IEICE TRANS. INF. & SYST., VOL.E98–D, NO.8 AUGUST 2015

LETTER

Predicting User Attitude by Using GPS Location Clustering

Rajashree S. SOKASANE[†a)], ***Nonmember and*** **Kyungbaek KIM**[†b)], ***Member***

SUMMARY In these days, recognizing a user personality is an important issue in order to support various personalized services. Besides the conventional phone usage such as call logs, SMS logs and application usages, smart phones can gather the behavior of users by polling various embedded sensors such as GPS sensors. In this paper, we focus on how to predict user attitude based on GPS log data by applying location clustering techniques and extracting features from the location clusters. Through the evaluation with one month-long GPS log data, it is observed that the location-based features, such as number of clusters and coverage of clusters, are correlated with user attitude to some extent. Especially, when SVM is used as a classifier for predicting the dichotomy of user attitudes of MBTI, over 90% F-measure is achieved.

key words: *user personality, location clustering, feature extraction, attitude, SVM*

1. Introduction

Psychological studies on personality have provided evidence on the influence of different personality traits over leadership, performance and group interaction styles [3]. For example, MBTI (Myers-Briggs Type Indicator) theory [1] is used to assess the personality preferences of user; MBTI is frequently used in the areas of pedagogy, career counseling, team building, group dynamics, professional development etc [1]. Personality prediction has many useful purposes in marketing, organization development, customized user interface (UI) and personalized recommendations. In recent years, there has been an increased interest in human computer interface (HCI) [5] on the importance of personality profiles; models of user personality preferences can be used to adapt personalized services.

Recent research devoted towards predicting user personality is based on smart phone usage data, such as information extracted from call detail records (CDRs), the usage of short message services (SMS) and the usage of web, music, video, maps, proximity information derived from bluetooth, usage of internet etc [2]. These research focused only on how to use *smart phone usage data* to predict user personality. In addition, number and types of applications installed on a smart phone [8] can be used to predict user personality. However, recent smart phones equip with various kinds of sensors and they can observe *user behaviors* as sensor logs. These sensor logs can be used for predicting user personality preferences.

In this paper, we use location information gathered from GPS sensors for predicting user attitude. Especially, we propose a new method of extracting classification features to predict user attitude by applying location clustering techniques to GPS log data. The proposed method applies either K-mean or DBSCAN [7] clustering techniques to GPS log data, and calculates *number of clusters* and *coverage of clusters* as classification features. The *coverage of clusters* can be further categorized into *average/maximum distance covered* and *average/maximum weighted distance covered*, according to consider the frequency of visits to locations.

To evaluate the correlation between the proposed classification features and user attitude, we gathered GPS log data of 30 users for a month and personality type of each user with the help of MBTI, and evaluate the performance of various classifiers such as Naïve Bayes, Support Vector Machine (SVM), decision tree, and k-NN with the proposed features. From the extensive evaluation, we observed that SVM classifier with the proposed features achieves about 90% F-measure of predicting user attitude.

2. Background

2.1 MBTI as User Personality

Generally, the assessment of personality preferences is based on MBTI (Myers-Briggs Type Indicator) theory and Big-Five personality traits. The goal of the MBTI is to allow user to further explore and understand their own personalities including their likes, dislikes, strengths, weaknesses, possible career preferences, and compatibility with other people [1]. According to MBTI theory, the 16 distinctive personality types are generated by using the four pairs of preferences or dichotomies viz. First dichotomy Extraversion (E) - Introversion (I) represents attitude. Second dichotomy Sensing (S) - iNtuition (N) are the information-gathering functions. Third dichotomy Thinking (T) - Feeling (F) are the decision-making functions; and the fourth dichotomy Judgment (J) - Perception (P) indicates lifestyle.

The extraversion-introversion dichotomy is used as a way to describe how people respond and interact with the world around them [1]. The extravert's flow is directed outward toward people and objects, enjoy more frequent so-

Manuscript received December 4, 2014.
Manuscript revised April 6, 2015.
Manuscript publicized May 18, 2015.

[†]The authors are with the Department of Electronics and Computer Engineering, Chonnam National University, Gwangju, Korea.
a) E-mail: sokasaners@gmail.com
b) E-mail: kyungbaekkim@jnu.ac.kr (Corresponding author)

DOI: 10.1587/transinf.2014EDL8245

cial interaction, and feel energized after spending time with other people. The introvert's is directed inward toward concepts and ideas, enjoy deep and meaningful social interactions, and feel recharged after spending time alone. Some characteristics of extraverts such as, meeting other people, enjoying events and visiting new places usually involve the locations of a user. As a result, location information of a user may helpful to identify attitude of a user.

