2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)

Sentiment Classification on Thai Social Media Using a Domain-Specific Trained Lexicon

Tanatorn Tanantong
Department of Computer Science, Faculty of Science and Technology, Thammasat University
Pathum Thani, Thailand
tanatorn@sci.tu.ac.th

Nuttapong Sanglerdsinlapachai
National Electronics and Computer Technology Center
Pathum Thani, Thailand
nuttapong.sanglerdsinlapachai@nectec.or.th

Udomsak Donkhampai
Zanroo Company
Nonthaburi, Thailand
udomsak@zanroo.com

***Abstract*—Social media contains many valuable documents which can be used to retrieve insights into any targeted topic. Sentiment analysis is a task to extract opinions of people out of their writing. Although a method using a sentiment lexicon is suitable for sentiment classification on social media documents, a lexicon created without consideration of any specific topic domain may not reflect actual intention in the domain of the considered documents. This study compares a sentiment lexicon created from documents in a specific domain with a generic lexicon. The experimental results show that using a specific lexicon improves classification performance, particularly on documents having non-neutral sentiment.**

Keywords—sentiment analysis, Thai social media, domain-specific lexicon, lexicon-based sentiment classification

I. Introduction

The more increasing of social media on the Internet, the more user-generated content there. Social media, e.g., Facebook and Twitter, as well as web boards and blogs, are public spaces on the Internet that users can express their opinions freely. Hence a vast amount of user-generated documents concerning various topics, e.g., products, films, hotels, and medical treatments, are published online. Those documents are valuable resources for retrieval insight into any topic. Sentiment analysis or opinion mining [1, 2, 3] is one of natural language processing (NLP) tasks used to gain such insight [4].

Sentiment analysis classifies given documents into different sentiment classes. Polarity orientations, i.e., positive, neutral, and negative classes, are frequently considered [1, 5]. Two classification approaches are used for sentiment analysis, i.e., a lexicon-based approach and a machine-learning-based approach. The former relies on terms with corresponding polarity orientations in a predefined vocabulary, called a sentiment lexicon [6, 7]. The latter uses many features extracted from amount of labelled documents to train a machine-learning model, and uses the model as a sentiment classifier [2, 8, 9].

For documents collected from social media, a lexicon-based sentiment classification method may be more suitable compared to a machine-learning-based one. A reason is that the typical length of documents is quite short, which is not suitable for training a machine-learning model [8]. Another reason is that sentiment orientation labelling on the vast amount of documents is time-consuming. However a sentiment lexicon, generated without consideration of document topic domains (i.e., a generic sentiment lexicon), may be not suitable compared to a sentiment lexicon created using documents from the same domain of the targeted documents (i.e., a specific lexicon).

This study aims to make a comparison between a generic lexicon (i.e., Polyglot [10]) and a specific lexicon, used in a lexicon-based sentiment classification method on Thai social media documents. The rest of this paper is organized as follows: Section II describes the proposed methods, both a sentiment lexicon creation and a lexicon-based sentiment classification. Section III presents our experiments, including, dataset, experimental settings, results and discussion. Section IV concludes our study and findings.

II. Proposed Method

A. Sentiment Lexicon Creation

In order to create a sentiment lexicon using a set of annotated document, each document is first tokenized into a list of terms and the polarity score for an individual term is then calculated. The polarity score for a term depends on frequencies of its occurrences in each polarity class of documents, i.e., positive, neutral, and negative. Precisely, the polarity score of a term t appearing in a training set T, is calculated by

$$score_T(t) = \frac{p^+(t) - p^-(t)}{p^+(t) + p^0(t) + p^-(t)}, \quad (1)$$

where $p^+(t)$ (respectively, $p^0(t)$ and $p^-(t)$) is the proportion of the number of term occurrences in the documents labelled as positive (respectively, neutral and negative) to the number of those documents.

B. Lexicon-Based Sentiment Classification

Using term polarity scores from a pre-created sentiment lexicon, the polarity score of a given document is computed as a sum of polarity scores of all terms occurring in the document. Assume that the document d consists of the terms t_1, t_2, …, and t_n, the polarity score of d is calculated by

$$score(d) = \sum_{i=1}^{n} score(t_i), \quad (2)$$

where $score(t_i)$ equals $score_T(t_i)$ if t_i appears in the training set T and equals zero, otherwise.

