ADAPTING GLOVE AND LSTM METHODS TOWARDS DETECTION OF ADVERSE DRUG REACTION

Hussein Ali Abdulhussein, Nazlia Omar

Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Malaysia.

p103920@siswa.ukm.edu.my, nazlia@ukm.edu.my

ABSTRACT

Identifying adverse drug reactions (ADRs) is one of the most critical data points for assessing the patient's perception of medicine. Most studies have focused on extracting adverse drug reactions from social media platforms where users discuss specific medications. Some research has utilized trigger terms to identify ADRs. These tests demonstrated a remarkable level of ADR extraction efficiency. However, these restrictions are insufficient because they must be perpetually updated to reflect the discovery of new adverse effects or substances with medical relevance. Based on word embeddings, this paper proposes a method for enhancing ADR detection. The study utilized a standard dataset and implemented several preprocessing techniques, such as stop-word removal, tokenization, and lemmatization. Using the recommended GloVe word embeddings, the Short-Term Long Memory (LSTM) classifier was trained. The experimental results demonstrated that GloVe outperformed the baseline method, attaining an accuracy of 99% on the dataset. This advantage accentuates the use of GloVe when embedded correspondences are precisely detected, as opposed to a predefined list of trigger terms.

Keywords: ADR, LSTM, Word Embeddings, GloVe

I. INTRODUCTION

Identifying adverse drug reactions is one of the essential responsibilities of those employed in the pharmaceutical industry. Before making medication available to the public, researchers conduct multiple clinical trials to identify potential side effects included in the medication's directions. In contrast, clinical studies cannot determine all potential adverse effects of a drug. Some of these effects only manifest after long-term medication use; some individuals not enrolled in the clinical studies have experienced these side effects. After receiving the go-ahead, a drug has the potential to pose significant threats to human health and even cause mortality as a result of its adverse effects. In addition, pharmacovigilance confronts the complex problem of identifying potential adverse effects in the post-approval phase of a drug's use (Alimova & Tutubalina 2018).

Medical reviews, a newly developed category of product reviews, have garnered increased interest from researchers. By delineating the effects of the medicines on their bodies, consumers can provide feedback on pharmaceutical products (Ebrahimi et al. 2016b). Some adverse effects and other medical conditions are discussed concerning the topic. Consequently, a task known as Adverse Drug Reaction (ADR) Detection aims to identify ADR mentions (Kumar et al. 2019).

According to the WHO, ADRs are any undesirable, unintended, and unintended effects of medication that occur at levels used for prevention, diagnosis, and treatment. These side effects can range from mild to fatal (Zhang et al., 2020). Several studies have collected data by scouring social networks such as Twitter or online drug databases to identify adverse medication reactions. In addition, researchers sifted through regular user remarks and reviews to extract ADR mentions from this data collection. For example, an adverse drug reaction (ADR) for "dizzy" can be found in research that begins with "After taking this pill, I felt dizzy," indicating that the user is reporting a side effect that occurred after taking a particular prescription (Nafea et al. 2021).

According to published research, researchers focused on machine learning techniques to identify ADRs ((Ebrahimi et al. 2016; Pain et al. 2016; Plachouras et al. 2016; Kiritchenko et al. 2018; Yousef et al. 2019; Nafea et al. 2021). These techniques train a classification model using annotated drug data from medical evaluations. Moreover, the programmers or implementers of the training stated that several characteristics might indicate ADR co-occurrences. One of these components is referred to as the Trigger Terms, and they are essentially a collection of terms typically associated with ADRs. Numerous researchers have utilized this attribute in conjunction with numerous categorization strategies. One of the problems associated with ADR extraction is the imprecise detection of ADRs, despite many complex issues. This study intends to solve the accuracy issue by offering a strategy to enhance accuracy. Specifically, this study wants to do this by seeking to improve upon accuracy.

Combining LSTM and Word embeddings is a popular technique for natural language processing (NLP) applications because it maximizes the advantages of both approaches. The LSTM recurrent neural network (RNN) type can simulate long-term relationships in sequential data, such as language. In contrast, word embeddings are dense vector representations of words that can capture the semantic relationships between words in a text corpus. Combining these two techniques can result in more precise and efficient NLP models.

