A Survey on Cyber Bullying Detection on Social Media based on Denoising Auto-Encoder

Umme Khadija & Dr. Ruksar Fatima
[1]*M. Tech Student KBNCE, Kalaburagi*
[2]*CSE Dept., KBNCE, Kalaburagi*

Abstract:

Opinion mining or sentiment analysis is considered as an important application of NLP (Natural Language Processing). Opinion mining is extracting the views that people express online. Those Websites which permits social interaction and collaboration can be considered as social media site, including networking sites such as Facebook, MySpace, and Twitter. Such sites offer today's youth a platform for amusement, entertainment, thrill, correspondence and communication with friends and furthermore have developed radically and exponentially as of late. This is the reason, there are various side effects, as cyberbullying has emerged as a serious issue afflicting children, adolescents and young adults. Machine learning techniques have conceivable ability to make automatic detection of bullying messages in social media, and this could develop a healthy and comparatively safe social media environment.

Keywords

NLP, cyberbullying, Social Network, Mining, collaboration.

1. INTRODUCTION

Opinion extraction, sentiment mining, affect analysis, analysis of emotion, review analyzing or mining, etc. are numerous comparative names with somewhat unique errands and with slightly different tasks. However, they are presently considered within the shades of opinion mining or sentiment analysis. This can be considered as an important application of NLP (Natural Language Processing). Some of social media sites, includes social networking sites such as Facebook, MySpace, and Twitter; video sites such as YouTube; photo sharing such as Flicker, Photobucket, or Picasa; gaming sites and virtual worlds such as Kaneva, Club Penguin, Second Life, and the Sims; live casting such as Ustream or Twitch; instant messaging like Google talk, yahoo messenger or skype and blogs. Authors in and [12] use reciprocally the terms Big Social Data and Social Big Data to refer about the data which is generated by social media. To show the tremendous measure of

data that is generated by social networking, affirms that via Facebook (the most popular social media [11]), 10 million photos are getting uploaded almost every day. focus attention to highlight the aspect that more than 250 million tweets are sent by Twitter every day consistently as well, also 3000+ photos are getting uploaded across Flicker every minute consistently without overlooking above 150 million web blogs posted daily. The expansion of Social Big Data is extremely valuable in numerous fields such as sociology, human science, psychology, governmental issues, politics and the very important commercial area. [16]

Nonetheless, social media over and above have some consequential side effects as cyberbullying, that might have intense unfavorable impact and can transform the life of people, especially the children, adolescent and teenagers. Cyberbullying detection can be administrated from social media as it can be particularized as a supervised learning problem. A classifier is initially trained on a cyberbullying corpus that is labeled with mark by humans, and then later the learned classifier is then used as a result to recognize & perceive bullying message. Three categories of information are commonly used inclusive of user demography, text, and social network features of cyberbullying detection . Since the text content substance is the most definitely dependable as well as reliable, our work here spotlights on the text-based content cyberbullying detection.

2. SENTIMENT ANALYSIS APPROACHES

Sentiment Analysis or Opinion Mining intends to adjudge the viewpoint stance of a person with respect to any subject. There are two fundamental approaches used for opinion mining, those are:
• Machine Learning
• Semantic Orientation.
The machine learning approach is said to be belonging to supervised classification approach. This approach is more precise as each of the classifiers is initially trained on a collection of representative data which is known as corpus. In this way, it is called "supervised learning". In a machine (supervised) learning based classification, two sorts of documents

are required: training set as well as test set. A training set is utilized to prepare and learn the classifier and a test set is utilized to test the performance on execution of the automatic classifier. A huge number of machine learning techniques are available to access which classifies the opinions. Machine learning techniques like Naïve Bayes, Maximum Entropy (ME) and Support Vector Machines (SVM) have achieved extremely significant success in text categorization.

3. PROCESS

Fig I. shows the procedure followed:
1. Tweets for various cases acts as input
2. The tweets are collected from the twitter using OAuth API.
3. For each of the tweets the stop words are removed and clean tweets are also obtained.
4. Each tweet is then compared against the words belonging to 5 different categories.
5. Compute the probability using Naïve Bayes formula.
6. Compute the contingency
7. Compute the Reduced contingency
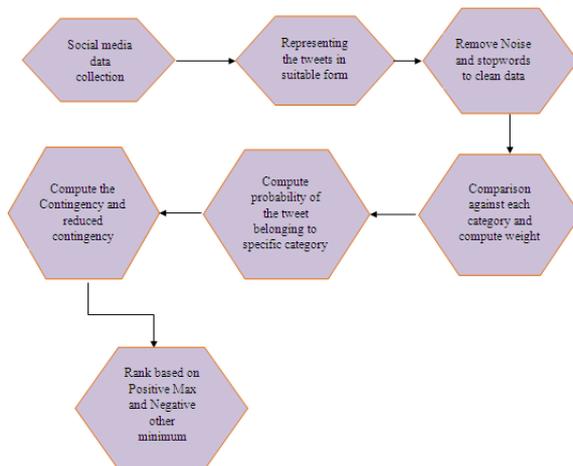8. Rank the Tweets based on maximum positive and negative minimum

