DSKTLAB: Vietnamese Sentiment Analysis for Product Reviews

Quynh-Trang Thi Pham[1], Xuan-Truong Nguyen[1], Van-Hien Tran[1], Thi-Cham Nguyen[1,2], Mai-Vu Tran[1]

[1]Data Science and Knowledge Technology Laboratory

University of Engineering and Technology, Vietnam National University, Hanoi

[2]Haiphong University of Medicine and Pharmacy

{trangptq_58, hientv, truongnx_58, vutm}@.vnu.edu.vn, nthicham@hpmu.edu.vn

Abstract— Sentiment analysis or opinion mining is one of the major tasks of NLP (Natural Language Processing). Sentiment analysis has gained much attention in recent years. In this paper, we aim to tackle the problem of sentiment polarity categorization, which is one of the fundamental problems of sentiment analysis. We proposed a Vietnamese Sentiment Analysis system with detailed descriptions, which classifies the opinion of a review into one of three types: "positive", "negative" and "neutral". Data used in this study are provided by The Fourth International Workshop on Vietnamese Language and Speech Processing (VLSP2016). Our system has achieved an accuracy score of 81% on the sample set and 76.38% on the final test set.

Keywords—Vietnamese sentiment analysis, maximum entropy, SVM, product reviews

I. INTRODUCTION

Sentiment is an attitude, thought prompted by feeling of people. In recent years, the explosion of social networking sites, blogs and review sites provide a lot of valuable information [1]. Millions of people express their uninhibited opinions about product features, etc. via posting their own content through various social media like online social networking sites. Therefore, sentiment analysis has a strong fundament with the available support of massive online data.

However, these types of online data have some problems that prevent the process of sentiment analysis. Firstly, the quality of their opinions cannot be guaranteed since people can freely post their own content. For example, some spammers post spam or fake opinions on forums. Secondly, because of a free writing style, people use abundant emoticons, abbreviations, slang, etc. It is very difficult for analysis and understanding exactly product reviews in review-level categorization.

From a technical point of view, there are some approaches on resolving sentiment analysis, namely *machine learning*, *lexicon-based*, *statistical* and *rule-based* approaches. The machine learning method uses several learning algorithms to determine the sentiment by training on a labeled dataset. The lexicon-based approach involves calculating sentiment polarity for a review using the semantic orientation of words or sentences in the review. The rule-based approach considers different rules for classification like dictionary polarity, negation words, emoticons, mixed opinions. Statistical models represent each review as a mixture of latent aspects and ratings.

In this paper, we focus on classifying the opinion of a review into one of three types, which are "positive", "negative" and "neutral". Based on sample datasets provided by VLSP campaign, we built a Vietnamese Sentiment Analysis system using various machine learning methods, along with effective selected features. In section 2, we describe the specific task and datasets provided. Section 3 presents our approach with a hybrid machine learning model to resolve the task. The evaluation and analysis is given in Section 4. Finally, section 5 discussed our conclusions and directions for future work.

II. SA TASK DESCRIPTION

The main aim of the campaign VLSP 2016 on sentiment analysis for Vietnamese language is to evaluate the ability of classifying Vietnamese reviews/documents into one of three categories: "positive", "negative", or "neutral". This task only focuses on one language - Vietnamese. Vietnamese has specific characteristics different from other languages as English. Unlike English, Vietnamese is more very difficult and complicated about grammar, vocabulary, etc. Therefore, sentiment analysis methods successfully applied in the English corpus could not make the corresponding result for data use Vietnamese. It requires a separately optimal model which is more suitable for Vietnamese analysis sentiment.

Sample dataset provided by VLSP campaign includes three files: POS file (positive reviews), NEU file (neutral reviews), NEG file (negative reviews) with the same 1700 reviews on each file. These files are collected from comments on mobile review forums and labeled by people. The data is balanced for training our model. The test dataset contains 1050 reviews for evaluating the model. To understand and resolve the problem better, we carry out statistical analysis on training dataset in the below table.

Measure	NEG file	NEU file	POS file
The number of sentences	2841	3510	2792
Sum of characters	195082	297831	179714
Average length of each sentence	69	85	64

Sum of all words	38734	58343	35346
Sum of unique words	7256	9078	6592

TABLE 1. SOME STATISTICS OF THE TRAINING DATASET.

In general, measures between NEG file and POS file are more balanced, while these figures for NEU file are larger than two remaining files. It shows that when giving a review on positive or negative opinion, a user tends to write more concisely. Reviews on neutral opinion are often longer because they can contain both positive and negative opinions although they are neutral when combining two opinions. Besides, a neutral review often includes comparable sentences, analysis sentences on aspects of products. The average length of each review is quite short so it can be difficult to hold syntactic information as well as other important information.

