A Social Navigation Guide using Augmented Reality

Felix Mata
National Polytechnic Institute (UPIITA)
Mobile Computing Laboratory, Mexico
mmatar@ipn.mx

Christophe Claramunt
Naval Academy research Institute
Brest Naval, 29240, France
claramunt@ecole-navale.fr

ABSTRACT

Social networks provide rich data sources for analyzing people activities. This paper introduces a mobile recommender system that suggests places to visit to tourists acting in the city of Mexico. The system developed generates itineraries based on the implicit users' behaviors. Recommendations are automatically extracted and analyzed from Twitter thanks to the application of Bayes and Tree algorithms. Suggested itineraries are cross-analyzed to take into account user profiles and preferences. The recommender system provides an augmented reality navigation system that suggests itineraries to the users according to some places of interest. The preliminary prototype developed is an Android app so-called "Turicel Social".

Categories and Subject Descriptors

D.0 [General]: Location-based services. H.5.2 [User Interfaces]: Graphical User Interface

General Terms

Algorithms, Design, Experimentation

Keywords

Recommender systems, Outdoor navigation, Augmented Reality.

1. INTRODUCTION

The fast growth of online social networks has recently favored the development of social recommendation systems that rely on similar users' interests [11]. The algorithms applied are mainly based on collaborative filtering techniques [2, 3] which predict a user's interest regarding a given item by mining the patterns from previous rating information of similar users. Recommender systems have been mainly developed in the context of desktop computing and e-commerce applications [1], while their potential for the delivery of location-based services might offer several avenues to explore. However, users acting in mobile environments are in a very different position than the ones taking decisions in a desktop environment. This leads to the second motivation of the prototype development achieved in this paper: combining a recommender system with an Augmented Reality (AR) interface that implicitly superposes recommendations by augmenting the sense of reality and superimposing virtual objects and cues upon the real world [4]. While AR has been the object of increasing development in outdoor and indoor environments [5, 6, 8, 9], still the integration with recommender systems is a promising avenue of research.

The prototype presented in this paper is composed of four stages: 1) tweets retrieval; 2) data analysis; 3) data classification; 4) navigation-based AR interface. Firstly, the system developed collects tweets related to tourist activities. These tweets are analyzed and classified according to a place-based ontology performed by the application of C4.5 and Bayes classifier. C4.5 sorts the data based on possible categories, while the Bayes algorithm determines the most popular places visited by people at specific times in Mexico City locations. An AR-based navigation presents the suggested places by either locations or route itineraries. Finally, the user can filter the results according to her/his profile and preferences. The remainder of the paper is organized as follows. Section 2 describes the principles of our approach. Section 3 introduces the algorithms applied and the AR navigation interfaces while section 4 concludes the paper and outlines further work.

2. METHODOLOGY

2.1 Tweet filtering and sintactical analysis

Tweets are retrieved thanks to a series of tourism-related keywords of Mexico City. The prototype developed follows a knowledge discovery process as applied to a database repository [10] and is developed in three steps: A) **preprocessing** that includes data cleaning, integration, selection and processing B) **pattern matching** where the decision tree and Bayes algorithms are applied and C) data assessement and **presentation**.

The data preprocessing step A) is divided into two successive phases: 1) **Tokenization of tweets** where words are identified and separated 2) **Cleaning tweets** where symbols and signs are removed. A Java-based tokenizer and tagger (ark-tweet-nlp at https://code.google.com/p/ark-tweet-nlp) is applied. The output gives a set of words free of symbols and signs that are not relevant to the case. The data retrieval is applied to the Twitter social network and to several generic accounts specialized in tourism such as @PaseosDF, @MexDFTurismo, @turisteando_df and @turismo_DF. Most of the tweets made in these accounts are explicitly or implicitly related to places of interest and are issued by a large amount of users. At the cleaning phase, the date and hour related to each tweet is retrieved. Tweets that explicitly have a location are matched to the Geonames place names large database that gives several

SIGSPATIAL'14, Nov 04-07 2014, Dallas/Fort Worth, TX, USA ACM978-1-4503-3131-9/14/11.
http://dx.doi.org/10.1145/2666310.2666364

hundreds of possible locations in the city of Mexico (http://www.geonames.org/). A few examples of tweets retrieved are shown in Table 1.

