Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning

Felix Gräßer
Institut für Biomedizinische Technik
Technische Universität Dresden
felix.graesser@tu-dresden.de

Surya Kallumadi
Department of Computer Science
Kansas State University
surya@ksu.edu

Hagen Malberg
Institut für Biomedizinische Technik
Technische Universität Dresden
hagen.malberg@tu-dresden.de

Sebastian Zaunseder
Institut für Biomedizinische Technik
Technische Universität Dresden
sebastian.zaunseder@tu-dresden.de

ABSTRACT

Online review sites and opinion forums contain a wealth of information regarding user preferences and experiences over multiple product domains. This information can be leveraged to obtain valuable insights using data mining approaches such as sentiment analysis. In this work we examine online user reviews within the pharmaceutical field. Online user reviews in this domain contain information related to multiple aspects such as effectiveness of drugs and side effects, which make automatic analysis very interesting but also challenging. However, analyzing sentiments concerning the various aspects of drug reviews can provide valuable insights, help with decision making and improve monitoring public health by revealing collective experience.

In this preliminary work we perform multiple tasks over drug reviews with data obtained by crawling online pharmaceutical review sites. We first perform sentiment analysis to predict the sentiments concerning overall satisfaction, side effects and effectiveness of user reviews on specific drugs. To meet the challenge of lacking annotated data we further investigate the transferability of trained classification models among domains, i.e. conditions, and data sources. In this work we show that transfer learning approaches can be used to exploit similarities across domains and is a promising approach for cross-domain sentiment analysis.

KEYWORDS

Text classification; Sentiment Analysis; Clinical Decision Support System (CDSS); Health Recommender System

ACM Reference Format:

Felix Gräßer, Surya Kallumadi, Hagen Malberg, and Sebastian Zaunseder. 2018. Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning. In *DH'18: 2018 International Digital Health Conference, April 23–26, 2018, Lyon, France.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3194658.3194677

DH'18, April 23–26, 2018, Lyon, France

ACM ISBN 978-1-4503-6493-5/18/04...$15.00
https://doi.org/10.1145/3194658.3194677

1 INTRODUCTION

Pharmaceutical product safety currently depends on clinical trials and specific test protocols. Such studies are typically done under standardized conditions in a limited number of test subjects within a limited time span. As a consequence, the discrepancies in patient selection and treatment conditions can have significant impact on the effectiveness and potential risks of adverse drug reactions (ADRs). Therefore, post-marketing drug surveillance, i.e. pharmacovigilance, plays a major role concerning drug safety once a drug has been released. Furthermore, clinical decision support systems (CDSS) which provide assistance with diagnosis and treatment decisions are expected to play an increasingly important role in healthcare. Approaches such as therapy recommender systems, which aim at helping to find an optimal personalized therapy option for a given patient and time, benefit from feedback on therapy outcome [8]. These systems, however, typically rely on structured data, i.e. data categorized into a number of classes on predefined scales. The amount of such data often is limited because it requires intense preparation which is not standard in clinical routine. Here, other sources, such as user reviews, offer great potential.

However, one major requirement for automatic processing and analysis of the information contained in large amounts of unstructured information is the transformation of inherent aspects into numerical ratings. One typical way of doing so, in the context of product ratings, is sentiment analysis, which is an extensively studied domain in processing free-text in web media analyses [11]. Many approaches to sentiment analysis are based on sentiment lexicons. These approaches recognize sentiment terms and patterns of sentiment expressions in natural language texts by matching textual units with opinion words in lexicons annotated for sentiment polarity. However, studies showed that sentiment analysis is often domain-dependent since the polarity of single terms can differ depending on the context they are used in [4, 6]. Furthermore, the language in online forums is highly informal and user-expressed medical concepts are often nontechnical, descriptive, and challenging to extract. Which is why typical lexicons are of limited use for drug review analyses. An alternative approach treats the task as classification problem. Here, machine learning is used to train classifiers on domain-specific data sets to detect the polarity at sentence or document level. Such approaches have the additional advantage to be capable of performing medical sentiment analysis over multiple facets, i.e. sentiments can be learned on specific aspects such as effectiveness and side effects.

