# Thai Sentiment Analysis via Bidirectional LSTM-CNN Model with Embedding Vectors and Sentic Features

Thititorn Seneewong Na Ayutthaya
*Faculty of Information Technology*
*King Mongkut's Institute of Technology Ladkrabang*
Bangkok 10520, Thailand
59606006@kmitl.ac.th

Kitsuchart Pasupa
*Faculty of Information Technology*
*King Mongkut's Institute of Technology Ladkrabang*
Bangkok 10520, Thailand
kitsuchart@it.kmitl.ac.th

***Abstract*—Sentiment analysis is one of the most frequently performed tasks in Natural Language Processing that plays an important role in marketing research. It allows us to understand customer sentiment. The outcomes from this kind of analysis can be used to improve products and services. Recently, a Word2Vec model, a technique for word embedding (converting text into a number) has been developed and used successfully to a degree to get the sentiment of customers from the text responses that they provided. This work attempted to incorporate two more features–part-of-speech and sentic features–to make the analysis more accurate. The part-of-speech feature identifies the type of words that better convey various sentiments, while the sentic feature identifies the emotion underlying certain words. Combining Bidirectional Long Short-term Memory and Convolutional Neural Networks models with several combinations of the features mentioned, we performed a sentiment analysis of Thai children stories and found that the combination of all three features gave the best result at 78.89 % F1-score.**

***Index Terms*—Sentiment Analysis, Deep Learning, Word Embedding, Part-of-Speech, SenticNet2**

## I. Introduction

Currently, we live in the social media era. A lot of people use a social media platform to communicate and express their opinions every day. Information can be rapidly spread across social networks around the world. Many companies have invested in social media marketing and aim to listen to their customers' feedbacks on these platforms which are very important to them as they can use these feedbacks to improve their products or services. These feedbacks are mostly in text. In the past, humans are required to extract and analyse these feedbacks manually. The task is very labourious and time-consuming. Natural Language Processing (NLP) allows us to understand the context of a text. An important task of NLP is sentiment analysis that allows us to extract sentiment from text. This task can be done by computer in a very short time so that companies can respond to feedbacks rapidly.

In the past, sentiment analysis focused on extracting discriminative features such as term frequency and term presence [1]. These features can implicitly indicate text sentiment. Apart from these features, Part-of-Speech (POS) can be used as an additional feature to improve the performance of sentiment analysis. POS indicates how a word is used in a sentence based on grammar rules. Agarwal et al. shows that using term frequency in combination with POS can result in a better performance than using term frequency alone [2]. Moreover, identified opinion words or phrases are very useful for identifying sentiment. In addition, negation can invert the sentiment polarity of words. After all of the relevant features are obtained, they can be used to train a model such as Naïve Bayes and Support Vector Machine (SVM).

TABLE I
THE HOURGLASS OF EMOTIONS.

| | Pleasantness | Attention | Sensitivity | Aptitude |
|---|---|---|---|---|
| 3 | ecstasy | vigilance | rage | admiration |
| 2 | joy | anticipation | anger | trust |
| 1 | serenity | interest | annoyance | acceptance |
| 0 | - | - | - | - |
| -1 | pensiveness | distraction | apprehension | boredom |
| -2 | sadness | surprise | fear | disgust |
| -3 | grief | amazement | terror | loathing |

Although using a term frequency and POS as a set of features to perform sentiment analysis can identify text sentiment to a degree, this set of feature still cannot extract information related to emotion or polarity of words. This kind of information can be represented by sentic vector [3]. This vector is based on the Hourglass of Emotion [4] that is directly related to activities in the human brain. It has four dimensions composed of sensitivity ($Snst$), aptitude ($Aptit$), attention ($Attnt$), and pleasantness ($Plsnt$). Combinations of these four dimensions can represent different types of emotion as shown in Table I. Moreover, these four elements can be used to calculate the polarity of a word ($p$) that ranges from $-1$ (extremely negative emotion) to $+1$ (extremely positive emotions) as shown in Equation 1 where $N$ is the total number of concept and $c_i$ represents the $i$-th input concept.

$$p = \sum_{i=1}^{N} \frac{Plsnt(c_i) + |Attnt(c_i)| - |Snst(c_i)| + Aptit(c_i)}{3N} \tag{1}$$

978-1-7281-0164-4/18/$31.00 ©2018 IEEE

Recently, deep learning techniques have become popular especially in the field of computer vision. It can perform many tasks outstandingly such as image classification and object detection. A Convolutional Neural Network (CNN)–one of deep learning techniques–has achieved a very good result with only 15.30 % error on one of the most challenging data set called "ImageNet" which contains more than ten million images, while the second can achieve 26.2 % error by using Scale-invariant Feature Transform (SIFT) and Fisher Vector [5]. Many researchers have applied deep learning techniques to real-world problems including NLP [6]–[8].

