ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

Sentimental Classification of User Reviews in Social Networks for Cancer Diseases

Vijaya Lakshmi D M, Deepa D, Tamilarasi A
Department of Computer Science and Engineering
Kongu Engineering College
Erode,India

Abstract

Sentiment Analysis is also known as opinion mining deriving the opinion of attitude of a speaker. A basic task in sentiment analysis is classifying the polarity of a given text. The applications for sentiment analysis are endless. More and more sentiment analysis are used in social media monitoring to track customer reviews, survey responses, competitors etc. Existing tools for sentiment analysis have a very low accuracy when used in web contexts. SHC-pt is a tool for sentiment analysis for cancer reviews which were extracted from Facebook and Twitter. This was possible because a specific lexicon for cancer domain is used. This lexicon considers semantic emoticons, hash tags, priority, dictionary, n-gram and syntactic phrases that contributed to the good result of SHC-pt. In the proposed work it can be extended to some other diseases by considering Facebook and Twitter comments. The accuracy of these works are critically examined with the help of performance metrics like Precision, Recall, F-Measure and Accuracy and inferences were made.

Keywords: *Sentimental analysis, Opinion mining, SHC-pt Senti health Cancer tool, Lexicon.*

I. INTRODUCTION

Sentiment Analysis is widely used to analyze opinions from people about a target, for example a product or a service. Existing SA techniques can be divided into three categories, depending on the level at which the analysis is made in the text document level, sentence level and entity/aspect level. At the document level, the opinion of a document is classified as positive, negative or neutral. This type of analysis it is not possible to classify a document that covers more than one entity, because each document is interpreted as having a text referencing just a single entity. Differently of that, the analysis in sentence level classifies an opinion into three classes positive, negative or neutral. Both analysis in the document level and sentence level uses only the language constructs to classify an opinion. However, the analysis in the entity and aspect level considers that for every opinion there is a target. Therefore it seeks to identify the target of each existing opinion in the text. This allows to analyze more than one opinion in a same sentence For example, the phrase “Although a bad service, I still like that restaurant.” Here it has more positive opinion than negative about the restaurant, but it has two fact aspects to be evaluated one is the service offered and the other is about restaurant. These are the targets of the opinion. Some works were done about SA, especially comparing the tools proposed.

Although there are several studies about SA, few scientific studies uses SA to classify a person emotional state considering the person himself as the target of analysis. However, it is possible to know if a person has more positive or negative thoughts by analyzing his texts. For example, if most of texts are negatives in a window of time, this person probably is in a negative emotional state. Unlike it considers texts and reviews extracted from posts appearing in Facebook communities of cancer patients. The reviews regards the authors of the texts as the targets of the analysis and uses SA solutions to classify the sentiment of the authors.SA techniques applied on posts in cancer online communities may not be used only to detect pessimistic emotional state but also to detect changes in a person mood in consequence of his interactions with other patients in a community and the person who writes a text without the specific purpose of reporting their emotional state may end up with revealing or unintentionally. Although he is more positive or

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

negative. This can be used for giving emotional support to the patients. A chronic patient may face various difficulties such as physical pain, stress, extreme anxieties, anger, depression, and frustration. These difficulties could cause suffering during the treatment and even take the patient to interrupt his treatment. So many patients seeks support in social networks to obtain information, encouragement, motivation, feedback, emotional support, tangible support, and network support that is exchanged among peers to have a better quality of life during their treatment. Thus the automatic analysis of patients mood can be very useful for assistants, family and patients themselves. SA methods are good options for this analysis. However, most works of SA assess opinions on a target which is different from the author of the emitted opinion. In addition, few studies focus on analyzing sentiment of cancer patients and their families, who go through strong experiences too and usually influenced emotionally by the context surround the patients. Finally, there are few works proposing context driven SA solutions aiming at improving accuracy of the classification result. In most cases the techniques are general listed and such case do not perform well on specific contexts. The objective of this work is to present a SA tool, named Senti Health Cancer (SHC-pt), that improves the detection of emotional state of patients in Brazilian online cancer communities, by inspecting their user reviews and posts.

The rest of this report is organized as follows: Section 2 discusses the related work on automatic summarization and sentence similarity measures. The proposed work to perform Extractive Text Summarization in mobile devices using fuzzy relational clustering algorithm is discussed in Section 3. In Section 4, the experimental evaluation for the sample dataset is discussed. Finally, additional discussions, conclusion and future work are presented in Section 5.

