Asian Journal of Electrical Sciences
ISSN: 2249-6297 (P) Vol.3 No.2, 2014, pp.51-55
© The Research Publication, www.trp.org.in
DOI: https://doi.org/10.51983/ajes-2014.3.2.1923

Ranking and Optimizing of Location Based Services by User's Behavioral Patterns Using Data Mining Techniques

S. Ravichandran[1] and M. Umamaheswari[2]
[1]Research Scholar in Department of Computer Science, Bharathiar University, Coimbatore, India
[2]Prof. & Dean in Department of Information Technology,RRASE College of Engineering, Chennai, India
E-mail: ravi17raja@gmail.com, karpagaravi15@gmail.com,druma_cs@yahoo.com

***Abstract* - The utilization of the mobile technology devices is growing more and more, and the world is shrinking into a mobile. Recent advances in mobile technologies have meant that the technical capability to record and transmit location data for processing is appearing in off-the-shelf handsets. This opens possibilities to get the profile of the people based on the places they visit, people they associate with, or other aspects of their complex routines determined through persistent tracking using GPS enabled mobile phones. There must be recording of the users behaviors about the location information and stored it into the database. The services can be provided to offer customized information based on the results of such behavioral profiling could become commonplace. The GPS locations of the people have been mined to reveal places of interest and to create simple profiles. The information drawn from the profiling activity ranges from intuitive through special cases to insightful. User can send their desired query to find out the particular shop or object in real-time. This provides an efficient response to the user regarding their query, and it distinguishes which place is famous for which thing based on the previous queries from several users.**

Keywords: Behavioral profiling, location-based services, mobile, normality mining, privacy

I.INTRODUCTION

Developments in mobile phone devices are rapidly reforming our relationship with technology. The changes are not just technological they are driving changes in cultural and social paradigms, and further empowering the consumer to seek new experiences and services. The drive from industry to stay at the cutting edge has seen mobile phones turned into feature-packed computing devices within decade. With Internet capabilities, high-resolution cameras, GPS, and growing selections of third-party software applications, these devices are no longer simply mobile phones. Indeed, they are now more like mobile computers on which we can make phone calls. It appears that mobile handsets are the first wave of successful "wearable computers" at least in the sense that they comprise a relatively powerful computing device, which people habitually carry with them. Such devices will almost certainly be viewed retrospectively as the forerunner to "ubiquitous computing" and "ambient intelligent environments," other paradigm shifts predicted in our evolving relationship with technology. As it stands, new generations of handset are heralding a new era of information access and disclosure.

Advances in mobile technologies have meant that being able to track or locate people has been possible for some time. However, such information is usually only readily available to mobile phone operators. More recently, the advent of data-enabled mobile phones, and the emergence of popular social networking Internet sites have realized a dramatic increase in the volume of information people willingly disclose about themselves. In many cases, the disclosure is to large numbers of complete strangers. In February 2009, Nokia forecast that 50% of its handsets sold in 2009 would include a GPS unit. The 3G iPhone, with integrated GPS, supposedly held a 4%–6% share of the handset market in the U.K. toward the end of 2008. New services are likely to appear, which encourage people to reveal where they are at any time in the name of safety, convenience, or for social use. A danger here is that this begins to remove the division between safety, convenience, or entertainment and the invasion of personal privacy.

II.ADVENT OF NORMALITY MINING

Collating data and inferring new information, or knowledge, from it through the application of data mining techniques is not new. Database marketing, for example, makes use of data mining and (new) knowledge discovery to develop models of customer behavior. Typical barriers to accessing this information are technical ones if the data cannot be collected, then it cannot be exploited. However, today's information society is making enormous headway in the ability to collect data from disparate sources.

Such data-mining and knowledge-discovery techniques typically require as much data as possible in order to develop a suitable model. Indeed, our concept of "normality mining" is drawn from the fact that everything we do in our normal lives is of some potential interest and value. Collaborative social systems, termed "participatory sensing, for example, have begun to reveal the potential of massively distributed sensor systems, which collate information on the seemingly begin to enable a service. This type of "crowd-sourcing" or "wisdom of crowds" has given rise to services such as real-time traffic information based on live-user data. In this scenario, for example, the mundane chore of being held up in traffic is very much of interest. Further exploitation of location also gives rise to the potential for context specific search and targeted marketing (advertising).