A widely used deep learning model for NLP is Long Short-Term Memory (LSTM) algorithm [9]. It is one of the Recurrent Neural Network (RNN) models that can learn sequential data, hence it can capture sequential information and structure of the text. Recently, LSTM has been extended to Bidirectional LSTM (Bi-LSTM) [10]. Conventional LSTM can learn only sequential data from left to right hand side. However, in linguistics, there might be some patterns from right to left direction too. Bi-LSTM has an ability to learn in both directions and combine the pieces of information together to give a prediction. There have been results showing that the performance of Bi-LSTM on text data sets was better than that of LSTM [11]. Apart from RNN, a conventional deep learning technique called "Convolutional Neural Network" (CNN) that is based on Feed-forward Neural Network has also been used for this task [12]. Similar to CNN merging neighbouring pixels in an image, convolution operation performs merging operations on nearby words. There have been several works that attempted to concatenated RNN and CNN together to create a model that would have an ability to learn both sequential data and relevant neighbouring data [13]–[15] .

In general, conventional deep learning model cannot learn directly from text data because it was intentionally developed for computer vision. In order to do a workaround, a pre-processor is required to transform a word into a vector. This process is called "Word Embedding". A widely used technique for word embedding is Word2Vec [16]. It utilises a neural network with a single hidden layer to train a model in order to obtain word embedding that can be used in various tasks in NLP such as text classification, sentiment analysis, and machine translation. Having been trained, the model will be able to convert a word into a vector. In order to get a good word embedding model, the model must be trained from a very large corpus that contains a lot of words. It is too time-consuming to train such a model. Therefore, transfer learning technique is used instead. It is a technique that uses previous knowledge learned from a task to solve other new tasks. This can be done by using a pre-trained model.

Most researches on sentiment analysis utilise deep learning technique to perform the task [17], [18]. There are pieces of evidence showing that combining POS with word embedding can improve sentiment prediction performance [19]. In this study, we aimed to improve the performance of sentiment analysis in Thai language. As mentioned earlier, an analysis will still be lack of emotion information if we use only word embedding and POS embedding features. To the best of our knowledge, there has been no attempt to use sentic feature in conjunction with Word2Vec and POS of Thai language.

## II. Methodology

This section explains the framework of sentiment analysis (as shown in Figure 1). It is divided into three main parts: pre-processing with Tokensation and POS-Tagging, feature extraction, and deep learning model.

### A. Pre-processing

*1) Tokenisation:* It is required to segment words in every sentence because the word must transform into the vector. However, in Thai language, adjacent words are not separated by a space which is a kind of punctuation that is used to segment words in a sentence like in English language. Therefore, an algorithm is needed to segment words in sentences. In this work, words in every sentence have been segmented by "KUCUT" software that utilises an unsupervised algorithm to perform the task [20].

*2) POS-Tagging:* Since this experiment used POS embedding, we must identify the type of words in every sentence, but the types of all of the words in every sentence in the data set that we used in this experiment have already been identified by a Thai part of speech tagging tool called Jitar [21] which uses an algorithm based on a Tri-gram Hidden Markov Model to classify the types of word. Jitar classified POS types into 49 types.

### B. Features

*1) Word Embedding Feature:* To transform each word in the sentence into the vector, Thai2Vec–one of word embedding techniques–is used. Thai2Vec is a pre-trained word embedding that is trained with Thai-Wikipedia data by ULMFit method [22]. Thai2Vec contains 60,000 words in the corpus. Each word is represented by a 300-dimensional vector.

*2) POS Embedding Feature:* POS embedding feature is similar to word embedding features. Instead of using word represented in sentences in the corpus, types of POS are used as words in the corpus. This allows the model to understand the structure of the sentence in POS aspect.

