Evaluation and Extraction of Patterns to find Human Behavior through Sentimental Analysis

***Mohammad Shabaz [1]Dr. Urvashi Garg**

Abstract

Pattern evaluation in human behavior for detecting human's opinion, emotions and feeling has a wide importance in this era. Emergence of Digitization, social media, news, foreign relations and many other directed us to find, evaluate and extract the sentiments of an individual or organization from a certain data or data-set, images etc based on certain parameters related to human behavior. Evolution of data helps us to detect the gaps occur in storage and processing. It is very difficult to process unstructured data and finds the patterns. With the large growth of digital data, it has overwhelmed all the traditional systems and methods. The driven data finds approach for computing. Thus before finding patterns it is required to formalize the data. In this article we are going to elaborate some of the parameters which enable us to evaluate and extract parameters to find human behavior.

Keywords: Sentimental Analysis, Human Behavior, Digitization, Social Media.

Introduction

Data management systems start automating traditional tasks. The data mainly consisted of digits and character strings. This ever-growing data provides a tidal wave of chances and issues in terms of visualization, data capture and many more. The Hindu mythology also gives the reference of their ancient records such as their ancient wars and gods. The biggest evidence of such preference of data is Puranas. To sort, analyze, and process the data, software based instruction are used. The response to these new technologies was predictable. Traditionally, there had been a disassociation between programs and data. As new applications appeared, the disassociation between programs and data became problematic. Fundamental methodology to perform sentimental analysis on text data can be done using AS Sentimental approach in which the sentences or paragraphs are tokenize into words [1].The problem occur when a particular word is not correctly spelled, same problem also occurs when we perform image segmentation since the boundaries are not defined, thus we are not able to apply any filter for classifying these incorrect spelled words or sentences or images to detect the expressions. With words it is very important to find the right expression. When we are talking about textual data then we can stem a particular sentence which results in the formation of adjectives and adverbs from the sentence, complete the process of tokenization of words and target those words from the defined repository which matches most of the characters of targeted words or sentences.

Literature Review and Related Work

Mohammad Shabaz and Ashok Kumar [1]: The authors in this article performed sentimental analysis using a novel approach called AS. They extract the data from social networking sites and tokenize it into words and compare them with predefined repository then using association rule mining to evaluate the sentiment count in numbers. The main limitation of this article is the validity. This article also defines the evolution of data since the data taken from Social networking sites are completely unstructured.

Jim Gray [2]: In this article the author give a detail study on the evolution of data from early 4000BC to 1995 and there-after. The author outlines that the relational systems offered huge improvements in ease-of-use, graphical interfaces, client-server applications, distributed databases, parallel data search, and data mining.

Mohammad Arif, Mandeep Singh and Rajdavinder Boparai [3]: The author's uses cross platform languages to detect sentiments. This paper mentioned the difficulty of incorrect spelling of words and works hard to remove all such language barriers which act as speed breakers in the smooth running of the process of finding sentiments. The main limitation of this article is Unicode matching. Different regional languages have different syntax thus to classify them into particular repository is difficult to achieve. The authors in this article uses cross platform languages to detect sentiments on regional languages. This paper mentioned the difficulty of incorrect spelling of words as the data chosen is Unicode based and works hard to remove all such language barriers which act as speed breakers in the smooth running of the process of finding sentiments.

Carlo Strapparava and Rada Mihalcea [4]: The authors have found that problems faced by the students includes load of subjects, assignments, secession tests, sleeping deficiency. Later on, they performed a review on other approaches executed on different datasets like Movies, novels, elections etc. Most work has been performed to detect sentiments is on English language. The main limitation in the article is noisy nature of data-set. Thus this article tells about natural language processing on different data-sets.

Wiebe, J. [5]: The author states about the English language which is universally accepted and much of the work of sentimental analysis is done using it.

Proposed work and Methodology

The methodology for evaluation and extraction of patterns has been discussed in the following points:

1. Get the data in any format and save it on any file (say text-file).
2. Set the Positive (p) and Negative (n) word and segmented image blob repository.

3. Set the value of longest sequence matching of pixels in case of image and characters in case of text.
4. Tokenize the Input into segments.
5. Start traversing and comparing.
6. Repeat the process until end of file.

