An Empirical Study on Sentiment Analysis for Vietnamese Comparative Sentences

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

***Abstract*—This paper presents an empirical study on sentiment analysis for Vietnamese language focusing on comparative sentences, which have different structures compared with narrative or question sentences. Given a set of evaluative Vietnamese documents, the goal of the task consists of (1) identifying comparative sentences in the documents; (2) recognition of relations in the identified sentences; and (3) identifying the preferred entity in the comparative sentences if any. A relation describes a comparison of two entities or two sets of entities on some features or aspects in the sentence. Such information is needed for sentiment analysis in comparative sentences, which is very useful not only for customers in choosing products but also for manufacturers in producing and marketing. We present a general framework to solve the task in which we formulate the first and the third subtasks, i.e. identifying comparative sentences and identifying the preferred entity, as a classification problem, and the second subtask, i.e. recognition of relations, as a sequence learning problem. We introduce a new corpus for the task in Vietnamese and conduct a series of experiments on that corpus to investigate the task in both linguistic and modeling aspects. Our work provides promising results for further research on this interesting task.**

***Index Terms*—Sentiment Analysis, Opinion Mining, Comparative Sentences, Support Vector Machines, Conditional Random Fields.**

I. INTRODUCTION

Sentiment analysis and opinion mining have become a hot research topic and attracted many researchers in natural language and data mining communities in recent years [1], [2]. The aim of a sentiment analysis system is to analyze opinionated texts, such as opinions, emotions, sentiments, and evaluations. Such analyses can provide useful information for both customers and manufactures. For customers, the system can help to choose a product or a service. For manufactures, the system can help to market products, understand customers, and suggest strategies for developing new products or services.

Most existing work in sentiment analysis and opinion mining focuses on sentiment classification, the task of classifying a given text as either positive or negative (or neutral). For example, the sentence "*It was a wonderful trip.*" can be labeled as positive, while the sentence "*That hotel provides very bad services.*" can be labeled as negative. Various methods have been proposed to deal with the sentiment classification task, including supervised methods [3], [4], [5], [6], unsupervised methods [7], and semi-supervised methods [8], [9], [10], [11].

Although mining comparative sentences is an important task in sentiment analysis and opinion mining, little work has been done on this task. Comparative sentences have specific structures in comparison with other types of sentences. Comparative sentences compare two entities or two sets of entities in some features or aspects. Sentiment analysis on comparative sentences consists of three subtasks, i.e. identifying comparative sentences, recognition of relations, and identifying the preferred entity. While the goal of the first subtask is to identify comparative sentences in the input text, the goal of the second subtask is recognizing compared entities, compared features, and comparing words in an identified comparative sentence. The third subtask using identified information to determine which entity is preferred by the writer. For example, the sentence "*The display quality of mobile phone X is better than that of mobile phone Y.*" compares two entities "mobile phone X" and "mobile phone Y" regarding their "display quality". From the comparing word "*better than*", we know that "mobile phone X" is the preferred entity.

In this paper, we study the comparative sentence sentiment analysis task for Vietnamese language. We present a framework to deal with the task in which we model the first subtask and the third subtask as a classification problem and model the second subtask as a sequence learning problem. We also introduce a corpus for the task consisting of Vietnamese sentences in the domain of electronic devices, and present a series of experiments conducted on that corpus. While several studies have been done on mining comparative sentences for English [12], [13], [14], [15], Arabic [16], Chinese [17], and Korean [18], this is the first work conducted for Vietnamese.

The rest of this paper is organized as follows.

Corresponding author: Ngo Xuan Bach
Email: bachnx@ptit.edu.vn
Manuscript received: 4/2018, revised: 5/2018, accepted: 8/2018

Section II describes related work. Section III presents our framework for Vietnamese comparative sentence sentiment analysis. Section IV introduces our corpus and experiments. Finally, Section V concludes the paper.

