Research Article
Drug Side Effect Detection as Implicit Opinion from Medical Reviews

Monireh Ebrahimi, Amir Hossein Yazdavar, Naomie Binti Salim, Shirin Noekhah and
Deborah Libu Paris
Faculty of Computing, Universiti Teknologi Malaysia, Johor, Malaysia

Abstract: The enormous growth of online reviews in social media provides a valuable resource for human decision-making activities in diverse domains such as the medical domain. Extracting explicit and implicit opinions is one of the main tasks in the opinion mining area. As implicit opinion mining is a complicated task, limited work has been done on it, especially in the medical domain, as implicit opinion is a domain dependent task. Side effects are one of the critical concepts the recognition of which is a challenging task since it coincides with disease symptoms both lexically and syntactically. To the best of our knowledge, limited work has been done on side effect extraction from drug reviews. This study tries to extract drug side effects as implicit opinions from drug reviews of drugratingz.com by using the rule-based and SVM techniques. Due to the novelty of this issue, corpus construction is also carried out. The results proved that the combination of lexical, syntactical, contextual and semantic features leads to the best results in the SVM technique in comparison with the rule-based algorithm in terms of side effect detection. In this study, we develop a system to detect side effects in the drug reviews as a subtask of detecting implicit opinions in medical sources and discriminate between side effects and disease symptoms. The proposed technique, as an implicit opinion mining system, can help patients to investigate the drug before taking it and help physicians and drug producers to consider user feedback in their decision-making.

Keywords: Drug reviews, implicit opinion, rule-based method, side effect, SVM algorithm

INTRODUCTION

The large amount of medical online information and rapid growth of social media in the medical domain mean that people no longer use a drug before investigating what other patients/physicians say about the specific drug on the Internet. Information extraction and data mining of biomedical text refer to the systematic extraction of structured data from semi-structured or unstructured documents. Medical documents can be divided into different categories based on their aspects, such as published papers and electronic health records (Simpson and Demner-Fushman, 2012). Most drug reviews can be extracted from medical websites, such as Drugratingz.com and Druglib.com.

Extracting and analyzing review opinions manually from the huge amount of sentences are a tedious if not impossible task. People can express their ideas and experiences about consuming a drug implicitly or explicitly through opinionated websites. Unlike explicit opinion mining, no significant amount of research has been done on implicit opinion mining in the medical domain, as it is a domain dependent task. Obviously,

considering both implicit and explicit opinions can improve the accuracy of the detection process. Side effects (such as anxiety, insomnia and headache, etc.) as a result of using a specific drug, play a critical role in the analysis of medical reviews. Patients usually describe their experience about their disease and their pre-conditions and post-conditions before/after using the specific drug. Although side effects can imply both positive and negative opinions about a drug, talking about a drug's side effects is rarely positive, with positive terms being more related to drug effectiveness.

The biomedical domain in comparison with other domains benefit from the large volume of knowledge resources and tools. The Unified Medical Language System (UMLS) is the main biomedical lexica and tool, which was generated by the US National Library of Medicine (NLM) and has been used by many researchers in this field (Li, 2011). The SeReMeD tool, MetaMap configurable program (Aronson, 2001), World Health Organization (WHO) Adverse Drug Reaction Terminology and National Drug Formulary Reference Terminology (NDFRT), as drug related terminology of UMLS, are other tools and resources that can be used.

Corresponding Author: Naomie Binti Salim, Department of Computing, Faculty of Computing, University Teknologi Malaysia, Johor, Malaysia, Tel.: +607 5538726

There are various challenges in terms of side effect extraction as implicit opinion detection. In drug reviews, we should differentiate between the side effects of the drugs and the information that is narrative and contains patients' experiences or disease symptoms. Name variation, abbreviations and acronyms, lack of a complete dictionary and context dependency of the meaning, etc., are among the other challenges in the detection of side effects. In this study, we propose two techniques based on regular expression and machine learning approaches for the purpose of detecting drug side effects and compare the results of both techniques. Moreover, we try to differentiate between drug's side effect and disease symptoms with high accuracy and less false positive detection samples.

LITERATURE REVIEW

Our work in side effect detection in drug reviews as implicit opinion is related to three research fields, implicit sentiment analysis, classification technique and medical domain reviews. Therefore, we review the related techniques of implicit and explicit sentiment analysis in medical domain. Furthermore, we review the approaches of sentiment extraction and classification.