Conclusion

In past recent years a relatively small data used to get collected and analyzed by online analytical processing (olap) and data mining technologies. But in recent years of digital era, an enormous amount of data is generated from various sources like sensor devices, internet and other channels [6]. This data can again not be managed by traditional relational database management system (rdbms) and data warehouse techniques. This upcoming technology of handling the huge data has become a market of great profit in business organizations. Therefore, corporate like oracle, google, yahoo and amazon have come up with a solution becoming the pioneer in the field of big data. They introduced advanced technologies like hadoop, map-reduce etc [7]. Incorrect spellings cannot be detected easily especially when performing sentimental analysis after tokenization of sentences into words and images. This article reveals the methodology of tokenization of words, images and detects using longest sequence matching [8].

Results

The proposed methodology is implemented on r-studio to tokenize the words using “quanteda” package. The following results are generated which are shown in Table 1.

S.No	Text	Minimum character size (MCS)	Positive word	Negative word
1	He is a very bad boy	3	0	1
2	Sheetal has been brutaeally rapied .	3	0	2
3	Lovea is blind	3	1	0
4	Shutup you Mentally sick person.	4	0	2
5	Internet is a resourceful place for information and Knowledge.	5	1	0

Table 1: Results obtained on Evaluation and Classification of Positive and Negative words from Noisy Text.

The results on images were in progress soon near in our next article it will be shown.

Acknowledgement

I would like to thank with whole heartedly Mr. Mohd Rafiq, Mr. Dilbag Singh, Mr. Munish Sabarwal Dr. Sanjeet Singh for their continue support and encouragement.

Reference

[1] Jim Grey,"Data Management: Past Present and Future", IEEE Conference 301 howard street sanfrancisco,pp. 38-46, June 1996.

[2] Mohammad Shabaz and Ashok Kumar, "AS: a novel Sentiment analysis approach", *International Journal of Engineering and Technology,* vol. 2.27, pp. 46-49, 2018.

[3] Mohammad Arif, Mandeep Singh and Rajdavinder Boparai, "Emotion detection on cross platform languages ", *International Journal of Engineering and Technology,* vol. 2.27, pp. 27-31, 2018.

[4] Carlo Strapparava, Rada Mihalcea, "Learning to Identify Emotions in Text", SAC'08 Fortaleza, Brazil 2008 https://doi.org/10.1145/1363686.1364052.

[5] Wiebe, J., "Learning subjective adjectives from corpora", Proceed-ings of the 17th National Conference on Artificial Intelligence (AAAI-2000), Austin, Texas, 2000

[6] Walter, K., & Mihaela, V., "Sentiment analysis for hotel reviews", *computational linguistics-applications, Jacharanka Conference,* 2011

[7] Pang B, Lee L," A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts", *Association for Computational Linguistics*, Stroudsburg, PA, USA, 2004. https://doi.org/10.3115/1218955.1218990.

[8] R. Piryani, D. Madhavi, and V. K. Singh, "Analytical mapping of opinion mining and sentiment analysis research during 2000–2015," *Information Processing & Management*, vol. 53, no. 1, 2017, pp. 122–150. https://doi.org/10.1016/j.ipm.2016.07.001.

[9] Medhat Walaa, Hassan Ahmed, Korashy Hoda, "Sentiment analysis algorithms and applications: A survey",Ain Shams Engineering Journal, *Elsevier,* Vol. 5, May 2014.

[10] Z. Nanli, Z. Ping, L. Weiguo and C. Meng, "Sentiment analysis: A literature review", 2012 International Symposium on Management of Technology (ISMOT), 2012. https://doi.org/10.1109/ISMOT.2012.6679538.

[11] Moreo A, Romero M, Castro JL, Zurita JM. "Lexicon-based com-mentsoriented news sentiment analyzer system" Expert Syst Appl, 39:9166–80, 2012. https://doi.org/10.1016/j.eswa.2012.02.057.

[12] Muhammaad Zubair, Aurangzeb Khan, Shakeel Ahmad, Fazal-MasudKundi and Asghar, 2014.A Review of Feature Extraction in Sentiment Analysis. ISSN 2090-4304 Journal of Basic and Applied Scientific Research.