Received 11 August 2015
Revised 19 April 2016
Accepted 12 May 2016

Recognition of side effects as implicit-opinion words in drug reviews

Monireh Ebrahimi, Amir Hossein Yazdavar and Naomie Salim
Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru, Malaysia, and
Safaa Eltyeb
College of Computer Science and Information Technology, Sudan University of Science and Technology, Khartoum, Malaysia and Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

Abstract

Purpose – Many opinion-mining systems and tools have been developed to provide users with the attitudes of people toward entities and their attributes or the overall polarities of documents. In addition, side effects are one of the critical measures used to evaluate a patient's opinion for a particular drug. However, side effect recognition is a challenging task, since side effects coincide with disease symptoms lexically and syntactically. The purpose of this paper is to extract drug side effects from drug reviews as an integral implicit-opinion words.

Design/methodology/approach – This paper proposes a detection algorithm to a medical-opinion-mining system using rule-based and support vector machines (SVM) algorithms. A corpus from 225 drug reviews was manually annotated by a medical expert for training and testing.

Findings – The results show that SVM significantly outperforms a rule-based algorithm. However, the results of both algorithms are encouraging and a good foundation for future research. Obviating the limitations and exploiting combined approaches would improve the results.

Practical implications – An automatic extraction for adverse drug effects information from online text can help regulatory authorities in rapid information screening and extraction instead of manual inspection and contributes to the acceleration of medical decision support and safety alert generation.

Originality/value – The results of this study can help database curators in compiling adverse drug effects databases and researchers to digest the huge amount of textual online information which is growing rapidly.

Keywords SVM, Drug review, Drug side effect, Medical-opinion mining, Regular expression, Rule based

Paper type Research paper

1. Introduction

The explosive growth of social media on the World Wide Web dramatically changes people's methods of expressing their opinions, and, consequently, how they make decisions in their lives. The medical domain is not immune of this fact. With the vast amount of medical online information and the rapid growth of social media in this field, people no longer take a drug before going to the internet to learn something about it. However, extracting and analyzing opinions manually from the huge volume of texts is a formidable or even impossible task. The aim of the automatic sentiment analysis

This work was partially supported by the Fundamental Research Grant Scheme (FRGS) funded by the Malaysian government. The authors hereby thank Dr Abolfath Ebrahimi and Dr Narjes Ebrahimi for taking their time to make the gold standard available.

Online Information Review
Vol. 40 No. 7, 2016
pp. 1018-1032

1468-4527
DOI 10.1108/OIR-06-2015-0208

(opinion mining) system is to provide the user with the attitudes of people toward entities and their attributes or the overall polarity of a document. Opinions can be classified into two groups according to the type of presentation in the sentences (Liu and Zhang, 2013). An explicit opinion is one that is expressed explicitly in a subjective sentence, e.g., "Methadone works great for me." On the other hand, an implicit opinion is one that is implied in an objective sentence. These sentences usually describe desirable or undesirable facts, e.g., "Methadone makes me feel nauseous." Unlike implicit-opinion mining, a significant amount of research has been done on explicit-opinion mining.

One issue that remains almost unexplored in medical-opinion mining is detecting implicit-sentiment words and phrases that show desirable or undesirable facts about one medicine in the drug review. One group of these words and phrases are drug side effects that cover a large portion of drug reviews. Side effects can imply both positive and negative opinions about one drug. Even so, talking about drug side effects is rarely positive, and positive terms are more related to drug effectiveness. On the other hand, most drug reviews are narratives and contain patients' experiences. Thus, just detecting these kinds of words individually and without considering the context cannot be an appropriate solution to this problem. For instance, some sentences show the general status of patients or their symptoms before taking a drug. As an example, in the sentence "I take alprazolam when I'm having anxiety about whatever it may be," "anxiety" is a disease symptom, and the reason for taking this psychoactive drug by the patient. Indeed, in this sentence the anxiety does not imply any opinion about the related drug. In contrast, in the sentence "Two months into taking this pill, I started having severe anxiety and anxiety attacks," which is part of a review on a birth control drug, anxiety is a drug side effect and is used to show the negative opinion of the patient about the drug's side effects. Therefore, despite the nonexistence of an opinion word, this sentence is opinionated, and a drug side effect should be considered as a negative implicit-opinion word in this context.

This study tries to resolve this problem by considering the context in which the medical concept occurs to differentiate between a disease-manifestation-related symptom and an adverse drug event to extract the second one as a sub-problem in medical-opinion mining.

To address this issue, we propose two approaches. The first is an unsupervised text-engineering technique, and the second is a supervised machine-learning method. In the first approach, we adapt the idea of using regular expressions for identifying contextual features from the clinical text suggested by Chapman *et al.* (2007). The second method is the machine-learning approach using support vector machines (SVM). Lastly, we compare the results of two approaches from the precision and recall point of view.

The paper is organized as follows. Section 2 presents the previous related work and discusses the research motivation. Section 3 presents the method applied and the experiments carried out. The results obtained are reported and discussed in Section 4 and Section 5. In Section 6 the conclusion and proposals for future work are expounded.

