Proceedings of the Federated Conference on Computer Science and Information Systems pp. 743–746 DOI: 10.15439/2019F319 ISSN 2300-5963 ACSIS, Vol. 18

Analysis of the Correlation Between Personal Factors and Visiting Locations With Boosting Technique

Ha Yoon Song
Department of Computer Engineering
Hongik Unversity, Seoul, Republic of Korea
Email: hayoon@hongik.ac.kr

JiSeon Yun
Department of Computer Engineering,
Hongik University, Seoul, Republic of Korea
Email: wltjs9214@gmail.com

***Abstract*—The paper analyzed the relationship between the person's fourteen characteristic factors and place to visit. The personal factors consist of personality, marital Status, final education, majors, religion, monthly income, commuting means and time, frequency of travel, userage of social media, time spent on social media per day, cultural type. In addition, the analysis was done on which factors have the greatest impact. The analysis involved thirty-four participants and the boosting technique was used as a method of analysis.**

I. Introduction

RECENTLY, A number of services provides useful information to people by predicting their moving pattern and location data, especially for Location Based Service (LBS). However, most of the studies predicting people's movements focus on analyzing past patterns of movement. Apart from this prediction method, we conducted another research on a relationship where a person visits with person's various factors [1] [2]. Factors such as a personality, marital status, and final education and so on clearly affect a person's favorite place to visit. In this study, the correlation between person's characteristic factors and place to visit are analyzed using Boosting techniques. In addition, the analysis of the greatest influential factors to location visit is also addressed. Section II will describe the Boosting technique to be used for correlation analysis. Section III will describe the person's characteristic data and location data used in the analysis. Section IV analyzes which factors have the greatest impact. Section V will describe the conclusions of this study and the future direction of study.

II. Boosting as an Analysis Method

A. Boosting

The analysis technique used for this study is Boosting, one of the ensembles techniques [3]. Boosting is one of the techniques of generating a number of classifiers by manipulating initial sample data similar to Bagging, but the biggest difference is that Boosting is a sequential method. Boosting is a technique to train several weak learners sequentially, to learn weight by adding weight to mispredicted data, and to predict using the finally generated learner. That is, the results of the previous learning will affect the next learning.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST)(NRF-2017R1D1A1B03029788).

TABLE I
Personality data of 5 volunteers using BFF

	O	C	E	A	N
Person1	3.3	3.9	3.3	3.7	2.6
Person2	2.7	3.2	3.2	2.7	2.8
Person3	4.3	3.1	2.3	3.2	2.9
Person4	4.2	4.3	3.5	3.6	2.6
Person5	4.0	3.7	4.0	3.9	2.8

B. XGBoost

There are a number of Boosting algorithms. In this research, we will use XGBoost boosting algorithm. XGBoost is an algorithm that visualizes how much the model relies on which factors [4] [5]. It also offers a variety of custom optimization options, including evaluation functions for flexibility. Therefore, it was appropriate to analyze which factors have the greatest impact on place to visit.

III. Input Data

A. Personality Data

Personality data was digitized into five personality types in the Big Five Factor (BFF). BFF was developed by psychologists P. T. Costa and R. McCrae in 1976 and is a personality psychological model that explains human personality in terms of five mutually independent factors [6] [7] [8] [9] [10] [11] [12]. O is Openness, C is Conscientiousness, E is Extroversion, A is Agreeableness, and N is Neuroticism. Table I is personality data of 5 volunteers using BFF.

