không cần cụ thể vào Vietnamese

An empirical study on life-long learning for sentiment classification on Vietnamese

No Author Given

No Institute Given

Abstract. Statistical machine learning techniques have achieved promising results for sentiment classification. In the Internet age, it is essential to transfer the knowledge from the domain we learned to adapt an unfamiliar domain and still have a high accuracy of predicting if a user review is positive or negative. Life-long learning had proved its performance in topic modeling and sentiment classification in English. However, for low-resource language such as Vietnamese, there are many problems to deal with to get the similar performance. This paper presents an empirical study on Vietnamese sentiment classification. Our proposed method is based on life-long learning but combined with black-list and segmentation techniques specifically for Vietnamese. The results are promising, showing that the method is able to make relatively accurate predictions even when no labeled data is given.

Keywords: sentiment classification; Vietnamese; supervised learning, life-long learning

1 Introduction

In this paper, we focus on document-level sentiment classification. Sentiment analysis, which consists of the sentiment classification task, remains a popular topic for research and developing sentiment-aware applications [1]. This is due to the rapid growth of e-commerce and the Web age, which make the sentiment knowledge quickly become an advantage to contribute more values to market predictions. Sentiment classification is the task of classifying an evaluative text check grammar or an opinion is expressing a positive, negative or neutral sentiment. ~~To be consistent with some previous work [2]~~, our focus is binary sentiment classification, in which the classes are positive(+) and negative(-). Many machine learning techniques has been adopted to improve the accuracy and performance in sentiment classification in recent years. câu này ở đây có ý nghĩa gì?

However, in real-life situation, the training data may not be fully a representative of the test data in a target domain for sentiment analysis because some sentiment words in testing may not appear in the training data. Therefore, transferring the knowledge learned from other domains is a potential solution to improve the result of classification. There were many approaches proposed to solve the problem, such as life-long learning [3], transfer learning [4], self-taught learning [5] and other domain adaptation techniques [4].

đây có phải lí do cần transfer learning?

Chen et al.~~(2015)~~ [3] has had a novel approach of life-long learning for sentiment classification, which is based on Naive Bayesian framework and stochastic gradient descent. The method accomplishes great performance but faces difficulties due to the use of unigram. Although bag-of-words is often used in language modeling, a single unigram might not represent the relationship between words, while many phrases can express the polarity of sentiments. While experimenting the method, many unigrams which are popular verbs that do not express polarity such as “make”, “use”, “get” make a negative effect on results while the importance of these words can be deduced by using bigrams. Another example is ’have to’, which is a common phrase in negative text (but much less important in positive text), cannot be taken advantage of with unigram feature. ~~Furthermore, the experiments were on a dataset which has a reasonable proportion of negative reviews among domains, from 14 to 30%; which might not represent well the real-life situation.~~

especially in isolated language, where words are not separated by white spaces.

not true

In contrast to the availability and popularity of opinion-rich resources in English, as a resource-poor language, there are quite few accomplishments in the field of Vietnamese sentiment analysis. Bach et al [6] has leveraged rating-based features to improve sentiment classification on Vietnamese hotel reviews on Agoda. Also with dataset retrieved from Agoda, Duyen et al. [7] has had an empirical study on sentiment analysis. using? To the best of our knowledge, there is no study found on the field of life-long learning and there is no suitable dataset with a reasonable amount of reviews and variance of products to run life-long learning. in Vietnamese

With more than 15,000 reviews crawled from Tiki with 17 distinctive domains, this paper presents an empirical study on life-long learning for sentiment classification on Vietnamese. We apply a similar life-long learning approach to Chen et al. [3] did while combine with bigram feature and some Vietnamese processing techniques. Our proposed method outperforms the baseline methods in both Vietnamese and English datasets, which illustrate the potential of the use of our ~~features~~. tự nhiên nhảy đâu ra feature ở đây?

em đâu có xài hết 15K này?

propose an improvement ...

