

Crowd Management using Sentiment Analysis

Prajaktee S. Rane, Rubeena A. Khan

Department of Computer Engineering, Modern Education Society's College of Engineering, Pune, Maharashtra, India

ABSTRACT

Today, many users are using Social networking sites such as Facebook, Twitter, LinkedIn etc. where user's give their own opinion on particular event or specific situation. This paper focus on Crowd management and control using sentiment analysis. Congestion in crowded area is identified is noted through public opinions made on social networking sites. The public opinions are ambiguous and it is hard to analyse the situation manually or through simple algorithms. Peoples post their feeling through Twitter, LinkedIn etc. Their response for a specific situation is either positive, negative or neutral. The public opinions are then collected, processed and analysed using data mining techniques. In this paper, Sentiment analysis is done by rule-based algorithm. By analysis crowded area, we can move crowd to uncrowded area to avoid undesirable situation.

Keywords: Crowd Management, Sentiment Analysis, Rule-Based Algorithm.

I. INTRODUCTION

Due to increase in population and number of peoples gather for some reason like railway station, religious gathering, music festival etc. Mass amount of peoples in crowd can cause congestion which may cause disaster.

Crowd management and control can be achieved by using Internet of Things like we can use RFID tags for finding whether density of peoples in specific area is more than expected then we can move crowd to uncrowded area. But the system becomes costlier.

Almost everyone is using social networking applications on smartphones, tablets. Peoples put their opinion on social networking sites on given topic for events happening around you. User's opinion having same interest are collected for sentiment analysis.

Collected data is then preprocessed to remove unwanted words like emotional icons, stop words like a, is, the. And finally classification is done by data mining and machine learning algorithm. Classification is done on basis of polarity. It is classified as either positive polarity or negative polarity.

Paper is arranged as follows: In Section 2 (**Preliminary**), some basic terms used in Sentiment Analysis are discussed. In Section 3 (**Architecture and Processes**), it describes architecture and processes used in Crowd management using sentiment analysis. Section 4 (**Related work**), it describes work related to proposed work. In Section 5 (**Proposed System**), it describes about our proposed system. Section 6 (**Conclusion**) is last part of paper which summarizes the study.

II. PRELIMINARY

Sentiment analysis is done on subjective sentences or objectives sentences. Peoples give their opinion through social media applications like Twitter, Facebook etc. This sentiment data can be used to analyse situation.

In this section, we will look at details of Sentiment Analysis like levels of sentiment analysis, polarity classification, algorithms used for Sentiment Analysis.

A. Levels of Sentiment Analysis. [7]

Sentiment analysis is getting real voice of peoples toward specific service, news, movie, event, issue etc. There are following levels of Sentiment Analysis.

1. Document level: In this level, people put sentiment on complete document like whether document is having positive, negative or neutral opinion. In this level, document should talk about single item.
2. Sentence Level: In this level, one-line comment written in the form of sentence is identified as the specified sentence represent its positive, negative or neutral sentiment. In this level of sentiment analysis, we should consider subjectivity of topic, negation and degree of polarity. Semantic orientation is one important part for sentence level sentiment analysis.
3. Entity and Aspect level: In aspect level opinion mining, we can find polarity towards different entities. It is concerned about opinion made by people.

B. Polarity Classification. [8,15]

Sentiment analysis classify document or sentence as positive, negative or neutral polarity. Polarity classification can be done by supervised or unsupervised technique.

C. Semantic Analysis and Semantic Orientation. [4,6]

Semantic analysis is combined effect of subjectivity and polarity of text and semantic orientation is polarity and strength of word. We mainly have to find semantic orientation of word, text or sentence. Before that we have find sentiment of words towards our goal.

There are two approaches for finding sentiment automatically.

1. Lexicon based Approach: In this approach semantic orientation of document or sentence is calculated from semantic orientation of words found in document or sentence. Classification of the words is done by supervised classification where labelled instances are present.
2. Statistical Approach or Semi-supervised: It is semi-supervised classification where dictionaries are manually created. It is lexicon based approach where we use seed words to expand dictionary. Dictionary contains list of lexicons (adjectives or adverbs) and their Semantic Orientation. After dictionary is built, retrieve sentiment from sentence and it is modified by adjective or adverb

which will increase or decrease intensify of sentiment. We find score of sentiment and these scores are aggregated into single score.

