Chapter 1

Introduction

1. Introduction

“What other people think” has always been important factor of information for most of us during the decision-making process. Long time before the widespread of World Wide Web, we often asked our friends to recommend an auto machine, or explain the movie that they were planning to watch, or confered Consumer Report to determine which television we would offer. But now with the explosion of Web 2.0 platforms such as blogs, discussion forums, review sites and various other types of social media … thus, comsumers have a huge of unprecedented power whichby to share their brand of experiences and opinions. This development made it possible to find out the bias and the recommendation in vast pool of people who we have no acquaintances.

In such social websites, users create their comments regarding the subject which is discussed. Blogs are an example, each entry or posted article is a subject, and friends would make their opinion on that, whether they agreed or disagreed. Another example is commercial website where products are purchased on-line. Each product is a subject that comsumers then would may leave their experience on that after acquiring and practicing the product. There are plenty of instance about creating the opinion on on-line documents in that way. However, with very large amounts of such availabe information in the Internet, it should be organized to make best of use. As a part of the effort to better exploiting this information for supporting users, researches have been actively investigating the problem of automatic sentiment classification.

Sentiment classification is a typical of text categorization which labels the posted comments is positive or negative class. It also includes neutral class in some cases. We just focus positive and negative class in this work. In fact, labeling the posted comments with cosummers sentiment would provide succinct summaries to readers. Sentiment classification has a lot of important application on business and intelligence [Bopang, survey sentiment]; therefore we need to consider to look into this matter.

As not an except, till now there are more and more Vietnamese social websites and comercial product online that have been much more intersting from the youth. Facebook[1] is a social network that now has about 10 million users. Youtube[2] is also a famous website supplying the clips that users watch and create comment on each clip… Nevertheless, it have been no worthy attention, we would investigate sentiment classification on Vietnamese data as the work of my thesis.

2. What might be involved?

As mentioned in previous section, sentiment classification is a specific of text classification in machine learning. The number class of this type in common is two class: positve and negative class. Consequently, there are a lot of machine learning technique to solve sentiment classification.

The text categorization is generally topic-based text categorization where each words receive a topic distribution. While, for sentiment classification, comsummers express their bias based on sentiment words. This different would be examine and consider to obtain the better perfomance.

On the other hands, the annotated Vietnamese data has been limited. That would be chanllenges to learn based on suppervised learning. In previous Vietnamese text classification research, the learning phase employed with the size of the traning set appropximate 8000 documents [Linh 2006]. Because anotating is an expert work and expensive labor intensive, Vietnamese sentiment classification would be more chanllenging.

3. Our approach

To date, a variety of corpus-based methods have been developed for sentiment classification. The methods usually rely heavily on annotated corpus for training the sentiment classifier. The sentiment corpora are considered as the most valuable resources for the sentiment classification task. However, such resources are very imbalaced in different languages. Because most previous work studies on English sentiment classification, many annotated corpora for English sentiment classification are freely available on the Internet. In order to face the challenge of limited Vietnamese corpus, we propose to leverage rich English corpora for Vietnamese sentiment classification. In this thesis, we examine the effects of cross-lingual sentiment classification, which leverages only English training data for learning classifier without using any Vietnamese resources. To archieve a better performance, we employ semi-supervised learning in which we utilize 960 unannotated Vietnamese reviews. We also examine the effect of selection features in Vietnamese sentiment classification by applying nature language processing techniques.

3. Related works

3.1 Sentiment classification

3.1.1 Sentiment classification tasks

Setiment categorization can be conducted at document, sentence or phrase (part of sentence) level. Document level categorization attempts to classify sentiments in movie reviews, product reviews, news articles, or Web forum posting [Bopang, 2002; BingLiu, 2004; Pang and Lee, 2004]. Sentence level categorization classify positve or negative sentiments for each sentence (Mullen and Collier, 2004, Pang and Lee, 2004]. The work on phrase level categorization capture multiple sentiments that may be present within a single sentence [Wilson et al. 2005]. In this study we focus on document level sentiment categorization.

3.1.2 Sentiment classification features

The types of features have been used in previous sentiment classification including syntactic, semantic, link-based and stylistics features. Along with semantic features, syntactic properties are the most commonly used as set of features for sentiment classification. These include word n-grams [Pang, 2002; Gamon, 2004], part-of-speech tagging [Pang, 2002].

