Medical Sentiment Analysis using Social Media: Towards building a Patient Assisted System

Shweta Yadav, Asif Ekbal, Sriparna Saha, Pushpak Bhattacharyya
Department of Computer Science and Engineering
Indian Institute of Technology Patna
Patna, Bihar, India
{shweta.pcs14, asif, sriparna,pb}@iitp.ac.in

Abstract

With the enormous growth of Internet, more users have engaged in health communities such as medical forums to gather health-related information, to share experiences about drugs, treatments, diagnosis or to interact with other users with similar condition in communities. Monitoring social media platforms has recently fascinated medical natural language processing researchers to detect various medical abnormalities such as adverse drug reaction. In this paper, we present a benchmark setup for analyzing the sentiment with respect to users' medical condition considering the information, available in social media in particular. To this end, we have crawled the medical forum website 'patient.info' with opinions about medical condition self narrated by the users. We constrained ourselves to some of the popular domains such as *depression*, *anxiety*, *asthma*, and *allergy*. The focus is given on the identification of multiple forms of medical sentiments which can be inferred from users' medical condition, treatment, and medication. Thereafter, a deep Convolutional Neural Network (CNN) based medical sentiment analysis system is developed for the purpose of evaluation. The resources are made available to the community through LRE map for further research.

Keywords: Medical sentiment, Medical blog, Machine learning, Deep learning, Convolutional Neural Network

1. Introduction

Attention towards sentiment analysis has been flourishing over the last two decades because of the immense popularity of social media. The phenomenal rise in blogging trend is observed in health communities such as medical forums which are swamped by millions of users (many of whom are patients) seeking for health-related information, sharing medical problems or experiences and opting for informational support or opinions from the other users (patients, health-professional or doctors). These self-narrated texts provide a platform to peek into a blogger's state-of-the-mind for several reasons:
(i) the subjective nature of the contents generated by blogger's; **(ii)** the temporal aspect of the blog which can be formed into thread; **(iii)** the abundance of blog data which allows cumulation of opinions, sentiments and thoughts in a very wide spectrum.
Medical sentiment analysis has its major applications in assessing the clinical records and in providing an automated decision support system for health professional. According to the study conducted by the Pew Internet & American Life Project[1], almost 80 percent of Internet users in US have explored health-related topic online. More often, people look for the information about specific medical problem (63%) over the internet. Nearly 47% of the users search for the medical treatment or procedure in the internet. With such a tremendous amount of freely available medical texts in the web, it is necessary to harness the crucial and important information. Analyzing these texts by capturing the sentiments is helpful because opinions are central to almost all human activities and are key influencer of our behaviors. Although, several techniques exist to capture sentiments in general domains, the sentiments expressed in medical narratives have not been well analyzed and exploited in the required measure as yet. The research in medical sentiment analysis mainly focuses on biomedical literature and Electronic Medical Record (EMR) documents. Recently, preliminary study was conducted by (Denecke and Deng, 2015) to capture the medical sentiment from clinical narratives and medical social media sources. Several shared-task challenges (Losada et al., 2017; Hollingshead et al., 2017) have also been conducted to study the social media texts to capture the user's opinion in medical setting. For this purpose, they generated domain-specific corpus containing clinical documents (nurse letters, radiology reports and discharge summaries) collected from the MIMIC II database. Furthermore, they extracted drug reviews from medical blogs such as WebMD, DrugRating. This study provides the quantitative assessment of sentiment in terms of six corpora with 1000 documents.
Literature survey shows that medical sentiment analysis, nowadays, is a topic of growing research interest. In this work, we have explored how sentiment analysis from medical forums can be effective for building a patient assisted healthcare system. We have provided a benchmark setup for mining patient opinions extracted from the medical web forums. Towards this end, we have studied different aspects of medical sentiment in health related texts that may relate to the following:

- **Health status:** Alteration in the health status which may vary over a particular time period.
- **Degree of medical condition that impacts patient life** (e.g., *severe headache impacts the patient's life more than the mild headache.*)
- **Consequence of a treatment** (e.g., there may be positive or negative impacts in a patient's treatment, such