Sentiment analysis of patient data in general and on drug experience in particular is a challenging research problem that is currently receiving considerable attention. One of the main issues, however, is the lack of annotated data, which is crucial for accurate sentiment classification. Especially, labeled data dealing with distinct aspects is rare. Moreover, the availability of labeled data is highly domain dependent. Patients suffering from certain conditions are more active in reporting experience on their treatment than others.

Consequently, this work studies (1) the possibility to apply sentiment analysis on drug reviews, and the identification of effectiveness of a drug as well as the type of side effect caused by a drug exploiting its reviews. Therefore, classification of side effects and effectiveness is treated as an aspect-based sentiment analysis problem. Furthermore, to address challenges related to the limited data availability, we investigate (2) the transferability of the trained models among domains, i.e. conditions, as well as (3) across data sources.

2 BACKGROUND AND RELATED WORK

Literature on drug reviews and pharmacovigilance can basically be divided into studies on identification of aspects such as automatic detection of ADRs or side effects and such works dealing with overall or aspect-based sentiment analysis.

Most approaches tackling ADR or side effect identification are lexicon-based and rely on mapping relevant terms and phrases from user data to specific vocabulary from various individual or combined lexicons [6, 10]. However, lexicon-based approaches suffer from phonetic and typographic misspellings. Therefore, recent works have also focused on machine learning techniques to overcome such limitations. Nikfarjam et al. applied association rule mining to find pattern, i.e. combinations in terms [14], or conditional random fields (CRFs), to extract mentions of ADRs [15]. Based on the underlying assumption that patients' posts about ADRs typically express negative sentiments, Korkontzelos et al. studied the effect of enriching a lexicon-based ADR identification method with sentiment analysis features [9]. Cavalcanti et al. demonstrate the extraction and classification of multiple aspects in drug reviews, e.g. adverse reactions, efficacy of a drug, symptoms and conditions, using a method based on syntactic dependency paths [2]. An extensive review on pharmacovigilance and ADR extraction techniques can be found in [16].

Works on drug review sentiment analysis can basically be divided into approaches applying lexicons with sentiment scores or such approaches learning sentiments employing supervised classification. In one of the earliest works on drug review sentiment analysis Xia et al. developed a topic classifier from patient data to eventually apply several polarity classifiers, one per topic [17]. Na et al. demonstrate a clause-level sentiment analysis algorithm considering multiple review aspects as overall satisfaction, effectiveness, side effects and condition. Here, a rule-based approach is employed that takes grammatical relations and semantic annotation into account and computes sentiment orientation of individual clauses based on a lexicon [13]. In [12], aspect-based sentiment analysis of patient reviews is studied on oncological drugs. Here, opinion words are identified and overall sentiments derived utilizing a lexical resource. Gopalakrishnan et al. analyze patient drug satisfaction by using a supervised learning sentiment analysis approach. In this study three levels of polarity were classified comparing SVM with neural network based methods [7].

Many research studies have attempted to improve domain adaption or cross-domain sentiment classification, although not on drug review aspect-level but among various entities as products, movies or restaurants. In [1] a comprehensive systematic literature review on cross-domain sentiment analysis is presented.

3 DATASET

We used data from two independent webpages for retrieval of user reviews and ratings on drug experience. Drugs.com is, according to the provider, the largest and most widely visited pharmaceutical information website providing information for both, consumers and healthcare professionals. It provides user reviews on specific drugs along with related condition and a 10 star user rating reflecting overall user satisfaction. Similarly, Druglib.com is a resource on drug information for both, consumer and healthcare professionals. It comprises considerably fewer reviews but reviews and ratings are provided in a more structured way. Reviews are grouped into reports on the three aspects *benefits*, *side effects* and overall *comment*. Additionally, ratings are available concerning overall satisfaction analogously to Drugs.com as well as a 5 step side effect rating, ranging from *no side effects* to *extremely severe side effects* and a 5 step effectiveness rating ranging from *ineffective* to *very effective*.