II. RELATED WORK

Jindal and Liu [13] describe a study on identifying comparative sentences in English documents. Their approach is a combination of class sequential rule mining and machine learning. Class sequence rules are found automatically using a class sequential rule mining system. Naive Bayes is then employed to build a classifier based on the rules. They achieve about 80% in the F_1 score on a corpus consisting of 5890 English sentences. Jindal and Liu [14] extract entities and features in comparative sentences using label sequence rules. They report an F_1 score of 72% on a corpus of nearly 600 English comparative sentences. Ganapathibhotla and Liu [12] introduce a method for mining opinions in English comparative sentences. Given a comparative sentence which contains two entities (or two sets of entities), a compared feature, and comparing words, the goal of the task is to identify which entity is preferred by the author. Their method is based on rules, which analyze characteristics of different types of English comparative sentences. Although that method achieves good results, it is too specific for English and difficult to adapt to other languages.

Xu et al. [15] present a method for mining comparative opinions in business intelligence. They introduce a graphical model using Conditional Random Fields [19] to extract and visualize comparative opinions between products from customer reviews. The goal of their system is to help manufactures discover potential risks, design new products, and suggest marketing strategies.

Among various work on mining comparative sentences for languages other than English, El-Halees [16] describes a study on opinion mining from Arabic comparative sentences. The work focuses on identifying comparative sentences and achieves 89% in the F_1 score on a corpus of 1048 Arabic sentences. Huang et al. [17] investigate the task of identifying comparative sentences in Chinese texts. They describe experiments with several linguistic and statistical features using various classifiers. Yang and Ko [18] introduce a hybrid method for identifying Korean comparative sentences in web documents. Their method first generates a set of comparative sentence candidates by using a set of predefined keywords and then exploits machine learning techniques to identify comparative sentences from candidates. They report 90% in the F_1 score on a corpus of 7384 Korean sentences.

In Vietnamese, several studies have been done on sentiment classification [20], [21], [22]. While Kieu and Pham [22] introduce a rule-based method to develop their system, Duyen et al. [21] describe a series of experiments on learning-based sentiment classification in Vietnamese. Bach et al. [20] introduce a weakly supervised method for sentiment classification in resource poor languages, and present experimental results on two datasets of Japanese and Vietnamese. To the best of our knowledge, however, the work presented in this paper is the first attempt on sentiment analysis for Vietnamese comparative sentences.

III. A SENTIMENT ANALYSIS FRAMEWORK FOR VIETNAMESE COMPARATIVE SENTENCES

In this section, we present our sentiment analysis system for Vietnamese comparative sentences. For the illustration purpose, we report here the results of the system when trained and tested with reviews in the domain of electronic devices. A system which analyzes other kinds of texts should have the same architecture as our system. Figure 1 illustrates the framework of our system. The system consists of a preprocessing module and three main modules: comparative sentence identification, relation recognition, and identifying the preferred entity.

- **Preprocessing**: this module conducts some preprocessing steps, including sentence detection, word segmentation, and part-of-speech tagging.
- **Comparative sentence identification**: this module receives a review sentence and identify whether it is a comparative sentence or not. In the case that the input sentence is a comparative sentence, the module also classifies it as either equal, non-equal, or superlative comparison.
- **Relation recognition**: this module receives an identified comparative sentence and recognizes entities, features, and comparing words in the sentence.
- **Identifying the preferred entity**: this module mines opinions from customer reviews using information from the previous modules and makes suggestions for customers or manufactures. Specifically, it identifies which entity is preferred by the writer.

A. Identifying Comparative Sentences

Like previous work for English [13], [14], we consider three types of comparative sentences, i.e. equative comparison, non-equative comparison, and superlative comparison.

- **Equative**: A sentence of this type describes an equative relation between two or more entities regarding a feature.

