

Real time IT-Job Matching Score Predictions

S. Melih Zenciroglu and Xing Zeng

KTH Royal Institute of Technology

1 Problem Statement

The IT-related job is highly demanding and competitive in the market, which is a high-pressure for IT-related job hunters. Besides, it is also a problem for companies to find a suitable candidate. So our task is to help both companies and job hunters match mutually by streaming data from twitter in spark. Our goal is to predict the match score between job-hunters and companies in real time by comparing IT job ads and job searchers' profiles once a company posts a job ad. Then listing people and their matching rates which exceed a certain threshold (e.g. 80% match).

2 Data

The goal of the project is to predict on real time data. As such, we will be utilizing real-time post ads data from Twitter API [1] to perform our predictions. Moreover, we will crawl potential-employee resume public data from the web [2] and then store it in Cassandra or Hbase.

The data source is:

- Twitter API [1]
- Resume data source web [2]

3 Tools

The tools utilized for this project will include:

- Stream processing: Spark Streaming
- Stream source: Kafka
- Storage: Hbase or Cassandra
- Language: Scala, Java

4 Methodology

The approach to this project can be summarized by the tasks below:

1. Crawl job searchers' profiles data from the web and store it in a database.
2. Query twitter API to get post data.

3. Get IT-job-ads data by filtering the post data.
4. Stream IT-job-ads data and job seekers' profiles data into Spark Streaming.
5. Set up spark streaming jobs to enable real time match score prediction by a matching algorithm (E.g. NRMP matching algorithm).
6. Evaluate the performance of the model.

References

- [1] Twitter API <https://developer.twitter.com/en/docs/twitter-api/getting-started/guide>
[2] <https://www.livecareer.com/resume-search/>