We gathered user comments and ratings from both pages using an automatic web crawler. The data was scraped from raw HTML using the Beautiful Soup library in Python. Crawling these domains resulted in two data sets comprising 215063 reviews from Drugs.com and 3551 reviews from Druglib.com. Furthermore, we derived three level polarity labels for overall patient satisfaction and three level effectiveness and side effect scores using thresholds as specified in table 1. Both data sets were further split into training and test partitions according to a stratified random sampling scheme with the proportion of 75% and 25%, respectively. As shown in table 1, the total number of individual drugs in the Drugs.com data amounts to 6345 in comparison to the 541 drugs contained in the data derived from Druglib.com. However, the average number of reviews per drug is still considerably higher in the Drugs.com data (58.86) than in the Druglib.com data (7.66). The amount of unique conditions contained in the Druglib.com data, on the other hand, seems to exceed the number of the Drugs.com data. However, it is to be noted that conditions in the first platform are user created in contrast to Drugs.com where conditions are selected from a defined list, and thus standardized. Therefore, in case of Druglib.com, conditions are not normalized but comprise manifold variations in spelling, synonyms and combinations of conditions.

4 APPROACHES

In this section a description of the methods used in this work is detailed. The objective of this study was threefold:

Table 1: Data Description

Data	#Train	#Test	#conditions	#drugs	length	rating	label	%
Drugs.com								
Overall Rating	161297	53766	836	3654	458.32 (240.76)	$rating \leq 4$	-1	25
						$4 < rating < 7$	0	9
						$rating \geq 7$	1	66
Side Effects (Annotated)	-	400	141	243	500.385 (209.42)	No Side Effects	0	32
						Mild / Moderate Side Effects	1	28
						Severe / Extremely Severe Side Effects	2	40
Druglib.com								
Overall Rating	3107	1036	1808	541	277.57 (283.21)	$rating \leq 4$	-1	21
						$4 < rating < 7$	0	10
						$rating \geq 7$	1	69
Benefits (Effectiveness)	3107	1036	1808	541	212.87 (198.51)	Ineffective	0	8
						Marginally / Moderately Effective	1	19
						Considerably / Highly Effective	2	73
Side Effects	3107	1036	1808	541	177.36 (197.93)	No Side Effects	0	30
						Mild / Moderate Side Effects	1	53
						Severe / Extremely Severe Side Effects	2	17

(1) Prediction of the overall patients' satisfaction with applied medications and sentiments on side effects and effectiveness by employing classification-based sentiment analyses.
(2) Evaluating the transferability of models among medical domains, i.e conditions, by learning a model on data from one condition (source domain) to classify overall patient satisfaction in data from another condition (target domain).
(3) Evaluating the transferability of models across data sources, i.e Drugs.com and Druglib.com, by learning a model on reviews from one data source (source data) to classify overall patient satisfaction and sentiments on side effects in data from another source (target data).

Whereas for the first two tasks the ground truth is available for both data sets, distinct reviews covering the aspects side effect and effectiveness along with labels are only available for the Druglib.com data. To evaluate the transferability of side effect prediction models across data sets, 400 randomly picked samples from the Drugs.com data were manually labeled concerning side effects by two independent annotators. The inter-rater agreement measured with the Cohen's Kappa statistic [3] is 81.84% which is considered as very strong agreement. The annotators discussed all mismatching entities and agreed on a consensus.