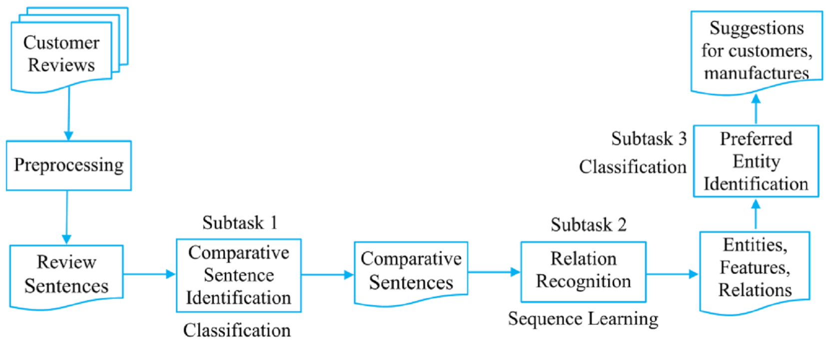

Fig. 1. A sentiment analysis framework for Vietnamese comparative sentences.

- **Non-Equative**: A sentence of this type describes a non-equative relation between two or more entities regarding a feature.
- **Superlative**: A sentence of this type describes a superlative relation between an entity and all other entities regarding a feature.

Figure 2 gives examples of comparative sentences of three types in Vietnamese and their translations into English. The first sentence states an equative relation between two entities, i.e. Nokia Lumia 920 and Samsung Galaxy S4, regarding their camera. The second sentence states a non-equative relation between Samsung Galaxy S4 and Samsung Galaxy S3 regarding their camera. In that sentence, the one of S4 is better than the one of S3. The last sentence sates a superlative relation between Iphone 5S and all other Iphones regarding the price.

We model the task of identifying Vietnamese comparative sentences as a classification problem, which labels each Vietnamese input sentence as either Equative, Non-equative, Superlative, or Non-comparative (sentences which do not state any comparative relation between entities).

Many learning algorithms have been proposed to deal with classification problems, including traditional methods such as k-NN, Decision Tree, Naive Bayes, and more advanced methods such as Maximum Entropy model (MEM) and Support Vector Machine (SVM). Any learning algorithm can be used in our proposed framework. In this work, we chose two classification methods, MEM [23] and SVM [24], to complete the framework. Both have been shown to be powerful and effective methods in various natural language processing and data mining tasks.

As features for classification models, we use words, syllables, and $n-$grams ($n = 1, 2, 3$) of them. Unlike English words, words in Vietnamese cannot be delimited by white spaces. Vietnamese words may consist of one or more syllables separated by white spaces.

B. Recognition of Relations

The goal of the relation recognition task is to recognize the relation stated in the input comparative sentence. Informally, the task is to identify entities, features, and comparing words in the sentence. Note that entities and features are enough to make clear relations in equative and superlative sentences in most cases. Hence, we only extract entities and features in equative and superlative sentences. Non-equative sentences, however, need more information to identify whether the relation is "*better than*" or "*worse than*". Therefore, we extract comparing words in addition to entities and features in non-equative sentences. A comparing word is defined as a word or a phrase which expresses comparing relation between entities. Figure 3 shows entities, compared features, and comparing words extracted from examples in Figure 2.

We model the task of relation recognition as a sequence learning problem, in which the input sentence is considered as a sequence of elements. Each element corresponds to a word in a word-based model or a syllable in a syllable-based model. We use the IOB notation to label each element by one of the following tags: B-Ent, I-Ent, B-Feat, I-Feat, BCWord, I-CWord, and O. Here, B-Ent means an element at the beginning of an entity; I-Ent means other elements of the entity. B-Feat, I-Feat, B-CWord, and I-CWord have the similar meaning for features and comparing words. Tag O is used for elements which are outside all entities, features, and comparing words. Figure 4 shows examples of how to model the task in a syllable-based model.

In our framework, we choose Conditional Random Fields (CRFs) [19] as the learning method. CRFs are undirected graphical models, which define the probability of a label sequence y given an observation sequence x as follows:

$$P(y|x, \lambda, \mu) = \frac{1}{Z(x)} exp(F(x, y, \lambda, \mu))$$

where $F(x, y, \lambda, \mu)$ is the total of feature functions:

$$F(x, y, \lambda, \mu) = \sum_j \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_k \mu_k s_k(y_i, x, i).$$

Equative:
Nokia Lumia 920 có camera ***tương tự*** Samsung Galaxy S4.
(The camera of Nokia Lumia 920 is ***similar to*** the one of Samsung Galaxy S4.)
Non-equative:
Camera của Samsung Galaxy S4 ***tốt hơn*** của Samsung Galaxy S3.
(The camera of Nokia Lumia 920 is ***better than*** the one of Samsung Galaxy S4.)
Superlative:
Điện thoại IPhone 5S có giá ***đắt nhất*** trong số các điện thoại Iphone.
(Iphone 5S is ***the most expensive*** one in the Iphone series.)