Denecke and Nejdl (2009) classified social media content into two groups- informative (use of adjectives) and affective (medical terminology)-by exploiting the machine learning technique. Goeuriot *et al*. (2011) in their analysis of drug reviews used linguistic features and determined the common sentiment aspects in the drug reviews. Noferesti and Shamsfard (2015) proposed a model for indirect opinion mining in medical documents and a novel approach to construct a knowledge-based corpus for indirect opinions, called OpinionKB. Katsahian *et al*. (2015) used an existing rating tool on a set of social network opinionated websites to evaluate the capabilities of these tools to help researchers to find the most adapted website to mine adverse drug reactions. While their approach is similar to subjective classification (which implies adjective classification), the side effects are medical terminology, which implies typically unfavorable opinion, so we cannot use this algorithm to determine opinionated sentences in the medical domain. Named Entity Recognition and Relation Extraction are two types of approaches which are widely used in sentiment analysis.

Side effect extraction is a special NER (Named Entity Recognition) problem, which is used in biomedical literature in pharmacovigilance. Li (2011) proposed a statistical algorithm by applying statistical NLP techniques to find adverse reactions to cholesterol-lowering drugs in drug reviews. They could also discriminate patient precondition from drug side effects by filtering the symptoms. This algorithm has some limitations, such as focusing on one type of drug for one disease, ignoring the common side effects of drugs for specific diseases and not being able to detect all the side effects of one drug. Skentzos *et al*. (2011) exploited TextMiner to find adverse drug reactions in Statins in the EMRs (Electronic Medical Records) of patients. NER systems have high accuracy in their results based on their evaluation (Simpson and Demner-Fushman, 2012). NER can be performed using a dictionary-based algorithm, rule-based algorithm or machine learning-based method. The primary entity recognition systems use rule-based methods to extract data from medical documents (Zweigenbaum *et al*., 2007). As their sentiment analysis is at the document level, the method suffers from the drawbacks of document level analysis. Although their method can detect side effects, it cannot discriminate the drug side effect from the disease symptoms.

In addition to the NER process, to have a more accurate side effect detection system, we need to recognize drug-treat relations and drug-cause relations (Cohen and Hersh, 2005). In a statistical-based study by Cao *et al*. (2007), they used the degree of co-occurrence of medical entities. Their method was simple and easy to apply, but it was not comprehensive. The Drug-Symptom and Disease-Symptom relations are the two most important and relevant areas that can be helpful for side effect detection. Wang *et al*. (2010) used co-occurrence criteria to identify the relation between entities and used the section name where the entities occur to increase the accuracy. However, this method is not applicable for unstructured narrative drug reviews. Regarding implicit opinion extraction, Zhang and Liu (2011) proposed a statistical approach to extract the noun product features of opinion from four different datasets including the drug review dataset. Yalamanchi (2011) focused on the opinion mining of drug reviews by proposing a system called Sideffective to extract and rank the negativity of reviews with respect to the side effects using their proposed Negativity Meter formula and evaluate the sentiment of the patient regarding a specific drug in terms of its side effects. The main drawback of this approach is the side effect extraction algorithm proposed by Rajagopalan (2011). Weeber *et al*. (2000) proposed a text-based discovery system (DAD), which has the ability to extract a new hidden association between the adverse drug reaction and the disease symptom. Niu *et al*. (2005) performed the polarity analysis of medical documents, as a classification problem by using SVM and extracting unigrams and bigrams, change phrases and UMLS features. In similar work, Swaminathan *et al*. (2010) identified the polarity and strength of biomedical relationships in journal articles and classified their relationships. These works are different from our methods in terms of the corpus content, SVM features set and its application.

Despite of the critical role of implicit opinion mining in medical documents, an efficient and accurate algorithm for extracting the side effects as an implicit opinion concept has not been proposed. Mostly, they consider the side effect to be the same as the disease symptoms; therefore, exploiting a combination of methods is crucial to achieve a more accurate detection model. In this study, we extract side effects by considering the context of the review from drug reviews by systematically detecting and filtering the disease symptom to improve the accuracy.

PROPOSED TECHNIQUE

In this section, we present our proposed implicit opinion extraction model for drug reviews. In this model, we consider the side effect of drug as implicit opinion, which can influence on the orientation of drug reviews.

