Natural Language Analysis of Online Health Forums

Abul Hasan, Mark Levene and David J. Weston

Department of Computer Science and Information Systems
Birkbeck, University of London
London WC1E 7HX, U.K.
abulhasan,mlevene,dweston@dcs.bbk.ac.uk

Abstract. Despite advances in concept extraction from free text, finding meaningful health related information from online patient forums still poses a significant challenge. Here we demonstrate how structured information can be extracted from posts found in such online health related forums by forming relationships between a drug/treatment and a symptom or side effect, including the polarity/sentiment of the patient. In particular, a rule-based natural language processing (NLP) system is deployed, where information in sentences is linked together though anaphora resolution. Our NLP relationship extraction system provides a strong baseline, achieving an F_1 score of over 80% in discovering the said relationships that are present in the posts we analysed.

Keywords: Natural Language Processing, Health related forums, Rule-based system

1 Introduction

Health related posts in medical forums often contain factual information regarding drug usage, the effectiveness of a drug when used to treat a symptom, and the experience of any adverse effects from the drug. In order to extract this type of information from free text in medical forums, natural language analysis of the text is required [5]. As has been pointed out by Karimi et al. [8], detecting, from unstructured text, the disease, treatment and symptom entities, their attributes and existing relationships, is a major research issue in the Natural Language Processing (NLP) domain. Although progress has been made in extracting entities, there is still the challenge of extracting specific relationships between these entities [7]. In particular, here we are interested in extracting relationships of the form (treatment, polarity/sentiment, symptoms/side effects), which represent relationships that provide us with information on the effectiveness or otherwise of various treatments, especially medication. When aggregated over many forum posts, such triples could inform practitioners and/or patients on the effectiveness of treatments beyond the information gathered from studies published in medical journals and by the pharmaceutical companies.

We report on an initial proof of concept of extracting such triples from about 1000 posts related to Parkinsons disease from the PatientsLikeMe website [14], providing details of the algorithm we deployed and the results from a comprehensive evaluation of the algorithm. It is important to note that patients' comments in a forum, such as the one we are analysing, will contain slang and verbose, informal, descriptions of treatments and side effects (for example, using "body shaking" instead of the more formal "tremor"). Such informal terms are not normally present in standard ontologies such as the *Unified Medical Language System* (UMLS) [1]. As a result of this difficulty, much of the previous research in this area has focused on extracting formal medical terms from the free text. For example, Gupta et al. [7] attempted to extract drugs and treatments (DTs), and symptoms and conditions (SCs) terms present in the forum text. On the other hand, Nikfarajam et al. [12] and Sampathkumar et al. [16] extracted adverse drug reactions (ADRs) from Twitter and the DailyStrength forum [4]. Their research shows that social media is a valuable source for finding information related to drugs, symptoms and side effects.

We now give a brief summary of our NLP relationship extraction system, whose aim is to build triples, which can be aggregated to provide useful statistical medical information relating to the effectiveness of various treatments. Our model makes use of the following *concepts*:

1. Drugs (X), or more generally, treatments.
2. Symptoms (Y), which the drug is meant to treat.
3. Side effects (Z), which are caused by the drugs.
4. Polarity (P), which indicates how positive a treatment is or how negative a side effect is from the patient's point of view.

The system first extracts different health related concepts from the forum posts, and then creates structured information by forming a relationship between a drug and a symptom or side effect, through polarity analysis of the text. We termed such a relationship formally as a *disease triple* henceforth simply a *triple*.

Despite machine learning techniques such as *conditional random fields* (CRFs)[10] being very effective in NLP information extraction, we have chosen to build a *rule-based system* [10]. The reasons for this choice are:

1. To the best of our knowledge, this is the first attempt to extract, from social media, relationships in the form of disease triples, which include patient sentiment. There is no baseline system for such work and in order to attain deep knowledge of the use of natural language, specialised rules are often needed.
2. Dictionaries, lexicons and ontologies in the medical domain are built for extraction tasks from documents written by experts [7]. However, patients are, in most cases, not familiar with this terminology, so they tend to use commonly understood terms. As a result matching to such pre-built dictionaries often results in poor performance. In order to build common domain knowledge, it is first necessary to manually analyse a significant number of posts, and extend publicly available lexicons and gazetteers using a specifically designed set of generalised rules for extracting structured information from the free text.