Fig. 2. Examples of Vietnamese comparative sentences.

Equative:
Nokia Lumia 920 có ***camera*** tương tự ***Samsung Galaxy S4***.
(Entities: Nokia Lumia 920, Samsung Galaxy S4; Feature: camera)
Non-equative:
Camera của ***Samsung Galaxy S4*** ***tốt hơn*** của ***Samsung Galaxy S3***.
(Entities: Samsung Galaxy S4, Samsung Galaxy S3; Feature: camera; Comparing word: tốt hơn (better))
Superlative:
IPhone 5S có ***giá*** đắt nhất trong số ***các dòng điện thoại Iphone***.
(Entities: Iphone 5S, các dòng điện thoại Iphone (Iphone series); Feature: giá (price))

Fig. 3. Examples of entities, features, and comparing words in comparative sentences.

Sentence 1: ***Nokia Lumia 920*** có ***camera*** tương tự ***Samsung Galaxy S4***.
Label Seq 1:B-Ent I-Ent I-Ent O B-Feat O O B-Ent I-Ent I-Ent O

Sentence 2: ***Camera*** của ***Samsung Galaxy S4*** ***tốt hơn*** của ***Samsung Galaxy S3***.
Label Seq 2:B-Feat O B-Ent I-Ent I-Ent B-CWord I-CWord O B-Ent I-Ent I-Ent O

Sentence 3:***IPhone 5S*** có ***giá*** đắt nhất trong số ***các dòng điện thoại Iphone***.
Label Seq 3:B-Ent I-Ent O B-Feat O O O O B-Ent I-Ent I-Ent I-Ent I-Ent O

Fig. 4. Examples of sequence labels in a syllable-based model.

Here $t_j(y_{i-1}, y_i, x, i)$ denotes a transition feature function (or edge feature), which is defined on the entire observation sequence x and the labels at positions i and $i-1$ in the label sequence y; $s_k(y_i, x, i)$ denotes a state feature function (or node feature), which is defined on the entire observation sequence x and the label at position i in the label sequence y; λ_j and μ_k are parameters of the model, which are estimated in the training process; and $Z(x)$ is a normalization factor.

CRFs have all the advantages of Maximum Entropy Markov models (MEMMs) but does not suffer from the label bias problem. They have been shown to be a suitable method for many sequence learning problems, especially in NLP tasks such as POS tagging, chunking, named entity recognition, syntax parsing, information retrieval, and information extraction [19], [25], [26].

C. *Identifying the Preferred Entity*

Given the relation extracted from the second subtask, i.e. two entities, feature, and the comparing word, the goal of this subtask is to identify which entity is preferred by the writer. For example, we have the input sentence "*The camera of Samsung Galaxy S4 is better than that of Samsung Galaxy S3*". In the second subtask, we extract the relation in the sentence, consisting of two entities, i.e. *Samsung Galaxy S4* and *Samsung Galaxy S3*, the comparing feature, i.e. *camera*, and the comparing word, i.e. "*better*". Based on that information, this subtask will determine the entity, which is preferred by the writer, i.e. Samsung Galaxy S4.

We also model this subtask as a binary classification, given two entities called Entity 1 and Entity 2, comparing feature, and comparing word, the model will predict which entity is preferred: label "+" for Entity 1 and label "–" for Entity 2. We determine Entity 1

TABLE I
STATISTICAL INFORMATION OF SENTENCE TYPES IN OUR DATASET

Sentence type	Number
Equative comparison	1000
Non-equative comparison	1000
Superlative comparison	1000
Non-comparative	1000
Total	**4000**

TABLE II
STATISTICAL INFORMATION OF ENTITIES, FEATURES, AND COMPARING WORDS

Type	Number
Entity	5119
Feature	2942
Comparing word	1087
Total	**9148**

and Entity 2 based on the order they appear in the sentence. Like the first subtask, we exploit two statistical learning models, i.e. Support Vector Machines and Maximum Entropy Model, to solve the task. As features, we use the two entities, the comparing word, and the comparing feature.

