ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 17, Issue 3, 2023

Enhanced Detection of Adverse Drug Reactions in Social Media Using Extended Trigger Terms

Ms.B.Swapna

Assistant Professor, Department of ECE,
Malla Reddy College of Engineering for Women.,
Maisammaguda., Medchal., TS, India (@gmail.com

Abstract: Medications may lead to a problem known as an Adverse Drug Reaction (ADR). Extraction of adverse drug reactions from social networks where people share their thoughts on a particular drug has been the subject of research. In order to extract entities, you need to look for certain phrases, called trigger terms, which might appear either before or after ADRs. It is recommended that these concepts be expanded, particularly when dealing with N-gram representations in multiples. With the use of the N-gram's multiple representation, this research hopes to suggest an expansion of trigger phrases. The suggested modification is utilized to train three classifiers: support vector machine, Naïve Bayes, and linear regression. The experiments make use of two benchmark datasets. In addition, Count Vector (CV) and Term Frequency Inverse Document Frequency (TFIDF) have been used as document representations. The suggested expanded trigger words achieve 88% and 69% of F1-scores for the first and second datasets, respectively, demonstrating that they exceed the baseline. This result suggests that the suggested expanded trigger phrases are useful for identifying novel ADRs.

Keywords: Acute Pharmacological Reaction, Trigger Words, Feature Extraction, SVM, and Linear Regression

Introduction

Many fields have been impacted by the meteoric rise of social media, including advertising, commerce, and the arts. People who use social media to voice their ideas have shown a growing interest in the medical field, for example (Denecke and Deng, 2015). Drug reviews, which detail the experiences of actual drug users, are one kind of such viewpoint. Patients may have a variety of Adverse Drug Reactions (ADRs). In the case of "this medicine made me sleepy," for instance, Before doing sentiment analysis, which entails categorizing people's attitudes as positive or negative, it is crucial to extract these ADRs, which represent important entities (Sohn et al., 2011).
The majority of ADR extraction investigations have used machine learning methods with models trained on past instances. (Liu and Chen, 2015) These models have the ability to extract new or unknown samples. Nonetheless, a feature space that may be created during model creation is the most important component of these methods. Specific things are described by their features, which are descriptive qualities (Alshaikhdeeb and Ahmad, 2017; 2018).

Contextually addressing the feature space

trigger phrases, which are required particular keywords that appear either before or after ADRs, in order to extract them. Researchers have used a set of phrases to activate ADR extraction in their studies (Ebrahimi et al., 2016; 2016). Because ADRs are abundant and have several synonyms and interpretations, trigger phrases still need a variety of expansions.
New trigger words with different N-gram topologies (unigram, bigram, trigram, and quadgram) are the target of this research. In the experiments, two benchmark datasets are used. In addition, Count Vector (CV) and Term Frequency Inverse Document Frequency (TFIDF) have been used as document representations. Furthermore, we have a look at three classifiers: Support Vector Machine (SVM), NaÅve Bayes (NB), and Linear Regression (LR).

Related Works

Many studies have been proposed to extract adverse drug entities by using a wide range of features, along with machine learning techniques. For example, Yu (2016) used a set of features to identify drug–effect relation. The set of features contain Bag-

