

Sentiment Analysis

Er. Divya Kadyan¹ and Dr. Sanjay Singla²

Abstract- It is very important to understand how people will communicate during disaster time with the disaster management. Social media plays an important role in broadcasting the information, and twitter is the most commonly used social media. During the 2015 Chennai Floods, Twitter was utilized for spreading information, sharing firsthand observations. In this paper, we will analyze various classifiers that can be used to detection text based emotions during the Chennai floods. In addition to this a prototype to analyses this will also be studied in this paper.

Keywords- Sentiment Analysis, Social Media, Machine Learning Algorithms.

1. INTRODUCTION

Most of the users use twitter information during the crisis. Tweets not only contains words but some of the tweets contains the emoticons too. However emotion based tweets has become the unexplored topic in present time [1, 2]. We seek the better communication of users with management in crisis using twitter data collected during the 2015 Chennai Floods. This recent Chennai huge amounts are the eighth-most high-priced healthy disaster to possess attack the world with 2015, claims UNITED KINGDOM reinsurance agent Aon Benfield. Asia endured the \$3 million loss in order to the financial system from serious rain fall and also surging with November and also beginning December, the corporation stated with the monthly statement in global catastrophes. The Wall membrane Block Log statement mentioned of which wildfires with Philippines with Jan price this Southeast Oriental financial system \$14 million, thus turning it into probably the most expensive healthy disaster involving 2015 [12]. Aon's recent statement mentioned global devastation losses with April are anticipated in order to top 10 million. As much as this recent water and also avalanche with Tamil Nadu, particularly with Chennai, Aon Benfield's estimate will be of which 600 men and women could have past away soon after rain fall lashed India's the southern area of shoreline and also areas of Sri Lanka. More than 100, 000 buildings had been harmed caused by inundation over the a couple countries.

Analyzing emotions based on text helps us to know how people react during crisis using social media. Emotions helped a lot in recognizing the text e.g. "fear" describes the mental illness [3, 4]. Due to large dataset of people tweets, it become necessary to get appropriate information, so machine learning algorithms provides best information.

This paper will discuss the need of the classifiers for the training of the disaster dataset. ML provides better learning rate so provided here. [5].

¹ *Computer Science Engineering IET Bhaddal Technical Campus Rupnagar, India*

² *Computer Science Engineering IET Bhaddal Technical Campus Rupnagar, India*

1.1. Twitter

Twitter is a well-known microblogging service exactly where people create reputation communications (called "tweets"). Most of these twitter updates some- periods show opinions concerning unique subjects. We propose an approach to instantly extract notion (positive as well as negative) from twitter [7]. This is very valuable since it allows opinions to get aggregated without having guide book treatment. Customers can use notion research to research offerings just before making a decision. Online marketers can use this particular to research public opinion in their business as well as solutions, or to assess customer satisfaction. Businesses also can make use of this to gather critical opinions concerning troubles inside recently produced solutions.

However in suggested examine twitter updates is going to be utilized to realize emotions.

1.2 Defining Sentiment

Sentiment analysis is a text classification problem which deals with extracting information present within the text [6, 8]. This extracted information can be then further classified according to its polarity as positive, negative or neutral. It can be defined as a computational task of extracting sentiments from the opinion. Some opinions represent sentiments and some opinions do not represent any sentiment.

- Sentiments: Opinions or in other sense can be recognized as someone's linguistic expressions of emotions, beliefs, evaluations etc.
- Analysis: To capture the opinions from a pool of users whether the opinion is positive, negative or neutral [9].
- Benefit: Provide efficient information in decision making

1.3 Characteristics of Tweets

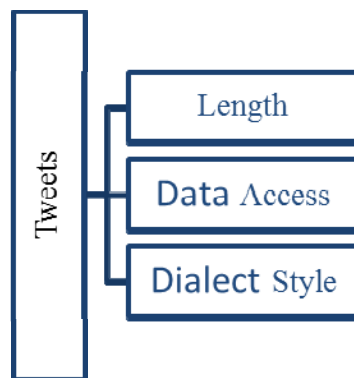


Figure 1. Features of Tweets

Twitter messages have many unique attributes as shown below [10, 11]:

Length: The absolute maximum length of a Twitter communication can be a hundred and forty personas.

Data access: Yet another difference could be the specifications of information readily available. While using Twitter API, it is quite all to easy to acquire countless twitting pertaining to teaching.

Dialect style: Twitter customers submit mail messages from many different media, as well as their particular cellular phones. Your frequency of misspellings along with slang throughout twitting is much above throughout other names.

2. MACHINE LEARNING METHODS

Machine learning algorithms are built to measure the probability of uncertainty based on past experience. The ML algorithms learn the patterns in data and then gives the output. High the learning observations good is the ML algorithm. Basic features of the ML algorithms are as following:

- Learning speed
- Memory utilization
- Accuracy
- Understandability

2.1 KNN Method

K nearest neighbours is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique [13].