Indeed, location-based services offer a new dimension to classic database marketing by disclosing the physical location. While this enables enhanced services or experiences, more marketing activities which explicitly address physical locations also become possible

A. Profiling Systems for Location Based Services

Modern mobile handsets, or "smartphones," allow observational access to domains of behavioral data not previously available even with constant observation and self "diary" reports. Data collection, which requires little or no user input, coupled with the strong, unique relationship between the handset and its user, make for behavioral data access on an unprecedented scale. GPS-enabled phones have helped generate a new era of information disclosure, and new services are likely to appear, which encourage people to reveal where they are at any time in the name of safety, convenience, or for social use. While these services will undoubtedly be useful, the seemingly harmless data may not just reveal where you are or have been it could expose aspects of your private life that at first glance may not be apparent. It is possible to aggregate pieces of information over time and use data-mining techniques to extract a "behavioral profile" from the data. Problems could arise if, for example, this information is used by third parties to vary their services or prices specific to the individual, in some cases to their detriment. This is especially so given that the end "user" may have no idea that this is happening, or indeed how it is happening. As an example application, consider "recommender systems." These are designed to help the user find what he is actually looking for. The idea of recommender systems is not new and a lot of such systems are now well known in practice.

The analysis of user behavior is accomplished by classical research separated into an active search, in which the user is asked explicitly about his behavior and a passive search where the behavior is derived by indirect information collection and interpretation. With the recent generations of mobile handsets, it is possible to download new software and services to the phone (e.g., the "apps" from the iPhone store), which can "enhance" the user experience. Recommender services exist in this context already, with most allowing the user to rate the suggested content to allow the service to learn the user's preferences. This is a type of active engagement. Consider the deployment of a recommender system, which attempts to passively build a profile, to take the burden away from the user. New generations of handsets are ideal for this purpose because not only do they have the capability to capture and send data, but also they are generally carried by an individual everywhere they go, practically at all times. This means GPS-enabled devices can persistently monitor where the user goes, data mining can be applied, and profiles can subsequently be drawn from the data. In the case of monitoring multiple users, social and business relationships can also be inferred. The advantage to the user is that, for example, when traveling to a new city, the user can ask the service individual specific, yet abstract questions: "where are restaurants that I would like?" or "where do people like me go to here?" etc. The downside is that the data used to create these profiles will inevitably reveal much more about the user.

To further explore this emerging technology, and to assess the potential impact on the privacy of the user in light of the European Data Protection Legislation, four people from three different European Member States took part in an experiment during April/May 2009. Over this period, they were persistently tracked via GPS-enabled mobile phones and their location data were stored in a central database for automated and manual processing. The aim of this processing was a first attempt to mine new information from the data relevant to forming behavioral profiles of the individuals. The information was based on where they had been and was used to investigate in what ways the location data collected over this experimental period supports our hypotheses that simple profiles of individual or group behavior can be drawn over short periods.

III. RELATED WORK

The application of GPS for varied services is manifold while several authors have examined the use of GPS data for simple location and tracking applications, the exploitation of the technology for profiling purposes is more limited. Data mining for profiling has been used extensively in many application scenarios, including e-commerce where data is easily obtainable, for a review of the techniques employed. While the application of profiling to GPS data is less well explored, the rapid adoption of mobile technology and the increasing functionality of smart phone handsets capable of relaying GPS data have led to a change in this trend. It has used GPS data and diary logs to seek an understanding of the social implications of tracking and monitoring subjects. Greaves and it has considered how a vehicle driving profile of a person can be derived from GPS data. In multiuser collaborative applications, predictive models have been derived based on in-car GPS to estimate performance measures of traffic flow and to identify black spots. Other studies have considered deriving context significant information from data including user action intentions and behaviors. The use of GPS to data stamp location information on to digital photographs has also been exploited to profile tourists by deriving movement information from uploaded images.

The discussion of such tracking from ethical and social stand-points has begun, however, none have examined in detail the inherent privacy issues associated with emerging applications, which will utilize behavioral profiles drawn from the GPS data.

IV. DATA COLLECTION AND PROCESSING

The design of the overall system employed was the result of multiple iterations based on conflicting factors. The resulting implementation wills each user was issued with a HTC P6500 handset with integrated GPS, running Windows Mobile 6.1; the unit provides an ideal development platform for this study, simplifying the hardware solution into a mainly integrated device, less prone to damage. Two pieces of custom software were run on all devices, the first logged each time the user authenticated themselves with the device (typically through the finger print reader). The second polled the internal GPS unit at 10-s intervals for updated data. The software then calculated the distance traveled and if greater than 100 m, then the cell tower details from the Windows Mobile Radio Interface layer were collected along with the last time of authentication from the log. Since the cell tower information only gives details about the id of the cell tower, the opencellid.org service was used to retrieve the GPS co-ordinates of the cell tower (if available). All of these discrete events were then sent for storage in a MS-SQL database.