3. Once the baseline is established it is possible to export the set of designed rules and resulting extended gazetteers to be used in a more sophisticated machine learning technique such as word2vec [11], to attain transferability and scalability when extracting triples from other forums.

Our overall contribution is the summarisation of health related forum posts by identifying relationship between concepts in the form of disease triples to provide a coherent structure, which can be used to extract meaningful medical statistical information.

1.1 A motivating example

Let us examine the following example post to motivate the research:
"I take 600mg of gabapentin at bedtime, helps me shake and kick less; and a donepezil 10mg, settles me down allowing sleep. Clonazapam works great but I can't take the groggy, foggy head the next day."

After various pre-processing steps and the application of linguistic rules, we create the disease triples as shown in Table 1. Most of the existing work concentrates on processing a single sentence. However, our method is a formal approach that uses information from the whole post, which may contain several sentences, by making use of *anaphoric relations* [6] present in the sentences. At this stage we do not identify dosages or temporal information present in the text; these will be tackled in future work.

Table 1: +, -, symp, side, drug , list, con and intens, denote positive polarity, negative polarity, symptom, side effects, drug/treatment, list of nouns, conjunction and intensifier, respectively.

Sentence or Sentence segment	**Disease triple**
I take 600mg of gabapentin$_{\text{drug}}$ at bedtime, helps$_{+}$ me shake$_{\text{symp}}$ and$_{\text{list}}$ kick$_{\text{symp}}$ less$_{-}$;	(gabapentin,+,shake) (gabapentin,+,kick)
and$_{\text{con}}$ a donepezil$_{\text{drug}}$ 10mg, settles$_{+}$ me down allowing sleep$_{\text{symp}}$.	(donepezil,+,sleep)
Clonazapam$_{\text{drug}}$ works$_{+}$ great$_{\text{intens}}$	(Clonazapam,+,?)
but$_{\text{con}}$ I can't take$_{-}$ the groggy$_{\text{side}}$, foggy head$_{\text{side}}$ the next day.	(Clonazapam$_{\text{drug anaphora}}$,-,groggy) (Clonazapam$_{\text{drug anaphora}}$,-,foggy head)

The rest of the paper is organised as follows. Section 2 describes prior research related to our work. Section 3 discusses our text processing architecture and describes the details of our algorithm. The data set and its annotation are discussed in Section 4. In Section 5 we provide our findings and results from the experiment on the posts from PatientsLikeMe. Finally, in Section 6 we give our concluding remarks and discuss future work.

2 Related Work

A comprehensive review of NLP systems used for text processing in the medical domain can be found in [8]. Prior research in finding useful health related information from social media, mainly focused on *named entity extraction* (NER) tasks such as discovering ADRs in relation to a drug or treatment as reported in [8]. For example, Nikfarjam et al. [12] built a system, called ADRMine, using CRFs for recognising ADR mentions from social media. Features such as contextual, lexical and semantic parts-of-speech (POS) tags features were added to an existing CRF classifier. ADRMine also added word embedding features, created from word2vec [11], trained on Twitter and the DailyStrength [4] corpora. Moreover, Nikfarjam et al. [12] added clusters of words formed from word2vec to their supervised model as an additional feature. In a recent paper, Korkontzelos et al. [9] analysed the effect of sentiment analysis features in ADR classification, which made use of rules such as "*negation*" to improve the performance of their system. Dai et al. [3] also investigated features to use for finding ADR in Twitter posts.

