Title

Real-time Sentiment Analysis of Twitter Data Streaming

Introduction

This project is to perform a basic sentimental analysis of real-time tweets. Here the Bag-Of-Word (BOW) method is used to decompose the tweets in a fixed time interval into two classes of sentimental polarities. Technically, the real-time text data stream is handled using the Kafka distributed streaming platform in order to balance the streaming speed and processing speed. From these real-time analysis, it is obvious to track the sentiment trend over a time period.

My Contribution

* Set up the data streaming pipeline using Kafka distributed streaming platform, Twitter streaming API and Spark Streaming.
* Code the BOW sentiment analysis method using Spark RDDs and MapReduce API.
* Data visualization using python Matplotlib package.

Reference