2. Related work

There are a great many social media sites on which people can post their opinions, experiences, and knowledge about drugs. These websites can be dedicated to drugs or cover different kinds of product. In a related study, Goeuriot *et al.* (2011), in their content analysis of drug reviews in three drug review websites, investigated

user-generated posts based on the number of opinion words, frequency of medical concepts, and linguistic features, such as length of review, sentence length, and proportion of different part of speech (POS) tags. Using the subjectivity lexicon, they observed that 8 percent of words are sentiment words (43 percent positive, 57 percent negative), whereas 64 percent are medical concepts. They also concluded from their observations on linguistic features, that drug reviews are more similar to spoken language than research papers, although they are both full of medical terms. Their analysis also shows that drugs related to depression, anxiety, weight loss, and pain relief are most frequently reviewed, and the most frequent negative-opinion terms are as follows: pain, depression, agony, and anxiety.

Because of these observations, it seems that the reason for the frequency of such opinion words is the frequency of the related drug. Thus, the drug has a treatment relationship with most of these words, and we should assign these kinds of words to the neutral class. However, these terms can be regarded as popular side effects of many drugs, and, thus, may be assigned to the negative category. Therefore determining sentiment word orientation based on opinion lexicon without considering the context is definitely not applicable to drug reviews.

Thus, drug reviews are full of adverse effect of drugs that are used by people to express their negative opinion toward a drug or its side effect aspect. However, people also speak about their disease symptom to describe their precondition. In this regard, the symptoms which are the arguments of "drug cause" relation should be detected precisely.

However, sentiment analysis is highly sensitive to the domain, this sensitivity is the result of different words and even language constructs which is exploited in different domain for giving opinion (Liu and Zhang, 2013). In addition, the same expression can show different sentiment orientation in different domain. Despite the importance of these systems specially in analyzing the drug reviews a little study has been done so far. SideEffective is the name of a system developed by Yalamanchi (2011) to search and analyze patient reviews about one drug in the drug reviews and ranking them based on the negativity. To do so, he proposed a novel approach which exploits a thesaurus service and training to reach an almost complete opinion lexicon. Finally, he determine the negativity of each review using the aggregation of sentiments of features, sentences and finally the review using his built opinion lexicon. Although this work take side effect into consideration but it does not have any mechanism to discriminate symptom from side effect or recognize patient's precondition. In addition he did sentiment analysis in review level (document level). So his work suffers from shortcomings which are inherent in document-level analysis. Indeed, this approach makes this assumption that the whole review is about one drug whereas in one drug review people may compare some drugs with each other and even speak about their experience about another similar drug or before taking the drug.

Like other extractions in the biomedical domain, this problem can take the advantage of more sophisticated statistical, rule based, knowledge based and particularly temporal extraction approaches. Next sub section provides an overview of previous work.

To resolve this issue, Wang *et al.* (2010) propose a combination of statistical and NLP based approaches. They use co-occurrence to detect two disease symptom and drug-potential adverse effects in electronic health records (EHR). To overcome the limitations of statistical methods, they use the structural feature of EHR and determine the type of relations based on the section where it occurs. Although they show the improvement of performance using filtering by section, this method is not

applicable to unstructured narrative drug reviews. In another statistical based by Li (2011) extracts side effect of statin and non-statin cholesterol lowering drug by considering the difference of patient precondition and drug side effect. They filter patient precondition by removing the symptoms which occur in both statin and non-statin drug reviews. However, this work is not extendable to our work because of two reasons. First, the idea behind this method is comparison of two drug type reviews of one disease. Second, their approach filters the common side effects of two types of drugs along with preconditions.

Rule-based approaches are the other flexible solutions to this challenging issue. By using this method, we can define the linguistic patterns to detect relation types precisely. Because of the flexibility and accuracy of rule-based methods, in this work we will adapt the proposed algorithm (ConText) by Chapman *et al.* (2007) to detect drug adverse reactions in drug reviews. A detailed description of our scheme is given in Section 6.

To sum up, knowing the polarity of medical text can play a pivotal role in decision-making, question answering, pharmacovigilance, and other things. However, existing general-opinion-mining systems are not sophisticated enough for such a domain because of its special characteristics. The next section describes the proposed methods and the materials used.

3. Materials and methods

As mentioned above, the main contribution of this work is extracting drug side effects more precisely as implicit opinion, taking their differences with disease symptoms into account by considering the context in some drug reviews of www.drugratingz.com. We apply a rule-based algorithm, where regular expressions and a list of trigger terms are used to detect drug adverse side effects and discriminate them from disease symptoms. In a later step, a combination of lexical, syntactical, contextual, and semantic features leads to the best results in the SVM technique.

3.1 Collecting drug reviews

In this study, 225 drug reviews are randomly selected from www.drugratingz.com for manual annotation. These reviews are related to such diverse categories as pain relief and antidepressant drugs. Indeed, the comment sections of drug reviews in this websites are full of sentences that contain drug side effects, and the role of our algorithm is to identify these side effects correctly. In our proposed rule-based algorithm, we use 70 reviews to generate rules manually and another 155 reviews as a test set. A five-fold cross-validation method is used to evaluate the SVM model.