[1] http://www.pewinternet.org/

Classification 1: Medical Condition			Classification 2: Medication	
Medical Blog	Label		Medical Blog	Label
This morning I had an **attack** of it that was **very sudden and very intense**. I felt an **incredible surge of unsteadiness**.	Exist		Hi been on Sertaline now for abut 4 weeks. Maybe nearer 5. I started on 25mg and now been on 50 for around 2 or 3 weeks. My **mood** has **definitely improved** and I am **alot calmer**	Effective
Previously I have taken flixonase and beconase which has given **no long term relief** 10 days ago I went back to the doctor and was given Betnesol. This has **immediately relieved me all symptoms**.	Recover		Had anxiety for few months on citalopram propanolol and 2mg diazapam. Took first diazapam today as I have health anxiety and was **scared** to take them. **Nothing seems to help been in bed for two days can't sleep** waking at 3 4 5 am	Ineffective
I recently started lexapro 3 days, I'm absolutely lost I feel **weak and shaky** everyday and **can't eat right** I don't sleep normal. I'll **die young** and the cause will be cardiac arrest	Deteriorate		I've been feeling a bit off still.. Day 6 that I haven't taken my citalopram. Anxiety is down, but now I'm starting to **feel more and more off.. Random high chest pains** Plus feeling a **bit foggy and spacey**..	Serious adverse effect (SAE)

Table 1: Exemplar description of annotation scheme. The words in bold represent possible sentiments.

as *'flixonase and beconase which has given no long term relief'*.)

- **Opinions towards a treatment** (e.g., a patient can have an adverse reaction after consumption of drug)
- **Certainty of a diagnosis:** (e.g., Health professional can be certain of some diagnosis.)

Medical sentiment can be studied at the various aspects like 'medical condition', 'treatment', 'procedure', etc. which can directly impact the users' health conditions. By analyzing patient status periodically, any progress or deterioration can be identified. Any user expresses his/her medical condition implicitly or explicitly. Implicit sentiment in medical context concerns mentioning of the symptoms, for instance consider the blog: *'I recently started **lexapro** 3 days, I'm on **extreme weight loss***. Here, '**weight loss** as in such does not reflect anything negative but in the above sentence, it represents the adverse drug reaction where sentiment is implicitly defined to be negative. These require additional information for making correct interpretation. In case of explicit sentiment, it is relatively much easier to analyze the health conditions. For example, consider "*I recently started **lexapro** 3 days, I'm **absolutely lost** I feel **weak and shaky** everyday". Here, **absolutely lost, weak and shaky*** provide symptoms which are explicitly defined.

For this work, we have collected a corpus consisting of $7,490$ user blog posts from popular medical forum 'patient.info' which is split on the basis of two major medical sentiment aspects, namely 'medical condition' and 'treatment'. The corpus is manually annotated with a predefined set of categories. Finally a Convolutional Neural Network (CNN) based model is developed for medical sentiment classification.

To the best of our knowledge, there is no existing benchmark setup available for medical sentiment analysis. We believe that creating such a resource might be beneficial for building a patient assisted healthcare monitoring system. Below the contributions of our work are summarized:

1. Introduction of a novel annotation scheme for analyzing medical sentiment;
2. Development of an annotated medical sentiment corpus;
3. Building a deep CNN based medical sentiment classification system;
4. More deeper analysis of the sentiments with respect to medical domain.

The paper is structured as follows: Related works is presented in Section 2. The annotation scheme is introduced in Section 3. Section 4 discusses the method for corpus collection and annotation. In Section 5, we describe our method for capturing the medical sentiments. The results of the annotation study are presented in Section 6. At the end of the paper, we provide pointers for future work.