The side effect detection technique tries to integrate implicit opinion detection into the medical opinion mining system by differentiating between a disease-Manifestation Related Symptom (MRS) and a drug-Adverse Drug Event (ADE). To accomplish this task, we use the two most popular techniques-rule-based approach and SVM technique. The rule-based approach uses the regular expression proposed by Chapman *et al*. (2007) to extract the contextual features from clinical documents, while SVM, as a machine learning system, is used to classify the side effect and disease symptoms. The corpus contains 225 drug reviews from different categories collected from DrugRatingz.com. Finally, we compare the results of both techniques to find the most accurate approach based on their precision (detecting of minimal disease symptom or other medical terms as a side effect-high precision), recall (ensuring drug side effects are not missed-high recall) and the f-measure. The drug reviews were stored in XML file format using windows-1252 encoding into GATE corpus.

DRUG REVIEWS TEXT SEGMENTATION, TERM EXTRACTION AND MAPPING

As the pre-processing step, we exploit the ANNIE plugin of GATE language processing tools to tokenize and segment the text into sentences and determine the POS tags of each token. Then, this output is used by the Tagger_MetaMap plugin, as the input of the rule-based algorithm and to construct some learning features of the SVM algorithm. The Tagger_MetaMap plugin (Gooch and Roudsari, 2011) wraps the MetaMap Java API client to allow the specified annotation content to be processed by MetaMap and the results are converted to GATE annotations and features. We also assign words and phrases using the proper semantic concept of UMLS Metathesaurus, which contains 1.7 million biomedical concepts that are assigned to at least one of 134 semantic types that are grouped into 15 semantic groups (Denecke and Nejdl, 2009). Most of the medical text words that indicate the side effects and symptoms are in the Disorder semantic group of UMLS.

In text segmentation, the ANNIE English Tokeniser is applied to split the reviews into tokens. The output is used in medical concepts mapping, rule-based (for writing the left hand side of the regular expression) and SVM algorithm (unigram feature). We also use RegEx Sentence Splitter for the regular expression based splitter in GATE to determine the boundary of the sentence, the scope of trigger terms and for concepts mapping using the Tagger_MetaMap plugin. In addition, the POS tagging is performed using the ANNIE POS Tagger (Fig. 1). All the processing is used as features by the SVM algorithm.

DEVELOPING RULE-BASED AND SVM ALGORITHM

The main goal of this study is to discriminate between the drug side effects and the disease symptoms. To achieve this goal, we develop two algorithms-rule-based and SVM. In the rule-based approach combined with the lexicon-based approach, we use a combination of some simple regular expressions and semantic rules according to the structure of the sentence, since considering the context and syntactic similarity of symptoms and side effects can be detected by applying these rules. Two main regular expressions to accomplish this task are as follows:

RE1: *<trigger term><nW><indexed term>*
RE2: *<indexed term><nW><trigger term>*

where, n is the number of single words or UMLS concepts. A manual list of trigger terms (such as to treat, Pill, Cause and Make, etc.) for recognition of side effects is generated individually for both disease symptoms and side effects (Appendix B). This list is also applied for selecting the features of the SVM algorithm.