From the data distribution view, the corpus has 342 symptoms of disease and 372 side effects of drugs that are distributed randomly in five corpus folds. Figure 1 presents the distribution of symptoms and adverse effects in our corpus.

In addition, a list of drug categories covered in this study and the number of drug reviews for each category are shown in Figure 2. It is worthwhile to recall that drugs related to depression, anxiety, weight loss, and pain relief are most frequently reviewed (Goeuriot *et al.*, 2011). To some extent, the drug category distribution over our corpus matches this observation.

3.2 Data pre-processing

Some pre-processing algorithms, including tokenization, sentence splitting, and POS tagging, should be run on the corpus before developing the following phases. To achieve this, some of the GATE processing resources are applied. ANNIE English

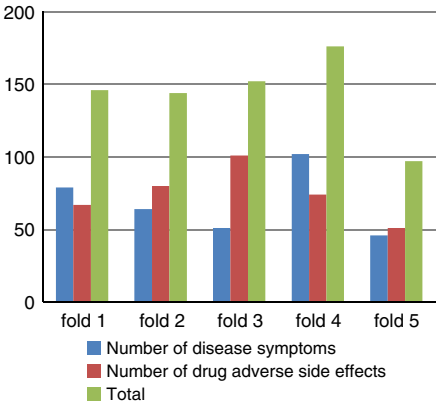

Figure 1.
Disease symptom and drug side effect distribution over the corpus

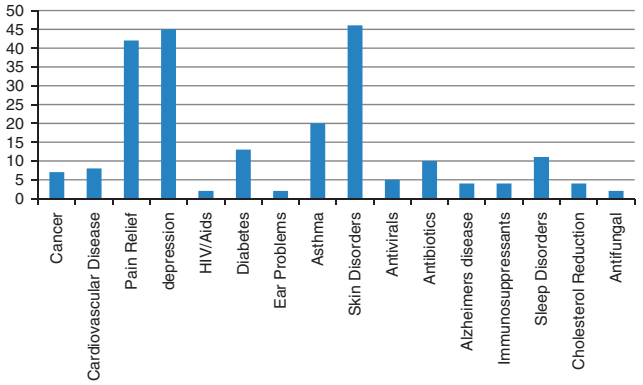

Figure 2.
Drug category distribution over the corpus

Tokenizer is used to split the drug review text into simple tokens, such as words, numbers, and punctuation. The results of this process are exploited by a sentence splitter and are required in medical concept mapping, rule based (for writing the left-hand side of the regular expression), and SVM algorithm (unigram feature).

Similarly, the RegEx Sentence Splitter is a regular expression-based splitter in GATE that identifies end-sentence boundaries. We have made some slight changes to regular expressions of this processing resource to make its output appropriate for our problem. The result of this algorithm is used by our proposed rule-based algorithm to identify the scope of trigger terms and is also required in medical concepts mapping by the Tagger_MetaMap plug-in. A part of speech tagging is done by the use of ANNIE POS Tagger. POS tags are added as category features on the annotations of type tokens. These results are used later as a feature by the SVM algorithm. Figure 3 displays these processes.

3.3 Term extraction and mapping to a medical concept

The next phase of this work is assigning words and phrases in the drug reviews to their corresponding semantic types in the Unified Medical Language System (UMLS) Metathesaurus. Most of the words used in medical texts to indicate side effects and symptoms fall within the Disorder semantic group of UMLS. Figure 4 shows an example of drug review sentences that are tagged with UMLS. These semantic tags are

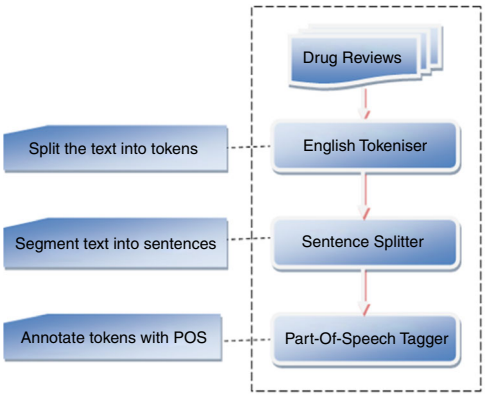

Figure 3.
Text segmentation processes

used by the SVM algorithm as part of the learning features. However, only disorder group tags are exploited as inputs to a proposed rule-based algorithm.

Among the existing programs, MetaMap is an effective configurable and open source program for indexing medical concepts in texts. The Tagger_MetaMap plugin (Gooch and Roudsari, 2011) for GATE wraps the MetaMap Java API client to allow the content of specified annotations to be processed by MetaMap and the results converted to GATE annotations and features.

3.4 Developing rule-based algorithms

The goal of this method is the discrimination of drug side effects from symptoms in the output of the previous phase, using the combination of some simple regular expressions and semantic rules. The idea supporting this scheme is the lexical and syntactic similarity of symptoms and side effects and the necessity of considering context to cope with this problem.

