THINK INDIA JOURNAL

ISSN:0971-1260
Vol-22- Issue-14-December-2019

A Framework On Text Mining And Structural Data Mining Techniques For The Detection Of Grievances In Social Media

1. B. Roopadevi 2. S.Swapna 3.P.Sindhuja

INTRODUCTION

The social nuisance is an inhuman act of the individual causing damage to the whole community and burdens the society which is punishable by law. Structuraldata mining followed by entiment analysis on online social media helps to detect the grievance patterns on texts. Taking on social networking sites, enables users to post their viewsand convey information about the author. Electronic media allows extraction of information and data from various databases which combines solutions from various fields. Aggregating and integrating the data to judge the opinion uses geospatial analysis which establishes correlations between the posts and crimes in various cities and towns.

Sentiment analysis techniques has to be conducted to analyse the vocabulary and intensity of grievance of a post of a particular location which reveals the crime rate of a location in real-time and helps to detect the patterns of crime. When analysing data, natural languageprocessing techniques are used to process the unorganised data. They have to be refined by the experts making it clear. For the data analytics we employ machine learning techniques to create a traditional programming on the data automation. We employ machine learning algorithms for

1. Representation of knowledge including datasets, rules, decisiontrees, vector machines etc.

2. Evaluation must be done on candidate programs for accuracy, prediction, probability, cost, margin etc

3. Optimization of programs generated for the process of searching.

BACKGROUND WORK

Sentiment analysis requires NLP, Text analysis to extract information from social sites, determiningauthor's attitude and contextual popularity of the text. Geographically analysing crime related tweets/posts identify crime prone areas and used data mining techniques to study and detect the data. Measuring emotions in online texts is done by opinion mining by NLPand ML techniques for the automation of sentiments in the knowledge from different resources. Adopting a dictionary based approach for determining sentiments uses a new approach with 3 dimensions of valence, arousal and dominance. Matchingtechniques uses mapping of data, pattern matching etc to clearly identify the sentiment in the text. The methods employed are arithmetic mean and normal distribution to calculate mean valence and arousal. Then the values have to be plotted in a 2D circumflexemotional model to determine the position within the model.Video-to-text processing, image-to-text processing,

ISSN:0971-1260
Vol-22- Issue-14-December-2019

and data from various online sources would also help improve accuracy. This type of study would help with informing others of the crime pattern both within and around their location, ultimately assisting them with staying in a safe zone by monitoring various social media outlets.

The term sentiment classification is defined as detecting sentiment polarity ofthe subjective sentences. This sentiment classification is also divided into two categories: Binary sentiment classification and multi-class sentiment classification.

i)Binary sentimentclassification involves classifying sentiments either positive or negative.

ii)Multi-classsentiment classification involves classifying sentiments into one of five categories: strong positive, positive, neutral, negative and strong negative.

The most common machine learning techniques used for sentiment classificationinclude naive Bayes, maximum entropy, and support vector machine. Mostsentiment analysis algorithms use simple terms to express sentiment. However thecultural factors, linguistic nuances, and differing contexts prevent researchers fromdrawing the sentiment accurately.

ANEW Based Approach:

The ANEW is being developed to provide aset of normative emotional ratings for a large number of words in the English language. This was developed to aid researchers when studying emotions; it is often used to determine a tweet's sentiment. Siddharth and Dr. Healey adopted a dictionary-based approach fordetermining the sentiment of tweets. They used the ANEW dictionary to provide pre-existing,normative emotional ratings for 1034 words along the three dimensions ofvalence, arousal and dominance. They used an independent matching technique to mapall of the words in a tweet that were found in ANEW. They used two approaches (thearithmetic mean and normal distribution) to calculate both the mean valence and arousal.

Constructing and Validating a Valence-Aware Sentiment Lexicon:

A Human-Centered Approach Manually created (much less, validating) a comprehensive sentiment lexicon is a labor intensive and sometimes error prone process. There is a great deal of overlap in the vocabulary covered by such lexicons; however, there are numerous items unique to each. We begin by constructing a list inspired by examining the well-established sentiment word-banks .To this, we next incorporate numerous lexical features common to a sentiment expression in microblogs, including a full list of emoticons10 (for example, “:-)” denotes a “smiley face” and generally indicates positive sentiment), sentiment-related acronyms and initialisms11 (e.g., LOL and WTF are both sentiment laden initialisms), and commonly used slangs with sentiment value (e.g., “hmm”, “wah” and “ooh”). This process provided us with over 9,500 lexical feature candidates. Next, we assessed the general applicability of each feature candidate to sentiment expressions. We used a wisdom-of-the-crowd13 (WotC) approach (Surowiecki, 2004) for the sentiment valence (intensity) of each context-free candidate feature. We collected intensity ratings on each of our candidate lexical features from ten independent human raters (for a total of 90,000+ ratings). Features were rated on a scale from “[–4]

ISSN:0971-1260
Vol-22- Issue-14-December-2019

Extremely Negative" to "[4] Extremely Positive", with allowance for "[0] Neutral (or Neither, N/A)". left us with just over 7,500 lexical features with validated valence scores that indicated both the sentiment polarity (positive/negative), and the sentiment intensity on a scale from –4 to +4. For example, the word "okay" has a positive valence of 0.9, "good" is 1.9, and "great" is 3.1, whereas "terrible" is –2.5, the frowning emoticon ":(" is –2.2, and "sucks" and "sux" are both –1.5. This standard list of features, with associated valence for each feature, comprises VADER's sentiment lexicon

Deep Learning for Sentiment Analysis:

Richard, Alex, Jean, and Jason introduced the Recursive model, a state-of-the-art in sentiment analysis. Theyalso introduced both the Recursive Neural Tensor Networks (RNTN) and the StanfordSentiment Treebank. The Treebank includes fine-grained sentiment labels for over200,000 phrases in the parse trees of over 11,000 sentences. When the RNTN model istrained on the new Treebank, it outperformed all previous methods on several metrics.