IV. EXPERIMENTS

This section describes our experiments on sentiment analysis for Vietnamese comparative sentences. We first introduce our corpus for the task. We then describe experimental settings and evaluation methods. Finally, we present experimental results on three subtasks.

A. Dataset

Our dataset was retrieved from VnReview[1] and Tinhte [2], two websites of technology products. We extracted Vietnamese technical reviews of electronic products such as computers, smartphones, and cameras. We then conducted preprocessing steps, including sentence detection[3], word segmentation, and part-of-speech tagging[4]. We also removed sentences which are not standard Vietnamese, i.e. sentences without tone marks. Vietnamese language consists of several tone marks. Some people, however, write sentences without using them to save time. Tables I and II show statistical information of our corpus. Our dataset consists of 4000 Vietnamese sentences, which contain 5119 entities, 2942 features, and 1087 comparing words.

B. Experimental Settings

For the first subtask, i.e. comparative sentence identification, we conducted experiments using all 4000 sentences. We randomly divided 4000 sentences into 5 folds and conducted 5-fold cross-validation test. The performance of our classification system was measured using accuracy, precision, recall, and the F_1 score.

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

Precision, recall, and the F_1 score were measured on each type of sentence. Let we consider sentences belonging to the equative type as an example, precision, recall, and the F_1 were calculated as follows:

$$precision = \frac{\#\text{of correctly classified equative sentences}}{\#\text{of predicted equative sentences}},$$

$$recall = \frac{\#\text{of correctly classified equative sentences}}{\#\text{of actual equative sentences}},$$

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

For the second subtask, i.e. relation recognition, we conducted experiments using 3000 comparative sentences, including equative, non-equative, and superlative types. We randomly divided 3000 comparative sentences into 5 folds and conducted 5-fold cross-validation test. The performance of our recognition system was measured using precision, recall, and the F_1 score, which were computed in a similar manner to the precision, recall, and the F_1 score in the first subtask.

For the third subtask, i.e. identifying the preferred entity, we conducted 5-fold cross-validation using non-equative sentences. The performance of the system was measured using accuracy.

C. Results

1) Comparative Sentence Identification: First, we conducted experiments on comparative sentence identification using SVM[5] with two feature extraction methods, i.e. syllable-based and word-based. For each feature extraction method, we conducted experiments with three feature sets: 1-grams; 1-grams and 2-grams; 1-grams, 2-grams, and 3-grams. Experimental results are shown in Table III. We can see that syllable-based method got better results than word-based method in all three cases of feature sets. For both syllable-based and word-based feature extraction methods, using 1-grams and 2-grams achieved the best results. Our best model, i.e. 1-grams and 2-grams extracted on syllables, achieved 86.30% accuracy.

Second, we conducted experiments to compare two learning algorithms, i.e. SVM and MEM, for Vietnamese comparative sentence identification. We also compared two algorithms using two feature extraction methods and three feature sets. As shown in Figure 5,

[1]http://vnreview.vn
[2]https://www.tinhte.vn
[3]http://mim.hus.vnu.edu.vn/phuonglh/softwares/vnSentDetector
[4]http://mim.hus.vnu.edu.vn/phuonglh/softwares/vnTagger

[5]We used LIBSVM [27] with RBF kernel.

