Preference Extraction using Places of Visits

Misbahur R. Warsi
Z.H.College of Engg & Tech. A.M.U.
Department of Computer Engg.
A.M.U., Aligarh

ABSTRACT

Recently there has been interest in extracting preferences of an individual from his actions on mobile, laptop, desktop etc. This information is used in personalizing delivery of service to the person. In this paper, a scheme is presented to extract preferences from information about the places a person has visited. Mobile phone is used to record latitude/longitude of the places during his visit. Information about the area of visit is retrieved using map service. Characteristic of the area is determined and from this preference of the individual is extracted.

General Terms

Preference Analysis, Preference Extraction, Social Networks, Personalization.

Keywords

Visit Information Based Preference Extraction, Location Capture, Map Tagging, Map Division, Tag Score, Cell of Interest, Region of Interest.

1. INTRODUCTION

Recently there have been lots of interest in studying preferences and behavior of a person from his actions and in this IT infrastructure has played a major role in data collection. Advances in technology have brought people closer to computing more than ever which is evident from growth in usage of devices such as mobile phones, tablet, laptop etc. These devices are not only providing computing to a person but are also recording their action and thus accumulating vast amount of data. This data is analyzed to study the preferences of a person. For example : 1) By recording and analyzing the nature of content accessed by a person on internet, one can understand what is liked by a person 2) By analyzing reviews written by a person about hotel rooms where he has stayed, his preference about hotels can be determined.

Existing research on preference extraction can be analyzed from following perspectives – 1) Which data in the problem domain is considered for preference extraction and how it is represented and modeled 2) What is the algorithm/computational scheme for processing such a data 3) What type of preferences are being extracted. Nature and quality of data used in preference analysis decides the quality of result. Due to widespread use of social networking platform, it became one of the major source of data for preference analysis [1][2][3][6][8]. Social networking platform provided the data such as user messages, likes, comments, follow, graffiti wall etc which are useful in preference analysis. Users' actions and comments on websites have also been input in preference analysis. Content most accessed or liked on a web site, review comment posted on

website [11] or blog written by a person are some of other useful data. Some of the algorithms used in existing work are Cosine Similarity algorithm (used in [2] to calculate similarity between users) and PersonalRank algorithm. In [5], Yang Junchao et al have used MapReduce model in online shopping preference analysis. Work in this area can be further classified whether the work deals with specific types of preference such as hotel preference [4], electric vehicle preference [7], music preference [9], customer and shopping preference [8][5], preference analysis for email spamming [10] etc.

This paper presents details of a scheme which does preference analysis based upon place visit history of a person. Several devices such as mobile phone, camera, tablet etc. which a person normally keeps while on the move are capable of identifying their location. Much work has been done on developing several location based services. Few examples of location based services are: 1) Locating nearby restaurant, hospital, cinema hall etc. 2) Calculating distance in a workout 3) Tracking or locating whereabouts of a person or place. In this paper, location data is being captured and being used for preference analysis. As person moves and visits various places, data is collected using mobile devices. Subsequently, characteristics of places visited by the person are determined and based upon this preferences of the person is identified. Contributions of the paper is as given below –

- Visit Information Based Preference Extraction (VIBPE) – Present work extracts preference from the places visited by an individual. On the other hand, most of the existing work determines preferences by capturing user's action on computing device.
- For VIBPE scenario, this paper describes elements of problem domain and presents a computational scheme for same.

Next section (Section 2) in the paper presents details about the proposed scheme. It describes architecture of the system, theoretical aspect and algorithms used in the scheme. Finally last section (Section 3) concludes the paper.

