UIT-VSFC: Vietnamese Students' Feedback Corpus for Sentiment Analysis

Kiet Van Nguyen*, Vu Duc Nguyen†, Phu X. V. Nguyen‡, Tham T. H. Truong§ and Ngan Luu-Thuy Nguyen¶
University of Information Technology
Vietnam National University - Ho Chi Minh City, Vietnam
Email:*kietnv@uit.edu.vn, {†14521100, ‡14520685, §14520824}@gm.uit.edu.vn, ¶ngannlt@uit.edu.vn

***Abstract*—Students' feedback is a vital resource for the interdisciplinary research combining of two fields: sentiment analysis and education. To strengthen the sentiment analysis of the Vietnamese language which is a low-resource language, we build a Vietnamese Students' Feedback Corpus (UIT-VSFC), a free and high-quality corpus for research on two different tasks: sentiment-based and topic-based classifications. In this paper, we present the methods of building annotation guidelines and ensure the annotation accuracy and consistency of this corpus. The resource consists of over 16,000 sentences which are human-annotated on the two tasks. To assess the quality of our corpus, we measure the inter-annotator agreements and classification accuracies on our UIT-VSFC. As a result, we achieved 91.20% of the inter-annotator agreement for the sentiment-based task and 71.07% of that for the topic-based task. In addition, the best results are of baseline model as the Maximum Entropy classifier with 87.94% and 84.03% of the overall F1-score of the sentiment-based and topic-based tasks respectively. These results illustrate that the corpus is reliable and helpful resource for research.**

I. INTRODUCTION

Students' feedback is not only a powerful source for improvement of education quality, but also an important resource for research on sentiment analysis and education. Creating a corpus related to education is not only beneficial for the development of sentiment analysis for the Vietnamese language but also is really helpful for research on education, especially in Vietnam. We have two primary motivations for building this corpus. First, there are few of Vietnamese corpora published for scientific research on natural language processing. Second, education domain is one of the fascinating domains due to the prized contributions to the development of the country in general, and the development of education in particular. The development of tools to improve the education quality is completely essential and beneficial for a range of institutes or universities. After building the dataset, we plan to build automatic students' feedback classification tools to support effective management of training and learning activities of a university.

The UIT-VSFC corpus is annotated for two different tasks: sentiment-based and topic-based classifications. Ensuring the quality of corpus is one of the most important issues in the annotation process of the corpus. Therefore, we build a clear and complete annotation guidelines to control the annotation consistency and accuracy of the corpus. The annotation guidelines are designed for two separate tasks: sentiment-based and topic-based. The guidelines are not only useful for annotators but also important documents for researchers concerning sentiment analysis and education. We evaluate our corpus measuring the annotation agreements and evaluating classification experiments.

In this paper, our two key contributions can be summarized as follows.

1) One of the most primary contributions is to obtain Vietnamese students' feedback corpus (UIT-VSFC) for sentiment analysis and education research. As a result, we finally achieved over 16,000 sentiment-labeled and topic-labeled sentences. The corpus of the two tasks is publicly available for the research purpose at the NLP@UIT group's website[1].
2) Besides, we analyzed the characteristics of Vietnamese students' feedback and pointed out the challenges of annotating the corpus, which is crucial to building clear and complete annotation guidelines for annotators to control the quality of the corpus and support to build classification models on the corpus.

The remainder of the paper is organized as follows. In Section 2, we discuss the importance of students' feedback in opinion mining to education. The characteristics of Vietnamese students' feedback are presented in Section 3. We define two sub-tasks and annotation guidelines in Section 4. Then the corpus construction process is shown in Section 5. Corpus is evaluated based on inter-annotation agreements and classification experiments in Section 6. In Section 7, we analyze the results. Finally, we draw a conclusion and future work in Section 8.

II. STUDENTS' FEEDBACK AND SENTIMENT ANALYSIS

Students' feedback is one of the most powerful influences on learning and achievement, but this impact can be either positive and negative [1]. There are two main types of feedback: (1) feedback from lecturers to students which is to help students aware of their weakness and strength for improvement of study and (2) feedback from students to lecturers that allows lecturers to reflect and improve their teaching. In particular, students give opinions about a range of different issues. The feedbacks are collected at the end of each semester through a course survey. For example, students' feedback can express

[1]The website address of the research group of NLP@UIT: https://sites.google.com/uit.edu.vn/uit-nlp/

978-1-5386-6113-0/18/$31.00 ©2018 IEEE

what students do and do not understand and what they like or dislike about the lecture as well as what the lecturers' teaching is excellent or terrible.

