ISBN 978-81-934533-1-5

Sentiment Analysis of Movie reviews based on the movie genre classifying on age group for Indian Viewers from a popular booking website in Python

Ashish Modi[1], Elizabeth L George[2]
[1 2]Assistant Professor
[1 2]Nagindas Khandwala College, Malad – west, Mumbai
[2]e.leahgeorge@gmail.com

***Abstract*—Sentiment Analysis and Opinion Mining is an adverse area of research that deals with the judgement and evaluation criteria in an Emotional communication. The research initiates by classifying the movies into different genre based on the age group. In this research we have assumed only five categories such as : Animation, Comedy , Action, Romance , Epics for different age group such as Kids, Teens, Young adults, Adults and Family. , This research understands the polarity of sentiments based on movie review and the rating given by different users based on comments and rating stars. Sentiment analysis is generally classified into three different level namely they are document level, sentence level and entity-aspect level. Several research papers are published based on machine learning techniques such as support vector machine, maximum entropy(max ent) and naïve bayes classifiers and random forest classifiers as the most widely used algorithm in sentiment analysis.**

Our approach is to take the unstructured data of movie reviews based on a popular booking website book my show, convert into structured form with a defined bag of words then extracts features from structured review using syntactic feature selection method that uses phrase patterns such as "n+aj" (noun followed by positive adjective) typically represent positive sentiment orientation, while "n+dj" (noun followed by negative adjective) often express negative sentiment then classification technique is applied on extracted features to classify them into its sentiment polarity that is namely either positive , negative or neutral. TextBlob package and Python data analysis tool kit (PANDAS) is used to classify the sentiment polarity.

***Keywords*—Sentiment analysis, machine learning, Support Vector machine, Naïve bayes classifier, Sentiment polarity.**

I. INTRODUCTION

Opinion mining is a growing field that identify the thoughts and sentiments of people, which they express in form of their feedbacks or reviews on various things. Today due to vast use internet and social platforms, people are having a huge amount of space where they can publically express their opinions.

Sentiment Analysis (SA), also called Opinion Mining, is currently one of the most studied research fields. Different techniques and software tools are being developed to carry out Sentiment Analysis.Sentiment analysis of online user generated content is important for many social media analytics tasks. A lot of work has been carried out for extracting people sentiments from textual data.

Opinion Mining and Sentiment Analysis covers techniques and approaches that promise to directly enable opinion-oriented information-seeking systems. The focus is on methods that seek to address the new challenges raised by sentiment-aware applications, as compared to those that are already present in more traditional fact-based analysis.

In this paper , the main focus is on Sentiment Analysis that identifies the sentiment expressed in a text then analyzes it. Therefore, the target SA is to find opinions, identify the sentiments they express, and then classify their polarity.

Sentiment Analysis has three main classification levels they are document-level, sentence-level, and aspect-level SA. Document-level SA aims to classify an opinion document as expressing a positive or negative opinion or sentiment. It considers the whole document a basic information unit (talking about one topic). Sentence-level SA aims to classify sentiment expressed in each sentence. Sentence-level SA will determine whether the sentence expresses positive or negative opinions.

Sentences can also be classified as short documents. Aspect-level SA aims to classify the sentiment with respect to the specific aspects of entities. Our research focuses on sentence level SA. [1]

The main aim of this paper is to analyse the existing methods and choose the best approach in finding the techniques of sentiment classification.

1. Methodology : Sentiment analysis classification

There are numerous number of articles presented every year in the SA fields. The survey conducted by Waala Medhat et al [1] gives us a panoramic view of the different available benchmarks data sets in related fields to SA including emotion detection, building resources and transfer learning.

Machine learning technique applied on Movie review dataset proved that machine learning technique performs well than human generated result [2]. Text databases are increasing day by day due to large collection of information in from of electronic document so information retrieval is the process through which information is retrieved from large collection of textual database. Support vector machine, Maximum Entropy(MaxEnt) and naïve bayes classifiers are the most