2.2 Assumptions Related to Locations

When we consider user attitudes in the user location domain, the characteristics of user attitude are related to visited locations and their distribution. Based on this consideration, we state following assumptions.

Assumption 1: Extraverts like to visit new locations; as a result they may visit many locations as compared to introverts.

Assumption 2: Extraverts are more distributed over locations and introverts are less distributed over locations; as a result more distribution of extraverts leads to travel more distance as compared with introverts.

3. Feature Extraction from Location Clustering

According to the assumptions, in order to extract quality classification features for predicting user attitude the following aspects need to be considered; number of interesting locations and distribution of interesting locations.

Firstly, we apply location clustering techniques over GPS log data. A GPS point with latitude and longitude can be mapped into a two dimension space, and it is possible to make clusters of GPS points. Each cluster can be considered as an interesting location of a user. If the *number of clusters* of a user is big, the user likes to visit many locations. In this manner, the number of clusters can be considered as a classification feature of user attitude.

Number of clusters (N_i): It is defined as the number of n participating location clusters of a user i.

$$N_i = |U_i| \tag{1}$$

$$U_i = \{L_1, L_2, L_3, ...L_n\} \tag{2}$$

$$L_j = \{P_1, P_2, P_3, ...P_m\} \tag{3}$$

$$P_k = (longitude, latitude) \tag{4}$$

where U_i denotes a set of location clusters of a user i, L_j denotes a location cluster of a user and P_k denotes a GPS point.

As a second feature, we consider coverage of cluster which means how interesting locations are distributed. To represent the coverage, the center point of a set of location clusters is considered. If the location clusters of a user are far from the center point, the user likes to travel more distance. In this manner, the *average/maximum distance covered* by a user can be considered as a classification feature.

Average/Maxumum distance covered (D_i^{avg} / D_i^{max}): It is defined as the average/maximum of distances covered by a set of location clusters of a user i.

$$D_i^{avg} = \frac{\sum_{j=1}^{n} d_i(L_j)}{N_i} \tag{5}$$

$$D_i^{max} = max(d_i(L_j)) \tag{6}$$

$$d_i(L_j) = |C_i - T_j| \tag{7}$$

$$C_i = Avg(T_j) \forall L_j \in U_i \tag{8}$$

$$T_j = Avg(P_k) \forall P_k \in L_j \tag{9}$$

where $d_i(L_j)$ denotes the Euclidian distance between the center point of a set of location clusters of a user i (C_i) and the center point of a location cluster L_j (T_j).

The distance covered by a user considers that each location cluster has same importance for a user, but actually the frequency of a location cluster is very different to other clusters. That is, during considering the distance covered by a user, we may need to consider the different importance of each location cluster. To do this, we use the number of points of a location cluster as a weighted factor for calculating *average/maximum weighted distance covered* by a user.

Average/Maximum weighted distance covered (WD_i^{avg}/ WD_i^{max}): It is defined as the average/maximum of weighted distances covered by a set of location clusters of a user i.

$$WD_i^{avg} = \frac{\sum_{j=1}^{n} wd_i(L_j)}{N_i} \tag{10}$$

$$WD_i^{max} = max(wd_i(L_j)) \tag{11}$$

$$wd_i(L_j) = |WC_i - T_j| \tag{12}$$

$$WC_i = \frac{\sum_{j=1}^{n} (\mid L_j \mid \times T_j)}{\sum_{j=1}^{n} \mid L_j \mid} \tag{13}$$

where $wd_i(L_j)$ denotes the Euclidian distance between weighted center point of a set of location clusters of a user i (WC_i) and the central point of a location cluster L_j (T_j).

Through the process of extracting features, we can obtain five features such as *number of clusters* (N_i), *average/maximum distance covered* (D_i^{avg}, D_i^{max}) and *average/maximum weighted distance covered* (WD_i^{avg}, WD_i^{max}). Among these features, *number of clusters* implies how many locations visited by a user (assumption 1); and *average/maximum distance covered* and *average/maximum weighted distance covered* implies how much distance is travelled by a user to visit some locations, to join some events or to meet some people (assumption 2).

4. Evaluation

4.1 Evaluation Setup

To evaluate the performance of predicting user personality, especially attitude, by using location clustering based features, we gathered GPS data for a month from 2013 October to 2013 November. By implementing Android based user behavior logging application and distributing it to each

participated user, we collected 30 users' smart phone usage data such as call logs, SMS logs, SNS service logs and GPS logs. GPS log data is collected with an interval of 15 minutes. Among the all of the collected GPS logs, we used only the GPS logs inside of Gwangju city area with the range of 10km x 10km; and the size of the filtered GPS logs is 100k. After applying the location clustering over the filtered GPS logs, we observed that, location clusters created by E user and I user are distributed around 1.75km and 0.75km distance over the target region respectively. We also collected MBTI-based personality type data of each user. As we discussed in Sect. 2, MBTI types are based on a set of 4 dichotomies. Table 1 explains classification with respect to each pair of dichotomy out of 30 users based on their personality type.