Table 1 Place-related Tweets

Id	Date	Time	Tweet
1	23/05/2014	22:27	Excelente obra de teatro mexicano, ubicado en Insurgentes pasando Río Mixcoac #DF, #TimeOut,…
2	23/05/2014	10:30	Los tacos del #DF: delicias en la calle! pasando eje central dirección patriotismo #Turismo @turismocdmx @TorrucoTurismo http://ow.ly/i/5W0j7
3	23/05/2014	22:21	Un buen lugar para hospedarse en av. Chapultepec y Tolstoi cerca #Polanco http://ow.ly/y7VEL #turismo #mexico

The tokenization of the third record in Table 1 gives:
tweet{palabra { un, buen, lugar, para, hospedarse, en, cerca, Polanco, turismo, mexico, Chapultepec, av., y, Tolstoi }}

Stop words are removed (i.e., articles, pronouns, prepositions etc). Stop words have been derived from the stops-words of the Google code project (https://code.google.com/p/stop-words/). Next, these location-based tweets are related to events of interest. A previously built dictionary filters the tweets using a keyword match process based on a repository derived from the Tourism Secretary of Mexico (http://datatur.sectur.gob.mx/wb/datatur/). The services and events extracted from this repository are categorised and linked to a list of words (extract given in Figure 1).

2.2 Tree decision algorithm

At the preprocessing step, tweets are categorised. Tweets should be grouped when related to a given place or to the neighbourhood of a given place, event and time. We observe than 65% of tweets contain named entities. The classification is performed by the application of the C4.5 classifier, an extension of ID3 algorithm [7], that generates a decision tree.

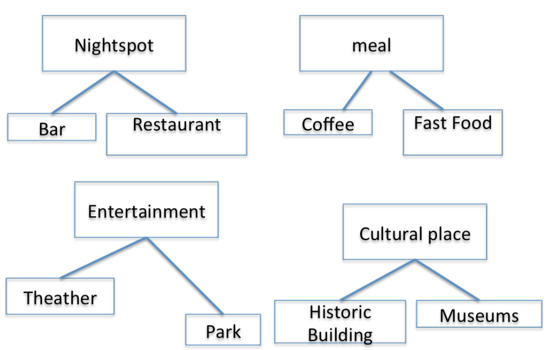

Figure 1 Categorization of events

Let us consider the tweet "*Los tacos del #DF: delicias en la calle! pasando eje central dirección patriotismo*". The words of that tweet are searched and matched to the hierarchy categorization. The "*tacos*" word is matched to the dictionary while "*eje central*" and "*patriotismo*" are matched to Geonames and labeled as *GeoRef1* and *GeoRef2* (two geographical references max are found for each tweet). Words that are not in the dictionary are discarded. The data structure resulting from this step is shown in Table 2.

Table 2 Categorized Tweets

Id	Date	Hour	Category	GeoRef1	GeoRef2
1	23/05	22:27	Entertainment	Insurgentes	Mixcoac
2	23/05	10:30	Meat	Viaducto	Eje Central
3	23/05	22:21	Lodging	Chapultepec	Tolstoi

At each node of the tree, C4.5 chooses the attribute that most effectively splits the data set of samples into subsets enriched in one class or the other by the evaluation of the entropy and information gain. The attribute with the highest entropy difference is chosen as the decision parameter. The C4.5 classifier recursively categorizes the tweets into smaller subsets. The output gives a classification of these tweets according to the places, times and events of interest identified. The C4.5 classifier is performed in five steps. The first one (I) derives the entropy values of the target attribute (i.e., the attribute that effectively splits the tweets). In fact the information distribution *E(S)* of the two classes generated is derived by the calculation of the entropy according to equations (1) and (2):

