Chapter 36
Implementation of n-gram Methodology for Rotten Tomatoes Review Dataset Sentiment Analysis

Prayag Tiwari

National University of Science and Technology MISiS, Department of Computer Science and Engineering, Moscow, Russia

Brojo Kishore Mishra

C. V. Raman College of Engineering, Department of Information Technology, Bhubaneswar, India

Sachin Kumar

Indian Institute of Technology Roorkee, Center for Transportation Systems, Roorkee, India

Vivek Kumar

National University of Science and Technology MISiS, Department of Computer Science and Engineering, Moscow, Russia

ABSTRACT

Sentiment Analysis intends to get the basic perspective of the content, which may be anything that holds a subjective supposition, for example, an online audit, Comments on Blog posts, film rating and so forth. These surveys and websites might be characterized into various extremity gatherings, for example, negative, positive, and unbiased keeping in mind the end goal to concentrate data from the info dataset. Supervised machine learning strategies group these reviews. In this paper, three distinctive machine learning calculations, for example, Support Vector Machine (SVM), Maximum Entropy (ME) and Naive Bayes (NB), have been considered for the arrangement of human conclusions. The exactness of various strategies is basically inspected keeping in mind the end goal to get to their execution on the premise of parameters, e.g. accuracy, review, f-measure, and precision.

INTRODUCTION

The Internet has altered the way people express their points of view. It is now done through the help of blog entries, online gatherings, item audit sites and so on. People rely on upon this client made dataset. When some person needs to buy a thing, they, as a rule, need to know its reviews through online before taking a decision. The measure of customer made dataset is excessively broad for a typical customer, making it impossible to examine. So, to computerize this, distinctive supposition analysis procedures are utilized. Sentiment analysis, otherwise called opinion mining, dissects individuals' opinion and additionally feelings towards datasets, for example, items, associations, and their related attributes. Machine learning proposal makes use of a planning set to add to a supposition classifier those gatherings suspicions. Sentiment analysis (Liu, 2012) is seen to be done in three distinct levels, for example, aspect level, document level and sentence level. Document level characterizes whether the record's opinion is negative, neutral or positive. Sentence level figures out if the sentence communicates any negative, positive or neutral opinion.

There is generally two types of machine learning techniques (Han et al., 2006) which has been used more often in sentiment analysis are unsupervised learning and supervised learning method. In supervised learning, we are provided a dataset and already having idea that what and how our output would look like and the idea that there is a relationship between the output and input. On the other hand, unsupervised learning (Kumar and Toshniwal, 2016) enables us to get problems with having little or do not have an idea that what and how our results supposed to look like. We can obtain structure from data where we don't know the effect of the variables.

The film reviews are generally in the text format and not structured in nature. Therefore, the stop words and other undesirable data are expelled from the reviews for further investigation. These frameworks are then offered a contribution to many machine learning methods for the arrangement of the surveys. Distinctive parameters are then used to assess the execution of the machine learning calculations.

The primary commitment of the paper can be expressed as takes after:

- There are many different kinds of machine learning techniques has been suggested to classify the film reviews of Rotten Tomatoes dataset using n-gram method viz., Bigram, Unigram, Trigram, an amalgamation of bigram and trigram, unigram and bigram and unigram and bigram and trigram.
- There are three machine learning techniques which SVM, NB and ME for purpose of classification by the help of n-gram proposal.
- The implementation of machine learning methods is estimated with the help of variables like recall, accuracy, precision and f-measure. The output acquired in this work demonstrates the better accuracy by comparing by other research works.

LITERATURE SURVEY

The literature review on the sentiment analysis shows the good research has been done by the various researchers based on sentiment analysis on document level.

In this paper was suggested diverse multi-mark order on sentiment analysis (Liu and Chen, 2015). They have utilized eleven multilevel characterization method seemed at on two smaller scale blog dataset furthermore eight distinctive assessment networks for examination. Aside from that, they have addition-

11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/implementation-of-n-gram-methodology-for-rotten-tomatoes-review-dataset-sentiment-analysis/252052

Related Content

Implicit Processes and Emotions in Stereotype Threat about Women's Leadership
Gwendolyn A. Kelsoand Leslie R. Brody (2015). *Exploring Implicit Cognition: Learning, Memory, and Social Cognitive Processes (pp. 118-137).*
www.irma-international.org/chapter/implicit-processes-and-emotions-in-stereotype-threat-about-womens-leadership/120856

Inferring Intent and Action from Gaze in Naturalistic Behavior: A Review
Kristian Lukander, Miika Toivanenand Kai Puolamäki (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 1464-1482).*
www.irma-international.org/chapter/inferring-intent-and-action-from-gaze-in-naturalistic-behavior/252092

Explore the Use of Handwriting Information and Machine Learning Techniques in Evaluating Mental Workload
Zhiming Wu, Tao Linand Ningjiu Tang (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 1423-1439).*
www.irma-international.org/chapter/explore-the-use-of-handwriting-information-and-machine-learning-techniques-in-evaluating-mental-workload/252090

Auto Associative Extreme Learning Machine Based Hybrids for Data Imputation
Chandan Gautamand Vadlamani Ravi (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 867-893).*
www.irma-international.org/chapter/auto-associative-extreme-learning-machine-based-hybrids-for-data-imputation/252061

Gaming Cultural Atonement: Healing Collective Personae With Mediated Biofeedback
Stephen Brock Schafer (2019). *Media Models to Foster Collective Human Coherence in the PSYCHecology (pp. 237-262).*
www.irma-international.org/chapter/gaming-cultural-atonement/229339