2. Related Works

Recent years have witnessed rapid proliferation in research on identifying and assembling subjective expressions or other non-factual expressions of textual contents characterizing peoples' opinions, feelings or emotions using medical blog texts. In general, we can categorize the existing works utilizing medical blog texts into three groups as follows:

- **Polarity Classification:** Some of the popular works include the studies carried out by Xia et al. (2009) which aimed to classify the patient opinions in eight categories and observed its polarity (positive, negative). Sokolova et al. (2011) also focused on classifying the tweets on the basis of sentiment (positive, negative and neutral). They used bag-of-words (BoWs) as features to learn several classifiers such as naive Bayes, decision trees and support vector machines. Study conducted by Biyani et al. (2013) used online cancer community user data to determine the polarity. They have adapted supervised machine learning techniques using hand-crafted features, which cover both domain-dependent as well as domain-independent features. They identified sentiments on two discourse

functions such as expressive and persuasive. A supervised machine learning model (multi-nominal naive Bayes) is developed using frequency-based features.

- **Adverse Drug Relation:** For medical domains, social media texts (corresponding to medical forums) have been utilized in the works such as DS (Leaman et al., 2010; Nikfarjam and Gonzalez, 2011; Liu and Chen, 2013), MedHelp (Yang et al., 2012) and PatientsLikeMe (Wicks et al., 2011). Non-medical social media forums like Twitter (Nikfarjam et al., 2015) have been exploited to capture adverse drug effect. With the availability of the extensive Adverse Drug Reaction (ADR) lexicons such as Side Effect Resource (SIDER)[2] (Kuhn et al., 2010), Coding Symbols for a Thesaurus of Adverse Reaction Terms (COSTART), Consumer Health Vocabulary (CHV) (Zeng-Treitler et al., 2008) and Medical Dictionary for Regulatory Activities (MedDRA) (Mozzicato, 2009), some prominent studies (Leaman et al., 2010; Yates and Goharian, 2013) focus on exploiting these pre-existing lexicons to identify ADR mentions in user posts. Some of the other popular studies include the works of (Na et al., 2012; Sharif et al., 2014) utilizing machine learning based NLP techniques to identify the ADR.

- **Emotion Classification:** Sokolova and Bobicev (2013) studied different forms of emotions from medical web documents. They analyzed the categories such as encouragement (hope, happiness), confusion (worry, doubt, concern), gratitude, facts, and facts+encouragement. They applied naive Bayes classifier with the features derived from lexicon WordNetAffect. Study conducted by Melziet et al. (2014) on emotion classification learned SVM using the feature set consisting of BoWs, n-grams and specific attributes.

3. Benchmarking and Annotation Scheme

In this section, we define the benchmark setup by studying the sentiment expressed in medical blog posts. Here, we focus on fine-grained medical sentiment aspects of the users' health status and treatment. Our intention is that the annotation scheme should be able to capture multiple perspectives of user health status. Below we provide two important medical aspects with the possible categories of sentiment values:

Classification 1: Medical Problem: Exist, Recover, Deteriorate.

Classification 2: Medication: Effective, Ineffective, Serious adverse effects.

We categorize medical problems into the following three possible sentiment classes:
Exist: Here user shares the symptoms (negative sentiment) of any medical problem.
Recover: The user shares the recovering status (positive sentiment) from the previous medical problems.
Deteriorate: The user describes its medical condition to be worsen (negative sentiment) over the span of medical treatment.

The other classification strategy concentrates on the effect of the medication. We describe below the possible sentiment values :
Effective: User shares the positive sentiment in the form of usefulness of the treatment.
Ineffective: The no effect of the treatment is reported in the user narration.
Serious adverse effect: User shares the negative opinion towards the treatment mainly in the form of adverse drug effect. The blog post falling under this category has to have the explicit mention of the drug name in the text.
From the examples as presented in Table-1, we analyze that sentiment in clinical narratives cannot always be manifested in single terms or phrases, rather it heavily depends on the context. The concept of medical sentiment is very complex and has multiple facets making it very interesting, but also challenging for automatic analysis.