ISSN 2454-9940
www.ijasem.org
Vol 17, Issue 3, 2023

of-Words (BoW), where multiple numbers of topologies of N-gram, including unigram and bigram, have been used. Part-of-Speech (POS) tagging has been utilized to indicate the syntactic tags of terms. Semantic correspondences have been shown using the WordNet lexicon. We have used four classifications—decision tree, maximum entropy, NB, and SVM—to find the drug-effect association.
Mishra et al. (2015) used statistical characteristics to extract drug-related entities from drug reviews. These features included word frequency and weighting terms. The Word Net lexicon is used in conjunction with statistical characteristics to ascertain semantic relatedness. It has also been used to identify items linked to drugs using an SVM classifier.
An automated method for detecting the effects of drugs was developed by Pain et al. (2016) using data obtained from Twitter. As trigger phrases, they used a collection of keywords and hashtags. The suggested characteristics, when applied to an SVM classifier, may detect a wide variety of drug-effect entities.
In their 2016 study, Ebrahimi et al. used a collection of medical concepts to identify medication adverse effects in medical reviews. The researchers used named entities as trigger phrases. The syntactic tag of phrases has also been identified via POS tagging. The detection of pharmacological side effects has been accomplished with the help of two classifiers: a rule-based classification technique and support vector machines (SVM).
In order to extract negative drug occurrences from Twitter reviews, Plachouras et al. (2016) used an N-gram representation in conjunction with a collection of trigger phrases or Gazetteers characteristics. The final extraction using the suggested characteristics was made possible by using the SVM classification approach.
A mix of morphological and semantic variables was used by Moh et al. (2017) to detect adverse drug occurrences. Negations and question marks are morphological elements that are used in this context. When working with SentiWordNet, the semantic feature is used. Classification tasks have also made use of SVM and NB.
A deep learning method for ADR extraction has been suggested by Lee et al. (2017). Using Twitter's unlabeled data, the suggested method trained a semi-supervised Convolutional Neural Network (CNN). When contrasted with the conventional supervised methods, the suggested method performed much better.
A deep learning method for ADR extraction based on word embedding was also suggested by Cocos et al. (2017). The scientists have trained a Recurrent Neural Network (RNN) using a large quantity of social data, mostly from Twitter, to create word embeddings. We have compared the suggested strategy to the state of the art, which was using lexicon-based methods. The suggested strategy outperformed the others, according to the results.

A deeb neural network for ADR extraction was suggested by Wang et al. (2019). Using a pre-trained model for embedding biological words is the key to the suggested approach. The suggested approach has therefore been evaluated using a benchmark dataset. The results demonstrated that the suggested technique outperformed the baseline ones.

Materials and Methods

Materials: Dataset

The dataset used in the experiments is a set of drug reviews from different categories and from patients' comments about ADRs available on different discussion forums on health websites and social media in English language. Thus, two datasets are used:

Dataset 1: Ebrahimi *et al*. (2016) is annotated by a medical expert. A total of 225 drug reviews are randomly selected from www.drugratingz.com for manual annotation. These reviews are related to diverse categories, such as pain relief and antidepressant drugs. Indeed, the comment sections of drug reviews in this website are full of sentences containing drug side effects and the role of the algorithm is to identify these side effects correctly. A total of 70 reviews are to generate rules manually and 155 reviews are assigned as a test set

Dataset 2: The annotated ADR review dataset is used in Yates and Goharian (2013). The review dataset is collected from drug review social media sites, namely, askapatient.com, drugs.com and drugratingz.com

Table 1 shows the details of each dataset.

The proposed method consists of three main phases (Fig. 1). The first phase is related to the datasets used within the experiments, along with the required preprocessing tasks that are intended to turn the data into an appropriate form. The second phase is related to feature extraction, i.e., trigger terms, which represent the core of this study. Lastly, the third phase involves the application of machine learning techniques to classify ADRs based on the utilized features. Each phase is discussed in further detail in the next subsections.

Phase 1: Input and Preprocessing

Both datasets have to undergo the preprocessing task. The text segmentation stage aims to run some of the preprocessing algorithms on the corpus to prepare it for the next phases. The aforementioned tasks can be illustrated as follows:

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 17, Issue 3, 2023

1. Sentence splitting: This task aims to split a text into a series of sentences by identifying sentence boundaries. For this purpose, the Natural Language Tool Kit (NLTK) library is used to achieve this task
2. Tokenization: This task aims to split a text stream into a series of tokens. Similarly, the NLTK library is used
3. Stemming: This task aims to reduce inflected (or sometimes derived) words to their word stem, base, or root form of a particular set of words by removing various suffixes while preserving the meaning of the word
4. Stop word removal: This task aims to remove the frequent words of a language that does not carry any significant information on their own. These words are often removed at the preprocessing stage to reduce the number of less informative features known as noise data
5. POS tagging: This task aims to identify the words with their POS categories, such as nouns, verbs, adjectives and adverbs

Phase 2: Extended Trigger Terms

In this phase, a combination of lexical, syntactical and contextual expressions and trigger terms is used to detect the adverse side effects of drugs. Trigger terms are extracted from the state of art by analyzing the sentences containing ADR and trigger term. Two lists, namely, existing terms and new trigger terms, are created (Tables 2, 3 and 4).

