A novel meta-embedding technique for drug reviews sentiment analysis

Aye Hninn Khine, Wiphada Wettayaprasit, Jarunee Duangsuwan
Artificial Intelligence Research Lab, Division of Computational Science, Faculty of Science, Prince of Songkla University,
Hat Yai, Thailand

Article Info

ABSTRACT

Traditional word embedding models have been used in the feature extraction process of deep learning models for sentiment analysis. However, these models ignore the sentiment properties of words while maintaining the contextual relationships and have inadequate representation for domain-specific words. This paper proposes a method to develop a meta embedding model by exploiting domain sentiment polarity and adverse drug reaction (ADR) features to render word embedding models more suitable for medical sentiment analysis. The proposed lexicon is developed from the medical blogs corpus. The polarity scores of the existing lexicons are adjusted to assign new polarity score to each word. The neural network model utilizes sentiment lexicons and ADR in learning refined word embedding. The refined embedding obtained from the proposed approach is concatenated with original word vectors, lexicon vectors, and ADR feature to form a meta-embedding model which maintains both contextual and sentimental properties. The final meta-embedding acts as a feature extractor to assess the effectiveness of the model in drug reviews sentiment analysis. The experiments are conducted on global vectors (GloVE) and skip-gram word2vector (Word2Vec) models. The empirical results demonstrate the proposed meta-embedding model outperforms traditional word embedding in different performance measures.

Corresponding Author:

Jarunee Duangsuwan
Division of Computational Science, Faculty of Science, Prince of Songkla University
Hat Yai, Thailand
Email: jarunee.d@psu.ac.th

1. INTRODUCTION

Although sentiment analysis has been extensively studied and largely explored in other domains, a limited number of works have been done in the medical area. This is due to the lack of publicly available domain-specific resources and datasets [1]. The significant difference in drug reviews from other domains is the implicit sentimental words. The patients may write their symptoms and effects caused by a particular medication or treatment. For instance, consider this review: "I recently started taking Provera last week. I am with mood swings frequently." Here, "mood swings" as such do not bring any negative meaning in general but in the context of drug reviews, it represents the adverse drug reaction (ADR) which is strongly related to implicitly defined negative sentiment [2]. In such cases, domain knowledge is critical in solving sentiment analysis problems in the medical domain. Previous research works proved that domain-specific resources paved the way to boost the performance of sentiment analysis in drug reviews. Most of the studies [3]–[7] focused on the web data such as blogs, social media, and forums.

Deep learning models have reported promising results on natural language processing (NLP) tasks including sentiment analysis and the popularity of pre-trained word embedding models in the feature extraction process. The use of word embedding in the deep learning models for medical sentiment analysis has been reported recently in [2]–[6]. Word embedding refers to the dense vector representation of words inspired by the distributional hypothesis [7] "The words that occur in the same context have similar meanings". Therefore, word embedding models excel at predicting words that appear in their context. Mikolov *et al.* [8] proposed two word2vec learning approaches: continuous bag-of-word (CBOW) and Skip-gram. CBOW predicted the target word from its given context words while Skip-gram predicted the context words from a given target word. Another kind of word embedding is known as global vectors (GloVE) which is as famous as word2vec was introduced by Pennington *et al.* [9] to fulfill the lack of global statistics in the Skip-gram word2vec model. Later, the researchers found that traditional word embedding such as word2vec and GloVe led to poor representation for sentiment analysis. Sentimentally dissimilar words such as "happy" and "sad" have similar vector representations since traditional word embedding models consider only contextual properties. Therefore, sentiment knowledge integrated embedding models have been proposed recently to tackle this issue [10]–[15].