B. Other Personal Factors

The person's factor without personality were collected through a questionnaire made directly by Google Form and quantified the categories for each factor. Table II is the person's characteristic factors without personality of four volunteers obtained from the questionnaire. Age refers to age, with 1 in the teens, 2 in the 20s, 3 in the 30s and 4 in the 40s and older. Job represents a job and has been assigned a category by adding 'students' to the International Classification of Work (ISCO) standard [13]. 1 is for students, 2 is for managers, 3 is for technical workers, 5 is for office workers, 6 is for service and sales, 7 is for functional workers, 8 is for device and machine operation, and 9 is for simple labor workers. Marriage indicates marital status, 1 is married and 2 is unmarried. Edu represents final education, 1 is below high school graduation, 2 is a high school graduate, 3 is a university graduate, 4 is master's degree and 5 is doctoral degree. Major represents the major, 1 is the humanities, 2 is the sociality, 3 is the educational, 4 is the engineering, 5 is the natural science, 6 is the medicine and 7 is the art. Religion represents religion, 1 is Atheist, 2 is Christianity, 3 is Catholicism (the Catholic Church), and 4 is Buddhism. Salary represents monthly income, with 1 being less than 500,000 won, 2 being less than 1 million won, 3 being more than 1 million won, 4 being more than 2 million won and 4 being less than 3 million won and 5 being more than 3 million won. Vehicle indicates means of commuting, 1 is walking, 2 is cycling, 3 is using self-driving, and 4 is public transportation. Comm T indicates commuting time, 1 is within 30 minutes, 2 is less than one hour, 3 is less than one hour, and 4 is more than two hours. Travel indicates the frequency of travel, 1 is less than 1 time, 2 is less than 4 times, 3 is less than 4 times, and 4 is more than 6 times. Social M indicates usage of social media, 1 is on social media, and 2 is not on social media. Social M2 represents the daily usage of social media, 1 is less than 30 minutes, 2 is less than 1 hour for 30 minutes, 3 is less than 1 hour and 4 is more than 3 hours. Finally, Culture represents cultural type, 1 corresponds to a mixture of static activity, 2 to dynamic activity, and 3 to both static and dynamic activities.

TABLE II
PERSON'S FEATURE DATA FROM THE QUESTIONNAIRE

	Volunteer1	Volunteer2	Volunteer3	Volunteer4
Age	2	2	3	2
Job	1	1	3	1
Marriage	2	2	2	2
Edu	2	2	4	4
Major	4	4	4	4
Religion	1	3	2	4
Salary	2	2	5	2
Vehicles	4	4	2	4
Comm T	3	3	2	3
Travel	2	2	2	3
Social M	1	2	2	1
Social M2	3	0	0	2
Culture	3	3	2	2

C. Location Categories

The SWARM application was used to collect location data. SWARM is an application that records the location of a visit when a user visits a site. Location data was created by categorizing each visit data to ten industry classification and accumulating number of visits for each category [14] [15]. Ten Industry categories include Foreign Institutions, Retail, Service industry, etc. Finally, Location data is obtained by calculating the ratio of number of visits of each category compared to the total number of visits. Table III is part of location data of four volunteers.

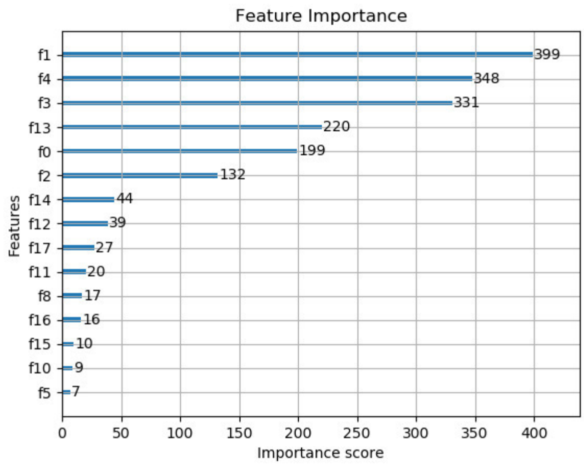

Fig. 1. Feature Importance Graph for Location Categories

IV. EXPERIMENTAL RESULT

We used XGBoost mentioned in section II as an analysis technique. An independent variable is a person's characteristic data, which was created by merging personality data obtained using BFF, and data for the rest of the factors obtained through a questionnaire. Table IV is characteristic data of three volunteers. Dependent variable is location data. A regression model was created by inserting dependent variable and independent variables into `XGBRegressor()` in XGBoost. Then, ran `feature_importances` on this regression model and found what factors among the various characteristics of person including personality are most effective for location data.

Figs. 1 show the result of performing feature importance analysis using XGBoost for Foreign Institutions. The y-axis (Features) represents each factor included in the person's characteristic factors. The x-axis (Importance) represents the effectiveness of the independent variable for the dependent variable. Labels f0 through f17 are in the order of the factors listed in Table IV. For example, in Fig. 1, Foreign Institutions of Feature Importance shows that f1 (C, Conscientiousness) has the greatest impact on location data classified as Foreign institutions.