Semantic features intergrate manual or semi-automatic annotate to add polarity or scores to words and phrases. [Turney, 2002] used a mutual information calculation to automatically compute the SO score for each word and phrase. While [Bing Liu, 2004; Bing Liu , 2005] made use the symnonym and antonym in WordNet to recognize the sentiment.

3.1.3 Sentiment classification techniques

There can be classified previously into three used techniques for sentiment classification. These consists of machine learning, link analysis methods, and score-based approaches.

Many studies used machine learning algorithms such as support vector machines (SVM) [Pang, 2002; Whilelaw, 2005; Xiao jun, 2009] and Naïve Bayes (NB)[Pang, 2002; Pang and Lee, 2004, Efron 2004]. SVM have surpassed in comparision other machine learning techniques such as NB or Maximum Entropy [Pang, 2002].

Using link analysis methods for sentiment classification are grounded on link-based features and metrics. Efron [2004] used co-citation analysis for sentiment classification of Web-site opinions.

Score-based methods are typically used in conjunction with semantic features. These techniques classify review sentiments throughby total sum of comprised positive or negative sentiment features [Turney, 2002; Fei, 2004].

3.1.4 Sentiment Classification Domains

Sentiment classification has been applied to numerous domains, including reviews, Web disscussion group, etc. Reviews are movie, product and music reviews [Pang, 2002; Bing Liu, 2004, 2005; Xiao jun, 2009]. Web discussion groups are Web forums, newsgroups and blogs.

In this thesis, we investigate sentiment classification using semantic features in compare to syntactic features. Becaused of the outperformance of SVM algorithm we apply machine learning technique with SVM classifier. We study on product reviews that are available corpus in the Internet.

3.2 Cross-domain text classification

Cross-domain text classification can be consider as a more general task than cross-lingual sentiment classification. In the case of cross-domain text classification, the labeled and unlabeled data originate from different domains. Conversely, in the case of cross-lingual sentiment classification, the labeled data come from a domain and the unlabeled data come from another.

In particular, several previous studies focus on the problem of cross-lingual text classification, which can be consider as a special case of general cross-domain text classification. Bel et al.(2003) study practical and cost-effective solution. There are a few novel models have been proposed as the same problem, for example, the information bottleneck approach (Ling et al., 2008), the multilingual domain models (Gliozzo and Strapparava, 2005), the co-training algorithm (Xijao Wan, 2009).

Chapter 3

The semi-supervised model with supportive knowledge

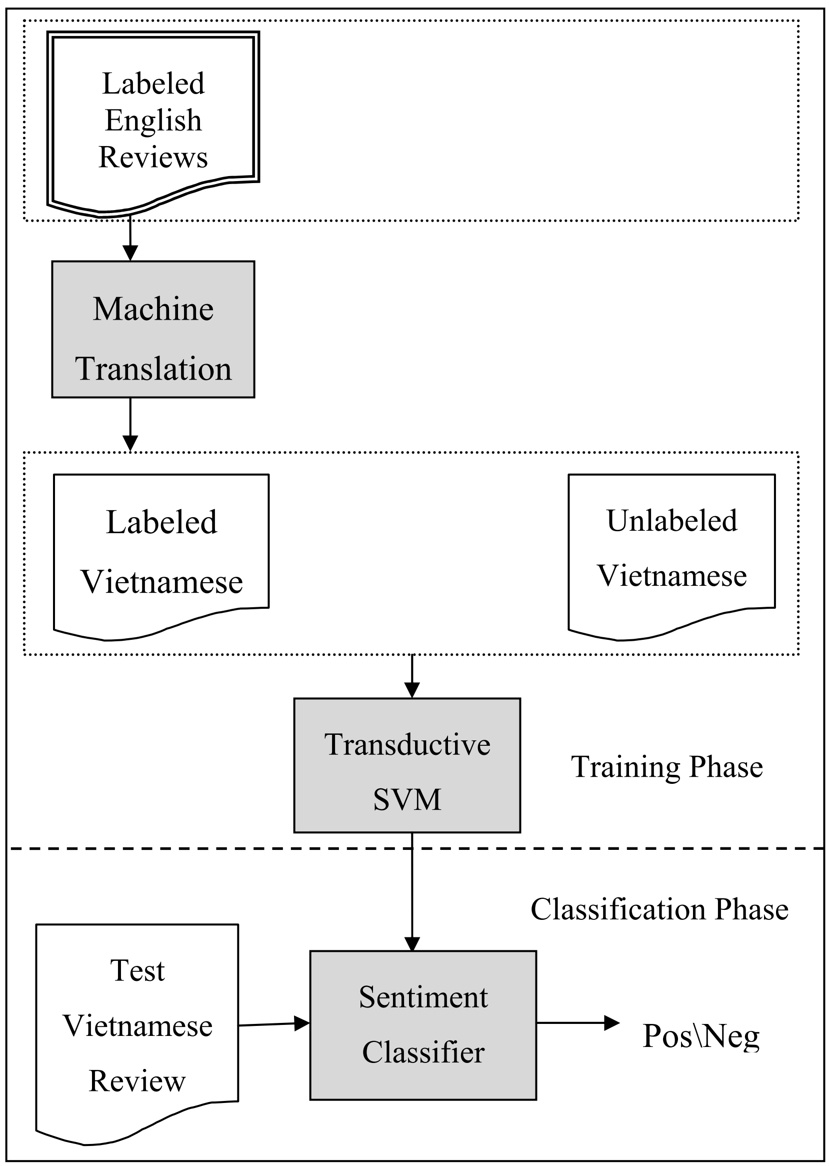
In this chapter, we describe the model that we proposed in section 3.1. Section 3.2 covers the machine translation which we employed. Section 3.3 describe some supportive information such as segmentation and part of speech tagging for Vietnamese languages in order to improve the classifier performance.