This study proposes a method to generate an improved meta-embedding model from sentiment and medical information. The main research question of the study is whether the meta-embedding outperforms traditional word embedding for domain-specific sentiment analysis. Traditional word vectors are refined to attain sentiment and medical knowledge by exploiting the polarity scores from general and domain-specific lexicons and ADR features. The proposed method is motivated by the previous research [15] working on refining pre-trained word embeddings based on the polarity values provided by a generic sentiment lexicon. In this work, we consider the importance of domain knowledge in a lexicon since previous studies [16]–[19], have confirmed that a domain-specific lexicon improves the performance of sentiment analysis. The proposed lexicon in this study is generated from the drug reviews corpus collected from medical blogs. Instead of relying on a single sentiment lexicon, the corresponding polarity scores from the four generic lexicons and one medical domain-specific lexicon are assigned to each word. The final polarity score is aggregated based on the assigned polarity scores. The common practice is that one lexicon is utilized without modifications across various domains ignoring the domain sensitivity of the scores. Therefore, the same lexicon may achieve well on a single domain while performing poorly on another domain unless some value adjustments are made. Therefore, the proposed word embedding refinement method considers both domain sensitivity and sentiment knowledge in word embedding refinement. Since a previous work [20] shows that meta-embedding is superior to individual embedding sets, refined word embedding is concatenated with original embedding to maintain the contextual properties of words. Meta-embedding is an ensemble of individual word embedding models to produce a single word embedding. Yin and Schütze [20] introduced three ensemble methods for meta-embedding (Concatenation, singular value decomposition, and 1ToN). We apply the concatenation approach in this study. The experiments were conducted to evaluate the meta-word embedding model obtained from the proposed method against conventional word embedding on the drug review sentiment analysis using a convolutional neural network (CNN) as the classifier. The empirical results exhibit that the proposed model provides higher performance to detect the users' sentiments in drug reviews.

The remaining sections of the paper are written. The detailed architecture of the proposed system is presented in section 2 followed by the comparative results demonstrated in section 3. The paper concludes in section 4.

2. METHOD

The architecture of the proposed meta-embedding learning for sentiment analysis is presented in Figure 1. There are three main steps in the proposed methodology. Step 1 is domain-specific lexicon generation. Step 2 is generating Meta embedding matrix. The final step is drug reviews sentiment classification using CNN.

2.1. Domain-specific lexicon generation

Firstly, the drug reviews are gathered from three medication-related blogs [21]–[23]. The total number of drug reviews in the corpus is 221,507. The drug review contains a mix of upper-case and lower-case letters and punctuations which influence the capability of the model to learn. The preprocessing tasks before assigning sentiment values involve lowercase conversion, punctuation removal, and work tokenization using the natural language toolkit (NLTK) tokenizer. The total number of words after preprocessing is 62,536. After preprocessing, the polarity value of each word is assigned based on the information from the five sentiment sources. The lexicons used in this study are sentiwordnet (SWN) [1], AFINN [24], multi-perspective question answering (MPQA) [25], affective norms for English words (ANEW) [26], and SocialSent: Drug [19] lexicons.
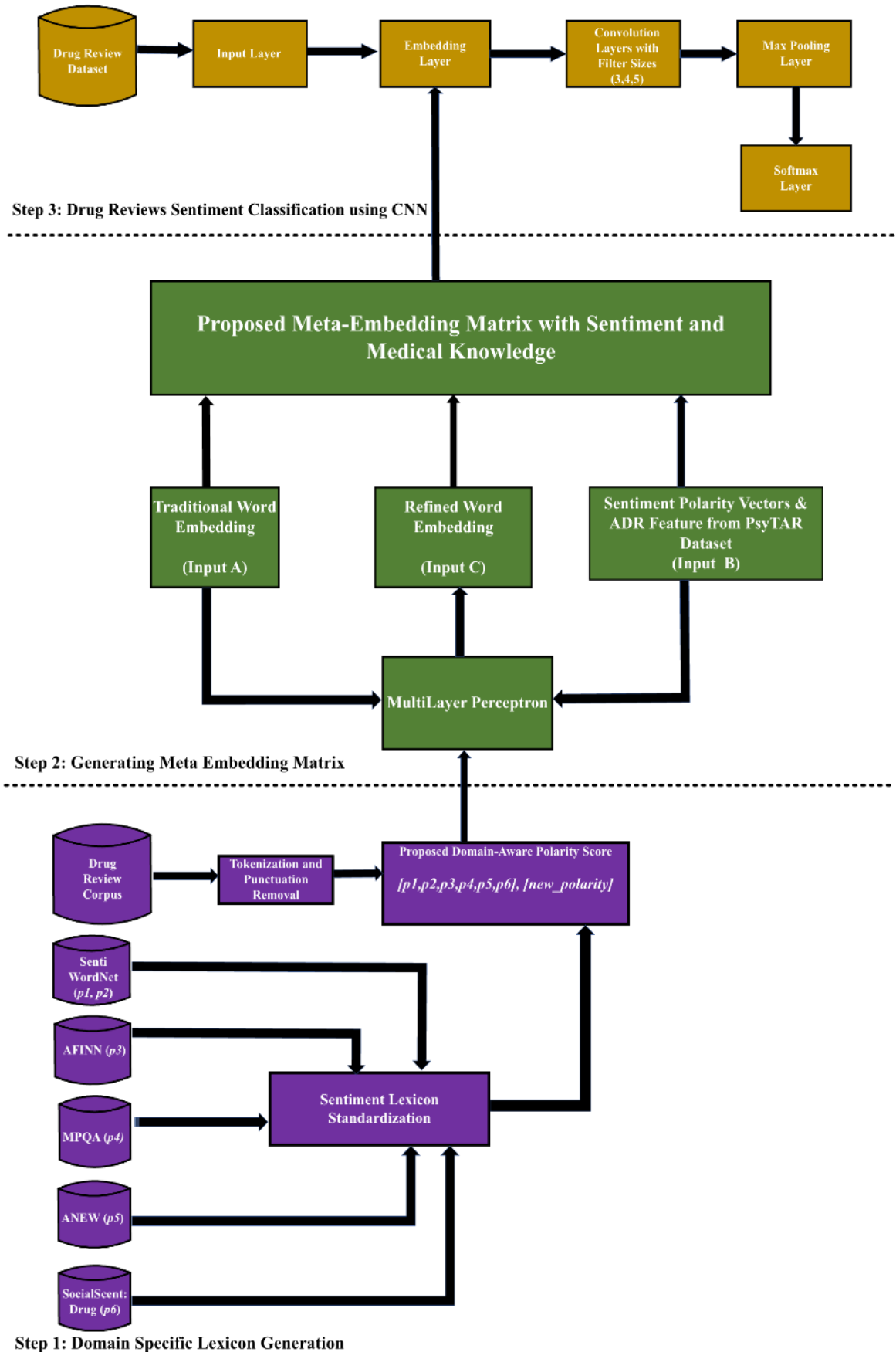
Figure 1. Proposed methodology for drug review sentiment classification with meta-embedding