The sentiment of the document d is classified according to the resulting polarity score. If the absolute value of the score is less than a pre-determined threshold th, the sentiment of d is classified as neutral. If the score is greater than th (respectively, less than $-th$), the sentiment is positive (respectively, negative). To illustrate sentiment classification according to the polarity score, the sentiment of d, denoted by $sentiment(d)$, is considered by

978-1-7281-6486-1/20/$31.00 ©2020 IEEE

$$sentiment(d) = \begin{cases} \text{positive}, & \text{if } score(d) > th, \\ \text{neutral}, & \text{if } -th \le score(d) \le th, \\ \text{negative}, & \text{if } score(d) < -th. \end{cases} \quad (3)$$

III. Experiment

A. Dataset

The dataset used in this study is collected from varied on-line sources using keywords concerning a mobile service provider in Thailand for crawling. The sources consists of social media websites (e.g., Facebook, Instagram, Twitter, and YouTube), web boards (e.g., Pantip), and others (e.g., blogs and news webpages). Documents in the dataset are mainly written in Thai and English, and are manually annotated with their sentiments, i.e., positive, neutral, and negative. The total 4608 documents are separated into 124 (2.7%), 3852 (83.5%), and 632 (13.7%) documents with their sentiments being positive, neutral, and negative, respectively.

B. Experimental Settings

The lexicon-based sentiment classification using a lexicon trained from the specific domain (the specific-lexicon classifier) were compared to that using a generic lexicon (the generic-lexicon classifier), which was not take a domain of a training data in consideration. The generic lexicon used in this study was provided by Polyglot, the natural language processing library that we used for the tokenization and language detection processes. For the specific-domain trained lexicon, the whole dataset was used as the training set.

The classifications were evaluated using two measurements, i.e., accuracy values (for the whole dataset) and f-measure values (for each polarity class, i.e., positive, neutral, and negative). The values of the threshold for sentiment classification were varied from 0.0 (no threshold) to 0.95, increasing by 0.05.

TABLE I. Classification Results Obtained from the Generic-Lexicon Classifier and the Specific-Lexicon Classifier, in Terms of Accuracy and F-measure for Positive, Neutral, and Negative Sentiment Classes

Threshold	Generic-Lexicon Classifier				Specific-Lexicon Classifier			
	Accuracy	*F-measure (Positive)*	*F-measure (Neutral)*	*F-measure (Negative)*	*Accuracy*	*F-measure (Positive)*	*F-measure (Neutral)*	*F-measure (Negative)*
0.00	19.2%	2.6%	12.7%	**14.5%**	15.9%	5.6%	0.0%	17.7%
0.05	52.2%	**3.0%**	34.3%	13.2%	40.8%	6.1%	23.2%	22.2%
0.10	74.8%	2.1%	42.9%	6.0%	53.3%	6.5%	31.1%	25.1%
0.15	82.2%	0.5%	45.2%	3.4%	67.9%	7.3%	38.5%	29.6%
0.20	83.2%	0.0%	45.4%	1.2%	77.7%	8.7%	42.7%	33.2%
0.25	83.3%	0.0%	45.4%	0.2%	87.4%	**13.5%**	46.2%	**36.3%**
0.30	83.4%	0.0%	**45.5%**	0.2%	**87.7%**	12.8%	**46.4%**	32.0%
0.35	**83.6%**	0.0%	**45.5%**	0.2%	86.5%	7.5%	46.2%	21.6%
0.40	**83.6%**	0.0%	**45.5%**	0.0%	84.8%	5.6%	45.8%	9.3%
0.45	**83.6%**	0.0%	**45.5%**	0.0%	84.0%	3.6%	45.6%	4.0%
0.50	**83.6%**	0.0%	**45.5%**	0.0%	83.7%	0.0%	45.5%	0.6%
0.55	**83.6%**	0.0%	**45.5%**	0.0%	83.6%	0.0%	45.5%	0.2%
0.60	**83.6%**	0.0%	**45.5%**	0.0%	83.6%	0.0%	45.5%	0.0%
0.65	**83.6%**	0.0%	**45.5%**	0.0%	83.6%	0.0%	45.5%	0.0%
0.70	**83.6%**	0.0%	**45.5%**	0.0%	83.6%	0.0%	45.5%	0.0%
0.75	**83.6%**	0.0%	**45.5%**	0.0%	83.6%	0.0%	45.5%	0.0%
0.80	**83.6%**	0.0%	**45.5%**	0.0%	83.6%	0.0%	45.5%	0.0%
0.85	**83.6%**	0.0%	**45.5%**	0.0%	83.6%	0.0%	45.5%	0.0%
0.90	**83.6%**	0.0%	**45.5%**	0.0%	83.6%	0.0%	45.5%	0.0%
0.95	**83.6%**	0.0%	**45.5%**	0.0%	83.6%	0.0%	45.5%	0.0%

