BOTTARI: Location based Social Media Analysis with Semantic Web

Irene Celino[1], Daniele Dell'Aglio[1], Emanuele Della Valle[2,1], Marco Balduini[2], Yi Huang[3], Tony Lee[4], Seon-Ho Kim[4], and Volker Tresp[3]

[1] CEFRIEL – ICT Institute, Politecnico of Milano, Milano, Italy
[2] Dip. di Elettronica e dell'Informazione – Politecnico di Milano, Milano, Italy
[3] SIEMENS AG – Corporate Technology, Muenchen, Germany
[4] Saltlux – Communicating Knowledge, Seoul, Korea

Abstract. Location-based services are influencing our lives and the way we experience the surrounding environment; smartphone and tablet applications supply a huge amount of information: shops around us, traffic conditions, etc. A recent trend in this kind of services is to provide personalized information, such as friends' position or events users could be interested in.

In this paper we present BOTTARI, an Android application that exploits social media and context to provide point of interest (POI) recommendations to user in a specific geographic location. BOTTARI exploits a number of semantic techniques (sentiment analysis, inductive reasoning, stream reasoning) for social media analysis and suggests POIs on the basis of users' tastes and influencing people's opinion.

1 Description of BOTTARI and its innovation

The Insa-dong area of Seoul is one of the most popular visitor attractions in South Korea: it is the focal point of Korean traditional culture and crafts, with a multitude of shops, restaurant and points of interest (POIs).

BOTTARI is a location-based mobile application developed for Android tablets targeted to Korean users moving in Insa-dong. It provides personalized POI recommendations, to help users find their way when they are in a specific location. BOTTARI collects relevant information from social media such as Twitter and blog posts, elaborates it and provides contextualized suggestions.

The Korean word "bottari" refers to a bundle or container made from patterned cloth that is used to transport a one's belongings when travelling; the BOTTARI application lets the user "transport" the location-specific knowledge, derived from social media, when moving in the physical space.

BOTTARI has two main sources of information. One is a curated dataset about the Insa-dong area, and collects information about some hundred POIs, each one described by a few dozen attributes (location, description, place category, price range, reviews, contacts, etc.); this dataset content is quite static and is used as "background" information about the POIs. This data are expressed in RDF, described with regards to an OWL ontology and sums up to more than 20 thousand triples.

The other dataset is gathered from social media. Apart from blog posts, which are manually collected from Korean web sites, the main source consists in tweets collected from Korean users (i.e., all tweets are written in Korean language) since April 2010; those short messages are acquired by means of the Twitter APIs, are further elaborated to identify the tweets talking about POIs in Insa-dong and processed to assess the "sentiment" they express (positive judgement vs. negative rating). The results are expressed in RDF and described with regards to the OWL ontology illustrated in [1]; those triples are then stored in a SOR repository[5]; the triple count is currently 0.6 billion, but it is continuously and steadily increasing.

(a) (b)

(c) (d)

Fig. 1. Screenshots of BOTTARI: (a) augmented reality display of recommended POIs, (b) POI selection and (c) visualization of the selected POI details, (d) trends in user sentiment about the POI.

Figure 1 shows some screenshots of the application. BOTTARI provides to its users four different types of recommendations:

– *Interesting* recommendations suggest POIs indicated for foreign visitors in Korea; this feature calls for analysis and retrieval of POIs attributes;
– *Popular* recommendations suggest the POIs which show the highest level of reputation on social media; this feature call for a complete analysis of the social sentiment about POIs;
– *Emerging* recommendations suggest the most popular POIs in a delimited period of time (e.g. last 6 months); this feature calls for the identification of "hypes" and new trends in the social sentiment;

[5] SOR is a Saltlux product: `http://semanticwiki-en.saltlux.com/index.php/SOR`.

- *For me* recommendations suggest POIs of interest for the current user; this feature calls for personalized recommendations.

The innovation brought by BOTTARI consists in offering a location-based service through a simple and intuitive interface for a natural user experience; the application provides advanced semantic features, hiding the complexity of their computation from the user's sight. The details of the internal functioning of the BOTTARI application are given in the following section.

2 Semantic features in BOTTARI

The four types of recommendations provided by BOTTARI requires different levels of semantic technologies. In this section, we explain how we addressed the different challenges.