II. RELATED WORK

Korkontzelos et al. (2016) Look at the ways that sentiment analysis tools may be used to distinguish between (ADRs) and mentions of an indication. The following analysis was carried out using a set of messages from the DailyStrength online community and tweets that had been annotated for mentions of adverse drug reactions (ADR) and indicators. The results show that the incorporation of sentiment analysis attributes can improve the effectiveness of a cutting-edge adverse drug reaction (ADR) recognition method. This is especially important considering the growing use of social media and online forums for people interested in health; this improvement helps pharmacovigilance practice.

ADRs, or adverse drug reactions, have a significant role in morbidity and mortality in patients. ADRs are often detected using traditional methods during clinical trials, although many unreported ADRs may still be present after the medicine is released into the market. To foresee potential ADRs, this study suggests an encoder-decoder system based on attention processes and the LSTM model. We use the masked strategy to generate the target data, and we often evaluate the effectiveness of our suggested model using the 5-fold cross-validation technique. Our approach outperforms the conventional methods in future ADR predictions (Qian et al. 2022).

CRFs are used by ADRMine, a concept extraction technique based on machine learning, to extract ADR references from extremely casual social media content. It groups words based on unsupervised, pre-trained word representation vectors (embeddings) generated from unlabeled user posts on social media using a deep-learning approach. For modelling the semantic similarity of words, this characteristic is innovative. This approach relies heavily on unlabeled data, which makes it suitable for social media mining and minimizes the need for massive, annotated training data sets (Nikfarjam et al. 2015).

Yousef et al. (2019) regulated ADR extraction from online communities where people discuss potential medications. To extract entities from texts discovered on social media, obtaining systems relied on important words and detailed trigger circumstances that may occur before or after ADRs are used. This article proposes new trigger words with various N-gram topologies, such as unigram, bigram, trigram, and quad ram. Using special terminology and trigger phrases that may come before or after Adverse Drug Reactions (ADRs), entities from social media texts are extracted. This article proposes new trigger words with various N-gram topologies, such as unigram, bigram, trigram, and quad ram. These longer trigger words were used in the studies to improve the accuracy of ADR extraction from social media communications. This study recommends training the SVM, LR, and NB classificatory models on the suggested extension. The TFIDF and TF were also used as two different document representations. Secondary data from drug websites was used to conduct the trials. Despite the possibility of ignoring the semantic component, the main problem with this study is its reliance on trigger terms.

Nafea et al. (2021) try to offer a semantic method based on latent semantic analysis (LSA) to enhance the detection of adverse drug reactions (ADRs). A benchmark dataset was employed, along with several pre-processing techniques such as stop word removal, tokenization, and stemming. Three classifiers—the Support Vector Machine (SVM), Nave Bayes (NB), and Linear Regression (LR)—were trained on the proposed LSA. Documents were described using the words "term frequency" (TF) and "term frequency-inverse document frequency" (TF-IDF). Instead of using a predetermined list of trigger words, it places more emphasis on the use of LSA in situations when semantic correspondences may be correctly identified.

It mechanically extracts adverse drug reactions (ADRs) from customer reviews made on various medicine social media platforms to uncover negative effects that are not reported to the US Food and Drug Administration (FDA) but are commended by consumers. It makes use of many lexicons, looks for trends, and compiles a list of synonyms that includes different forms of medical jargon. ADRs are divided into "expected" and "unexpected" types. Using background (drug) language, the strength of the identified unexpected ADRs is evaluated (Yates & Goharian 2013). As it shows table 2.1 show the methods of previous studies used to detect ADRs.

Table 1 Critical Review

Author	**Method**	**Features**	**Data**	**Finding**
Korkontzelos et al. (2016)	sentiment analysis features	Trigger terms	Twitter	As demonstrated by a statistically significant rise in F-measure from 72.14% to 73.22% on Twitter, this strategy can aid in reducing the number of adverse drug reactions (ADRs) discovered.
Qian et al. (2022)	Encoder-decoder framework based on attention mechanism and the long short-term memory (LSTM)	Trigger terms	Drug websites	While "Multilabel" performs better than "BIOHD" at accurately expressing both continuous and discontinuous ADR remarks, LSTM-CRF can achieve a higher score than CRF.