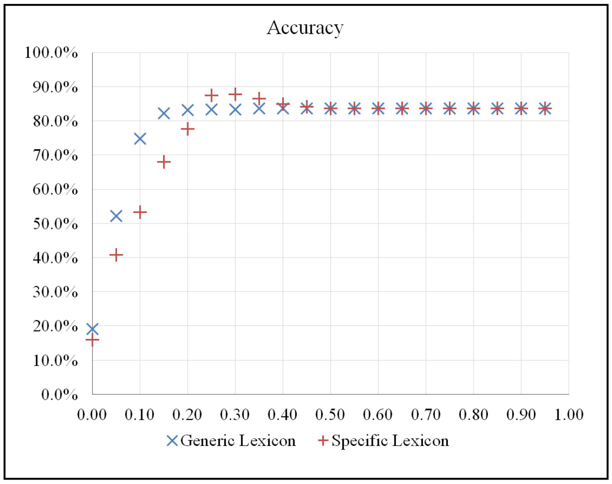

Fig. 1. Accuracy values obtained from the classifier using the generic lexicon and the classifier using the specific lexicon.

C. Results and Discussion

Table I shows results obtained from the generic-lexicon classifier (cf. from the second column to the fifth column) and from the specific-lexicon classifier (cf. from the sixth column to the ninth column). The specific-lexicon classifier yields the highest accuracy of 87.7%, while that obtained from the generic-lexicon classifier is 83.6%. For the f-measure, the specific-lexicon classifier yields the highest values of 13.5%, 46.4%, and 36.3% for the positive, neutral, and negative, respectively, classes. The generic-lexicon classifier yields the highest f-measure values of 3.0%, 45.5%, and 14.5% for the positive, neutral, and negative, respectively, classes.

To illustrate the effect of the threshold for sentiment classification, Fig.1 (respectively, Fig.2, Fig.3, and Fig.4) shows the values of accuracy (respectively, f-measure for positive, neutral, and negative class) at each value of the threshold, obtained from the specific-lexicon classifier and the generic-lexicon classifier.

In Fig.1, when the threshold value is increased, the accuracy values obtained from both the classifiers tend to increase until the threshold value of 0.35 and 0.30 for the generic-lexicon classifier and the specific-lexicon classifier, respectively. When the threshold is in the range of 0.00-0.20, the accuracy value obtained from the generic-lexicon classifier is higher than that from the specific-lexicon classifier. On the other hand, when the threshold is in the range of 0.25-0.40, the specific-lexicon classifier yields the higher accuracy value. For the threshold value of greater than 0.40, the accuracy values obtained from both the classifiers are the same value of 83.6%.

In Fig.2, the f-measure values for positive class are plotted. The f-measure value for positive class obtained from the specific-lexicon classifier is greater than that from the generic-lexicon classifier when the threshold is in the range of 0.00-0.45. When the threshold is greater than 0.45, both the classifiers yield the same f-measure value of 0.00. The highest f-measure for positive class obtained from the specific-lexicon classifier is 13.5% when the threshold is 0.25.

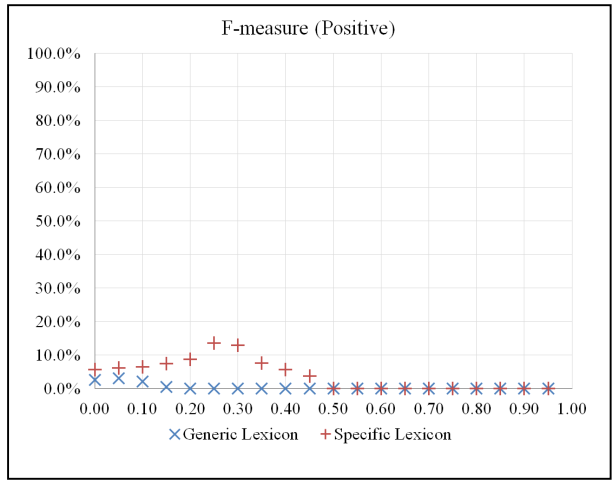

Fig. 2. F-measure values for positive class obtained from the classifier using the generic lexicon and the classifier using the specific lexicon.