On the other hand, Sampathkumar et al. [16] mined ADRs from health forums using a supervised *hidden Markov model* (HMM) [10]. The HMM provides statistical structure for the forum messages, where drug and side effects keywords representing the causal relation between the drug, side effects and other words, were encoded as hidden states. Concepts were extracted the from messages using existing medical lexicons. The model was trained with the positive samples of ADRs, and learnt the association between drugs and side effects through the presence of keywords. After training the forum messages, hidden states offered predictions from the preprocessed observed messages. The authors conducted various experiments by varying different components of the system. One of their findings was that the F-score of the supervised classification model is significantly lowered as the size of the dictionaries is reduced.

Comparing the results from the above models, it seems that the model using HMMs achieved better performance than that using CRFs. The reason for CRFs not performing as well, can be attributed partially to the pre-built dictionaries used for labelling the text, containing noisy data [7]. However, the lack of a common data set in this domain makes the judgement in comparing different models a difficult task.

Closely related to our work, is a semi-supervised algorithm deployed by Gupta et al. [7], where they extracted SCs and DTs from social media using lexico-syntactic patterns. At first, they labelled the concepts using dictionaries constructed from publicly available sources. Then, flexible patterns were created by looking at two to four words before and after the labelled tokens. Patterns were scored by a frequency measure, i.e. the top-k most occurred patterns were chosen. They applied these patterns to all the sentences and extracted the matched phrases. These learned phrases were added to the dictionary, and the process was repeated until convergence. This method resulted in an improvement of the F_1-score by approximately 5% over the baseline lexicon-based approach. The authors also reported on the discovery of new DTs and SCs terms that were not present in the seed dictionaries. Also closely related is the recent work of Pain et al.[13]. They investigated classification methods for identifying drug/effect relations and their corresponding polarities for tweets.

We note that extracting entities from individual sentences in a post, is an important step towards analysing natural language in medical forums. However, linking the sentences in a post would allow a more comprehensive analysis of the text. Disambiguation of semantic relations between two expressions (sentences in our case) is known as *anaphora resolution*, where a later expression (the *anaphor*) has some semantic relation to an earlier expression (the *antecedent*). The rule-based system described in [6] is a knowledge-centric and pattern-based approach for disambiguating anaphoric references in clinical records. For our work, we took a slightly different approach by considering protagonist theory [6], which suggests that narrative events are centered on one or more key actors. Our analysis of the posts reveals that a drug mention found in a sentence is often referred to in subsequent sentences. Thus, in the case of drug mentions, we seek to find the antecedents of anaphoric expressions.

3 Methodology

The overall methodology is schematically shown in Figure 1 and the corresponding pseudo-code of the algorithm is presented in Algorithm 1 below. In the following subsections we describe this methodology in more detail.

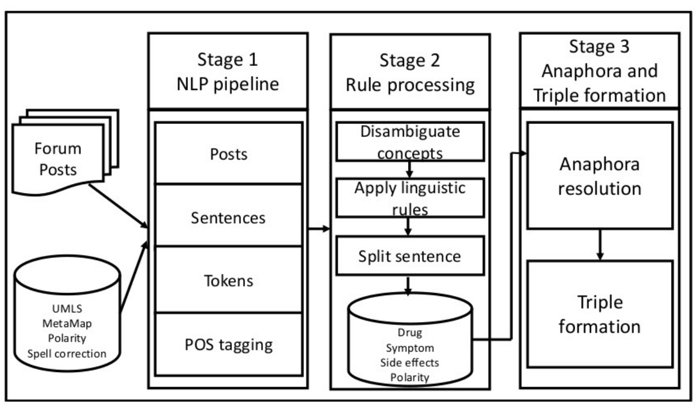

Fig. 1: Text processing architecture

3.1 NLP pipeline

The text processing pipeline, constructed using GATE [2], splits the posts into sentences, tokenises the text and labels the tokens with their POS tags. At this stage, our system recognises drugs, symptoms, side effects and opinions present in the text by using different lexicons and plug-ins to construct the text processing pipeline. Apart from using some publicly available resources, we constructed gazetteers using domain knowledge and extended these by augmenting the terms during the training phase of the system.