We applied two different location clustering techniques, DBSCAN and K-mean, to the filtered GPS logs and extract five features (number of clusters, average/maximum distance covered, average/maximum weighted distance covered). In order to evaluate the effectiveness of the extracted features, we conducted personality prediction with various classifiers such as naive Bayes, decision tree, k-nn and SVM. With these classifiers, we conducted 2-fold cross validation. We randomly assign GPS data into two sets d_1 and d_2. We then train on d_1 and test on d_2, followed by training on d_2 and testing on d_1. With 2-fold cross validation each GPS data point is used for both training and validation on each fold.

The performance of the presented methods was evaluated by using precision (P), recall (R) and F-measure. Recall measures how many of the related personalities in a collection have actually been judged as relevant. Precision measures how many of the personalities judged in analysis are actually relevant. The F-measure can be interpreted as a weighted average of the precision and recall.

4.2 Results and Analysis

Table 2 shows the performance of MBTI-based personality preferences such as E & I, S & N, F & T and J & P with different classifiers. All classifiers can classify the personality preferences attitude (E & I) and lifestyle (J & P) better than information-gathering (S & N) and decision-making (F & T). Especially, SVM classifier gives better classification results than all other classifiers. The main reason of better performance of SVM is related to the clusters of data in feature space. That is, the clusters are hard to be separated linearly and a hyperbolic plane is required to separate them.

Next, to evaluate the performance of extracted features, we classified features into three feature sets F1, F2 and F3. Feature set F1 consists of all 5 features. Feature set F2 consists of *number of clusters*, *average distance covered* and *maximum distance covered* and the feature set F3 consists of *number of clusters*, *average weighted distance covered* and *maximum weighted distance covered*. Figure 1 shows the performance of personality prediction with different feature sets by using SVM classifier. From Fig. 1 we observed that SVM performs well with the feature set F3, than with the feature sets F1 and F2. The features WD_i^{avg} and WD_i^{max} are derived from D_i^{avg} and D_i^{max} respectively. The dependency between features may degrade the performance of predicting user personality and the performance of F1 becomes worse than F2 and F3. Generally, F3 achieves the best performance, and it is because the frequency of visiting a location is an important feature.

We applied DBSCAN and K-mean over GPS logs. From Fig. 2 we can observe the effect of clustering techniques over classification of personality preferences; DBSCAN works better than K-mean. We have collected the GPS log with 15 minute time interval; as a result with K-mean clustering points are dispersed and hard to find the exact interesting location.

Through the evaluation, we observed that the features extracted from location clusters are very useful to predict user attitude. From the obtained results we can say that, we

Table 1 Classification of users with personality preferences

Personality Preferences	E	I	N	S	F	T	J	P
Number of users	17	13	23	7	22	8	20	10

Table 2 Performance of personality prediction with different classifiers

Algorithm		E	I	N	S	F	T	J	P
Naive Bayes	Precision	70	70	0	76	0	72	82	75
	Recall	82	54	0	100	0	95	90	60
	F-measure	75	60	0	86	0	82	85	67
Decision Tree	Precision	69	53	25	77	30	75	70	40
	Recall	52	69	28	73	37	68	70	40
	F-measure	60	60	27	75	33	71	70	40
k-NN (k = 1)	Precision	68	67	28	80	09	64	72	50
	Recall	81	50	33	77	14	52	84	33
	F-measure	74	57	30	79	11	57	78	40
SVM	Precision	89	100	100	92	100	84	100	100
	Recall	100	84	71	100	50	100	100	100
	F-measure	94	91	83	95	67	91	100	100

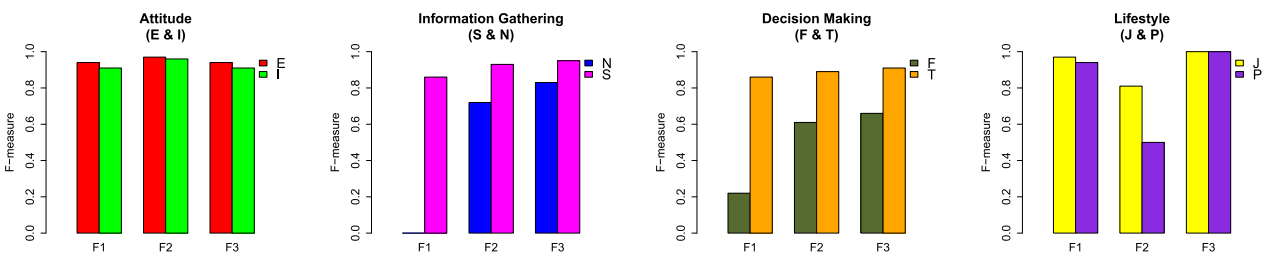

Fig. 1 Performance of personality prediction with different feature sets using SVM classifier

