Recommendation of Points of Interest from User Generated Data Collection

Karol Waga, Andrei Tabarcea, Pasi Fränti

Speech & Image Processing Unit, School of Computing, University of Eastern Finland, Joensuu
{kwaga, tabarcea, franti}@ cs.uef.fi

Abstract—Systems that aim to predict user preferences and give recommendations are now commonly used in many systems such as online shops, social websites, and tourist guides. In this paper, we present a context aware personalized recommendation system on web and mobile, which recommends relevant location-based data from user collection and consisting of GPS routes and photos. We recommend three types of items: services, photos and GPS routes that are points of interests in user's surrounding. We score all items from database based on four aspects of relevance: location, content, time and network. In order to personalize the results we built user profile based on user's activity in the system. We study performance of the system within MOPSI.

Keywords—Recommendation, relevance, user collection, GPS trajectories, routes, context aware computing, location based systems

I. INTRODUCTION

A vast availability of location-acquisition devices and technologies (smartphones, GPS, GSM networks, mobile internet) allows people to record their activity by taking photos and tracking [16], especially if the service offers the possibility of sharing them with friends or with people who share the same interests. Gathering such data about activities of users allows discovering patterns of users' movements as well as information about points of interest in the area.

Recommendation systems produce personalized search results relying on a variety of contextual information. We have designed a recommendation system based on the four aspects of relevance: content, time, location and social network discussed in [3]. The system recommends items from a user-generated location-based dataset which consists of geotagged photos, trusted services and routes. Current version of the recommendation system is based on an earlier prototype as described in [12]. The goal of the recommendation is to suggest to user in certain location at given time where to go next, considering three types of items: services, photos and routes. Our solution is implemented within MOPSI system. The system includes various location-based services and applications such as search engines, data collection, user tracking and route recording. It has applications integrated both on web and in mobile phones. MOPSI contains services, photos and routes databases. Two latter ones are collected by service users utilizing a mobile application. The collections are shown in Fig. 1. Our user profile database used for giving personalized recommendation contains data about activities of MOPSI users within the service.
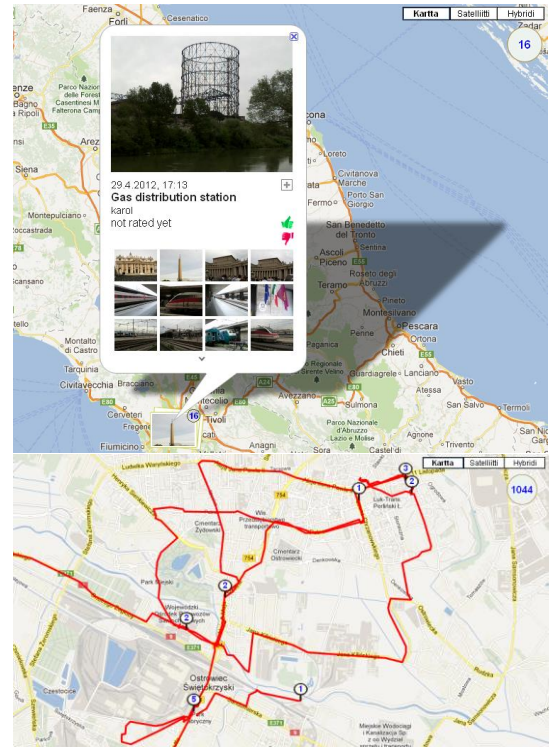

Figure 1: Examples of photos and routes from MOPSI user collection

II. RELATED WORK

Using location based-based data and user's location as input to a context-aware recommendation algorithm has been studied in several research projects.

For example, the system described in [8] recommends online content and offline events based on current user location, adding location to multi-dimensional personalization. Events are also recommended in [6], which employs a multi-stage collaborative filtering process. Our system also recommends location-based data and considers user profiles and personalization, but the data is user-generated (photos and routes).

CityVoyager, a system described in [9], proposes recommendation of shops based on user location history. Visited locations (shops) are used as input to an item-based collaborative filtering algorithm. Similarly, we use location history as a relevance criterion.

Magitti, described in [1], predicts user activity from sensing context and from patterns of user behavior. The recommendation of content is generated automatically using a combination of various models, including collaborative filtering, distance, stated or learned preferences. We also recommend the content

automatically, with minimum input from users.