3.1 The semi-supervised model

In document online, the amounts of labeled Vietnamese reviews have been limited. While, the rich annotated English corpus for sentiment polarity identification has been conducted and publicly accessed. Is there any way to leverage the annotated English corpus. That is, the purpose of our approach is to make use of the labeled English reviews without any Vietnamese resources'. Suppose we has labeled English reviews, there are two straightforward solutions for the problem as follows:

1) We first train the labeled English reviews to conduct a English classifier. Lastly, we use the classifier to identify a new translated English reviews.
2) We first learn a classifier based on a translated labeled Vietnamese reviews. Lastly, we label a new Vietnamese review by the classifier.

As analysis in Chapter 2, sentiment classification can be treated as text classification problem which is learned with a bulk of machine learning techniques. In machine learning, there are supervised learning, semi-supervised learning and unsupervised learning that have been wide applied for real application and give a good performance. Supervised learning requires a complete annotated training reviews set with time-consuming and expensive labor. Training based on unsupervised learning does not employ any labeled training review. Semi-supervised learning employ both labeled and unlabeled reviews in training phase. Many researches [Blum,1998] [Joachims,1998] [Nigam, 2000] have found that unlabeled data, when used in conjunction with a amount of labeled data, can produce considerable improvement in learning accuracy.



The idea of applying semi-supervised learning has been used in [xiajun wan, 2009] for Chinese sentiment classification. [xiajun wan, co training] employ co-training learning by considering English features and Chinese features as two independent views. One important aspect of co-training is that two conditional independent views is required for co-training to work. From observing data, we found that English features and Vietnamese features are not really independent. As the wide – application of English and the Vietnamese origin from Latin language, Vietnamese language include a number of word-borrows. Moreover, because of the limitation of machine translator, some English words can have no translation into target language.

In order to point out the above problem, we propose to use the transductive learning approach to leverage unlabeled Vietnamese review to improve the classification performance. The transductive learning could make use full both the English features and Vietnamese features. The framework of the proposal approach is illustrated in Figure 3.1.

The framework contains of a training phase and classification phase. In the training phase, the input is the labeled English reviews and the unlabeled Vietnamese reviews . The labeled English reviews are translated into labeled Vietnamese reviews by using machine translation services. The transductive algorithm is then applied to learn a sentiment classification based on both translated labeled Vietnamese reviews and unlabeled Vietnamese reviews. In the classification phase, the sentiment classifier is applied to identify the review into either positive or negative.

For example, a sentence follow:

"*Màn hình máy tính này dùng được lắm, tôi mua nó được 4 năm nay*" (This computer screen is great, I bought it four years ago) will be classified into positive class.

3.2 Review Translation

Translation of English reviews into Vietnamese reviews is the first step of the proposed approach. Manual translation is much expensive with time-consuming and labor-intensive, and it is not feasible to manually translate a large amount of English product reviews in real applications. Fortunately, till now, machine translation has been successful in the NLP field, though the translation performance is far from satisfactory. There are some commercial machine translation publicly accessed. In this study, we employ a following machine translation service and a baseline system to overcome the language gap.