A ~~brief~~ summary of our contribution is expressed below:

- We combined the bigram ~~language model~~ with the Naive Bayesian optimization framework, which has better leveraged the phrases that contains sentiment and improved the performance on both Vietnamese and English datasets.
- With the crawled dataset ~~which express the real-life situation for Vietnamese sentiment classification~~, we experimented different kinds of processing techniques, some of which are typical to Vietnamese, an isolated language. Our experiment results in this paper show different results when applying these techniques. cái này đọc xong chả thấy đóng góp chỗ nào

2 Related Work

Our work is related to life-long learning, multi-task learning, transfer learning and domain adaptation. Chen and Liu has exploited different types of knowledge cite?

for life-long learning on mining topics in documents and topic modeling [8, 9]. Chen and Liu 2015 [3] also propose the first life-long learning approach for sentiment classification. Likewise, Ruvulo and Eaton[] developed a method for online multi-task learning in the lifelong learning setting, which maintains a sparsely shared basis for all task models. About domain adaptation, most of the work can be divided into two groups: supervised (Finkel and Manning 2009[], Chen et al. 2011) and semi-supervised(Kumar et al. 2010[], Huang and Yates 2010[]).

There are also many previous work on transfer learning and domain adaptation for sentiment classification. Yang, Si and Callon 2006[] proposed an approach based on feature-selection for cross-domain sentence-level classification. Other approaches include structural correspondence learning (Blitzer, Dredze and Perreira 2006), spectral feature alignment algorithm (Pan et al. 2010[]), CLF (Li and Zong 2008). Similar methods can be found in Liu 2012[].

In the field of sentiment analysis for Vietnamese, Duyen et al. [7] has published an empirical study which compared the use of Naive Bayes, MEM and SVM with hotel reviews. Also using the corpus from Duyen, Bach et al. 2015[] proposed the use of user-ratings for the task. Term feature selection approach is selected by Tran et al. 2011[], while Kieu and Pham [] investigate a rule-based system for Vietnamese sentiment classification. As that being said, to the best of our knowledge, there is no previous work on domain adaptation or life-long learning as well as a appropriate dataset for Vietnamese. why?

3 Our Proposed Method

trình bày sơ lược về LL, sau đó phần áp dụng của em như thế nào?

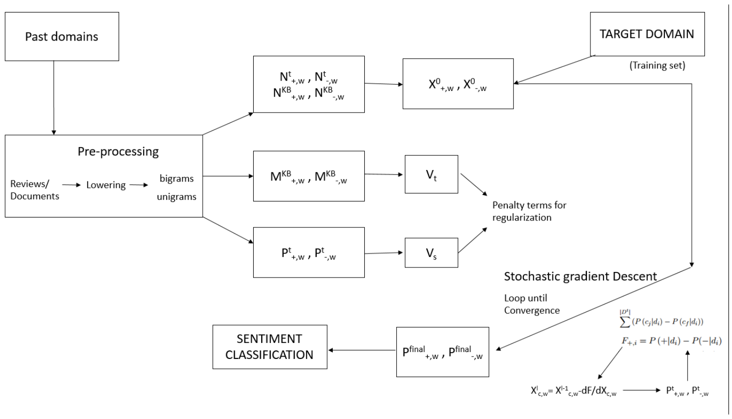
In this part, we introduce our system for Vietnamese sentiment classification in a life-long learning setting, which is a combination of components to analyze reviews from many domains. The system takes customer reviews from multiple types of products as input, in which each review is labeled as positive, negative or neutral based on how users rated them. Each review a.k.a evaluative document can contains multiple sentences. On our dataset, the average unigram per document on each domain varies from 66 to just above 75 unigrams. The input collection includes a number of past domains and one target domain. The knowledge gained from the input domain will be used to add values to the learning task on the target domain. Finally, with refined parameters, the system will classify reviews on the testing part of the target domain. The figure below indicate how the system work and the important parameters of the system. The system contains 3 main modules: pre-processing, optimization, sentiment classification.