Table 2 shows the trigger terms extracted by Ebrahimi *et al*. (2016). Among the terms, "caused" or "makes" are associated with ADRs in their dataset.

Table 3 shows the trigger terms in Yates and Goharian (2013) dataset, whose terms are similar to those in Ebrahimi *et al*. (2016) dataset. These terms (e.g., "caused" and "made me") are also related to ADRs.

Building the Extended Trigger Terms

This study utilizes a statistical technique, namely, Point-wise Mutual Information (PMI), to identify new and extended trigger terms. This technique aims to examine the co-occurrence among terms. Both datasets are being annotated already. As such, PMI is applied to terms that frequently occur with ADRs. PMI can be computed on the basis of the following equation (Zhang *et al*., 2009):

$$PMI(ADR,t_i) = \log \frac{P(ADR,t_i)}{PMI(ADR) \times P(t_i)} \quad (1)$$

where, *P (ADR)* refers to the probability of individual ADR occurrence; *P(ti)* denotes the probability of the individual occurrence of certain terms and *P(ADR, ti)* corresponds to the co-occurrence among *ADR* and certain terms. The highest value of PMI indicates a high correlation among the two terms.

The results of *PMI* on both datasets reveal trigger terms that are similar to the baseline. Therefore, our study implements a manual filtering task to exclude the ones used by the baseline. With this filtering approach, new and extended trigger terms are identified. Table 4 shows a sample of these proposed extended terms associated with some example patterns from both datasets.

The sentences are represented in a feature vector that contains the selected features. Such a representation aims to articulate the distinctive terms in separated attributes. In this regard, every sentence is examined on the basis of the occurrence of such terms (i.e., whether or not a term occurs in a sentence). Here, the features are the distinctive terms and two frequency topologies, namely, Term Frequency –Inverse Document Frequency (TFIDF) and count vector, are depicted.

In count vector, the aim is to simulate the occurrence of terms as binary representation:

> *"If a term occurs, it is represented as '1'; otherwise, it is represented as '0'".*

TFIDF aims to represent the frequency of terms as real values that indicate the ratio of occurrence between a term and a sentence, along with a term with other sentences, which can be computed using the following equation (Chen *et al*., 2016):

$$TFIDF = (t/d) = tf_{td} \times \log \frac{N}{N_t}, \quad (2)$$

where, tf_{td} refers to the occurrence of the term in a particular document. The document in our study is a metaphor for a sentence. N is the number of the total documents (i.e., sentences) and Nt is the document that contain the term t.

Apart from frequency, the terms have been articulated in the N-gram representation by using four topologies, namely, unigram (i.e., one term), bigram (i.e., two terms), trigram (i.e., three terms) and quadgram (i.e., four terms).

Phase 3: Training Model

In this phase, machine learning is applied to classify ADRs. Classification methods, including SVM, NB and LR, are used to evaluate the performance based on f-measure.

The first classification method is SVM, which works by determining an accurate separator between data instances in a 2-dimensional space. Such a separator can be computed using the following equation (Ebrahimi, *et al*., 2016):

$$f(\vec{x}) = \text{sgn}((\vec{x} \times \vec{w}) + b) = \begin{cases} +1: & (x \times w) + b > 0 \\ -1: & Otherwise \end{cases}, \quad (3)$$

where, $d+$ ($d-$) denote the shortest path between the positive and negative examples.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 17, Issue 3, 2023

NB is working by identifying the probabilities of classes for the data instances. Such a probability can be calculated using the following equation (Elhadad *et al*., 2019):

$$P(C_i \mid d) = \frac{P(C_i)P(d \mid C_i)}{P(d)} \quad (4)$$

where, $P(C_i|d)$ is the posterior probability of class C_i given the predictor (x, attributes).