In the early days of research, researchers focus on identifying sentiment mainly on news text [6], [7], [8], [9]. In recent years, there have been many the well-known shared tasks about opinion mining on Twitter, for examples, the SemEval-2013 Task 2 [10], and SemEval-2014 Task 9 (Rosenthal et al., 2014) did experiments on two sub-tasks consisting of a expression-level and a message-level polarity; and the SemEval-2015 Task 10 [11] added subtasks on topic-based message polarity classification. The SemEval-2016 Task 4 [12] dropped the phrase-level subtask and the strength of association subtask, and focused on sentiment with respect to a topic. The latest shared task of SemEval-2017 Task 4 [13] adding a new language, Arabic, and information about the Twitter users. Especially, Twitter provided good corpora for sentiment analysis and opinion mining [21]. In addition, there were several related work on opinion mining in e-learning data [2], [3], [4], [5].

Vietnamese is a language which has few resources for research on natural language processing. There are merely a few of research works on opinion mining, for examples, computer customers' reviews (laptops and desktops) [14], hotel reviews [15], electronic devices [16]. As can be seen that these works are primarily focused on e-commerce products, and there have been no sentiment experiments related to education domain as students' feedback data. Students' feedback can help the lecturers and the education managers understand learners' behaviors as well as learning and teaching activities which learners desire. Therefore, we determine to build a new Vietnamese students' feedback corpus to serve research community freely about opinion mining and education.

III. Description of Vietnamese Students' Feedback

In this section, we introduce the characteristics of Vietnamese students' feedback. Their pieces of feedback are short messages written freely by students. They contain a range of acronyms, spelling mistakes, emotions, icons and other characters that express exceptional meanings. For examples, there are many acronyms in the feedback, gv (giảng_viên - means: lecturers), sv (sinh_viên - means: students), hk1 (học_kỳ 1 - the first semester), etc. These abbreviations is very beneficial for building an acronym dictionary. Students' feedback also contains emotions including positive emotions: "<3", ":-)", ":D, ":v", ":)", "=)", etc, and negative emotions: ":(", ":-(", ":@", ":p" and so on. However, there are emotions that appear in both of polarities, for examples, the positive sentence "em chưa bao_giờ có ý_định nghỉ_học lớp của thầy :v" (means: "I have never intended to be quit your class :v") versus the negative sentence "Đôi_lúc thầy khó_tính quá :v" (means: "Sometimes you are too strict :v").

Students are concerned issues consisting of lecturers, curricula, facilities and others. For lecturers, there are two types of feedback for the lecturers. Firstly, students describe the actions, words, gestures and attitudes of their teacher in teaching activities which is either negative or improper or inconsistent. Besides, students give suggestion to their lecturers about what to make it better. For curriculum, students present their views about subjects, lessons, homework, assignments, syllabuses, etc. In addition, students alse recommended (using verbs such as suggest, think, see, need, etc.) ways to solve those problems better. For facility, students describe current matters of university facilities which influence teaching and learning activities. Their descriptions usually relate to the quality of facilities, or their suggestions may be good to solve these problems. Students show their comments related to other topics or no mention of any topic.

IV. Task Definition and Annotation Guidelines

A. Task Definition

Before presenting the procedure of building the corpus, we introduce two subtasks: sentiment-based and topic-based.

- Sentiment-based task: Given a Vietnamese students' feedback sentence, determine whether it expresses a positive, negative, or neutral/objective sentiment. For example, annotated sentences are shown in the Table I and the sentiment polarities are shown in the third column of the table respectively.
- Topic-based task: Given a Vietnamese students' feedback sentence, decide whether it expresses information related to Lecturer, Curriculum, Facility, or Others. For instance, we present labeled examples with related topics which are shown in the fourth column of Table I respectively.

B. Annotation Guidelines

Based on the characteristics of students' feedback, we build annotation guidelines for sentiment and topic tasks.

1) Sentiment-Based Task: Each sentence is labeled one of three sentiment polarities including positive, negative, and neutral. Annotation guidelines are briefly described as follows.