The system in [16] recommends friends and places using individual location history. In the recommendation process, three factors are considered: user's particular location history, similarity between users in terms of location history, estimation of user's individual interests in an unknown region by comparing the location history and interest of other users. The collaborative filtering and recommendation algorithm is described in [15].

The service proposed in [13] uses web 2.0 technologies along with location-based services. A multi-dimensional collaborative filtering algorithm is designed in order to achieve dynamic personalized information which is delivered to mobile devices.

Location-Aware Recommender System (LARS) [5] uses location-aware ratings for recommendation. The ratings and items are considered to be spatial or non-spatial, also using the travel penalty approach which favors recommendation candidates close to user's location. The datasets used are social data from Foursquare and a part of MovieLens movie recommendation data. We also have a ranking-based system in which travel distance from user's position is an important factor.

"I'm feeling LoCo" [7] proposes a ubiquitous location-based recommendation algorithm that focuses on user experience by considering user preferences, time, location and similarity measures automatically, having Foursquare as a dataset. We also focus on user experience and aim that user input is minimal. The information from the user's social network, form of transportation and phone's sensors is inferred to provide recommendation of places from the dataset.

Recommending GPS trajectory is one of the main enhancements of the system described in [12]. GPS trajectory data mining is also the focus of [14], which extracts attributes from GPS trajectories in order to extract collective intelligence and recommend itineraries.

Another example is the system documented in [2], which recommends tourist locations based on user's visiting history in a geographically remote region. A set of geotags is used to compute location similarity and novel places are recommended to the user.

The agent system based described in [4] is based on prediction of the user's future behavior. The system understands the context from the GPS receiver and the prediction is performed by Dynamic Bayesian Networks. Finally, [10] proposes location-dependent collaborative filtering recommendation by using mobile user's location history and behavior prediction.

III. SYSTEM DESCRIPTION

In this section, we provide description of what are the typical use case scenarios, what our system actually recommends, and how it uses the four aspects of relevance identified in [3] as the recommendation context.

A. Use case scenarios

The work reported in this paper is part of MOPSI system. It implements various location-based services and applications such as search engines, data collection, user tracking and route recording. We have had so far 245 individuals using it. The recommendation system is available both on mobile device and on desktop computer. Mobile user interface is shown on Fig. 2 and web interface is shown on Fig. 3.
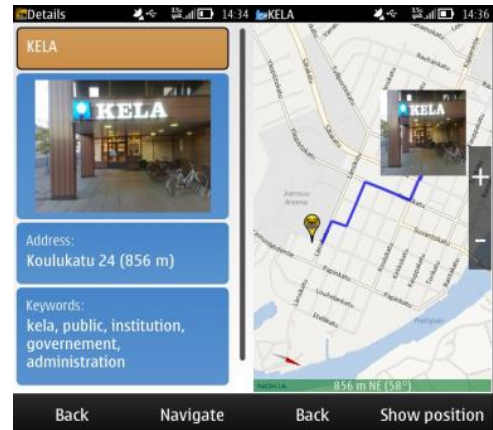

Figure 2: Recommendation results in mobile application. Left screenshot shows details of one of the recommended items. Right screenshot presents navigation screen to the selected item.
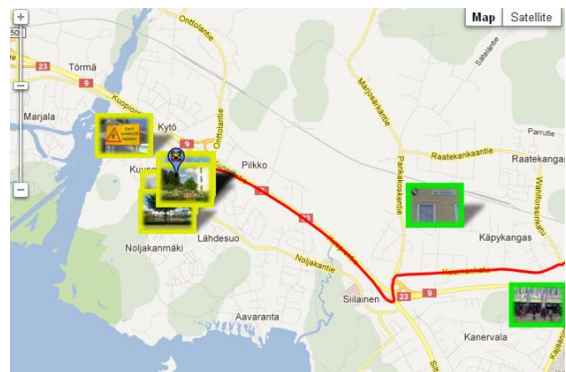

Figure 3: Recommendation system results on website.

The most common use case scenario of our system is user asking for recommendation using mobile device when the user is in a location he or she is not familiar with. Mobile access to the service is important since this is the most natural environment where the system can be most beneficial in real life. The key functionality of the system is that the user can ask for recommendation in any location so that he or she can easily visit suggested places immediately. The recommendation system suggests what the user can see and do in the surroundings. The results are displayed as list with the most recommended item on top. Depending on the type of recommended item, there are various additional statistics. The location of all items is also shown on map.

We provide recommendations also on website. In this case, usage scenario is that users search for what is interesting in the area before visiting it.