Fig I. Process to detect bullying messages

4. LITERATURE REVIEW

A. Survey on Sentiment Analysis:

Paper [40], Nasukawa et al. demonstrates an approach to sentiment analysis in which they extract sentiments for particular subjects associated with its polarities i.e. negative or positive from a document, in spite of classifying the whole document. Initial work on sentiment analysis focused on identifying polarity of reviews of product Opinions and movie reviews from IMDB (Internet Movie data base) at entire document level [7]. Later work handles sentiment analysis at sentence level. Recent studies are focus was shifted from sentence level to phrase-level [2] and short-text forms in response to the popularity of micro blogging services such as Twitter [6,1,3,5]. In , Ding et al. proposed an effective method for identifying semantic orientations of opinions expressed by reviewers on product features. In sentiment analysis, Pandey et al. presents a system, which can extract micro messages relevant to any specific topic from a blogging service such as Twitter and then analyze the messages to determine sentiments they carry and to classify them as neutral, positive or negative.

B. Survey on social media big data:

According to indication in [12], the Social Big Data, produced represent all the data as well as information generated through the social media. This data is then perceived by: the substantial volume, the commotion that can be propagated (spam) and the dynamic characteristic property (the frequent changes day by day) [16]. They can likewise be perceived by an arrangement set of connections or links (due to connections between users), a structure or frame which is not aligned pattern in nature (as a result of the length of messages which is required by particularly some kind of micro-blogging, the existence of spelling inaccuracies or other) and the absence of fulfillment (as a result of satisfying the user requirements for privacy of their data) [15]. The mentioned attributes of Social Big Data qualify them unconventionally discrete from other version data on which elementary data mining techniques are enforcedly applied [16]. For mining such sort of data, various research issues, methodologies, approaches, procedures and techniques are derived.

C. Survey on Effect of Social Media

Sentiment Analysis is used for the prediction of polarity of textual data into positive, negative and neutral classes. This textual data can be gathered from social media (e.g. twitter). Following Table 1[9] shows how social media creates its impact in various situations.

D. Survey on Cyberbullying

To originate a productive potent efficient practical cyberbullying identification model, we suggest expanding on discoveries within the psychology community. There are various reviews investigating the psychological dimensions of social collaborations

that can be used to determine and analyse the cyberbullying risk elements, risk factors, risk aspects, which a model ought to consider. A large portion of the work in determining bullying among youth has concentrated on traditional bullying, conventional tormenting, or cyberbullying by the means of versatile mobile or visit chat-based scenes, e.g., [10]. Previous contributions have concentrated different aspects of bullying and cyberbullying in their studies, e.g., whether guardians' viewpoint of adolescents' online conduct is causational with teenagers' vulnerability to cyberbullying [10], probabilities of exploitation and victimization and passionate effect as well as emotional impact in light of age and gender, and measuring the correlation between seriousness of online hostility and the quantity of spooks or bullies included . While the outcomes about pervasiveness and determinants of cyberbullying change in the psychology literature, there are some critical patterns and regions of understanding [13] among these outcomes.

5. CONCLUSION

Text mining is likewise alluded as textual data mining, which can be generally considered as text analytics. It refers to the process of deriving insights from text, include task such as sentiment analysis which suffers from problem of high dimensionality. Studies showed that machine learning techniques have conceivable ability to make automatic detection of bullying messages in social media, and this could develop a healthy and comparatively safe social media environment.

6. REFERENCES

[1] A. Agarwal, B. Xie, I. Vovsha, O. Rambow and R. Passonneau, "Sentiment Analysis of Twitter Data," In Proceedings of workshop on languages of social media, 2011. pp. 30-38.

[2] A. Agarwal, F. Biadsy, and K. R. Mckeown. "Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams." Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2009. pp. 24-32.

[3] A. Go, R. Bhayani, L. Huang, "Twitter sentiment classification using distant supervision", CS224N Project Report, Stanford, 2009. pp. 1-12.

[4] A. Harb, M. Plantié, G. Dray, M. Roche, F. Trousset, and P. Poncelet. "Web Opinion Mining: How to extract opinions from blogs?" In Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology, ACM, 2008. pp. 211-217.

[5] A. Pak, and P. Paroubek "Twitter as a corpus for sentiment analysis and opinion mining," Proceedings of the seventh International Conference on Language Resource and Evolution (LREC'10), 2010. pp. 19-21.

[6] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. "From tweets to polls: Linking text sentiment to public opinion time series." ICWSM 11, no. 122-129 ,2010. pp. 1-2.

[7] B. Pang, L. Lee, and S. Vaithyanathan. "Thumbs up? : sentiment classification using machine learning techniques."Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002. pp. 79-86.

[8] C. Whitelaw, N. Garg, and S. Argamon "Using appraisal groups for sentiment analysis," presented at the Proceedings of the 14thACM international conference on Information and knowledge management, Bremen, Germany, 2005.

[9] D. M. Haughton, J. J. Xu, D. J. Yates, X. Yan "Introduction to Data Analytics and Data Mining for Social Media Minitrack", 49th Hawaii International Conference on System Sciences, 2016.p. 1414.

[10] D. M. Law, J. D. Shapka, and B. F. Olson. "To control or not to control? Parenting behaviours and adolescent online aggression." Computers in Human Behavior 26.6,2010 pp.1651-1656.

[11] D. Quinn, C. Liming, and M. Maurice "Does age make a difference in the behaviour of online social network users?"Internet of Things (iThings/CPSCom), 2011 International Conference on