LR works by determining the linear equation of class probability, which can be depicted as follows (Montgomery, 2015):

$$y = a + bx, \quad (5)$$

where X is the dependent variable, a is the y-intercept and b is the slope of the line.

The evaluation involving f-measure can be depicted by the following equation:

$$F1 - score = 2 \times \frac{\frac{TF}{TF + FF} \times \frac{RF}{TF + FN}}{\frac{TF}{TF + FF} \times \frac{RF}{TF + FN}} \quad (5)$$

where, TP is the correctly classified ADR, FP is the incorrectly classified ADR and FN is the correctly classified ADR in accordance with the total number of ADRs.

The three classifiers are trained on the extracted patterns produced by the proposed trigger terms and the benchmark ones. This training aims to build a model that can classify new data in the testing phase. During the training, the model of each classifier learns the cases of the potential occurrence of ADRs. Table 5 shows the experimental settings.

Table I: Dataset details

Attributes	Dataset 1 (Ebrahimi *et al*., 2016)	Dataset 2 (Yates and Goharian, 2013)
Number of total reviews	225 (labelled 157)	2500 (labeled 246)
Number of sentences	1212	944
Number of ADR	372	982

Table 2: List of benchmark and proposed trigger terms based on dataset 1 (Ebrahimi, *et al*., 2016)

Benchmark trigger terms (Ebrahimi *et al*., 2016)	Proposed trigger terms
Caused	Causing effects of the drug
Causes	Getting off
Can cause	Wonder for
Caused an	Have been suffering
Cant cause	Have not got
Caused me	Short term effects
Makes you	Suffering with
Made me	Chronic
Make me	I had problems with
Makes people	Really helps

Table 3: List of benchmark and proposed trigger terms based on (Yates and Goharian, 2013)

Benchmark trigger terms (Yates and Goharian, 2013)	Proposed trigger terms
Caused	Started having
Causes	Began having
Makes you	Still have
Made me	Also have
Side effects	Had some
Side effect	Have some
Any side effects	Was having
I have	I am having
Get	Extreme
Side effect	Felt like

Table 4: Samples of extracted trigger terms

Comment_No	Sentence_No	Patterns	Trigger Terms	ADR Found
10	6	DT nn have caused (ADR)	have caused	Bad reaction
11	3	Prp suffer (ADR)	suffer	High blood
15	1	Jj side effect in(ADR)	side effect	arthritis
20	12	makes you (ADR)	makes you	Binge
24	3	Prp had never suffered from(ADR)	had never suffered	depression
26	2	Prp made me (ADR)	made me	Depressed
60	3	my side effect (ADR)	my side effect	stress

Table 5: Experimental settings

Experiment	Description
Feature	1. Baseline trigger terms with TFIDF (Unigram, Bigram, Trigram and Quadgram) 2. Baseline trigger terms with count vector (Unigram, Bigram, Trigram and Quadgram) 3. Proposed trigger terms with TFIDF (Unigram, Bigram, Trigram and Quadgram) 4. Proposed trigger terms with count vector (Unigram, Bigram, Trigram and Quadgram)
Classifiers	1. SVM 2. NB 3. LR
Dataset	1. Benchmark trigger terms dataset 1 (Ebrahimi, *et al*., 2016) 2. Benchmark trigger terms dataset 2 (Yates and Goharian, 2013)
Training and Testing	70% for training and 30% for testing

