A dynamic self-adapting recommendation engine for the sensor web

Rohana Rezel[1], Steve Liang[2]

1 Geomatics, University of Calgary, rdrezel@ucalgary.ca
2 Geomatics, University of Calgary, steve.liang@ucalgary.ca

Abstract

Web-enabled sensor networks are growing rapidly, mirroring the exponential growth of the World Wide Web (WWW). The complexity of these networks necessitates mechanisms to enable users to gain quick access to relevant and reliable information. We are proposing a self-correcting recommendation engine that would suggest new sensors to the web using a set of metrics, including user behaviour and collaborative filtering.

Background and Relevance

The next generation web-enabled sensor networks (sensor webs)(Botts et al. 2008)are anticipated to mimic the growth of the WWW. They are expected to grow exponentially in terms of complexity, with millions of heterogeneous nodes and billions of users connected at any given time. Moreover, the traditional model where sensors could be considered experts is being broken down as practically any individual with a mobile phone being able to function as a sensor (Kansal et al. 2007).

In common with the WWW, the proliferation of sources of varying degrees of quality would lead to sensor networks that are data rich and information poor. Some users of such networks would be faced by information overload, while others would have to deal with information poverty as they would not be aware what sensors exist where.

Therefore, users would start requiring recommendation services to help them identify which sensors could be useful and/or interesting to them. There are recommendation engines already available for the traditional web, but there has been no research done on extending these to the sensor web. Catalogue Services are already available for the sensor web (Nebert & Whiteside 2004) and there are nascent research efforts to develop search engines for it (Reddy et al. 2007). It is envisaged that advancements made with recommendation engines may be used to enhance the capabilities of the search engines.

There are a number of recommendation engines available for different applications (e.g. Amazon for books and other merchandise, StumbleUpon[1] and reddit[2] for websites), and they incorporate elements of both user behaviour analysis and collaborative filtering. However, they only deal with content matching, but a sensor recommendation engine would also have to deal with the spatial aspects.

[1]http://stumbleupon.com
[2] http://reddit.com

Methods and Data

Methods

The recommendation engine will be built by harnessing the collective intelligence of the users by enabling a folksonomy (Mathes 2004) and a rating system. A self-correcting weighting algorithm, based on neural networks, is proposed to train the engine.

The proposed content recommendation engine would have a graphical front end that allows users to tags nodes (analogous to del.ici.ous[3] in the WWW world) and rate nodes (analogous to the Web2.0 sites Digg[4] and reddit). There will be search/browse function that would enable the users to access content as the system is trained.

The proposed recommendation algorithm uses five metrics which have been successfully used in WWW systems, some of which are mentioned above:

1. User behaviour analysis
2. Collaborative filtering
4. Submitter Reputation
3. Sensor Reputation
4. Popularity
5. Freshness

User behaviour analysis

The user behaviour analysis will consist of both spatial closeness matching and tag similarity matching. Spatial interpolation methods, such as Kriging and Inverse Distance Weighting, will be used to calculate the score based on distance. Methods proposed by Cattuto et al. (2008) will be used to assess tag similarity.

Collaborative filtering

Collaborative filtering will be achieved using the weighted Slope One predictor proposed by (Lemire & Maclachlan 2005). Slope One algorithm is considered to be one of the simplest algorithms to predict a user's opinion based on the said user's opinion and other users' opinions. The predictor is of the form $f(x) = x + b$ which precomputes the average difference ratings given to an item by two users. For example, if user A likes items X and Y, and users B likes items X,Y and Z, then there's a high probability that user A would also like item Z.

Submitter Reputation

One successful strategy adopted by community driven websites is the concept of user karma, i.e., a score assigned to users on the basis of the quality of their participation. It was popularized by the pioneering collaborative news aggregation portal Slashdot[5]. It is envisaged that the recommendation engine will use the reputation of the submitter, based on the submitter's karma, to initially assign a score to a sensor. However, given that there's a significant number of sensors provided by experts (e.g. WMO, NASA, Environment Canada, USGS), these experts can be assigned a high karma score by default.

---

[3] http://delicious.com
[4] http://digg.com
[5] http://slashdot.org

Sensor Reputation, Popularity and Freshness

Reputation will be based on the weighted sum of user ratings. Popularity will be a function of the number of users who tag/rate a node, and Freshness will depend on the time of the last update of a sensor.

Self-adopting weighting algorithm

The system will be a learning system, adopting weights given to different scores based on a neural network.

System Design Considerations

The system will be of a very high computational complexity. For example, for a system of m nodes and n users, the size of the Slope One matrix will be mn^2. Therefore, a series of optimizations is proposed to reduce computational complexity of the engine.

Moreover, any wisdom-of-the-crowds system is open to gaming, *i.e.*, a segment of users (often a tiny fraction), who try to skew the results by manipulating algorithm to promote certain nodes. For example, Google bombing (Judit Bar-Ilan 2007) is a classical example of people gaming the Google Pagerank algorithm. We will design anti-gaming mechanisms in order to prevent a small amount of users manipulate the algorithm and skew the recommendation results.

Data

The data will consist of the sensor metadata, and user generated data. The sensor metadata would include sensor location, the data the sensor is providing (temperature, ground-level O_3 concentrations, road traffic images) and information about the sensor operator. The user generated data would include tagging (using both controlled vocabularies and free-text) and ratings.

Results

The results will be evaluated by measuring the average scores users assign to recommended nodes after they have used the system for a reasonable period of time, and a critical mass of users have signed up to the system.

The expected results are that as the user numbers increase and user-hours spent using the system (thereby passively training it) increase, the average score assigned to recommended nodes will go up.

Conclusions

We have proposed a recommendation engine that harnesses the collective intelligence from sensor web users and helps them to discover sensors that maybe of utility or interest to them. The recommendation engine uses a neural network to aggregate a set of metrics. The expected results are that as more user-hours are spent on the system, the performance of the engine will increase.

Future work will add temporal aspects to the recommendation engine.

References

Botts, M. et al., 2008. *OGC« sensor web enablement: Overview and high level architecture*, Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-54249109045&partnerID=40.

Judit Bar-Ilan, 2007. Google Bombing from a Time Perspective. *Journal of Computer-Mediated Communication*, 12(3), 910-938.

Kansal, A., Goraczko, M. & Zhao, F., 2007. Building a sensor network of mobile phones. In *IPSN '07: Proceedings of the 6th international conference on Information processing in sensor networks*. Cambridge, Massachusetts, USA: ACM, pp. 548, 547. Available at: http://dx.doi.org/10.1145/1236360.1236433.

Lemire, D. & Maclachlan, A., 2005. Slope one predictors for online rating-based collaborative filtering. Available at: http://citeseer.ist.psu.edu/lemire05slope.html.

Mathes, A., 2004. Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Available at: http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html.

Nebert, D. & Whiteside, A., 2004. OpenGIS Catalog Service Specification 2.0 GC Document 04-021r2.

Reddy, S. et al., 2007. Sensor-Internet Share and Search—Enabling Collaboration of Citizen Scientists. In *Proceedings of the ACM Workshop on Data Sharing and Interoperability on the World-wide Sensor Web*. Cambridge, Mass., USA: ACM, pp. 11–16. Available at: http://www.isi.edu/ johnh/PAPERS/Reddy07a.html.