D. Techniques used for Sentiment Analysis.

There are number of techniques like machine learning approach, Natural language processing, text mining. Here, we use following machine learning algorithms for sentiment analysis.

Rule Based Approach [2,6,10]: Rule based approach is domain specific approach. It is typically {IF:THEN} form of expression. (Example: IF <condition> THEN Result)

Rule is applicable if a specific or specified number of conditions are satisfied. On left hand, side it specifies number of conditions to satisfy and class label is specified on right hand side.

Rule based approach must have dictionary of subjective words. Rule based approach is used to find polarity of pattern of words found in sentence or document. Following criteria are considered for generating a rule.

Support- Absolute number of inputs used in training data set which satisfies specific rule.

Confidence- It is probability that right hand side of rule is satisfied if left hand side is satisfied.

Unsupervised Algorithm [3,4]: Here, we use lexical type of analysis. In unsupervised machine learning classification is done on unlabeled data.

In this type first extract potential words from sentence and then calculate semantic orientation according to well-known emotional words. For this purpose, different search engines are used.

E. Advantages of Sentiment Analysis.

There are following advantages of sentiment analysis

1. It listens to public voice, what the public feels about current issue.
2. System can measure positive and negative comments made by public.
3. Accordingly, administration or system can take necessary action in case of congestion due to crowd.

4. It is fast process where we can get continuous input from desired location for specific issue.
5. It is cheapest method to implement.

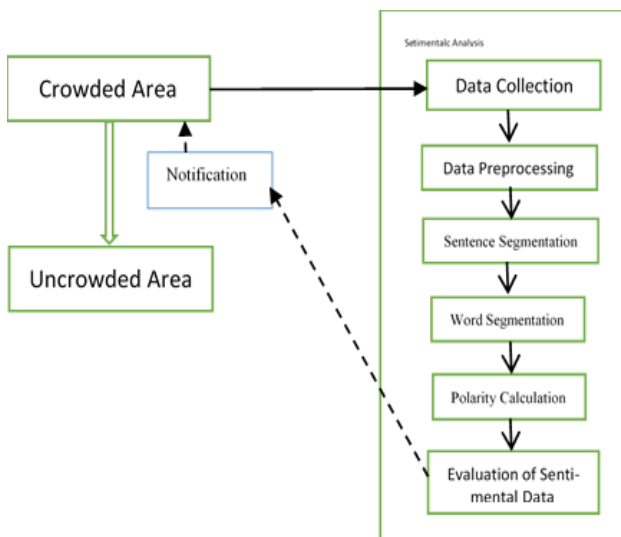
III. ARCHITECTURE AND PROCESSES

A. Data Collection.

If some mischievous behavior occurs at surrounding, peoples puts their opinion on social networking sites. For example, on Twitter, users Tweets for event or some undesirable behavior occurs at the surrounding.

In this paper, if user feels like crowd in the particular area is more, then peoples put their opinion regarding congestion in area. This can be done by identifying location and firing query depending on specified location. They give their opinion as positive, negative or neutral sentiment. User can enter different kind of sentences having similar feeling on social networking sites.

Data collection is continuous process in which we get the data of particular interest from social networking sites.



Data collection should be done which satisfies some criteria like location, area where they are blocked etc.

B. Data Pre-processing. [2,4,6]

Find Similar words: Collect or extract information through social networking sites depending on some search criteria like location, time, event. This fetched information may contain raw content.

Removing unwanted words: When user enter his sentiments through social media, they also include some words which does not specify any sentiments like

- Emotional icons (There are 170 emoticons which does not specify any emotions; So remove them)
- URLs – Replace it with word |URL|
- Stop words – Words like “a”, “is”, “the” does not indicate any sentiment.
- User names and Hash tags - @ symbol before username and # for topic are replaced by AT_USER
- Repeated Letters- croooooowd is replaced by crowd.
- Slang words: are non-English words.

Grouping the words: In this step, group the words having same theme together into tokens.

C. Sentence and Word Segmentation. [4,17]

Sentence Segmentation: In this, message is divided into sentences. Dots are used as delimiter whereas commas and semicolons aren’t used as delimiter because they are used to express sentiment.