II. RELATED WORK

The automatic analysis of user generated contents such as online news, reviews, blogs and tweets are extremely valuable for tasks such as mass opinion estimation, corporate reputation measurement, political orientation categorization, stock market prediction, customer preference and public opinion study. Public health surveillance is critical to monitoring the spread of infectious diseases using traditional surveillance systems. To address this problem twitter and facebook message has been used. Dunkel Schetter et al. (2013) presented a work that past research indicates that social support is beneficial to cancer patients in adjusting to the stress of the disease. In this a brief review of research on social support is provided as a framework within which support among cancer patients can be examined. Research on cancer is then reviewed and selected results from an investigation of 79 cancer patients are reported. The findings indicate that health care providers are particularly important sources of support to cancer patients of several types. Emotional support is seen as especially helpful and the types of support seen as most helpful by those with cancer depends on who provides them. In addition, variability in stress among cancer patients mediated the frequency of interpersonal problems, and the association between support and various indices of adjustment. Implications of these results for future research on social support in stressed populations, especially cancer patients, are discussed. Beaudoin et al. (2008) proposed an approach that it considers the impact of internet use and online social capital on the health outcomes of supporters of cancer patients. Structural equation modelling offers support for internet use and online social capital stress and depression. Specifically, asynchronous online communication and offline communication stimulated by online communication had direct positive paths to social interaction and social support which, in turn, were both predictive of healthier or lower levels of stress and depression. These findings support previous research, which has indicated positive associations between mass media use and social capital and between social capital and health outcomes. In contrast, internet information-seeking by information type had a negative direct path to social support and negative indirect paths to stress and depression, indicating that information-seeking online has a detrimental impact on social capital and the health indicators.

Bouma et al. (2015) developed a work that the effect of internet-based support programs on psychosocial and physical symptoms resulting from cancer diagnosis and treatment is analyzed. Selection of studies was based non randomized controlled trials, performed in adult cancer patients, comparing quantitative psychosocial and physical outcomes of an internet-based support program with comparison group. Literature search yielded 2032 studies of which 16 fulfilled the eligibility criteria. Three different internet-based support programs were identified: social support groups, online therapy for psychosocial/physical symptoms, and online systems integrating information, support, and coaching services. Outcomes improved by these programs in nine studies. Especially fatigue, social support, and distress improved, regardless of the program type. All online systems showed positive effects, mainly for social support and quality of life. This analysis indicates that internet-based support programs are effective in improving psychosocial and physical symptoms in cancer patients. Kubzansky et al. (2010) prospectively examined social ties and survival after breast cancer diagnosis. Participants included 2,835 women from the Nurses' Health Study who were diagnosed with stages 1 to 4 breast cancer between 1992 and 2002. Of these women, 224 deaths (107 of these related to breast cancer) accrued to the year 2004. Social networks were assessed in 1992, 1996, and 2000 with the Berkman-Syme Social Networks Index. Social support was assessed in 1992 and 2000 as the presence and availability of a confidant. Cox proportional hazards models were used in prospective analyses of social networks and support, both before and following diagnosis, and subsequent survival.

Hwang et al. (2010) proposed a work that the social support for weight loss shared by members of a large Internet weight loss community and later conducted a mixed-methods study with surveys (n=193) and interviews (n=13) of community members along with a content analysis of discussion forum messages (n=1924 messages). Qualitative data were analyzed for social support themes. Survey respondents were primarily white (91.4%) and female (93.8%) with mean age 37.3 years and mean body mass index 30.9. They used forums frequently, with 56.8% reading messages, 36.1% replying to messages, and 18.5% posting messages to start a discussion related to weight loss on a daily or more frequent basis. Major social support themes were encouragement and motivation, mentioned at least once by 87.6% of survey respondents, followed by information (58.5%) and shared experiences (42.5%). Subthemes included testimonies, recognition for success, accountability, friendly competition, and humour. Members valued convenience, anonymity, and the non-judgmental interactions as unique characteristics of Internet-mediated support. Chen et al. (2009), developed a influence maximization problem of finding a small subset of nodes (seed nodes) in a social network that could maximize the spread of influence. Then the study of efficient influence maximization from two complementary directions. One is to improve the original greedy algorithm of and its improvement to further reduce its running time, and the second is to propose new degree discount heuristics that improves influence spread. Later evaluated our algorithms by experiments on two large academic collaboration graphs obtained from the online archival database arXiv.org. Experimental results show that our improved greedy algorithm achieves better running time comparing with the improvement of with matching influence spread, our degree discount heuristics achieve much better influence spread than classic degree and centrality-based heuristics, and when tuned for a specific influence cascade model, it achieves almost matching influence thread with the greedy algorithm, and more importantly the degree discount heuristics run only in milliseconds while even the improved greedy algorithms run in hours in our experiment graphs with a few tens of thousands of nodes.