$$E(S) = \sum_{i=1}^{c} -pi \log_2 pi \qquad (1)$$

$$E(T, X) = \sum_{c \in X} P(C)E(C) \qquad (2)$$

Where c denotes the number of possible states, T is the first attribute and X the second one.
The next step (II) divides the data by attribute and calculates the information gain at each division. The gain is the difference between the amount of information needed for classification prior to and after the division. It is derived by determining the difference between the entropy of the target and the weighted sum of the entropies when the set is divided as follows:

$$Gain(T, X) = E(T) - E(T, X) \qquad (3)$$

At step (III) the attribute with the highest gain is chosen (i.e., attribute that splits the most effectively the tweets). The process is applied recursively (IV) until (V) all data are classified. A branch with a null value of entropy is a leaf node. Table 3 illustrates a subset of the classification results where tweets are classified per event, place and time.

Table 3 Classified tweets

		Event or place Type				
		Theather	Food	Trip	GeoRef1	GeoRef2
Hrr	22:27	1	0	0	Insurgentes	Mixcoac
	10:30	0	1	0	Viaducto	Eje Central
	22:21	0	0	1	Chapultepec	Tolstoi
Gain= 1.58						

In table 3 the "*Time*" attribute has three different values this giving a probability *1/n* of an event to happen at that time, *n* being the number of events. Let us consider this small example to illustrate the application of the algorithm. By applying equation (1) this gives *Entropy(Hour)* =-$[(1/3\log_2 1/3)]-[(1/3\log_2 1/3)]-[(1/3\log_2 1/3)]$. By applying equations (2) and (3) we have *Gain (Hour, Theather) = 1.58*. The same process is applied to attributes *Food*, *Trip*, *GeoRef1* and *GeoRef2*. The attribute *Time* is the attribute that gives the higher information gain (Figure 2). This gives the root node and the process is applied recursively.

2.3 Bayes classifier

The application of the Bayes algorithm provides the recommender dimension of the prototype development. The Bayes algorithm is combined with a matching process that relates the categorized tweets to user profiles and preferences. The user profile is composed by users' prefered activities (e.g., musem, park) and favorite foods. These values are explicitly requested by the system when the user log in for the first time and locally stored.

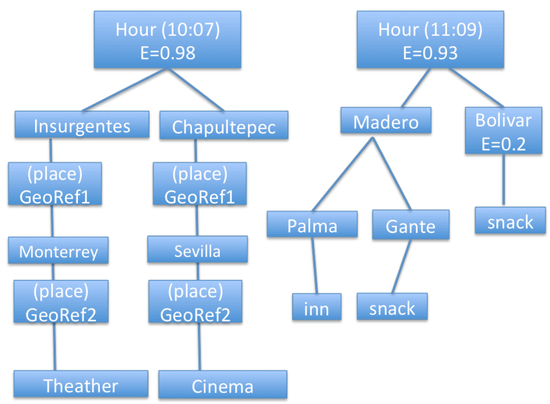

Figure 2 C4.5 generated tree (extract)

The principles of the approach are as follows. Let us consider *p* tweets classified as *Entertainment*, *q* as *Food* and so on. Each tweet might refer to a given place and/or event. When considering a given user searching for an event, place and time, the objective is to identify the best recommendation. Four parameters are considered: place, time, date, and category. The interest of the first three are computed using the Bayes algorithm while category is applied as a keyword-match between the user profile and the categorized tweets. In fact the Bayes algorithm derives which tweets are the most populars to be recommended according to the attributes derived from the user profile.

The Bayes search process determines the probability that an event occurs on a given day, time or place. Three probability values are derived, one per attribute. These probabilities are multiplicated to obtain the tweets with the highest probability values, that is, the candidate places and events to recommend. Next, these recommendations are keyword matched to the user profile in order to give the final recommendation. The whole process is based on the Bayes algorithm. The probability for each event to occur in a specific location at a given time should be derived. The goal is to find the most popular event for a specific location on a particular day (e.g., *play theather* in *Chapultepec* on *Thursday*; *fast food* in *Chapultepec* on *Thursday*). The probability for each of these events to occur at a given time is derived. Frequency tables are derived (i.e., number of times that an attribute appears on a tweet), this giving the probability for each event to occur at a place, time and date. Matching events allow to derive probability values. First, the probability that an event occurs on a given day at a given place is derived according to equation (4):