3.2 Rule processing

We disambiguate multiple concepts recognised in the previous stage by applying different linguistic rules, then split the sentences and compute the polarity score of each resulting segment.

Disambiguation of concepts We extracted four types of concepts from each sentence, i.e. drugs, symptoms, side effects and polarities. We now briefly discuss the concept extraction and disambiguation process.

- **Drug extraction.** We used a subset of the RxNorm terminology [17] as our drug gazetteer, where the vocabulary consisted of drugs and treatments used for Parkinson's disease; RxNorm is a normalised naming system for generic and branded drugs. For our purpose, it is very important to distinguish between the mentions of generic and branded drugs to avoid extracting the same type of drug mention multiple times. As such we have constructed two dictionaries for generic and branded drugs. A feature is added to the drug token according to the drug gazetteer it is extracted from. If a sentence contains both prescription and generic drug mentions, the generic drug mention is subsumed. Spelling mistakes in drug names are corrected using a normalised edit distance of two based on *Levenshtein's edit distance* [2].
- **Symptom extraction.** We used the MetaMap [1] annotation plug-in to annotate symptoms (sign or symptom semantic types) in sentences. We also constructed a separate symptom gazetteer using domain knowledge, and terms such as "voice", "smell" and "restless" were also added to the gazetteer. By default the polarity feature of symptoms recognised by the MetaMap program are set to negative. The symptom gazetteer contains explicit polarity for each symptom. Not all the symptoms present in the gazetteer are negative, for example concepts such as "sleep" and "energy" are marked as positive. The polarity feature is extracted from the gazetteer for each symptom.
- **Side effect extraction.** For side effects, a gazetteer using COSTART (Common Standard Thesaurus of Adverse Reaction Terms) [17] is constructed. We extended the dictionary by adding new terms during the training phase.
- **Polarity extraction.** We have collected the polarity terms from the MPQA Subjectivity Lexicon [18], which contains more than 8000 words annotated manually by the authors as positive, negative or neutral. The lexicon also includes the POS information for the terms. Symptom and side effect terms present in the dictionary are not labelled as polarity terms. The prior polarity score for positive and negative terms are set to +1 and -1 respectively, and neutral words are given as score 0. Our system matches a token with the lexicon if the POS information present in the lexicon is the same as the POS category of the token in the sentence.

Linguistic rules. Matching polarity words is not enough in order to extract opinion from a sentence, and it often produces wrong result. A description of how valence of a lexical item is modified by the presence of different lexical items such as "negation", "modifiers" and "presuppositionals" can be found in [15]. Similar to [15] we applied following heuristic linguistic rules:

- **Negation.** If a negation word such as "not" and "don't" precedes in 0 to 3 tokens of a polarity, symptom or side effect concept, then the polarity of each concept is reversed and will generate the feature `negation_concept` (for example, "doesn't_work").
- **Modifiers.** Modifiers such as intensifying adverbs (for example, 'very", "strongly"), diminishing adverbs (for example, 'little", "kind of") increases or decreases the sentiment value of a concept and generate a feature `modifier_concept` (for example, "very_tired"). This rule modifies polarity of the first matched concept at a distance of 0 to 3 tokens.
- **Presuppositional.** The polarity of presuppositional items such as "barely" is multiplied with that of the concepts and polarity is flipped as a result. A feature such as `pre_concept` (for example, "barely_tremor") is generated after applying this rule. As for modifiers, this rule also changes the polarity of the first matched concept at a distance of 0 to 3 tokens.