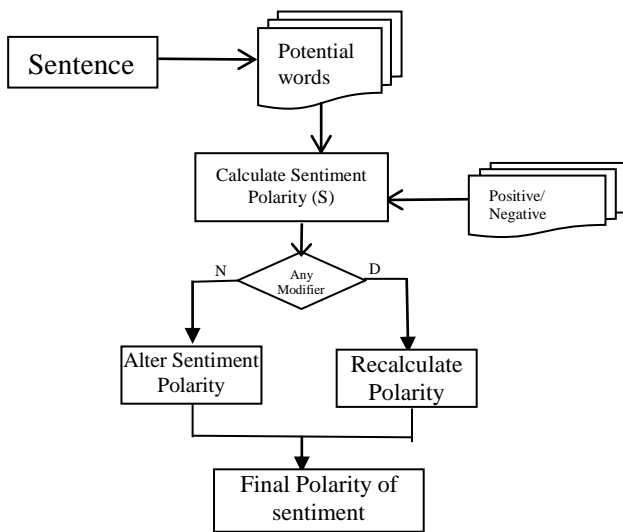
Word Segmentation or Tokenization: Words are extracted from sentences. Word segmentation is done by using whitespaces which are present in between two words. These words are nouns, verbs, adverbs, adjectives etc. We define positive and negative seed sets (PS, NS). These positive and negative sentiment sets are expanded by adding synonyms and antonyms of seed taken from data set Thesaurus. (<http://thesaurus.com>)

D. Polarity Calculation. [2,3]

Sentiment polarity of word is determined for potential word. If it appears more frequently in positive seed sets than negative seed sets, then word represent positive sentiment otherwise it represents negative sentiment.

To find sentiment polarity of potential word, we need to consider sentiment word, modifiers of sentiment like negation and depth of feeling. For example, Potential word ‘crowded’ represent negative sentiment but if we use ‘not crowded’, it will represent positive sentiment. If we use ‘not too much crowded’ will decrease strength of sentiment.

In addition to this, location of potential word in sentence also alter sentiment polarity of word.



E. Semantic Rule Construction. [2]

Formation of sentiment rule is important in rule based approach. Sentiment rule is built by considering feeling expressed by sentiment word let S, Negation word which will alter meaning of sentiment let N and degree of sentiment which represent intensity of sentiment let D. Sentiment rule is formed by combining these three factors S, N and D. S is main part of sentiment which are nouns and verbs defined in sentence, N and D are used as modifiers of sentiment which are adverbs and adjectives which adds special meaning to nouns and verbs.

F. Evaluation of Sentiment Data. [2,4]

Classification technique is used for monitoring correct result. Final output is used to detect the crowd in specified region. The different classification models are used for real-time crowd monitoring.

Final polarity is calculated by comparing polarity labels with user's feeling. This is calculated by following formulae.

$$R = (1/F) \sum (S)(I)$$

R = Resultant polarity

F- Set of Feelings

S- Strength of feeling

I - Polarity Label where $I = \{-1,1\}$

1 - Region is crowded region

1 - Region is not crowded.

The system continuously monitors the system. The system then sends signal to Electronic Bulletin Board. Arrow direction on bulletin board will move the traffic from crowded area to uncrowded area.