We also use SVM to classify the side effects and disease symptoms by exploiting a proper set of features and performing NER. Support Vector Machines (SVM) is an effective and possibly the most popular supervised learning method used in NLP learning problems. SVM has achieved state of the art performance for many NLP learning tasks (Liu, 2007). It would be useful to note that SVM only works on numeric data and builds two-class classifiers. To deal with multiple classes classification problems several methods such as "one against others" and "one against another" can be used (Cunningham *et al*., 2011). In general, SVM determines
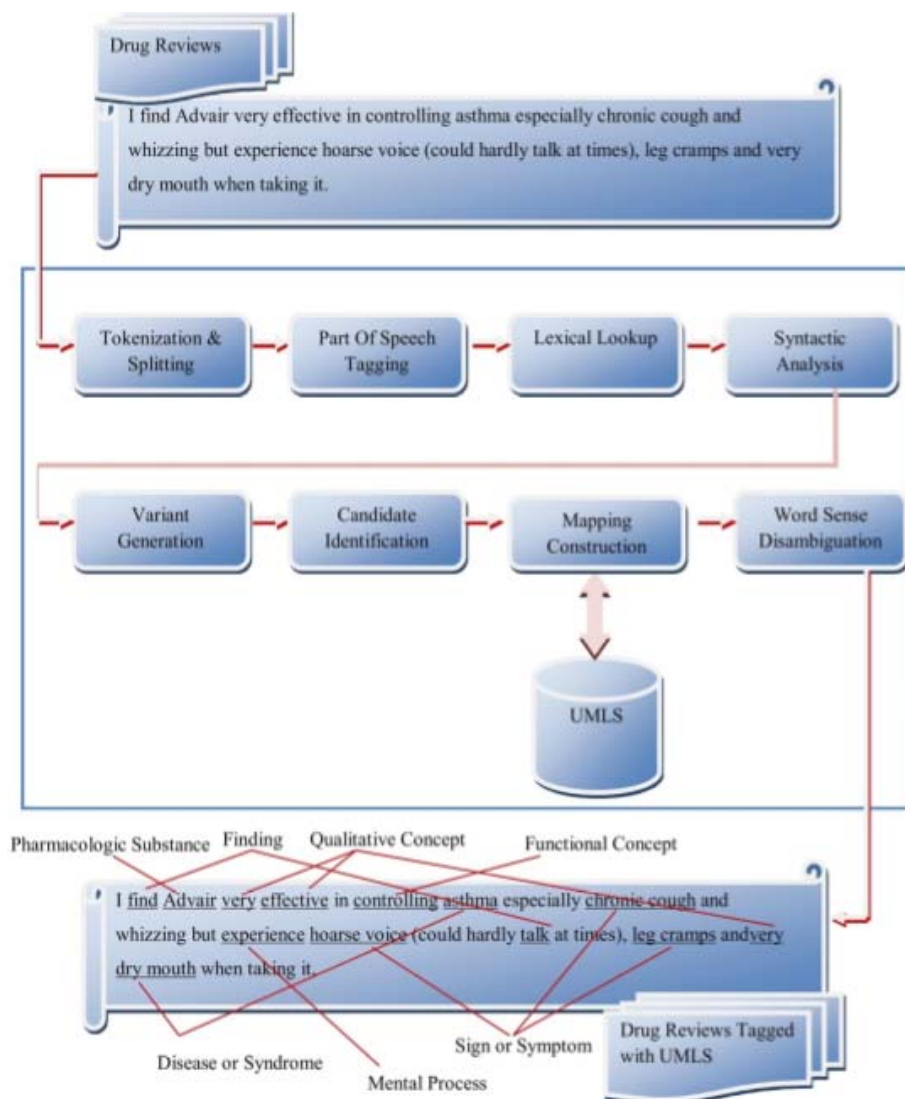
Fig. 1: Example of drug review tagged with UMLS using MetaMap

the optimal hyperplane that separates a set of positive examples from a set of negative examples with a maximum margin in a training data. There are 3 cases that require 3 different approaches when SVM is exploited: a linear SVM for a linear separable case, linear SVM for linearly non separable case and finally non-linear SVM for linearly non separable problems. To deal with the second case, the margin concept is softenized to allow some error data point at the wrong side of the margin boundaries by adding extra cost of errors to the objective function. Finally, to handle the problems with nonlinear decision boundaries, SVM map data points into a higher dimensional space called the feature space a linear hyperplane can separate positive and negative examples by using kernel functions. More mathematical detail on SVM method can be found in Liu (2007). The extraction process is done by SVMLibSvmJava using the GATE learning plugin and JAPE rules. Linear and polynomial kernel functions and different values of uneven margin parameters are used when building the SVM and 5-fold cross validation along with representing the most salient NLP feature of the learned model, by using the VIEWPRIMALFORMMODELS mode of GATE batch learning processing resources, which is used to evaluate the proposed learning model.