- Positive polarity: Sentences in which students use to express their satisfaction and compliments about elements of teaching activities such as lecturers, subjects, assignments, etc. For instance, the sentence "Giảng_viên rất nhiệt_tình và tận tâm" (means: "Lecturers are really enthusiastic and dedicated.") is annotated with the positive polarity.
- Negative polarity: Sentences express students' dissatisfaction, requests, and complaints related to Lecturer, Curriculum, Facility, and so on. An example sentence which is "Nội_dung môn_học này quá nhiều." (means: "The course content is too much for students.") is a sentence expressed the negative emotion.
- Neutral polarity: Sentences are incomplete or unclear in meaning or do not contain opinions, for instance, the sentence "giờ_giấc và cách giảng dạy" (means: time and teaching methods) is a neutral sentence which is actually a noun phrase and does not contain any words expressing emotions. Or, sentences are complete but do not express any students' feelings. For example, the sentence "Em

Table I
ANNOTATED EXAMPLES WITH SENTIMENT AND TOPIC CLASSIFICATIONS.

No.	Sentence	Sentiment	Topic
1	Giảng viên hướng dẫn tận tình và chu đáo. (The lecturer is enthusiastic and thoughtful.)	positive	Lecturer
2	Nội dung môn học chưa đủ và chưa đúng với đề cương. (Course contents are inadequate and incompatible with the syllabus.)	negative	Curriculum
3	Nhà trường cần cải thiện hệ thống điện và máy chiếu. (Schools need to improve electrical systems and projectors.)	negative	Facility
4	Em không có bất cứ một lời phê bình nào. (I have no comments.)	neutral	Others

cảm_ơn thầy." (means: "I thank you.") do not express any emotions, therefore, this sentence is a neutral feedback.

There are sentences which express both negative and positive emotions which are difficult to judge. They are often complex sentences with conjunctions such as nhưng (but), tuy_nhiên (however), etc. In this cases, the annotators are asked to choese the stronger polarity based on their understanding. We can see that the sentence "Thầy dạy dễ hiểu nhưng nhiều lúc hơi nhanh." (means: "The teacher's taught was easy-to-understand but sometimes fast.") is assigned to the negative polarity, although it contains the positive clause "The teacher's taught was easy-to-understand."

2) Topic-Based Task: In this task, we divided the topics which students are most interested in into four major classifications consisting of Lecturer, Curriculum, Facility, and Others:

- Lecturer: Sentences express emotions about teaching methods, attitudes, knowledge, and so on. Detecting these feedback related to lecturers help improve the quality of lecturers' teaching activities. As an example, the given sentence "Thầy có phương_pháp giảng_dạy rất hay." (means: "You have a great teaching method.") portrays the compliment about a lecturer, therefore, this sentence is assigned to the Lecturer topic.
- Curriculum: Sentences express the sentiment about matters related to subjects, curriculum, assignments, grades, session time, knowledge, labs, etc. For example, the given sentence "Môn_học này rất cần_thiết và bổ_ích." (means: "This subject is very necessary and beneficial." is assigned to the Curriculum category.
- Facility: There are students' feedback related to facilities such as computers, electric lights, projectors, fans, etc. For example, the given sentence "Máy_tính phòng lab chạy quá chậm." (means: "The lab computers run slowly.") is labeled with Facility.
- Others: Sentences are unclear or belong to other than the above topics, for instance, the sentence "Em không hài_lòng." (means: "I'm not satisfied.") do not express any certain topics, so it belongs to the others type.

V. CORPUS BUILDING PROCESS

Corpus creation includes four main phases: collecting data, pre-processing, and corpus annotation.

A. Collecting Data

Firstly, we collect students' feedback from a university in Vietnam, which were collected through student surveys by the university at the end of each semester from 2013 and 2016 with more than 16,000 students' sentences. Table II shows the detailed information of the collected data such as numbers of lecturers, numbers of students, numbers of subjects and numbers of students' responses . With the variety of students' responses on many subjects over years, this dataset is an important resource for research.

Table II
THE STATISTICS FOR THE NUMBER OF THE COLLECTED STUDENTS' FEEDBACK.

Academic Year	2014-2015	2015-2016	2016-2017
#Lecturers	175	184	227
#Students	2,235	2,856	3,789
#Subjects	143	160	175
#Responses	6,038	6,288	13,417

B. Pre-processing

We conduct the preprocessing phase to have normalized sentences due to taking advantage of the power of NLP tools effectively such POS taggers and parsers in the process of building models for sentiment analysis. Firstly, we segment students' feedback into sentences. After that we replace abbreviations with complete words and correct the misspelling. The abbreviations and spelling mistakes are stored in a dictionary for media data mining. In addition, we must protect the confidentiality of personal information by anonymizing person names with other symbols. As a result, we obtain over 16,000 normalized sentences in this phase.