Bronchi et al. (2011) proposed a problem of finding a set of users in a social network, such that by targeting this set, one maximizes the expected spread of influence in the network. Here it includes the study of influence maximization from a novel data-based perspective. In particular we introduce a new model, which they call credit distribution that directly leverages available propagation traces to learn how influence flows in the network and uses this to estimate expected influence spread. Although cancer classification has improved over the past 30 years, there has been no general approach for identifying new cancer classes class discovery for assigning tumors to known classes. Here a generic approach to cancer classification based on gene expression monitoring by DNA microarrays is described and applied to human

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

acute leukemias as a test case.A class discovery procedure automatically discovered the distinction between acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) without previous knowledge of these classes. An automatically derived class predictor was able to determine the class of new leukemia cases. The results demonstrate the feasibility of cancer classification based solely on gene expression monitoring and suggest a general strategy for discovering and predicting cancer classes for other types of cancer, independent of previous biological knowledge.Goo et al. (2013) proposed a new problem that is on social network influence maximization. The problem is defined as by a given target user w, finding the top-k most influential nodes for the user. Different from existing influence maximization works which aim to find a small subset of nodes to maximize the spread of influence over the entire network (i.e., global optima).Our problem aims to find a small subset of nodes which can maximize the influence spread to a given target user (i.e., local optima). The solution is critical for personalized services on social networks, where fully understanding of each specific user is essential. Although some global influence maximization models can be narrowed down as the solution, these methods often bias to the target node itself. Then presented a local influence maximization solution. Firstly provided by a random function with low variance guarantee to randomly simulate the objective function of local influence maximization. Later presented the efficient algorithms with approximation guarantee. For online social network applications, we present a scalable approximate algorithm by exploring the local cascade structure of the target user. Later we test the proposed algorithms on several real world social networks.

III. PROPOSED WORK

The proposed work generates the data collected from online social networks were published by the user as a public text in a public group. A Java application was developed to connect with the social network's API and collect posts from selected groups. The online social network selected by us as source of texts was the Facebook because it contains user groups in the theme of this work and also because it is one of most currently used online social network by both cancer patients and by their families. Two Face-book's groups were selected based on volume of data and on public access of messages. The data were collected between March to December 2016. The Senti Health method used in this work uses information about the application context and users communication styles, e.g. Hash tags and emoticons, to improve the classification performance. As an instance of SH method the tool Senti Health-Cancer (SHC-pt) is launched. The tool contributes to automate the method and to evaluate the Facebook reviews. The Senti Health method proposed in this work uses information about the application context and internet user communication styles, **Eg.** hash tags and emoticons, to improve the classification performance. As a instance of SH method the tool Senti Health-Cancer (SHC-pt) is launched. The tool contributes to automate the method and to evaluate it. Here the emotional state of each patient is identified. And each word in a review is classified whether it is positive, negative and neutral. Then the senti score calculation for each word in a review is done. Later the negative words are taken separately and false report is calculated. If the patient is in negative mindset means recommendations are provided to change the patient mood.

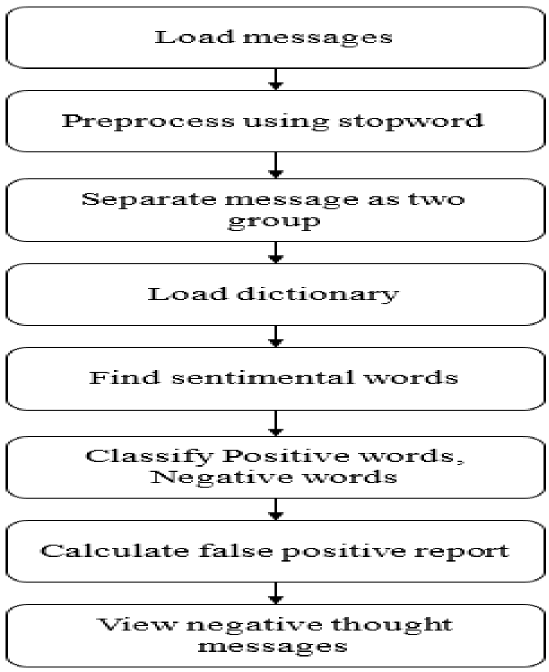

Figure 3.1 Senti health tool Architecture

3.1 SA Tools

The method SH uses the text file to analyze the feelings of the messages: "dictionary.txt",

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

"emoticon.txt", "hash tags.txt" and "ngrams.txt". Each of these files has in each line, a term or set of terms and information about the terms separated by a blank space or a colon (":").

