Chapter 5
Classification Approach for Sentiment Analysis Using Machine Learning

Satyen M. Parikh
Ganpat University, India

Mitali K. Shah
Ganpat University, India

ABSTRACT

A utilization of the computational semantics is known as natural language processing or NLP. Any opinion through attitude, feelings, and thoughts can be identified as sentiment. The overview of people against specific events, brand, things, or association can be recognized through sentiment analysis. Positive, negative, and neutral are each of the premises that can be grouped into three separate categories. Twitter, the most commonly used microblogging tool, is used to gather information for research. Tweepy is used to access Twitter's source of information. Python language is used to execute the classification algorithm on the information collected. Two measures are applied in sentiment analysis, namely feature extraction and classification. Using n-gram modeling methodology, the feature is extracted. Through a supervised machine learning algorithm, the sentiment is graded as positive, negative, and neutral. Support vector machine (SVM) and k-nearest neighbor (KNN) classification models are used and demonstrated both comparisons.

DOI: 10.4018/978-1-7998-4042-8.ch005

INTRODUCTION

Sentiment Analysis

The word opinion mining is utilized in many different applications for sentiment analysis. Its purpose is to check the opinion of the people on the product's benefits and product's characteristics (features). Essentially, an evaluation of opinion is a kind of common approach used to determine whether or not people are interested in the product. The emotions or feelings of people are expressed textually. These aspects are obtained from different sites or mobile applications. (Duygulu et al., 2002)

Analysis of sentiment is used to keep in touch or show the feelings of people. People give product's positive and negative opinion for the services of item. When making any choice for purchase the product, these opinions are very valuable for the customer (Zhang et al., 2013). It is very difficult to correctly interpret the right view in the sense of large textual data set and unstructured data sets. To identify the characteristics of an unstructured dataset, it is essential to structure a productive technique (Yu et al., 2010).

Preprocessing, feature extraction and classification are three main parts and analysis methods for sentiment analysis (Xu et al., 2013). Some association only works with positive and negative text. They skip the neutral text recognition. Such texts use the binary classification limits. Several researchers introduce polarity problems for perform three classes (Ren & Wu, 2013). For entropy and SVM classifiers, the neutral classes, presentation etc. are very essential. This improves the accuracy of overall classification. These principles ought to be considered by the neutral classes for the execution. The first algorithm understands the concept of neutral language and extracts people's remaining opinions. In simply single phase, this algorithm completes the three-level classification (Poria et al., 2013).

In every class, in the next methodology, the probability and distribution are calculated (Garcia-Moya et al., 2013). When the data is the most neutral with the variance between positive and negative outcomes, it becomes more difficult to implement this methodology (Cheng et al., 2013). Sentiment analysis is used in some various ways. It is useful for marketers to measure the credibility and accomplishment of some new product dispatch and to see what edition of the product is in demand and show the new product's famous highlights (Hai et al., 2013)

The below are some of the key features of tweets:

- Message Length
- Writing technique
- Availability
- Topics