Nikfarjam et al. (2015)	Conditional random fields with novel feature ADRMine. clustering words by pre-trained word representation vectors	Trigger terms	Twitter	being able to obtain an F-measure of 0.82. The feature analysis's findings indicate that the suggested word cluster characteristics considerably boost the extraction's efficacy.
Masino et al. (2018)	ConvNet processes tweets as word vectors, which are generated through unsupervised learning.	Trigger terms	Twitter	a classification score of 0.46 and a sensitivity of 0.78 for tweets containing ADR.
Yousef et al. (2019)	SVM, LR, NB	Trigger terms	Twitter dataset	achieved an F-measure of 0.69
Nafea et al. (2021)	LSA, SVM, LR, NB	Trigger terms	Twitter dataset	The suggested LSA achieved 82% of the F-measure for the dataset, outperforming the baseline extended trigger terms, according to the results.
Yates & Goharian (2013)	ADRTrace automatically extracts adverse drug reactions (ADRs)	various lexicons identify patterns, generate a synonym set	Twitter dataset	For identifying anticipated and unanticipated adverse drug reactions (ADRs) from customer evaluations on multiple pharma social media sites, the program obtained good accuracy and recall rates.

III. METHODOLOGY

The research methodology consisted of five phases, as depicted in Figure 1. In the first section, we will prepare medication evaluations with annotations. (Yousef et al. 2019) modified a dataset created by Yates & Goharian (2013) that contains the necessary data. The second phase of the endeavour will consist of data preprocessing operations such as stopword removal, tokenizer, and lemmatizer. In the third phase, semantic similarity is examined using the prescribed word embedding, namely GloVe. A gloVe is a popular option for NLP tasks because it can extract word similarity efficiently and generate accurate word embeddings.

Additionally, it is relatively simple to use, making it a popular option among researchers and practitioners. GloVe's ability to analyze the co-occurrence statistics of words in a large corpus of text, thereby efficiently capturing word similarity and semantic relationships, is one of its advantages. This results in more precise word embeddings than other methods. In the fourth segment, the LSTM algorithm will categorize the data. One of the main advantages of LSTMs is their ability to address the vanishing gradient problem that may arise during training. LSTMs solve this issue by enabling more efficient gradient flow and weight updates, enhancing performance. Finding the optimal hyperparameter for the proposed ADR model is the final step.