The database can now store the data which has been selected from the individual user according to their interest on the places and the product, shop, etc. The user who is new to that city, who authenticates him selves and query about the details of the city, from inside the city, to find the nearest available places from that GPS point, the k-NN algorithm is used.

A. k-NN Algorithm:

KNN is a non-parametric lazy learning algorithm. That is a pretty concise statement. When you say a technique is non-parametric, it means that it does not make any assumptions on the underlying data distribution. This is pretty useful, as in the real world, most of the practical data does not obey the typical theoretical assumptions made (eg gaussian mixtures, linearly separable etc) . It is also a lazy algorithm. What this means is that it does not use the training data points to do any generalization. In other words, there is no explicit training phase or it is very minimal. This means the training phase is pretty fast. More exactly, all the training data is needed during the testing phase. (Well this is an exaggeration, but not far from truth).

a. Determine the parameter K = number of nearest neighbors
b. Calculate the distance between the query-instance and all the training samples
c. Euclidean Distance $\text{Dist}(X_1,X_2)=\sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2}$
d. Sort the distances for all the training samples
e. Determine the nearest neighbor based on the K-th minimum distance.
f. Get all the Categories of training data for the sorted value which falls under K.
g. Use the majority of nearest neighbors as the prediction value.

Choice of k

a. smaller k ⇒ higher variance (less stable)
b. larger k ⇒ higher bias (less precise)

Proper choice of k depends on the data: Adaptive methods, heuristics, Cross-validation.
In the contents of the database, the data are given ranking according to the user interestingness which has been entered by the user when they all passing through all situations. In this stage, the Page rank algorithm which is used to rank the web pages as per the user's frequent visiting of pages. This algorithm is applied to the database to provide the ranking among the user registered places or product, etc.

B. Ranking Algorithm

PageRank is a probability distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page. PageRank can be calculated for collections of documents of any size. The PageRank computations require several passes, called "iterations", through the collection to adjust approximate PageRank values to more closely reflect the theoretical true value. Ranking is given by,

$$PR(A) = (1-d) + d\,(PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))$$

Where,

- PR(A) is the PageRank of page A,
- PR(Ti) is the PageRank of pages Ti which link to page A,
- C(Ti) is the number of outbound links on page Ti and
- d is a damping factor which can be set between 0 and 1.

The PageRank theory holds that even an imaginary surfer who is randomly clicking on links will eventually stop clicking. The probability, at any step, that the person will continue is a damping factor *d*. Various studies have tested different damping factors, but it is generally assumed that the damping factor will be set around 0.85.The damping factor is subtracted from 1 (and in some variations of the algorithm, the result is divided by the number of documents (*N*) in the collection) and this term is then added to the product of the damping factor and the sum of the incoming PageRank scores.

By using these two algorithms, this proposed system will provide the efficient response to the query from the user about the nearest place and the famous product or a shop, etc.

The process was to determine the location identified as being of interest. It is expected that in a fully fledged profiling system, this would be an automated process utilizing multiple online resources such as maps, databases, and business listings. In such a system, established knowledge discovery, data-mining and data-fusion techniques, need to be complemented by emerging techniques, notably those which capitalize on spacio-temporal characteristics to enable the nontrivial processing. The system architecture for this processing is shown in the figure1.

This architecture shows that the users are requesting for the location based query to the server through a mobile and the location server will apply data mining to get most recently, frequently user information of the group of people.

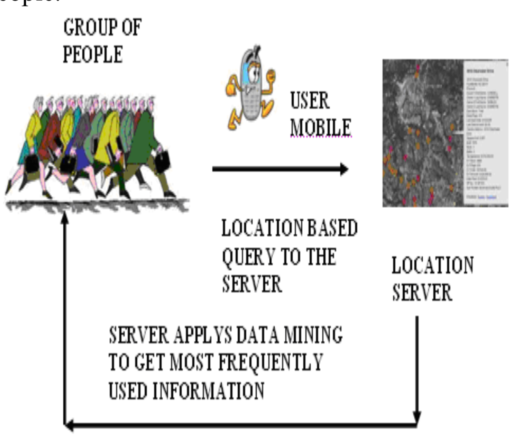

Fig.1 The system architecture

Although some algorithm have been utilized to mine the data. Much qualitative assessment has also been performed to aid in the exploitation

V.RESULTS

The results of the proposed system will be shown from the simulation work as a mobile using the WTK5.0, toolkit available to have the mobile simulation and can provide coding in Java.