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 17, Issue 3, 2023

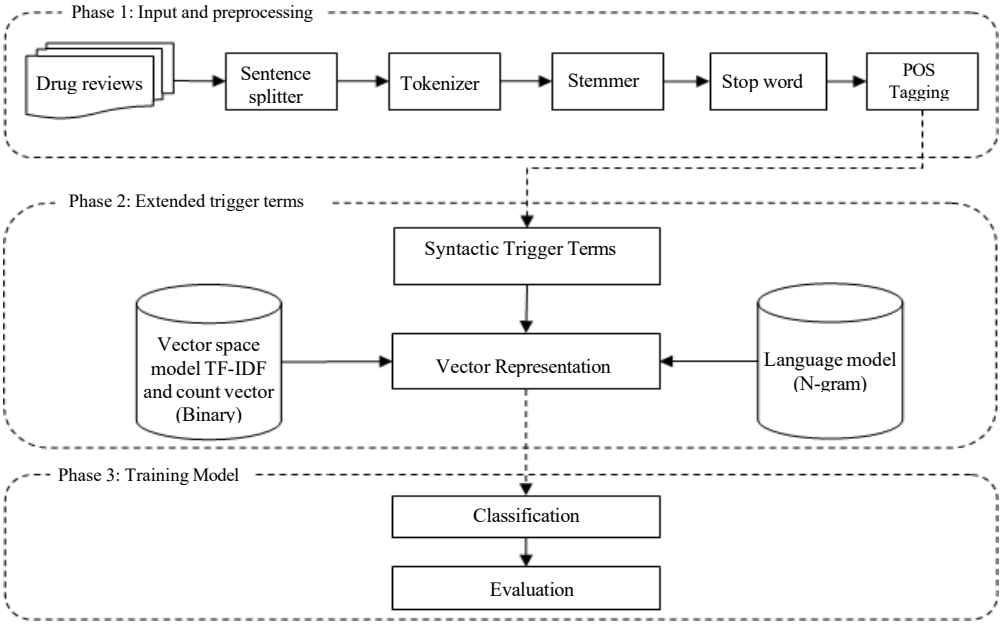

Fig. 1: Proposed extended trigger terms

Results

An assortment of n-gram topologies, including unigram, bigram, trigram, and quadgram, as well as various representations (such as count vector vs. TFIDF), trigger terms (baseline vs. suggested), and experimental designs have all been thoroughly investigated. Furthermore, the outcomes of several classifiers, such as SVM, NB, and LR, have been calculated using the widely used information retrieval measure F1-score. The outcomes of the first dataset are shown in Table 6. The quadgram has the greatest values among the topologies in terms of the F1-score, which is affected by the grams of words (Tables 6 and 7). When comparing the results of the two experiments in Tables 6 and 7, it is clear that the trigram and quadgram approaches yielded comparable results. The value of investigating multigram terms is suggested by this discovery. In comparison to the TFIDF, the count vector has higher F1-score values (Table 6). Each and every classifier, both existing and planned, is subject to this stipulation. This finding further supports the idea that counting vectors, which use binary format, are more important than TFIDF, which use numerical representation. However, the suggested and baseline findings should be compared to verify the proposed trigger conditions. Apparently, all the trials suggest that the proposed technique outperforms the baseline. In particular, the highest result is achieved by using the proposed trigger terms via the count vector with the SVM and the quadgram term. The result of F1-score is 88% (Table 6), demonstrating the effectiveness of the proposed trigger terms.

In Tables 8 and 9, similar to the result of dataset 1, the increase in the term gram affects the F1-score; that is, the quadgram findings indicate the highest values. Tables 8 and 9 show that SVM and NB performed similarly for both studies when trigram and quadgram were used. This result suggests the usefulness of the suggested technique in evaluating multigram phrases. Unlike the findings of dataset 1, the F1-score values of the TFIDF (Table 9) are somewhat higher than those of the count vector (Table 8). When looking at how the suggested technique compares to the baseline, it's clear that the proposed method is superior. In particular, the greatest results are reached by the suggested trigger words using the TFIDF with SVM and both trigram and quadgram terms and the result of F1-score is 69%. Dataset 1 yields a better result (88% vs. 69% in dataset 2). The label variance overlap between the two datasets is the cause of this discrepancy.