*3) Sentic Feature:* It is used to represent emotions in a vector form which is inspired by Hourglass of Emotions. It is based on SenticNet2 [23]. SenticNet2 is only available in English, hence bidirectional translation between Thai and English is employed in order to convert SenticNet2 into Thai [24]. SenticNet2 consists of 14,244 English terms and was bidirectional translated into 16,584 Thai terms[1]. Sentic feature is a five-dimension vector. The first four elements are pleasantness, attention, sensitivity, and aptitude values, respectively. The last element represents a polarity value. All the values range between $-1$ and 1.

[1]Thai-SenticNet2 is available to download at author's website.

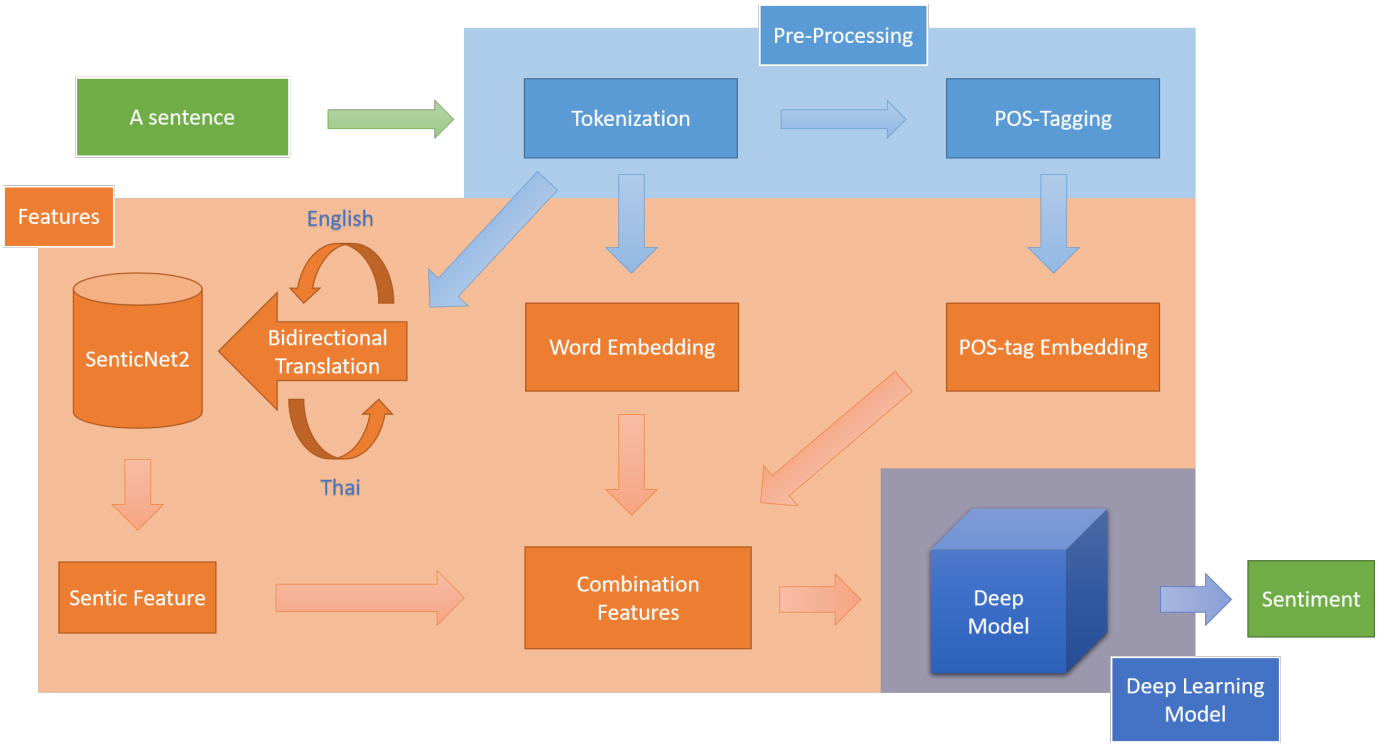


Fig. 1. Framework of sentiment analysis.

### C. Deep Learning Model

In this work, Concatenated Bidirectional Long Short-term Memory (Bi-LSTM) with Convolutional Neural Network (CNN) model, here refers as Bi-LSTM-CNN, is used to classify sentiment into three class, i.e. Negative, Neutral, and Positive. The idea is to learn contexts with Bi-LSTM, then capture local features with CNNs.