The resultant services from this proposed system are the registration, Authentication to enter the data and also the query, user activity monitoring to get the details, then store it in the location database and providing ranking to the data to be easily available to the user at the querying time.

In the Location base query system, users have to register for their future query search. Without registering a user, can't access the clocking agent. For register, the user should give his details such as his name, address, age, sex, etc., once a user register his details he can get useful information from the clocking server. Each user will identify by a unique username and password.

If a client wants to arise a query, first he should be authenticated by the server. For this, he has to login by his user name and password. After he got sign-in, he can arise query to the server. This query will go the clocking agent and the clocking agent will send the query to the location server. User will be moving from one place to another place & requesting the services. This process is called as Location based Services. For example U2 starts out from location A and moves to B for a lunch; then to F for shopping; and to I for watching a movie. After the movie, U2 goes on to station K for a subway to final destination D. So our system is designed in a way that the user activity is monitored as per the request along with the Location.

Location Based services is the services which is provided to all the users & is utilized as the request given by each user. Location based services is maintained in the Location server as Database, which is updated with all the data regarding the place, services offered in that place. The Database which includes Recreation, Shopping, Education, Banks, Sports, Restaurants & all other commonly used services by any user. The user Behavior is monitored by the request provided by the user. This Behavioral pattern analysis will fetch the interest shown by the user in a particular place, for a particular time. This pattern helps the data provider in reading the Periodic interest of each & every user in a particular place. So this process finally could able to judge the frequent utilization of a particular data from the server.

This process will fetch the best data from the Data Server from the frequent data access by the different users. Based on the frequent data utilization by the different users, the best data is given as result to the current user. The Location based main server will store all the frequent data utilization by different users. Once the new user request for a query, the main database server will fetch the maximum number of Hits (number of frequent utilization) by the overall users. According to the number of Hits, ranks are maintained, based on the top Rank score, the results are retrieved to the Current user.

VI.CONCLUSION

It is evident that an enormous amount of information about the individual is buried in the data available from persistently tracking people. From this, results of profiling range from intuitive through special cases to insightful. It is, however, evident that firm conclusions on some personal and sensitive information such as family life, health, and data confidence, may only come from data collection over months, even years. Indeed, the importance of selecting a large enough sample and the difficulty in determining whether a sample is large enough in order to identify a location, especially given the imprecise clustering nature of the PoI data points, or to draw a reasonable conclusion from the data cannot be understated. If we consider the potential application of a simple recommender system, then one of the typical recommendations is likely to be restaurants in an unknown town or city. In this study, few occasions of eating out actually occurred and so even extrapolating this over several months, it could be difficult to draw good conclusions. In any case, the behavior of the individual may well differ greatly when they are in an unknown location for example, they may prefer local cuisine rather than whatever they go out to eat normally. Equally, they may prefer to spend more money on food and other goods when traveling. This type of idiosyncratic behavior will be complex to model and profile, and really could only come together after vast amounts of time monitoring the individual. What is clear is the real potential for incorrect conclusions being reached based on the data, and thus, inaccuracies pervading profiles. The impact of this on the individual could be significant, especially as the user may have no real means of redress or willingness to check and rectify information.

This paper has demonstrated that over the interests of the individuals involved to make sure that they are aware of how this information is being further exploited by those offering the services they are using. In the case where a service is offered by the mobile-phone provider, it is clear that since they hold a record of the identification (e.g., real name) of the individual for billing purposes, they can trivially link the phone and, thus, the data to an individual. The person, as the source of the data, is easily identifiable. Clearly, the individual could be identified from the data and the information can be considered as personal data. It is worth pointing out, however, that such a broad interpretation of the concept of identifiability, with regard to the personal data, has received great criticism from the industry.

ACKNOWLEDGMENT

The authors are thankful to L. Caronia and A. H. Caron, S. Mann and my guide for providing the necessary facilities for the preparation of the paper. Also thanks to AJEAT staffs to publish this paper. At last, I extend my heartfelt salutations to our beloved Parents, my Wife and to the almighty to establish this paper in successful manner.

