



An Approach for Sentiment Analysis using Neural Network

Varsha Hole, Madhuri Chavan, Tanuja Gavhane, Prof. Sunil Yadav

Department of Computer Engineering, Siddhant College of Engineering, Sudumbare,
Pune, Maharashtra, India

ABSTRACT

Sentiment analysis has a wide range of applications specially to know product reviews, movie reviews, and political sentiment. With the growth of social media as important medium for people to express their opinion, Sentiment analysis along with usage of cell phones has also become important to find out popularity of an issue or product, specifically, to know future about events/issues of society. In current study, we present a way to predict future events of society were by performing sentiment analysis on unstructured knowledge available to us. For this research we take assistance from the Machine Learning algorithmic rules and neural networks. The paper explains how results of sentiment analysis can be derived for every unknown issue (which may or about to occur in near future) of society using existing unstructured knowledge that already exists on various social media platforms for the user. The part of unstructured knowledge is text of people from online blogs, tweets, reviews, Facebook posts and comments.

Keywords: *Sentiment Analysis, Text Mining, Machine learning, Neural Network*

I. INTRODUCTION

Social Media platforms have led to abundant availability of data about people's opinion. Using various sentiment analysis techniques, the polarity of the opinions can be examined and can be classified as positive, negative and neutral. Through secure use of the data about opinions on various subjects provided by these online media platforms, many industries and job creation can flourish better if data is used efficiently to perceive the opinion of people about many unknown issues of society. With the help of such data some sudden unwanted issues can predicted beforehand can prevented from happening.

This information of such kind i.e. opinion prediction is very beneficial for some digital marketing companies. These corporations track sentiments of people to predict mood of larger public about a product or towards elected authorities in a given country, or to predict allegiances of people towards various sports teams.

A common approach to sentiment analysis consists of systematically reviewing content from websites, especially social networks like Facebook, Twitter, and Google+, and using an algorithm to determine the opinions of the masses. There is term coined for sentiment analysis at such scale performed on data extracted from text mining i.e. Opinion Mining [12]. First step for opinion prediction on unknown matter is to determine polarity of opinion on specified existing text.

II. LITERATURE SURVEY

Towards Using Visual Attributes to Infer Image Sentiment of Social Events

Author: Unaiza Ahsan, Munmun De Choudhury, Irfan Essa

Description: Huge spread and pervasive adoption of smart phones has caused instantaneous sharing of images that capture events starting from mundane to life- changing happenings. Author tried to propose the framework that captures sentiment information of such pictures event. This approach extracts associate intermediate visual illustration of event pictures supported the visual attributes that occur within the pictures going on the far side sentiment-specific attributes. Author tried to map the highest foretold attributes to sentiments and extract the dominant feeling related to an image of an event.

Understanding Pending Issue of Society and Sentiment Analysis Using Social Media

Author: Jong-Seon Jang, Byoung-In Lee, Chi-Hwan Choi, Jin-Hyuk Kim Description: The recognition of smart telephones and social media is growing every day with the development technology. In this study, the unfinished problem of society is centered through reading the non-established facts of the 'Sejong metropolis dot-com' in the social media of Sejong metropolis targeted by analyzing the non-structured information of the 'Sejong town dot-com' within the social media of Sejong town. By using Naive based Machine Learning formula, author derived the results of the Sentiment Analysis of every unfinished issue of society. Through secure use of the govt. provided information, following new industries and job creation could make the most of this info to rise, perceive the unfinished issue of society, as an example, the prediction and bar of sudden irregular incident will increase.

Sentiment Analysis of Short Informal Texts

Author: Svetlana Kiritchenko, Xiaodan Zhu, Saif M. Mohammad.

Description: Author describe a progressive sentiment analysis approach that detects the sentiment of short informal matter messages like tweets and SMS (message-level task) and also the sentiment of a word or a phrase at intervals a message (term-level task). The method applied text classification approach using a range of surface type, semantic, and sentiment options. A novel high-coverage tweet-specific sentiment lexicons is used to derive sentiment options square measure primarily derived from novel high-coverage tweet-specific sentiment lexicons. These lexicons square measure mechanically generated from tweets with sentiment-word hash tags and from tweets with emoticons. To adequately capture the sentiment of words in negated contexts, a separate sentiment lexicon is generated for negated words.