3.1.1 Dictionary Set

The "dictionary.txt" is the same dictionary of terms in Portuguese used by Senti Strength and has 1964 terms. In each row there is a word and the emotional strength of this word indicated by a numeral ranging from −5 to 5. The higher this number, the more positive is the sense of the word and the smaller the greater is its negativity. The "emoticon.txt" contains the same 125 textual emoticons and sentimental strength used by Senti Strength tool. An example of a positive emoticon is **"(: 1"**.where **"(:"** is the emoticon and **"1"** is the positive emotional strength of this emoticon. [4] The difference between this Senti Strength dictionary and ours is that we also considered the question mark as an emoticon with a sentimental strength equal to zero ("? 0"). The file "hashtags.txt" has 6 hash tags applied in cancer groups in social networks, such as "#obrigadodeus" and "#obrigadodoador". The sentimental strength of these hashtags are separated by a blank space. The file "ngrams.txt" contains n-grams and four information in each line: an n-gram, its sentimental strength, a number 1 or 0 indicating whether the n-gram is or is not priority and another number 1 or 0 indicating whether variations of this n-gram are con-side red or not. For example, the line **"happy:4:1:0"** indicates that the n-gram **"happy"** has a sentimental strength **+4**, it is priority and variations of it (happiness, unhappiness, happily) are considered.

3.1.2 URLs

Before analyzing a message, SHC-pt.checks if there exists any URL on it. If so, all URLs are removed from the text. This is done because the URLs are formed by a set of characters without sent-mental significance.

3.1.3 Priority Sentences

Priority sentences have some of the characteristics: emoticon, hash tag, exclamation point, capitalized word or some n-gram established as a priority in the file "ngrams.txt". It is found that these specific features, when present in a sentence were able to represent the author's mood by them-selves. Thus when a message has priority sentences, only these need to be considered for the analysis of the sentiment of the whole message. Consequently we discard the non-priority sentences when there is at least one priority sentence in the message. Each priority sentence is classified separately as belonging to one of the following classes: positive, negative or neutral, as explained in the following sections. After that the message is labeled with the present class. For example, if a message has two positive priority sentences one neutral and one negative, the sentiment of this message is classified as positive. If there is a tie in the numbers of positive and negative priority sentences the message is classified as indefinite.

If at least one priority sentence is classified as positive or negative, the amount of neutral priority sentences in the message is considered zero. This is because when a person does not want to show positive nor negative sentiments in a message, he does not write any sentence with these sentiments, it stays neutral from start to end of the message. If there are no priority sentences the entire message is used to calculate the sentiment and only the dictionaries "ngrams.txt" and "dictionary.txt" are used for SA. For that each word is scored based on the scores found in these dictionaries and the message's score corresponds to the sum of words' score. Finally the message sentiment will be classified as positive if the sum is greater than zero; negative if it is less than zero and neutral otherwise.

3.1.4 Emoticons and Hashtags

To calculate the sentiment of each priority sentence it is checked if any emoticon in file "emoticons.txt" is contained in the sentence. While calculating the sentiment for the priority sentence, the presence of the emoticons on these sentence will be checked against the "emoticons.txt". If so, only these emoticons are considered to define the sense of the whole sentence. Moreover SHC-pt. follows a similar procedure with hashtags if there is any hashtag in the sentence that also occurs in "hahstags.txt", the other terms in the sentence are disregarded and only the hashtag is used to classify the sentence as positive, negative or neutral. If both emoticons and hashtags are present in the sentence only the emoticons are considered.

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

3.1.5 Question Mark

Although interrogative sentences can be classified as positive, negative or neutral by summing the scores of its terms, these sentences usually express no sentiment. Thus considering that when there is an emoticon in a sentence, only this is used in the sentiment classification, we added a question mark in the file “emoticons.txt” with sentiment strength equal to zero so that interrogative sentences are considered as having neutral sentiment. If the text contains no emoticons or hashtags included in the files “emoticons.txt” and “hahstags.txt”, the files “dictionary.txt” and “ngrams.txt” are used to classify the whole sentence.