2.1.1. Sentiment lexicons standardization

From step 1 of Figure 1, the sentiment lexicons have various formats. Some lexicons specify polarity scores with different numerical ranges while others describe sentiment categories such as positive, negative, and neutral. The different format makes it difficult to compare among the sentiment lexicons. So, each lexicon is standardized to make each word have common symmetric values. In this study, the sentiment lexicons are standardized to a real-valued score ranging from -1.0 and 1.0. Each sentiment lexicon is transformed,

− Sentiwordnet (SWN) – SWN is derived from the WordNet that associates each synset with three sentiment scores: positivity, negativity, and objectivity. The negative score was subtracted from the positive score to provide a real-value polarity score (p_1) within a range of {-1, 1}. The average sentiwordnet score (p_2) of each word is also considered in this study.

− AFINN – AFINN is a collection of words and phrases manually rated for the degree with an integer between -5 (the most negative) and +5 (the most positive). These values (p_3) were normalized from {-5, 5} to {-1, 1} using the Min-Max transformation method.

− MPQA – MPQA also known as subjectivity lexicon is a list of subjective words with five categories denoting strongly positive, weakly positive, neutral, weakly negative, and strongly negative. These categories (p_4) were converted to 1.0, 0.5, 0, -0.5, and -1, respectively.

− ANEW – This lexicon is derived from extended affective norms of English words (E-ANEW) and includes 13,915 words with three real-valued scores for valence, arousal, and dominance. Among them, the valence represents the intensity of positive and negative sentiments in the range of 1 to 9. The valence value of 1 indicates the most negative and 9 is the most positive sentiments. Therefore, valence is considered as the value of polarity. And these values (p_5) were normalized from {1, 9} to {-1, 1}.

− Socialscent: Drug-It is a domain-specific lexicon generated by the SentProp algorithm from the subreddit communities. The drug lexicon contains 4,908 words that associate a real-valued sentiment score between -5 to +5. The scores (p_6) were normalized from {-5.5} to {-1, 1}.

2.1.2. New domain polarity value

After the polarity transformation of sentiment lexicons, each tokenized word from the drug review corpus is looked up in each sentiment lexicon. The polarity values of each word matching more than one sentiment lexicon are averaged to yield the new polarity value using the (1). The new polarity values range from -1 to 1.

$$New\_Polarity = \frac{1}{n}\sum_{i=1}^{6}(p_i)$$ (1)

Where n is the number of sentiment lexicons where the word is found. Finally, we have achieved a domain-specific lexicon containing 62,536 words. For example, the word "miserable" has polarity scores as follows: [$p_1, p_2, p_3, p_4, p_5, p_6$] in Step 1 of the Figure 1 refers to [SWN, Average SWN, AFINN, MPQA, ANEW, SocialScent: Drug]. Therefore, the polarity representations of the word "miserable" are [-0.875, -0.688, -0.6, -0.79, -0.631, -0.804] and the [*new_polarity*] value is the mean of those values which is [-0.776]. This can avoid the influence of a single lexicon in sentiment analysis. The new polarity value is used as the ground truth value in refining word embedding which will be discussed in the next section.