EXPERIMENTAL RESULTS

In this section, the results of the pre-processing step along with the results of side effect recognition using two proposed techniques – rule-based and SVM – are discussed and compared with each other. Due to the novelty of the side effect detection area, one of our contributions is the construction of the annotated corpus. Due to the effect of the Tagger_MetaMap accuracy on the efficiency of our proposed method, its accuracy will be evaluated.
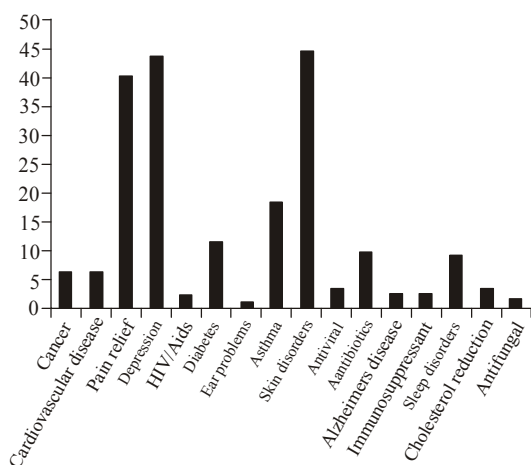
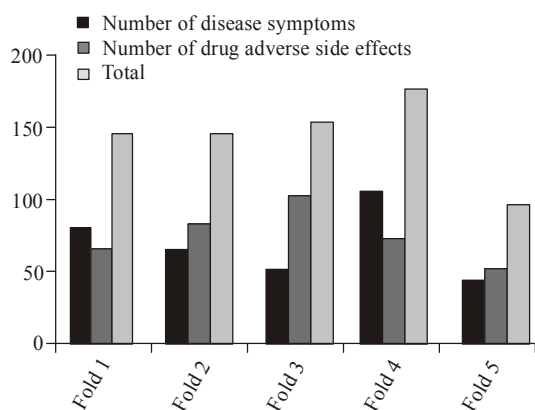Fig. 2: Distribution of drug categories over the corpus

Fig. 3: Disease symptom and drug side effect distribution over the corpus

Dataset: The corpus contains 225 manually random annotated drug reviews from different categories collected from DrugRatingz.com. In the rule-based method, 70 reviews are considered as a development set and 155 as a test set. In addition, we use 5-fold cross-validation to evaluate the SVM model. The MetaMap program performs the mapping task of medical terms of drug reviews to UMLS concepts and a list of trigger terms is constructed for further analysis. Figure 2 illustrates a list of drug categories of this study along with the number of drug reviews for each category (Goeuriot *et al*., 2011).

Building corpus results: Due to the novel nature of the side effect detection domain, there is no standard annotated dataset for testing and evaluating the proposed method. So, in this study, in addition to the systematic evaluation mechanism, by applying well-done methods, such as SVM, we use a pharmacist (Dr. Abolfath Ebrahimi, graduated from Faculty of Pharmacy at Shiraz University of Medical Sciences) to identify all the disease symptom and drug side effects in each drug review of the corpus based on the annotation guideline provided (Appendix A). The result of the annotation is used to measure the precision, recall and F-measure. In order to use this annotated corpus as a training set for SVM, we divide it (342 disease symptoms and 372 drug adverse side effects) into 5 almost equal parts (Fig. 3).

Side effects and disease symptom discrimination: We developed the rule-based and SVM algorithm to perform the discrimination task. The experimental results prove that SVM significantly outperforms the rule-based algorithm. In our proposed rule-based approach, we exploit two regular expressions that use the prepared list of trigger terms to determine whether the mapped term is a disease symptom or an adverse drug side effect. The Confusion matrix is shown in Table 1.

Based on the results, we can see that the rule-based method is too low because of certain drawbacks, such as sensitivity to MetaMap performance; using descriptive sentences instead of phrase for disease symptom and side effect description; existence of non-disorder phrases mapped to the UMLS semantic types, which belong to the Disorder semantic group and vice

Table 1: Confusion matrix for rule-based algorithm results

Class label	Correct	Partially correct	Missing	False positive
Disease symptom	78	8	120	92
Side effect	62	5	233	29

Table 2: Confusion matrix for results of the side effect recognition using the rule-based algorithm

Class label	Correct	Partially correct	Missing	False positive
Side effect	62	5	233	29

Table 3: Results of side effect recognition using SVM algorithm

Feature set	Class label	Strict			Lenient			SVM Configuration		
		P	R	F1	P	R	F1	Kernel	τ	PB
1+2+3+8	Side effect	0.71	0.46	0.54	0.90	0.59	0.70	L	0.4	0.2
1+2+3+6+8	Side effect	0.72	0.48	0.57	0.91	0.62	0.72	L	0.4	0.2
2	Side effect	0.51	0.20	0.28	0.61	0.23	0.33	L	0.4	0.2
1	Side effect	0.49	0.31	0.38	0.64	0.40	0.49	L	0.4	0.2

Table 4: Comparison between SVM and rule-based algorithm for side effect and disease symptom discrimination

