위치 클러스터를 사용한 개인 특성화 식별

Towards User Personality Identification by Using Location Clustering

Rajashree S. Sokasane(소가서네 라저스리)*, Kyungbaek Kim(김경백)**
*전남대학교 전자컴퓨터공학부 박사 학생 sokasaners@gmail.com
**전남대학교 전자컴퓨터공학부 조교수 kyungbaekkim@jnu.ac.kr

Abstract: Nowadays, identifying a personality type of a user by using its smartphone usage has getting more attention in the field of research. Existing approach towards detecting a user's personality is based on the usage of smartphones, such as call detail records (CDRs), the usage of short message services (SMSs) and the usage of Internet. In this paper, we focus on the possibility of using clusters of user locations as a new feature in the direction of prediction of user's personality preference. We applied two basic location clustering techniques (K-mean and DBSCAN) to GPS log data of each distinct user which has different personality. As a result, we observed that location clusters are correlated to users' personality and DBSCAN is better to represent user personality than K-mean.

1. Introduction

Nowadays usage of smartphone is tremendously increased. Determining a user's personality type by using their smartphone usage and providing services according to their personality preferences, is become more attention-grabbing topic in today's research space. In recent years, there has been an increased interest in the Computer Science community (and particularly HCI) [6] on the importance of personality profiles: models of the users' personality traits and preferences can be used to adapt and personalize services. For example, Personality Psychology has provided evidence on the influence of different personality traits over leadership, performance and group interaction styles [3]. Assessment of personality preferences is based on MBTI (Myers-Briggs Type Indicator) theory [1]; the indicator is frequently used in the areas of pedagogy, career counseling, team building, group dynamics, professional development etc. [5].

Existing approaches headed for detecting a user's personality is based on smartphone usage data, such as information extracted from call detail records (CDRs) [1], the usage of short message services (SMSs) and the usage of web, music, video, maps, proximity information derived from bluetooth etc., in addition to the traditional call and SMS usage information [2]. Unlike existing approaches, in our approach we use clustering user's locations by using user's GPS log information collected via user's smartphone.

In this paper, we focus on the possibility of using clusters of user locations as a new feature. To show the possibility of detecting personality preference using location clusters, we gathered user's GPS log data and personality type of user with the help of MBTI [1]. We applied K-mean [10] and DBSCAN [9] clustering techniques to GPS log data of each distinct user which has different personality preference. From our experimental study, we observe some patterns after applying clustering technique, which shows correlation between location clusters and personality type; and we found that DBSCAN works better than K-mean with respect to representation of user personality.

The paper is organized as follows. Section 2 describes past work relating to MBTI theory, importance of personality prediction and its usage. Section 3 describes the proposed approach in detail, and the clustering techniques used with it. Section 4 describes evaluation study with the details of dataset used for and results we obtained with proposed approach. Finally, in section 5 we provided conclusion and the future plan.

2.Background

2.1MBTIasUserPersonality

Understanding your preferences can help you in all sorts of ways, such as finding fulfilment in your career, improving relationships, or developing your leadership skills. The purpose of the Myers-Briggs Type Indicator (MBTI) personality inventory is to make the theory of psychological types described by C. G. Jung understandable and useful in people's lives [5]. The 16 distinctive personality types are generated by using the four pairs of preferences or dichotomies viz. 1. Extraversion (E) - Introversion (I), 2. Sensing (S) - iNtuition (N), 3. Thinking (T) - Feeling (F), and 4. Judgment (J) - Perception (P).

The preferences for extraversion (E) and introversion (I) are often called "attitudes". The extravert's flow is directed outward toward people and objects, and the introvert's is directed inward toward concepts and ideas [1]. Sensing and intuition are the information-gathering (perceiving) functions. They describe how new information is understood and interpreted. Sensing perceives facts and tangible reality, iNtuition perceives possibilities and delves into the unknown. Thinking and feeling are the decision-making (judging) functions. Thinking makes decisions on the basis of objective logic, and seeks to establish what is true or correct. Feeling makes decisions on the basis of subjective values, and seeks to establish what is important or worthy of attention. According to Myers, judging (thinking or feeling) types like to "have matters settled" and perceptive (sensing or intuition) types prefer to "keep decisions open" [1].

In our work, we focus on the E and I personality preference. We provide the possibility of using clusters of user locations as a new feature in the direction of detection of user's personality preference (E and I).

2.2RelatedWork

In recent years, the field of Human-computer interaction (HCI) has emphasized the importance of identifying the users' personality traits and preferences in order to build adaptive and personalized systems with an improved user experience [6]. In [2], authors have analyzed the relationship be-tween smartphone usage and self-perceived personality. Applications usage, call and SMS logs contained several meaningful relationships to the Big-Five personality framework. Their study is based on a large-scale dataset of 8 months of real usage of smartphones by 83 people and personality surveys that are suitable for large mobile or online studies. Their feature set was enriched with information automatically extracted from a variety of sources and collected from smartphones of a large number of users, using non-intrusive software. The idea of predicting people's personalities from their cellphone stems from recent advances in data collection, machine learning, and computational social science showing that it is possible to infer various psychological states and traits from the way people use every day digital technologies [8]. The study provides the evidence that personality can be predicted from standard carriers' mobile phone logs [2] [8]. Another work [4], suggest that variables derived from the users' mobile phone call behavior as captured by call detail records and social network analysis of the call graph can be used to automatically infer the users' personality traits as defined by the Big Five model.