Both approaches, sentiment analysis regarding overall patients' satisfaction and the aspect-based analysis of patients' sentiments on side effects and medication effectiveness were converted to classification problems. In case of overall patient satisfaction, the user ratings were converted to three disjoint classes representing the polarity of a patient's sentiment regarding the applied medication (negative, neutral, positive). In addition, also the severity of side effects and the level of effectiveness were transferred to three disjoint classes as described in table 1.

For all prediction tasks we apply a n-grams approach to represent the user reviews. That means both, single tokens, e.g. words, (unigrams) as well as two or more adjacent tokens (bigrams, trigrams), e.g. 2- or 3-word expressions, were used to derive features for classification. Based on the total collection of occurring n-grams, i.e. the corpus, each review can be represented as a sparse vector of token counts.

Initially, all reviews were preprocessed according to a standard scheme: Alphabetic characters were transferred to lowercase and special characters, punctuation and numbers were removed. Subsequently, the preprocessed documents were tokenized on spaces to obtain the overall vocabulary and a feature space representation of each review. No stop words were removed from the texts. However, to reduce the feature space, terms that have a relative document frequency higher than a given threshold were discarded when building the vocabulary.

Using the extracted feature representations, logistic regression was employed for building sentiment models for the various prediction tasks. Model hyperparameters were tuned using a 5-fold cross validation grid search on the respective training data, targeting the best Cohens's Kappa score. Optimized hyperparameter include n-gram number of adjacent tokens, token document frequency threshold, and logistic regression regularization strength. As shown in table 1, besides the annotated subset from the Drugs.com data, labels are considerably unbalanced. To compensate for this disproportionate distribution, classification errors were penalized with a weight inversely proportional to the class frequency during training. All experiments were evaluated by computing confusion matrices and deriving both, accuracy and Cohens's Kappa scores.

5 EXPERIMENTS AND RESULTS

5.1 In-domain Sentiment Analysis

In an initial experiment, overall performance when applying sentiment analysis to drug reviews was studied. Therefore, one model for each data set (Drugs.com and Druglib.com), to classify overall patient satisfaction reviews, is trained and evaluated utilizing the corresponding training and test data. Additionally, as in case of the Druglib.com data the *comments* section might only contain

Table 2: In-domain Sentiment Analysis

Aspect	Source	Acc. / Kappa
Overall Rating	Drugs.com	92.24 / 83.99
Overall Rating	Druglib.com	69.88 / 28.45
Overall Rating (all)	Druglib.com	75.19 / 43.59
Benefits (Effectiveness)	Druglib.com	77.70 / 44.13
Side Effects	Druglib.com	76.93 / 60.13

supplementary remarks, a combination of all three reports (*benefits*, *side effects* and *comments*) of a patient on a respective drug were concatenated to represent the overall patient satisfaction review.

Furthermore, we studied the expression of sentiments on the two aspects *side effects* and *effectiveness* within patient generated texts. Therefore, two logistic regression models were optimized and trained on the *benefits* and *side effects* training data derived from Druglib.com, respectively. Both, predicted *effectiveness* and *side effect* labels were compared against the actual labels obtained from the user ratings.

As detailed in table 2, overall patient satisfaction can be mined from patient texts with very high accuracy and Cohen's Kappa score in case of the Drugs.com data. The significantly worse performance reported for the Druglib.com data is assumed to have two main reasons. First, the data set is considerably smaller, which hampers the modelling. Moreover, the *comments* section is mainly used for supplementary information on personal experience and drug application and not explicitly for comments on satisfaction. When combining all three aspects, i.e. patient reports, classification performance could be improved over the previous result. In both approaches concerning the Druglib.com data the largest error contribution results from neutral ratings classified as positive which cannot be improved by data combination. The performance improvement, however, results from the reduction of misclassified negative ratings.