C. Annotation Process

To ensure the quality of corpus, we try to build the basic guidelines as completely as possible before starting the annotation process. In addition, we also update the annotation guidelines during the annotation process, which is necessary because there are many complicated cases. In this annotation process, we ask the annotators to discuss difficult cases if they find and update the guidelines. After updating the guidelines, we re-train the annotators with the revised guidelines. Eventually, we evaluate the corpus based on inter-annotator agreements and classification experiments.

D. The resulted corpus

We obtained over 16,000 sentiment-labeled and topic-labeled sentences. To better understand the characteristics of the corpus, we continue to provide the detailed statistical information. Table III shows the distribution of sentiment and topic classes of the corpus. We recognize two important things: firstly, both positive and negative labels account for a large proportion with over 45% for each , while neutral data constitutes less than 5% of the total, and secondly, students' comments focus on two aspects such as Lecturer and Curriculum. The majority parts of the corpus belongs to the Lecturer type (71.76%), which makes the corpus highly imbalanced.

Table III
DISTRIBUTION OF THE CORPUS ON SENTIMENTS AND TOPICS (%).

	Positive	Negative	Neutral	Total
Lecturer	33.57	25.38	1.81	71.76
Curriculum	3.40	14.39	1.00	18.79
Facility	0.11	4.21	0.08	4.40
Others	1.61	2.01	1.43	5.04
Total	49.69	45.99	4.32	100

The distribution of sentiment-based sentences according to the length of the sentence was presented in Table IV. Students often use short comment sentences in their feedback with the length from one to fifteen words, accounting for more than 83% of the corpus. In terms of sentiments, sentences with negative polarity tend to be longer than those with positive polarity. This is due to the fact that negative sentences usually include reasons and suggests for solution. These negative sentences are remarkably valuable in the process of improving the quality of teaching and learning process of universities.

Table IV
DISTRIBUTION OF SENTIMENT-BASED SENTENCES ACCORDING TO THE LENGTH OF THE SENTENCE (%).

Lengths	Positive	Negative	Neutral	Overall
1-5	17.26	9.75	2.31	29.32
6-10	21.0	15.34	41.17	37.55
11-15	7.19	8.59	0.51	16.29
16-20	2.37	5.17	0.15	7.69
21-25	1.06	2.85	0.07	3.98
26-30	0.37	1.72	0.07	2.16
>30	0.40	2.57	0.04	3.01
Total	49.6	45.99	4.32	100

Table V shows the distribution of topic-annotated sentences according to their lengths. Students often express sentences in their topic-based feedback with one-to-fifteen-words lengths, more than 80% of the corpus. Besides, feedback related to three main topics has a tendency to use longer sentences (more than five words) to describe their comments.

VI. CORPUS EVALUATION

To evaluate the quality of our corpus, we conducted evaluations based on the inter-annotator-agreement measures and classification experiments.

Table V
DISTRIBUTION OF TOPIC-BASED SENTENCES ACCORDING TO THE LENGTH OF THE SENTENCE (%).

Lengths	Lecturer	Curriculum	Facility	Others
1-5	21.31	4.39	0.95	2.67
6-10	28.25	6.33	1.49	1.48
11-15	11.69	3.36	0.82	0.42
16-20	5.11	1.96	0.45	0.17
21-25	2.38	1.16	0.29	0.15
26-30	1.30	0.64	0.15	0.07
>30	1.72	0.95	0.25	0.09
Total	71.76	18.79	4.40	5.05

A. Inter-Annotator Agreements

In order to calculate inter-annotator agreements, we follow the agreement measure A_m [20] which is defined as the proportion of agreement after the chance agreement is removed from consideration, given by the following formula.

$$A_m = \frac{P_0 - P_e}{1 - P_e} \tag{1}$$

Where, the observed agreement (P_0) is the proportion of sentences which both of the annotators agreed on the classes pairs and the chance agreement (P_e) is the proportion of items for which agreement is expected by chance when the sentences are seen randomly. Table VI shows the agreement measure of the corpus on three annotators for two tasks. With regard to the sentiments, the A_m agreement and the observed agreement are very high with 91.20% and more than 95% respectively. Besides, the A_m agreement of topics is 71.07%, whereas the topic observed agreement is higher than that with 90.33%. As a result, the agreement of sentiment is much higher than that of topics, over 20%. This implies that the annotators identify the sentiment polarities easier than the topics.