Google Translate 1: Still, Google Translate is one of the state-of-the-art commercial machine translation system used today. Google Translate not only has effective performance but also runs on many languages. This service applies statistical learning techniques to build a translation model based on both monolingual text in the target language and aligned text consisting of examples of human translation between the languages. Different techniques from Google Translate, Yahoo Babel Fish was one of the earliest developers of machine translation software. But, Yahoo Babel Fish has not translated Vietnamese into English and inversely.

Here are two running example of Vietnamese review and the translated English review. HumanTrans refers to the translation by human being.

Positive example: “Giá cả rất phù hợp với nhiều đối tượng tiêu dùng”

HumanTrans: “The price is suitable for many consumers”

GoogleTrans: Price is very suitable for many consumer object

Negative example: “Chỉ phù hợp cho dân lập trình thôi”

HumanTrans: “It is only suitable for programmer”

GoogleTrans: Only suitable for people programming only

3.3 Features

3.3.1 Word Segmentation

While Western language such as English are written with spaces to explicitly mark word boundaries, Vietnamese are written by one or more spaces between words. Therefore the white space is not always the word separator [Cam Tu, Word Segmentation].

Vietnamese syllables are basic units and they are usually separated by white space in document. They construct Vietnamese words. Depending on the way of constructing words, there are three type words, they are single words, complex words and reduplicative words. The reduplicative words are usually used in literary work, the rest widely applies.

For example, in the sentence

Sentence:	Tôi (*I*)	thích (*like*)	sản phẩm (*product*)	của (*this*)	hãng (*brand*)	Nokia
Type:	single word	single word	complex word	single word	single word	single word

Due to distinguishing the different usages of "khăn" (*tissue*) in "Bạn nên dùng khăn mềm lau chùi màn hình" (*You should clean the screen soft tissue*). The sentence does not indicate any sentiment orientation. Inversely, the word "khó_khăn" (*difficult*) in "Tôi thấy sử dụng công tắc bật tắt rất khó khăn" (*I found using the power switch is very difficult*) that indicates negative orientation. In order to fingure out that problem we perform segmentation on Vietnamese data before learning classifier.

3.3.2 Part of Speech Tagging

[Oanh, An experiment on POS, 2009]

Part of Speech tagging is a problem in Nature Language Processing. The task is signing the proper POS tag to each word in its context of appearance. For Vietnamese language, the POS tagging phase, of course, is performed after the segmentation words phase. For example, given a sentence:

Sentence:	Tôi thích sản phẩm của hãng Nokia (*I like Nokia products*)					
Segmentation phase	Tôi	thích	sản_phẩm	của	hãng	Nokia
POS phase	P (đại từ)	V (động từ)	N (danh từ)	E (giới từ)	N (danh từ)	Np (Danh từ riêng)

This serves as a crude form of word sense disambiguation: for example, it would distinguish the different usages of "đầu tiên" in "Nokia 6.1 là sản phẩm đầu tiên ra mắt thị trường" (indicating orientation) versus "Việc đầu tiên tôi muốn nói đến…" (it is a start a sentence)

3.3.2 N-gram model

N-gram model is type of probabilistic model for predicting the next item in a sequence. Till now, n-grams are used widely in natural language processing. An n-gram is a subsequence of n items (gram) from a given sequence. The items can be phonemes, syllables, letters or words according to the application. In the language identification systems, the characteristic should be base on the position of letters, therefore the items usually letters. On the other hand, in the text classification, the items should be words.

An n-gram of size 1 refers to a unigram, of size 2 is a bigram and similar to larger numbers. For this study, we focused on features based on unigrams and bigrams. We consider bigrams because of the contextual effect: clearly "tốt" (good) and "không tốt" (not good) indicate opposite sentiment orientation. While, in Vietnamese language "không tốt" is composed by two words "không" and "tốt". Therefore, we attempt to model the potentially important evidence.

As analysis above, due to the different of Vietnamese language to Western language such as English, we first apply in which each syllable are an item or a gram. And then, we use each word as an item in n-gram model after segmentation Vietnamese words. We also do another experiment by using a pair word and pos as an item.