- **Pre-processing**: This component is to process the raw reviews from the past domains and gain the parameters for learning task on the target domain. Since Vietnamese is an isolated language, we follow the maximum entropy approach of Dinh and Vu [10] to segment each document on the Vietnamese dataset. The segmentation task can help better modeling the sentiment adjectives which often contains two or more syllables, hence, provide a better vocabulary set for classification on Vietnamese using unigram feature. On both English and Vietnamese dataset, each review is tokenized into unigrams and bigrams to help us experiment both cases of modeling. After the text processing tasks, the system will try to extract the knowledge that will be used to optimize the classifier on the target domain. The main types of knowledge consists of:

- The Prior probability $P_+^t(w|c)$ and $P_-^t(w|c)$ where t is a past learning task. Because we choose to combine Naïve Bayes and bigram feature, we also store $P_+^t(w_i|w_{i-1})$
- Number of times a word appears in a positive or negative review: $N_{+,w}^t$, $N_{-,w}^t$. Similarly, the number of occurrences of w in the positive and negative documents are respectively $N_{+,w}^{KB} = \sum N_+^t$ and $N_{-,w}^{KB} = \sum N_-^t$
- Number of past tasks in which $P_{w|+} > P_{w|-}$ or vice versa: $M_{+,w}^{KB}$, $M_{-,w}^{KB}$. The two values are used to leverage domain knowledge via a penalty terms to penalize the words that less than a reasonable number of domains

– **Optimization**: This component follow the stochastic gradient descent with similar regularization techniques proposed by Chen et al. [3]. The component will use the knowledge above to update the virtual count X until convergence to get the best classifier for the target domain.

– **Sentiment classification**: To have a truly overall view of life-long learning for Vietnamese sentiment classification, we test the classifier which is just optimized for the target domain on different cases including segmentation and without segmentation task, unigram feature and bigram feature, with and without stop words.



vẽ lại hình này

We recommend the use of bigram feature instead of only unigram features on this type of sentiment classification. Wang and Manning [11] has proved that using bigram always improved the performance on sentiment classification. Due to the fact that there are a lot of phrases that can express sentiment in the documents, the bigram feature can help improve the original Naive Bayes framework applying unigram feature with classifier for phrase such as 'have to' in English or 'không thích'(dislike) in Vietnamese. Hence, using bigram feature captures modified and nouns. Bigram feature also help capture (adjective +

noun), (verb + adverb).e.g phrases in both Vietnamese and English. The model we choose to combine Naïve Bayes and bigram feature is described below:

$$P_{+|d} = \frac{P_+}{P_-}.P_+(w_0).P_+(w_1|w_0).P_+(w_2|w_1)\ldots P_+(w_n|w_{n-1}) \tag{1}$$

$$P_{-|d} = \frac{P_-}{P_+}.P_-(w_0).P_-(w_1|w_0).P_-(w_2|w_1)\ldots P_-(w_n|w_{n-1}) \tag{2}$$

whereas $P_+(w_{i+1}|w_i) = \frac{\lambda+N_{+,w_i w_{i+1}}}{\lambda+N_{+,w_i}}$ (and similar to $P_-(w_{i+1}|w_i)$).

4 Dataset and Experimental Study

4.1 Dataset

Labeled Vietnamese reviews We crawled the reviews from Tiki.vn, which is a large e-commerce website with quality reviews from the customers. It is a large corpus of 17 diverse domains or products. We follow the previous work [12] [13] to treat reviews with more than 3 stars as positive reviews and fewer than 3 stars as negative ones. The number of positive and negative reviews are shown as in the table ~~below~~: ref canh lại bảng cho đẹp