TABLE III
COMPARATIVE SENTENCE IDENTIFICATION USING SVM

Feature extraction method	Feature set	Accuracy(%)
Syllable-based	1-grams	83.27
	1-grams + 2-grams	**86.30**
	1-grams + 2-grams + 3-grams	84.31
Word-based	1-grams	82.59
	1-grams + 2-grams	**86.11**
	1-grams + 2-grams + 3-grams	83.22

TABLE IV
SENTENCE IDENTIFICATION RESULTS USING SVM FOR EACH SENTENCE TYPE

Sentence type	Pre(%)	Re(%)	F_1(%)
Equative comparison	86.93	92.00	89.38
Non-equative comparison	82.18	80.51	81.32
Superlative comparison	93.70	89.97	91.79

TABLE V
EXPERIMENTAL RESULTS ON RELATION RECOGNITION USING DIFFERENT FEATURE SETS

Model	Precision(%)	Recall(%)	F_1(%)
Window size = 1	90.00	81.33	85.89
Window size = 2	91.21	81.66	86.17
Window size = 3	91.36	**81.73**	**86.28**
Without POS tags	**91.71**	77.52	84.02

SVM outperformed MEM in all cases. In the best case, i.e. using 1-grams and 2-grams extracted on syllables, SVM achieved 86.30% accuracy while MEM achieved only 81.00% accuracy.

We also evaluated the effectiveness of our method on each type of sentence. Table IV shows the F_1 scores on three types of sentences, i.e. equative, non-equative, and superlative sentences[6]. We achieved 89.38%, 81.32%, and 91.79% in the F_1 score on three types of sentences, respectively. There are two reasons which may explain why superlative comparison sentences have the highest F_1 score. The first reason is that superlative comparison sentences usually contain some specific phrases, such as "*the best*", "*the worst*", and "*all others*". The second one is that the structure of superlative sentences is different from the structure of equative and non-equative sentences. While equative and non-equative sentences compare two entities (or two sets of entities), superlative sentences compare an entity with all the others.

2) Relation Recognition: For the relation recognition task, we conducted experiments using CRF[7] with four different feature sets. With each word in the sentence, we extracted features in a window size of N, i.e. N preceding words and N next words and their part-of-speech tags. The first three feature sets corresponded to the window size $N = 1$, $N = 2$, and $N = 3$. The last feature set was the third one ($N = 3$) without part-of-speech tags. Table V shows experimental results on relation recognition. In general, the window sizes did not affect very much to the experimental results. Using window size 2 achieved better results than using window size 1. Using window size 3 got the best results. Without POS tags, the performance of the system decreased significantly.

Table VI shows the F_1 scores measured on entities, features, and comparing words, separately. Three models using window sizes 1, 2, and 3 achieved nearly the same results: about 93% on entities, 78% on features, and 73% on comparing words. The model without POS tags got much lower F_1 scores than three previous models.

Table VII compares experimental results between three sentence types, equative comparison, non-equative comparison, and superlative comparison[8]. Similar to the first subtask, we achieved the highest results on superlative comparison sentences on both entities and features.

3) Identifying the Preferred Entity: We conducted experiment with two statistical learning methods, i.e. Support Vector Machine (SVM) and Maximum Entropy Model (MEM). For SVM, we used LIBSVM[9] [27] with RBF kernel. For MEM, we used Weka[10]. Experimental results are shown in Table VIII. Similar to the first subtask, SVM outperformed MEM significantly (92.30% compared with 85.50%). From the experimental results of all three subtasks, Conditional Random Fields and Support Vector Machines have been shown to be effective machine learning techniques to deal with the task of sentiment analysis for Vietnamese comparative sentences.

V. CONCLUSION

We have presented an empirical study on sentiment analysis for Vietnamese comparative sentences, which consists of three subtasks: identifying comparative sentences; recognition of relations in identified

[6]We report the scores of the best model, i.e. using SVM with 1-grams and 2-grams extracted from syllables.

[7]We used CRF++, an implementation of Taku Kudo which is available at http://taku910.github.io/crfpp/.

[8]Comparing words were only recognized in non-equative sentences.