For example, the sentence "*Tôi thích sản phẩm của hãng Nokia*" has the unigrams, bigrams, unigrams after segmentation words and unigrams after POS tagging as following:

Unigrams	**Bigrams**	**Unigrams after segmentation words**	**Unigrams after POS tagging**
Tôi, thích, sản, phẩm, của, hãng, Nokia	Tôi_thích, thích_sản, sản_phẩm, phẩm_của, của_hãng, hãng_Nokia	Tôi, thích, sản_phẩm, của, hãng, Nokia	Tôi-P, thích-V, sản_phẩm-N, của-E, hãng-N, Nokia-Np

Chapter 4

Experiments

4.1 Experimental set up

We establish experiments on Window NT operating systems and run on Java framwork with Java 1.6.0_03.

The tools employed in the experiments are illustrated in Table 4.1

No.	Name	Description
1	jTextOpMining	Author: Nguyen Thi Thuy Linh The utility: This module classifyies a review to be a positive or negative review. This tool is built on Java framework.
2	jTextPreProcessing	Author: Nguyen Thi Thuy Linh The utility: This modulde preprocess data. It removes noise, segment text, part of speech tagging text and exact features. This tool is contructed on Java 1.6.0_03 framework
3	svm_light	Author: Throasten Joachims Site: http://svmlight.joachims.org/ The utility: This tool learn a classifier and classifies a

		review into a positive or negative review.
4	Segmentation	Author: Site: http://vlsp.vietlp.org:8080/demo/?page=home The utility: This tool segment Vietnamese text
5	Pos	Author: Site: http://vlsp.vietlp.org:8080/demo/?page=home The utility: This tool part of speech tagging Vietnamese text

4.2 Data sets

The following three datasets were collected and used in the experiments:

Training English Set (Labeled English Reviews):

There are many labeled English copus available on the Web. We used the corpus contructed for multi-domain sentiment classification [Blitzer et al., 2007], because the corpus was large-scale and it was within domain that we experiment. The data set contains 7536 reviews, in which there are 3768 positive reviews and 3768 negative reviews for six different product types: camera, cell_phones, hardware, computer, electronics and software. In order to assess the performance of the proposed approach, each English review was translated into Vietnamese review in the training set. Therefore, we obtained a traning set consists labeled Vietnamese reviews.

Test Set (Labeled Vietnamese Reviews):

We collected and labeled 960 product reviews (580 positive reviews and 580 negative reviews) from popular Vietnamese commercial web sites. The reviews regard on such products as DVDs, mobile phones, laptop computers, television and fan electronic.

Unlabeled Set (Unlabeled Vietnamese Reviews):

We downloaded additional 980 Vietnamese reviews from Vietnamese commercial websites and employed that reviews to contruct the unlabeled set.

In addition, we collected and labeled 20 product reviews (10 positive and 10 negative reviews) from Vietnamese web sites. Those reviews will be employed to learn a classifier as a baseline.

Note that the training set and the unlabeled set are used in the training phrase, while the test set is blind to the training phrase.

4.3 Evaluation metric

As a first evaluation measure we simply take the classification *accuracy*, meaning the percentage of reviews classifed correctly. We also computed precision, recall and F-measure of the identification of the individual classes (positive and negative class). The metrics are defined the same as in general text categorization.

4.4. Features

Recall that the n-gram model we remind in Chapter 3. In this thesis, we use unigrams and bigrams as features. The features weight is calculated by TF (term frequency) weight that is often used in information retrival. This weight evaluate how important a word (or item) to a document in a corpus. The important increases proportionally to the number of times a word appears in the document. TF is defined as follows:

$$TF_{i,j} = \frac{\text{the number of occurences of the term } t_i \text{ in the document } d_j}{\text{the sum of number of occurences of all terms in document } dj}$$

4.5 Results

4.5.1 Effect of supportive knowlegde

In order to test our proposal, we built a classifier that use only 20 labled reviews from commercial Vietnamese websites and Unlabeled Set as a baseline method. And then, we compare the classification performance between the corpus making use of English labled data and the baseline method. The classification accuracies resulting are shown in line (1) and (2) respectively of Table 4.1. As a whole, our approach clearly surpass the baseline without the English corpus of 20%. Using the supportive knowlegde that is avaiable English corpus impove the classification performance significantly.

Furthermore, our approach also perform well in comparison to the supervised techniques that only employ the labeled data to learn the model shown in line (3). Because the number of unlabeled data is small for the number of labled data in the training set for semi-supervised learning, the classifciation performance is unremarkable increase.

