

PREDICTING A DRIVER'S PERSONALITY FROM DAILY DRIVING BEHAVIOR

Yuichi Ishikawa, Akihiro Kobayashi, Atsunori Minamikawa, & Chihiro Ono
KDDI Research, Inc.
Saitama, Japan
Email: yi-ishikawa@kddi-research.com

Summary: A wealth of literature has shown the predictive and preventive utility of the Big Five personality traits model (BIG5) for various kinds of unsafe driving. However, the commonly used method for BIG5 measurement requires subjects to answer long and stressful questionnaires, making its applicability limited. In this paper, we study the potential for predicting a driver's BIG5 traits from his/her daily driving behavior. We collected naturalistic driving data on (A) car usage behavior (driving frequency, distance, duration, etc.) and (B) driving operation behavior (operation of steering wheel, accelerator and brake pedal, etc.) from 140 Japanese subjects over two months. By analyzing the data while focusing on various specific driving conditions, we were able to find features which significantly correlate with BIG5 traits from both (A) and (B). In the evaluation, the features we found predicted whether the traits scores are above $\mu + \sigma$ or below $\mu - \sigma$ (μ : average, σ : standard deviation) at an accuracy of ROC-AUC 0.62~0.85, confirming the potential for predicting BIG5 traits from daily driving behavior.

INTRODUCTION

A wealth of literature has shown that, in addition to demographic attributes such as gender and age, drivers' psychographic attributes play a significant role in predicting and preventing various kinds of unsafe driving including risky driving, loss of vehicular control, aggressive driving (Dahlen et al., 2006), driving accident involvement (Arthur et al., 1996) and interacting with a mobile phone while driving (Parr et al., 2016). Among several psychometric instruments, one of the most widely used instruments in current research is the Big Five personality traits model (BIG5) (Norman, 1963), which quantifies personality based on the scores for five traits: Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N).

The most common approach to measuring BIG5 traits has been to use a questionnaire consisting of multiple questions (typically 30-300 questions) in which the subject is asked about his/her thoughts, beliefs, norm and so on (Widiger., 2017). However, relying entirely on the above method for BIG5 measurement has limited its potential, since answering such questions takes considerable time and concentration and imposes strain on subjects, making its applicability limited to people with special incentives (e.g., monetary rewards, passing an aptitude test for a license, etc.) and, at the same time, occasionally leading to distortion of a subject's answers. Therefore, to widen its applicability and avoid distortion, it is preferable not to rely on the conventional approach entirely, but to predict it from the driver's daily behavior.

With this motivation, this paper studies the potential for predicting a driver's BIG5 traits from his/her driving behavior. The study was conducted in 3 steps. First, we collected naturalistic

driving data on (A) car usage behavior (driving frequency, distance, duration, etc.) and (B) driving operation behavior (operation of steering wheel, accelerator and brake pedal, car speed, etc.) from 140 Japanese subjects over two months and also measured their BIG5 traits scores via the questionnaire. Second, we analyzed the correlation between driving behavior and BIG5 traits to find out features that significantly correlated with BIG5 traits from (A) and (B). The analysis was conducted assuming that (A) and (B) correlate with BIG5 traits to different extents depending on driving conditions such as traffic environment (congested or not, etc.), purpose of driving (commuting, recreation, shopping, etc.) and driving situation (left/right turn, slow speed, etc.). Based on this assumption, we analyzed the correlation focusing on various specific driving conditions. Finally, we conducted prediction of BIG5 traits scores using the features found in the second step, and evaluated prediction accuracy.

METHODS

Participants

We recruited 140 Japanese participants who committed to drive his/her car for more than 30km and 2 hours during the period of data collection (Feb. ~ Apr., 2018) and owned a car of the specific model designated. By doing this, we intended to exclude difference of driving operation behavior derived from different car models so that we can focus on correlation between driving operation behavior and participant's personality as much as possible. After recruitment, we measured their BIG5 traits scores by using the Japanese version of NEO-PI-R (Shimonaka et al., 1998) followed by data collection of daily driving behavior. Table 1 and figure 1 show the demographic and psychographic distribution of participants, respectively.