In this paper, we provide location clustering as a new feature with the possibility of detecting user's personality preferences based on user's GPS log data.

3. Location Clustering for Distinguish User Personality

In this paper, we introduced clustering user locations with the possibility of detecting user's personality preferences. We used user's GPS log data and MBTI based personality type to add possibility towards detection of user's personality preferences. As we described earlier in section 2.1, extraverts are more social; they involve in social activities such as making new friends, visiting new location and interacting with others. In contrast introverts are less social; they prefer explore ideas and concepts internally. We use clusters of user locations to represent the personality preference E and I. The extraversion people are social and distributed; as a result they visit many locations. The introversion people are less social and concerted; as a result they are less distributed over locations. Based on this hypothesis, we applied two basic location clustering techniques (K-mean and DBSCAN) to GPS log data for clustering user's locations, towards predicting user's personality preferences (E and I).

3.1 DBSCAN

Clustering algorithms are attractive for the task of class identification in spatial databases. Density-based clustering methods are an important category of clustering methods that are able to identify areas with dense clusters of any shape and size. DBSCAN (Density Based Spatial Clustering of Applications with Noise) is one of them. DBSCAN is simple and effective density-based clustering algorithm that illustrates a number of important concepts that are important for any density-based clustering approach. DBSCAN is efficient even for large spatial databases [9].

3.2 K-means

The process, which is called 'K-mean,' appears to give partitions which are reasonably efficient in the sense of within-class variance [10]. The aim is to divide the data into k distinct groups so that observations within a group are similar, whilst observations between groups are different. It is simple and computationally efficient, but can sometimes be sensitive to the selection of starting points. Running the k-means algorithm several times for different starting values can help check whether results are robust. K-means is generally applicable to data sets with continuous valued feature vectors, and in principle it is not suitable for data with nominal (categorical) coordinates.

4. Evaluation of Location Clustering

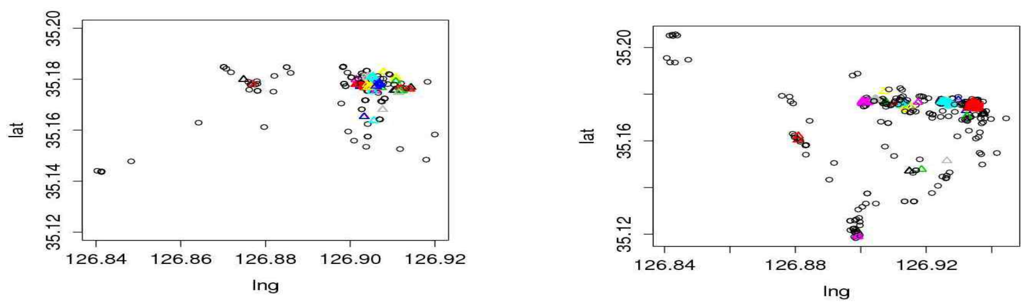
4.1 Dataset

To evaluate the performance of location clustering to distinguish user personality, we gathered data for a month from 2013 October to 2013 November. We collected number of users GPS log data for every 15 minute. And also we collected MBTI-based personality type data of each user. During our experimentation, we used INTJ and ESTP; the two distinct personality type users' GPS log data.

Note that with the proposed approach, we only analyze GPS log information to get location information to study only user's movement behavior. Therefore, the data cannot be used in any way to reveal the users' identity or to obtain personal information.

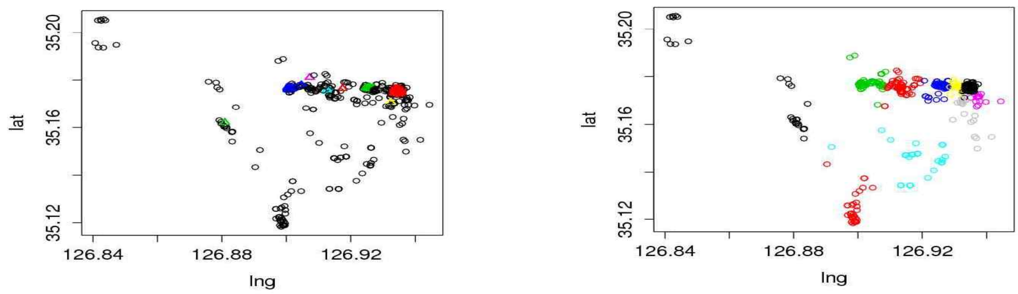
4.2Results

<Figure 1> shows the comparison between the results, obtained after applying location clustering by using DBSCAN with extraverted and introverted. We apply location clustering and calculated the mean of every cluster. And we observed that the extravert (personality type is ESTP) is distributed over 1.75 km distance and introvert (personality type is INTJ) are distributed over 0.75 km distance. From this observation, we can say that there is possibility of detecting user's personality preference based on location clustering. Figure 1(b) shows the result of location clustering based on GPS log data of extravert user who belongs to ESTP personality type; Figure 1(a) shows the result of location clustering based on GPS log data of introvert user who belongs to INTJ personality type. As per MBTI theory, Extravert's flow is directed towards people and objects; by using location clustering we observed that user visited multiple places/persons might be extravert. And Introvert's is directed inward towards concepts and ideas; they are less distributed or less social as compared to extraverts.