3.2 PREPROCESSING

The techniques used for preprocessing the text documents are,

3.2.1 Stop words removal

Stop word removal is the process of removing the words that carry less weightage in the document clustering. Stop words are not predefined; the words can be defined according to the users. There is no specific rule for defining the stop words. The stop words are removed as they increase the number of features unnecessarily and also mislead and weaken the performance of the underlying clustering process. The common words like who, like, better, the, and, between, is are considered as stop words.

3.2.2 Stemming

Stemming converts inflectional or derivationally related form of word into their root words. This process reduces the number of features in the feature space. There are many techniques to perform the stemming process. Porter stemmer is a stemming method used here to remove the stop words. For example, semantically, semantics, semantical all have the common root word semant.

3.2.3 Tokenization

In tokenization, a document gets split into independent terms called tokens. These tokens are taken into consideration for selecting the feature subset. The length of the token varies according to the user from unigrams to n-grams. Single term is being considered in this project.

3. 3 FIND SENTIMENTAL WORDS

The main idea of clustering topic nodes is to first measure the common reachability of topics nodes for a given sample node set. Group the topic nodes into different clusters based on the common reachability. The fundamental assumption is that if two nodes can be grouped into one cluster, then there is a high likelihood that these two nodes can reach a certain number of common neigh bours.

3.4 CLASSIFY WORDS

Propose a novel method to effectively migrate the local influence of topic nodes to the corresponding topic-related representative nodes without running the actual clustering algorithm. The key idea of influence migration is as follows: Given a topic node, a number of random walk paths are generated starting from the topic node.

3. 5 CALCULATE MEASUREMENT

Calculate the precision and recall value based on the review sentimental analysis. Find accuracy based on the analysis result.

IV.EXPERIMENTAL EVALUATION

4.1 PRECISION

In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to the query:

$$\text{Precision} = \frac{|\{\text{relevant document}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

Precision takes all retrieved documents into account, but it can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. This measure is called precision at n or P@n.For example for a text search on a set of documents precision is the number of correct results divided by the number of all returned results. Precision is also used with recall, the percent of all relevant documents that is returned by the search. The two measures are sometimes used together in the F1 Score (or f-

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

measure) to provide a single measurement for a system. Note that the meaning and usage of "precision" in the field of information retrieval differs from the definition of accuracy and precision within other branches of science and technology.

4.2 RECALL

Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

$$\text{Recall} = \frac{|\{\text{relevant document}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

For example for text search on a set of documents recall is the number of correct results divided by the number of results that should have been returned in binary classification, recall is called sensitivity. So it can be looked at as the probability that a relevant document is retrieved by the query. It is trivial to achieve recall of 100% by returning all documents in response to any query. Therefore, recall alone is not enough but one needs to measure the number of non-relevant documents also, for example by computing the precision.

4.3 F- MEASURE

A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

$$F = 2.\frac{[\text{ Precision . recall}]}{[\text{ Precision + recall}]}$$

This measure is approximately the average of the two when they are close, and is more generally the harmonic mean, which, for the case of two numbers, coincides with the square of the geometric mean divided by the arithmetic mean. There are several reasons that the F-score can be criticized in particular circumstances due to its bias as an evaluation metric. This is also known as the measure, because recall and precision are evenly weighted.

4.4 ACCURACY

Accuracy is perhaps the most intuitive performance measure. It is simply the ratio of correctly predicted observations. In the screenshot you saw before, the two green cells are the correctly predicted observations: 2894+11750=14,644. The total number of events is 16,281, hence the accuracy is 14,644/16,281=0.899, or approximately 90%.Using accuracy is only good for symmetric data sets where the class distribution is 50/50 and the cost of false positives and false negatives are roughly the same. It can be attractive at first because it’s intuitively easy to understand, however. You should not rely on it too much because most data sets are far from symmetric. Example: you are building a model which predicts whether a device is defective. The class distribution is such that 6 in 1000 devices is truly defective (positive). A model which simply returns “negative” – i.e. not defective – all the time gets it right 99.4% of the time and therefore has an accuracy of 0.994, when in fact it never correctly identifies a defective device.

Table 1. Performance metrices values for obtained cancer reviews

Precision	66.666
Recall	87.69
F-Measurre	82.498
Accuracy	76.47

4.5 SENTI SCORE CALCULATION

Input (pre-process review)

Such great restaurant. Really friendly server, but not high price. Tidy and delicate place, really tasty food.

