Concept of Behavioral-Targeted Recommendation System for Location-Based Services

Alexey Koren, Ina Sipunova
State University of Aerospace Instrumentation
Saint-Petersburg, Russia
{alexey.koren, ina.sipunova}@gmail.com

Abstract

Location-based services are one of the hottest topics from commercial point of view amongst mobile services. It is clear that using information about user's location it is possible to provide him with much more relevant information. But as soon as systems providing context-aware information on demand using geographical position are widely spread, there are practically no attempts to use geographical information to analyze and predict user's needs. In this paper we describe concept-level model of behavioral-targeted recommendation system for tourist-oriented location-based services basing on matching routes with thematically tagged areas.

INDEX TERMS: BEHAVIORAL TARGETING, LBS, RECOMMENDATION SYSTEM.

I. INTRODUCTION

While location-based services are amongst rapidly developing areas there are still few problems, and one from them is monetization through advertisement model. Nowadays users don't like to pay much for informational services, preferring free services with advertisement model. Here location-based services (LBS) have huge potential, because advertisement of places situated near user is much more targeted (please note - only "geo-targeted") than advertisement of places without taking into account user's location. At the same time there are no targeting based on individual preferences and interests, at least as soon service is not provided by some huge search engine or other entity which can aggregate data about user's web surfing and associate particular mobile LBS user with person who did search queries via desktop computer's browser. But in all mobile applications problem of targeting should be considered very carefully, because of small screen and psychological issues. Users treat mobile devices as very personal and really don't like any annoying things on the screen. This problem could be avoided if service doesn't show too much advertisement and this advertisement has extremely high level of targeting.

II. MAIN PART

A. Related work

While targeting in location-based systems is a pretty new problem, targeted computational advertisement (both behavioral targeted and context-based) in web services are well-known area and there are some good results showed. Jun Yan et al made research illustrating how much can behavioral targeting help online advertising [1]. Authors analyzed 7-day work interval of commercial search engine and came to following conclusions:

1) Users who clicked the same advertisement will have similar behaviors on the Web (over 90 times more similar);

2) CTR of online advertisement can be improved up to 670% by segmenting users for behavioral targeted advertising in a sponsored search;

3) Using short term user behaviors to represent users is more effective than using long term user behaviors for behavioral targeting

It is easy to conclude that user segmentation and behavioral targeting can really help increase advertisement efficiency.

Authors note that CTR can be increased further by using novel user segmentation and behavioral targeting algorithms. Nowadays a large variety of such algorithms based on different techniques: Poisson regression model [2], genetic programming [3], support-vector machines [4], Bayesian networks [5], etc.

As soon as there are good achievements in computational advertisement in web services we focused not only on creating efficient novel algorithm for targeting in location-based services but also on adapting location-based data for being placed in targeting algorithms described above or other existing algorithms.

B. General description of the approach

It is easier to measure relation of user by the manner in which he/she interacts with some objects in the system. The interaction can be positive or negative, for example mechanisms of implicit feedback claim to indicate explicitly object's quality or its actuality. It is very simply to handle such data, but the disadvantages of the approach consist in the fact that users are quite lazy and they give feedback very rarely. In addition, those who do it regularly are usually the most active part of the users, which gives non-representative selection because such active users are usually no more 10-20 percent of all users of some resource.

Therefore systems of explicit feedback are used extensively, which don't ask the user explicitly about anything, try to construct a model of his/her preferences, based on his/her observation of actions performed on the resource, such as a page view, further search queries on similar subject after viewing the resource (which usually means that the user wasn't satisfied with the information received) and etc.

In the web it is enough simply to emphasize objects, user's actions, which are necessary to observe: it is usually a web-pages, which encapsulate news, article, image or video. In the case of Geographic Information Systems such division is becoming a little more complicated. The division of territory on the so-called points of interest (POI) seems intuitive variant as soon as most Geographic Information Systems operate this conception. POI is a certain information object associated with geographical position. It can be not only sight existing physically, but also history associated with this place. Such points are generally localized narrowly and granulated highly. The alternative variant consists in the use of locations, in other words it consists in the use of big districts (for example historical center, business block or block with stormy night-life), in which user's stay gives information about his/her interests.

At the moment the works haven't published which evaluate the most effective size, form, pattern, algorithm of distribution of locations/POI. Therefore, in our model, we will operate zone conception as a geographical area which can be small and represent POI or larger and represent locations. Further analysis of experimental data will help us to choose the most effective method of distribution of such zones, which will possibly use joint application not only POI, but also locations.