2.2. Generating meta embedding matrix

For the word embedding refinement process, we propose a multilayer-perceptron (MLP) model to predict the polarity score of each word from the original word embedding. However, all words are not covered in our proposed lexicon, but are comprised in the vocabulary set of input word embedding, are defined as neutral, and set the polarity of 0. For each pre-trained word vector, lexicon vector obtained from sentiment lexicons, and ADR features are given to the input layer.

As shown in step 2 of Figure 1, each word vector [v_1, vn] is extracted from the original embedding where n is the size of the original word embedding. Secondly, the polarity scores of each word are extracted from the sentiment lexicons, and the lexicon vector is formed by concatenating the extracted scores. ADR is important information to detect the users' sentiments in the medical setting. ADR phrases from the psychiatric treatment adverse reaction (PsyTAR) dataset [27] contain adverse drug events related to patients' experiences with psychiatric medications. To obtain a binary ADR value (*adr*) of each word, PsyTAR is used to check whether a given word is ADR or not. If it is included in ADR phrases, 1 is assigned to *adr* value otherwise 0. For example, the word "miserable" is contained in the ADR phrases so that we assign 1 to its *adr* value. Therefore, the vector representation of "miserable" to the input layer is [$v_1, …, v_n, p_1, p_2, p_3, p_4, p_5, p_6, adr$] which is equal to [$v_1, . . . , v_n$, -0.875, -0.688, -0.6, -0.79, -0.631, -0.804, 1]. According to the size of refined word embedding, the model creates a *k-unit* hidden layer (k is defined as 100 in this study) with a tanh activation function described in (2).

$$h_v^\theta = \tanh(W_h^T \cdot v + b_h) \tag{2}$$

In (2), $W_h \in R^{d \times k}$ and $b_h \in R^k$ are the model parameters where d is the dimension of original input embedding, and k is that of refined word embedding. The sigmoid function is utilized in the output layer to calculate the score corresponding to the polarity value of the given vector. The model calculates the scores described in (3).

$$f_v^\theta = \sigma(W_o^T \cdot h_v^\theta + b_o) \tag{3}$$

W_o and b_o are the model parameters of the sigmoid output layer having dimension 1. Adam optimization is applied in the model training and mean absolute error (MAE) loss function with L2 regularization is used for calculating the difference between the ground truth polarity value and the predicted value. The maximum number of epochs is 300. After each epoch, we determine whether the stopping criterion is met, and the learning process is terminated if the MAE value is less than 0.05.

$$J(\theta) = \frac{1}{n}\sum_{k=1}^{n}\left|f_{v_k}^\theta - f_{v_k}^g\right| + \lambda||\theta||_2^2 \tag{4}$$

In (4), while $f_{v_k}^\theta$ indicates the predicted polarity value for input vector v_k, $f_{v_k}^g$ is the ground truth polarity according to our proposed new polarity. A k-dimensional matrix $[u_1,..., uk]$ is taken out from the output of the hidden layer h_v^θ after training when the input vector is given. The extracted refined word embedding is linearly concatenated with original word vectors, polarity vectors, and an *adr* value to form a final meta-embedding matrix which will be used in the embedding layer of the sentiment classification model.

$$Meta\text{-}embedding = [v_1,.....,v_n] \oplus [u_1,...,u_k] \oplus [p_1,p_2,...,p_6] \oplus [adr] \tag{5}$$

2.3. Drug review sentiment classification using CNN

CNN has been reported as one of the widely used classifiers in sentiment classification and recent works focused on sentiment analysis of drug reviews also applied CNN for classification purposes. In this model, we adopted the architecture described in [2]. The input layer accepts each drug review as the input to the classification model. Given a drug review consisting of n words $w_1, w_2,...,w_n$, each word w is transformed into a real-value vector x_i using meta-embedding matrix $W \in R^{kx|V|}$ where $|V|$ represents the fixed-length vocabulary, and k refers to the dimension of the word embedding model. The drug review representation matrix $x_{1:n_w}$ is defined as,

$$x_{1:n_w} = x_1 \oplus x_2 \oplus .... \oplus x_{n_w} \tag{6}$$

Then, the word embedding is fed as the input to the convolution layer where the convolution filter $F \in R^{mxk}$ (m refers to the size of the filter) is applied to each context window.

$$c_i = f(F * x_{i:i+m-1} + b) \tag{7}$$

Where f is a rectified linear unit (RELU) activation function, and b is a bias term. In the proposed model, filter sizes of (3), (4), and (5) are applied to increase the scope of the n-gram model. The max-pooling strategy is to minimize the size of the representation by identifying the most significant features generated by the convolution layer. These pooled features are fed into the fully connected softmax layer which output is the probability distribution over three sentiment classes. During the training phase, the Adam algorithm is used as the optimization method defining a learning rate of 0.001. The categorical cross-entropy loss is employed to calculate the error of the model.

