



The members of SoBigData have a consolidated experience in Data Analysis, Human Mobility Analytics, Text and Social Media Mining, Social Network Analysis, Social Data Analysis.

Stakeholders

Industry

Data Analysts

Research

Human Mobility Analytics

Users' behaviours Analysis
Social and demographic indicators
Sport Analytics
Air Traffic management
Transportation planning
Traffic management
Mobility Mining for Smart Cities

Text and Social Media Mining

Topic annotator
High-quality ranking function
Financial prediction model
Text mining analytical tools
Customer- and domain-specific text mining
Visual analytics
Complex graph mining
Semantic analysis

Social Network Analysis

Real-time social media monitoring
Real-time social media analytics
Data aggregation and visualisation
Social media summarisation tools
Community detection algorithms
Financial networks
News and financial behavior

Social Data

Social and demographic indicators
Health and well-being data analysis
Analytical Platforms and Data Infrastructures for Social Mining
Ethical Data Mining
Visual Analytics for Social Mining

What SoBigData can offer to you:

The RI will take care of the legal, ethical, methodological, and infrastructural issues arising from working with social data, in order to enable data scientists to focus on research itself. It will provide access to the following key types of social data:

- Mobile and sensor data
- Social networks data
- Social media data, including Twitter, Facebook, and FourSquare content, organised into topic- and problem-specific social media virtual collections
- Mobility data, e.g. London Transport Oyster Card records and vehicular GPS trajectories
- Open social data and relevant Linked Open Data resources
- Other social data (such as one of the largest databases of Pinterest records)

Stories

Tripbuilder Tour Planner

Tripbuilder Tour Planner is a unsupervised system for recommending personalized sightseeing tour. By exploiting Flickr and Wikipedia we build trajectories of Pols and enrich them with additional metadata: Wikipedia categories, popularity, transition time, distance, etc. The sightseeing tour recommendation problem is modeled as an instance of the Generalized Maximum Coverage (GMC) problem, where a measure of personal interest for the user given her preferences and visiting time-budget is maximized. The set of derived trajectories from the GMC solution is scheduled on the tourist's agenda by using a particular instance of the Traveling Salesman Problem (TSP). ...

Large scale organization of financial and economic networks

To identify and characterize the large scale organization of financial and economic networks, focusing in particular on the inference of (i) the block structure organization (e.g. core-periphery, bipartite, modular, etc) and (ii) the hierarchical structure and ranking of nodes. ...

Urban Profiles

Imagine that you visit a city, but instead of going to all museums like a tourist, you want to experience life in the city as a local with a lot of free time. You would like to answer the following questions: What are the different neighborhoods in the city and what are its representative venues (e.g., restaurants, cafeterias, parks)? What is a good tour of the city, given venue preferences and budget constraints?...

The objectives of the analysis is to illuminate life in different cities from each of the perspectives described above. Concretely, for each city we discover a set of regions, a description in terms of types of activities they host, representative venues, a set of tours that capture possible tour preferences, a set of citizen profiles for daily activity in a city. ...

Elianto: Crowdsourcing Entity enrichment of structured documents

Understanding textual information requires, besides simplistic bag-of-words strategies, identifying relevant pieces of text within a document and even more matching that text to actual entities. However not all entities are similarly important, some may be more relevant than others. In order to be able to learn how to identify these pieces of text, various strategies can be proposed, but often human supervision is required.

Elianto is an open-source web framework for the production of human annotated rank-enriched datasets for Entity Linking and Salient Entities tasks. It is a tool that allows users to annotate pieces of text with entities (e.g., persons, places, events), and with their saliency (e.g. most relevant, relevant, not relevant). ...

Borders of Human Mobility

The availability of massive network and mobility data from diverse domains together with novel analytical paradigms have placed human relationships or their mobility patterns at the center of investigation. We propose a general method to determine the influence of social and mobility behavior over a specific geographical area in order to evaluate to what extent the current administrative borders represent the real basin of human movements. We build a network representation of human movement starting with vehicle GPS tracks and extract relevant clusters, which are then mapped back onto the territory, finding a good match with the existing administrative borders. ...

Please go to www.sobigdata.eu for full stories