2. ABOUT THE WORK

2.1 Architecture

Architecture of the system is shown in Fig 1. *MapDivision* divides map into smaller areas. Preference extraction is done based upon characteristics of smaller areas visited by a person. Location of the person is captured using mobile phone. An application running on mobile phone periodically captures location information in terms of longitude/latitude as he makes visit to a place. In *LocationProcessing*, location information is mapped to area/locality visited by the person and duration of stay in that area/locality is also determined.
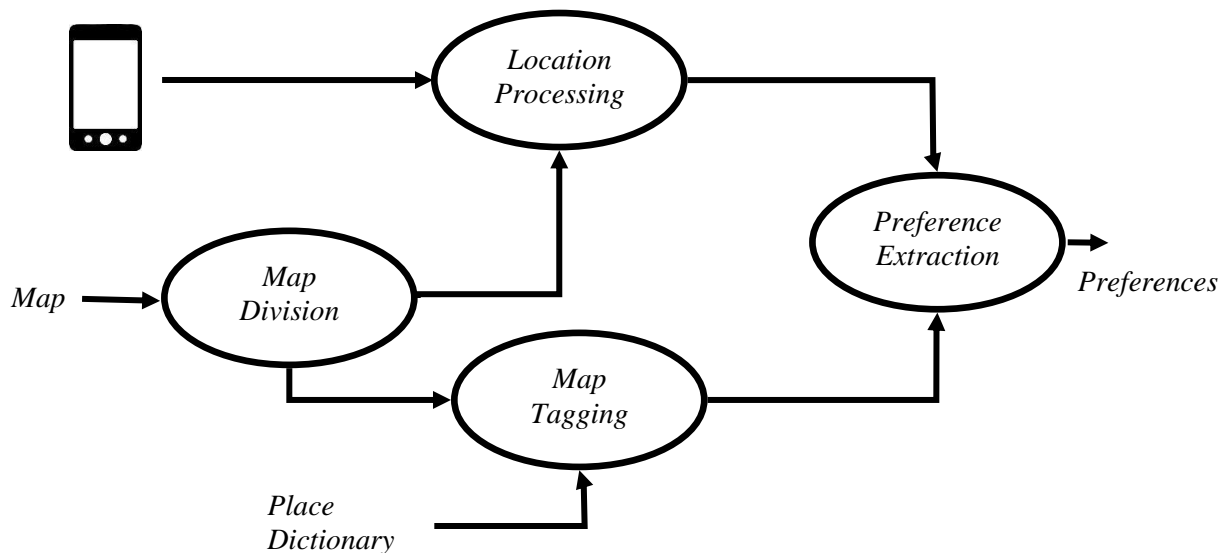
Fig 1: Architecture of the System

MapTagging deals with determination of characteristics of an area on map. It associates a tag with the area which specifies its nature or characteristics. A tag is represented with set of keywords picked up from *PlaceDictionary*. *PlaceDictionary* contains those words which typically describe a place or area such as market, shop, cinema hall, hospital, hill etc. *PreferenceExtraction* determines preferences from tags of places of visit.

Table 1. Definition of Region

Region Id	Cells Part of Region
1	<4, 9>, <5, 10>, <5, 11>, <4, 10>, <4,11>
…	…
…	…

2.2 Map Division

Map is divided into smaller areas known as cell having rectangular shape like a grid pattern. Size of the cell is small such that it accommodates roughly 3-4 shops. Cell is identified by its row position and column position which are also known as CellCoordinate (Table 2). Multiple cells grouped together form region. A region need not be in rectangular shape. A region is identified by a number and is defined by a group of cells as shown in Table 1. For example, region id 1 consists of 5 cells whose coordinates are mentioned in the table.

A cell captures detail about a very small area and at a lower level. When a person makes visit, he spends comparatively larger duration of time in some cells which are of interest to him. Such cells are identified and their characteristics are used to determine preferences of the person. On the other hand, a region captures detail at higher level. Characteristics of a region are determined from characteristics of its constituent cell. For example if a person visits a restaurant in a market, cell characteristics conveys that he loves dining and region characteristics conveys that he likes marketing. In other words, cell and region capture detail at two granular levels. In order to simplify computation, a restriction has been applied that a cell can belong to at most one region.

2.3 Location Processing

Location information which has been periodically captured using mobile phone is processed and as a result it gets converted in terms of cell coordinate and region ids that have been visited. Duration of stay in corresponding cell and region is also computed. Steps of *LocationProcessing* are shown in Fig 2. Each coordinate in *LocInfo* is converted to cell coordinate. Since location is captured at every *LocIntv*, each cell coordinate corresponds to stay in that cell for *LocIntv* duration. If stay in cell is longer that *LocIntv*, it repeats in sequence which is further collapsed to determine total duration of stay.