(a) DBSCAN Eps=0.0005 Minpts=5 (I) (b) DBSCAN Eps=0.0005 Minpts=5 (E)

<Figure 1. Introverted Vs Extraverted>



(a) DBSCAN Eps=0.002 Minpts=10 (E) (b) K-mean k=10 (E)

<Figure 2. DBSCAN vs. K-mean>

<Figure 2> shows location clustering results after applying DBSCAN and k-mean clustering techniques respectively. "Location Clustering" is considered as new feature towards identification of personality preferences. DBSCAN and K-mean were applied on range of dataset to see which clustering algorithm works well; and the experimental results shows that DBSCAN shown in figure 2(a) works better than k-mean shown in figure 2(b) and produces good results. With k-mean clustering, points in a cluster are dispersed and hard to find exact location. DBSCAN is coarse-grained; and works better to represent user personality. Also, to get better results, we not only choose proper method but also we need to set proper parameters to proper method. Figures show that DBSCAN with distance 50m (Eps=0.0005) better results than DBSCAN with distance 200m (Eps=0.002).

5. Conclusion

In this paper, we presented location clustering as a new feature with the possibility of identifying personality preferences of a user. We applied K-mean and DBSCAN clustering techniques to GPS log data of each distinct user which has different personality preference. Evaluation section shows analysis of location clustering towards identifying user personality and the results obtained with it. From the results obtained in evaluation section, we can say that there is possibility of detecting user's personality preference based on location clustering. We observe that "location clustering" is working as a feature for identifying the personality preferences of a user; and location clusters are correlated to users' personality preference. After applying K-mean and DBSCAN, we observed that DBSCAN clustering method works well than k-mean clustering method to represent user personality. By using location clustering, we may able to detect whether the user is extravert or introvert in nature.

Future work will focus on creating Bigdata storage platform to save GPS log data, MBTI data and smartphone log data; and design an engine to analyze the Bigdata, to obtain more expressive results.

Acknowledgment

This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2014-H0301-14-1014) supervised by the NIPA(National IT Industry Promotion Agency)

References

[1] Myers, Isabel Briggs, and Mary H. McCaulley. "Myers-Briggs Type Indicator: MBTI". Consulting Psychologists Press, 1988.

[2] Chittaranjan, Gokul, Jan Blom, and Daniel Gatica-Perez. "Who's who with Big-Five: Analyzing and Classifying personality traits with smartphones." InWearable Computers (ISWC), 2011 15th Annual International Symposium on, pp. 29-36. IEEE, 2011.

[3] P. Balthazard, R. E. Potter, and J. Warren. "Expertise, extraversion and group interaction styles as performance indicators in virtual teams: how do perceptions of its performance get formed?" SIGMIS DB, 35(1):41 - 64, 2004.

[4] de Oliveira, Rodrigo, Alexandros Karatzoglou, Pedro Concejero Cerezo, Ana Armenta Lopez de Vicuña, and Nuria Oliver. "Towards a psychographic user model from mobile phone usage." In CHI'11 Extended Abstracts on Human Factors in Computing Systems, pp. 2191-2196. ACM, 2011.

[5] http://en.wikipedia.org/wiki/Myers-Briggs_Type_Indicator
[6] K. M. Lee and C. Nass. “Designing social presence of social factors in human computer interaction" In CHI ’03: Proc. of the SIGCHI conf. on Human factors in computing systems, pages 289 - 296, New York, NY, USA, 2003.
[7] Nov, Oded, Ofer Arazy, Claudia López, and Peter Brusilovsky. "Exploring personality-targeted UI design in online social participation systems." InProceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 361-370. ACM, 2013.
[8] de Montjoye, Yves-Alexandre, Jordi Quoidbach, Florent Robic, and Alex Sandy Pentland. "Predicting personality using novel mobile phone-based metrics." InSocial Computing, Behavioral-Cultural Modeling and Prediction, pp. 48-55. Springer Berlin Heidelberg, 2013
[9] Ester, M., Kriegel, H.-P., Sander, J., and Xu, X., \A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," In Proc. 2nd Int'l Conf. on Knowledge Discovery and Data Mining, Portland, Oregon, pp. 226-231, Aug. 1996.
[10] Some methods of classification and analysis of multivariate observations In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (1967), pp. 281-297 by J. B. Macqueen