When compared to the baseline words, the suggested trigger terms generally do a better job of identifying ADRs. In terms of extracting ADRs, this study suggests that offering an extended trigger is effective. It is important to compare the suggested method not just to the traditional baseline that used old approaches like SVM and NB, but also to more contemporary methods that used considerably more advanced techniques. Actually, an F1-score of 64.5% was achieved by Lee et al. (2017) using a CNN deep learning technique to extract ADRs. These findings show that the suggested strategy is still competitive when compared to the ones produced by other methods.

Table 8: Count vector results of dataset 2 (Yates and Goharian, 2013)

	Count vector (F1-score)					
	SVM		NB		LR	
N-gram	Baseline	Proposed	Baseline	Proposed	Baseline	Proposed
Unigram	0.54	0.55	0.38	0.42	0.55	0.57
Bigram	0.59	0.65	0.51	0.57	0.59	0.64
Trigram	0.59	*0.67*	0.52	*0.61*	0.61	0.66
Quadgram	0.56	*0.67*	0.52	*0.61*	0.61	*0.67*

Table 9: TFIDF results of dataset 2 (Yates and Goharian, 2013)

	TFIDF (F1-score)					
	SVM		NB		LR	
N-gram	Baseline	Proposed	Baseline	Proposed	Baseline	Proposed
Unigram	0.54	0.54	0.38	0.44	0.54	0.54
Bigram	0.58	0.67	0.51	0.58	0.59	0.66
Trigram	0.60	***0.69***	0.54	*0.61*	0.60	*0.68*
Quadgram	0.60	***0.69***	0.54	*0.61*	0.60	*0.68*

However, other studies such as Cocos *et al*. (2017) and Wang *et al*. (2019) whom utilized much sophisticated deep learning approaches, have obtained an F1-score higher the proposed method as 75.5% and 84.4% respectively. Yet, their approaches were requiring a pre-trained data of embedding for the medical words. Considering the feature engineering that has been utilized by the proposed method, it is clear that the proposed method is still considered to be less complicated.

Conclusion

This study proposed an extended set of trigger terms for detecting ADRs. These trigger terms were compared with the baseline ones by using two benchmark datasets. Experiments involved three classifiers, namely, SVM, NB and LR and multiple N-gram topologies, including unigram, bigram, trigram and quadgram. The proposed trigger terms achieved higher results than the baseline ones when quadgram and SVM classification were used. Further studies on feature types would facilitate the process of detecting ADRs.

References

Alshaikhdeeb, B. and K. Ahmad, 2017. Feature selection for chemical compound extraction using wrapper approach with Naive Bayes classifier. Proceedings of the 6th International Conference on Electrical Engineering and Informatics, Nov. 25-27, IEEE Xplore press, Langkawi, Malaysia, pp: 1-6. DOI: 10.1109/ICEEI.2017.8312421

Alshaikhdeeb, B. and K. Ahmad, 2018. Comparative analysis of different data representations for the task of chemical compound extraction. Int. J. Advanced Sci. Eng. Inform. Technol., 8: 2189-2195. DOI: 10.18517/ijaseit.8.5.6432

Chen, K., Z. Zuping L. Jun and Z. Hao, 2016. Turning from TF-IDF to TF-IGM for term weighting in text classification. Expert Syst. Applic., 66: 245-260. DOI: 10.1016/j.eswa.2016.09.009

Cocos, A., A.G. Fiks and A.J. Masino, 2017. Deep learning for pharmacovigilance: Recurrent neural network architectures for labeling adverse drug reactions in twitter posts. J. Am. Med. Informatics Assoc, 24: 813-821. DOI: 10.1093/JAMIA/OCW180

Denecke, K. and Y. Deng, 2015. Sentiment analysis in medical settings: New opportunities and challenges. Artificial Intel. Med., 64: 17-27. DOI: 10.1016/j.artmed.2015.03.006

Ebrahimi, M., A.H. Yazdavar, N. Salim and S. Eltyeb, 2016. Recognition of side effects as implicit-opinion words in drug reviews. Online Inform. Rev., 40: 1018-1032. DOI: 10.1108/OIR-06-2015-0208