Table 2. Definition of Some Terms

Term	Definition
CellCoordinate	Row Number and Column Number of Cell
LocInfo	<Latitude1, Longitude1> <Latitude2, Longitude2> ...
LocIntv	Fixed time interval between two successive points on time scale at which location is captured
cVisitHis	Cell Visit History having following format - <CellCoordinate1, Duration of Visit1> <CellCoordinate2, Duration of Visit2> ....
rVisitHis	Region Visit History having following format - <Region Id1, Duration of Visit1> <Region Id2, Duration of Visit2> ....

2.4 Map Tagging

In map tagging, regions and cells of map are tagged with set of keywords. Tags describe a region or a cell. For example if a region is tagged with *<shopping, dining>*, it describes that it is a market with shops and restaurant. Similarly if a cell is tagged with *<shopping,medicine,doctor>*, it describes that the cell has medical store and doctor's clinic. The tag of a cell is

determined from description of geographical places which are within the cell. Synonyms are also considered while determining the tags. For example, the terms hotel, guest house and lodge are used interchangeably and therefore all these terms are included in tags.

Algo : **LocationProcessing**
Input *: LocInfo, LocIntv*
Output *: cVisitHis, rVisitHis*
Steps *:*
For each location lc in LocInfo do
{
Determine cell coordinate c corresponding to lc
If c is different from last coordinate in cVisitHis
*   Then Add <c, LocIntv> to cVisitHis*
*   Else Add LocIntv to duration part in last data*
*   in cVisitHis*

Determine region id r corresponding to cell c
If r is different from last region id in rVisitHis
*   Then Add <r, LocIntv> to cVisitHis*
*   Else Add LocIntv to duration part in last data*
*   in rVisitHis*
}

Fig 2: Location Processing Algorithm

Algorithm for map tagging is shown in Fig 3. It is done in two steps – one for determining cell tags and another for region tags. In first step for each cell, nearby places and their description is obtained from mapping service. The words in place dictionary matching with place description are selected and they form the tag of corresponding cell. In second steps, region tags are determined by taking union of tags of constituent cells.

2.5 Preference Extraction and Results

As mentioned earlier, cells and regions are tagged which their characteristics. Preference extraction is done from tags of selected cells (known as *CellOfInterest*) and selected regions (known as *RegionOfInterest*) from among the ones visited by the person. Major steps in preference extraction are shown in Fig 4. In this scheme, few parameters associated with cells and regions are considered and their values are calculated which fall in the range of 0-10. *CellOfInterest* and *RegionOfInterest* is defined for each parameter and includes those cells and region whose value of corresponding parameter is 6 or more. Let *CellOfInterest$_i$* and *RegionOfInterest$_i$* correspond to i_{th} parameter.

Though number of parameters is possible but in this work following parameters have been considered to determine *CellOfInterest* and *RegionOfInterest* –

- *Average Duration of Stay* – It is average duration of stay in a cell or in a region which might have been visited several times.
- *Maximum Duration of Stay* – It is the maximum value of the duration of stay in a cell or a region.
- *Frequency of Visit* – It is the number of visit to a cell or to a region.

Having determined *CellOfInterest* and *RegionOfInterest*, next step in the algorithm deals with tag score calculation. In tag score calculation, only those cells are considered which belong to *CellOfInterest* and only those regions are considered which belong to *RegionOfinterest*. Tag score is calculated as

the weighted sum of parameter values. Following weights are considered in the expression –

W_{1i}, *1<=i<=3*   *Weight associated with cell and i_{th} parameter*

W_{2i}, *1<=i<=3*   *Weight associated with region and i_{th} parameter*

Following factors have also been considered in preference extraction –

- Some cells and regions which do not contribute to preference of an individual are excluded from calculation of preference extraction. Example of such cells and regions could be the one which correspond to residence or office of person. These cells do not carry preference information but score high in most duration and most frequency criteria.
- Visit information corresponding to certain times which does not carry preference information is also excluded from calculation of preference extraction. Example of this could be visit information pertaining to office hours on week days.
- While determining which visit information to include in preference extraction, focus is on those which carry preference information such as visit information corresponding to weekend or holidays.