2.1 Semantic Information Retrieval to get *Interesting* POIs

The first kind of recommendations requires to suggest the user with a subset of the POIs that matches (1) the user current location and (2) the category of "attractions of interest for foreign visitors". To provide those recommendations, we used the Semantic Information Retrieval features of SOR, which provides a geographic extension of SPARQL that allows to query both the "semantic" description of POIs and their physical location.

2.2 Sentiment Analysis of social media to get *Popular* POIs

The second type of recommendations requires an analysis of the social media. The tweets and blog posts are processed by a sentiment analysis algorithm that detects if the message talks about a POI and, in case, if it expresses a positive or negative rating on the POI.

We adopted a twofold approach to compute the "sentiment"; we applied the two methods both separately and in cooperation, to improve the precision of results. On the one hand, we used a pure machine learning approach using SVMs with syllable kernel. On the other hand, we used a rule-based approach: the messages are analysed with respect to some rules about the structure and language. Those rules are both manually coded and generated by machine learning algorithms; this is a NLP technology of Saltlux, specialized on the Korean language.

Once the sentiment is elicited, this information is attached as metadata to the message description in the triple store. The *popular* recommendations are then generated by querying the knowledge base and suggesting the POIs with the highest number of positive ratings; the geographic features of SOR are also used to filter POIs and recommend only those around the user current location.

2.3 Stream Reasoning to get *Emerging* POIs

The opinion of users on POIs can change over time: the third kind of recommendations suggests the users with POIs that are “on fashion” in the latest period of time. To this end, we adopted stream reasoning [2] to identify trends and changes in the sentiment about the POIs.

Figure 2 illustrates how we elaborated the social media stream to derive synthetic and aggregated information that help the user in choosing the POI to visit. Because of the sentiment analysis elaboration, the stream of messages annotated with the user sentiment is not in real-time, but it is “re-streamed” from its storage. The queries enabled by the C-SPARQL Engine [3, 4] let find the emerging opinion of users about POIs: from left to right in the figure, our engine counts the positive opinions about a POI per each day, so to create top-10 lists; the positive messages per day can be further aggregated by week or by month and be visualized as plot lines or heatmaps.

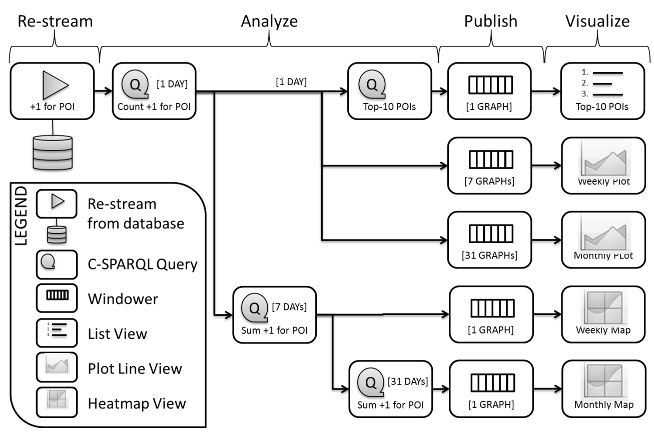


Fig. 2. Query chain to elaborate social media streams with C-SPARQL

2.4 Inductive Reasoning to get *For me* POIs

Finally, POI recommendations can be personalized: the user can be suggested with POIs that could be interesting for her. To this end, we adopted inductive reasoning on social media to compute BOTTARI’s *for me* recommendations.

We exploited the SUNS approach (Statistical Unit Node Set) described in [5, 6]. SUNS is a machine learning approach for exploiting the regularities in large data sets in relational and semantic domains. The approach can be used to detect interesting data patterns and predict unknown but potentially true statements.

In BOTTARI we applied SUNS to estimate the probability that a user will like a POI, based on the sentiment the same user expressed about other POIs and the opinion that other users expressed about that POI. In this sense, we provide a personalized collaborative filtering recommendation engine, to suggest users with the most interesting POIs with respect to their preferences.

3 Design and development of BOTTARI

The front-end of BOTTARI is a dedicated Android application which uses localization and augmented reality to provide its functionality to users in mobility in the Insa-dong area of Seoul; the visual appearance of the front-end is exemplified in Figure 1. The BOTTARI back-end is where the semantic technologies are adopted for social media analysis. To design and develop the back-end we exploited the potentialities of the LarKC platform [7, 8].