In topic-based classification, the SVM classifier have been reported to use bag-of-unigram features to achieve accuracies of 90% and about for particular categories [Joachims, 1998, Nguyen Thi Thuy Linh, 2006] – and such results are for setting with more than two classes. This provides suggestive evidence that sentiment categorization is more difficult than topic classification, which coresponds to the mention above. Nonetheless, we still wanted to investigate ways to improve our sentiment categorization results; these experiments are reported below.

Table 4.1: The effect of supportive knowledge

No	Technique	Training size	# of features	Accuracy	Pre	Recall
(1)	Semi-supervised	7536 + 980	20428	**0.7125**	0.7107	0.7167
(2)	supervised	7536	20023	0.7062	0.7045	0.7104

(3)	Semi-supervised	20 + 980	2232	0.5181	0.5194	0.4851

4.5.2 Effect of extraction features

In order to improve the sentiment classification results, we performed tests based on the standard dataset that was descripted.

a, Using stopword lists.

In text categorization research [Joachims, 1998, Linh, 2006], they used some stoplists in their experiments. In topic based classification, important word is related the topic that it belongs, we want to receive much more that words. Generally, the more important words the large weight number they have. While, stopword appears almost documents, therefore, removing stopword in order to removing meaningless for classfication. In this study, we also make a test the effect of stopwords in documents. The classification results are illustrated in line (4) of Table 2. The result is smaller than using unigram alone. Does the important word is not effective in sentiment classification.

From the analysis above, we then test the influence of the vector weight. Recall that we represent each document d by a feature-count vector (n1(d), … nm(d)). In order to investigate whether reliance on frequency information could account for the higher accuracies of SVMs, we set ni(d) and nj(d) in the same weight. In other hand, if feature fi appears three times and feature fj appears one time in document d, fi and fj were weighted in the same number. Interestingly, this is in direct opposition to the observations of McCallum and Nigam (1998) with topic classification. We speculate that this indicates a difference between sentiment and topic categorization – perhaps due to topic being conveyed mostly by particular content words that tend to be repeated . As can be seen from line (2) of Table 4.2, the performance is not better than using only unigram with features frequency.

Table 4.2: The effect of selection features

No	Features	# of features	Accuracy	Pre	Recall	training time (CPU)	Count
(1)	unigram	20428	**0.7125**	0.7107	0.7167	671.66	freq
(2)	unigram	20428	0.6958	0.6992	0.6875	1107	pres
(3)	bigram	231834	0.7115	0.7192	0.6938	1450.44	freq
(4)	remove_stop + unigram	20409	0.6656	0.7076	05646	757.48	freq
(5)	Seg + unigram	23661	0.6958	0.6983	0.6896	523.27	Freq
(6)	pos + unigram	34906	0.6771	0.6693	0.7000	1807.66	freq
(7)	Subpos + unigram	40164	0.6628	0.6852	0.6021	1387.37	freq

b, Segmentation and Part of speech tagging

In line (5), we segment Vietnamese words and set each word be a features (unigram model). In complex words, the syllables are connected by "_". We apply the Segmentation module belonging to VLSP project [1]. The results is showned in Table 4.2.

Another step, we experimented with apending POS tags to every word by POS tag module of VLSP project. The POS tags module tags each word into subPos (see Appendix B) and the number of features will increase. Since observing data, we found that it is unnesscessary to use subPos as features, pos list (see Appendix B) is enough for distinguishing. A pair word and pos are formated as follow: [word]-[Pos].

As can be seen from line (6) of Table 4.2, a better performance is achieved by using only pos list, not subPos list. However, the effect of this pos information seems to be a wash: comparing line (1) and (6) of Table 4.2.

c, Bigrams

We set up an experiment using bigram model in which each feature is unigram or bigram. The connection between bigrams is "_". The result is shown in line (3) of Table 4.2. Seen from the table, the number of features in bigram experiment much more than the one in unigram experiment. It is also consuming time in training phase. However, the result is not better than unigram model. Since, we experiment no bigram model after segmentation words or POS tagging.