Table VI
INTER-ANNOTATOR AGREEMENTS ON THE UIT-VSFC CORPUS (%).

Task	P_0	P_e	A_m
Sentiment-based	95.27	46.19	**91.20**
Topic-based	90.33	66.59	**71.07**

B. Experimental Results

1) Experimental Setting: We implement two classifiers as Naive Bayes and Maximum Entropy on the corpus. In addition, we use the five-fold cross-validation scheme for our experiments. We use four evaluation metrics such as accuracy, precision, recall and F1-score in our experiments. To create the training, development, and test datasets for future experiments, we take the original corpus and split it randomly into three parts which cover approximately 70.0%, 10.0%, and 20.0% of the corpus, respectively. Table VII shows statistics of these datasets in number of tokens and number of sentences.
We used two available tools for our classification experiments, the Datumbox framework [18] and the Stanford Classifier [17] for Naive Bayes and Maximum Entropy respectively.

Table VII
STATISTICS OF TRAINING, DEVELOPMENT, AND TEST SETS.

	Train	Dev	Test
Number of distinct tokens	2518	1,158	1,585
Number of tokens	163,485	21,640	45,017
Number of sentences	11,426	1,538	3,166

Table VIII
SENTIMENT AND TOPIC DETECTION ON TWO CLASSIFIERS

	Sentiment-based Task	Topic-Based Task
Naive Bayes	86.1	83.0
MaxEnt	87.9	84.0

Besides, we used the vnTokenizer tool [19] for Vietnamese word segmentation. Due to the first baseline experiments to evaluate our corpus, we only used simple features as n-grams (bigrams) in all our experimental settings.

2) Sentiment Classification Results: The performance of MaxEnt classification on sentiment-labeled data is better than that of Naive Bayes, as can be seen in Table VIII. The precision, recall and F1-score of the sentiment MaxEnt model on the test dataset are shown in Table IX. The results on both positive and negative are high with F1-score, 91.32% on the positive examples and 90.52% on the negative examples, whereas the performance on the neutral examples is low with the only 33.99%. This is because the corpus is unbalanced, for instance, the neutral data occupy 4.32% of the corpus.

Table IX
CLASSIFICATION MEASURES OF SENTIMENT DETECTION WITH THE MAXENT CLASSIFIER ON THE TEST DATASET (%).

	Precision	Recall	F1-score
Positive	91.69	90.94	91.32
Negative	87.69	93.54	90.52
Neutral	50.00	25.75	33.99
Average	87.71	88.66	**87.94**

3) Topic Classification Results: The performance of MaxEnt classification on topic-labeled data is better than that of Naive Bayes, as can be seen in Table VIII. Table X shows the precision, recall and F1-score of the topic MaxEnt model. The result of the Lecturer class is high with the F1-score of 91.12%, whereas the figure for the Curriculum class are low with 67.19% of the F1-score. The overall measure of topics is over 84%, lower than that of sentiments. In general, the quality of the classifications still needs further improvements.

Table X
CLASSIFICATION MEASURES OF TOPIC-BASED WITH THE MAXENT CLASSIFIER ON THE TEST DATASET (%).

	Precision	Recall	F1-score
Lecturer	90.17	92.10	91.12
Curriculum	67.07	67.31	67.19
Facility	90.65	86.90	88.73
Others	45.61	32.70	38.10
Average	83.78	84.40	**84.03**

VII. ANALYSIS

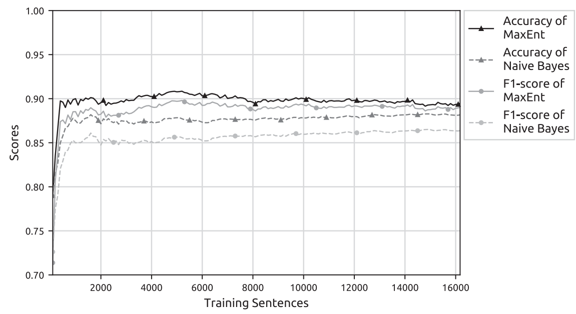

Figure 1. Learning curves of the sentiment classification models on the corpus.