We propose to use the mechanism of events, which are generated under certain user's actions, for example under zone entrance and exit. The event contains the vector, which describes positiveness or negativity of user's attitude towards a certain set of concrete objects or thematic categories. Then this vector goes to the input of training algorithm, which retrofits model describing the predilections of man.

The application of vectors of events will allow us to use not only mechanisms of explicit feedback, which work for all users, because it doesn't require them additional actions, but also mechanisms of implicit feedback, which will give very exact information about preferences of actual user as a minimum.

In the following sections are descriptions of the model of data, vectors of events. At the moment the choice of effective training algorithm is practically impossible in default of experimental data, therefore preliminary variants are only given with analysis of potential pluses and minuses.

C. Data model and categorization of objects

It is important to select the right model of data presentation and keeping for its effective analysis. The challenge consists in keeping of thematic marks and categories, to which the user's attitude will be the object of research.

We consider three basic types of organization of this kind of information in the Web: folksonomy, taxonomy and ontology.

Recently it is becoming more popular to represent data in the semantic format. As a general rule, in this case, ontologies are used to describe some subject field. There are a large number of software modules and open web-standards to work with it (e.g. RDF, OWL).

The main benefit of ontology is it's semantic. It is able not only to work with data effectively (from the viewpoint of volume of accessible information), but also to add on graph, using mechanisms of reasoning and presuming that all types of objects and relationships have been described formally. Ontologies are usually assigned in hierarchical format, identifying some entities as subclasses of other; it gives all benefits of taxonomy.

There are several ontologies for tourism domain are already developed, survey can be found in [6]. Usage of existing ontology or derived from existing will simplify development as well as give us addition compatibility with other platforms.

Concerning community-generated classification, the direct construction or editing of ontologies by user communities haven't yet demonstrated it's validity, however it is possible to use the «Semantic tagging» concept, to involve users in the process of knowledge creation, clarification and formalization. Further we will be able to extract chosen relations and classes to add it into core ontology.

Thus the most attractive model of classification is ontology in combination with semantic tags. In this case ontology can describe field of knowledge quite clearly, semantic tags express individual user's presentations about the object. These presentations differ from ontology either because of another user's opinion or high specifics at the level of concrete proper names or limited specialized dictionary.

To construct effective model it is necessary to take into account community content somehow transferring information from semantic folksonomy to ontology while controlling the level of importance of each folksonomy concept not only for subject that uses it for classification, but also for basic mass of users.

We will analyze two approaches. The first is based on the appending popular concepts from semantic folksonomy to the general ontology (Ontology Learning). The second is based on the customization of the ontology for each concrete user.

1) There are the common ontology for all, the pool, in which all tags of users "enter", and some sluice that analyses the popularity and importance of tags, trying to carry on the normalization at the level of word forms and to combine synonyms.

2) Each user has an individual system. The user forms this system personally: he/she adds classes to the ontology, forms groups of words, which are matched to the certain theme; also he/she defines the hierarchy.

The hierarchy of ontology has tree structure and three four levels of abstraction. The ontology begins with root node L0, which is the top of tree. The weight of whole tree is accumulated at this node. The level L1 is the level of directions of human activity (sport, culture, art, etc). The level L2 is the level of concrete thematic directions (football, visual art, Gothicism, etc). The levels L3 and below (L3+) are the levels of concrete words, which are applied in the context L0-L1-L2, for example: exotic styles or branches, proper names. The levels L3+ is the levels, where practically all leaves of the tree are situated. Thus each of the nodes, as well as each of the leaves is the tag, which can mark some information object.

Let's consider how user's attitude towards thematic categories will be determined.

There are many events which are generated automatically on the basis of user's actions, expressing the level of his/her interest to certain thematic fields.

One of the main events is a entering of the person into tagged area. In most cases attending locations means certain level of person's interest to a thematic topics corresponding to this location, so person will be interested in attending similar or related places. For sure some areas could be visited frequently not because of interest but just because they are on the way to some place which is actually interesting for person. Exclusion algorithms for such places are out of scope of this paper. Other types of events are: information requests, explicit feedback etc.

Event can be shown in form of vector:

$$E = ((t_1, w_1), \dots, (t_n, w_n), W_E), \text{ where}$$

$t_1, \dots, t_n$ - tags of event (area keywords, words in search query etc.)