The LarKC platform, realized in the homonymous EU project[6], is aimed to reason on massive heterogeneous information such as social media data. The platform consists of a framework to build workflows, i.e. sequences of connected components (plug-ins) able to consume and process data. Each plug-in exploits techniques and heuristics from diverse areas such as databases, machine learning and the Semantic Web.

In BOTTARI, we designed a workflow that makes use of a number of plug-ins that implements the features explained in Section 2 (semantic information retrieval, stream reasoning, inductive reasoning). Every time a user of the mobile application requests a type of recommendation, the BOTTARI LarKC workflow is invoked and instantiates the plug-ins to compute the requested data. Each plug-in encapsulate a part of the data processing and interacts with the other plug-ins through well-defined interfaces; the plug-ins used in BOTTARI are described in several technical reports[7]. The adoption of LarKC made it easy the composition and interplay of the different semantic technologies used in the BOTTARI back-end.

While the mobile application requires an Android device and the physical presence of the user in Insa-dong, the back-end application runs on a server. The interest reader can have a look "behind the scenes" of BOTTARI at `http://larkc.cefriel.it/lbsma/bottari/`.

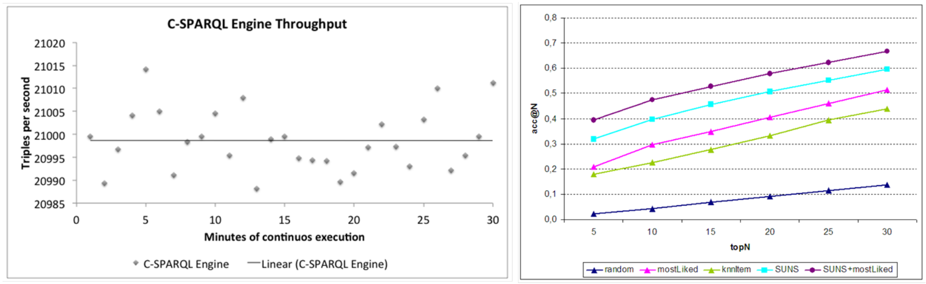
4 Evaluation and Outlook on BOTTARI

The semantic technologies behind BOTTARI have been evaluated in terms of the LarKC plug-ins used in the workflow. We started various evaluation campaigns on our implementation and their early results are available in some of LarKC project technical reports[8]. Hereafter, we present some of those early results.

In the stream reasoning processing, one of the main components is the C-SPARQL Engine [4]. In Figure 3(a), we report the results of a long lasting execution of a simple continuous query registered in the C-SPARQL Engine: the query matches all triples in the stream and regenerates them out. In our experimental settings (an Intel Core 2 Duo T7500 at 2.2GHz, 4GB of RAM DDR2 at 667 MHz, Hard Disk at 5400 rpm), the C-SPARQL Engine throughput was between 20,987 and 21,015 triples/second; the average throughput was 20,999 triples/second. This result proves that we are able to process the amount of data currently produced in social media streams.

---

[6] Large Knowledge Collider, cf. `http://www.larkc.eu/`.
[7] Cf. D3.9, D6.10 and D6.11 at `http://www.larkc.eu/deliverables/`.
[8] Cf. D3.9 and D6.11 at `http://www.larkc.eu/deliverables/`.



(a) (b)

Fig. 3. BOTTARI evaluation: (a) execution of a continuous query registered in the C-SPARQL Engine and (b) accuracy of recommendations

In Figure 3(b), we show the accuracy of recommendations: we compare two baseline algorithms – random guess (random) and item-based k-nearest neighbour (knnItem) – with C-SPARQL's *emerging* recommendations (mostLiked) and SUNS' *for me* recommendations; here, we do not consider the location dimension (i.e., in this evaluation we recommend POIs regardless the current position of the user). As expected, the random is the worst; C-SPARQL is slightly better than the similarity-based method: this might indicate the "bandwagon effect" that exists in many social communities; SUNS significantly outperformed all other methods (with a number of the latent variables greater than 100). The best ranking was produced by the combination of both SUNS and C-SPARQL: these results confirm again the effectiveness of combined approach of deductive and inductive reasoning [6].

The design of the BOTTARI service from the user point of view, as well as its prototypical implementation and its early evaluation, demonstrate that Semantic Web technologies can be successfully applied to concrete scenarios and can help in adding added-value functionalities. We foresee a potential market exploitation of this kind of location-based social media analysis applications and Saltlux is going to continue the development of BOTTARI to a commercial product for Korean customers.