As can be seen, Figure 1 shows the learning curves of the sentiment classification models on the development dataset. There are two important points in the figure. First of all, both evaluation measures of the models increase tremendously from 1 to 1,000 sentences, followed by a fluctuation on the dataset volumes of 1000–8000 sentences. After that, training data with over 8000 sentences witnesses a leveling off. Additionally, the MaxEnt models are much better than the Naive Bayes models in the sentiment-based task.

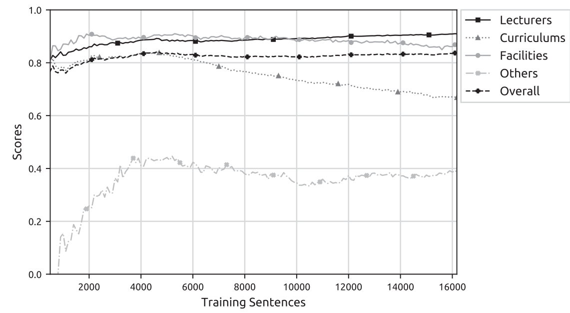

Figure 2. Learning curves of the topic classification models on the corpus.

Figure 2 shows the learning curves of the topic classification models in terms of the accuracy and F1-score measures on the corpus. It can be seen that the topic classification measures fluctuate on the dataset from 1 - 8,000 sentences, followed by an steady increase. Also, the measures of MaxEnt outnumber that of Navie Bayes in topic classification. Thus, MaxEnt is a better model with both the two tasks on the corpus. As can be seen from the Figure 3, the overall F1-score increase slightly due to the major effect of detecting the Curriculum and Facility topics. In particular, the F1-score of the Curriculum and Facility ones decrease when training data increases, while the Lecturer topic accounting for the major of the corpus show the reverse trend.

In addition, we analyze the common examples confusion of annotators, as can be seen in Figure 4, the annotator confusions of the Curriculum topic is very high, accounting for 33.00% of

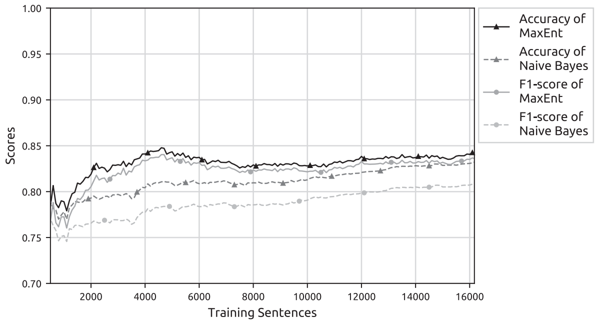

Figure 3. Learning curves (F1-score) of the topic classification with Maximum Entropy in terms of different topics on the corpus.

the total confusion examples and 18.79% of the corpus. This is a remarkable point to help us to consider more about analyzing the corpus to detect annotation errors and re-annotate the data if needed.

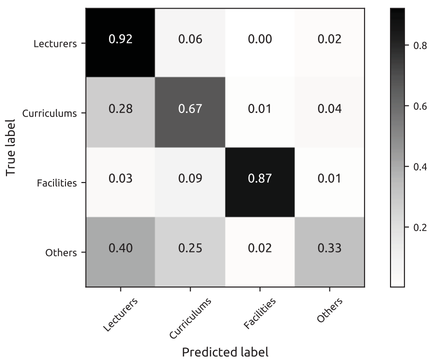

Figure 4. The topic confusion matrix of the MaxEnt classifier on the test dataset.

VIII. Conclusion and Future Work

We built an annotated corpus for sentiment analysis at Vietnamese sentence-level for free research. We archived about over 16,000 annotated sentences with the high annotation agreement of over 91% for sentiments and of more than 71% for topics. In addition, we obtained the best overall F1-score of sentiment polarities and four different topics of roughly 88% and over 84% on Maximum Entropy respectively. These are the first model results which are the baseline models to compare to coming experiments on this corpus. In future work, we are going to improve the quality of our corpus by error analysis of the corpus and re-annotating the corpus if needed. Besides, we will create a new corpus at document-level on the similar domain. We also plan to conduct experiments with deep learning models on this corpus.

References

[1] John Hattie and Helen Timeperley, The Power of Feedback, Review of Educational Research, 77, 81–112, 2007.

