

A Survey on Sentiment Analysis towards Health Care

Roopal Mamtora¹, Dr. Lata Ragha²

¹(Computer Engineering Department, SIES-GST, Nerul, Navi Mumbai, India)

²(Computer Engineering Department, FCRIT, Vashi, Navi Mumbai, India)

Abstract: Sentiment Analysis can be termed as the process of estimating or determining opinion of an individual/group of individuals into various classes. At times it is referred as Opinion Mining by various researchers and practitioners. This paper presents a short survey of the work done in the field of sentiment analysis. It tries to focus on application of sentiment analysis in different domains, some of which are well explored and the others which are less explored like health care.

Keywords: Emotion Recognition(ER), Health Care, Opinion Mining(OM), Sentiment Analysis(SA).

I. Introduction

In today's technological world, knowledge is available on the click of a mouse. People are interested in sharing data, view points, getting solutions to problems by various new and effective means like Wiki, blogs and discussion forums. The exponential rise in sharing knowledge has added unstructured and non-relevant data to the web which needs to be analyzed and made relevant to the person in need. Today, even if a person wants to buy any product, the normal approach is to look at related reviews, blogs, forums, ratings and come to a quick decision to buy it or not. People have gone to the extent of getting information about diseases, finding solutions to their problems & sharing their experiences about personal health (health forums). This emphasizes that the web has become an integral part of life for every aspect. One way to harness the important and crucial data is to apply Opinion Mining (OM)/Sentiment Analysis(SA). Generally the terms SA and OM are used as aliases of each other. SA is the process of identifying the subjectivity and polarity of the opinions[1]. This field has grown far from analyzing public opinion about every possible topic that can be thought of like products, movies, politics, finance, health care etc.

A survey by Ganeshbhai et al. [2] suggests that practitioners and researchers have classified SA into into different granularity levels, such as Document-level, Sentence-level and Aspect-level. From top to bottom it can be termed as specialization & from bottom to top it can be taken as generalization. Talking with respect to specialization, document level breaks down to sentence level which further is broken into aspect level. For example, talking about the document which contains reviews on any of the products can be analyzed as document level SA, which could tell us if the product is worth buying or not. This document level SA is a result of culmination of all sentence level SAs. In turn the sentence level SA is derived by considering various features of the products, that contribute towards the aspect level SA. Aspect level SA is alternatively referred as feature based SA. Aspect level SA can give crucial information about which features of a product are liked by the customers and which are disliked. Such important piece of information can be utilized for decision making in future & capture the competitive edge.

SA has various sub-domains like Emotion Recognition(ER), Building Resources and Transfer Learning. Each of the sub-domains is very promising area in itself and has a great scope for research. ER tries to figure out the basic or complex emotions like happy, sad, fear, frustration, confused, disgust, etc. Building Resources forms the base for ER / SA /OM to be performed. It builds the knowledge base for application of techniques to detect ER / SA /OM. Transfer Learning is cross-domain application which transfers the learning obtained from one domain to the other domain.

This paper gives brief about the journey of SA being applied to enormous application areas to ease decision making for the prospective customers, people trying to seek information about medical problems, even at times getting solutions or seeking empathy for their health issues.

The rest of paper is organized as follows: Section II discusses Literature Review, Section III describes about Research Issues and Challenges in context of medical SA and Section IV Conclusion.

II. Literature Review

Sentiment analysis has been applied to varied domains such as political debates, reviews for cars, hotels, computers, books, customers, market prediction, social advertising, box-office prediction, news, blogs, twitter messages, e-health, etc. using Chinese, English, Spanish languages. Researchers & practitioners have used supervised, semi-supervised, unsupervised learning techniques and even a combination of supervised and

unsupervised approaches for analyzing sentiments. The most commonest techniques utilized were lexical approach, support vector machine(SVM), k-nearestneighbors,bayesian networks, convolutional neural networks(CNN), hidden markov models (HMM), and artificial neural networks (ANN)[3-8].

There is a constant rise in number of people who resort to internet for seeking health related information.The information could be about some diet related issues or trying to know symptoms of a disease ,first aid cure for problems , knowing the adversities of drugs being consumed.It could also be about checking the validity of the treatment or even in the worst case searching reviews given by people suffering from some crucial disease.This culture seems to be prevalent among the age group of 25-45. Approximately 63% of people look for some or the other kind of information about specific medical problems, whereas 47% search for treatments or cures on the problems.With this huge percentage of people , it becomes important to analyze the contents available and determine sentiments embedded in them. By human nature , sentiments have a great impact over an individual's thought process[9].This gives a broad scope to sentiment analysis to be applied to health care.

Lot of reasearchers have put in efforts in the\is field. The authors Shweta Y et al. [9], collected users' posts in medical blogs(forums) constrained to popular domains like depression, anxiety, asthma, and allergy. They proposed a method which used a deep convolutional neural network(CNN) to predict the possible medical sentiment category for both 'medication' and 'medical condition' of the patients who have posted the blogs .The categories considered for medication were-Effective, Ineffective ,Serious adverse effects and for medical condition were-Exist, Recover, Deteriorate. Their proposed method with CNN gives better performance than the considered baseline methods like SVM,Random forest & MLP(Multi-layer perceptron).