Access to the recommendation results requires connection to server, where the computations are done. In order to provide recommendation results immediately we calculate them in advance. Results are recalculated every time location of the user changes. The decision whether to recalculate or not a new set of recommended items is made on server-side by the recommendation synchronizer (see Fig. 4). The decision is based on the difference between user's current location and location in which recommendation was given previously to the user. Moreover, if user is in the same place, the time that passed since last recommendation is considered to assure that results are up to date.

B. Recommended items

MOPSI contains three databases that are used as a source for recommended items.

The first database contains trusted services verified by administrators illustrated by the green markers on the list and map in Fig. 3. These services represent variety of categories from restaurants, bars, and cafeterias, through grocery stores, pharmacies, and ATM machines, to car repairs, and museums. Service data include location, contact information, and relevant keywords as well as rating given by users.

The second database (user photo collection) contains photos users have taken using mobile phones and uploaded with several related information, such as location, time, and description as well as rating given by other users. Example of such collection is shown on top of Fig.1. In recommendation results these items are shown as yellow markers on the list and map in Fig. 3. MOPSI users collected 12.095 photos (6.8.2012).

The third database contains routes that users have recorded by mobile application. A route collection sample is shown at the bottom of Fig. 1. In recommendation results the routes are presented as red markers on the list and red lines on map shown in Fig. 3. Therefore it contains information about users' movements and places they have visited.

A route can be described using following characteristics: start time, location (set of route points), duration, length, transportation mode, novelty and attractiveness. Transportation mode in our system is one of the following: walking, running, cycling, or using motor vehicle. Detailed explanation on how we detect the transportation mode can be found in [11]. The route database in MOPSI contains 7.576 routes with 4.819.423 of individual points (6.8.2012).

C. Recommendation methods

In our recommendation system, we give personalized recommendations by combining various paradigms of recommendation systems. We combine collaborative filtering with information about user profile and context. As a source of recommendation we use the three different user generated data collections described above.

The challenge is how to select the most relevant items

to users. First we define the context for each recommendation request. In our previous work [3] we identified four aspects of relevance: location, content, time and network. Location is physical location of the user represented by geographical coordinates (latitude and longitude). Content is determined currently based on the description of the photos, by the keywords attached with the services or by the area covered by routes. Time is considered only for routes and photos and measures the age of the item and the season of the year when item was collected. One way we utilize the social network are the ratings given by other users to services and photos. For routes we consider that route novelty and attractiveness of the destination are concepts which use the network aspect. The use of the network aspect of relevance constitutes an integral part of the system based on collaborative filtering.

Bearing these contexts in mind, we create profile for each user of MOPSI. The user profile contains user behavioral data, such as location and previous usage of service, i.e. how user interacted with the system. Namely, we store data about location users have visited and searches they have performed (location and keyword used in previous search requests).

D. System implementation

In this section, we describe in details how we implement the system. Brief summary of the algorithm is given. The system architecture is presented in Fig. 4.

Users access our recommendation system using MOPSI mobile application available for Symbian, Android, iOS and Windows Phone operating systems. The same application is used for generating the data collections. Alternatively, users can use the website to get recommendations. Both website and mobile applications communicate with recommendation synchronizer as shown in Fig. 4.

The recommendation synchronizer is responsible for preparing recommendation results and keeping them up to date. The recommendation results for each user are stored in a file which contains also location and time what recommendation was calculated at. Whenever the user changes position, a request is sent to the recommendation synchronizer. Moreover, if the user is less than 3 clicks away from recommendation button the same request is sent. In practice it means that the synchronizer is called whenever the user enters main screen or screen with the recommendation button. The synchronizer decides whether to recalculate recommendations or not. The recommendation is recalculated whenever user position changes significantly or available recommendation results are outdated.

The recommender (see Fig. 4) contains implementation of the recommendation algorithm. The algorithm consists of three major steps. Firstly, the items available in databases are filtered by location. Secondly, the items that were left after filtering are scored using criteria derived from the aspects of relevance. Thirdly, the scores of all

three different types of items are merged together, sorted in descending order and first 20 items with the highest scores are returned as the recommendation results.
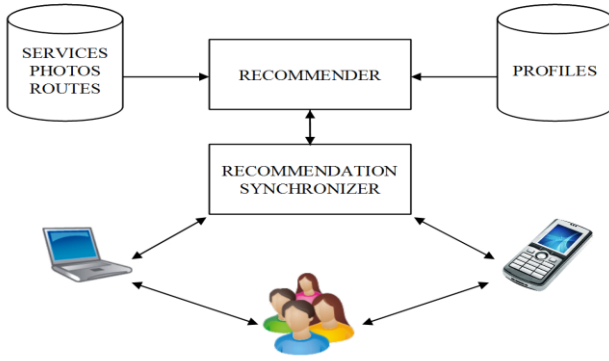

Figure 4: Architecture of the system.

