Trend-Aware User Modelling with Location-Aware Trends

Marcel KANTA*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
kanta07@student.fiit.stuba.sk

The Web brings people information, There is too much information and people face these days a phenomena called information overload. To cope this problem we can recommend users relevant information based on their personality and context. To do this it is important to create user models. Twitter is highly significant source of social information and interaction. Therefore in our work we use Twitter as a source for user modelling and we focus on creating a user model to help people with information overload.

When we proposed a conceptual modelling technique, we found similar successful works [1, 2, 3] that inspired our model. Conclusions were that personalised recommendation is more important than trend-aware recommendation, but integrating trend-aware and personal recommendation can improve recommender results. We want to incorporate trend-awareness and personalisation. On top of that we use location-awareness to improve the results, thus the user model will be more precise. The idea is based on the assumption that employing location of trends will improve the quality of user model. Applications that will incorporate our user model will have more precise results compared with traditional non-location-aware model.

We formally define our user model as follows:

$$P(u) = \{(c, w(u,c,l)), l \mid c \in C, u \in U, l \in L\}, \tag{1}$$

where c stands for concept, w for weighting function, u for user and l for location. We introduce location l, which means that every concept and user belongs to a region and its parent regions that are represented by quadtree structure (see Fig. 1). We use TF-IDF and t-TF-IDF [2] for region and time as a weighting function w mentioned in equation 1. t-TF-IDF is a trend-aware modification of standard TF-IDF. It uses temporal stability of concepts in form of computing a standard deviation of appearance of concepts in time quantums. There can be also a trend model that is location-aware.

Our hypothesis is that location-awareness will improve the quality of user model. The principle of location awareness is that user model is modelled in regions, so weighting is done with region in mind; weighting only per region. This is the key that enables location-awareness. Our location-aware model uses at most $M(\log n)$ times more data than traditional model where n is maximal number of regions.

---

* Supervisor: Marián Šimko, Institute of Informatics and Software Engineering

Spring 2012 PeWe Workshop, April 1, 2012, pp. 37–38.

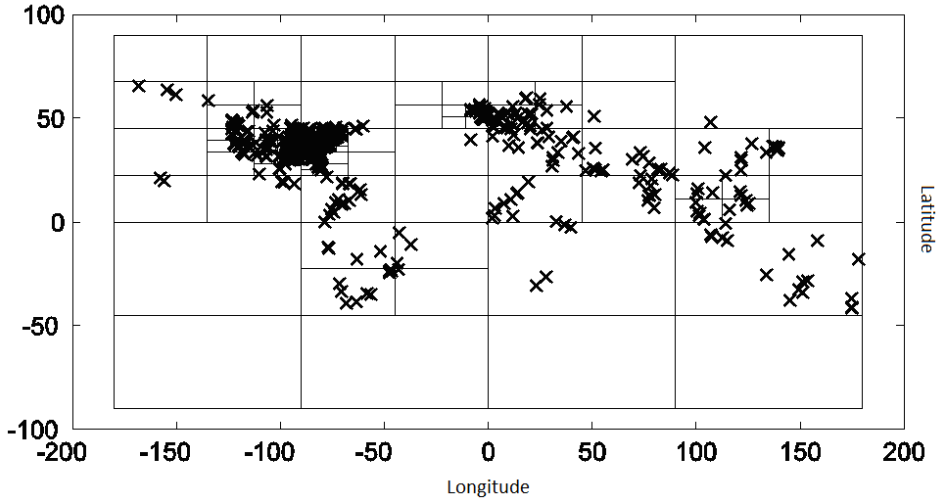

Figure 1. Our quadtree regions generated from user dataset[1]

Our user model will be quantitatively validated on a Twitter dataset. We chose Map-Reduce as a platform for evaluation. We use Hadoop implementation and Hive for SQL-like syntax. We plan a general synthetic evaluation approach used in machine learning, where we create user model on train set and then test (evaluate) it on other, test examples. In testing phase we recommend top *n* items that are matched by cosine similarity and then test, if user actually posted a link that matches one of the recommended links. For validating, we will use Precision P@n. We evaluate results on the whole dataset, on every user, so we get an average from all precisions. The point is to have the best possible precision. We will show the difference between location aware and not-aware user model. This is how we will prove the quality of location-aware user model.

This approach does not depend only on domain of microblogging, it may be used everywhere where we want to model entities with semantic information, where location matters. Also, regions of our approach does not have to be based on geographical coordinates, it may be a tree of clustered data, e.g. students in classes and recommendations in different classes and schools.

Acknowledgement: This work was partially supported by the Slovak Research and Development Agency under the contract No. APVV-0208-10.

References

[1] Abel, F., Gao, Q., Houben, G.J., and Tao, K.: Analyzing User Modeling on Twitter for Personalized News Recommendations. In *Proc. of the Conf. on User Modeling, Adaption and Personalization*, Springer, (2011), pp. 1–12.

[2] Gao, Q., Abel, F., and Houben, G.: Interweaving Trend and User Modeling for Personalized News Recommendation. In *Proc. of the Int. Conf. on Web Intelligence and Intelligent Agent Technology WIIAT 2011*, IEEE, (2011), pp. 100–103.

[3] Morales, G.D.F., Gionis, A. and Lucchese ,C.: From Chatter to Headlines: Harnessing the Real-Time Web for Personalized News Recommendation. In *Proc. of the Fifth Int. Conf. on Web Search and Web Data Mining* (WSDM), ACM (2012), pp. 153–162.

[1] *Twitter dataset:* http://wis.ewi.tudelft.nl/umap2011/