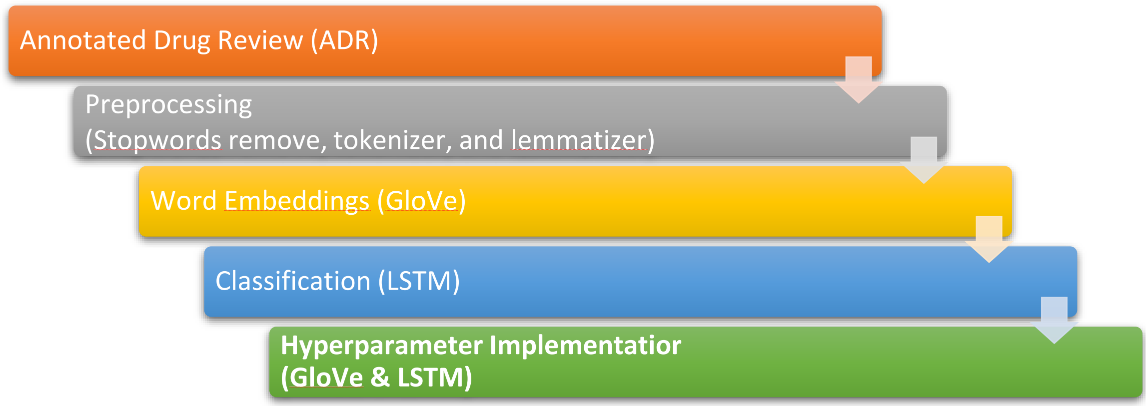

Figure 1 Research Methodology

A. Dataset Collection

The dataset used in this study was constructed utilizing the dataset benchmark established by (Yates & Goharian 2013). This study's data consists of 2,500 evaluations (along with 246 labelled documents). Each document contains at least one statement. The documents constitute 945 statements. These utterances are extracted directly from the tweets of various Twitter users. In addition, a total of 982 ADRs are accumulated across all documents. These evaluations of the paper are written solely in English. Review reviews were compiled using data from social media sites that provide medication evaluations, namely drugs.com, askpatient.com, and drugratingz.com.

B. Stop-Word Removal

This exercise aims to eliminate general terms that have no special significance. These phrases are typically omitted from preprocessing to reduce the quantity of ambient data or unhelpful characteristics. Before or after processing material from natural language, computers filter out stop words (Rajaraman & Ullman 2011). Typical examples include "the, a, an, and of," frequently appearing in the text. The word eradication mechanism filters and eliminates these words to improve the algorithm's performance.

C. Tokenization

During the tokenization procedure, the text must be divided into sentences, and the tokens comprising those sentences must be transformed into new token sequences (i.e., words).

D. Lemmatization

Lemmatization is a technique for processing natural language that converts words to their lemma or unmodified form. For instance, the lemma of the word "caring" is "care."

E. GloVe Hyperparameter

Gensim is a Python utility that is both open-source and free to use. Typically, this method represents documents as mathematical vectors that convey meaning. Consensus holds that Gensim's library was the first to implement the GloVe standard. The collection has grown and been updated over the years, making it an exhaustive resource. At the time of the investigation, Gensim version 4.0.1 was the most recent. The hyperparameters of the GloVe model will be learned and fine-tuned. The hyperparameters for GloVe are dataset size, BinaryCrossentropy, Adam Optimizer, and Epochs.

F. LSTM Hyperparameter

The essential LSTM model hyperparameters are the Activation function, number of units, and Dropout rate.

IV. RESULTS & DISCUSSION

A. Methods Comparison Results

This study compares FastText, Word2vec, GloVe&Word2vec, and the recommended GloVe results, so these techniques analyze text using LSTM classifiers. The F1-score functions as the comparison's premise. Table 2 presents the evaluation of the outcomes of the numerous suggested methodologies. The efficacy of the accuracy algorithm using GloVe attained the highest number (90%) and is superior to that of FastText and Word2vec. The significance of employing the GloVe representation technique as opposed to the quantitative representation is implied by this study's results.

Table 2 Methods comparison

Representation Method	F1-score 0	F1-score 1	Accuracy
FastText	0.74	0.00	0.59

Word2vec	0.76	0.00	0.61
GloVe&Word2vec	0.9	0.89	0.89
GloVe	0.92	0.88	0.90

According to the comparative results, each created representation technique can potentially reduce the feature spaces. Unsurprisingly, word vector approaches can considerably reduce feature sizes, given their exceptional abstract ability. The proposed GloVe representation utilizing LSTM can reduce the feature space and improve accuracy. The outcomes demonstrated the usefulness of context representation techniques.

There is no conclusive evidence that GloVe alone is superior to the combination of GloVe and Word2Vec. However, we can examine the distinctions and parallels between these two word embedding techniques.

A potential advantage of GloVe is that the directions in the embedding space can be meaningful, enabling analogous relationships between words to exist. For instance, in the GloVe vector space, the relationship "king - man + woman queen" holds.

B. Hyperparameter Implementation Results

This study manipulated the hyperparameter to enhance the precision of the results as shown in Table 2, the first hyperparameter was used.

Table 3 First Hyperparameter

GloVe Hyperparameter	Training data	Learning rate	Adam	BinaryCrossentropy	epochs
	0.7	0.0001	Default	Default	10
LSTM Hyperparameter	Number of units	Dropout rate	Activation function	Input shape	
	64	0.2	sigmoid	63,300	
Accuracy	0.81				

After a series of modifications and adjustments to the hyperparameter to enhance precision, Table 4 displayed exceptional results.