4.5.3 Effect of feature size

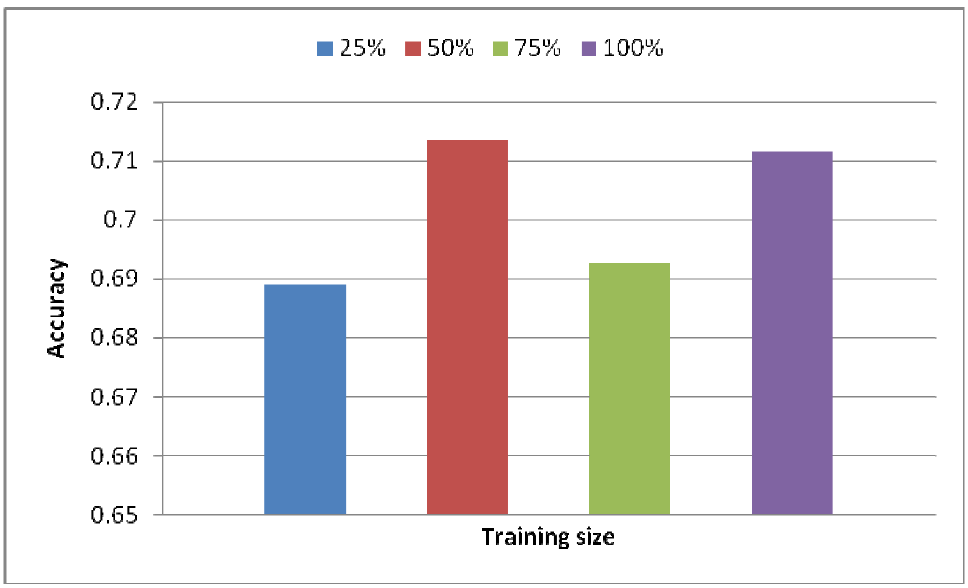


Figure 4.1: The effects of training size

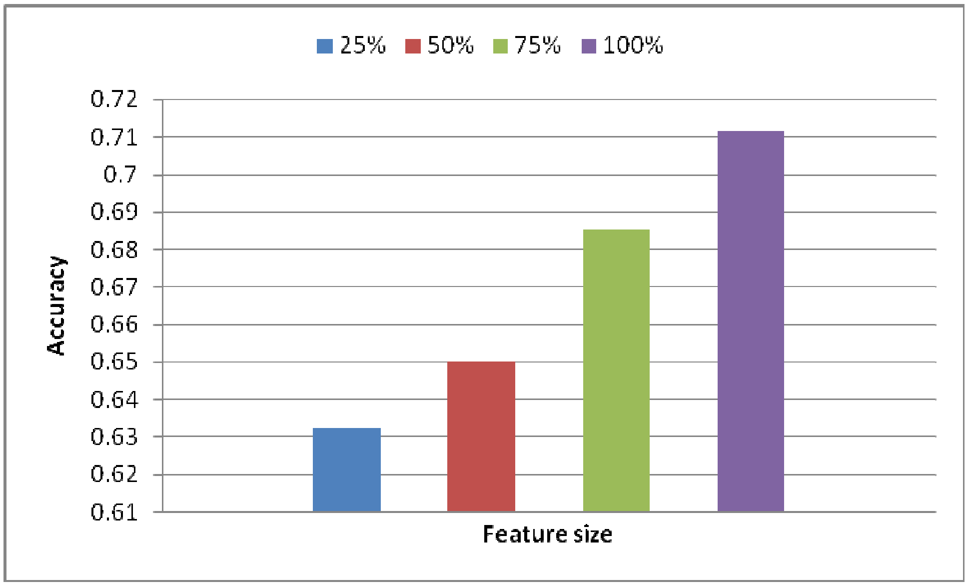


Figure 4.2: The effects of feature size

Appendix A

Stopword list

cả	chỉ	chính	chính vì	chính vì lẽ
cho	cho cả	cho dù	cho hay	cho hay những
có	có những	còn	cũng	cũng có
cũng có những	cũng không	cũng như	cũng như những	điều
điều không	do	dù	gì	giá
hay	hay không	hay những	hồ	hồ có
hoặc	hơn	không	không gì	lại
lại có	lại còn	lẽ	lẽ như	nên
nếu	ngay	ngay cả	ngay tại	như
như những	như thế	nhưng	những	nhưng cũng
nhưng không	nữa	tại	thế	thì
tuy	vậy	vì	vì lẽ	vì vậy

Appendix B

References

[1]. Nguyen Thi Thuy Linh, Phan Xuan Hieu, Ha Quang Thuy, Independent language for text classification, *Thesis*, 2006

[2]. X. Wan, Using Bilingual Knowledge and Ensemble Techniques for Unsupervised Chinese Sentiment Analysis, *Proceedings of the 2008 conference on Empirical Methods in Natural Language Processing*, pages 553-561, Honolulu, October 2008.

