S

Sentiment Analysis Using LSTM

Anweshan Mukherjee
St. Xavier's College (Autonomous), India

Rajarshi Saha
St. Xavier's College (Autonomous), India

Ashwin Gupta
St. Xavier's College (Autonomous), India

Debabrata Datta
St. Xavier's College (Autonomous), India

Anal Acharya
St. Xavier's College (Autonomous), India

INTRODUCTION

Sentiment Analysis is a field of study in Natural Language Processing (NLP) domain that focuses on determining the sentiment of data given as input; mostly, classified into three classes – negative, neutral and positive. It focuses on identification and classification of opinions or sentiments conveyed in the source text (Neethu, M. S., & Rajasree, R, 2013). Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, tweets, and database sources through NLP. It is also referred to as subjectivity analysis, opinion mining, and appraisal extraction. Analysis of user generated data to extract the sentiment or opinion of the crowd is of prime importance in the real world (Kharde, V., & Sonawane, P, 2016).

There are three basic methodologies of Sentiment Analysis:

1. Symbolic techniques or Rule-based approach
2. Machine learning techniques or Automatic approach
3. Hybrid techniques

Rule-based techniques require an outsized database of predefined emotions and sentiments and an efficient knowledge representation for classifying sentiments properly (Neethu, M. S., & Rajasree, R, 2013). In rule-based approach, we use a set of human-crafted rules or guidelines to help determine the subjectivity, polarity, or the subject of an opinion. Rule-based systems are very naive since they do not consider how words are combined in a sequence. Although more advanced processing techniques can be used, and new rules added to support new expressions and vocabulary. However, adding new rules may affect the previous results, and eventually the whole system may get very complex. Since rule-based systems often need fine-tuning and maintenance, they also need regular investment.

DOI: 10.4018/978-1-7998-9220-5.ch057

Machine learning approach in sentiment analysis involves using a training set to train and develop a sentiment classifier model that categorizes or classifies sentiments. It is simpler than rule-based approach since such a large database of predefined emotions or sentiments is not required (Neethu, M. S., & Rajasree, R, 2013). Machine learning techniques for sentiment classification are useful because they are able to capture the context accurately by modelling many features efficiently. They are capable of adapting to changing input and can calculate as a part of the process the degree of uncertainty of classification, making them a suitable technique for many applications (Boiy, E., & Moens, M. F, 2008). Machine learning techniques or Automatic approaches rely on different machine algorithms to classify opinions. In this technique, a sentiment analysis task is modelled as a classification problem, where a classifier is loaded with a text and returns a category, e.g., positive, negative, or neutral. This approach involves a training and a prediction process. In the training process the model learns to associate a particular input i.e., a text to the corresponding output or tag, based on the test samples used for training. The feature extractor then transfers the text input into a feature vector. Pairs of feature vectors and tags (positive, negative, or neutral) are input into the machine learning algorithm to generate a model. In the prediction process, the feature extractor is used to transform unknown text inputs into feature vectors. These feature vectors are then input into the model, which then generates the predicted tags (positive, negative, or neutral). For classifying the text, various statistical models may be used such as Naïve Bayes (NB), Support Vector Machines (SVM), Linear Regression, and Deep Learning.

Finally, hybrid approaches combine the desirable elements of rule-based and Machine Learning techniques into one system. The main advantage of this approach is that the results are often more accurate.

The research work highlighted in this paper uses a machine learning based approach for classifying texts into three classes – negative, positive and neutral. The methodology proposed here was also deployed into a web application so that anyone can use it, and if developed further, can be used in several critical applications such as in the field of psychology or, detecting the overall opinion from the reviews obtained for any product or, predicting the sentiment of a suspected criminal during interrogation.

Background

Pang et al. (2002) tried to tackle the problem of classifying a document based on its overall sentiment, instead of classifying it by topic. The document contained a review, and the aim was to determine whether the review is positive or negative. They used movie reviews as their data and found out that machine learning techniques outperformed human-produced rules and guidelines. However, it was also observed that the three Machine Learning methods that were employed, namely Naïve Bayes (NB), Maximum Entropy Classification (ME) and Support Vector Machines (SVM) do not perform as well on sentiment classification as they do on traditional topic-based categorization. On researching the effects of the three machine learning techniques (NB, ME and SVM) in the specific domain of movie reviews, they were able to achieve an accuracy of 82.9% using SVM and an unigram model. The results produced by machine learning techniques were better in comparison to the human generated guidelines. In terms of relative performance, Naive Bayes performed the worst and SVMs performed the best, although the differences are not very large. Despite trying out several different types of features, they were unable to achieve accuracies on the sentiment classification problem comparable to those reported for standard topic-based categorization.

In another research work, Turney (2002) introduced a simple unsupervised algorithm for rating a review as positive (thumbs up) or negative (thumbs down). The classification of a review was predicted by the average semantic orientation of the words and phrases in the reviews that contained adjectives