This approach follows the multi-class sentiment classification; it predicts five sentimentclasses: very positive, positive, neutral, negative and very negative. The sentimentprediction's accuracy can reach 80.7%.

To detect sentiment valences of words, the following methods can be used in combinations:

- Senti Word Net
- Emoticons
- AFNN
- Bayesian Networks

By using the combinations of above methods valence or polarity of a subjective sentences can be generated, which can be used to detect the reliability of words

CONDUCT SENTIMENT ANALYSIS

Sentiment analysis is used to determine an author's attitude with respect to either aparticular topic or a document's overall contextual polarity. As stated earlier, effort wasgiven to evaluate the sentiment which existed in messages using two sentiment analysistechniques, namely ANEW based technique and Deep Learning model.

In ANEW based technique, effort was made in mapping every term from ANEWto its equivalent in the tweet. This was then applied by Porter Stemming to improvethe mapping. Selection of matching words that existed in both ANEW and the tweet wasgiven further consideration. The mean valence μv and mean arousal μa for each selectedANEW term is fetched to compute average of the valence and arousal of all the ANEWterms.

ISSN:0971-1260
Vol-22- Issue-14-December-2019

CONCLUSION

Analysis of emotions and sentiments expressed by people on social networks and various web sources is a challenging task. Sentiments are the biggest source to analyze the reviews of any product and services. Sentiment analysis plays a biggest role in finding the emotions and sentiments in internet. Sentiment analysis is a useful technique for people whom they can express their views about any topic, text, content or product. Sentiment analysis will analyze these expressions of social networks by using various techniques and tools. Sentiment analysis classifications are divided into various levels.
i) Document level can check the whole document
ii) Sentence level analyze the sentence of a document
iii) Semantic level care about people opinions of social media about any product or any content related to that product.
iv) Sentiment lexicon focusing on the opinion words.
Sometimes words have neutral meaning neither positive nor negative inquiries and large organizational data are the sources of data useful to express various feelings and data. Comments, irregular opinions, fake comments, Sarcasm challenges related to social media like twitter and facebook which is a never ending and challenging process. Due to these requirements sentiment analysis has its significance and produces tremendous results in solving such tough problems. Several problems & challenges are still unsolved in the field of machine learning, however sentiment analysis techniques and approaches help to solve various problem domains in the machine learning. These solutions can be very useful to identify the grievances in social networking sites.

BIBLIOGRAPHY

1.Java, X.Song, T.Finn, and B.Tsang. Why We Twitter; Understanding Microblogging Usage and Communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD Workshop 2007 on Web Mining and Social Network Analysis, pages 56-65, 2007.

[2] J. Bollen, A.pepe, and H.Mao. Modelling public mood and emotion: Twitter sentiment and socio-economic phenomena. arxiv:0911.1583v0911 [cs.CY], pages 09-19,2010

3.SpotCrime.com. (2014, April 1). In Wikipedia, The Free Encyclopedia. Retrieved

4. B. Pang and L. Lee. A Sentimental education: Sentiment Analysis using subjectivity summarization based on minimum cuts. In Proceedings 42nd ACL, pages 271-278, 2004.

5. H. Tang, S. Tan and X. Cheng.A survey on sentiment detection of reviews. Expert Systems with Applications, 36(7):10760-1073, 2009.

6.Cataldi, M., Di Caro, L., Schifanella, C.: Emerging Topic Detection on TwitterBased on Temporal and Social Terms Evaluation. In: Proceedings of the TenthInternational Workshop on Multimedia Data Mining, MDMKDD 2010 (2010).

ISSN:0971-1260
Vol-22- Issue-14-December-2019

7.D. Zeng and H. Chen, R. Lusch, and S. H. Li, “Social Media Analytics and Intelligence,” IEEE Intelligent Systems, vol. 25, no. 6, pp. 13-16, 2010.

8. J. Leskovec, “Social media analytics: tracking, modeling and predicting the flow of information through networks,” in Proceedings of the World Wide Web Conference, pp. 277-278, 2011.

9.Cambria, E., Speer, R., Havasi, C., & Hussain, A. (2010). SenticNet.In Proc. of AAAI SCK-10.Davidov, D., Tsur, O., &Rappoport, A. (2010).

10.Enhanced Sentiment Learning Using Twitter Hashtags and Smileys. ICCL-10. De Choudhury, M., Counts, S., & Horvitz, E. (2013).

11.Predicting Postpartum Changes in Emotion and Behavior via Social Media.In Proc. CDeChoudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting Depression via Social Media. In Proc. ICWSM-13