4. CMS: Corpora for Medical Sentiment

Attributed to the fact of growing interest in users' self stated medical reviews, we crawl the medical forums where multiple users discuss on various medical conditions. We consider the following points while selecting the source of information from which to extract the corpus:

- It should be extremely popular and reliable site in search of medical issues with reasonable number of users.

- There should exist fair number of opinions which must either have discussions on medical conditions or medications.

In order to obtain potential and effective sources which satisfy the above requirements, we did exhaustive search exploiting multiple medical forums. The task was quite tedious as most of the forums either do not have sufficient number of users or the text was heavily noisy. After surveying several websites, we chose the 'patient.info' [3] medical forum. This forum contains on an average 1500 opinions per medical discussion group. We selected popular discussion groups such as *Anxiety*, *Depression*, *Asthma* and *Allergy* having $5,000$ blog posts on an average. In total we collected $10,000$ blog posts of which $5,188$ posts concern about the medical conditions and $2,302$ contain medication related blog posts which were collected during the period of 25^{th} September 2016 to 15^{th} November 2016. We removed $2,510$ blog-posts which did not have any mention of medication or medical condition. To ensure the confidentiality of user, all the user related information were removed. The statistics of corpus are presented in Table-2 and Table-3. A team of three expert annotators independently annotated the user posts with three classes on both the classification strategies. The Cohen's kappa approach (Cohen, 1960) was

[2]http://sideeffects.embl.de/

[3]https://patient.info/

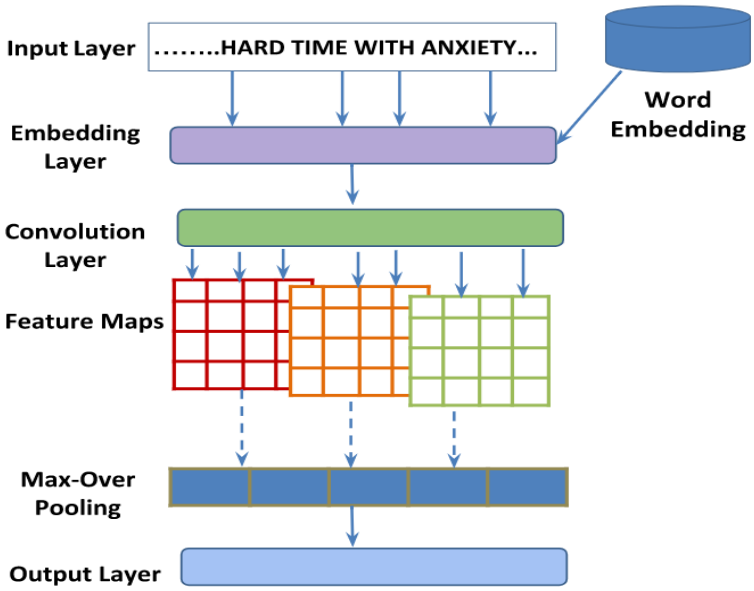


Figure 1: Proposed architecture for predicting the medical sentiment from blog-post.

Classification 1: Medical Condition				
Exist	Recover	Deteriorate	Avg # of sentences	Avg # of words
2396	703	2089	10	192

Table 2: Dataset statistics for classification-1

Classification 2: Medication				
Effective	Ineffective	Serious Adverse Effect	Avg # of sentences	Avg # of words
462	613	1,226	9	176

Table 3: Dataset statistics for classification-2

used to measure the inter-annotator agreement. We observe high agreement ratio of 0.79 for exact matching of the class with respect to each blog post.

5. Approach for Capturing Medical Sentiment

In this section we have presented the approach developed for extracting sentiments of users' posts in medical blogs.

5.1. Network for Identifying Severity Level

In this section we propose a method based on CNN that exploits sentiments from health forums (or, medical blogs) in augmentation layer. As presented in Figure-1, the proposed model has four different components which are similar to the conventional CNN components as proposed by (Kim, 2014). The first layer represents the input layer which takes a complete blog post in the form of vector representation (word embedding) and outputs a probability corresponding to the classification types. We use max-pooling over the whole blog post to obtain global features through all the filters. This pooled feature is fed into the fully connected neural network. In the output layer, we use the softmax classifier to automatically classify the post into three output classes. We describe below the layers of our proposed model in details:

1. **Input layer:** Each blog post is provided as the input to the model.