3.   RESULTS AND DISCUSSION

− Baseline Traditional Word Embedding Models: In this section, we evaluate our proposed meta-embedding in multi-class drug reviews sentiment classification. The proposed model is evaluated by GloVe model [9] pre-trained on the Twitter corpus (GloVe-Twitter) and two traditional word2vec models trained on the Skip-gram method: Word2Vec-Drug-Reviews and Word2Vec-Google-News [8]. GloVE-twitter model has 200 dimensions with a vocabulary of 1.2 million. Drug-Review word2vec is trained on the drug review corpus described in Section 2. This model contains a vocabulary of 60,000 and the dimension is 300.

Google-News word2vec is trained on GoogleNews corpus with 300 dimensions and the model contains a vocabulary of 100 billion.

− Dataset: The dataset proposed in [28] is used for the extrinsic evaluation of the proposed model. Each drug review presents a rating from 0 to 9 which reflects the patient's satisfaction level with the drug. The corpus contains a total of 215,603 drug reviews and is split into training and testing datasets in the ratio of 75:25. The users' ratings are converted to three-sentiment classes according to the original creators: positive, neutral, and negative. A drug review that has a rate less than or equal to 4 is labeled as negative, rating 5 and 6 is defined as neutral, and above 7 is labeled as positive. The sentiment class distribution is imbalanced by presenting the majority as the positive class.

− Evaluation Metrics: To assess the proposed Meta embedding against traditional word embedding models, different evaluation metrics are presented. We exhibit evaluation metrics of sentiment analysis, which are widely applied in previous literature. Evaluation metrics are accuracy, precision, recall, and F1-measure. Precision is referred to measure the exactness, and recall is used to measure the completeness of the model. F1-measure is the mean of precision and recall.

Figures 2 to 4 report the results of the drug review benchmark using GloVE-Twitter, Word2Vec-Drug-Reviews, and Word2Vec-Google-News as baseline traditional word embedding models respectively. To evaluate the proposed model against the baseline method on the extrinsic sentiment classification task, we define the inputs to the embedding layer of the classification model as described in Table 1. Input A represents the traditional word embedding, B represents the lexicon and ADR features, and C is the refined word embedding achieved from the proposed refinement approach. Model 1(A) is the baseline method where only the traditional word embedding model is used in the embedding layer. In model 2(A+B), traditional word vectors are concatenated with the lexicon and ADR features. The use of lexicon and ADR features is excluded in model 3(A+C) where traditional and refined word vectors are concatenated. In model 4(A+B+C) which is our proposed method where traditional word vectors, refined word vectors, lexicon, and ADR features are concatenated to form an improved meta-embedding matrix used in the embedding layer of the CNN model for feature extraction.
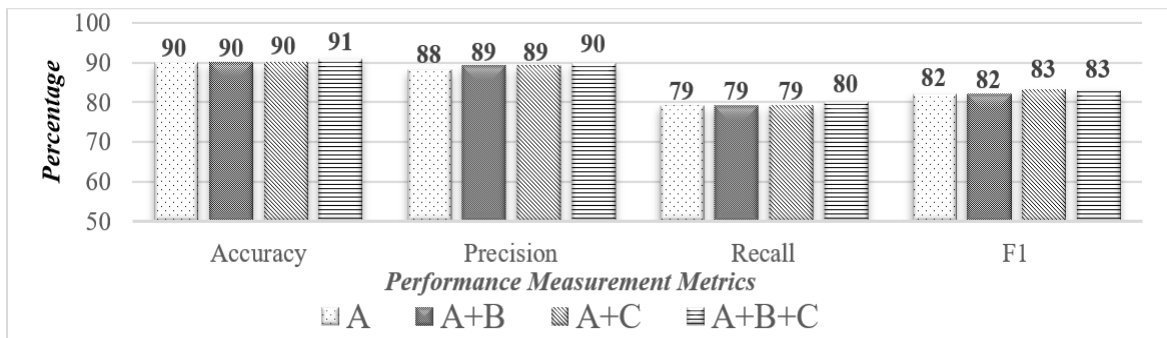

