Opinion Mining of Twitter Data using Hadoop and Apache Pig

Abstract

Twitter, one of the largest and famous social media site receives millions of tweets every day on variety of important topic. This large amount of raw data can be used for industrial, Social, Economic, Government policies or business purpose by organizing according to our need and processing. Hadoop is one of the best tool options for twitter data analysis and hadoop works for distributed Big data, Streaming data, Time Stamped data, text data etc. This paper discuss how to use FLUME for extracting twitter data and store it into HDFS for opinion mining because twitter contains variety of opinions on various topics so we have to analyse these opinions using hadoop and its ecosystems to check every tweets polarity either tweets contains positive, negative or neutral opinions on particular topic. This paper provides an efficient mechanism to perform opinion mining by coming up with a finish to finish pipeline with the assistance of Apache Flume, Apache HDFS, and Apache Pig.

Here we have used dictionary based approach for analysis for which we have implemented pig statements through which we can analysis these complex twitter data to check polarity of the
tweets based on the polarity dictionary through which we can say that which tweets have negative opinion or positive opinion.

**References**


**Index Terms**

Computer Science  Information Sciences
Keywords

Hadoop, twitter, Flume, opinion mining, social analysis, apache pig.