2. **Word embedding layer:** This layer encodes every word into a real-valued vector. Given a blog text M consisting of n words $w_1, w_2, w_3, ...w_n$, each word w_i is transformed into real-valued vector x_i. Each word in M is looked up in the corresponding word embedding matrix $W \in R^{k \times |V|}$, where $|V|$ represents fixed length vocabulary and k is the word embedding size. The blog-post representation matrix $x_{1:n_W}$ can be constituted as:

$$x_{1:n_W} = x_1 \otimes x_2 \ldots \otimes x_{n_W} \tag{1}$$

where $\otimes$ represents the concatenation operator. We perform zero padding in case the number of the words in blog text is less than n to fix the length.

3. **Convolution layer:** Word embedding is fed as the input to the convolutional layer where filter $\mathbf{F} \in \mathbb{R}^{m \times k}$ is convoluted to the context window $x_{i:i+m-1}$ of h words for each blog-post as follows.

$$c_i = f(\mathbf{F}.x_{i:i+m-1} + b) \tag{2}$$

where f is non-linear function[4] and b is a bias term. The feature map f is generated by applying given filter $\mathbf{F}$ to every potential window of word in the blog-post.

$$\begin{aligned} f = [&g(\mathbf{F} \cdot x_{1:1+h-1} + b), g(\mathbf{F} \cdot x_{2:1+h-1} \\ &+ b) \ldots g(\mathbf{F} \cdot x_{n-h+1:n} + b) \\ = [&f_1, f_2, f_3 \ldots . f_{n-h+1}] \end{aligned} \tag{3}$$

[4] In our experiments we have used the rectified linear unit as a non linear function.

Classification Strategy	Classification Models		Precision	Recall	F1-Score
Medical Condition	Baselines	SVM	0.42	0.49	0.43
		Random Forest	0.45	0.48	0.46
		MLP	0.41	0.43	0.40
	CNN		0.68	0.60	0.63
Medication	Baselines	SVM	0.74	0.76	0.75
		Random Forest	0.72	0.73	0.73
		MLP	0.74	0.75	0.74
	CNN		0.86	0.77	0.82

Table 4: Performance comparison of CNN architecture with other baseline classifiers

Effective	**Ineffective**	**Adverse Drug Effect**
feeling wonderful More energy	feeling down Moods fluctuating	feel odd sensations skin
feel like normal person now	feel like death! feel	feeling like drunk night
feel most comfortable pacing	feel totally hopeless almost	feel horrible dizzy sickly
feel great better than done	feel really down and hard	feel super nauseous sleeping

Table 5: The set of informative 4-grams over different classes of Medication category

In order to increase the coverage of n-gram model, multiple filters with different window sizes can be applied.

4. **Pooling layer:** The function of the pooling layer is to gradually minimize the spatial size of the representation by identifying the most abstracted feature generated by the convolutional layer. It involves non-linear down sampling to extract most relevant set of features. In our work, we apply max-pooling operation over feature map and set the maximum value as a feature for this particular filter. The max-pooling operation is performed over feature map as follows:

$$\hat{d} = max(f_1, f_2, f_3.....f_{n-h+1}) \quad (4)$$

5. **Output layer:** The blog-level feature vector is passed to the softmax layer to label 'y' from a discrete set of classes for the corresponding blogs 'M'.

