3rd International Conference on New Horizons in Science, Engineering and Management and Humanities
IIMT College of Engineering, Greater Noida (India)
19th April 2019 www.conferenceworld.in
ISBN : 978-93-87793-82-8

EMOJI BASED SENTIMENT ANALYSIS USING KNN

B.GNANA PRIYA
Assistant Professor
Department of Computer Science and Engineering
Faculty of Engineering and Technology
Annamalai University

ABSTRACT

In today's world every minute we receive lots of comments in social media for any given topic. Many of us tend to express our feelings using emojis apart from the text. In this work we take into account only the emojis used in the comments and analyze the sentiment. We develop an emoji expression detection system which finds emotions such as smile, anger, crazy, funny, ,sick, sleepy, cool and sadness. We extract feature from the emoji image and use KNN algorithm for detection. Emojis are detected and feature point extraction method extracts the selected feature point that are used by the algorithm. we use dataset from Twitter . This method achieved a result of 80% for smile, 72% for anger and 65% for sadness.

Keywords*: Emoji, Emoji expression detection, Emotion prediction, Sentiment analysis, KNN*

1.INTRODUCTION

Sentiment analysis is a popular area of research and is an intelligent method of extracting various emotions and feeling of users. An emotion is a mental and physiological state which is subjective. It involves a lot of behaviors, actions, thoughts and feelings. Sentiment analysis is the automated process of understanding an opinion about a given subject from written or spoken language. Six emotional expressions are considered as most common and universal. The expressions are happiness, sadness, disgust, surprise and fear. We have got several emojis that can be used to express our emotions pictorially. Displaying emotions using emoji is a non-verbal communication technique If efficient methods can be brought about to automatically recognize these emojis expressions, striking improvements can be achieved in the area of computer interaction. Research in emotion recognition has being carried out in hope of attaining these enhancements.

Emoji were invented in 1999 by Shigetaka Kurita and were intended for a Japanese user base. Emoji are available in almost all messaging apps, and while different apps have distinct emoji styles, emoji can translate across platforms. Recent improvements in this area have encouraged the researchers to extend the applicability of emojis emotion recognition to areas like chat room . The ability to recognize emotions can be valuable in emoji recognition applications as well. Suspect detection systems and intelligence improvement systems meant for children with brain

ISBN : 978-93-87793-82-8

development disorders are some other beneficiaries . With the help of sentiment analysis systems the unstructured information available could be automatically transformed into structured data of public opinions

Emojis make the gathering of emotional responses an easy and enjoyable process. It also makes it entertaining to the user to present feedback. The technology referred to as emotional analysis, provides insights into how a customer perceives a product, the presentation of a product or their interactions with customer service representative. Emoji are in this sense an extension to coding standard Unicode, which defines every character in most languages in the world. The closest resemblance to emoji might be developed by Microsoft in 1990. The font includes various symbols, for example objects, shapes, arrows, smiley faces and etc. The repertory of the symbols can be found on current emoji keyboard. It is important to note that are emotions based on colored icon. Sentiment analysis on large-scale social media data is important to bridge the gaps between social media contents and real world activities including political election prediction, individual and public emotional status monitoring and analysis, and so on. Although textual sentiment analysis has been well studied based on platforms such as Twitter and Instagram, analysis of the role of extensive emoji uses in sentiment analysis remains light.

2.RELATED WORK

Anchal [2] analyze various techniques of sentiment analysis for opinion mining like machine learning and lexicon-based approaches. The various techniques used for Sentiment Analysis are analyzed to perform an evaluation study. Mohammed [3] discusses the characteristic of social networks and the effects of Emoji in text mining and sentiment analysis. Twitter is taken as information source for analysis. This analysis proves that the utilization of Emoji characters in sentiment analysis results in higher sentiment scores. Yuxiao[4] proposed a novel scheme for Twitter sentiment analysis with additional attention on emojis. They built a bi-sense emoji embeddings under positive and negative sentimental tweets individually and train a sentiment classifier with LSTM.

Schukla [5] presented a tool which judges the quality of text based on annotations on scientific papers. Its methodology collects sentiments of annotations in two approaches. It counts all the annotation produces the documents and calculates total sentiment scores. Kasper [6] proposed a Web Based Opinion Mining system for hotel reviews. It is capable of detecting and retrieving reviews on the web and deals with reviews. It has multi-topic domain and is based on multi-polarity classification. Mobile devices products reviews were analyzed by Zhang [7]. Nisha [8] uses Machine learning to predict the classification accuracy using Naïve Bayes algorithm. In addition, the research made a judgment of the product quality and status in the market is advantageous. This research used three machine learning algorithms :Naïve Base Classifier, K-nearest neighbour, and random forest to calculate the sentiments accuracy. Godbole [9] analyzes news sentiments and blogs. It splits work in the context of their specific

ISBN : 978-93-87793-82-8

task sentiment analysis for news and blogs into two categories. First category which regards with techniques for automatically creating sentiment lexicon and the second one which relates to systems that analyze sentiment for entire documents.