$$P(Ai/B) = \frac{P(B/Ai)P(Ai)}{P(B)} \qquad (4)$$

Where *P(Ai)* is the *a priori* probability, *P(B/Ai)* is the probability given by the hypothesis and *P(Ai/B)* the *a posteriori* hypothesis. Let us illustrate the process. The probability that an event occurs on Thursday at a given place is given as *P(GeoRef1yGeoRef2/ Thursday)*. The probability that an event occurs at 10:20 in some places is denoted as *P(GeoRef1yGeoRef2/10:20)*. The combination of these two probabilities gives the final probability. For example, the probabilities derived for the query "*Places to visit Thursday at 10:00*" according to Equation (4) are as follows: *Insurgentes, Mixcoac, 0.66; Chapultepec, Tolstoi, 0.66 and Eje Central, Viaducto, 0.33*. The first attribute values (*Insurgentes*, *Mixcoac*) denotes the event location, while the third one gives the probability of this event to occur at this location. The query is performed at the the interface level, times are given with a predefined interval of 30 minutes. For the example given the final result includes a play located at *Insurgentes* and *Mixcoac*, or a place to host at *Chapultepec* and *Tolstoi* streets (Figure 4).

3. PRELIMINARY RESULTS

The system starts with an interface that asks the user for an even or activity, time and place. The data retrieval process is performed using the twitter4j library that gives access to tweets (twitter.org). The experimental developments have been tested using Android phones version 4.0. A panel of 20 students in a period of 30 days has experimented the system. The search interface asks for an event or place at a given day and time. Figure 3 shows an example of query result, including the option to choose between AR or hybrid views. Users can select places using text- or image-based functions to go to this place, routes are suggested by either AR or AR-hybrid views according to the option selected.

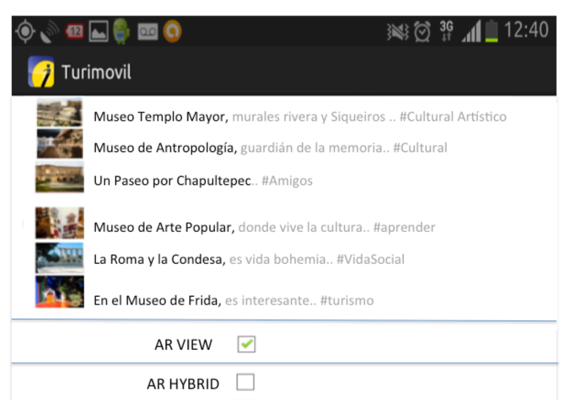

Figure 3 Search Results

Figure 4 shows an example of hybrid view (augmented view + Google maps View) when the user asks for "*Places to visit in Morning*", and the route generated.

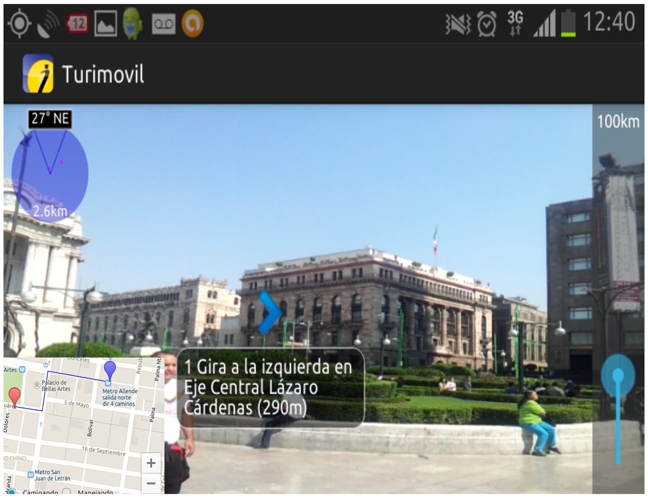

Figure 4 A guided itinerary

As illustrated in Figure 4, a blue arrow in the augmented view is shown in order to guide the user towards some popular places to visit. In the left corner of the screen, a route is displayed on top of a Google map and a textual route description is suggested. The user can also interact with the suggested AR Views. An additional view appears with a navigation drawer menu in the upper left corner. The items of this menu allow to show places, itinerary, personalized searches filtered by food, entertainment, cultural to mention some examples (Figure 5). As illustrated in Figure 5, the user can choose from a few menu options to explore the scenery, select additional places or to idenfity a place.