Ebrahimi, M., A.H. Yazdavar, N.B. Salim, S. Noekhah and D.L. Paris, 2016. Drug side effect detection as implicit opinion from medical reviews. Res. J. Applied Sci. Eng. Technol., 12: 1086-1094. DOI: 10.19026/rjaset.12.2850

Elhadad, M.K., K.F. Li and F. Gebali, 2019. Sentiment Analysis of Arabic and English Tweets. In: Web, Artificial Intelligence and Network Applications, Springer, Cham, ISBN-10: 978-3-030-15035-8, pp: 334-348.

Lee, K., A. Qadir, S.A. Hasan, V. Datla and A. Prakash *et al*. 2017. Adverse drug event detection in tweets with semi-supervised convolutional neural networks. Proceedings of the 26th International Conference on World Wide Web, Conferences Steering Committee, Apr. 03-07, ACM, Perth, Australia, pp: 705-714. DOI: 10.1145/3038912.3052671

Liu, A. and M. Chen. 2015. A research framework for pharmacovigilance in health social media: Identification and evaluation of patient adverse drug event reports. J. Biomed. Informatics, 58: 268-279.

Mishra, A., A. Malviya and S. Aggarwal, 2015. Towards automatic pharmacovigilance: Analysing patient reviews and sentiment on oncological drugs. Proceedings of the IEEE International Conference on Data Mining Workshop, Nov. 14-17, IEEE Xplore Press, Atlantic City, NJ, USA, pp: 1402-1409. DOI: 10.1109/ICDMW.2015.230

Moh, M., T.S., Moh and Y. Peng, 2017. On adverse drug event extractions using twitter sentiment analysis. Netw. Model. Analysis Health Informatics Bio., 6: 18-18. DOI: 10.1007/s13721-017-0159-4

Montgomery, A., 2015. Introduction to Linear Regression Analysis. 1st Edn., John Wiley and Sons.

Pain, J., L. Jessie, Q. Adam and B. Anja, 2016. Analysis of twitter data for postmarketing surveillance in pharmacovigilance, Proceedings of the 2nd Workshop on Noisy User-Generated Text, Dec. 11, Osaka, Japan, pp: 6-39.

Plachouras, V., J.L. Leidner and A.G. Garrow, 2016. Quantifying self-reported adverse drug events on twitter: Signal and topic analysis. Proceedings of the 7th International Conference on Social Media and Society, Jul. 11-13, London, United Kingdom, ACM, pp: 1-6. DOI: 10.1145/2930971.2930977

Sohn, S., J.P. Kocher, C.G. Chute and G.K. Savova, 2011. Drug side effect extraction from clinical narratives of psychiatry and psychology patients. J. Am. Med. Informatics Assoc., 18: i144-i149. DOI: 10.1136/amiajnl-2011-000351

Wang, C.S., P.J. Li, C.L. Cheng, S.H. Tai and Yea, H.K. Yang *et al*. 2019. Detecting potential adverse drug reactions using a deep neural network model. J. Med. Internet Res., 21: e11016- e11016.

Yates, A. and N. Goharian, 2013. ADRTrace: Detecting expected and unexpected adverse drug reactions from user reviews on social media sites. In: Advances in Information Retrieval, Serdyukov, P. (Ed.), Springer, Berlin, Heidelberg, ISBN-10: 978-3-642-36972-8, pp: 816-819.

Yu, A., 2016. Towards extracting drug-effect relation from Twitter: A supervised learning approach. Proceedings of the IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing and IEEE International Conference on Intelligent Data and Security, (IDS' 16), IEEE, pp: 339-344.

Zhang, W., T. Yoshida, X. Tang and T.B. Ho, 2009. Improving effectiveness of mutual information for substantival multiword expression extraction. Expert Syst. Appl., 36: 10919-10930. DOI: 10.1016/j.eswa.2009.02.026