Table 4 Hyperparameter implementation

Model	Hyper.	Value											
GloVe	train_df	0.7	0.7	0.7	0.7	0.7	0.5	0.75	0.8	0.85	0.9	0.95	1
	learning_rate	0.0001	0.001	0.0001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	Adam	default	default	default	default	default	default	default	default	default	default	default	default
	BinaryCrossentropy	default	default	default	default	default	default	default	default	default	default	default	default
	epochs	10	40	10	40	40	40	40	40	40	40	40	40
	nomber of units	64	64	64	64	64	64	64	64	64	64	64	64
	dropout rate	0.2	0.2	0.2	0.2	0.5	0.5	0.2	0.2	0.2	0.2	0.2	0.2
	Activation function	sigmoid	sigmoid	softmax	softmax	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid
LSTM	Input shape	63, 300	63, 300	63, 300	63, 300	63, 300	63, 300	63, 300	63, 300	63, 300	63, 300	63, 300	63, 300
accuracy		0.81	0.9	0.41	0.41	0.9	0.77	0.92	0.93	0.95	0.97	0.97	0.99

As seen in Table 4, the accuracy decreases to 41% when the Softmax activation function is used; this demonstrates that sigmoid is superior for classification because softmax is typically used for multi-class classification problems and sigmoid for binary classification problems. Increasing the learning rate to 0.001 improves performance. The increase in trained data results in a discernible improvement in accuracy rates. The Dropout rate of 0.2 prevents overfitting by arbitrarily setting 20% of LSTM outputs to zero during training. The performance of the proposed GloVe with LSTM in detecting ADRs was excellent.

C. Discussion

This literature review determines the originality of the proposed method. This section compares pertinent works with precision, as shown in Table 5.

Table 5 ADR summary

Author	Dataset	Representation approach	Feature	Accuracy
Yates & Goharian (2013)	Dataset for the Yates and Goharian (2013) benchmark	Rule-based	Trigger terms	0.78
Yousef et al. (2019)	Yates and Goharian's (2013) benchmark dataset was amended by Mohammad Yousef et al. (2019)	NB, LR, SVM	Trigger terms	0.69
Nafea et al. (2021)	Yates and Goharian's (2013) benchmark dataset was amended by Mohammad Yousef et al. (2019)	NB, LR, SVM	LSA	0.82
Proposed method	Yates and Goharian's (2013) benchmark dataset was amended by Mohammad Yousef et al. (2019)	LSTM	GloVe	0.99

As shown in Table 5, the accuracy of the proposed method exceeds that of the related work. It is important to note that the related study relied on Trigger terminology. The proposed method outperformed Yates and Goharian's (2013) approach on the same data set before the update by Yousef et al. (2019). Since the proposed method is based on GloVe, the efficacy of the proposed GloVe in detecting ADRs is generally superior to that of the baseline GloVe.

Due to its ability to decrease feature spaces and token model similarity, the GloVe representation is the most advantageous of the four examined methods. This discovery lends credibility to the proposal of using GloVe to retrieve ADRs. GloVe, whose semantic correspondences have been accurately identified, is regarded as superior to a predefined collection of trigger terms due to its superior performance. Due to its ability to manage synonymy issues within a given dataset, GloVe's word similarity analysis has attained a higher f1-score of classification than the basic vector space model or N-gram representation. In addition, GloVe is capable of working effectively with datasets spanning a broad range of topics, making it an ideal candidate for analyzing data on hazardous drug reactions, which span a broad range of medical topics.