Jianbo Yuan et al.[10],in their work collected data from the medical forums of potential Autism spectrum disorder(ASD) patients. They converted hand written medical forms into digital format (tif) by the scanning process and further preprocessed it. This digital format was then used for handwriting recognition and then classification using SVM to detect ASD. They used lexical features, LDA model and doc2vec features for extracting discriminative features.

R. G. Rodrigues et al.[11] Collected posts from two cancer communities in facebook which were in portuguese. The posts were analyzed by sentiment analysis tools that support the Portuguese language (Semantria and SentiStrength) and were also converted into English to be used with SA tools like- AlchemyAPI and Textalytics.The former posts were also analyzed by the proposed tool (SHCpt) to detect cancer patient's mood which considers the sentence level using a lexicon and heuristics for SA.

Keumhee Kang et al[12] in their work , have proposed a method to identify the users with depressive moods. They performed unimodal analysis to determine the hidden user's moods from text, emoticon, and images. A learning based text analysis was used to extract mood from text,a word-based emoticon analysis was done to extract moods from emoticons and a SVM based image classifier was used for mood extraction from images. The individually extracted moods from the respective analyses were integrated into a mood and again aggregated for every day, which allowed for continuous monitoring of user's mood trends.

Maria et al[13] and others , collected corpus from twitter specifically from the diabetes domain .They proposed a three step process to detect sentiments from these corpuses: pre-processing ,detecting aspects by means of the semantic annotation technique, calculating polarity of each aspect found on the SentiWordNet (SWN) lexicon.

Shimrit Fridenson et. al[14],proposed to determine basic and complex emotions in autistic children using multimodal approach.They first examined emotion recognition unimodally through faces, voices, and body language, and then went on to test the same multimodally through integrative scenarios with no verbal content cross-culturally in Israel, Britain, and Sweden.The authors were able to prove that children with Autism spectrum condition(ASC) showed ER deficits in all three modalities and their integration as compared to the normal children.

T. Ali and others [15] , attempted the problem of sentiment analysis on medical forums dedicated to Hearing Loss (HL). They performed experiments using three different classifiers: Naïve Bayes(NB), SVM and logistic regression (LR).They used WEKA tool with 10-fold cross validation classifier for this purpose & found that the performance of logistics regression was the best with the feature set that they selected.

Table no 1: Comparative Study of SA application to health care

Ref	Task	Year	Method	Modality	Data Source		Accuracy			
							Precision	Recall	F1-score	F2-score
[9]	sentiment analysis	2018	CNN	Unimodal(T)	data collected from the 'Patient.info' forum containing users' original posts	Medical condition	0.68	0.60	0.63	
						Medication	0.86	0.77	0.82	
[10]	autism detection	2017	SVM	Unimodal(T)	medical forms	LDA	0.58	0.84		0.77
						Doc2Vec	0.65	0.91		0.84
[11]	sentiment analysis	2016	lexicon and heuristics	Unimodal(T)	posts from cancer community on Facebook.				54.8 (avg. F1)	
[12]	sentiment analysis	2016	SVM	Multimodal (T+E+I)	Tweets from twitter		0.86	0.88	0.87	
[13]	sentiment analysis	2017	Aspect-level approach	Unimodal(T)	corpus from Twitter		0.82	0.81		
[14]	emotion detection	2016	Proposed ER battery	Multimodal (V+A)	School going children					
[15]	sentiment analysis	2013	NB,SVM,LR	Unimodal(T)	Posts from 3 health forums	NB	0.656	0.65	0.644	
						SVM	0.661	0.641	0.625	
						LR	0.649	0.645	0.641	

T:Text, V:Video, A:Audio, E:Emoticon, I:Image

III. Research Issues and Challenges

Medical SA seems to be a promising field for analysis of very critical information, but it is much more complicated than the standard SA. There are a lot of additional hurdles to be overcome as compared to the former field of work. A few of them are listed as follows:

- Medical data is not readily available since it involves the medical problems of an individual, which makes it very confidential in nature. So people do not tend to put the exact and clear form of information. Hence the data available is not adequate.
- Data here is very domain specific as it involves medical jargons which do not work efficiently with the lexicons available for standard SA. So there is a need to build specific lexical resources for medical SA.
- Medical forums mostly have data with implicit meanings which creates problems in classification.
- Data should be taken from a popular source which can ensure the validity of data to some extent as people always tend to take information or post the same on a platform which is popularly used.
- Preprocessing of the data is time consuming, as some times the handwritten medical documents need to be converted to images or other forms which can be used for processing.
- Most researchers have performed SA for English language, whereas vast information on the internet is available in languages other than English.
- It is not easy to transfer learning or perform cross domain knowledge transfer, as the domain is very specific.
- A lot of work has been done considering single modalities, most commonly text analytics. So there is a huge scope for multimodality to be considered.

