Semantic Enrichment of Mobile Phone Data Records Using Linked Open Data

Zolzaya Dashdorj, Luciano Serafini
SKILL, Telecom Italia and DKM, Fondazione Bruno Kessler and
ICT International Doctoral School, University of Trento, Trento, Italy
{dashdorj@disi.unitn.it,serafini@fbk.eu}

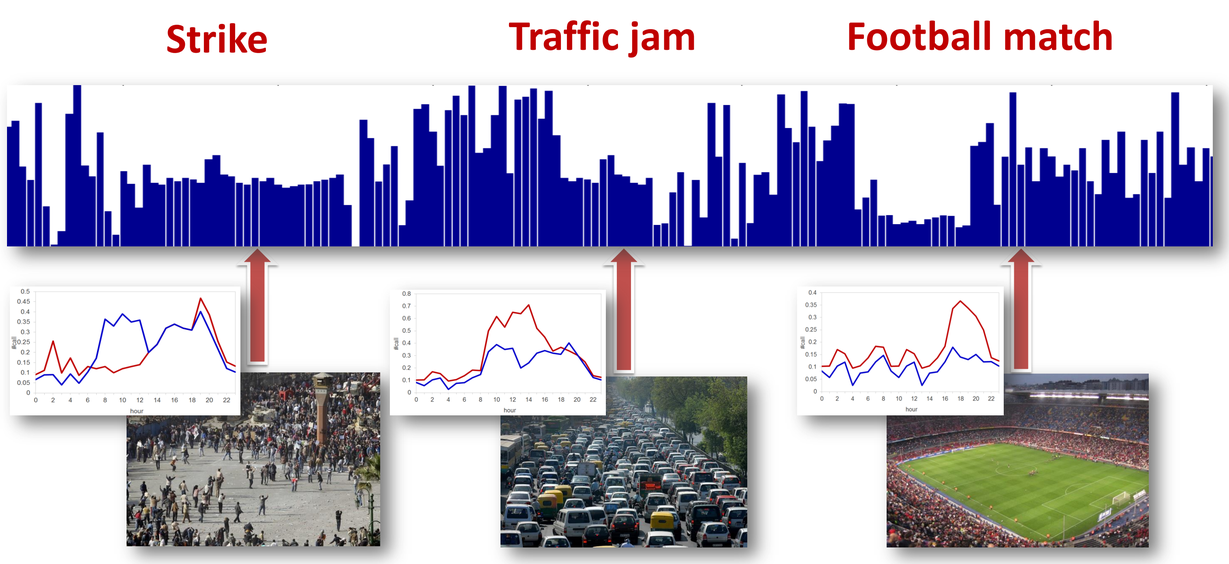
ISWC 2013 Sydney, Australia

The problem

Recognizing personal & social behaviors from mobile phone data (CDR) using geo/time referenced relevant knowledge [1]

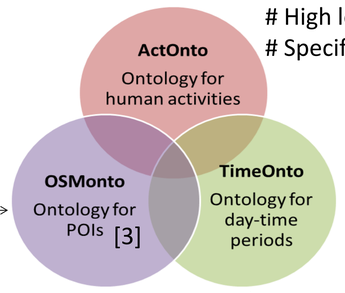
Challenges:
- ✓ Large amounts of CDR
- ✓ Privacy (largely anonymous data)
- ✓ Noisy and incomplete data
- ✓ Heterogeneous data (CDR and Linked Open Data (LOD))

A qualitative description of human **activities/events** interpreted in CDR



Methodology

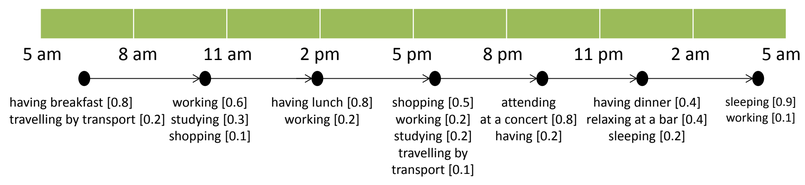
:Instance of Point of Interest (POI)



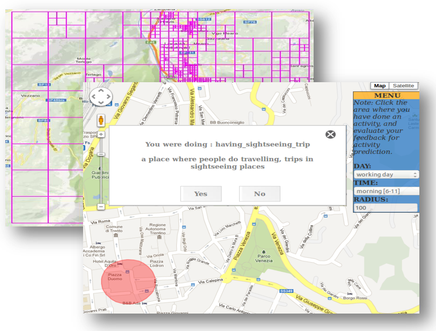
High level activities: 10
Specific activities: 220