		Strict			Lenient			Avarage		
Method	Class label	P	R	F1	P	R	F1	P	R	F1
Rule based	Overall	0.55	0.30	0.36	0.59	0.32	0.40	0.57	0.31	0.38
	Disease Symptom	0.44	0.38	0.41	0.48	0.42	0.45	0.46	0.40	0.43
	Side effect	0.65	0.21	0.31	0.70	0.22	0.34	0.67	0.22	0.33
	Overall	0.57	0.47	0.51	0.68	0.56	0.61	0.63	0.52	0.56
SVM	Disease Symptom	0.62	0.58	0.59	0.71	0.66	0.67	0.67	0.62	0.63
	Side effect	0.51	0.38	0.43	0.65	0.48	0.55	0.58	0.43	0.49

Table 5: Comparison between SVM and rule-based algorithm for side effect recognition

		Strict			Lenient			Avarage		
Method	Class label	P	R	F1	P	R	F1	P	R	F1
Rule based	Side effect	0.65	0.21	0.31	0.70	0.22	0.34	0.67	0.22	0.33
SVM	Side effect	0.72	0.48	0.57	0.91	0.62	0.72	0.82	0.55	0.65

versa; and typographical error, etc. Moreover, the small number of analysed drug reviews for constructing the trigger term list leads to low performance.

SVM is the second method that we developed to detect the side effects from the disease symptoms using different combinations of features and configurations under 5-fold cross validation throughout the evaluation. Based on the results of implementing this method the highest accuracy based on different SVM configurations (both linear and polynomial kernel functions have been used) using different NLP features is achieved by the linear SVM and the feature $\tau = 0.4$ and threshold Probability Boundary = 0.4. Moreover, among all the features mentioned in previous sections, the combination of unigram, UMLS semantic type and drug category extracted from the drug reviews structure for contextual filtering or trigger terms features achieves the best results.

Recognition and analysis of side effects: In this section, we examine the result of detecting only adverse drug side effects in drug reviews. In this case, the accuracy of the SVM model is higher than the rule-based scheme. The results achieved by the rule-based algorithm for side effect recognition are presented in Table 2.

For only side effect detection, the result significantly improves, which means that SVM is more powerful for detecting only side effects rather than differentiating it and disease symptoms. By only considering the side effect class, it improved F1 from 0.45 to 0.75 in strict mode and from 0.56 to 0.72 in lenient mode, as shown in Table 3. Moreover, including disease symptom annotations (annotated by expert) can improve the performance; using this feature increases the accuracy level (1: Unigram, 2: UMLS Semantic Types, 3: POS Category, 6: Disease Symptom and 8: Drug Category).

Evaluation metrics: Precision, recall and F-measure are three metrics applied to evaluate the results of our proposed system. They are calculated based on the following equations:

$$Precision = \frac{Correct + \frac{1}{2}Partial}{Correct + Spurious + Partial}$$

$$Recall = \frac{Correct + \frac{1}{2}Partial}{Correct + Missing + Partial}$$

$$F - measure = \frac{(\beta^2 + 1)P * R}{(\beta^2 P) + R}$$

Evaluation results: In this study, many experiments were done for the detection of side effects. The final comparison among all the techniques can be seen in Table 4 and 5. The tables prove that when SVM is used, precision, recall and F1 are improved almost two-fold. However, due to the drawbacks of the rule-based method, which were discussed in the previous section, this approach has low performance. In contrast, SVM deals better with this problem by using the lexical, semantic and contextual features.

By analysis of these tables, we can see that recall is much lower than the precision in the side effect recognition, which indicates that most of the side effects are missed by the algorithm (Table 1 and 2). Moreover, the low precision and recall of disease symptom recognition shows the high number of false positives and missing samples.

In this study, we developed two new techniques to recognize the side effects in the drug reviews, which can be used as a subtask in a medical opinion mining system. The results of both algorithms are encouraging and a good foundation for future research. Obviating the limitations and exploiting the combined approaches would improve the results.

CONCLUSION

In this study, we developed a SVM and rule-based algorithm in terms of drug side effect extraction as implicit opinions of drug reviews, which is a novel approach and can be used for further investigation in medical opinion mining systems. The performance of two techniques, rule-based and SVM, by using the precision, recall and F-measure, proved that drug review side effect extraction and discriminating

between drug adverse side effects and disease symptoms can be performed using the SVM algorithm, which significantly outperformed the regular expression based (rule-based) algorithm. We also constructed the annotated medical corpus, which is required in the evaluation phase, by using the knowledge and expertise of a pharmacist. In terms of using features for discriminating between disease symptoms and the adverse side effect unigram, UMLS semantic type, drug category and trigger terms achieved the best results and, in terms of adverse side effects recognition, unigram, UMLS semantic type, POS category, disease symptom tags and drug category are the best feature set. However, the result of applying the regular expression based approach proved that this method produces many false negatives in determining disease symptoms and drug adverse side effects, which reduces the accuracy of the prediction.

