International Conference on
Big Data Analysis and Data Mining

May 04-05, 2015   Kentucky, USA

Sentiment analysis of twitter data using parallel write approach of replica placement in Hadoop cluster

Divyesh Patel

Charotar University of Science & Technology, India

In recent years social networking has turned out to be very popular. Twitter, a micro-blogging service, is appraised to have about 200 million registered users and these users create about 65 million tweets a day. Twitter users usually express their views about topics of their interest. The challenge is that each tweet is partial up to 140 characters, and is hence very short. It may contain jargon and misspelled texts. Thus, it is hard to apply traditional NLP techniques which are designed for working with formal languages, into Twitter domain. One more challenge is that the total volume of tweets is tremendously high, and it takes a long time to process. In this project, we have described a Hadoop distributed system for real-time Twitter sentiment analysis. Our system consists of three components: A lexicon builder, a sentiment classifier and Hadoop new distributed file system for replica placement. These three components are capable of running on a large-scale distributed system since they are implemented using a Hive, Flume Map Reduce framework, HBase database model and other Hadoop environment. Thus, our sentiment classifier and lexicon builder are scalable with the number of machines and the size of data. The experiments also show that our lexicon has a good quality in opinion extraction, and the accuracy of the sentiment classifier can be improved by combining the lexicon with a machine learning technique. Another Part in our project is of HDFS; we are experiencing an information explosion era. Due to which huge amount of distributed data is being achieved and put in storage? To manage this type of data, application uses distributed file system. For replica placement the HDFS is used to save time and for better reuse. Recent state of the art design & implementation of the HDFS implement the new approach for efficient replica placement in Hadoop DFS which can improve throughout and data transfer rate.

Biography

Divyesh Patel has completed his BTech from Charotar University of Science and Technology, and currently pursuing MTech in Computer Engineering at Charotar University of Science and Technology. His research interests are Data Mining and Business Analytics. He played a role of General Secretary at his University during his graduation. He is expert in many fields and areas like management and engineering. He has done various projects with Government of Gujarat, India.

09it088@charusat.edu.in

Notes:

J Comput Sci Syst Biol 2015

ISSN: 0974-7230, JCSB an open access journal

Data Mining-2015
May 04-05, 2015

Volume 8 Issue 3

Page 30