Esuli [10] proposed the sentiment evaluation which refers to get the sentiment polarity -positive, negative, or neutral of a text reviews data and evaluate the sentiment score of the text review. The previous research on sentiment-based categorization of the input documents has implicated either the using models inspired mostly by cognitive linguistics [11] or the manual or semi-manual construction of discriminate word lexicons [12]. Turney [13] introduced a new method for sentiment extraction in real time in the domain of finance; which is working based on messages from web-based stock message boards, attempt to automatically label each message. Common sources of opinionated texts have been movie and product reviews [14], [15], [16], blogs [17], [18], [19] and Twitter posts [20], [21], [22].

3.KNN ALGORITHM

KNN is a supervised learning algorithm where the result is classified based on the majority vote from its K nearest neighbor category. The algorithm works based on minimum distance from the test data to the training samples to determine the K nearest neighbor. After getting K nearest neighbor a simple majority of them is taken to make prediction of test data. The KNN works as follows: The distance between the test data and all the training samples are calculated. The distance may be calculated by any standard means .Example Euclidean distance. The K nearest neighbor may be included if the distance of the training samples to the query is less than or equal to Kth smallest distance. We then gather a particular feature value of all the nearest neighbors training samples. We take the simple majority of this value as prediction and categorize our new test data.

The KNN algorithm can compete with the most accurate models because it makes highly accurate predictions. Therefore, we can use the KNN algorithm for applications that require high accuracy but that do not require a human-readable model. The quality of the predictions depends on the distance measure. Therefore, the KNN algorithm is suitable for applications for which sufficient domain knowledge is available. This knowledge supports the selection of an appropriate measure. The KNN algorithm is a type of lazy learning, where the computation for the generation of the predictions is deferred until classification.

4.PROPOSED WORK

The use of Emoji on the social network increased significantly in the past few years and one of the most popular social network platforms is Twitter. The usage of Emoji on Twitter has increased, similar to other social networks. we choose Twitter as data source for analyses. In order to evaluate the effects of using Emoji characters on the Sentiment Analysis models, the expressivity of the characters should be investigated. The

ISBN : 978-93-87793-82-8

aim is to study the effects of Emoji characters on Sentiment Analysis models. We analyzed whether people use Emoji more frequently in positive or negative life events. We observed that the usage of Emoji characters in sentiment analysis appeared to have higher impact on overall sentiments of the positive opinions in comparison to the negative opinion.

4.1 EMOJI PREPROCESSING

This module is used to convert the color image into gray scale conversion. Color image consist of all natural colors which is the combination of RED BLUE and GREEN. The color values of the pixels range from 0 to 255.The gray scale converted image contains only gray combination of Pixels.

4.2 EMOJI SEGMENTATION

Segmentation is one of the key problems in image processing. Image segmentation is the process that sub divides an image into its constituent parts or objects. Thresholding techniques are used for image segmentation.

Fig 1: Emoji Segmentation

4.3 FEATURE EXTRACTION

In feature extraction desired feature vectors such as color, texture, morphology and structure are extracted. Feature extraction is method for involving number of resources required to described a large set of data accurately. Statistical texture features are obtained by Gray level co-occurrence matrix formula for texture analysis and texture the specified position relative to others. Numbers of gray levels are important in GLCM . Different statistical texture features of GLCM are energy, sum entropy, covariance, information measure of correlation, entropy, contrast and inverse difference and difference entropy.

ISBN : 978-93-87793-82-8

Fig 2: Emoji Feature Extraction

4.4 CLASSIFICATION

Classification is one of the most often used methods of information extraction. In classification ,usually multiple features are used for set of pixels ,i.e., many images of a particular objects are needed. Image classification is the labeling of a pixel or a group of pixel based on its grey value .

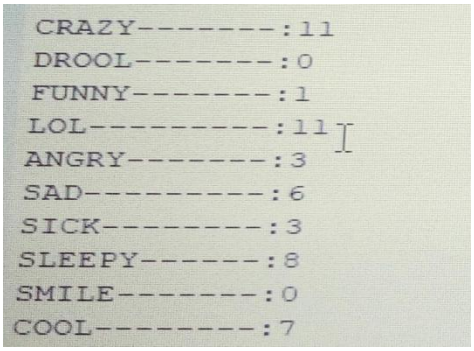

Fig 3: Emoji Classification based on Features

5.CONCLUSION

This work is based on image classification and learning the sentiments based on number of emojis we get in each category. This work is implemented in MATLAB. Initially a emoji detection step is performed on the input image. Afterwards an image processing based feature point extraction method is used to extract the feature points. Finally, a set of values obtained from processing the extracted feature points are given as input to recognize the emotion contained. Likewise, all the emojis in any given dataset are processed and the count on each emotion based on classification results are found. This gives us a successful outcomes in the area of emotion using emojis . In future, Text based sentiment analysis can be combined with emojis based emotion analyses to give better results.