Algo : **MapTagging**
Input *: Cells without tagging,*
*   Regions without tagging*
Output *: Cells with tagging,*
*   Regions with tagging*
Steps *:*
1. For each cell c do
* {*
*   Obtain nearby places for cell c*
*   For each nearby place p of cell c which lies*
*   within the boundary of the cell do*
*   {*
*   Select set of keywords k from place*
*   dictionary which are matching with*
*   description of p;*
*   Add k to the tag of c*
*   }*
*   Remove duplicate keywords from tag of c*
* }*
2. For each region r do
* {*
*   Let rTag= Union of all tags of constituent*
*   cells in r*
*   Remove duplicate keywords in rTag*
*   Attach rTag with region r*
* }*

Fig 3: Map Tagging Algorithm

Preferences retrieved using above schemes were verified with users and some preferences were accepted while some were not accepted by the user. Following observations (as shown in Fig 5) are noted – 1) Tag acceptance percentage increases with increase in number of days for which visit information is collected (Fig 5(a)). It is due to the reason that with larger value of number of days, frequency of visit to a place and average duration of visit is accurately captured. 2) Tag Acceptance percentage is not uniform across different areas (Fig 5(b)). This may be attributed to following reasons –

Maps have limited accuracy and information about many places is missing from the map; and a cell or a region contains more than one place but not all are visited by a person.

1. *Let NoP be the number of selection parameter*
   Let $Score_t$ be the score of the tag t
2. *For each paremeter i, $1 <= i <= NoP$ do the following*
   {

   For each cell j calculate PVC_{ij} from cVisitHis where
   PVC_{ij} = Value of i_{th} parameter for j_{th} cell
   For each region j calculate PVR_{ij} from rVisitHis where
   PVR_{ij} = Value of i_{th} parameter for j_{th} region
   Let $CellOfInterest_i$ = Set of those cells whose PVC_{ij} is 6 or more
   Let $RegionOfInterest_i$ = Set of those regions whose PVR_{ij} is 6 or more

   }
3. *Tag Score Calculation*
   Let $score_t[1,i]$ be the score of tag t due to i^{th} cell which is initialized to 0
   Let $score_t[2,i]$ be the score of tag t due to i^{th} region which is initialized to 0
   For each parameter i do
   {

   For each cell c in $CellOfInterest_i$ do $score_t[1, c] += W_{1i} PVC_{ic}$
   For each region r in $RegionOfInterest_i$ do $score_t[2, r] += W_{2i} PVR_{ir}$
   }
   Score of tag t, $Score_t$ = Max ($score_t[i,j]$)
4. *Tags with high value of score is taken as preference of the individual*