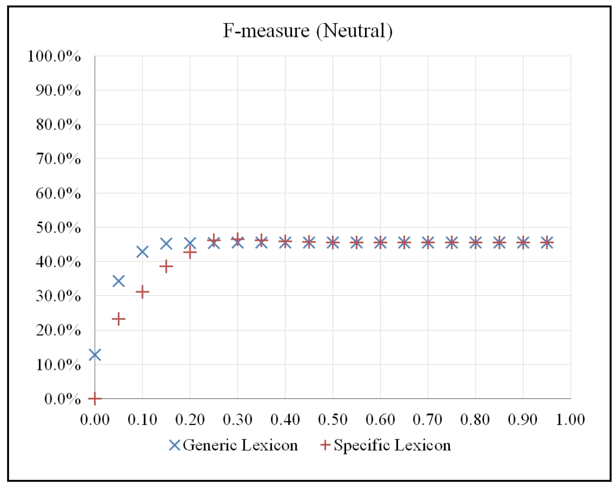

Fig. 3. F-measure values for neutral class obtained from the classifier using the generic lexicon and the classifier using the specific lexicon.

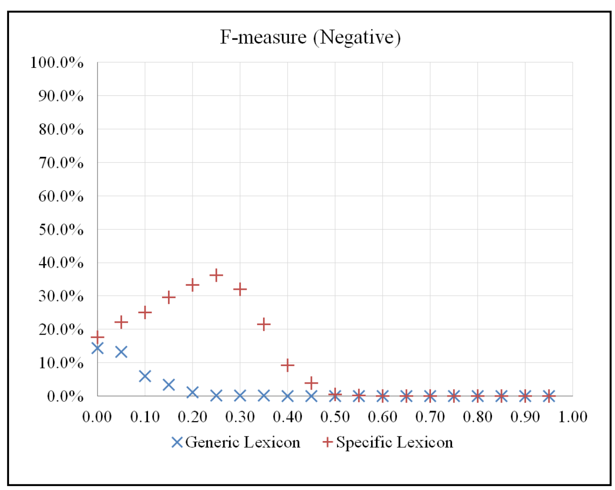

Fig. 4. F-measure values for negative class obtained from the classifier using the generic lexicon and the classifier using the specific lexicon.

In Fig.3, the change of the f-measure values for neutral class, according to the threshold, are similar to the accuracy. The f-measure values from both the classifiers increase until the threshold value of 0.30, and then converge to the f-measure value of 45.5%.

In Fig.4, the f-measure value for negative class obtained from the generic-lexicon classifier decreases when the threshold value increases. On the contrary, the f-measure value for negative class obtained from the specific-lexicon classifier increases until it achieves the highest value of 36.3% when the threshold value is 0.25, and then decreases when the threshold value is greater than 0.25. The f-measure value for negative class obtained from both the classifiers are 0.0% when the threshold value is greater than 0.55.

Observed from the four figures, the optimal threshold value, approximately 0.25, makes the performance of the specific-lexicon classifier achieve its highest point, higher than the generic-lexicon classifier. Using the specific lexicon, the calculated polarity scores of all positive and negative documents are between -0.60 and 0.60. The f-measure values for those two classes then become 0.0% when the threshold is greater than 0.60. Compared to the specific lexicon, the scores of those documents calculated using the generic lexicon are between -0.40 and 0.40. Since the range of polarity scores obtained from the specific-lexicon classifier is broader than that obtained from the generic-lexicon classifier, there is more possibility to correctly separate the document classes.