[3]. X. Wan, Co-training for cross-lingual Sentiment classification, *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, pages 235-243, Suntec, Singapore, 2-7 August 2009.

[4]. Bo Pang and Lillian Lee, A sentiment education: sentiment analysis using subjectivity summarization based on minimum cuts, *Proceedings of the ACL*, 2004.

[5]. Ahmed abbasi, Hsinchun Chen and Arab Salem, Sentiment Analysis in multiple languages: features selection for opinion classification in Web forums. *ACM Transactions on Information Systems*, vol.26, No.3, Article 12, June 2008

[6]. M. Hu and B. Liu. 2004a. Mining and summarizing customer reviews. In *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM Press New York, NY, USA.

[7]. M. Hu and B. Liu. 2004b. Mining Opinion Features in Customer Reviews. *In Proceedings of Nineteenth National Conference on Artificial Intelligence*.

[8]. Titov and R. McDonald. 2008. Modeling online reviews with multi-grain topic models. *In Proceedings of the 17h International Conference on World Wide Web*.

[9]. Sasha Blair-Goldensohn, Kerry Hannan, Ryan McDonald, Tyler Neylon, George Reis, Jeff Reynar,Building a Sentiment Summarizer for Local Service Reviews,

WWW Workshop on NLP Challenges in the Information Explosion Era (NLPIX), 2008.

[10]. J. Blitzer, M.Dredze and F. Pereira, 2007. Biograpies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification. *In Proceedings of ACL-07*

[11]. Ivan Titov, Ryan McDonald , A Joint Model of Text and Aspect Ratings for Sentiment Summarization, *Association for Computational Linguistics*, 2008.

[12]. M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger. Pulse: Mining customer opinions from free text. In *Proc. of the 6th International Symposium on Intelligent Data Analysis*, pages 121–132, 2005.

[13]. P.D. Turney and Micheal L. Littman, 2002, Unsupervised learning of semantic orientation from a hundred-billion-word corpus. Technical Repost EGB-1094, National Research Council Canada.

[14]. P. D. Turney, 2002. Thumbs up or thumbs down? Semantic orientations applied to unsupervised classification of reviews. *In Proceedings of ACL*

[15]. J.M. Wiebe, T. Wilson and M. Bell, 2001. Identifying collocations for recognizing opinions. *In Proceedings of the ACL/EACL Workshop on Collocation.*

[16]. L. Zhuang, F. Jing, and X. Zhu. Movie review mining and summarization. In *Proc. of the 15th ACM international conference on Information and knowledge management (CIKM)*, pages 43–50, 2006.

[17]. A. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In *Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2005.

[18]. T. Hofmann. Unsupervised Learning by Probabilistic Latent Semantic Analysis. *Machine Learning*, 42(1):177–196, 2001.

[19]. T. Joachims, 1999. Transductive inference for text classification using support vector machines, in Proceedings of ICML-99.

[20]. Bo Pang and Lillian Lee, Opinion mining and sentiment analysis

[21]. J. Carbonell, *Subjective Understanding: Computer Models of Belief Systems*. PhD thesis, Yale, 1979.

[22]. Y. Wilks and J. Bien, "Beliefs, points of view and multiple environments," in *Proceedings of the international NATO symposium on artificial and human intelligence*, pp. 147–171, USA, New York, NY: Elsevier North-Holland, Inc., 1984.

[23]. T. Joachims, 1998, Text categorization with support vector machines: Learning with many relevant features. In Proceedings of the European conference on Machine Learning (ECML), pages 137-142

[24]. T. Joachims, 1999, Making large-scale SVM learning practical. In Bernhard Sholkopf and Alexander Smola, editors, *Advances in Kernel Methods – Support Vector Learning*, pages 44-56. MIT Press

[25]. K. Nigam, John Lafferty, and Andrew McCallum, 1999, Using maximum entropy for text classification. *In Proceedings of the IJCAI-99, Workshop on Machine Learning for Information Filtering*, pages 61-67

[26]. A. McCallum and K. Nigam, 1998. A comparison of event models for Naïve Bayes text classification. In Proceedings of the AAAI-98 Workshop on Learning for Text Categorization, pages 41-48.

[27]. A. Blum and T.Mitchell, 1998, Combining labled and unlabled data with co-trainin, In *Proceedings of COLT-98*