IV. RELATED WORK

- Yuichi Kawamoto, Naoto Yamada, Hiroki Nishiyama, Yao Zheng, Yoshitak Shimizu and Nei Kato, [1] published “A Feedback Control Based Crowd Dynamics Management in IoT System” In the given system information regarding crowd is collected through Internet of Things. Here they use feedback control system to control and verify the effectiveness of crowd dynamic management. In this paper, result of system depends on sensitivity of users, lag time of changes and the power of instruction.
- Jianping Cao, Ke Zeng, Hui Wang, Member, Jiajun Cheng, Fengcai Qiao, Ding Wen, and Yanqing Gao presented [2] “**Web-Based Traffic Sentiment Analysis: Methods and Applications**”. In this paper, author focuses on intelligent transportation systems (ITSs). Author propose the traffic sentiment analysis (TSA) to tackle the problem of modern ITSs. Author applies rule-based approach for TSA. It demonstrates the cases like the “yellow light rule” and “fuel price” in China. These are used to find efficiency of system. Functions used in TSA are investigation, evaluation and prediction.
- “**A Survey of Opinion Mining and Sentiment Analysis**” [3] Bing Liu, Lei Zhang, University of Illinois at Chicago, Chicago. In this chapter, author discuss about identifying relevant sites and summarizing information and opinion contained in them which will be helpful for subjective biases and mental limitation.
- A. Mantejo-Raez, M.C.Diaz-Galiano, L. A. Urena-Lopez and F. Martinz-Santiago presented [4] “**Crowd Explicit Sentiment Analysis**”. This paper introduces CESA as sentiment analysis approach. From collection of document microblogs are identified. Microblogs shows clear polarity value in textual representation. This paper uses polarity classification for different languages. It introduces SVM machine learning algorithm for classification. It uses steam based approach. This method is only used for simple structure. We need to build solution for sophisticated sentences.

- Sasan Amini, Ilias Gerostathopoulos and Christian Prehofer [5] presented “**Big Data Analytics Architecture for Real-Time Traffic Control**”. This paper presents a comprehensive and flexible architecture for real-time traffic control based on Big Data analytics. This paper used Kafka platform for building data pipelines and stream processing. This is having limitation for large and heterogeneous data sets from multiple sources.
- Eleonora D’Andrea, Pietro Ducange, Beatrice Lazzerini, and Francesco Marcelloni [6] presented “**Real-Time Detection of Traffic From Twitter Stream Analysis**”. In this paper, author presented real-time monitoring system for traffic event detection from Twitter stream analysis. The system fetches opinions made by people through Twitter, and with some search criteria it processes tweets by applying text mining techniques; and finally performs the classification of tweets. Its implementation shows capability of the system of detecting traffic events almost in real time. Under worst case, system should react to traffic related events.
- Mr. Penubaka Balaji, Dr. O. Nagaraju, Prof.D. Haritha presented [7] “**Levels of Sentiment Analysis and Its challenges: A Literature Review**”. In this paper author discuss about sentiment analysis as machine learning approach and this approach is used to categorize and mine human opinions, sentiments. Author also represents different phases for sentiment analysis.
- Daniel Ansari [8] presented “**Sentiment Polarity Classification using Structural Features**”. In this paper, author finds position of phrase for prediction of phrase. Hybrid lexicon and simple machine learning based algorithm is used for classification. This paper uses application for nutrition supplement and hotel.

V. PROPOSED SYSTEM

Crowd Management is done by using Rule based algorithm in Sentiment Analysis. This eliminate disadvantages of existing system where there is lack of hardware resources to capture the current status of event, situation etc. In this system, public opinions are collected from social networking sites for specific event or situation. Data is then pre-processed by eliminating unwanted characters, symbols and finding

similar words. Then, we break down the document in sentences and words. Words are extracted from sentence and build positive and negative seed domain by observing antonyms and synonyms from Thesaurus and adding those words to positive and negative seed sets. Sentiment polarity of potential words are analysed as positive or negative sentiment. The polarity of sentiment is altered by negation. Strength of polarity is increased decreased by depth of sentiment represented by adverbs used in sentence. We use Rule based approach, sentiment of potential word is found by sentiment represented by potential word and their modifiers. Negative words and degree of potential word are modifiers of sentiment. Finally, resultant polarity of document is calculated by set of sentiments and strength of sentiment. Resultant polarity will be helpful to detect situation of specified area.

VI. CONCLUSION

Today, crowd management is an important issue to avoid congestions caused due to crowd and it will also cause accidents. This can be done by collecting and analysing public opinion. Due to increased usage of social networking sites and smartphones, the peoples immediately express their feeling for any event or situation. Peoples express their feeling which differs from each other. It is difficult to analyse actual situation by manually tracking their feelings. We use Rule based approach which detect polarity of sentiment as positive or negative sentiment. Also, it also considers adverbs which add special meaning to sentiment Negative words revert the polarity of sentiment. And degree adverbs increase or decrease polarity of sentiment. We suggest this system can be used for decision making for management authority.