V. CONCLUSION

In recent years, the proliferation of social networking sites has contributed significantly to expanding written information. Ordinary users in modern society could express their opinions on various topics. Identifying potentially hazardous drug reactions is one of these topics (ADRs). This study seeks to improve the accuracy of ADR identification by devising and recommending a semantic technique based on the Word Embeddings technology called Global Vectors (GloVe). Additionally, GloVe can improve the identification of adverse drug reactions (ADRs) by mapping words into a meaningful space in which the distance between words corresponds to their syntactic similarity. It is accomplished by identifying connections between words, including synonyms. After describing all of the experiments conducted during this study, it was discovered that the suggested GloVe detected ADR accurately. Compared to the accuracy obtained by the baseline trigger terms from the literature by Yousef et al. (2019) and Nafea et al. (2021), the results demonstrate that the suggested GloVe outperforms the baseline by reaching 99% of F1-scores using LSTM. These comparisons demonstrate the precision attained by the baseline trigger terms (2021). In detecting ADRs, the performance of the suggested GloVe is generally superior to that of the GloVes used as a baseline. This finding provides support for the viability of using GloVe to collect ADRs.

ACKNOWLEDGEMENT

The faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, are thanked for allowing the authors to conduct this research.

REFERENCE

Alimova, I. & Tutubalina, E. 2018. Automated detection of adverse drug reactions from social media posts with machine learning. Networks, S. (ed.). *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*vol. 10716 LNCS, , pp. 3–15. Cham: Texts . Springer.

Ebrahimi, M., Yazdavar, A.H., Salim, N. & Eltyeb, S. 2016. Recognition of side effects as implicit-opinion words in drug reviews. *Online Information Review* 40(7): 1018–1032.

Kiritchenko, S., Mohammad, S.M., Morin, J. & De Bruijn, B. 2018. NRC-Canada at SMM4H shared task: Classifying tweets mentioning adverse drug reactions and medication intake. *CEUR Workshop Proceedings* 1996: 1–11.

Korkontzelos, I., Nikfarjam, A., Shardlow, M., Sarker, A., Ananiadou, S. & Gonzalez, G.H. 2016. Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. *Journal of Biomedical Informatics* 62: 148–158.

Kumar, N.H., Narendra, B., Upendra, K. & Rajesh, K. 2019. A review on Adverse drug reactions monitoring and reporting. *International Journal of Pharmacy Research \& Technology* 9(2): 12–15.

Masino, A.J., Forsyth, D. & Fiks, A.G. 2018. Detecting Adverse Drug Reactions on Twitter with Convolutional Neural Networks and Word Embedding Features. *Journal of Healthcare Informatics Research* 2(1–2): 25–43.

Nafea, A.A., Omar, N. & AL-Ani, M.M. 2021. Adverse Drug Reaction Detection Using Latent Semantic Analysis. *Journal of Computer Science* 17(10): 960–970.

Nikfarjam, A., Sarker, A., O'Connor, K., Ginn, R. & Gonzalez, G. 2015. Pharmacovigilance from social media: Mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association* 22(3): 671–681.

Pain, J., Levacher, J., Quinquenel, A. & Belz, A. 2016. Analysis of Twitter Data for Postmarketing Surveillance in Pharmacovigilance. https://aclanthology.org/W16-3914 [27 February 2023].

Plachouras, V., Leidner, J.L. & Garrow, A.G. 2016. Quantifying Self-Reported Adverse Drug Events on Twitter. *Proceedings of the 7th 2016 International Conference on Social Media & Society - SMSociety '16* 1–10. http://dl.acm.org/citation.cfm?doid=2930971.2930977.

Qian, J., Qiu, X., Tan, X., Li, Q., Chen, J. & Jiang, X. 2022. An Attentive LSTM based approach for adverse drug reactions prediction. *Applied Intelligence* 1–15.

Rajaraman, A. & Ullman, J.D. 2011. Mining of massive datasets. *Mining of Massive Datasets* 9781107015: 1–315.

Yates, A. & Goharian, N. 2013. ADRTrace: Detecting Expected and Unexpected Adverse Drug Reactions from User Reviews on Social Media Sites. Serdyukov, P., Braslavski, P., Kuznetsov, S.O., Kamps, J., Rüger, S., Agichtein, E., Segalovich, I. & Yilmaz, E. (eds.). *Advances in Information Retrieval*, pp. 816–819. Berlin, Heidelberg: Springer Berlin Heidelberg.

Yousef, R.N.M., Tiun, S. & Omar, N. 2019. Extended trigger terms for extracting Adverse Drug Reactions in social media texts. *Journal of Computer Science* 15(6): 873–879.