MoDisSENSE: A Distributed Platform for Social Networking Services over Mobile Devices

Ioannis Mytilinis, Ioannis Giannakopoulos, Ioannis Konstantinou, Katerina Doka, Nectarios Koziris
{gmytil,ggian,ikons,katerina,nkoziris}@cslab.ece.ntua.gr

Motivation and Overview

- Terabytes of data varying with respect to volume, variety and velocity are produced in a daily basis.
- Social network APIs facilitate the access to social network data, causing a viral growth to third-party social applications.
- Combining real-time, large scale data from heterogeneous sources presents opportunities for smart services that leverage user experience.

To this end, we present MoDisSENSE: a distributed analytics platform for social network based, geo-location services over mobile devices. The MoDisSENSE data sources are:

- Structured, user-related information provided through the application GUI and enriched through social media accounts (e.g., friends, POIs visited, etc.)
- Spatio-temporal data provided by mobile devices (GPS traces)
- Textual information extracted by social networks (e.g., status updates, comments, etc.)

Spatio-temporal and textual analysis are performed in a distributed fashion, in order to extract knowledge and make smart suggestions to the user.

MoDisSENSE Application

MoDisSENSE supports:

- Personalized and socially charged search of POIs based on:
  - location
  - keywords
  - overall sentiment
  - social media friends’ opinion of the place
  - time window
  
  e.g., “list all meat restaurants offering lamb near Acropolis that my friends have visited and commented positively about in the last month”
  
  “show all places near my current location where most of my friends are gathered right now”
- Discovery of new POIs and trending events through analysis of GPS traces from mobile devices
- Inference of semantic trajectories and automatic creation of a blog with user’s activities within the day through combination of GPS traces with background information
- Capability of editing the blog or inserting additional information through the MoDisSENSE UI

Application clients developed for Android, iOS and the Web.

MoDisSENSE Architecture

Frontend: Web and mobile applications
Backend: Platform where data are being stored and processed.
- Storage backend: Hybrid datastore for flexibility and performance, comprising of
  - PostgreSQL Server for efficient indexing of structured data
  - HBase cluster for scalability and high write throughput of GPS and textual data
- Processing backend: Hadoop cluster for large-scale, distributed processing
  - Text Processing: Sentiment analysis using the Mahout Naïve Bayes implementation.
  - GPS Traces Processing: MapReduce implementation of the DBSCAN clustering algorithm.
- Query Answering: Answers queries from HBase by taking advantage of the devised schema and HBase coprocessors.

Performance

- Scalability achieved through
  - Data replication
  - HBase coprocessors
  - Sharding of data among NoSQL nodes
- Throughput of up to 40 queries/second.
- Socially charged query response time in the order of milliseconds for up to 5000 network friends.