TABLE III
SAMPLE LOCATION CATEGORY VISITING RATES OF FOUR VOLUNTEERS

	Volunteer1	Volunteer2	Volunteer3	Volunteer4
Foreign Institutions	0.01705	0.00551	0.13559	0.25833
Retail	0.05634	0.67250	0.04237	0.01667
Service industry	0.02965	0.00162	0.02260	0.00333
Restaurant	0.19496	0.07620	0.40960	0.15167
Pub	0.02743	0.01232	0.00847	0.02000
Cafe	0.19422	0.07847	0.07910	0.06167
Cinema	0.01705	0.00551	0.00565	0.01000
Educational institution	0.43662	0.14008	0.27401	0.47333
Hospital	0.00741	0.00292	0.02260	0.00000
Historic sites	0.01927	0.00486	0.00000	0.00500

V. CONCLUSION

In this study, we analyzed the correlation between various factors of people and place to visit through boosting. As a result, we were able to see how each characteristic of a person affects each place visit. However, there are many similar results for each of the ten place data, and the accuracy was not high. Therefore, we analyzed the reason in many ways. Firstly, many biased results were obtained because most of the volunteer were students in the process of collecting data. Therefore, in the next study, we will recruit the volunteer by various occupations and ages. Secondly, the number of volunteers was few, and the data of place to visit were also insufficient. This is because the volunteer does not use the SWARM application properly. SWARM does not automatically collect the places visited, but it is inconvenient because user has to check-in themselves actively. Therefore, in the next study, we will recruit more volunteers and make detailed guidance on how to collect data with SWARM. Thirdly, there are several parameters when generating the XGBoost predictive model. When using XGBoost, tuning hyperparameters means that they are the most essential and important. There might be more way to tune XGBoost parameters for future research. Lastly, the accuracy is not great because most ambiguous places to sort are put into service industry or Historic sites in the process of applying the visited places to the industry classification. Therefore, location category classification should be improved by other than current industry classification standards.

TABLE IV
PERSONAL FACTORS OF THREE VOLUNTEERS

	Volunteer1	**Volunteer2**	**Volunteer3**
O (f0)	3.3	2.7	4.3
C (f1)	3.9	3.2	3.1
E (f2)	3.3	3.2	2.3
A (f3)	3.7	2.7	3.2
N (f4)	2.6	2.8	2.9
Age (f5)	2	2	3
Job (f6)	1	1	3
Marriage (f7)	2	2	2
Edu (f8)	2	2	4
Major (f9)	4	4	4
Religion (f10)	1	3	2
Salary (f11)	2	2	5
Vehicles (f12)	4	4	2
Comm T (f13)	3	3	2
Travel (f14)	2	2	2
Social M (f15)	1	2	2
Social M2 (f16)	3	0	0
Culture (f17)	3	3	2

Location-based services (LBS) is one of the emerging issues that have great potential for future service. In particular, understanding human mobility patterns is a key part of LBS. We can analyze human mobility patterns by using the correlations between various factors of people and visiting locations analyzed in this study. Therefore, this analysis result might be extended and can be utilized in LBS. It is also expected to be useful for recommendation systems. A recommendation system is a kind of information filtering technology that recommends information that might be of interest to a specific user, such as video recommendations of Netflix and YouTube. People with specific factors will be able to correlate the frequent visits to specific places and apply them to the recommendation system.

Location-based services (LBS) is one of the emerging issues that have great potential for future service. In particular, understanding human mobility patterns is a key part of LBS. We can analyze human mobility patterns by using the correlations between various factors of people and visiting locations analyzed in this study. It is also expected to be useful for recommendation systems. A recommendation system is a kind of information filtering technology that recommends information that might be of interest to a specific user. People with specific factors will be able to correlate the frequent visits to specific places and apply them to the recommendation system.

REFERENCES

[1] S. Y. Kim and H. Y. Song, "Predicting human location based on human personality," *International Conference on Next Generation Wired/Wireless Networking*, 2014. doi: https://doi.org/10.1007/978-3-319-10353-2-7