According to Figure 2, the input is fed into the Bi-LSTM model as a sequential feature vector–sequence of words in a sentence. Bi-LSTM allows the model to learn sequential data from both left to right and right to left directions. This is done by concatenating both output from bidirectional models together. This output will be fed into convolutional layer–that can extract features. This layer performs merging operations of nearby word. The information will be fed into a fully connected layer, therefore, the dimension of each batch should be fixed. It should be noted that dimension in each batch might not be the same because each sentence might have different length–different number of word in a sentence. Therefore, we processed the output from convolutional layer by dynamic 1-D max pooling operation that is for temporal data. This can be done by setting the 1-D max pooling filter size to the number of word in each sentence. Hence, the size of the feature is reduced to the number of filter in the convolutional layer that is 100. Finally, the output will be fed to the last layer that contains Softmax activation function to give a probability of each class.

## III. Experiment Settings

In this experiment, we compared the performances of Bi-LSTM-CNN model with individual word embedding, POS embedding, or sentic feature as well as Bi-LSTM-CNN with several combinations of pairs of these features and Bi-LSTM-CNN with all of these features. In the part of deep model training, we tuned the parameters of the model by using Adam Optimizer [25], set the learning rate to 0.001, and set the batch size to 16. Here, the dimension of POS embedding feature vector is set to 50.

The original data set that we considered using in this experiment was from 40 Thai children tales consisting of 1,964 sentences [26], [27]. Each sentence was already labeled with one sentiment out of the total of three sentiments–negative, neutral, and positive–by three experts. Only the sentences that had been labeled similarly by all three experts were used to train the model and in performance evaluation. To conclude, the real data set that we used in this experiment had 1,115 sentences consisting of 298 that were labeled negative, 508 that were labeled neutral, and 309 that were labeled positive.

In this experiment, we randomly split the data 10 times into training, validation, and test sets with the ratio of 80:10:10, respectively. We obtain the optimal parameter, i.e. number of epoch, based on minimum loss of validation data. Once we obtain the optimal model, it is evaluated by test data. Here, we evaluate the performance by F1-score.