IV. RECOMMENDER

In this section, we present in details the recommendation algorithm. The main conceptual steps of our algorithm are briefly described in previous section.

The algorithm input has three parameters. The first parameter identifies the user to whom recommendation is personalized for. The second parameter is the location of the user. The third parameter is the time of the recommendation request.

Firstly, we process each type of recommended items. We retrieve items from database. The full list of available items is then filtered by location to limit number of items for scoring. The remaining items are scored using several criteria based on aspects of relevance described earlier. These criteria are presented in details later in this section. The process of selection and scoring is performed for services, photos and routes separately.

Secondly, we merge all items retrieved in the previous step and sort them by total score in descending order. The recommendation results are the 20 top items in the list.

In the following subsections we describe in details how the three types of items are scored.

A. Services

Services are scored using search history, distance and rating criteria. We use both search history of all users as well as search history of the user who requested recommendation. The general search history is used to check what searches were performed in nearby locations and find what is in the area. If service keywords can be found among the keywords searched nearby, then the service is promoted by giving extra points. Furthermore, extra points are given for services with high frequency and keywords searched recently. Search history of the user is utilized in the same way.

Total score of search history consists of the following components:

$$S_H = S_{GN} + S_{GS} + S_{GF} + S_{UN} + S_{US} + S_{UF} \qquad (1)$$

where S_{GN}, S_{GS} and S_{GF} are the raw counts for keyword matches in nearby locations, within recent time and frequency of keywords in general search history and S_{UN}, S_{US} and S_{UF} are the same type of raw counts, but related to user history.

For location score we calculate distance from user to each recommendation item. By use of distance, we introduce location relevance aspect to the system.

Users can rate services through web and mobile interface of MOPSI. Services are rated by users in scale of 0 to 5. Rating and search history scores introduce content, social network and time aspects of relevance.

B. Photos

Photos are scored using search history, location, rating and time. Search history and location are used the same way as in case of services. The only difference is that service keywords are replaced by words used in photos descriptions. Rating of photos is cumulative, using a thumbs up/thumbs down system, for example a photo liked by 5 users and disliked by 2 has a rating of 3. The total score represents the rating score.

Additional score is given for recency, because the relevance of a photo decreases with time, as the places or views captured by users may change over time. Moreover, the season when the photo is taken is important for the relevance of photo, as for example winter activities are less relevant during summer.

More recent photos in the user collection are considered more relevant than old ones and the newer the photo is, the higher score it receives. Additional difference is that the score is also influenced by time of the year when the photo was taken.

Total score based on time (S_T) is for each photo calculated as follows:

$$S_T = S_A + S_Y \qquad (2)$$

where S_A is the recency and S_Y is the score for season of the year when photo was taken.

Moreover, photos are clustered into location clusters based on distances between them. Distance between photos that create a cluster is automatically decided based on distances between all photos selected for scoring. From such clusters we select only photos with highest scores. In this way we avoid recommending too many photos from the same location because we assume that the photos in same location present the same object.

C. Routes

Routes are scored using location, time and attractiveness. They are selected for scoring based on their starting point and its proximity to user's location. Main objective is to suggest user next places to visit. Therefore, only routes longer than 1 km are considered.

Location score of a route is the distance to its starting point from user location. Time score is the same as for photos.

A route is considered attractive if there is extensive collection of photos, services and other routes near its destination (destination attractiveness). Number of photos along a route increases its attractiveness (route popularity). Attractiveness score uses content aspect of relevance. Total route attractiveness score S_A is calculated using the following formula:

$$S_A = S_D + S_R \qquad (3)$$

where S_D is destination attractiveness score and S_R is popularity of the route.

Similarly as in photos, clustering is used for routes. In this case we build location clusters from end points of routes. From each cluster, we select route with the highest score.

D. Total score

All the above scores are normalized to the scale [0..1] using the following formula:

$$N = \frac{S - MIN(S)}{MAX(S) - MIN(S)} \qquad (4)$$

where S is the raw score, N is the normalized score, $MIN(S)$ and $MAX(S)$ are the minimum and maximum scores for each of the criterion respectively.