Output:

Such great restaurant.

Really friendly server, but not high price.

Tidy and delicate place, really tasty food.

Step 2: output of step 1 store into array and compare each line to dictionary (SD, sdd, nd).
Example

Input:

Such great restaurant.

Output:

Such- level 3- give the score to 2

Great- positive word-give score to 1

Restaurant- product feature-not give any score.

Score is=2*1=2.

Input:

Really friendly server, but not high price.

Output:

Really-level 2-give score to 4

Friendly-positive word-give score to 1

Server-product feature-not give any score.

Not-negation word-give score to -1

High- "*high*" is a negative sentiment word so give score to −1, because "***but***" is a twist conjunction after a positive word (friendly).

Score is 4*1+ (-1*-1) =5

Input:

Tidy and delicate place, really tasty food.

Output:

Tidy-positive word-give score to 1

Delicate –positive word-give score to 1

Tasty-positive word –give score to 1

Really-level2-give score to 4

And-used to conducting these three positive sentiment words so the score is

1+1+1*4=6

Final score is=1/number of clauses (2+5+6)

Number of clauses is 3.

Then sentiment score is 10/1+e (final score)-5.

V. CONCLUSION AND FUTURE WORK

Existing tools for Sentiment Analysis have a very low accuracy when used in web texts of the cancer context, written in Portuguese language. SHC-pt is a tool for Sentiment Analysis to work at sentence level using a lexicon and heuristics to analyze people's texts involved with cancer. These texts were collected from user posts on Facebook's Brazilian cancer groups. Unlike other tools SHC-pt in this work considers the author of messages as the target of analysis. Many experiments were done with two online social groups from Facebook.

The results show that SHC-pt outperformed all other tools addressed in all experiments. This was possible because, a specific lexicon for cancer domain were used. This lexicon considers terms, semantic emoticons, hashtags and n-grams that contributed to the good result of SHC-pt. In the proposed work the SHD-pt for diabetes will be developed with specific lexicon for diabetes domain. Considering Facebook and Twitter comments.

VI. REFERENCES

[1] Abeed sarker., Azadeh Nikfarjam., Graciela H. Gonzalez., (2016) 'Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts' 2016.Sciencedirect, pp 148 –158

[2] Abeed sarker., Graciela H. Gonzalez., (2016) 'Portable automatic text classification for adverse drug reaction detection via multi-corpus training' 2016.Sciencedirect, pp 196 – 207

[3] Abbasi A., Dhar M., Hassan A., (2014) 'Benchmarking twitter sentiment analysis tools'. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation, Iceland ,pp 823–829

[4] Abdullah N., Basar S., Roffeei S., (2015) 'Seeking social support on facebook for children with autism spectrum disorders'2015.Sciencedirect, pp 375–385

[5] Bifet A., Frank E (2010) 'Sentiment knowledge discovery in streaming data', In: Discovery Science, 2010. Springer, pp 1–15

[6] Brynielsson J., Johansson F., Jonsson C., Westling A (2014) 'Emotion classification of social media posts for estimating people's reactions to

communicated alert messages during crises', Security Inf 3(1):1–1

[7] Camilo junior G., Ramon Gouveia Rodrigues., Rafel Marques das., (2016) 'SentiHealth-Cancer:A Sentiment analysis tool to help detecting mood of patients in online social networks training' 2016.Sciencedirect, pp 80–95

[8] Chun SA., Geller J., Ji X (2012) 'Epidemic outbreak and spread detection system based on cancer data'. In: Proceedings of the First international conference on Health Information Science, Beijing, China, pp 152–163

[9] Dekhil M., Ghosh R., Hsu M., Liu B., Zhang L (2011) Combining lexicon-based and learning-based methods for Cancer sentiment analysis

[10] Kiritchenko S., Mohammad SM., Zhu X (2014) 'Sentiment analysis of short informal texts', J Artif Intell Res:723–762

[11] Lee L., Pang B., Vaithyanathan S (2002) Thumbs up?: 'sentiment classification using machine learning techniques'. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing-vol. 10, pp 79–86

[12] Lillian Lee., Pang Bo (2008), 'Opinion Mining and Sentiment Analysis', Foundations and Trends in Information Retrieval, Vol. 2, No, 1/2 pp. 1-135.

[13] Liu B., Zhang L (2012) A survey of opinion mining and sentiment analysis. In: Mining Text Data, pp 415–463