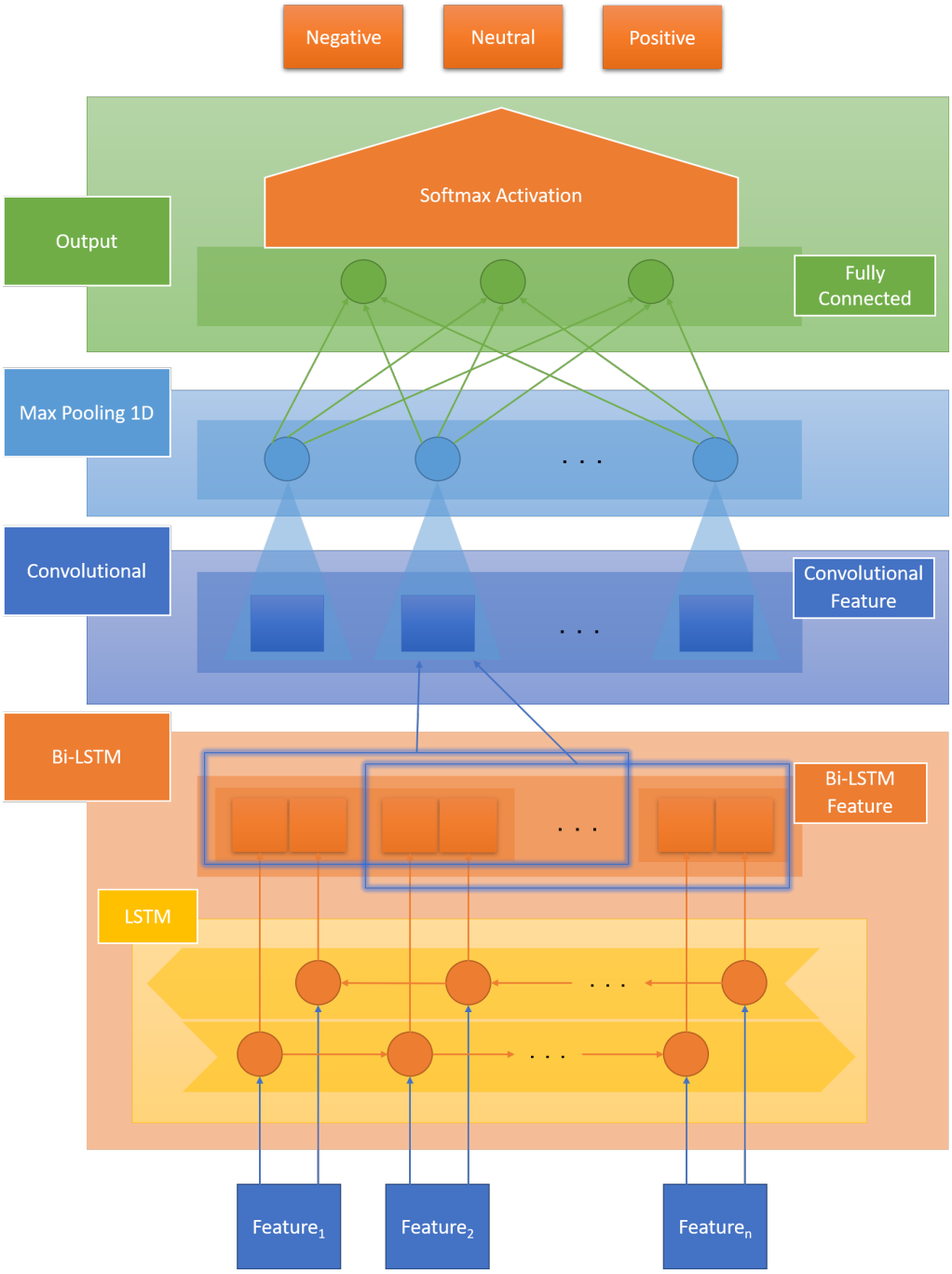


Fig. 2. Bidirectional LSTM-CNN Model.

## IV. Result and Discussion

Performance of using individual features and combining features are presented in Table II. It can be observed that according to the F1-score each combination the deep model with an individual feature achieved, the feature individual feature that yielded the best performance (75.75 %) was word embedding feature. The second best individual feature was the sentic feature (69.26 %), and the least performance came from the individual POS embedding feature (48.88 %). The reason that word embedding feature gave the best prediction accuracy was because word embedding was the result of an examination of a large corpus and the pre-trained dimensions of this feature was much larger than the dimensions of the other two features, i.e., the vector of a word processed by word embedding represented the word more thoroughly than the vectors of the word processed by the other two features did. On the same token, the vector of a word processed by POS embedding which only represented one of the 49 grammatical

TABLE II
PERFORMANCE OF BI-LSTM-CNN ALGORITHM ON VARIOUS TYPES OF FEATURES.

| Features | F1-score (%) |
|---|---|
| POS Embedding | 48.88 |
| Sentic | 69.26 |
| Sentic + POS Embedding | 74.04 |
| Word Embedding | 75.75 |
| Word Embedding + POS Embedding | 78.31 |
| Word Embedding + Sentic | 77.47 |
| Word Embedding + POS Embedding + Sentic | **78.89** |

types of the word, so it was not able to capture the sentiment as well as the other two features.

Considering POS embedding features can improve overall performances, i.e. combining word embedding with POS embedding features increases F1-score from 75.75 % to 78.31 % while taking POS embedding feature into account with sentic features can improve F1-score from 69.26 % to 74.04 %. This indicates that considering word type in the model can enhance overall performance. Similarly, combining word embedding with sentic features can improve the overall performance to 77.41 %. It means that considering emotion factor can assist the model to better understand the sentiment of each sentence. Furthermore, combining all features yields the best F1-score with 81.25 %.

## V. CONCLUSION

This research presents an approach to perform sentiment analysis in Thai. Three different features are evaluated with Bi-LSTM-CNN model on 40 Thai Children stories. The results show that combining all features–word embedding, POS embedding, and sentic features–yield the best results. This indicates that POS can enhance the model to understand semantic and role of word better and considering emotion of word can help the model to understand emotions of text.

## ACKNOWLEDGEMENT

This research is supported by the Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang.

## REFERENCES

[1] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014.

[2] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in *Proceedings of the Workshop on Languages in Social Media*. Association for Computational Linguistics, 2011, pp. 30–38.