To reach this goal, we examine some drug reviews to generate some rules to indicate the role of the context on determining the side effects and symptoms. We use two regular expressions for detecting the symptoms and side effects:

$$\text{RE1}: \ < \text{trigger term} > \ < W^* > \ < \text{indexed term} >$$

$$\text{RE2}: \ < \text{indexed term} > \ < W^* > \ < \text{trigger term} >,$$

where W^* is any number of single words or UMLS concepts.

Indeed, the default scope of each trigger term is the end of the sentence unless another scope is defined explicitly for it. Our algorithm utilizes a manual list of trigger terms for recognition. Tables I and II list samples of the trigger terms used in this study.

Based on our observations, in drug reviews, when a medical term that belongs to the "Disorder" semantic group of UMLS appears after such terms as "because" and "for," it typically shows the disease symptom. We call these cue words, which can be used for discrimination: trigger terms. In this sentence, for example, "This drug is great for anxiety," anxiety will be assigned to disease symptoms using RE1, considering "for" as a disease symptom trigger.

In fact, the proposed algorithm is a combination of the lexicon-based (trigger terms extraction) and rule-based algorithm. To implement this approach, two GATE processing

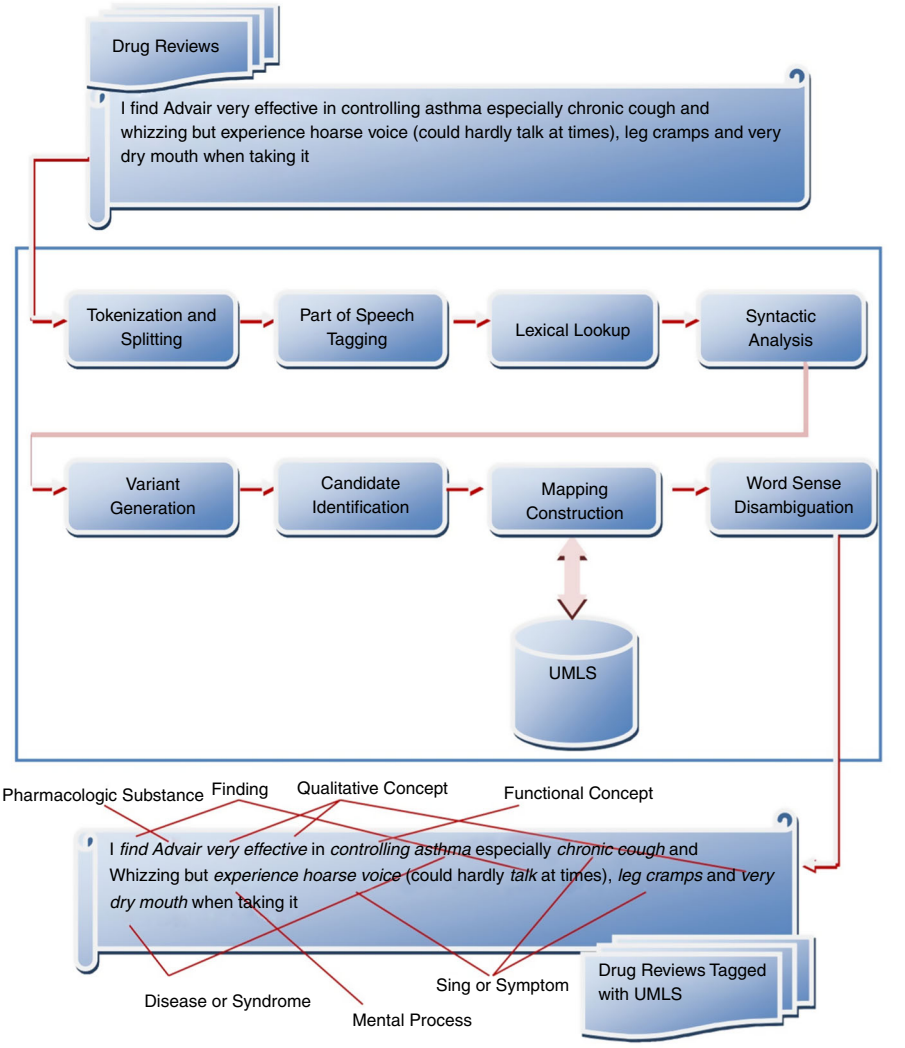

Figure 4.
Example of drug review tagged with UMLS using MetaMap

resources, namely, ANNIE Gazetteer and JAPE Transducer, have been, respectively, used. The Gazetteer processing resource uses the list of trigger terms, which is called the Gazetteer list, to look up trigger terms in the corpus. In this case, we define four Gazetteer lists to find the side effect and symptom indicators in the first and second regular expressions. In the rule-based part, the JAPE Transducer processing resource is used to detect the side effects and symptoms based on patterns in annotations using RE1 and RE2. Figures 5 and 6 show more details of this algorithm.