Image Sentiment Analysis from a Mid-level Perspective

Author: Jianbo Yuan, Quanzeng You, Sean Mcdonough, Jiebo Luo Description: Visual content analysis is very popular but difficult to implement. Due to this cause, the popularity of social networks, pictures grow to be a convenient carrier for information diffusion among on-line users. To understand the diffusion styles and absolutely unique factors of the social snap shots,

author desires to interpret the pictures initial. Almost like content material, snap shots conjointly carry completely different levels of sentiment. But, absolutely distinctive from text, sentiment analysis uses handy accessible linguistics and context records; to extract and interpret the sentiment of a photo remains very tough. In this paper, yuan proposed a picture sentiment prediction framework that takes advantage of the mid-level attributes of a picture to predict its sentiment. This makes the sentiment classification outcomes a variety of explicable than directly victimization the low-level options of a snap.

Modified Naïve Bayes Classifier for Categorizing Questions in Question-Answering Community

Author: Yeon, Jongheum, Sim, Junho, Lee Snggu

Description: Extended (weight is given) Naive Bayes Classification is used to obtain good result values for classification of nonstructured data in social media. Assigning weights in accordance with the frequency of each attribute allowed author to derive a value for better results.

III. PROPOSED SYSTEM

In proposed work, A non-structured data of social media is classified using Extended (considered that weight value is given) Naive Bayes algorithm. A good result can be derived by assigning weights according to the frequency of each [3, 4]. The makes use of over Social media as Twitter yet fb are entirely famous between latest generation then increases the lookup possibilities to decide as what people sense and then emote regarding entities and events. Twitter is used widely by people to share opinions on daily basis. To develop the sentiment analyser based on presidential debates, a new framework is proposed by kim [8].

Proposed work is very much similar to the SentiBank approach [13] which extracts data based on representation of images and then predicts their sentiment using features. Proposed methodology is different and includes the new terminology image type. Previously, prediction is based on visibility (objects and faces are very clear), so object detectors can be used for detection. The focus of proposed work is on event sentiment detection and condition where faces and object picture may not clearly visible. The details of proposed work is depicted in following figure1.

A. NaiveBayes Class:

This is the main piece of the Text Classifier. It equipment strategies such as train() and predict() which are responsible for training a classifier and then the usage of that because predictions. It should be noticed that this category is additionally accountable for calling the appropriate external strategies to pre-process and tokenize the record earlier than training/prediction.

B. NaiveBayesKnowledgeBase Object:

The outturn of training is a NaiveBayesKnowledgeBase Object which stores whole the integral statistics yet probabilities to that amount are back with the aid of the Naive Bayes Classifier.

C. Document Object:

The training and the prediction texts within the implementation phase are stored as Document Objects (DO). The DO stores all the tokens of the document, their information and the goal type of the document.

D. FeatureStats Object:

The FeatureStats Object stores numerous records which are generated during Feature Extraction process. This type of information is joint counts of Features and Class (from which the joint probabilities and likelihoods are estimated), the individual Class counts and the number of observations that are used for training.

E. FeatureExtraction Class:

In this stage, the information required by classification process are stored in caches and returned in a FeatureStats Object to skip the recalculations and to make process fast.

F. TextTokenizer Class:

The responsibilities of this class are pre-processing, clearing and tokenizing the texts and convert it into Document objects.

Main components of the proposed system are:

1. Pre-Processing: The process includes➤ **(SWR) Stop word removal**➤ **Extra symbol removal**➤ **Part Of Speech tagging**

It utilizes English parser mode and find out the actual parts of speech.

In this process, standard Penn Treebank POS tag sets are used.

2. Feature Extraction: In this phase, features are extracted from original dataset. The result then used to distinguish the positive and negative polarity of a sentence to determine the actual sentiment of the individuals.

3. Training and Classification: ANN algorithm is considered as a best approach for classification in proposed system. It follows two steps i.e. Training and Testing. Training phase includes positive and negative comments of IMDB review dataset. After that weights are assigned to each comment and also fuzzy logic is applied to remove the negative results (like not, never) that improves the overall efficiency. It increases the accuracy in terms of correlations and dependencies. At final we create the dictionary with positive comments and in subsequent section, reviews are tested.

IV. Conclusion

Sentimental Analysis and data provided by social media platform can equip us with the capability to analyze possibility of occurrence of some unknown events in the future. The proposed framework relies on social media platforms, to gather data regarding opinions of people for a product or company or political party. One very crucial limitation of this study was that data used for Sentiment Analysis was classified only in the positive and negative and neutral categories. Sentiment analysis should provide us with a very fine-grained classification of sentiments. In future studies, we need to work on minimizing error rate. One of the way it can be by using a more robust and very fine sentiment analyzer. Besides, we can also take samples of existing knowledge from social media for much longer period.