Figure 2. Performance of CNN sentiment classification model over drug review benchmark using GloVe embedding and improved meta-embedding (GloVe-Twitter)
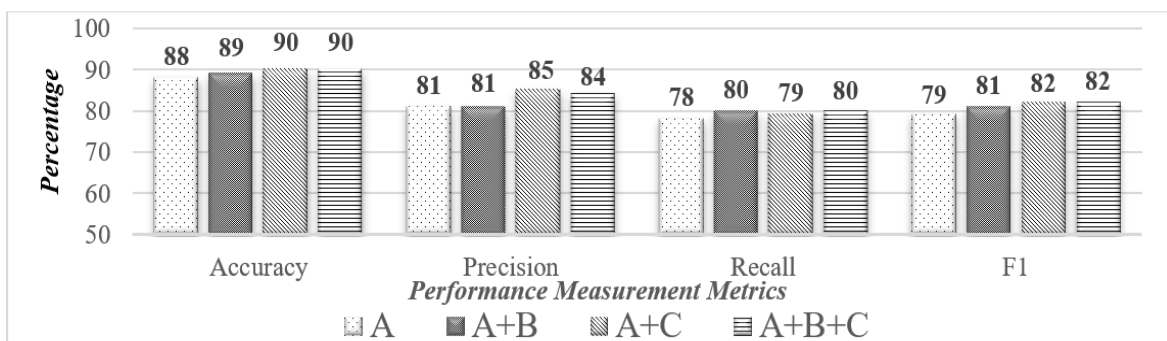

Figure 3. Performance of CNN sentiment classification model over drug review benchmark using word2vec-embedding and improved meta-embedding (Word2Vec-Drug-Reviews)
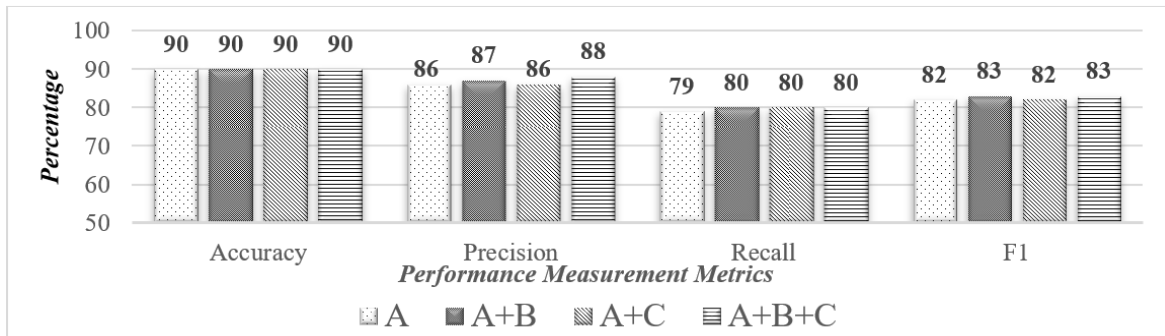
Figure 4. Performance of CNN sentiment classification model over drug review benchmark using word2vec-embedding and improved meta-embedding (Word2Vec-Google-News)

Table 1. Summary of CNN drug review sentiment classification models

Model	Inputs	Description
1	A	Only traditional word2vec is fed to the embedding layer $[v_1,.....,v_n]$
2	A+B	Traditional word2vec is concatenated with the lexicon and ADR features and the resulting vectors are fed to the embedding layer. $[v_1,.....,v_n]\oplus [p_1,p_2,...,p_6]]\oplus[adr]$
3	A+C	Traditional and refined word vectors are concatenated, and the resulting vectors are fed to the embedding layer $[v_1,.....,v_n]\oplus[u_1,...,u_k]$
4 (Proposed model)	A+B+C	Traditional word vectors, refined word vectors, lexicon, and ADR features are concatenated, and the resulting meta-embedding matrix is fed to the embedding layer. $[v_1,.....,v_n]\oplus[u_1,...,u_k]\oplus [p_1,p_2,...,p_6]\oplus[adr]$

According to Figure 2 of GloVe-Twitter, model 4(A+B+C) gives the highest accuracy, precision, recall, and F1-measure denoting 91%, 90%, 80%, and 83%, respectively. Meanwhile, model 1(A) presents 90%, 88%, 79%, and 82%, respectively. In Figure 3 of Word2vec-Drug-Reviews, model 4(A+B+C) exhibits the highest accuracy, recall, and F1-measure representing 90%, 80%, and 82%, respectively. But model 3(A+B) obtains the highest precision at 85%. Model 1(A) indicates 88% accuracy, 81% precision, 78% recall, and 79% F1-measure, respectively. In Figure 4 of Word2vec-Google-News, all models report the same accuracy of 90%. However, model 4(A+B+C) achieves the highest precision, recall, and F1-measure with 88%, 80%, and 83% while model 1(A) obtains lower performance with 86%, 79%, and 82%, respectively.