Product Types	Positive	Neutral	Negative
TrangDiem(Cosmetics)	3629	792	154
Dungcuhocsinh(Tools for students)	1803	164	37
Sanphamvegiay(Papers)	1778	144	343
Butviet(Pens and pencils)	1044	125	28
Dodungnhabep(Kitchen)	987	100	24
DauGoi(Shampoo)	347	59	18
Tainghe(Headphones)	698	90	18
DoDungChoBe(Baby)	658	61	14
Filehosobiahoso(Files)	157	47	14
Phukiendienthoaimaytin-hbang(Accessories)	583	32	13
Nuochoa(Perfume)	207	21	10
Thietbilamdep(Beauty equipment)	127	16	8
Butxoaxoakeo(Eraser)	311	43	7
Mayxaymayep(grinders)	395	38	7
Binhdunsieutoc(kettle)	107	13	5
Dungcuanuong(dining substances)	114	19	5
TranhDongHo(Dong Ho paintings)	274	10	5
Total (15394 reviews)	13219	1774	401

bôi đậm

thiếu gạch

Table 1. Names of 17 domains and the number of positive, neutral and negative reviews

Compared to some other datasets for multi-domain sentiment classification, ours express the real life situation where the number of reviews among domains are quite variant. As that being said, to experiment life-long learning, a mass of reviews among multiple product types are required, although there is no Vietnamese sentiment dataset that can meet the requirements. Compared to other e-commerce websites, each review published on Tiki was checked by the website administrators, which help reducing the rate of fake and low-quality reviews or reviews which does not contain tone marks.

The information packed in a single review in our dataset consists of product name, author name, rating, headline, bought-already, time of review, detailed review. Other review information including rating or headline can also be useful for other tasks of sentiment analysis. From 17 domains, 10 domains with reasonable size of the negative set are selected as another group for evaluation. The table 1 presents numbers of reviews on each type of label on each domain in the dataset. There are a total of 15,394 reviews with approximate 2.6% of negative ones.

Because of this proportion and the fact that the difference between number of reviews on each type may affect the classification task, for each domain, we selected randomly a maximum amount of 100 reviews on each class.

Labeled English reviews The corpus from [3] was utilized to compare directly with their life-long learning approach in English sentiment classification. The corpus contains reviews of 20 different products crawled from Amazon.

thêm chi tiết

4.2 Evaluation Metrics

nếu ngắn vậy có thể gom với 4.1

The evaluation method used is 5-fold cross validation. While dividing a domain into groups, we tried to keep the class distribution to avoid the case of no negative review on a segment due to the small proportion of negative class mentioned above. F1-measure on negative and positive class in types of Micro-average and Macro-average are applied.

4.3 Baseline

Our method is compared to VietSentiWordnet by Vu et al 2014 []. It is noted that VietSentiWordnet can only work on a single domain data. nêu tóm tắt cách làm ntn?

On English, we compare our model to Chen et al. [3] to illustrate the benefits of our models applied on life-long learning.

4.4 Bigram feature improves the classification on English dataset

We compare our result to the original life-long learning approach of Chen et al. [3](LSC) on both natural class distribution and balanced class distribution. On balanced class distribution, the accuracy our method exceeds LSC by 12% to get to a high of 95.84%. On the other hand, the F1 score for negative class outperform LSC with an average of 87.93%.

gom 4.5 với 4.6 lại

4.5 Vietnamese sentiment classification on the group of 10 domains

In this group, we compare the use of our proposed method (life-long learning for Vietnamese sentiment classfication-LLVi) with VietSentiWordnet in 4 types: without segmentation and bigram feature(LLVi-0), with bigram feature (LLVi-B), with segmentation (LLVi-S), with segmentation and bigram feature (LLVi-SB).

The table 2 has obviously shown that while the segmentation task only helps improving the performance on life-long learning with unigram feature, the use of bigram actually improved the result up to about 10%. Moreover, it is clear that the life-long learning approach has a huge advantage over VietSentiWordnet, which only works on the target domain.

4.6 Vietnamese sentiment classification on the group of 10 domains

On the total of 17 domains, we compare the results between LLVi with segmentation and unigram feature, LLVi with bigram feature and without segmentation, LLVi with bigram and with segmentation.