3.5 Developing SVM algorithm

SVM has been exploited previously in many applications to perform NER and text classification tasks. In this work, SVM has been used to accomplish side effect

Trigger term for symptom	Sample selected sentence from drug reviews	Detected symptom
RE1		
I have been	I have been suffering with headaches for over 10 years	Headache
For	This drug is a wonder for chronic back pain	Back pain
I have had	I have had anxiety since I was a teenager	Anxiety
Diagnosis of	I now have the diagnosis of bi-polar	Bi-polar
helps with	It helps with the pain	Pain
Because	Started Prozac mainly because of irritability	Irritability
To combat	I have been taking 60mg of Fluoxetine daily for 15 years to combat depression and anxiety and winter blues	Depression, anxiety, winter blues
To treat	I have been through a lot of various meds to treat my bi-polar symptoms	Bi-polar symptoms
Cure	This drug literally cured my depression	Depression
I have	I have an anxiety sleep disorder	Anxiety sleep disorder
RE2		
Medication	Methadone is the very best pain medication	Pain
Doctor	He got Methadone from his pain doctor	Pain
Drug	The best pain drug I have ever used	Pain
Pill	It does not impair your activities in the way that other narcotic pain pills do, so this is my answer for migraines!!!	Pain
Free	Takes about 2 hours and I'm back to pain free	Pain
Management	how I could have made it without the pain management	Pain

Table I. Sample of trigger terms for symptoms

Trigger term for side effect	Sample selected sentence from drug reviews	Detected side effect
RE1		
Cause	It could cause allergic reactions in some individuals	Allergic reactions
Make	Makes me sleepy though, but I can function on it	Sleepy
Side effect	As for side effects I had a small bit of the jitters the first week	Jitter

Table II. Sample of trigger terms for side effects

recognition (label = side effect) and their discrimination from disease symptoms (label 1 = side effect; label 2 = disease symptom) in drug reviews.

First, we should extract features that can differentiate between the classes. Intuitively, disease symptoms and drug adverse side effects are common in most of their lexical, syntactical, and semantic features. Thus, consideration of these features for recognition and additional contextual features for discrimination seems crucial to achieve good results. To verify this supposition, unigrams, POS tags, UMLS semantic types, disease symptoms and side effect trigger terms (identified in the previous algorithm), and drug categories are selected as features (see Table VI). In addition, manually annotated disease symptoms are used as a feature for side effect recognition tasks. Also, the context window is defined for some of the features. The extraction procedure is run on the corpus, using GATE plug-ins and JAPE rules.

To perform the learning, SVMLibSvmJava of the GATE learning plug-in is applied. Linear and polynomial kernel functions and different values of uneven-margin parameters are used when building an SVM model.

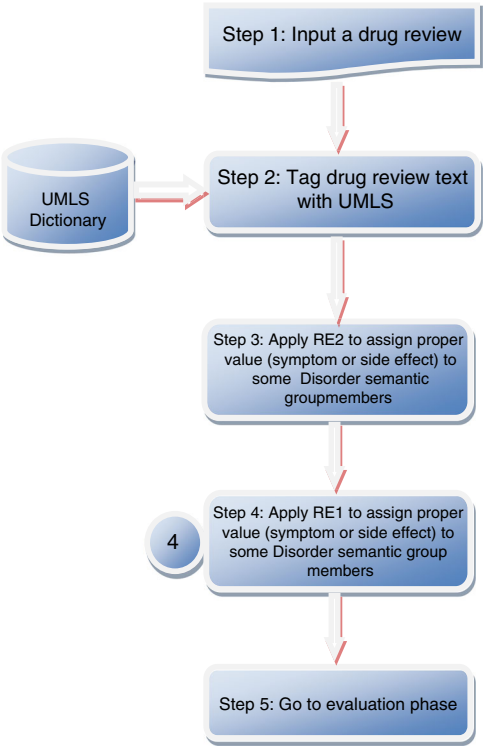

Figure 5.
Rule-based algorithm

Lastly, the proposed learning model is evaluated by using a five-fold cross-validation. The identification of the features combination, which leads to a good generalizable result, is of great importance to reach the goal of this study.

4. Results

Because of the novel nature of this problem, there is no annotated corpus for testing and evaluating this work. Thus, a group of pharmacists participated in this research by identifying all the disease symptoms and drug side effects in every drug review comment sentence in the corpus. These tagged values are used to calculate the precision, recall, and *F*-measure.

To calculate these three values, we compared the output tags, which are assigned by our algorithms with the corresponding gold standard values, according to three criteria: namely, strict, lenient, and average measures. Partially correct responses are considered as spurious and correct in strict and lenient measures, respectively. Likewise, the average measure allocates a half-weight to partially correct responses.

The proposed rule-based approach is a simple regular expression-based algorithm that uses the prepared list of trigger terms as contextual parameters to decide whether a disorder term recognized by MetaMap is a disease symptom or a drug adverse side effect. The results of this method are summarized in Table III by presenting the precision, recall, and *F*-measure of this method in strict, lenient, and average modes (Table IV).

SVM is the second approach to dealing with the disease symptom and drug side effect extraction problem, which is proposed, developed, and evaluated in this study.