[3] S. Poria, E. Cambria, G. Winterstein, and G.-B. Huang, "Sentic patterns: Dependency-based rules for concept-level sentiment analysis," *Knowledge-Based Systems*, vol. 69, pp. 45–63, 2014.

[4] E. Cambria, A. Livingstone, and A. Hussain, "The hourglass of emotions," in *Cognitive Behavioural Systems*. Springer, 2012, pp. 144–157.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th Advances in Neural Information Processing Systems (NIPS'2012)*, 2012, pp. 1097–1105.

[6] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Proceedings of the 29th AAAI International Conference on Artificial Intelligence (AI'2015)*, vol. 333, 2015, pp. 2267–2273.

[7] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for twitter sentiment classification," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL'2014)*, vol. 1, 2014, pp. 1555–1565.

[8] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[9] X. Wang, Y. Liu, S. Chengjie, B. Wang, and X. Wang, "Predicting polarities of tweets by composing word embeddings with long short-term memory," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, vol. 1, 2015, pp. 1343–1353.

[10] B. Plank, A. Søgaard, and Y. Goldberg, "Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss," *arXiv preprint arXiv:1604.05529*, 2016.

[11] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu, "Text classification improved by integrating bidirectional lstm with two-dimensional max pooling," *arXiv preprint arXiv:1611.06639*, 2016.

[12] A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2015, pp. 959–962.

[13] S. Lin, H. Xie, L.-C. Yu, and K. R. Lai, "SentiNLP at IJCNLP-2017 Task 4: Customer feedback analysis using a Bi-LSTM-CNN model," *Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP'2017)*, pp. 149–154, 2017.

[14] Z. Xie, Z. Zeng, G. Zhou, and T. He, "Knowledge base question answering based on deep learning models," in *Proceedings of the International Conference on Computer Processing of Oriental Languages (ICCPOL'2016) and National CCF Conference on Natural Language Processing and Chinese Computing (NLPCC'2016)*. Springer, 2016, pp. 300–311.

[15] L. Luo, Z. Yang, H. Lin, and J. Wang, "DUTIR at the BioCreative VI Precision Medicine Track: Document triage for identifying ppis affected by genetic mutations," in *Proceedings of the BioCreative VI Challenge Evaluation Workshop*, 2017, pp. 119–122.

[16] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the International Conference on Empirical Methods in Natural Language Processing (EMNLP'2014)*, 2014, pp. 1532–1543.

[17] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov, "Semeval-2016 task 4: Sentiment analysis in twitter," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 1–18.

[18] S. Rosenthal, N. Farra, and P. Nakov, "Semeval-2017 task 4: Sentiment analysis in twitter," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 2017, pp. 502–518.

[19] S. M. Rezaeinia, A. Ghodsi, and R. Rahmani, "Improving the accuracy of pre-trained word embeddings for sentiment analysis," *arXiv preprint arXiv:1711.08609*, 2017.

[20] S. Sudprasert, "KUCut Thai word segmentor (online)," 2004, https://bitbucket.org/veer66/kucut/. Accessed 21 August 2018.

[21] de Kok D, "Jitar HMM part of speech tagger (online)," 2014, https://github.com/danieldk/jitar/. Accessed 21 August 2018.

[22] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL'2018)*, vol. 1, 2018, pp. 328–339.

[23] E. Cambria, C. Havasi, and A. Hussain, "Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis," in *Proceedings of the 25th International FLAIRS Conference (FLAIRS'2012)*, 2012, pp. 202–207.

[24] R. Lertsuksakda, P. Netisopakul, and K. Pasupa, "Thai sentiment terms construction using the hourglass of emotions," in *Proceeding of the 6th International Conference on Knowledge and Smart Technology (KST'2014)*, 2014, pp. 46–50.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[26] K. Pasupa, P. Netisopakul, and R. Lertsuksakda, "Sentiment analysis of Thai children stories," *Artificial Life and Robotics*, vol. 21, no. 3, pp. 357–364, 2016.

[27] P. Netisopakul, K. Pasupa, and R. Lertsuksakda, "Hypothesis testing based on observation from Thai sentiment classification," *Artificial Life and Robotics*, vol. 22, no. 2, pp. 184–190, 2017. [Online]. Available: http://rdcu.be/nu1P