Split sentence. In this step, we first split a sentence and then calculate the final polarity score. These two steps are described below:

- **Sentence segmentation.** We used common conjunctions (*"and"*, *"but"*) and *"until"* to segment a sentence. Our analysis revealed that *"and"* is sometimes used to connect two or more words to form a list rather than connecting two different parts of a sentence. We constructed a rule to find such conjunctions. For example if POS tags of the two tokens in either end of *"and"* are same, then it is denoted as a list of tokens and we do not segment sentence in such cases.
- **Final polarity score.** We add the polarity scores of all the opinion concepts in a sentence or segment of a sentence. The polarity of a triple is positive if total score is more than 0, negative if it is less than 0. If the score is 0, then polarity of the triple will be that of the symptom or side effect.

3.3 Triple formation

At this stage, we first perform anaphora resolution and then form triples. These two steps are described in the following two paragraphs.

Anaphora resolution. Messages in a forum contain sentences referring to the concepts mentioned earlier in the text. Our rule for finding anaphoric references for drugs is: if the current sentence has a drug mention, then the drug is carried forward to the next sentence in the text. Using this rule, if a triple has a drug mention and a subsequent triple contains the default drug concept ("?"), then we

```
Data: P is a list of Posts
foreach p in P do
    Split p into a list of sentences, S;
    foreach s in S do
        Tokenise s into a list of tokens, T;
        foreach t in T do
            Append the POS information;
            Identify the concept class, C matching with the gazetteers;
            Let, C = {X, Y, Z, P} where X, Y, Z, P are drug, symptom, side
              effect, and polarity, respectively;
            Disambiguate concepts t;
            Calculate the polarity, p by applying linguistic rules for t in P;
        end
        Split s into a list of segments, G using conjunctions, and, but, until;
        foreach g in G do
            Compute polarity score, SC;
            Create a list of triples, L;
            A triple is either (X, Y, SC) or (X, Z, SC);
            Where, "?" is the placeholder for a missing concept;
            foreach l in L do
                Perform anaphora resolution for missing X;
            end
        end
    end
end
```

Algorithm 1: Text processing algorithm

replace the default with the drug found, and repeat the same process for all the sentences in text until we find a new drug mention. If we find multiple mentions of drugs then multiple triples are created containing each drug mention. We also look for the patterns such as "from X to Y", which indicates that the person actually stopped using drug "X" and moved on to using "Y". In such case, we add a feature to the drug ("X") token indicating that the algorithm should stop creating triple for this drug in subsequent sentences.

Disease triples. We create a list of concepts by ordering them according to their offset from the beginning of a sentence. Triples are formed using the following format:

1. *Triple 1*: Drug, Polarity, Symptom
2. *Triple 2*: Drug, Polarity, Side effect.

The algorithm iteratively finds drug, polarity and symptom or side effect concepts using the order shown in the formation of triple. A triple is formed by taking three consecutive concepts of a different kind. In our algorithm, consecutive concepts (for example, drugs) of same kind, signals the starting point of a new triple. The algorithm places a default concept which is "?" in case of missing concepts.

4 Experimental Setup and Evaluation

The following subsections describe the data we collected and the procedure we used for verifying our annotation of the unstructured text. We then describe our experiment and its results.

4.1 Data set

PatientsLikeMe [14] is an online health discussion forum, where patients with chronic health conditions can share their experiences living with disease. For our study, we extracted user comments from discussion threads related to Parkinson's disease. After registering with this website, pages were automatically scraped and the posts were anonymised by removing user IDs. A total of 1058 posts were collected from the period of April, 2016 to June, 2016.

500 posts were used for training and 400 for testing the system. The remaining posts were used for the annotation validation, described in the next section, where 58 posts were used to train the annotators and the remaining 100 posts for cross validation of the annotations.

4.2 Annotation validation

The annotation for the dataset was carried out by the first author. In order to verify the fidelity of these annotations an experiment was conducted using a small subset of the data, where the level of agreement between the annotator and other annotators was measured.