REFERENCES

[1] L. Caronia and A. H. Caron, “Constructing a specific culture: Young people’s use of the mobile phone as a social performance. Convergence,” Int. J. Res. New Media Technol., vol. 10, no. 2, pp. –61, 2004.

[2] J. Garc´a-Montes, D. Caballero-Mu˜ oz, and M. Perez-Alvarez, “Changesın in the self-resulting from the use of mobile phones,” Media, Cult. Soc., vol. , no. 1, pp. 67–82, 2006.

[3] S. Mann, “Wearable computing: A first step towards personal imaging,” IEEE Comput., vol. 30, no. 2, pp. 25–32, Feb. 1997.

[4] N. Streitz and P. Nixon, “The disappearing computer,” Commun. ACM, vol. 48, no. 3, pp. 33–35, 2005.

[5] S. Figge, “Situation-dependent services—A challenge for mobile network operators,” J. Bus. Res., vol. 57, no. 12, pp. 1416–1422, 2004.

[6] M. Parameswaran and A. B. Whinston, “Research issues in social computing,” J. Assoc. Inf. Syst., vol. 8, pp. 336–350, 2007.

[7] L. A. Kurgan and P. Musilek, “A survey of knowledge discovery and data mining process models,” Knowl. Eng. Rev., vol. 21, pp. 1–24, 2006.

[8] G. Mort and J. Drennan, “Mobile digital technology: Emerging issues for marketing,” J. Database Mark., vol. 10, no. 1, pp. 9–23, 2002.

[9] S.-H. Liao and C.-H. Wen, “Mining demand chain knowledge for new product development and marketing,” IEEE Trans. Syst., Man, Cybern., Part C, vol. 39, no. 2, pp. 223–2, Mar. 2009.

[10] D. Cuff, M. Hansen, and J. Kang, “Urban sensing: Out of the woods,” Commun. ACM, vol. 51, pp. 24–33, 2008.

[11] N. Eagle, “Behavioral inference across cultures: Using telephones as a cultural lens,” IEEE Intell. Syst., vol. 23, no. 4, pp. 62–64, Jul. /Aug. 2008.

[12] G. Adomavicius and A. Tuzhilin, “Toward the next generation of rec- ommender systems: A survey of the state-of-the-art and possible extensions,” IEEE Trans. Knowl. Data Eng., vol. 17, no. 6, pp. 734–749, Jun. 2005.

[13] N. Poolsappasit and I. Ray, “Towards achieving personalized privacy for location-based services,” Trans. Data Privacy, vol. 2, no. 1, pp. 77–99, 2009.

[14] L. Perusco and K. Michael, “Humancentric applications of precise location-based services,” in Proc. IEEE Conf. E-Bus. Eng., IEEE Comput. Soc., Beijing, 2005, pp. 409–418.

[15] J. Shawe-Taylor, T. De Bie, and N. Cristianini, “Data mining, data fusion, and information management,” Proc. Inst. Elect. Eng. Intell. Transp. Syst., vol. 153, no. 3, pp. 221–2, Sep. 2006.

[16] N. Eagle and A. Pentland, “Social serendipity: Mobilizing social software,” IEEE Pervasive Comput., vol. 4, no. 2, pp. –34, Jan.–Mar. 2005.

[17] M. Raento, A. Oulasvirta, and N. Eagle, “Smartphones: An emerging tool for social scientists,” Sociol. Methods Res., vol. 37, no. 3, pp. 426–454, 2009.

[18] K. Michael, A. McNamee, M. Michael, and H. Tootell, “Location-based intelligence modeling behavior in humans using GPS,” in Proc. Int. Symp. Technol. Soc., New York, 2006, pp. 8–11.

[19] S. Greaves and C. De Gruyter, “Profiling driving behaviour using passive global positioning system (GPS) technology,” presented at the Inst. Transp. Engineers Int. Conf., Melbourne, Australia, 2002.

[20] N. Eagle and A. Pentland, “Social serendipity: Mobilizing social software,” IEEE Pervasive Comput. (Special issue: the smart phone), vol. 4, no. 2, pp. –34, Jan.–Mar. 2005.

[21] J. Himberg, K. Korpiaho, H. Mannila, K. J. Tikanma, and H. Toivonen,“Time series segmentation for context recognition in mobile devices,” in Proc. IEEE Int. Conf. Data Mining (ICDM), San Jose, pp. 203–210, 2001.