TABLE V
FEATURE IMPORTANCE OF EACH LOCATION CATEGORIES

	feature 1	feature 2	feature 3	feature 4	feature 5
Foreign Institutions	C	N	A	Comm T	O
	0.219472	0.191419	0.182068	0.121012	0.109461
Retail	O	N	Travel	E	Social M2
	0.243421	0.154605	0.115132	0.108553	0.105263
Service industry	C	O	Social M2	Edu	Comm T
	0.250000	0.180147	0.161765	0.147059	0.088235
Restaurant	O	A	Vehicles	Salary	E
	0.255132	0.184751	0.114370	0.102639	0.099707
Pub	O	Salary	C	N	A
	0.367742	0.258065	0.129032	0.109677	0.064516
Cafe	O	C	N	Salary	E
	0.255435	0.217391	0.125000	0.089674	0.084239
Cinema	C	N	A	Vehicles	Religion
	0.243590	0.153846	0.123932	0.085470	0.085470
Educational institution	O	C	Salary	Edu	N
	0.215827	0.165468	0.158273	0.146283	0.083933
Hospital	O	Salary	C	Travel	A
	0.247664	0.219626	0.140187	0.126168	0.079439
Historic sites	O	Salary	N	C	Comm T
	0.219780	0.179487	0.175824	0.157509	0.131868

[2] M. J. Chorley, R. M. Whitaker, and S. M. Allen, "Personality and location-based social networks," *Computers in Human Behavior*, vol. 46, pp. 45 – 56, 2015. doi: https://doi.org/10.1016/j.chb.2014.12.038. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563214007559

[3] F. Schapire, Robert E, *Boosting: Foundations and Algorithms*. MIT Press (MA), 2014.

[4] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," *CoRR*, vol. abs/1603.02754, 2016. doi: 10.1145/2939672.2939785. [Online]. Available: http://arxiv.org/abs/1603.02754

[5] "Xgboost," https://xgboost.readthedocs.io/en/latest/index.html, accessed: 2019-01-15.

[6] P. T. Costa and R. R. McCrae, "Four ways five factors are basic," *Personality and Individual Differences*, vol. 13, no. 6, pp. 653 – 665, 1992. doi: https://doi.org/10.1016/0191-8869(92)90236-I. [Online]. Available: http://www.sciencedirect.com/science/article/pii/019188699290236I

[7] J. Hoseinifar, M. M. Siedkalan, S. R. Zirak, M. Nowrozi, A. Shaker, E. Meamar, and E. Ghaderi, "An investigation of the relation between creativity and five factors of personality in students," *Procedia - Social and Behavioral Sciences*, vol. 30, pp. 2037 – 2041, 2011. doi: https://doi.org/10.1016/j.sbspro.2011.10.394 2nd World Conference on Psychology, Counselling and Guidance - 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877042811022191

[8] D. Jani, J.-H. Jang, and Y.-H. Hwang, "Big five factors of personality and tourists' internet search behavior," *Asia Pacific Journal of Tourism Research*, vol. 19, 05 2014. doi: 10.1080/10941665.2013.773922

[9] D. Jani and H. Han, "Personality, social comparison, consumption emotions, satisfaction, and behavioral intentions: How do these and other factors relate in a hotel setting?" *International Journal of Contemporary Hospitality Management*, vol. 25, no. 7, pp. 970–993, 2013. doi: 10.1108/IJCHM-10-2012-0183. [Online]. Available: https://doi.org/10.1108/IJCHM-10-2012-0183

[10] O. P. John and S. Srivastava, "The big-five trait taxonomy: History, measurement, and theoretical perspectives," 1999.

[11] L. R. Goldberg, "'the structure of phenotypic personality traits": Author's reactions to the six comments." *American Psychologist*, vol. 48, pp. 1303–1304, 12 1993. doi: 10.1037/0003-066X.48.12.1303

[12] Y. Amichai-Hamburger and G. Vinitzky, "Social network use and personality," *Computers in Human Behavior*, vol. 26, no. 6, pp. 1289 – 1295, 2010. doi: https://doi.org/10.1016/j.chb.2010.03.018 Online Interactivity: Role of Technology in Behavior Change. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563210000580

[13] "International standard classification of occupation (isco)," https://www.ilo.org/, accessed: 2018-12-20.

[14] E. B. Lee and H. Y. Song, "An analysis of the relationship between human personality and favored location," *The Seventh International Conference on Advances in Future Internet*, 2015.

[15] Song, Ha Yoon and Kang, Hwa Baek, "Analysis of relationship between personality and favorite places with poisson regression analysis," *ITM Web Conf.*, vol. 16, p. 02001, 2018. doi: 10.1051/itmconf/20181602001. [Online]. Available: https://doi.org/10.1051/itmconf/20181602001