Therefore, the word meta-embedding presents a better performance over conventional word embedding in the drug reviews sentiment classification. This is because the meta-embedding model makes use of polarity and domain-specific knowledge which are lacking in conventional word vectors. The proposed method increases the size of word embedding by concatenating it with refined word embedding, polarity, and ADR features. The proposed model can capture not only contextual relationships of words but also domain-specific sentiment and medical properties. The more knowledge the word embedding model has, the better performance can be achieved for the sentiment classification in CNN.

4. CONCLUSION

This paper proposed an approach for learning meta-embedding from sentiment and ADR features. The polarity score estimation model is developed using MLP to refine traditional word vectors. Domain-aware polarity score obtained from the different sources of sentiment lexicons is used as the ground truth value for refinement. The refined word embedding model is concatenated with traditional word vectors, sentiment, and ADR features to form an improved meta-embedding matrix which can be used in the embedding layer of the sentiment classification model. The experiments are conducted using a different combination of vectors to prove the effectiveness of sentiment and ADR features in our model. Empirical results show that the proposed model improves classification performance in terms of accuracy, precision, recall, and F1-measure. For future research, the representation of missing vocabularies in word embedding models and in merging multiple sentiment lexicons for enhancement of domain polarity score should be studied for improvement. The use of medical ontologies such as unified medical language system (UMLS) can be applied to improve word embedding more reliably for medical settings in NLP tasks.

ACKNOWLEDGEMENTS

REFERENCES

[1]   K. Denecke and Y. Deng, "Sentiment analysis in medical settings: New opportunities and challenges," *Artificial Intelligence in Medicine*, vol. 64, no. 1, pp. 17–27, May 2015, doi: 10.1016/j.artmed.2015.03.006.

[2]   S. Yadav, A. Ekbal, S. Saha, and P. Bhattacharyya, "Medical sentiment analysis using social media: Towards building a patient assisted system," 2018.

[3]   C. Colón-Ruiz and I. Segura-Bedmar, "Comparing deep learning architectures for sentiment analysis on drug reviews," *Journal of Biomedical Informatics*, vol. 110, p. 103539, Oct. 2020, doi: 10.1016/j.jbi.2020.103539.

[4]   S. Yadav, A. Ekbal, S. Saha, P. Bhattacharyya, and A. Sheth, "Multi-task learning framework for mining crowd intelligence towards clinical treatment," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 2018, pp. 271–277, doi: 10.18653/v1/n18-2044.

[5]   S. Yadav, A. Ekbal, S. Saha, and P. Bhattacharyya, "A unified multi-task adversarial learning framework for pharmacovigilance mining," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 5234–5245, doi: 10.18653/v1/p19-1516.

[6]   T. Huynh, Y. He, and A. Willis, "Adverse drug reaction classification with deep neural networks," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics*, Dec. 2016, pp. 877–887.

[7]   Z. S. Harris, "Distributional structure," in *Papers on Syntax*, Springer Netherlands, 1981, pp. 3–22.

[8]   T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations ofwords and phrases and their compositionality," *Advances in Neural Information Processing Systems*, 2013.

[9]   J. Pennington, R. Socher, and C. Manning, "Glove: global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543, doi: 10.3115/v1/D14-1162.

[10]   A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 2011, pp. 142–150.

[11]   L.-C. Yu, J. Wang, K. R. Lai, and X. Zhang, "Refining word embeddings using intensity scores for Ssentiment analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 671–681, Mar. 2018, doi: 10.1109/taslp.2017.2788182.

[12]   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 (Volume 1: Long Papers)*, 2014, pp. 1555–1565, doi: 10.3115/v1/p14-1146.

[13]   M. Kasri, M. Birjali, and A. Beni-Hssane, "Word2Sent: A new learning sentiment-embedding model with low dimension for sentence level sentiment classification," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 9, Dec. 2020, doi: 10.1002/cpe.6149.