$w_1, \dots, w_n$ - tags' weights that represent its value comparing to other tags. If it isn't set then default weight is used.

W_E - summarized weight of the event. If it is set then tags' weights means not absolute value for ranking algorithm but only used to determine absolute weight as percentage of all tags' weights sum.

It is easy to see that all weights are optional so it is possible to reduce pattern of input data to simple keyword list. This should help with getting data from external sources

Depending on any input vector E, state of behavioral model of person is updating. Simplest behavioral model is an ontology described above added weights representing person's attitude to every node (thematic category).

Weights' updating is a cascade process; it means that during any node's weight update we also update weighs of all its ancestors. It gives as an opportunity to track:

1) value of interest to very abstract or common groups of terms or thematic categories like art, architecture etc., which are not suitable to be used as tags;

2) total value of information accumulated as sum of all tree nodes' weights.

Given model gives a method of track person's interests and recommend informational objects depending on his or her interests, but it takes into account only interest to particular topic and its ancestors, ignoring possible semantic relations to topics situated in distant parts of tree.

For recovering such relations we are planning to use association rules analysis methods like market basket analysis. Thanks to hierarchical structure of ontology we are also able to use generalized association rules [7], tracking also objects' places in hierarchy in addition to direct associations. We are planning to test A priori [8], FPG [9] and quantitative association rules [10].

III. Conclusion and Future Work

In this paper we describe concept-level model of behavioral-targeted recommendation system for tourist-oriented location-based services basing on matching routes with thematically tagged areas. Most of proposed techniques were derived from web search and recommendation system domain, where there are certain successes and achievements in this field, but its application for location-based services are relatively new and promising direction.

First future step is to move from concept to prototype of recommendation system. Here, for choosing best methods and algorithms (ontology learning, classification, association mining etc) real world tourists' routes data is needed very toughly. That is a very common problem for evaluating new kind of services, but we plan to get experimental data by the end of summer 2010.

Second step is to build modeling framework for experiments and evaluation of new versions of the method, but, again, estimation of model relevance is only possible after getting real-world data.

One more important direction is to move from very concrete area of tourists and geographical location to any areas or location, trying to apply model to virtual environment like distributed smart spaces.

References

[1] Jun Yan, Ning Liu, Gang Wang, Wen Zhang, Yun Jiang, Zheng Chen. How much can Behavioral Targeting Help Online Advertising? WWW 2009, April 20–24, 2009, Madrid, Spain. ACM.

[2] Ye Chen, Dmitry Pavlov, John F. Canny. Large-Scale Behavioral Targeting. KDD'09, June 28–July 1, 2009, Paris, France. ACM.

[3] Anisio Lacerda, Marco Cristo, Marcos Andre Goncalves, Weiguo Fan, Nivio Ziviani, Berthier Ribeiro-Neto. Learning to Advertise. SIGIR'06, August 6–11, 2006, Seattle, Washington, USA. ACM.

[4] Andrei Broder, Marcus Fontoura, Vanja Josifovski, Lance Riedel. A Semantic Approach to Contextual Advertising. SIGIR'07, July 23–27, 2007, Amsterdam, The Netherlands. ACM.

[5] Berthier Ribeiro-Neto, Marco Cristo, Paulo B. Golgher, Edleno Silva de Moura. Impedance Coupling in Contenttargeted Advertising. SIGIR'05, August 15–19, 2005, Salvador, Brazil. ACM.

[6] Tourism ontology and semantic management system state-of-arts analysis. Kathrin Prantner, Ying Ding, Michael Luger, Zhixian Yan. IADIS International Conference WWW/Internet 2007

[7] R. Srikant, R. Agrawal. "Mining Generalized Association Rules", In Proc. of the 21th International Conference on VLDB, Zurich, Switzerland, 1995

[8] R. Agrawal, R. Srikant. "Fast Discovery of Association Rules", In Proc. of the 20th International Conference on VLDB, Santiago, Chile, September 1994.

[9] Buffet, O., and Aberdeen, D. 2006. The factored policy gradient planner (ipc'06 version). In Proceedings of the Fifth International Planning Competition (IPC-5).

[10] R. Srikant, R. Agrawal. "Mining quantitative association rules in large relational tables". In Proceedings of the ACM SIGMOD Conference on Management of Data, Montreal, Canada, June 1996