IV. Conclusion

From the comparative study, it is observed that the task of sentiment analysis in context to health care has been mostly performed with a single modality using forums or posts from online communities. These could be forums related to health care, specific problems like diabetes, adverse drug effects, depression, hearing loss, cancer, autism etc. SA done with aspect level approach outperforms the other baseline methods considered like lexicon, heuristics, SVM, NB, LR, CNN. It is also dominant that SA with combined multimodalities give great results even with the most simple method like SVM. Each modality adds certain additional information to obtain better results. Therefore, multimodality in health care remains an open research problem.

References

- [1]. R.Nithya , D. Maheshwari, Sentiment Analysis On Unstructured Review , Proc. 2014 *International Conference on Intelligent Computing Applications (ICICA'14)*,pp.367-371.
- [2]. Ganeshbhai, S.Y. and Shah B.K , Feature Based Opinion Mining: A Survey ,Proc. *Advance Computing Conference (IACC)*,2015 IEEE International ISBN: 978-1-4799-8046-8,pp. 919 - 923.
- [3]. Walaah Medhata,Ahmed Hassanb and Hoda Korashyb, Sentiment analysis algorithms and applications: A survey , *Elsevier Science,Ain Shams Engineering Journal , Volume 5, Issue 4*,2014,pp.1093–1113.
- [4]. Kumar Ravia and Vadlamani Ravia, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications , *Elsevier Science,Journal Knowledge-Based Systems , Volume 89 Issue C*,2015 ,pp. 14-46.
- [5]. Jasmine Bhaskar, K. Sruthi, Prema Nedungadi, Hybrid Approach for Emotion Classification of Audio Conversation Based on Text and Speech Mining , *Elsevier Science,Procedia Computer Science, Volume 46*, 2015, Pages 635-643
- [6]. Amir Zadeh; Rowan Zellers; Eli Pincus; Louis-Philippe Morency, Multimodal Sentiment Intensity Analysis in Videos:Facial Gestures and Verbal Messages , *IEEE Intelligent Systems,Year: 2016, Volume: 31, Issue: 6*, Pages: 82 – 88.
- [7]. Soujanya Poria , Erik Cambria , Newton Howard , Guang-BinHuang , AmirHussain, Fusing audio, visual and textual clues for sentiment analysis from multimodal content ,*Elsevier Science,Neurocomputing , Volume 174, Part A*, 22 January 2016,Pages 50–59.
- [8]. Verónica Pérez Rosas, Rada Mihalcea, Louis-Philippe Morency, Multimodal Sentiment Analysis of Spanish Online Videos , *IEEE Intelligent Systems,Year: 2013, Volume: 28, Issue: 3*,Pages: 38 -45.
- [9]. Shweta Yadav, Asif Ekbal, Sriparna Saha and Pushpak Bhattacharyya , Medical Sentiment Analysis using Social Media: Towards building a Patient Assisted System,Proc. *Language Resources and Evaluation Conference 2018*.
- [10]. Jianbo Yuan, Chester Holtz, Tristram Smith and Jiebo Luo,Autism spectrum disorder detection from semi-structured and unstructured medical data,*EURASIP J Bioinform Syst Biol. 2017 Dec; 2017: 3*.
- [11]. R. G. Rodrigues, R. M. das Dores, C. G. Camilo-Junior,T. C. Rosa, SentiHealth-Cancer: a sentiment analysis tool to help detecting mood of patients in online social networks, *International Journal of Medical Informatics, vol. 85, no. 1*, pp.80–95, 2016.
- [12]. Keumhee Kang,Chanhee Yoon,Eun Yi Kim, Identifying Depressive Users in Twitter Using Multimodal Analysis,2016 Proc. *International Conference on Big Data and Smart Computing (BigComp)*,Electronic ISSN: 2375-9356.
- [13]. María del Pilar Salas-Zárate, José Medina-Moreira, Katty Lagos-Ortiz, Harry Luna-Aveiga, Miguel Ángel Rodríguez-García, and Rafael Valencia-García , Sentiment Analysis on Tweets about Diabetes: An Aspect-Level Approach, *Computational and Mathematical Methods in Medicine Volume 2017 (2017)*, (Article ID 5140631), 9 page.s
- [14]. Shimrit Fridenson-Hayo , Steve Berggren, Amandine Lassalle, Shahar Tall , Delia Pigat , Sven Bölte, Simon Baron-Cohen and Ofer Golan, Basic and complex emotion recognition in children with autism: cross-cultural findings ,(Molecular Autism -2016) 7:52.
- [15]. T. Ali, D. Schramm, M. Sokolova, and D. Inkpen, Can i hear you? Sentiment analysis on medical forums, Proc. *International Joint conference on Natural Language Processing*, pp. 667–673, Nagoya, Japan, October 2013.