SOCIAL NETWORK SENTIMENT ANALYSIS APPLICATION

Nemchenko M.S., Liubchenko N.Yu.
National Technical University
«Kharkiv Polytechnic Institute»,
Kharkiv

In recent years, most people have been attracted to social-networking platforms. Most use social sites to convey their emotions, beliefs or opinions regarding things, places or personalities. Everyday users make millions of posts to broadcast their opinions on a variety of topics and share their personal feelings. The wealth of information generated on social networks makes it a «Big Data» source of thoughts.

There is a space for performing challenging researches in broad areas by computationally analyzing opinions and sentiments [1, 2]. Therefore, a gradual practice has grown to extract the information from data available on social networks for the prediction of an election, to use for educational purposes, business intelligence, customer relationship management, recommendation systems, and marketing.

When it comes to handling and discerning such amounts of information produced by some outside system or systems as a continuous and infinite stream, Stream Processing (SP) is used. SP technologies are able to scan huge volumes of data and find actionable insights in near-real time. However, SP cannot exist by itself, it is integrated with the rest of the Big Data infrastructure to deliver real-time processing capabilities and can be added to an existing Big Data infrastructure as a new service. Designing and implementing such applications can be a real challenge as there is no general-purpose architectural and technological solution.

Thus, a high-throughput, fault-tolerant, scalable application was created to perform opinion mining with near instantaneous insights. The application, or rather the set of applications, making up a whole Big Data pipeline, pulls posts related to some specified keywords, drops them to an aggregator, classifies them as positive or negative, and exports results to operational storage. On top of that, a web application was developed in order to provide end users with a convenient and informative interface to interact with the system.

The application consists of several modules: social network consumer that uses Apache Kafka persistent message queue system, Apache Spark stream processor, Machine Learning unit, Redis in-memory lookup database, Cassandra operational storage, HDFS as delivery of ingestions and an asynchronous web client.

References:
1. Chetviorkin I. Extraction of Russian Sentiment Lexicon for Product Meta-Domain / I. Chetviorkin, N. Loukachevitch // In Proceedings of COLING 2012: Technical Papers. – 2012. – P. 593-610. **2.** Bing L. Sentiment Analysis and Subjectivity / L. Bing // Handbook of Natural Language Processing, Second Edition. – 2010. – 38 p.