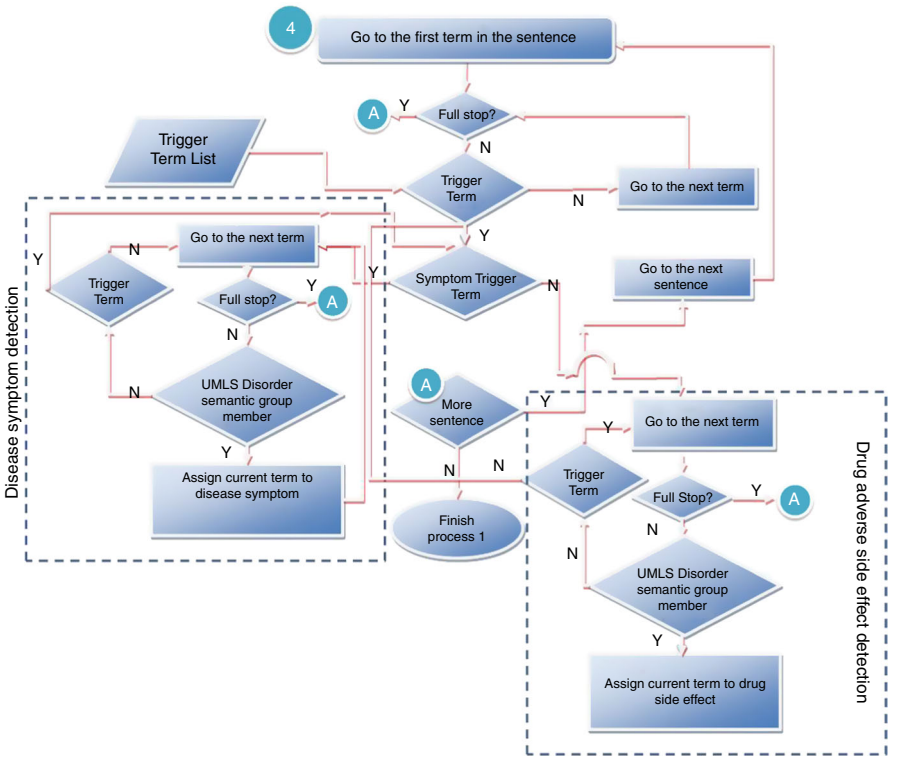

Figure 6.
Rule-based algorithm (applying RE1)

Class label	Strict			Lenient			Average		
	P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
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

Table III.
Results from rule-based technique

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

Table IV.
Confusion matrix for rule-based algorithm results

Upon examining the results obtained in this study, Table V lists the performance of different SVM configurations using different NLP features that are listed in Table VI. For each feature, we consider a context window to capture the features of some immediately preceding and succeeding tokens of the current tokens as their features.

From the configuration perspective, linear and polynomial kernel functions are used. The value of the uneven margins parameter for the SVM is shown by τ. If the training data have a small number and a large number of positive and negative examples, respectively, setting the value of τ to a value less than 1 (the value of τ for standard

Table V.
Summary of SVM results for different feature sets and configurations

Feature set	Class label	Strict			Lenient			SVM configuration		
		P	*R*	*F*1	*P*	*R*	*F*1	Kernel	τ	P B
1	Overall	0.65	0.23	0.34	0.78	0.27	0.40	L	0.4	0.4
	Disease symptom	0.66	0.26	0.36	0.78	0.30	0.43			
	Side effect	0.62	0.20	0.30	0.76	0.25	0.37			
1+2	Overall	0.56	0.47	0.51	0.66	0.55	0.60	L	0.4	0.4
	Disease symptom	0.61	0.58	0.58	0.68	0.65	0.65			
	Side effect	0.51	0.38	0.43	0.65	0.48	0.55			
1+2+4	Overall	*0.55*	*0.48*	*0.51*	*0.65*	*0.57*	*0.61*	*L*	*0.4*	*0.4*
	Disease symptom	*0.60*	*0.57*	*0.57*	*0.68*	*0.64*	*0.65*			
	Side effect	*0.50*	*0.42*	*0.45*	*0.63*	*0.52*	*0.56*			
1+2+8	Overall	*0.57*	*0.47*	*0.51*	*0.68*	*0.56*	*0.61*	*L*	*0.4*	*0.4*
	Disease symptom	*0.62*	*0.58*	*0.59*	*0.71*	*0.66*	*0.67*			
	Side effect	*0.51*	*0.38*	*0.43*	*0.65*	*0.48*	*0.55*			
1+2+4	Overall	0.61	0.29	0.39	0.70	0.33	0.45	*P* ($d=3$)	0.4	0.4
	Disease symptom	0.69	0.34	0.45	0.75	0.37	0.49			
	Side effect	0.53	0.25	0.34	0.66	0.31	0.42			
1+2+4	Overall	0.50	0.32	0.39	0.60	0.39	0.47	L	0.2	0.4
	Disease symptom	0.57	0.36	0.43	0.64	0.40	0.49			
	Side effect	0.43	0.28	0.34	0.57	0.37	0.44			
1+2+4+7	Overall	0.61	0.42	0.49	0.68	0.48	0.55	L	0.4	0.4
	Disease symptom	0.63	0.54	0.54	0.67	0.59	0.59			
	Side effect	0.62	0.32	0.41	0.70	0.38	0.48			
1+2+4+8	Overall	0.61	0.41	0.48	0.74	0.49	0.58	L	0.4	0.4
	Disease Symptom	0.66	0.45	0.52	0.74	0.50	0.59			
	Side effect	0.58	0.38	0.46	0.73	0.48	0.58			
2	Overall	0.48	0.25	0.32	0.56	0.29	0.38	L	0.4	0.2
	Disease symptom	0.45	0.34	0.35	0.55	0.40	0.41			
	Side effect	0.51	0.20	0.28	0.61	0.23	0.33			