Another problem is negation opinion. For example, a positive opinion is "Sản phẩm này *đẹp*"/"This product is *nice*" but this opinion becomes negative when adding "không"/"not" as: "Sản phẩm này *không đẹp*"/ "This product is *not nice*". Besides, the word "không"/"no" is written in a free style writing in Vietnamese such as: 'khong', 'ko', 'khg', 'k', etc. These problems also hinder the process of sentiment analysis. After the preprocessing (POS tagger), we statistic the number of unique nouns, verbs, adjectives in each file and the appearance frequency of each word in the below table.

Type of Word	NEG file	NEU file	POS file
Noun	766/15	825/22	679/15
Verb	2718/3	3357/4	2456/3
Adjective	1654/2	1994/2	1463/2

TABLE 2: SOME STATISTICS OF THE TRAINING DATASET.

Basically, the neutral file uses a number of very abundant adjectives. It can be easy to understand since neutral reviews contain both negative opinions and positive opinions. We also built two dictionaries: sentiment dictionary and intensity dictionary for supporting our model. From our model built, it is evaluated on the test dataset, depending on classifying result of each review. In the next section, we will present our approach to resolve the task.

III. OUR APPROACH

A. Features Selection

Based on survey training dataset and recent researches on sentiment analysis [2,3,4], we decided to utilize two main features for machine learning models, including n-gram feature and phrase feature.

Firstly, n-gram is a very successful approach to represent text because of holding word sequences. The benefit of using n-gram instead of single words as features comes in being able to capture some dependencies between the words and the importance of individual phrases. Thus, N-gram is used for developing features for supervised machine learning model such as SVM, Maxent. We intend to use Bigram for machine learning models. However, since the number of n-gram features is quite large because of abundant sample dataset, we filter n-gram features based on chi-square measure to remain important n-gram features.

Secondly, to use phrase feature, we built manually two dictionaries: Intensity Dictionary (191 words) and Sentiment Dictionary (5760 words). Some keywords in Intensity Dictionary such as 'rất'/'very', 'khá'/'quite' and Sentiment Dictionary contains some terms like 'đẹp'/'nice', 'tốt'/'good'. Since a sentiment word and an intensity word often appear together in a sentence, we find out the occurrence of them in a maximum length of 3 words. We use it as an important feature for machine learning models.

B. Classifier Selection

In this task, we select three typical machine learning methods using for analysis sentiment task: Maximum Entropy, Support Vector Machine and Perceptron. We access three models separately and compare the performance among them.

IV. EVALUATION

A. Experimental Data

As said above, we received sample datasets including three files: POS file (positive reviews), NEU file (neutral reviews), NEG file (negative reviews) with the same 1700 reviews on each file. Firstly, we use 10-folds cross-validation to evaluate three machine learning methods. These results are presented in the below table:

	POS	NEU	NEG
Perceptron	65.88	51.07	61.78
SVM	78.59	72.12	73.14
MaxEnt	85.59	79.66	82.41

TABLE 3: 10 FOLDS CROSS-VALIDATION ON THREE VARIOUS METHODS

As can see from the above table, MaxEnt method gives the best result. Thus, we intend to use Maxent model for classifying review opinions for test dataset (1050 reviews provided by VLSP). The result is given in the table below:

	Precision	Recall	F1
Accuracy	76.38		
Positive	75.85	89.71	82.2
Negative	79.88	76.0	77.89

TABLE 4: THE EVALUATION RESULT ON THE TEST SET.

B. Experimental Results and Analysis

In general, the result is quite satisfactory. It proves the effective of two main features selected as well as Maxent machine learning method.

V. Future Work and Conclusion

We intend to select and build proper features to improve the performance of the current system. Besides, we will use some hybrid models to achieve better results in the future.

References

[1] Thi-Ngan Pham, Thi-Hong Vuong, Thi-Hoai Thai, Mai-Vu Tran, Quang-Thuy Ha, Sentiment Analysis and User Similarity for Social Recommender System: An Experimental Study

[2] Huyen-Trang Pham, Tien-Thanh Vu, Mai-Vu Tran, Quang-Thuy Ha, A Solution for Grouping Vietnamese Synonym Feature Words in Product Reviews

[3] Yelena Mejova, Padmini Srinivasan, Exploring Feature Definition and Selection for Sentiment Classifiers, 2011, Association for the Advancement of Artificial Intelligence

[4] Subhabrata Mukherjee, Pushpak Bhattacharyya, Feature Specific Sentiment Analysis for Product Reviews.

[5] Xing Fang and Justin Zhan, Sentiment analysis using product review data, 2015 Fang and Zhan, Journal of Big Data.