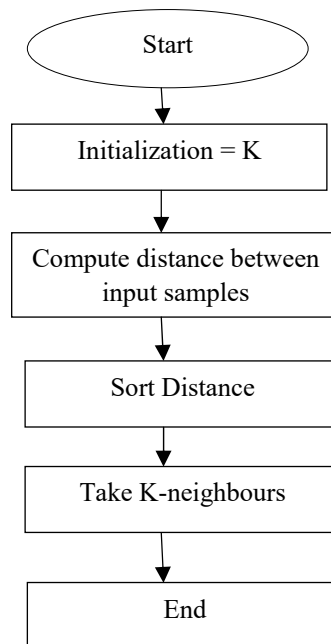


Figure 2. KNN Classifier

Classification KNN predicts the classification of a point X_{new} using a procedure equivalent to this:

1. Find the NumNeighbors points in the training set X that are nearest to X_{new} .
2. Find the NumNeighbors response values Y to those nearest points.

Assign the classification label Y_{new} that has smallest expected misclassification cost among the values in Y

2.2. Naïve Bayesian Method

Naïve Baye's methods are supervised learning methods based on baye's theorem.

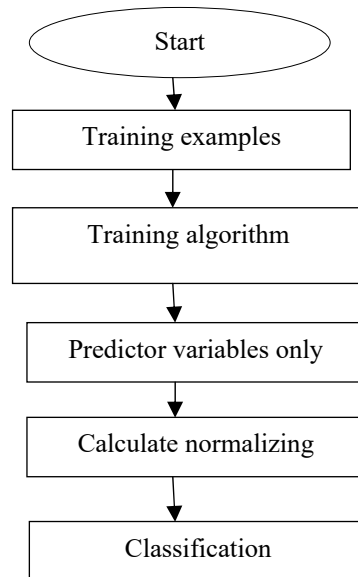


Figure 3. Naïve Bayesian Classifier

Bayes theorem can be applied as in the following manner [14]:

$$A(u | c_1, \dots, c_m) = \frac{A(u) \prod_{i=1}^m A(c_i | u)}{A(c_1, \dots, c_m)}$$

Using the naive independence assumption that

$$A(c_o | u, c_1, \dots, c_{o+1}, \dots, c_m) = A(c_o | u),$$

for all o , this relationship is simplified to

$$A(u | c_1, \dots, c_m) = \frac{A(u) \prod_{i=1}^m A(c_i | u)}{A(c_1, \dots, c_m)}$$

Since $A(c_1, \dots, c_m)$ is constant given the input, we can use the following classification rule:

$$A(u | c_1, \dots, c_m) \propto A(u) \prod_{i=1}^m A(c_i | u)$$

2.3 HMM Method

A Markov model is a probabilistic process over a finite set, $\{S_1 \dots S_k\}$, usually called its states. Each state is called a process [14].

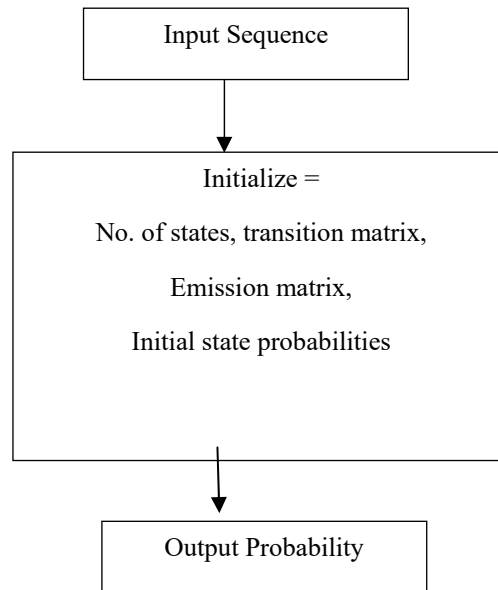


Figure 4. HMM Model

Order 0 Markov Models

An order 0 Markov model has no "memory":

$at(cy = Do) = at(cy' = Do)$, for all points y and y' in a sequence.

Order 1 Markov Models

An order 1 (first-order) Markov model has a memory of size 1. It is defined by a table of probabilities $at(cy=Do | cy-1=Dk)$, for $o = 1..1$ & $k = 1..1$. You can think of this as 1 order 0 Markov models, one for each Dk . Examples of Markov Model

Let $X(n)$ be the weather of the n th day which can be $M = \{\text{sunny, windy, rainy, cloudy}\}$. One may have the following realization: $X(0) = \text{sunny}$, $X(1) = \text{windy}$, $X(2) = \text{rainy}$, $X(3) = \text{sunny}$, $X(4) = \text{cloudy}$,

2.5 Weighted K-Mean

K means is a MATLAB library which handles the K – means problem, which organizes a set of N points in M dimensions into K clusters.