RECOMMENDATIONS

As mentioned before, side effect detection is considered to be implicit opinion detection, for which limited works have been undertaken. Therefore, as future work, we plan to enlarge the corpus to include hundreds of drug reviews of diverse drug categories to have better analysis and evaluation results by using more than one medication expert annotator. The accuracy of the proposed rule-based approach was low, so we will try to generate more complicated regular expression rules and define a more sophisticated scope for trigger terms by considering the syntactic features. Moreover, we can improve the SVM feature selection algorithm to find the best SVM feature set and increase the accuracy. Finally, we can exploit other machine learning techniques and select the best one.

Conflict of interest: This study is supported by the Ministry of Education Malaysia and Soft Computing Research Group (SCRG) of University Teknologi Malaysia (UTM). This Study is also supported in part by grant from Vote 4F373.

Appendix: A
Annotation guideline: The purpose of this study is to determine all drug adverse side effects and disease symptoms in the comment section of some drug reviews considering the patient's opinion. As an example, in this drug review "anxiety attacks" and "anxiety" are disease symptoms.
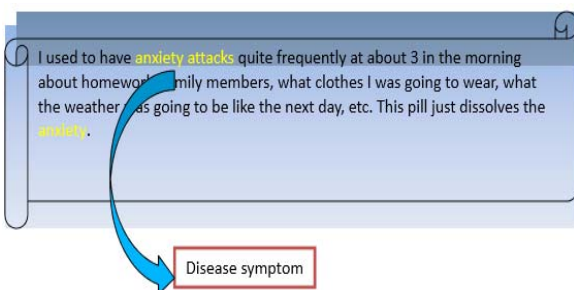
So, the annotator (medication expert) should highlight all of drug adverse side effects in "COMMENTS" column of file With RED colour and conversely make the disease symptoms GREEN.

Appendix: B
Side effect and disease symptom trigger terms:

	Trigger Term for symptom
	Coverage
	Medication
	Medications
	Med
	Meds
	Dr
	dr.
	Doctor
	Specialist
	Drug
	Pill
	Pills
	Aid
	Reliever
	Free
RE1	Management
	Syndrome
	Trigger Term For drug adverse side effect
	Causes
	Causes
	Was cause
	Caused
	It Caused
	Make
	Makes
	Made me
	Made you
	Makes me
	Make some people
	Makes you feel
	Makes me feel
	Gave me
	Can have
RE1	Can have serious consequences such as
	Down side to
	The down side
	Side effect
	Side effects
	Side effects involving
	Side effects such as
	Any side effect just
	Any side effect (just
	Any side effect only
	Any side effect (only
	No side side effect except
	Very little side effect other than
	Without any negative side effects other than
	The main side side effect was
	The only negative is that
	Only other side effect is
	Complaint
	My only complaint is
	Trigger Term For Symptom
	I have been
	I have had
	When I' am having
	When I am having
	When I got
	I have
	I too have
	I used to have
	I've had
	I have been suffering with
	You have the drug to
	You have a lot of
	You have

Appendix: B

RE1	Management of
	Healing of
	Diagnosis of
	Assisted with
	Helped with
	Helps with
	Helps
	For
	Since
	Because
	To stop
	To combat
	To fight
	To cure
	Cure
	Cured
	To treat
	Treated
	I was treated for
	To control
	Dissolve
	Dissolves
	Dissolved
	Clear up
	Slows down
	Effective in
	Effective for
	Works great for
	Works wonder for
	Works great on
	Works wonder(s) for
	Relieving
	Self-treating my
	With
	Bye bye
	Situation that previously would have caused
	Trigger Term For drug adverse side effect
	Nature
RE2	After injection
	After injections
	Through the first… injection
	Through the first injections
	Through the first injection

REFERENCES

Aronson, A.R., 2001. Effective mapping of biomedical text to the UMLS metathesaurus: The MetaMap program. Proceeding of the AMIA Symposium.