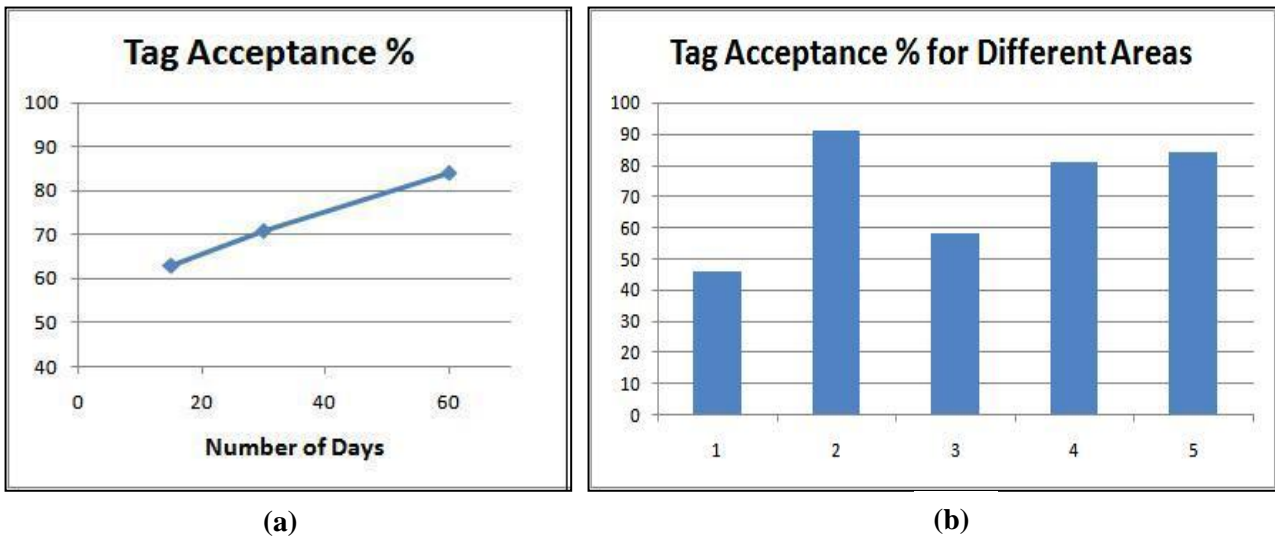**Fig 4: Computation of Preference**

(a)

(b)

Fig 5: Variation in Tag Acceptance %

3. CONCLUSION

In this paper, a scheme is presented to extract preferences of a person. In this scheme, places visited by the person are captured using mobile phone and from their nature and characteristics, preferences are determined. The scheme considers place of visit at two granular levels – cell and region. Characteristics of both the cell and region are considered in preference extraction. Mapping service may not have information about all the places (shops, cinema hall, clinic etc.) and therefore scheme is dependent upon amount of information available with mapping service.

Proposed algorithm for computation of preference offers scope for further experimentation by carrying out variations in future. Variations in the algorithm can be carried out along following directions – 1) more parameters associated with cells and regions may be introduced and the algorithm may be studied for different set of parameters 2) the algorithm may be studied for different functions to calculate tag score. Identification of visit information which carries more preference information is also a challenge and this aspect of the problem can also be studied.

4. REFERENCES

[1] Cheng-Hung Tsai; Han-Wen Liu; Ping-Che Yang; Tsun Ku; Wu-Fan Chien; "Social persona preference analysis on social networks"; 2015 International Conference on Connected Vehicles and Expo (ICCVE), 2015.

[2] Bo Shen; Bao-Wen Hu; Huan Zhang; "Method for the analysis of the preferences of network users"; IET Networks, Volume: 5, Issue: 1; 2016.

[3] Cheng-Hung Tsai; Han-Wen Liu; Tsun Ku; Wu-Fan Chien; "Personal preferences analysis of user interaction based on social networks"; 2015 International Conference on Computing, Communication and Security (ICCCS); 2015.

[4] Kai Zhang; Keqiang Wang; Xiaoling Wang; Cheqing Jin; Aoying Zhou; "Hotel recommendation based on user preference analysis"; 2015 31st IEEE International Conference on Data Engineering Workshops; 2015.

[5] Yang Junchao; Luo Jiangtao; Shen Jian; Deng Shengxiong; "Online Shopping Preference Analysis of Campus Network Users Based on MapReduce"; 2014 International Conference on Cloud Computing and Big Data; 2014.

[6] Shen Bo; Zhang Huan; Hu Baowen; "Research on Analysis Method of Network User Preference"; 2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing; 2014.

[7] Wonsuk Ko; Tae-Kyung Hahn; "Analysis of Consumer Preferences for Electric Vehicles"; IEEE Transactions on Smart Grid; Volume: 4, Issue: 1; 2013.

[8] J. S. Kim; M. H. Yang; Y. J. Hwang; S. H. Jeon; K. Y. Kim; I. S. Jung; C. H. Choi; W. S. Cho; J. H. Na; "Customer Preference Analysis Based on SNS Data"; 2012 Second International Conference on Cloud and Green Computing; 2012.

[9] Kunsu Kim; Donghoon Lee; Tae-Bok Yoon; Jee-Hyong Lee; "A music recommendation system based on personal preference analysis"; 2008 First International Conference on the Applications of Digital Information and Web Technologies (ICADIWT); 2008.

[10] Frank Jiang; Jin Gan; Yuanyuan Xu; Guandong Xu; "Coupled behavioral analysis for user preference-based email spamming"; 2016 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC); 2016.

[11] Kaiqi Zhao; Gao Cong; Quan Yuan; Kenny Q. Zhu; "SAR: A sentiment-aspect-region model for user preference analysis in geo-tagged reviews"; 2015 IEEE 31st International Conference on Data Engineering; 2015.