Ten researchers from Birkbeck's department of Computer Science and Information Systems volunteered for the validation experiment. Annotators were trained by showing annotated posts (20 posts were chosen from the annotator training set of 58) and explaining each concept and triple types. The remaining 100 posts were divided randomly into five sets of 20. Each of the ten annotators were randomly assigned two sets such that each set would get two annotations from different annotators. The agreement between the actual annotator and volunteers were calculated using both Cohen's kappa statistic [19] and accuracy. A very high level of agreement in recognising drug, symptom, positive and negative strings was achieved. However, agreement and accuracy in recognising triples and side effects were somewhat lower (72.09% and 76.18% respectively). It was subsequently determined that two of the annotators had not fully understood the task. Table 2 shows the results with and without these two 'outlier' annotators. It can be seen that there is a better agreement in identifying triples, however agreement in identifying side-effects has been reduced. This is due to the small number of side-effect concepts in the validation set (10 in total).

4.3 Evaluation

To evaluate our proposed approach the standard measures of accuracy, precision, recall and F_1 [19] were used. Each post was split into segmented sentences as described in Section 3.2. Triples formed from a segmented sentence are then

Table 2: Annotation validation result. Kappa-O and Accuracy-O are the results after discounting the 2 'outlier' annotators.

Concept	Kappa	Accuracy	Kappa-O	Accuracy-O
Drug and Treatment	87.37%	94.40%	86.84%	94.16%
Symptom	91.49%	96.90%	92.73%	97.32%
Side effects	76.18%	99.40%	71.05%	99.23%
Positive polarity	89.71%	96.26%	89.06%	95.97%
Negative polarity	90.72%	97.15%	89.77%	97.02%
Triples	72.09%	88.79%	76.18%	90.72%

merged with those from other segments and subsumed in case of repetition. A sentence can contain zero of more concepts and consequently zero or more disease triples, as shown in Table 3.

Table 3: Test set summary

Posts	Sentences	zero triples	one triple	more than one triple
400	2564	544	1447	573

5 Training and Test Results

Training of the system was conducted incrementally over five iterations. The training data was split into 5 sets of 100 posts each. At the beginning of the first iteration, we analysed the posts from the first set, i.e. annotated them with the concepts, generated rules and extended dictionaries, and then evaluated the system's output with the actual annotation. After we achieved a satisfactory performance, which meant a precision of approximately 80% or above, we moved on to the next iteration and followed the same procedure. The evaluation was carried out at each iteration by cumulatively adding a new set of posts to the posts from previous iterations.

It is interesting to see the overall performance from the training data, Table 4. This was achieved in a principled manner, as described in Section 3.2, involving the development of as few rules as possible. It should be noted that without anaphora resolution very few triples would have been successfully identified. In addition, dictionaries were extended when necessary. Though we are very successful in recognising concepts, the system makes a few mistakes in disambiguating polarity terms. As a result, the performances at triple level are lower, which resembles that of recognising positive and negative polarity terms. This result is in line with our hypothesis (see Section 1) that we can establish a relation between drug and symptom and drug and side effects through the polarity of a sentence. Although the polarity dictionary [18] has been extended by incorporating common phrases and is also supported by set of generalised rules, there is still room for improvement.

Table 4: Training and Test results

Dataset	**Training**				**Test**			
Concept	Accuracy	Precision	Recall	F_1	Accuracy	Precision	Recall	F_1
Drug	95.06%	99.63%	99.12%	97.29%	90.71%	88.29%	95.14%	91.59%
Symptom	95.71%	85.64%	99.06%	91.86	94.26%	84.08%	87.36%	85.69%
Side effects	99.06%	87.95%	98.50%	92.92%	98.42%	80.25%	93.53%	86.38%
+ polarity	90.19%	80.98%	96.15%	87.92%	86.44%	72.68%	94.42%	82.13%
- polarity	90.61%	79.60%	94.04%	86.22%	87.08%	73.57%	88.52%	80.35%
Triple 1	83.06%	81.01%	95.23%	87.54 %	73.93%	71.11%	96.02%	81.71%
Triple 2	84.76%	82.28 %	94.96%	88.16%	74.47%	71.31%	96.81%	82.13%

For testing the system was run over the remaining 400 post test dataset, without any modification to the system. The results are shown in Table 4. We can see that the system has generalised well. In general the disease triple identification has had the greatest fall in performance, since recognising these relationships is dependent on accurately identifying the concepts from which they are comprised.