Acknowledgments

This work was partially supported by the EU project LarKC (FP7-215535).

References

1. Celino, I., Dell'Aglio, D., Della Valle, E., Huang, Y., Lee, T., Kim, S.H., Tresp, V.: Towards BOTTARI: Using Stream Reasoning to Make Sense of Location-Based Micro-Posts. In Garcia-Castro, R., et al., eds.: ESWC 2011 Workshops, LNCS 7117, Springer, Heidelberg (2011) 80–87
2. Della Valle, E., Ceri, S., van Harmelen, F., Fensel, D.: It's a Streaming World! Reasoning upon Rapidly Changing Information. IEEE Intelligent Systems **24**(6) (2009) 83–89

3. Barbieri, D.F., Braga, D., Ceri, S., Della Valle, E., Grossniklaus, M.: C-SPARQL: a Continuous Query Language for RDF Data Streams. Int. J. Semantic Computing **4**(1) (2010) 3–25
4. Barbieri, D.F., Braga, D., Ceri, S., Della Valle, E., Grossniklaus, M.: Incremental Reasoning on Streams and Rich Background Knowledge. In: Proc. of ESWC2010. (2010)
5. Tresp, V., Huang, Y., Bundschus, M., Rettinger, A.: Materializing and querying learned knowledge. In: Proc. of IRMLeS 2009. (2009)
6. Barbieri, D.F., Braga, D., Ceri, S., Della Valle, E., Huang, Y., Tresp, V., Rettinger, A., Wermser, H.: Deductive and Inductive Stream Reasoning for Semantic Social Media Analytics. IEEE Intelligent Systems **25**(6) (2010) 32–41
7. Fensel, D., et al.: Towards LarKC: a Platform for Web-scale Reasoning. In: Proc. of ICSC 2008. (2008)
8. Cheptsov, A., et al.: Large Knowledge Collider. A Service-oriented Platform for Large-scale Semantic Reasoning. In: Proceedings of WIMS 2011. (2011)

Appendix – Addressing Evaluation Requirements

In the following, Table 1 and Table 2 summarize how BOTTARI addresses the minimal and the additional requirements, as they are listed in the Semantic Web Challenge Criteria. We provide a qualitative rating (L = low, M = medium, H = high) and a textual explanation.

Criteria	Rating	Motivation
End-user application	H	BOTTARI is intended as an application for users in mobility and its design is driven by users' needs
Information sources		
- diverse ownership or control	H	the data sources come from location-specific information (POIs description) and public social media streams
- heterogeneous	M/H	the data are indeed heterogeneous (POI descriptions, social media, users' profile)
- real-world data	H	all data used in BOTTARI are real and are considered at real scale
Meaning of data		
- Semantic Web technologies	H	both POI descriptions and social media are represented and stored in RDF and described by OWL ontologies; SPARQL and its extensions are heavily used in the data processing
- data manipulation/processing	H	social media are processed to perform sentiment analysis and to provide recommendations for users
- alternative technologies	M/H	our evaluation shows that the results are indeed better than those of other approaches

Table 1. Minimal requirements

Criteria	Rating	Motivation
attractive and functional Web interface	H	the BOTTARI interface is designed to be intuitive and easy to use, hiding the complexity of the technology behind it
scalable application	H	our evaluation demonstrates that our approach is able to handle the data scale of social media at a very good rate
rigorous evaluation	H	the early evaluation results are given in this paper; further analysis are being performed
novelty	M/H	we are applying sentiment analysis, stream reasoning and inductive reasoning together to provide added value to the user
beyond information retrieval	H	we provide location-based information retrieval, but we also apply sentiment analysis, stream reasoning and inductive reasoning
commercial potential	H	BOTTARI is currently a prototype, but Saltlux intends to continue its development to a commercial product for Korean customers
ratings or rankings	H	we use both social media and location information to "rate" the POIs to be recommended to users
use of multimedia	L/M	BOTTARI is an augmented reality application that, whenever possible, uses intuitive and graphical visualization means to ease the user experience
dynamic data	H	we combine static information about POIs with dynamic information coming from social media streams
results accuracy	M/H	our evaluation shows the accuracy of our approach
multiple languages and accessibility	L/M	the user interface is developed in Korean and in English

Table 2. Additional Desirable Features