] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135, 2008.

[19] I. Rish. An Empirical Study of the Naive Bayes classifier. In *Proceedings of IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, 2001.

[20] J. Rothfels and J. Tibshirani. Unsupervised Sentiment Classification of English Movie Reviews using Automatic Selection of Positive and Negative Sentiment Items. *Technical Report*, Stanford University, 2010.

[21] R. Socher, J. Pennington, E.H. Huang, A. Ng, and C. Manning. Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 151–161, 2011.

[22] R. Socher, A. Perelygin, J.Y. Wu, J. Chuang, C. Manning, A. Ng, and Christopher Potts. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In *Proceedings of EMNLP*, pp. 1631–1642, 2013.

[23] O. Tackstrom and R. McDonald. Semi-supervised Latent Variable Models for Sentence-level Sentiment Analysis. In *Proceedings of ACL*, pp. 569–574, 2011.

[24] V.N. Vapnik. Statistical Learning Theory. *Wiley-Interscience*, 1998.

[25] L. Xiang. Ideogram Based Chinese Sentiment Word Orientation Computation. *Computing Research Repository*, 2011.

[26] S. Zhou, Q. Chen and X. Wang. Active Deep Networks for Semi-Supervised Sentiment Classification. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, pp. 1515–1523, 2010.