:Instance of activity/event

Possible actions ranked



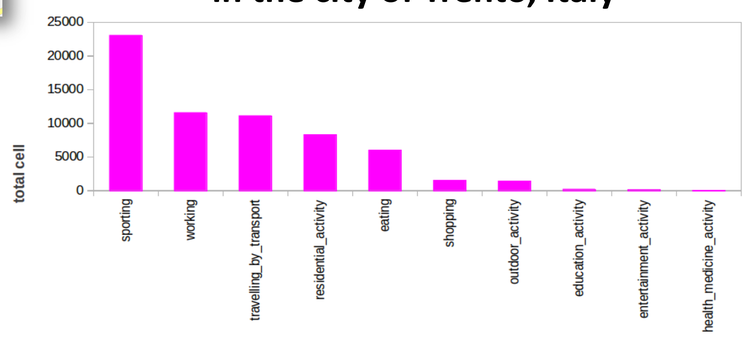
BHRModel application



Trento's area of about 157.9 km^2 with 333,809 POIs (cleaned to 159,314) is partitioned in 27,632 cells, with an average of 7.1 POI/cell.

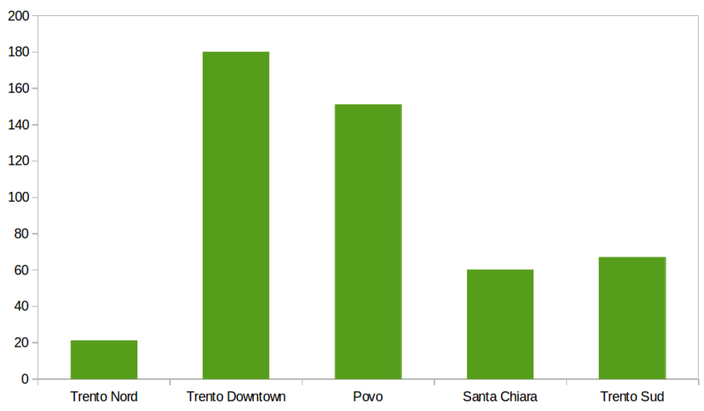
We created a web application that predicts possible human activities/events happened in a territory (selected the city of Trento, Italy for the experiment). A geographical area of the territory is divided into a grid [2].

Histogram of activities in the city of Trento, Italy



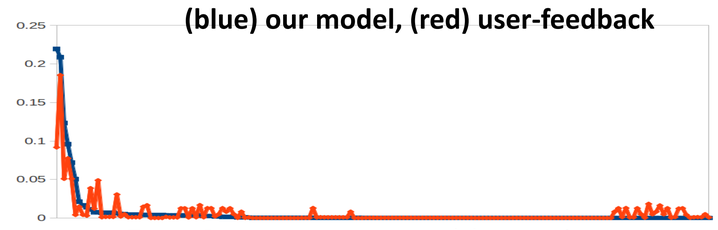
Experiment & Evaluation

Histogram of feedback in the city of Trento, Italy

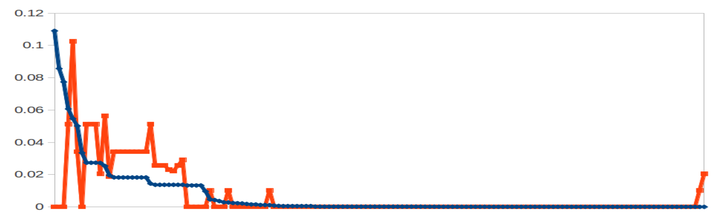


Territory: the city of Trento, Italy
Total feedback: 481
Participants: 32 people
Duration: a week
Clusters by the feedback location: 5 clusters
Overall accuracy (specific activities) : **70.89**%
Overall accuracy (high level activities): **80.23**%
Overall accuracy (among top-5) : **61.95**%

Divergence of probability of activities from (blue) our model, (red) user-feedback



Trento Downtown area on weekday mornings



Trento Povo area on weekday afternoons

Conclusions & Future work

✓Ability to predict possible human activities/events that happen in a given region at a given time in which CDR occur.
✓Ability to provide a qualitative description of human activities/events given in terms of semantically rich concepts that refer to an ontology of human actions and events and environmental descriptions.

Future work:
✓Enrich the model making use of the other geo/time-referenced knowledge available on the web, such as info on weather forecast, social events, news events and other statistical information about a region and so on.
✓Improve the classification task of behaviors and the identification and prediction from the CDR with the help of machine learning methods.

References

[1]. Z.Dashdorj and L.Serafini. Semantic interpretation of mobile phone records exploiting background knowledge. In Intl.Conf. ISWC 2013, Doctoral Consortium, 2013.
[2]. S.Phithakkitnukoon, T.Horanont, G.Di Lorenzo, R.Shibasaki, and C.Ratti. Activity-aware map: identifying human daily activity pattern using mobile phone data. I n the Proc. of the 1st Intl. Conf. Human Behavior Understanding, pages 14–25, 2010.
[3]. M. Codescu, G. Horsinka, O. Kutz, T. Mossakowski, and R. Rau. Osmonto - an ontology of openstreetmap tags. In State of the map Europe (SOTM-EU), 2011