Sentiment analysis related to the specific aspects *effectiveness* and *side effects* shows promising results. Especially the *side effects* comments seem to provide valuable features that facilitate mining sentiments on side effects. Here, errors are mainly due to misclassification of neighbouring classes, namely excessive missclassification as *mild / moderate side effects*. In case of *effectiveness* classification the largest error contribution stems from *marginally / moderately effective* reviews classified as *considerably / highly effective*, whereas *considerably / highly effective* labeled reviews can be classified correctly with 95% accuracy. However, it must be kept in mind that also comments on *benefits* not necessarily relate to effectiveness only but may also encompass other aspects.

5.2 Cross-domain Sentiment Analysis

In this experiment we studied the performance of models built on data from one condition, i.e. the source domain, and evaluated on data related to other conditions, i.e. the target domain. To do so, overall patient satisfaction models were trained on drug review subsets related to one selected condition only. These domain models were then evaluated on other condition related subsets. Domains, i.e. subsets of particular conditions, were selected by extracting five of the most frequent disorders present in the Druglib.com data set from diverse medical fields. These are Contraception (38436), Depression (12164), Pain (8245), Anxiety (7812) and Diabetes, Type 2 (3362), with frequency in descending order. In-domain performances, i.e. training and testing of data from the same condition, are reported as averaged k-fold cross validation results (k=5).

The results summarized in table 3 demonstrate that the selected training domain has considerable impact on the classifier performance when applied to data from other domains. Especially, in-domain training and testing clearly outperforms all cross-domain setups. This finding clearly emphasizes the hypothesis of domain-specific vocabulary. For Contraception and Diabetes, even the overall rating classification using the entire data could be outperformed. However, the model trained on Depression data only seems to generalize better on the other domain data than e.g. a model trained on Diabetes data only. Furthermore, there are combinations showing better performances than others, e.g. Depression and Anxiety compared to Contraception and Anxiety, which is assumed to be due to underlying coherences of side effects or expressions and domain specific vocabulary used by patients. Moreover, the medical field dealing with Depression and Anxiety is closely related. From drugs concerning Depression (115) and Anxiety (81), 33 drugs are applied in both conditions whereas for Contraception (181) and Anxiety there is no overlap. Furthermore, the confusion matrices show that main classification errors occurred on neutrally labeled reviews for all domain combinations. Transferring the task to a binary classification problem without classification of neutral entities would result in substantially higher accuracy and Cohen's Kappa values.

5.3 Cross-data Sentiment Analysis

Finally, we study the transferability of the trained models among data sources. Overall patient satisfaction models were trained on both associated training data sets and evaluated on drug reviews from the other, independent data source test set. As discussed in 5.1, in case of the Druglib.com data a combination of all three reports (*benefits*, *side effects* and *comments*) were concatenated to represent the overall patient satisfaction review. Additionally, the performance of a classifier trained on *side effect* comments from the Druglib.com data is evaluated on the manually annotated data from Drugs.com.

Transferring a sentiment model trained on the significantly larger Drugs.com data to the Druglib.com data shows promising classification capabilities. Evaluating the model trained on the much smaller Druglib.com data with the Drugs.com data, however, doesn't perform satisfactorily. We assume such findings, on the one hand, to result from the limited training data size. On the other hand, differing data properties are likely to restrict the transferability. As stated previously, in contrast to the Druglib.com data Drugs.com reviews are highly unstructured covering multiple aspects in an entire review.

As summarized in table 4, applying the model trained on the *side effect* aspect to the Drugs.com reviews also performs poorly. The largest fraction of the classification error stems from reviews labeled as reporting *no* or *severe / extremely severe side effects* as *mild / moderate*. The features extracted from the Druglib.com data obviously don't contain sufficient discriminating power to classify the unstructured Drugs.com review which are not dealing with a