5.2. Hyperparameter Settings in CNN

The values of hyper-parameters are determined from preliminary experiments by evaluating the model's performance using 5-fold cross validation by varying the convolution feature sizes (100, 200 & 300). Word embedding is generated through pre-trained Google news word embedding model. Specifically, all the deep learning models use the 300-dimension word embedding, feature map size of 300 on multiple filters with window sizes of 3, 4, 5. We use Adam (Kingma and Ba, 2014) as our optimization method with a learning rate of 0.001. Training was performed using stochastic gradient descent over mini-batches considering the Adadelta (Zeiler, 2012) update rule. As a regularizer, we use dropout (Hinton et al., 2012) with a probability of 0.5. After training, we choose the best performing model to be evaluated on the test sets. The model introduced in this paper is implemented on Theano [5].

[5]http://deeplearning.net/software/theano/

6. Experimental Results and Analysis

To evaluate the effectiveness of our algorithm, we have developed three strong baselines models:

Baseline 1: The first baseline model is constructed by training SVM (Cortes and Vapnik, 1995).

Baseline 2: In this model, we use Random Forest (Breiman, 2001) based classification model.

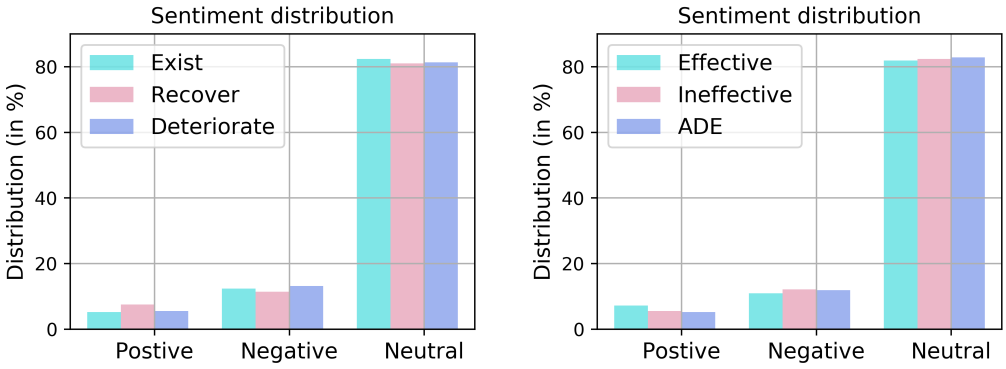
Baseline 3: Multi-layer perceptron (MLP) (Collobert and Bengio, 2004) is utilized to learn the model. In order to learn the baseline classifiers, we used the following feature set which is specific to the forum data.

- **N-grams:** This feature plays a very important role in capturing the contextual information of the blog. We generated uni-grams, bi-grams, tri-grams with respect to the target words within the window size of $[-2, 2]$.
- **Medical abbreviated feature:** Generally, users tend to use abbreviated wordforms to describe medical condition or treatment for e.g., ECG/EKG for Electrocardiogram. We created medical abbreviation dictionary by crawling medical acronym and abbreviation related website[6]. We generated binary feature which sets the feature value to 1, if the target word is present in the dictionary else the value is set to 0.
- **Sentiment feature:** We designed three real valued features which compute the positive, negative and neural sentiment scores of the blog by finding number of positive, negative and neutral words in a document. The sentiment score was calculated by using most recent and popular lexicon, SentiWordnet [7]. These three sentiment scores were calculated by the following equation:

$$Score^{(K)}(blog) = \sum_{i=1}^{n} SC^{(K)}(w_i)/n \quad (5)$$

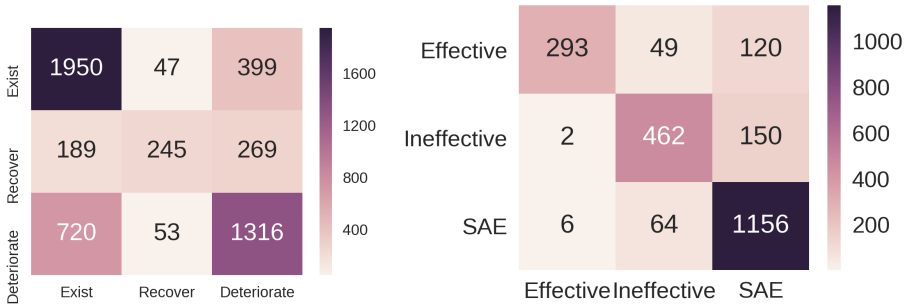
[6]http://www.health.am/acronyms/

[7]http://sentiwordnet.isti.cnr.it/



(a) Medical Condition (b) Medication

Figure 2: Sentiment word distribution through SentiWordNet



(a) Medical Condition (b) Medication

Figure 3: Confusion Matrix for both the classification strategies

Here $SC^{(K)}$ denotes the sentiment score of the word for k^{th} sentiment where $K \in \{+, -, neutral\}$ and n is the number of words (w) in a medical blog.

We have reported the results obtained by our CNN based sentiment classification model along with other baseline models in Table-4. The CNN system that uses only the pre-trained embedding achieves 63% and 81% F1-Score values on *medical condition* and *medication* classification strategies, respectively. The confusion matrix for both types of classification problems is presented in Figure-3. Our CNN based model obtains significant performance improvements over all the three baselines for both the classification strategies. Feature ablation experiments are also conducted to analyze the importance of different features selected. Analysis shows that medical abbreviation and sentiment scores are not effective features in medical setting.

6.1. Major Analysis

Our analysis on the user-generated medical blog reveals that the usual health status information is presented in an elusive way by the user. The word usage in the medical blog is more implicit and requires deeper analysis of metaphor and sarcasm. We have illustrated these scenarios in Figure-2, where our system was unable to capture the implicit negative or positive sentiment present in the users' posts and thus the posts were classified into neutral.

The general SentiWordNet(SWN) lexicon is observed to be not assisting the system in capturing the sentiment in medical setting. For example, consider a text from medical blog-post

"all the sudden my heart like drops and feels like its going to stop.".

Here, the phrase 'heart like drops' and 'going to stop' are user's narrated symptoms presenting the examples for implicit negative sentiment. However, the SWN lexicon provides the label neutral to these words as in general these do not carry any positive or negative sentiment.

After deep analysis of data, we observed that majority of the medical sentiment occurs in the vicinity of the term 'feel' and its variations. We have generated the 4-grams taking 'feel' as the target word and as shown in Table-5, we have observed that these 4-gram words can provide an effective clue in capturing the sentiment. Further, more semantics and context-dependent features are required to capture the peculiarity of the medical blog text.

We have also observed that deeper understanding of sentiment in MS analysis further requires consideration of the context which may not be available on the blog. For example, problems with a 'head' can be captured with multiple symptoms: 'headache, nausea, fever'. Thereby it is highly required to utilize background knowledge in order to cluster these symptoms to the similar medical condition.

7. Conclusions and Future Works

We have presented a large corpus of annotated data collected from the 'Patient.info' forum containing users' original posts written on the 'Anxiety', 'Depression', 'Asthma'

and 'Allergy' forums. This paper provides fine-grained annotation scheme to capture the sentiment in medical setting which concentrates on detailed medical aspects such as 'medication' and 'medical condition' instead of conventional polarity (positive or negative) to judge user's health status. We have also presented a deep convolutional neural network based classification framework to predict the possible medical sentiment category for both 'medication' and 'medical condition' classification schemas. We are able to obtain significant performance improvements over the baseline in all the cases. In future, we aim to develop the medical sentiment specific lexicon and would like to propose a method to capture implicit, metaphoric & sarcastic phrases.

8. Acknowledgements

Asif Ekbal greatfully acknowledges the Young Faculty Research Fellowship (YFRF) Award, supported by Visvesvaraya PhD scheme for Electronics and IT, Ministry of Electronics and Information Technology (MeitY), Government of India, being implemented by Digital India Corporation (formerly Media Lab Asia).

9. Bibliographical References

Biyani, P., Caragea, C., Mitra, P., Zhou, C., Yen, J., Greer, G. E., and Portier, K. (2013). Co-training over domain-independent and domain-dependent features for sentiment analysis of an online cancer support community. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 413–417. ACM.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32, Oct.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Collobert, R. and Bengio, S. (2004). Links between perceptrons, mlps and svms. In *Proceedings of the twenty-first international conference on Machine learning*, page 23. ACM.