[9]https://www.csie.ntu.edu.tw/~cjlin/libsvm/

[10]http://www.cs.waikato.ac.nz/ml/weka/

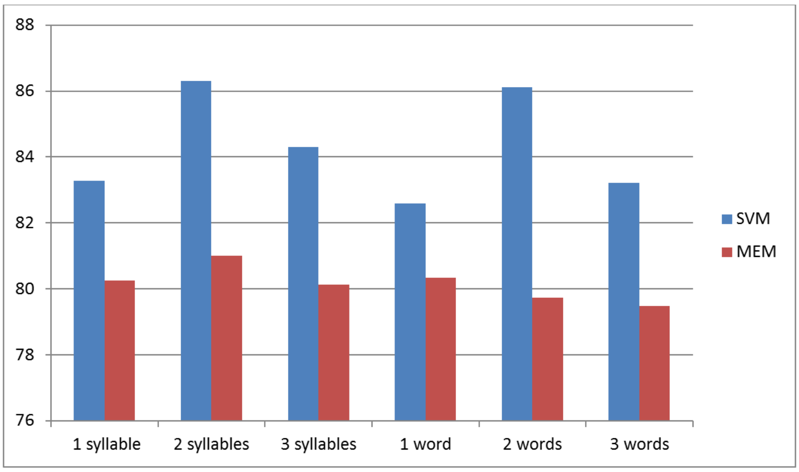

Fig. 5. Comparative sentence identification using SVM vs. MEM.

TABLE VI
EXPERIMENTAL RESULTS OF RELATION RECOGNITION IN DETAIL

Model	**Entity**			**Feature**			**Comparing word**		
	Pre(%)	**Re(%)**	**F_1(%)**	**Pre(%)**	**Re(%)**	**F_1(%)**	**Pre(%)**	**Re(%)**	**F_1(%)**
Window size = 1	95.56	**91.75**	**93.62**	85.86	69.60	76.88	78.43	68.37	73.06
Window size = 2	95.42	91.54	93.44	86.70	70.96	78.04	79.23	**68.97**	**73.74**
Window size = 3	95.44	91.32	93.33	**87.06**	**71.51**	**78.52**	**79.35**	68.42	73.48
Without POS tags	**96.83**	86.98	91.64	86.82	67.18	75.75	76.50	65.87	70.79

TABLE VII
RECOGNITION RESULTS ON THREE TYPES OF SENTENCES

Model	**Entity**			**Feature**		
	Pre(%)	**Re(%)**	**F_1(%)**	**Pre(%)**	**Re(%)**	**F_1(%)**
Equative	**95.78**	82.35	88.56	83.33	63.39	72.00
Non-equative	95.10	91.35	93.19	83.80	65.50	73.53
Superlative	95.50	**92.79**	**94.12**	**88.49**	**73.00**	**80.00**

TABLE VIII
EXPERIMENTAL RESULTS ON PREFERRED ENTITY IDENTIFICATION

Model	**Tool**	**Accuracy(%)**
MEM	Weka	85.50
SVM	LIBSVM	**92.30**

comparative sentences; and identifying the preferred entity. We described a general framework to solve the task and introduced an annotated corpus, which consists of 4000 Vietnamese sentences in the domain of electronic devices. Experiments showed that our model achieved promising results on this interesting task. For the first subtask, we got 86.30% accuracy. For the second subtask, our model achieved 93.33%, 78.52%, and 73.48% in the F1 score on recognition of entities, features, and comparing words, respectively. For the third subtask, we got 92.30% accuracy.

We have investigated three subtasks independently. For each subtask, we used gold input sentences to conduct experiments instead of using the output of the previous subtask. Only comparative sentences were recognized in the second subtask and non-equative comparative sentences were processed in the third subtask. In the future, we plan to investigate all three subtasks in a unified system.

REFERENCES

[1] B. Liu, *Sentiment Analysis and Opinion Mining: Synthesis lectures on human languages technologies*. Morgan and Claypool publishers, 2012.

[2] S. Poria, E. Cambria, D. Hazarika, N. Majumder, A. Zadeh, and L. Morency, "Context-dependent sentiment analysis in user-generated videos," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017, pp. 873–883.

[3] D. Bespalov, B. Bai, Y. Qi, and A. Shokoufandeh, "Sentiment classification based on supervised latent n-gram analysis," in *Proceedings of the International Conference on Information and Knowledge Management (CIKM)*, 2011, pp. 375–382.