SVM) typically leads to a better result (Cunningham *et al.*, 2011). BP in the last column of Table V is a threshold probability boundary that determines the confidence threshold on the start and end tokens of chunk. Only those boundary tokens with a confidence of more than this value are selected as candidates for the entities. As indicated in Table V, linear SVM with $\tau = 0.4$ and threshold probability boundary $= 0.4$ obtained the best results.

From the SVM feature set selection perspective, Table V also provides different combinations of features and corresponding results. The first model shows the results of considering only a unigram for solving the problem (baseline). As the second model shows, considering UMLS semantic types increases the precision and recall significantly. Using contextual features in the third and fourth models slightly increases performance. Nevertheless, using these contextual features together in the same model counter intuitively slightly degrades the performance.

In conclusion, a combination of novel feature sets including unigram as a lexical feature, UMLS semantic type as a semantic feature by using the domain knowledge, and, finally, a drug category that is extracted from a drug review structure for contextual filtering or considering trigger terms leads to the best result.

In this study, many experiments were conducted to solve the problem. The overall results can be summarized in Table VII. The table indicates, the results achieved by SVM are more promising. The next section discusses the proposed models presented on previous sections in terms of the results, limitation and source of errors.

5. Discussion

In this study, many experiments were conducted to extract the adverse side effects of drugs as implicit-opinion words from drug reviews to enhance the usability of

Table VI. List of NLP features used to build different SVM models

Number	Feature name	Feature extraction tool	Context window From	Context window To
1	Unigram	GATE ANNIE English Tokeniser PR	−1	+1
2	UMLS semantic types	GATE Tagger_MetaMap Plugin	−1	+1
3	POS category	GATE ANNIE POS Tagger	−1	+1
4	Side effect or disease symptom trigger terms	GATE ANNIE Gazetteer	−1	+1
5	Side effect or disease symptom trigger terms	GATE ANNIE Gazetteer	−5	+2
6	Disease symptom	Expert Annotation	−1	+1
7	Drug category	GATE JAPE Transducer	−1	+1
8	Drug category	GATE JAPE Transducer	0	0

Table VII. Comparison between SVM and rule-based algorithms for side effect and disease symptom discrimination

Method	Class label	Strict *P*	Strict *R*	Strict *F*1	Lenient *P*	Lenient *R*	Lenient *F*1	Average *P*	Average *R*	Average *F*1
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
SVM	Overall	0.57	0.47	0.51	0.68	0.56	0.61	0.63	0.52	*0.56*
	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

opinion-mining systems in the biomedical domain. The overall results for them can be concluded as in Table VII. As the tables indicate, experimental results show that SVM outperforms the rule-based algorithm significantly.This conclusion is more severe for side effect recognition task that when SVM is used, precision, recall and *F*1 are improved almost two-fold.

As shown in Table III and Table V, among the precision and recall evaluation measures, recall is much lower than precision in side effect recognition. This indicates that most side effects are missed by the proposed rule-based algorithm, as shown in Table IV. In addition, the low precision and recall of disease symptom recognition shows the high numbers of false positives and false negatives that are evident in this table.

A further manual analysis of results clears the challenges and reasons behind the low performance of the proposed rule-based algorithm. The main reasons include the sensitivity of proposed rule-based algorithm to MetaMap performance, using descriptive sentences instead of a phrase to describe disease symptom and side effects in some cases by patients, existence of some non-disorder phrases which are mapped to the UMLS Disorder semantic group and vice versa, some typographical error; and other factors as illustrated in Table VIII. Another problem is the small number of analyzed drug reviews for constructing the list of trigger terms, which leads to low performance. In contrast, SVM deals better with this problem using lexical, semantic and contextual features.

Based on the literature, using machine-learning methods in information extraction may provide easier portability solutions, even though rule-based information extraction systems provide more-reliable results (Spasić *et al.*, 2010). Surprisingly, our experimental results show that the SVM algorithm provides a more-reliable and portable solution to this problem. The main reason to support this phenomenon is the small size of the development data set, which leads to the construction of a not-comprehensive small list of trigger terms. In addition, in some cases more complicated rules should be considered instead of just simple regular expression-based methods. The other important reason is related to the strong dependency of a rule-based algorithm in an extraction of disorders to MetaMap outputs, as discussed earlier. However, the proposed SVM algorithm deals better with the small data set using a five-fold cross-validation, and its dependency on the MetaMap result is much less. Indeed, SVM only uses MetaMap UMLS semantic types as a feature, and the extraction task will be accomplished by SVM chunk learning using the whole feature set, including MetaMap.