Final score of each service is then calculated using the following formula:

$$S_{SERVICE} = N_H + 2N_L + N_R + 1 \qquad (5)$$

where N_H stands for the normalized score for search history, N_L for location, and N_R for rating. Instead of using time relevance, a constant of one point is added in order to promote services for recommendation, because they are assumed to originate from a trusted source and therefore more relevant than older photos from user collection. The location score is multiplied by two to emphasize the importance of the location.

Final score of each photo item is calculated in the same way as services, having an additional time score:

$$S_{PHOTO} = N_H + 2N_L + N_R + N_T \qquad (6)$$

where N_T stands for time.

Final score of each route is calculated using the following formula:

$$S_{ROUTE} = 2N_L + N_A + N_T \qquad (7)$$

where N_A stands for attractiveness score.

V. SYSTEM EVALUATION

As we described in [12], the evaluation of user satisfaction is an important part of the evaluation of recommendation system. Moreover, as stated there, the feedback from users can be used to improve the system performance. The current improvements of our system are based on previously conducted experiments and on collected feedback. For example, recommending too many photos from one location was pointed out as a drawback of our previous system.

For testing our current system we used the same qualitative experimental settings as in previous work. We chose the city of Joensuu as the location of our tests, because the biggest number of service users has data collected in the area. Within Joensuu borders we selected several locations of various types. The locations list included city center, living areas, industrial area and recreational areas. We checked whether the recommendation is relevant for users. We also evaluated the items that were scored, but not recommended, in order to find out what relevant results may have been overlooked by the implemented scoring function.

Recommendations in city center always give many cafeterias, pizzerias and bars taken from service database. In addition, the system provides additional recommendations such as sport places and shops taken from photo collection. Also services that are not in service database, but are in photo collection of users, are recommended. Experiments have shown that all factors have impact on the recommendation results. For example, in suburban area with many housing blocks of flats and services nearby, there are many eating places recommended, but they are chosen based on rating and search history. Same example shows well that our recommendation system chooses relevant photos from the collection, such as shop, kiosk and mailbox, whilst general photos of streets, houses and people are skipped, although located nearby. In suburban are, on the other hand, where user generated collections are smaller, location has bigger impact on the relevance score.

In the course of experiments with our system, we noticed that it gives results expected by users mainly in case of services such as restaurants, bars and outdoor activities places (skiing, swimming). It does not perform as well in case of shops though, because of lack of data in users collections. There are cases where recommendation system does not suggest relevant only because it lacks description input by the user.

Experiments indicate that including time score is very important. The example in Fig. 5 shows recommendation in the same location in the middle of lake close to group of islands both in winter and in summertime. The place is popular destination for cross country skiers in winter. In summer the place is possible to access by own boat only and therefore rarely visited. Our system recommends the islands as place to visit when the lake is frozen and skiing tracks exist, because there are many routes recorded by users in wintertime. On the other hand, in summer it recommends barbecue places and swimming places on lakeside in town as no user activity is visible close to the islands.

Fig. 6 demonstrates the use of the clustering for photo recommendation. On the left of the figure we see photos recommended when location clustering is not applied. There are several photos located in a very small perimeter. Most of the photos are inside or around the same building. On the right side of the figure, only one photo is chosen to represent the cluster formed. Moreover, our system selects photos with best scores from cluster. In this way, we

avoid recommending irrelevant test photos described above that has high distance score, but theirs other scores are low.
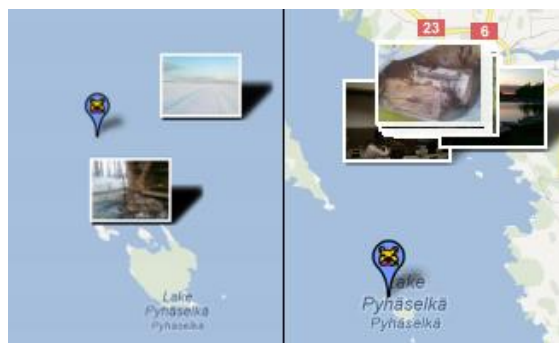

Figure 5: Recommended photos in winter (left) and summer (right).

Figure 6: Photo recommendation without location clusters (left) and with clustering applied (right).