Cortes, C. and Vapnik, V. (1995). Support vector machine. *Machine learning*, 20(3):273–297.

Denecke, K. and Deng, Y. (2015). Sentiment analysis in medical settings: New opportunities and challenges. *Artificial intelligence in medicine*, 64:17–27.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

Kristy Hollingshead, et al., editors. (2017). *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality*. Association for Computational Linguistics, Vancouver, BC, August.

Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1746–1751.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

Kuhn, M., Campillos, M., Letunic, I., Jensen, L. J., and Bork, P. (2010). A side effect resource to capture phenotypic effects of drugs. *Molecular systems biology*, 6(1):343.

Leaman, R., Wojtulewicz, L., Sullivan, R., Skariah, A., Yang, J., and Gonzalez, G. (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*, pages 117–125. Association for Computational Linguistics.

Liu, X. and Chen, H. (2013). Azdrugminer: an information extraction system for mining patient-reported adverse drug events in online patient forums. In *International Conference on Smart Health*, pages 134–150. Springer.

Losada, D. E., Crestani, F., and Parapar, J. (2017). Clef 2017 erisk overview: Early risk prediction on the internet: Experimental foundations.

Melzi, S., Abdaoui, A., Azé, J., Bringay, S., Poncelet, P., and Galtier, F. (2014). Patient's rationale: Patient knowledge retrieval from health forums. In *eTELEMED: eHealth, Telemedicine, and Social Medicine*, volume 2014, page 140.

Mozzicato, P. (2009). Meddra. *Pharmaceutical Medicine*, 23(2):65–75.

Na, J.-C., Kyaing, W. Y. M., Khoo, C. S., Foo, S., Chang, Y.-K., and Theng, Y.-L. (2012). Sentiment classification of drug reviews using a rule-based linguistic approach. In *International Conference on Asian Digital Libraries*, pages 189–198. Springer.

Nikfarjam, A. and Gonzalez, G. H. (2011). Pattern mining for extraction of mentions of adverse drug reactions from user comments. In *AMIA Annu Symp Proc*, volume 2011, pages 1019–1026.

Nikfarjam, A., Sarker, A., O'Connor, K., Ginn, R., and Gonzalez, G. (2015). Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association*, page ocu041.

Sharif, H., Abbasi, A., Zafar, F., and Zimbra, D. (2014). Detecting adverse drug reactions using a sentiment classification framework. In *Proceedings of the sixth ASE international conference on social computing (SocialCom)*.

Sokolova, M. and Bobicev, V. (2011). Sentiments and opinions in health-related web messages. In *RANLP*, pages 132–139.

Sokolova, M. and Bobicev, V. (2013). What sentiments can be found in medical forums? In *RANLP*, volume 2013, pages 633–639.

Wicks, P., Vaughan, T. E., Massagli, M. P., and Heywood, J. (2011). Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nature biotechnology*, 29(5):411–414.

Xia, L., Gentile, A. L., Munro, J., and Iria, J. (2009). Improving patient opinion mining through multi-step classification. In *International Conference on Text, Speech and Dialogue*, pages 70–76. Springer.

Yang, C. C., Yang, H., Jiang, L., and Zhang, M. (2012). Social media mining for drug safety signal detection. In *Proceedings of the 2012 international workshop on Smart health and wellbeing*, pages 33–40. ACM.

Yates, A. and Goharian, N. (2013). Adrtrace: detecting expected and unexpected adverse drug reactions from user reviews on social media sites. In *European Conference on Information Retrieval*, pages 816–819. Springer.

Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.

Zeng-Treitler, Q., Goryachev, S., Tse, T., Keselman, A., and Boxwala, A. (2008). Estimating consumer familiarity with health terminology: a context-based approach. *Journal of the American Medical Informatics Association*, 15(3):349–356.