[4] T. Nakagawa, K. Inui, and S. Kurohashi, "Dependency tree-based sentiment classification using crfs with hidden variables," in *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2010, pp. 786–794.

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

[6] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP)*, 2013, pp. 1631–1642.

[7] J. Rothfels and J. Tibshirani, "Unsupervised sentiment classification of english movie reviews using automatic selection of positive and negative sentiment items," Stanford University, Tech. Rep., 2010.

[8] S. Li, Z. Wang, G. Zhou, and S. Lee, "Semi-supervised learning for imbalanced sentiment classification," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2011, pp. 1826–1831.

[9] R. Socher, J. Pennington, E. Huang, A. Ng, , and C. Manning, "Semi-supervised recursive autoencoders for predicting sentiment distributions," in *Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP)*, 2011, pp. 151–161.

[10] O. Tackstrom and R. McDonald, "Semi-supervised latent variable models for sentence-level sentiment analysis," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 2011, pp. 569–574.

[11] 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)*, 2010, pp. 1515–1523.

[12] M. Ganapathibhotla and B. Liu, "Mining opinions in comparative sentences," in *Proceedings of the International Conference on Computational Linguistics (COLING)*, 2008, pp. 241–248.

[13] N. Jindal and B. Liu, "Identifying comparative sentences in text documents," in *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2006, pp. 244–251.

[14] ——, "Mining comparative sentences and relations," in *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2006, pp. 1331–1336.

[15] K. Xu, S. Liao, J. Li, and Y. Song, "Mining comparative opinions from customer reviews for competitive intelligence," *Decision Support Systems*, vol. 50, no. 4, pp. 743–754, 2011.

[16] A. El-Halees, "Opinion mining from arabic comparative sentences," in *Proceedings of the International Arab Conference on Information Technology (ACIT)*, 2012, pp. 265–271.

[17] X. Huang, X. Wan, J. Yang, and J. Xiao, "Learning to identify comparative sentences in chinese text," in *Proceedings of the Pacific Rim International Conferences on Artificial Intelligence (PRICAI)*, 2008, pp. 187–198.

[18] S. Yang and Y. Ko, "Extracting comparative sentences from korean text documents using comparative lexical patterns and machine learning techniques," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 2009, pp. 153–156.

[19] 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)*, 2001, pp. 282–289.

[20] N. Bach and T. Phuong, "Leveraging user ratings for resource-poor sentiment classification," in *Proceedings of the 19th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES)*, 2015, pp. 322–331.

[21] N. Duyen, N. Bach, and T. Phuong, "An empirical study on sentiment analysis for vietnamese," in *Proceedings of the International Conference on Advanced Technologies for Communications (ATC)*, 2014, pp. 309–314.

[22] B. Kieu and S. Pham, "Sentiment analysis for vietnamese," in *Proceedings of the International Conference on Knowledge and Systems Engineering (KSE)*, 2010, pp. 152–157.

[23] A. Berger, V. Pietra, and S. Pietra, "A maximum entropy approach to natural language processing," *Computational Linguistics*, vol. 22, no. 1, pp. 39–71, 1996.

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

[25] F. Peng and A. McCallum, "Information extraction from research papers using conditional random fields," *Information Processing Management*, vol. 42, no. 4, pp. 963–979, 2006.

[26] F. Sha, "Shallow parsing with conditional random fields," in *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2003, pp. 213–220.

[27] C. Chih-Chung and L. Chih-Jen, "Libsvm: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (ACM TIST)*, vol. 2, no. 3, pp. 1–27, 2011.

Ngo Xuan Bach received his B.Sc. degree in computer science from the University of Engineering and Technology (UET), Vietnam National University (VNU), Hanoi, in 2006. He received his M.Sc. and Ph.D. degrees in information science from the School of Information Science, Japan Advanced Institute of Science and Technology (JAIST), in 2011 and 2014. He is now with Faculty of Information Technology, Posts and Telecommunications Institute of Technology (PTIT), Hanoi. His research interests include statistical natural language processing and machine learning.