TABLE II. EXAMPLES OF TERMS IN THE SPECIFIC LEXICON CREATED USING THE DATASET

Terms	Meaning (Part of Speech)	Assigned Polarity Score
ยกเลิก	cancel (verb)	-0.802
เสีย	not usable (adjective)	-0.590
สัญญาณ	signal (noun)	-0.404
ไม่	not (adverb)	-0.350
เน็ต	network, Internet (noun)	-0.292
ปัญหา	problem (noun)	-0.165
ฟรี	free (adjective)	-0.116
บริการ	service (noun)	-0.091
โปร	promotion (noun)	0.136
คุ้ม	valuable (adjective)	0.292
แท้	genuine (adjective)	0.522

Table II shows examples of terms in the specific lexicon that occur frequently, i.e., their individual numbers of occurrences being more than those of 90% of all terms. Even some terms are noun, polarity scores are assigned to them through sentiment lexicon process. These terms may capture common sentiment expression in the specific domain. Therefore, the results obtained from the specific-lexicon classifier are more accurate than those from the generic-lexicon classifier.

IV. CONCLUSION

Each type of documents on social media is concerned with a specific domain and has a specific characteristic. A sentiment lexicon created for general purpose may not be suitable for documents on social media. This study makes a comparison between the classifier using the generic sentiment lexicon and that using the specific sentiment lexicon for sentiment classification on Thai social media documents. A basic lexicon-based sentiment classification method, summing polarity scores from terms appearing in a document, is used. Different values of the classification threshold, i.e., 0.00, 0.05, 0.10, ..., and 0.95, are considered to increase the classification performance.

The experimental results show that the specific-lexicon classifier yields higher performance, in terms of the accuracy and the f-measure, compared to the generic-lexicon classifier when an optimal value of the threshold is set. Particularly, the f-measure for the negative sentiment class (respectively, the positive sentiment class) is 36.1% (respectively, 13.5%) higher at the threshold value of 0.25.

Since the neutral documents are the large amount of the dataset, the classification accuracy depends on the neutral sentiment class. The further investigation includes a consideration of the balance of the sentiment classes and also uses of other generic sentiment lexicons. Using data from a different domain to create a trained lexicon and using cross-validation method to conduct experiments are also taken into consideration for the future work.

ACKNOWLEDGMENT

This work was supported by the National Innovation Agency (Public Organization), Grant No. P12-ST-61-05-043 and by the government budget of Thammasat University, Grant No. 7/2562. It was also partially supported by the Zanroo Company.

REFERENCES

[1] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” Found. Trends Inf. Ret., vol.2, no.1-2, pp.1-135, 2008.

[2] T. Tanantong and S. Thaskhwan, “Sentiment classification on Thai university mentions in twitters,” In Proceedings of the 11th National Conference on Information Technology, pp 79-84, NCIT, 2019.

[3] T. Tanantong, S. Thaskhwan and S. Kreangkriwanich, “A social media monitoring system for Thai university,” TNI J. of Eng. and Technol., vol. 7, no. 2, pp. 1-17, 2019

[4] B. El Haddaoui, R. Chiheb, R. Faizi, and A. El Afia. “Toward a sentiment analysis framework for social media,” In Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications, pp. 1-6, ACM, 2018.

[5] M. Hu and B. Liu, “Mining and summarizing customer reviews,” In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.168-177, ACM, 2004.

[6] R. Moraes, J. F. Valiati and W. P. G. Neto, “Document-level sentiment classification: An empirical comparison between SVM and ANN,” Expert Syst. Appl., Vol. 40, No. 2, pp. 621-633, February, 2013.

[7] S. Rustamov, E. Mustafayev and M. A. Clements, “Sentiment analysis using Neuro-Fuzzy and Hidden Markov Models of text,” In Proceedings of IEEE Southeastcon, Jacksonville, pp. 1-6, IEEE, 2013.

[8] W. Chamlertwat, P. Bhattarakosol, T. Rungkasiri, and C. Haruechaiyasak, “Discovering consumer insight from twitter via sentiment analysis,” J. Univers. Comput. Sci., vol. 18, no. 8, pp. 973-992, 2012.

[9] N. Sanglerdsinlapachai, A. Plangprasopchok and E. Nantajeewarawat, “Exploring linguistic structure for aspect-based sentiment analysis,” Maejo Int. J. Sci. Technol., vol. 10, no. 2, pp. 142-153, May, 2016.

[10] Rami Al-Rfou, Bryan Perozzi, and Steven Skiena, “Polyglot: Distributed word representations for multilingual NLP”. arXiv preprintarXiv:1307.1662, 2013.