[2] Alvaro Ortigosa and José M Martín and Rosa M Carro, Sentiment Analysis in Facebook and Its Application to e-Learning, Computers in Human Behavior, 31, 527–541, 2014, Elsevier Science Publishers B. V.

[3] M. Martin and A. Ortigosa and R.M. Carro, Sentbuk: Sentiment analysis for e-learning environments, In International Symposium on Computers in Education (SIIE), 212–217, 2012.

[4] Zied Kechaou and M Ben Ammar and Adel M Alimi, Improving e-learning with sentiment analysis of users' opinions, Global Engineering Education Conference, 2011, 1032–1038.

[5] Dan Song and Hongfei Lin and Zhihao Yang, Opinion mining in e-learning system, In Network and Parallel Computing Workshops, 6, 2007, 788–792.

[6] Janyce Wiebe and Theresa Wilson and Claire Cardie, Annotating expressions of opinions and emotions in language, Language Resources and Evaluation, 39, 165–210, 2005, Kluwer Academic Publishers.

[7] Bo Pang and Lillian Lee and Shivakumar Vaithyanathan, Thumbs up?: Sentiment classification using machine learning techniques, In Proceedings of the Conference on Empirical Methods in Natural Language Processing - Volume 10 (EMNLP '02), 2002, 79–86, Association for Computational Linguistics.

[8] Minqing Hu and Bing Liu, Mining and summarizing customer reviews, In Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04), 2004, 168–177, New York, USA.

[9] Stefano Baccianella and Andrea Esuli and Fabrizio Sebastiani, SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, 2010, 2200–2204, Valletta, Malta.

[10] Preslav Nakov and Sara Rosenthal and Zornitsa Kozareva and Veselin Stoyanov and Alan Ritter and Theresa Wilson, SemEval-2013 Task 2: Sentiment analysis, In Proceedings of the 7th International in Twitter Workshop on Semantic Evaluation, 2013, 312–320, Atlanta, Georgia, USA.

[11] Sara Rosenthal and Preslav Nakov and Svetlana Kiritchenko and Saif Mohammad and Alan Ritter and Veselin Stoyanov, SemEval-2015 Task 10: Sentiment analysis in Twitter, In Proceedings of the 9th International Workshop on Semantic Evaluation, 2015, 451–463, Denver, Colorado, USA.

[12] Preslav Nakov and Alan Ritter and Sara Rosenthal and Fabrizio Sebastiani and Veselin Stoyanov, 2016 task 4: Sentiment analysis in Twitter, In Proceedings of the 10th International Workshop on Semantic Evaluation, 2016, 1–18, San Diego, California, USA.

[13] Sara Rosenthal and Noura Farra and Preslav Nakov, SemEval-2017 Task 4: Sentiment Analysis in Twitter, In Proceedings of the 10th International Workshop on Semantic Evaluation, 2017, 502–518, Vancouver, Canada.

[14] B. T. Kieu and S. B. Pham, Sentiment analysis for Vietnamese, In 2010 Second International Conference on Knowledge and Systems Engineering, 2010, 152–157.

[15] N. T. Duyen and N. X. Bach and T. M. Phuong, An empirical on sentiment analysis for Vietnamese, In 2014 International Conference on Advance Technologies for Communications (ATC 2014), 2014, 309–314.

[16] N. X. Bach and N. D. Tai and P. D. Van and T. M. Phuong, Mining Vietnamese comparative sentences for sentiment analysis, In 2015 Seventh International Conference on Knowledge and Systems Engineering (KSE), 2015, 162–167.

[17] D. Klein, The stanford classifier, ed: The Stanford Natural Language Processing Group, 2003.

[18] V. Vryniotis, Developing a Naive Bayes Text Classifier in JAVA, 2014, http://www.datumbox.com/machine-learning-framework/.

[19] Phuong H. Le and Huyen T. M. Nguyen and Azim Roussanaly and Vinh T. Ho, A Hybird Approach to Word Segmentation of Vietnamese Texts, Language and Automata Theory and Applications, 2008.

[20] Plaban Kr. Bhowmick and Pabitra and Anupam Basu, An Agreement Measure for determining Interannotator Reliability of Human Judgements on Affective Text, Proceedings of the Workshop on Human Judgements in Computational Linguistics, 2008, 58–65.

[21] Alexander Pak and Patrick Paroubek, Twitter as a Corpus for Sentiment Analysis and Opinion Mining, Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), 2010, Valletta, Malta.