Despite the high performance of SVM in handling problems with high-dimensional data, such as information extraction, this algorithm is highly sensitive to selected features. To handle this issue, the SVM model is evaluated using some combination of lexical, syntactic, and contextual features. The results show that the combination of unigram and UMLS semantic type with one of the contextual features (drug category or trigger terms lists) will lead to a better result.

However, the results of both algorithms are encouraging and a good foundation for future research. Obviating the limitations and using combined approaches would improve the results.

6. Conclusion

In this paper, we have investigated the utility of the regular expression and the machine learning in to deal with side effect recognition as implicit-opinion words in

Review	Category	Side effect	Disease symptom	MetaMap Semantic type	Preferred name	Reason of system failure
My acne is GONE	Skin disorders	–	False negative	Dsyn	Acne vulgaris	Small number of analyzed drug reviews for constructing the trigger terms list/small trigger terms list
I get a little dry lip but I use lip balm and I'm fine	Skin disorders	False positive	–	Dsyn	Little's disease	Sensitivity of the algorithm to MetaMap performance/existence of some non-disorder phrases which are mapped to the UMLS disorder semantic group
I get a little dry lip but I use lip balm and I'm fine	Skin disorders	True positive	–	Fndg	On examination – dry lips	–
I have depression with anxiety, and have been on this medication	Depression/ anxiety disorders	–	True positive	Mobd	Mental depression	–
I have depression with anxiety, and have been on this medication	Depression/ anxiety disorders	–	True positive	Mobd	Anxiety disorders	–
WellbutrinXL was the only one that left me with a clear head, and got me out of the pits of depression	Depression/ anxiety disorders	–	False negative	–	–	Sensitivity of the algorithm to MetaMap performance/ typographical error
I do have to take an occasional xanax or two, for anxiety	Depression/ anxiety disorders	–	False negative	–	–	Sensitivity of the algorithm to MetaMap performance/ typographical error
It stunts your growth and it weakens your immune system	Asthma	False negative	–	–	–	Using descriptive sentences to talk about side effect
It stunts your growth and it weakens your immune system	Asthma	False negative	–	–	–	Using descriptive sentences to talk about side effect

Table VIII. Samples of manual error analyses

drug reviews. The system's value in the real world is represented by its ability to detect new cases or modify the existing statistics of adverse drug effect.

In consideration of the experiments carried out, our main finding is that drug review side effect recognition can be handled by using the SVM algorithm, which significantly outperforms the regular expression-based algorithm. The best SVM feature set to discriminate disease symptoms from adverse side effects among examined features include unigram, UMLS semantic type, and drug category or trigger terms.

The results of the proposed algorithm are encouraging and a good foundation for future research. However, enlarging the corpus to have hundreds of drug reviews of diverse drug categories and also enlarging the development data set for a rule-based algorithm to construct a more-comprehensive list of trigger terms, generating more

complicated hand-crafted rules and defining a more sophisticated scope for trigger terms, considering the syntactic features, and using a sophisticated SVM feature selection algorithm for finding the best SVM feature set would improve the results. However, the result analysis highlights the potential of proposed algorithm to reach much better accuracy in future.

References

Chapman, W.W., Chu, D. and Dowling, J.N. (2007), "Context: an algorithm for identifying contextual features from clinical text", *Proceedings of the Workshop on Bionlp 2007: Biological, Translational, and Clinical Language Processing, Association for Computational Linguistics*, pp. 81-88.

Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N., Roberts, I., Gorrell, G., Funk, A., Roberts, A. and Damljanovic, D. (2011), "Developing language processing components with gate version 6 (a user guide)", available at: http://gate.Ac.Uk/sale/tao/tao.Pdf (accessed May 12, 2015).

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

Gooch, P. and Roudsari, A. (2011), "A tool for enhancing metamap performance when annotating clinical guideline documents with umls concepts".

Li, Y.A. (2011), *Medical Data Mining: Improving Information Accessibility Using Online Patient Drug Reviews*, Massachusetts Institute of Technology, Cambridge, MA.

Liu, B. and Zhang, L. (2013), "A survey of opinion mining and sentiment analysis", *Mining Text Data*, Springer, New York, NY, pp. 415-463.

Spasić, I., Sarafraz, F., Keane, J.A. and Nenadić, G. (2010), "Medication information extraction with linguistic pattern matching and semantic rules", *Journal of The American Medical Informatics Association*, Vol. 17 No. 5, pp. 532-535.

Wang, X., Chase, H., Markatou, M., Hripcsak, G. and Friedman, C. (2010), "Selecting information in electronic health records for knowledge acquisition", *Journal of Biomedical Informatics*, Vol. 43 No. 4, pp. 595-601.

Yalamanchi, D. (2011), *Sideffective-System to Mine Patient Reviews: Sentiment Analysis*, Rutgers university-graduate school-new brunswick.

Corresponding author
Naomie Salim can be contacted at: drnaomiesalim@gmail.com