Table 3: Cross-domain Sentiment Analysis

		Train Data					
		Contraception	Depression	Pain	Anxiety	Diabetes, Type 2	avg. test
Test Data	Contraception	95.57 / 92.39	64.40 / 35.66	59.36 / 22.59	60.59 / 24.59	62.12 / 33.63	68.41 / 41.77
	Depression	62.05 / 31.51	90.13 / 78.07	75.21 / 40.69	77.07 / 43.95	66.98 / 33.93	74.29 / 45.63
	Pain	66.53 / 27.11	78.80 / 42.43	92.65 / 79.32	80.72 / 37.50	57.70 / 20.67	75.28 / 41.40
	Anxiety	64.35 / 28.14	82.64 / 51.22	79.74 / 43.43	92.37 / 78.41	67.51 / 30.64	77.32 / 46.37
	Diabetes, Type 2	69.90 / 44.50	71.83 / 43.37	68.17 / 32.32	69.48 / 34.18	94.74 / 89.84	74.82 / 48.84
	avg. train	71.68 / 44.73	77.56 / 50.15	75.03 / 43.67	76.05 / 43.73	69.81 / 41.74	

Table 4: Cross-data Sentiment Analysis

Aspect	Train Source	Test Source	Acc. / Kappa
Overall Rating	Drugs.com	Druglib.com	75.29 / 48.08
Overall Rating (all)	Druglib.com	Drugs.com	70.06 / 26.76
Side Effects	Druglib.com	Drugs.com	49.75 / 25.88

single aspects only. Utilizing a larger training data set, leading to less ambiguous features, might improve the results.

6 CONCLUSIONS

Within this preliminary work, we studied the application of machine learning based sentiment analysis of patient generated drug reviews. Logistic regression models were trained using simple lexical features such as unigrams, bigrams and trigrams extracted from the reviews. Besides patient satisfaction, sentiment aspects concerning effectiveness and experienced side effects were analyzed. Depending on aspect and data source, promising classification results could be obtained.

As labeled data sets for building classification models are rare or are only available in unstructured fashion, we investigated various approaches for model portability. Whereas in-domain (i.e. condition) training and evaluation shows very good classification results, the performance of models trained on one specific condition and tested on another condition, varies among domains. However, conditions which belong to similar medical fields and are partly treated with equal medications, also show higher potentials for model transferability. Cross-data evaluation, i.e. training and testing classifiers on data from different sources, was only unsatisfactorily possible with the applied classifier and features. Therefore, we believe that employing more sophisticated features and applying more powerful machine learning models, e.g. deep learning approaches as proposed in [5], can improve the achieved results. Furthermore, the results clearly indicate that especially aspect-based sentiment analysis requires more extensive data sets to extract features with sufficient generalization capabilities. However, we believe that this work contributes to open up future research directions, improves automatic extraction of aspect-related sentiments from patient drug reviews and promotes pharamcovigilance and development of CDSSs such as therapy recommender systems.

REFERENCES

[1] T. Al-Moslmi, N. Omar, S. Abdullah, and M. Albared. 2017. Approaches to Cross-Domain Sentiment Analysis: A Systematic Literature Review. *IEEE Access* 5 (2017), 16173–16192. https://doi.org/10.1109/ACCESS.2017.2690342

[2] Diana Cavalcanti and Ricardo Prudêncio. 2017. Aspect-Based Opinion Mining in Drug Reviews. In *Progress in Artificial Intelligence*, Eugénio Oliveira, João Gama, Zita Vale, and Henrique Lopes Cardoso (Eds.). Springer International Publishing, Cham, 815–827.

[3] Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20, 1 (1960), 37–46. https://doi.org/10.1177/001316446002000104

[4] Kerstin Denecke. 2015. *Sentiment Analysis from Medical Texts*. Springer International Publishing, Cham, 83–98. https://doi.org/10.1007/978-3-319-20582-3_10

[5] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Domain Adaptation for Large-scale Sentiment Classification: A Deep Learning Approach. In *Proceedings of the 28th International Conference on International Conference on Machine Learning (ICML'11)*. Omnipress, USA, 513–520.