[14]   D. Tang, F. Wei, B. Qin, N. Yang, T. Liu, and M. Zhou, "Sentiment embeddings with applications to sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 2, pp. 496–509, Feb. 2016, doi: 10.1109/tkde.2015.2489653.

[15]   B. Naderalvojoud and E. A. Sezer, "Sentiment aware word embeddings using refinement and senti-contextualized learning approach," *Neurocomputing*, vol. 405, pp. 149–160, Sep. 2020, doi: 10.1016/j.neucom.2020.03.094.

[16]   J.-C. Na and W. Y. M. Kyaing, "Sentiment analysis of user-generated content on drug review websites," *Journal of Information Science Theory and Practice*, vol. 3, no. 1, pp. 6–23, Mar. 2015, doi: 10.1633/jistap.2015.3.1.1.

[17]   M. Z. Asghar, S. Ahmad, M. Qasim, S. R. Zahra, and F. M. Kundi, "SentiHealth: Creating health-related sentiment lexicon using hybrid approach," *SpringerPlus*, vol. 5, no. 1, p. 1139, Jul. 2016, doi: 10.1186/s40064-016-2809-x.

[18]   A. H. Khine, W. Wettayaprasit, and J. Duangsuwan, "Refining word embeddings using domain-specific knowledge for drug reviews sentiment classification," in *2021 25th International Computer Science and Engineering Conference (ICSEC)*, Nov. 2021, pp. 104–109, doi: 10.1109/icsec53205.2021.9684639.

[19]   W. L. Hamilton, K. Clark, J. Leskovec, and D. Jurafsky, "Inducing domain-specific sentiment lexicons from unlabeled corpora," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 595–605, doi: 10.18653/v1/d16-1057.

[20]   W. Yin and H. Schütze, "Learning word meta-embeddings," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016, pp. 1351–1360, doi: 10.18653/v1/p16-1128.

[21]   "Patient Info," 2023, [Online]. Available: https://www.patients.info.

[22]   "Drug Information Portal," 2023, [Online]. Available: http://journal.um-surabaya.ac.id/index.php/JKM/article/view/2203.

[23]   "Drugs.com Know More Be Sure," 2023, [Online]. Available: http://journal.um-surabaya.ac.id/index.php/JKM/article/view/2203.

[24]   F. F. Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," Mar. 2011.

[25]   L. Deng and J. Wiebe, "MPQA 3.0: An entity/event-level sentiment corpus," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2015, pp. 1323–1328, doi: 10.3115/v1/n15-1146.

[26]   S. Shaikh *et al.*, "ANEW+: Automatic expansion and validation of affective norms of words lexicons in multiple languages," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016, pp. 1127–1132.

[27]   M. Zolnoori *et al.*, "The PsyTAR dataset: From patients generated narratives to a corpus of adverse drug events and effectiveness of psychiatric medications," *Data in Brief*, vol. 24, p. 103838, Jun. 2019, doi: 10.1016/j.dib.2019.103838.

[28]   F. Gräßer, S. Kallumadi, H. Malberg, and S. Zaunseder, "Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning," in *Proceedings of the 2018 International Conference on Digital Health*, Apr. 2018, pp. 121–125, doi: 10.1145/3194658.3194677.

BIOGRAPHIES OF AUTHORS

Aye Hninn Khine 🆔 🔍 SC ◯ is a Ph.D. candidate at the Division of Computational Science, Faculty of Science, Prince of Songkla University, Thailand. She got a Bachelor of Computer Science and Master of Computer Science from the University of Computer Studies, Yangon. Her research interest is Natural Language Processing and Machine Learning. She can be contacted at 6110230008@email. psu.ac.th.

Wiphada Wettayaprasit 🆔 🔍 SC ◯ is an assistant professor at the Division of Computational Science, Faculty of Science, Prince of Songkla University, Thailand. She got BSc., MSc., and Ph.D. in Computer Science from Prince of Songkla University, University of Missouri-Columbia, USA, and Chulalongkorn University, respectively. Her research interests are Artificial Intelligence, Neural Networks, and Machine Learning. She can be contacted at wiphada.w@psu.ac.th.

Jarunee Duangsuwan 🆔 🔍 SC ◯ is an assistant professor at the Division of Computational Science, Faculty of Science, Prince of Songkla University, Thailand. She got BSc., MSc., and Ph.D. in Computer Science from Chiang Mai University, Prince of Songkla University, and the University of Reading, UK, respectively. Her research interests are Natural Language Processing and Machine Learning. She can be contacted at jarunee.d@psu.ac.th.