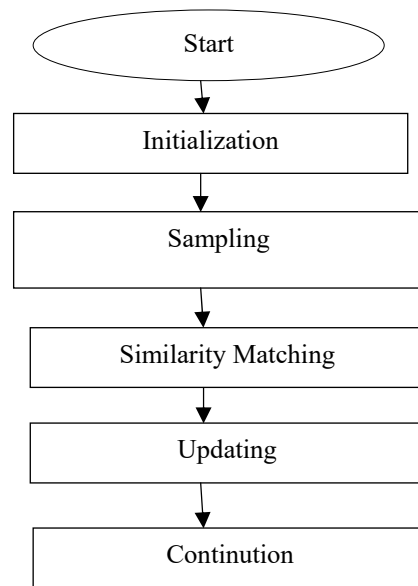


Figure 5. Weighted K-Mean method

The K means problem allows us to include some information, namely, a set of weights associated with the data points. The intent is that a point with a weight of 5.0 is twice as “important” as a point with a weight of 2.5, for instance. This gives rise to the “weighted K-Means problem [16].

In K-means problem, we are given a set of N problems in M dimensions, and a corresponding set of nonnegative weights $W(I)$. The goal is to arrange the points into k clusters, with each cluster having a representative point $Z(j)$, usually chosen as the weighted centroid of points in the cluster. In the proposed work also, k mean classifier is used to demonstrate a comparative study.

3. CONCLUSION

Sentiment analysis is an ongoing field of research in text mining field. SA is the computational study of Opinions, sentiments, subjectivity toward an entity. The entity can represent individuals, events or topics. The two expressions sentiment analysis and opinion mining of big data are interchangeable

The main goal of this approach is to present the machine learning algorithms that will be helpful in analyzing the sentiments based on text words and to realize that it is not only important to extract the sentiment of user at a certain time but also to capture the habits of user’s activities which will help in detecting the change of sentiments or the user sentiment state. It can also be explained as capturing user’s regular pattern.

REFERENCES

- [1] G. Mishne, “Experiments with Mood Classification in Blog Posts”, Proceedings of the Style2005”, The 1st Workshop on Stylistic Analysis of Text for Information Access, SIGIR 2005, Salvador, Brazil, 2005.
- [2] Y. Jung, H. Park and S. H. Myaeng, “A Hybrid Mood Classification Approach for Blog Text”, Lecture Notes in Computer Science, vol. 4099, pp.1099-1103, 2006.
- [3] H. Liu, H. Lieberman and T. Selker, “A model of textual affect sensing using real-world knowledge” Proceedings of the 2003 international conference on intelligent user interfaces, pp.125- 132, 2003.

- [4] C. O. Alm, D. Roth and R. Sproat, “Emotions from Text: Machine Learning for Text-based Emotion Prediction”, Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada, pp.579-586, 2005.
- [5] A. Neviarouskaya, H. Prendinger and M. Ishizuka, “Textual Affect Sensing for Social and Expressive Online Communication”, Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction, pp.218-229, 2007.
- [6] C. M. Lee and S. S. Narayanan, “Toward Detecting Emotions in Spoken Dialogs”, Journal of the American Society for Information Science. IEEE Trans. on Speech and Audio Processing, vol. 13, pp.293-303, 2005.
- [7] Information from Social networking site: www.twitter.com
- [8] Wiebe, J., Wilson, T., Cardie, C, “Annotating expressions of opinions and emotions in language”, Language Resources and Evaluation, vol. 39, pp. 165–210, 2005.
- [9] Martin, J.R., White, P.R.R, “The Language of Evaluation”, Appraisal in English, Palgrave, 2005.
- [10] Read, J., Hope, D., Carroll, J, “Annotating expressions of Appraisal in English”, The Proc. of the ACL Linguistic Annotation, Workshop, Prague , 2007.
- [11] Whitelaw, C., Garg, N., Argamon, S, “Using Appraisal Taxonomies for Sentiment Analysis”, Proc. of the 2nd Midwest Comp., Linguistic Colloquium, Columbus, 2005.
- [12] http://www.business-standard.com/article/current-affairs/chennai-floods-are-world-s-8th-most-expensive-natural-disaster-in-2015-115121100487_1.html
- [13] Tang B, He H, “ENN: Extended Nearest Neighbor method for pattern recognition”, IEEE Computational Intelligence Magazine, vol.10, pp. 52–60, 2015.
- [14] Metsis, Vangelis; Androutsopoulos, Ion; Paliouras, Georgios, “Spam filtering with Naive Bayes— which Naive Bayes”, Third conference on email and anti-spam (CEAS), 2006.
- [15] Boudaren et al., M. Y. Boudaren, E. Monfrini, and W. Pieczynski, “Unsupervised segmentation of random discrete data hidden with switching noise distributions”, IEEE Signal Processing Letters, vol. 19, No. 10, pp. 619-622, October 2012.
- [16] Drineas, P.; A. Frieze; R. Kannan; S. Vempala; V. Vinay, “Clustering large graphs via the singular value decomposition”, Machine learning, vol. 56, pp. 9–33, 2004.