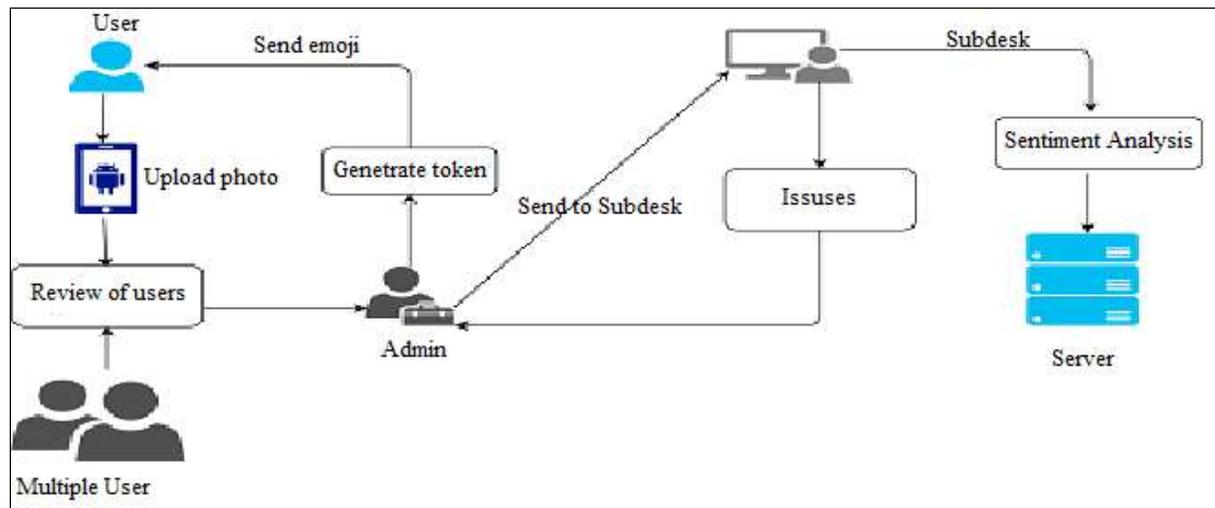


Figure 1. Architecture Diagram

References

1. IDG Korea, "Read Emotions in Article", Understand of Sentiment Analysis, IDG Tech Report, 2014.
2. Dave, Kushal, Steve Lawrence, and David M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", Proceedings of the 12th international conference on World Wide Web, ACM, 2003.
3. Yeon, Jongheum, Sim, Junho, Lee Snggu, "Modified Naïve Bayes Classifier for Categorizing Questions in Question-Answering Community", Information science society journal, Real and Letter of Computing, pp 95-99, 2010.
4. Kim, Hyeonjun, Jung, Jaeun, Cho Geunsik, "Spam - Mail Filtering System Using Weighted Bayesian Classifier", Information science society journal, Software and Application, pp 1092-1100, 2004.
5. Korean Society For Big Data Service: "Derive the Issue of Sejong City through Sejong City dot-com", Research Report of Sejong Metropolitan Autonomous City (Sejong City), 2015.
6. Kim Sangho, "Transportation Bigdata Analysis and Evaluation of Performance in Apache Spark", Chungbuk National University, 2015.
7. Kang Hanhoon, "An Improved Naïve Bayes Method and Senti Lexicon for Ranking and Sentiment Analysis of Places", Sejong University, 2012.
8. Lee, Eungyong, "New Possibilities and solution task of Big Data Era", Internet Security Issue, Korea Internet Security Agency(KISA), 2012.
9. Analysis of complaints after the e-People was established that complaint keyword, Research Report of Anti-Corruption and Civil Rights Commission, Complaints Information Analysis Center, 2013.
10. Seo, Jinwan Nam, Gibeom Kim, Gyewon, "Analysis and meaning of the situation that utilize social media of local autonomous entity", The Korean review of public administration, 2012.
11. Kim, Singon Cho, Jaehui, "Proposals for the introduction of Big Data of local autonomous entity", Journal of Korean Association for Regional Information Society, 2013.
12. S. K. Khatri and A. Srivastava, "Using sentimental analysis in prediction of stock market investment," 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, 2016, pp. 566-569.
13. Complaints Information Analysis Center: Analysis of complaints after the e-People was established that complaint keyword, Research Report of Anti-Corruption and Civil Rights Commission, 2013.