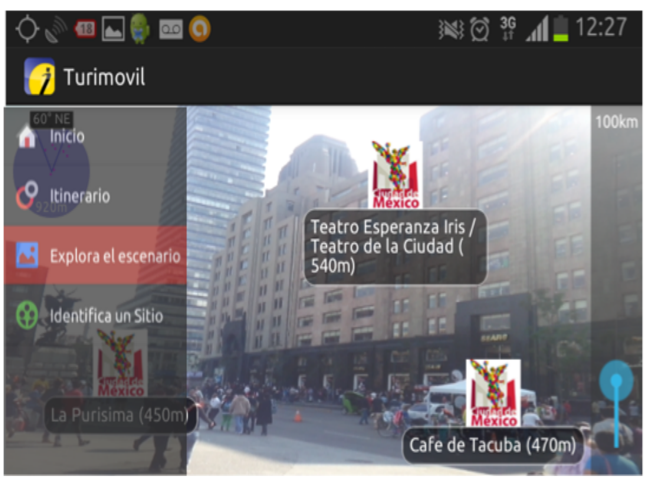

Figure 5 Popular places in an augmented View

4. CONCLUSION

This paper introduces a preliminary prototype of a social navigation guide. The system developed recommends places to visit or events to attend based on implicit information derived from a large tweet repository related to the city Mexico City. In constrast to previous works, this prototype provides an hybrid form of itinerary generation itineraries based on the combination of decision tree and Bayes algorithm. Recommendations are displayed using a AR navigation interface. This combines real sceneries with digital images of the environment, and information based on generated itineraries. The system developed has been tested on Cuernavaca an Mexico city. Current and further works are also oriented to additional validations and comparison with other recommender systems.

5. ACKNOWLEDGMENTS

The authors want to thank the SIP (Secretaría de Investigación y Posgrado), IPN (Instituto Politecnico Nacional), COFAA (Comisión de Operacion y Fomento a las Actividades Academicas del IPN) and UPIITA-IPN for their support.

6. REFERENCES

[1] F. Ricci, L. Rokach, B. Shapira, and P. Kantor. *Recommender Systems Handbook*. Springer, 2011.

[2] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.*, 17(6):734–749, June 2005.

[3] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, Jan. 2009

[4] J. Carmigniani, Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., and M. Ivkovic. Augmented reality technologies, systems and applications. *Multimedia Tools Appl.* 51, 1, 341-377, 2011.

[5] Lee J-Y, Lee S-H, Park H-M, Lee S-K, Choi J-S, Kwon J-S (2010) Design and implementation of a wearable AR annotation system using gaze interaction. *Consumer Electronics (ICCE)*, Digest of Technical Papers, pp.185–186, 2010.

[6] Feng Zhou, Henry Been-Lirn Duh, and Mark Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality* (ISMAR '08). IEEE Computer Society, Washington, DC, USA, 193-202, 2008.

[7] Quinlan, J. R. C4.5: *Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.

[8] Mata, F., Claramunt, C. and Juarez A. An experimental virtual museum based on augmented reality and navigation. In *Proceedings of 19th International Conference on Advances in GIS (ACM SIGSPATIAL GIS 2011)*, D. Agrawal et al. (eds.), Chicago, 1-4 November, ACM Press, pp. 497-500, 2011.

[9] Mata, F. and Claramunt, C. Augmented reality in outdoor spaces. In *Proceedings of 20th International Conference on Advances in GIS (ACM SIG'SPATIAL 2013)*, ACM Press, Orlando, November 5-8, pp. 534-537, 2013.

[10] Piateski, Gregory, and William Frawley. *Knowledge discovery in databases*. MIT press, 1991.

[11] X. Yang, H. Steck, and Y. Liu. Circle-based recommendation in on line social networks. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge discovery and data mining*, 2012.