ISBN : 978-93-87793-82-8

REFERENCES

[1] Penubaka Balaji,D. Haritha, "An Overview on Opinion Mining", International Journal of Pure and Applied Mathematics, Vol 118, 2018

[2] Anchal Kathuria, , Dr. Saurav Upadhyay, " A Novel Review of Various Sentimental Analysis Techniques" , IJCSMC, Vol. 6, Issue. 4, April 2017, pg.17 – 22

[3] Mohammed O. Shiha1, Serkan Ayvaz, "The Effects of Emoji in Sentiment Analysis", International Journal of Computer Electrical Engineering,March 2017

[4] Yuxiao Chen, Jianbo Yuan, Quanzeng You, "Twitter Sentiment Analysis via Bi-sense Emoji Embedding and Attention-based LSTM" , arXiv:1807.07961v2 [cs.CL] 7 Aug 2018.

[5]Schukla, A., "Sentiment analysis of document based on annotation", CORR Journal, Vol. abs/1111.1648, 2011.

[6] Kasper, W. & Vela, M., "Sentiment analysis for hotel reviews", proceedings of the computational linguistics-applications, Jacharanka Conference, 2011.

[7] Zhang, L., Hua, K., Wang, H., and Qian, G.,"Sentiments reviews for mobile devices products", The 11th International Conference on Mobile Systems and Pervasive Computing (MobiSPC-2014) ,procedia computer scinence, Volume 34, 2014.

[8] Nisha, J., & Dr.E. Kirubakaran, "M-Learning sentiment analysis with Data Mining Techniques", International Journal of Computer Science and Telecommunications, Volume 3, Issue 8, 2012.

[9] Godbole, N., Srinivasaiah, M., and Skiena, S., "Large-Scale Sentiment Analysis for News and Blogs", ICWSM'2007 Boulder, Colorado, USA, 2007.

[10] Esuli A., & Sebastiani, F.,"SentiWordNet: A High-Coverage Lexical Resource for Opinion Mining", Kluwer Academic Publishers. Printed in the Netherlands, 2006.

[11] Hearst, M., "Direction-based text interpretation as an information access refinement", In Paul Jacobs, editor, Text-Based Intelligent Systems. Lawrence Erlbaum Associates, 1992.

[12] Das, S., and Chen, M., "Yahoo! for Amazon: Extracting market sentiment from stock message boards", In Proc. of the 8th Asia Pacific Finance Association Annual Conference (APFA 2001), 2001.

[13] Turney, P., "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews". In Proc. of the ACL, 2002.

[14] Pang, B. and Lee, L.,Seeing stars," Exploiting class relationships for sentiment categorization with respect to rating scales". Proceedings of the Association for Computational Linguistics (ACL),2005

ISBN : 978-93-87793-82-8

[15] Wiebe, J. and Rilo, E. ," Creating subjective and objective sentence classifiers from un-annotated texts". In Proceedings of CICLing 2005,

[16] Yu, H. and Hatzivassiloglou, V., "Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences." In Proceedings of the 2003 conference on Empirical methods in natural language processing- Volume 10, Association for Computational Linguistics.

[17] Harb, A., Plantie, M., Dray, G., Roche, M., Trousset, F., and Poncelet, P., "Web opinion mining: how to extract opinions from blogs?". In Proceedings of the 5th international c nference on Soft computing as transdisciplinary science and technology CSTST'08. ACM, New York, NY, USA, pp. 211-217

[18] Yang, H., Si, L., and Callan, J. ," Knowledge transfer and opinion detection in the TREC 2006 blog track." In Proceedings of TREC 2006.

[19] Yang, C., Hsin-Yih Lin, K., and Chen, H. ," Emotion classification using web blog corpora." Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, 2007

[20] Gruhl, D., Guha, R., Kumar, R., Novak, J. and Tomkins, A. " The predictive power of online chatter". Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, 2005

[21] Jansen, B.J., Zhang, M., Sobel, K., and Chowdury, A. ," Twitter power: Tweets as electronic word of mouth." Journal of the American Society for Information Science and Technology,2009.

[22] Thelwall, M. and Prabowo, R.,"Identifying and characterising public science-related concerns from RSS feeds". Journal of the American Society for Information Science and Technology, 2007.