[6] Lorraine Goeuriot, Jin-Cheon Na, Wai Yan Min Kyaing, Christopher Khoo, Yun-Ke Chang, Yin-Leng Theng, and Jung-Jae Kim. 2012. Sentiment Lexicons for Health-related Opinion Mining. In *Proceedings of the 2Nd ACM SIGHIT International Health Informatics Symposium (IHI '12)*. ACM, New York, NY, USA, 219–226. https://doi.org/10.1145/2110363.2110390

[7] Vinodhini Gopalakrishnan and Chandrasekaran Ramaswamy. 2017. Patient opinion mining to analyze drugs satisfaction using supervised learning. *Journal of Applied Research and Technology* 15, 4 (2017), 311 – 319. https://doi.org/10.1016/j.jart.2017.02.005

[8] Felix Gräßer, Stefanie Beckert, Denise Küster, Susanne Abraham, Hagen Malberg, Jochen Schmitt, and Sebastian Zaunseder. 2017. Neighborhood-based Collaborative Filtering for Therapy Decision Support. In *Proceedings of the 2nd International Workshop on Health Recommender Systems co-located with the 11th International Conference on Recommender Systems Italy, August 31, 2017.* 22–26.

[9] Ioannis Korkontzelos, Azadeh Nikfarjam, Matthew Shardlow, Abeed Sarker, Sophia Ananiadou, and Graciela H. Gonzalez. 2016. Analysis of the Effect of Sentiment Analysis on Extracting Adverse Drug Reactions from Tweets and Forum Posts. *J. of Biomedical Informatics* 62, C (Aug. 2016), 148–158. https://doi.org/10.1016/j.jbi.2016.06.007

[10] Robert Leaman, Laura Wojtulewicz, Ryan Sullivan, Annie Skariah, Jian Yang, and Graciela Gonzalez. 2010. Towards Internet-age Pharmacovigilance: Extracting Adverse Drug Reactions from User Posts to Health-related Social Networks. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing (BioNLP '10)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 117–125. http://dl.acm.org/citation.cfm?id=1869961.1869976

[11] Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

[12] A. Mishra, A. Malviya, and S. Aggarwal. 2015. Towards Automatic Pharmacovigilance: Analysing Patient Reviews and Sentiment on Oncological Drugs. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. 1402–1409. https://doi.org/10.1109/ICDMW.2015.230

[13] Jin-Cheon Na and Wai Yan Min Kyaing. 2015. Sentiment Analysis of User-Generated Content on Drug Review. *Journal of Information Science Theory and Practice* 1, 1 (Mar 2015). https://doi.org/10.1633/JISTaP.2015.3.1.1

[14] Azadeh Nikfarjam and Graciela H. Gonzalez. 2011. Pattern mining for extraction of mentions of Adverse Drug Reactions from user comments. *AMIA Annual Symposium proceedings* 2011 (2011), 1019–1026.

[15] Azadeh Nikfarjam, Abeed Sarker, Karen O'Connor, Rachel E. Ginn, and Graciela Gonzalez. 2015. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. In *JAMIA*.

[16] Abeed Sarker, Rachel Ginn, Azadeh Nikfarjam, Karen OĂćâĆňâĎćConnor, Karen Smith, Swetha Jayaraman, Tejaswi Upadhaya, and Graciela Gonzalez. 2015. Utilizing social media data for pharmacovigilance: A review. *Journal of Biomedical Informatics* 54 (2015), 202 – 212. https://doi.org/10.1016/j.jbi.2015.02.004

[17] Lei Xia, Anna Lisa Gentile, James Munro, and José Iria. 2009. Improving Patient Opinion Mining Through Multi-step Classification. In *Proceedings of the 12th International Conference on Text, Speech and Dialogue (TSD '09)*. Springer-Verlag, Berlin, Heidelberg, 70–76. https://doi.org/10.1007/978-3-642-04208-9_13