6 Concluding Remarks

We have proposed summarising potential useful medical information in free-form unstructured text, with disease triples. We have developed a strong baseline system, achieving an F_1 score of over 80%, in identifying these disease triples using traditional NLP methods, and have demonstrated that this approach can generalise successfully. One current limitation of our system, which is left as future work, is that we are not yet recording useful temporal/quantative data such as dosages or frequency of recurrence of side effects.

Our next goal is to transfer the knowledge gained in our system to discover triples in other patient forums. In order to achieve this we anticipate the use of machine learning methods, such as word2vec [11], which can adapt to different usage of language than the PatientsLikeMe forum we have concentrated on.

References

1. Bodenreider, O.: The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Research 32(suppl 1), D267–D270 (2004)
2. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N., Roberts, I., Gorrell, G., Funk, A., Roberts, A., Damljanovic, D., et al.: Developing language processing components with gate version 6 (a user guide). University of Sheffield, Department of Computer Science (2011)
3. Dai, H.J., Touray, M., Jonnagaddala, J., Syed-Abdul, S.: Feature engineering for recognizing adverse drug reactions from twitter posts. Information 7(2), 27 (2016)
4. DailyStrength: `https://www.dailystrength.org/`, accessed: 2017-05-04

5. Denecke, K., Deng, Y.: Sentiment analysis in medical settings: New opportunities and challenges. Artificial Intelligence in Medicine 64(1), 17–27 (2015)
6. Gooch, P., Roudsari, A.: Lexical patterns, features and knowledge resources for coreference resolution in clinical notes. Journal of Biomedical Informatics 45(5), 901–912 (2012)
7. Gupta, S., MacLean, D.L., Heer, J., Manning, C.D.: Induced lexico-syntactic patterns improve information extraction from online medical forums. Journal of the American Medical Informatics Association 21(5), 902–909 (2014)
8. Karimi, S., Wang, C., Metke-Jimenez, A., Gaire, R., Paris, C.: Text and data mining techniques in Adverse Drug Reaction detection. ACM Computing Surveys (CSUR) 47(4), 56 (2015)
9. Korkontzelos, I., Nikfarjam, A., Shardlow, M., Sarker, A., Ananiadou, S., Gonzalez, G.H.: Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. Journal of Biomedical Informatics 62, 148–158 (2016)
10. Manning, C.D., Schütze, H., et al.: Foundations of Statistical Natural Language Processing, vol. 999. MIT Press (1999)
11. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems. pp. 3111–3119 (2013)
12. Nikfarjam, A., Sarker, A., O'Connor, K., Ginn, R., Gonzalez, G.: Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. Journal of the American Medical Informatics Association pp. 1–11 (2015)
13. Pain, J., Levacher, J., Quinqunel, A., Belz, A.: Analysis of twitter data for post-marketing surveillance in pharmacovigilance. In: Proceedings of the 2nd Workshop on Noisy User-generated Text. pp. 94–101 (2016)
14. PatientsLikeMe: `https://www.patientslikeme.com/`, accessed: 2017-04-21
15. Polanyi, L., Zaenen, A.: Contextual valence shifters. In: Computing attitude and affect in text: Theory and applications, pp. 1–10. Springer (2006)
16. Sampathkumar, H., Chen, X.w., Luo, B.: Mining Adverse Drug Reactions from online healthcare forums using Hidden Markov Model. BMC Medical Informatics and Decision Making 14(1), 91 (2014)
17. U.S. National Library of Medicine: `https://www.nlm.nih.gov/`, accessed: 2016-06-21
18. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP). pp. 347–354. Association for Computational Linguistics (2005)
19. Witten, I.H., Frank, E., Hall, M.A., Pal, C.J.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann (2005)