As seen in the table 3, the use of bigram make a huge improvement while the segmentation task hurts slightly on the performance.

Vietnamese, 10 domains	VietSentiWordnet	LLVi-0	LLVi-S	LLVi-B	LLVi-SB
Negative class, Micro average	40.85	80.39	81.25	91.48	90.41
Negative class, Macro average	33.21	66.91	72.51	75.19	74.89
Positive class, Micro average	NA	94.71	NA	97.86	NA
Positive class, Macro average	84.05	94.34	NA	97.86	NA

Table 2. F1-measure on 10 datasets of Vietnamese reviews (%)

chỉnh lại table

$Vietnamese, 17domains$	$LLVi-S$	$LLVi-B$	$LLVi-SB$
$Negativeclass, Microaverage$	75.09	89.45	89.24
$Negativeclass, Macroaverage$	70.83	79.03	78.31

Table 3. F1-measure on 17 datasets of Vietnamese reviews (unit: %)

5 Conclusion

viết lại phần này

In this paper, we conducted an experimental study on life-long learning for sentiment polarity classification. We firstly described the framework which our method is based on and the Vietnamese dataset we crawled to work with the task which contains 15,394 reviews among multiple products. After that, we propose a model which combine the life-long learning framework based on Naive Bayesian classification and bi-gram feature. The proposed method has been examined on both English and Vietnamese dataset.

The experiments examines the use of segmentation task and bi-gram feature on life-long learning classification. On both datasets, we outperforms the baselines significantly and show the potential of using bigram feature on life-long learning for sentiment classification.

In the future experiments, we will examine the use of emoticons on the Vietnamese dataset and prepare a larger dataset that may fit the real-life situation more.

References

1. Pang, B., Lee, L.: Opinion mining and sentiment analysis (2008)
2. Liu, B.: Sentiment analysis and opinion mining (2012)
3. Chen, Z., Ma, N., Liu, B.: Lifelong learning for sentiment classification. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics

and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Beijing, China, Association for Computational Linguistics (2015) 750–756
4. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. on Knowl. and Data Eng. **22** (2010) 1345–1359
5. Raina, R., Battle, A., Lee, H., Packer, B., Ng, A.Y.: Self-taught learning: Transfer learning from unlabeled data. In: Proceedings of the 24th International Conference on Machine Learning. ICML '07, New York, NY, USA, ACM (2007) 759–766
6. Bach, N.X., Phuong, T.M.: Leveraging user ratings for resource-poor sentiment classification. Procedia Computer Science **60** (2015) 322 – 331 Knowledge-Based and Intelligent Information Engineering Systems 19th Annual Conference, KES-2015, Singapore, September 2015 Proceedings.
7. Duyen, N.T., Bach, N.X., Phuong, T.M.: An empirical study on sentiment analysis for vietnamese. In: Advanced Technologies for Communications (ATC), 2014 International Conference on. (2014) 309–314
8. Chen, Z., Liu, B.: Mining topics in documents: standing on the shoulders of big data. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM (2014) 1116–1125
9. Chen, Z., Liu, B.: Topic modeling using topics from many domains, lifelong learning and big data. In Jebara, T., Xing, E.P., eds.: Proceedings of the 31st International Conference on Machine Learning (ICML-14), JMLR Workshop and Conference Proceedings (2014) 703–711
10. Dien, D., Thuy, V.: A maximum entropy approach for vietnamese word segmentation. In: Research, Innovation and Vision for the Future, 2006 International Conference on. (2006) 248–253
11. Wang, S., Manning, C.: Baselines and bigrams: Simple, good sentiment and topic classification. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Jeju Island, Korea, Association for Computational Linguistics (2012) 90–94
12. Blitzer, J., Dredze, M., Pereira, F.: Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, Association for Computational Linguistics (2007) 440–447
13. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? sentiment classification using machine learning techniques. In: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics (2002) 79–86