Cao, H., G. Hripcsak and M. Markatou, 2007. A statistical methodology for analyzing co-occurrence data from a large sample. J. Biomed. Inform., 40(3): 343-352.

Chapman, W.W., D. Chu and J.N. Dowling, 2007. ConText: An algorithm for identifying contextual features from clinical text. Proceeding of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing. Association for Computational Linguistics, Stroudsburg, PA, USA, pp: 81-88.

Cohen, A.M. and W.R. Hersh, 2005. A survey of current work in biomedical text mining. Brief. Bioinform., 6(1): 57-71.

Cunningham, S.J., M. Brady-Van den Bos and D.J. Turk, 2011. Exploring the effects of ownership and choice on self-memory biases. Memory, 19(5): 449-461.

Denecke, K. and W. Nejdl, 2009. How valuable is medical social media data? Content analysis of the medical web. Inform. Sciences, 179(12): 1870-1880.

Goeuriot, L., J.C. Na, W.Y.M. Kyaing, S. Foo, C. Khoo, Y.L. Theng and Y.K. Chang, 2011. Textual and informational characteristics of health-related social media content: A study of drug review forums. Proceeding of the Asia-Pacific Conference on Library and Information Education and Practice, pp: 548-557.

Gooch, P. and A. Roudsari, 2011. A tool for enhancing MetaMap performance when annotating clinical guideline documents with UMLS concepts. Proceeding of the IDAMAP Workshop at 13th Conference on Artificial Intelligence in Medicine (AIME'11). Bled, Slovenia.

Katsahian, S., E. Simond Moreau, D. Leprovost, J. Lardon, C. Bousquet, G. Kerdelhué, R. Abdellaoui, N. Texier, A. Burgun, A. Boussadi and C. Faviez, 2015. Evaluation of internet social networks using net scoring tool: A case study in adverse drug reaction mining. Stud. Health Technol. Inform., 210: 526-530.

Li, Y.A., 2011. Medical data mining: Improving information accessibility using online patient drug reviews. Ph.D. Thesis, Massachusetts Institute of Technology.

Liu, B., 2007. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Springer, Berlin, New York.

Niu, Y., X. Zhu, J. Li and G. Hirst, 2005. Analysis of polarity information in medical text. Proceeding of the AMIA Annual Symposium, pp: 570-574.

Noferesti, S. and M. Shamsfard, 2015. Resource construction and evaluation for indirect opinion mining of drug reviews. PLOS One, 10(5): e0124993.

Rajagopalan, S., 2011. Sideffective-system to mine patient reviews: Side effect extraction. Rutgers, The State University of New Jersey.

Simpson, M.S. and D. Demner-Fushman, 2012. Biomedical Text Mining: A Survey of Recent Progress. In: Aggarwal, C.C. and C. Zhai (Eds.), Mining Text Data. Springer Science+Business Media, LLC, pp: 465-517.

Skentzos, S., M. Shubina, J. Plutzky and A. Turchin, 2011. Structured vs. unstructured: Factors affecting adverse drug reaction documentation in an EMR repository. Proceeding of the AMIA Annual Symposium. pp: 1270-1279.

Swaminathan, R., A. Sharma and H. Yang, 2010. Opinion mining for biomedical text data: Feature space design and feature selection. Proceeding of the 9th International Workshop on Data Mining in Bioinformatics (BIOKDD'2010).

Wang, X., H. Chase, M. Markatou, G. Hripcsak and C. Friedman, 2010. Selecting information in electronic health records for knowledge acquisition. J. Biomed. Inform., 43(4): 595-601.

Weeber, M., H. Klein, A.R. Aronson, J.G. Mork, L.T. de Jong-van den Berg and R. Vos, 2000. Text-based discovery in biomedicine: The architecture of the DAD-system. Proceeding of the AMIA Symposium, pp: 903-907.

Yalamanchi, D., 2011. Sideffective-system to mine patient reviews: Sentiment analysis. Rutgers University-Graduate School-New Brunswick.

Zhang, L. and B. Liu, 2011. Extracting resource terms for sentiment analysis. Proceeding of the 5th International Joint Conference on Natural Language Processing. Chiang Mai, Thailand, pp: 1171-1179.

Zweigenbaum, P., D. Demner-Fushman, H. Yu and K.B. Cohen, 2007. Frontiers of biomedical text mining: Current progress. Brief. Bioinform., 8(5): 358-375.