VII. REFERENCES

- [1] Yuichi Kawamoto, Naoto Yamada, Hiroki Nishiyama, Nei Kato, Yoshitaka Shimizu, Yao Zheng [1] presented “A Feedback Control Based Crowd Dynamics Management in IoT System”. DOI 10.1109/JIOT, IEEE, Internet of Things Journal, 2017
- [2] Jianping Cao, Ke Zeng, Hui Wang, Member, Jiajun Cheng, Fengcai Qiao, Ding Wen, and Yanqing Gao “Web-Based Traffic Sentiment Analysis: Methods and Applications” IEEE Transaction on Intelligent Transportation Systems, VOL. 15, NO. 2, APRIL 2014 pp-844-853

- [3] A Survey of Opinion Mining and Sentiment Analysis- Bing Liu, Lei Zhang, University of Illinois at Chicago, Chicago- C. C. Aggarwal and C. X. Zhai (eds.), Mining Text Data, DOI 10.1007/978-1-4614-3223-4_13, Springer Science+Business Media, LLC 2012 415-462
- [4] A. Mantejo-Raez, M.C.Diaz-Galiano, L. A. Urena-Lopez and F. Martinz-Santiago “Crowd Explicit Sentiment Analysis”, Elsevier, Knowledge-Based Systems 69 (2014) 134–139
- [5] Sasan Amini, Ilias Gerostathopoulos and Christian Prehofer “Big Data Analytics Architecture for Real-Time Traffic Control”, Models and Technologies for Intelligent Transportation Systems (MT-ITS) 5th IEEE International Conference, August 2017
- [6] Eleonora D’Andrea, Pietro Ducange, Beatrice Lazzarini, and Francesco Marcelloni “Real-Time Detection of Traffic From Twitter Stream Analysis”, IEEE Transactions on Intelligent Transportation Systems, 2015
- [7] Mr. Penubaka Balaji, Dr. O. Nagaraju, Prof.D. Haritha. “Levels of Sentiment Analysis and Its challenges: A Literature Review”. International Conference On Big Data Analytics and Computational Intelligence, IEEE 2017 436-439
- [8] Daniel Ansari “Sentiment Polarity Classification using Structural Features”, IEEE 15th International Conference on Data Mining Workshops, 2015 pp-1270-1273
- [9] Long Chenga, Jianwei Niua, Linghe Kongb, Chengwen Luoc, Yu Gu d, Wenbo Hee, Sajal K. Dasf “Compressive sensing based data quality improvement for crowd-sensing applications”, Elsevier, Journal of Network and Computer Applications, 77 (2017), 123-134
- [10] Tanzim Mahmud, K. M. Azharul Hasan, Mahtab Ahmed, Thwoi Hla Ching Chak, “A rule based Approach for NLP based query Processing”, Electrical Information and Communication Technology (EICT), IEEE, Jan 2016, (78-82)
- [11] Michel Wiegand, Manfred Klenner, Dietrich Klakow, “Bootstrapping Polarity Classifiers with Rule-based Classification”, Springer, December 2013, Volume 47, Issue 4, pp 1049-1088.
- [12] David Milne, Ian H. Witten, “An Opensource Toolkit for mining `Wikipedia”, Elsevier, Artificial Intelligence 294 (2013), 222-239
- [13] Yisheng Lv, Yuanyuan Chen, Xiqiao Zhang, Yanjie Duan, and Naiqiang Li, “Social Media Based Transportation Research: the State of the Work and the Networking,” IEEE/CAA Journal of Automatica Sinica, Vol. 4, No. 1, January 2017
- [14] Xinhua Zheng, Wei Chen, Pu Wang, Dayong Shen, Songhang Chen, Xiao Wang, Qingpeng Zhang, and Liuqing Yang, “Big Data for Social Transportation”, IEEE Transactions on Intelligent Transportation Systems, Vol 17, No. 3., March 2016.
- [15] https://en.wikipedia.org/wiki/Sentiment_analysis
- [16] Geetika Gautam, Divakar yadav “Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis”, IEEE, 2014
- [17] Alvaro Ortigosa, Jose Martin, Rosa M. Carro “Sentiment Analysis in Facebook and its application in e-learning”, Ortigosa et al./Computers in Human Behavior, Elsevier, 2013