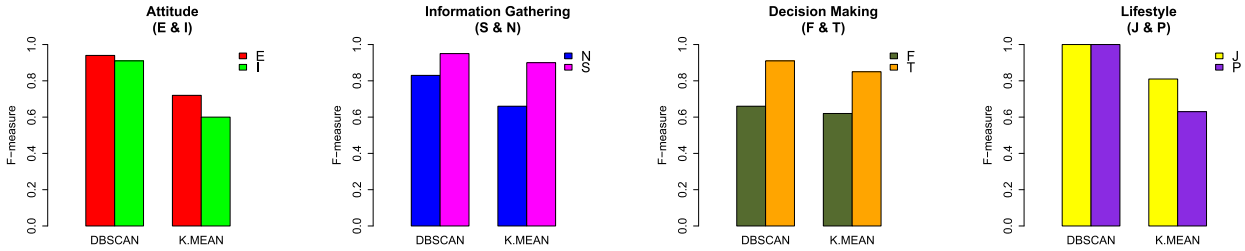

Fig. 2 Performance of personality prediction with different clustering techniques using SVM classifier and F3 feature set

can use location clustering and extracted features for predicting user attitude and other personalities to some extent.

5. Related Work

The field of 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 [5]. Presently, predicting personality traits based on smart phone usage data accesses call logs, SMS logs, GPS logs, social networking services logs, use of internet, use of Bluetooth and battery [2], [4], [6]. In a previous work, authors have analyzed the relationship between smart phone usage and self-perceived personality with the help of applications usage logs, call logs and SMS logs. Their feature set was enriched with features extracted from call data, SMS data [2]. Oliveira et al. [4] suggested 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.

Recently, some studies have tried to use GPS data with user personality traits [6], [9]. One of them focused on how to use GPS logs to identify personal mobility pattern with the collaboration of big five user personality [9]. In another work, authors use GPS logs to predict big five user personality [6]. In this work, GPS logs are used to trace out the location of call (radius of gyration, number of places from which calls have been made) as a classification feature. Unlike these previous researches, we mainly focus on how to use location clustering to extract more effective classification features for predicting user attitude based on MBTI theory.

6. Conclusion

A smart phone with various sensors can observe user behavior, and the observed behavior can be used for predicting user attitude. In this paper, a new method is proposed for extracting quality classification features by using location clusters with the observed user GPS data. Through the extensive evaluation based on real user dataset, we show that features extracted by GPS location clusters are effective to predict user personality, especially the attitude and the lifestyle dichotomy of MBTI. The research for extracting features from other sensor data is a natural extension of this work.

Acknowledgments

This work was supported by the National Research Foundation of Korea Grant funded by the korean Government (NRF-2014R1A1A1007734).

References

[1] I.B. Myers and M.H. McCaulley, "Myers-Briggs Type Indicator: MBTI," Consulting Psychologists Press, 1988.
[2] G. Chittaranjan, J. Blom, and D. Gatica-Perez, "Who's who with Big-Five: Analyzing and Classifying personality traits with smartphones," Proc. IEEE ISWC 2011, pp.29–36, 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, vol.35, no.1, pp.41–64, 2004.
[4] R. de Oliveira, A. Karatzoglou, P.C. Cerezo, A.A. Lopez de Vicuña, and N. Oliver, "Towards a psychographic user model from mobile phone usage," Proc. ACM SIGCHI 2011, pp.2191–2196, 2011.
[5] K.M. Lee and C. Nass, "Designing social presence of social factors in human computer interaction," Proc. ACM SIGCHI 2003, pp.289–296, 2003.
[6] Y.-A. de Montjoye, J. Quoidbach, F. Robic, and A.S. Pentland, "Predicting personality using novel mobile phone-based metrics," Social Computing, Behavioral-Cultural Modeling and Prediction, vol.7812, pp.48–55, 2013.
[7] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Proc. ACM KDD, pp.226–231, 1996.
[8] S. Seneviratne, A. Seneviratne, P. Mohapatra, and A. Mahanti, "Predicting User Traits from a Snapshot of Apps Installed on a Smartphone," SIGMOBILE Mob. Comput. Commun. Rev., vol.18, no.2, pp.1–8, 2014.
[9] S. Kim, H.Y. Song, and H.J. Koo, "Probabilistically Predicting Location of Human with Psychological Factors," Proc. ECCS 2012.