VI. CONCLUSIONS AND FUTURE WORK

We have designed a context aware personalized recommendation system. Database of recommended items has free form and is generated by the users of MOPSI without any data cleansing. In this paper we study how to mine knowledge from user generated collections. We recommend three types of items: services, photos and routes. The goal of the system is to recommend points of interests to visit in user's surrounding.

The conducted experiments demonstrate that our system selects relevant items to recommend. Changes of algorithm we proposed in comparison to previous version of the system are beneficial for recommendation result quality.

However, despite the fact of positive feedback from system users, there is room for further improvements. Recommending routes was introduced to the recommendation system recently and scoring criteria should be improved. In some tests relevant routes were missed, because the routes had lower score than other items so they were not selected for recommendation. Moreover, routes recommendation brings new challenges. Route processing and computing the attractiveness score

is time consuming. Therefore, we consider storing route statistics in database in a similar way as users' profiles are stored.

The clustering concept can be expanded for the photo collection. We can create not only location clusters, but also content clusters based on photo descriptions. The photo content can be analyzed automatically and such keywords assigned in the process could be stored. Such solution will prevent us from relying merely on user provided description of photo as descriptions are often missing in photo collections.

Furthermore, the weights for different elements of total score could be adjusted automatically.

REFERENCES

[1] Bellotti, V. et al., "Acitivity-based serendipitous recommendation with the Magitti mobile leisure guide", ACM SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, April 2008.
[2] Clements, M., Serdyukov, P., de Vries, A. P., and Reinders, M. J. T., "Personalised travel recommendation based on location co-occurrence", The Computing Research Repository, 2011.
[3] Fränti, P., Chen, J., and Tabarcea, A., "Four aspects of relevance in sharing location-based media: content, time, location and network", Int. Conf. on Web Information Systems and Technologies, Noordwijkerhout, The Netherlands, May 2011.
[4] Kim, Y., and Cho, S.-B., "A recommendation agent for mobile phone users using Bayesian behavior prediction", Int. Conf. on Mobile Ubiquitous Computing, Sliema, Malta, October 2009.
[5] Levandoski, J.J., Sarwat, M., Eldawy, A., and Mokbel, M.F., "LARS: a location-aware recommender system", Int. Conf. on Data Engineering, Washington D.C., USA, April 2012.
[6] Li, L.H., Lee, F.M. and Chen, Y.C, "A multi-stage collaborative filtering approach for mobile recommendation", Int. Conf. on Ubiquitous Information Management and Communication, Suwon, Korea, January 2009.
[7] Savage, N.S., Baranski, M., Chavez, N.E., and Höllerer, T., "I'm feeling LoCo: A Location Based Context Aware Recommendation System", Lecture Notes in Geoinformation and Cartography, 2012.
[8] Schilke S.W., Bleimann, U., Furnell, S.M., Phippen, A.D., "Mand interest-based recommendation", Internet Research, Vol. 14, Iss: 5, pp. 379 – 385, 2004.
[9] Takeuchi, Y., and Sugimoto, M. "CityVoyager: An Outdoor. Recommendation System Based on User Location History", Int. Conf. on Ubiquitous Intelligence and Computing, China, 2006.
[10] Tuan, C.C., Hung, C.F., and Kuei, T.C., "Location dependent collaborative filtering recommendation system", Int. Conf. on Future Network Technologies, Qingdao, China, August 2011.
[11] Waga, K., Tabarcea, A., Chen, M., and Fränti, P., "Detecting movement type by route segmentation and classification", in press.
[12] Waga, K., Tabarcea, A., and Fränti, P., "Context-aware recommendation of location-based data", Int. Conf. on System Theory, Control, and Computing, Sinaia, Romania, October 2011.
[13] Yang, F., and Wang, Z.M., "A mobile location-based information recommendation system based on GPS and WEB2.0 services", WSEAS Transactions on Computers, Vol. 8 Iss: 4, pp. 725 – 734, April 2009.
[14] Yoon, H., Zheng, Y., Xie, X., and Woo, W., "Social itinerary recommendation from user-generated digital trails", Personal and Ubiquitous Computing, June 2011.
[15] Zheng, V. W., Zheng, Y., Xie, X., and Yang, Q., "Learning from GPS data for mobile recommendation", Artificial Intelligence Journal, February 2012.
[16] Zheng, Y., Zhang, L., Ma, Z., Xie, X., and Ma, W.-Y., "Recommending friends and locations based on individual location history", ACM Transactions on the Web, Vol. 5, No. 1, February 2011.