Table 1. Demographic distribution of participants

Male					Female				
Total	~39 y/o	40~49 y/o	50~59 y/o	60 y/o~	Total	~39 y/o	40~49 y/o	50~59 y/o	60 y/o~
109	10	35	33	31	31	9	9	10	3

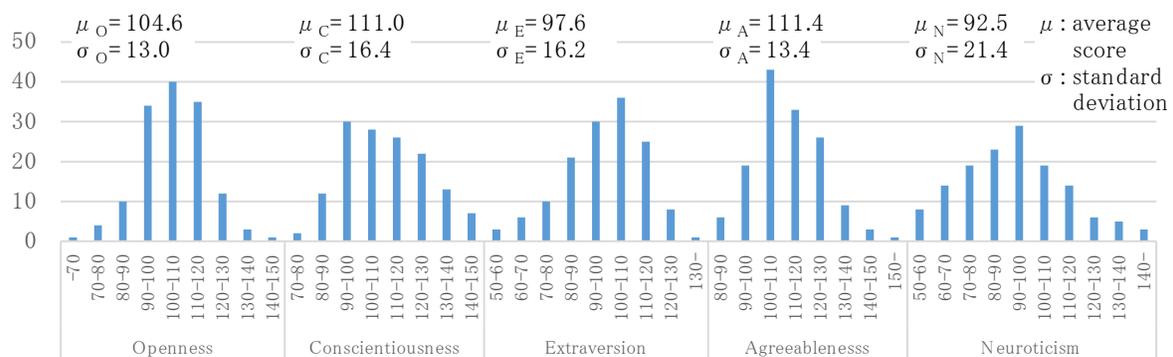


Figure 1. Psychographic distribution of participants (x-axis: trait score, y-axis: number of participants)

Collection of Driving Behavior Data

We provided all participants with devices which read CAN (Controller Area Network) data from the cars via an OBD (On-Board Diagnostics) interface and kept track of location and acceleration by using their own GPS receivers and accelerometers. Using the device, we collected the number

of hours and distance driven for each drive (from engine start to stop) and the number of times the car was driven to establish car usage behavior. For driving operation behavior, we collected the following data at 5Hz frequency: steering wheel angle, degree to which the brake and accelerator pedals were depressed, car speed and 3-axis acceleration. Participants were asked to have the devices attached to their cars at all times during the period of data collection.

Analysis Approach

As for car usage behavior, we assumed that the extent to which BIG5 traits affect car usage behavior varies with traffic environment and purpose of driving. For example, to commute, highly conscientious people would not use a car but take public transportation when a traffic jam is expected. On the other hand, for recreation or when a traffic jam is not expected, his/her conscientiousness would not affect whether to use a car so much. Based on this assumption, we extracted and analyzed the data for each day of the week (weekday/weekend) and time of day (morning/afternoon/night) in expectation that we could analyze car usage behavior in different traffic conditions and purposes of driving (e.g., On weekday mornings, traffic jams are highly likely and commuting is a common purpose of driving, whereas, on a weekend afternoon, congestion is not highly likely and the probable purpose of driving is recreation or shopping.).

Regarding driving operation behavior, motivated by existing research that argued that each BIG5 trait correlates with different kinds of unsafe driving (Dahlen et al., 2006), we assumed that BIG5 traits affect driving operation behavior to different extents depending on driving situations such as right/left turn, very slow/slow speed and reversing. Based on this assumption, we extracted and analyzed driving operation data for each driving situation.

RESULTS

Relationships between Big Five and Daily Driving Behavior

Car Usage Behavior. Correlation with car usage behavior was examined by looking at the average and variance of (1) drive count, (2) hours and (3) distance for each day of the week and time of day. We also looked at (4) drive periodicity and (5) tendency to drive long distances for each day of the week. (4) is the autocorrelation of (1)~(3) and (5) is the drive count percentage for different distance intervals (~10km, 10k~50km, 50~100km, 100km~). (1) ~ (5) were also examined for all days and all times. Table 2 presents the results showing correlation coefficients between BIG5 traits scores and (1) ~ (5) ($p < 0.05$).

From the results, though the coefficients are small ($|r|$ is around 0.2), all BIG5 traits scores significantly correlate with the above statistics. When examining car usage behavior for all-day and all-time, significant correlations were found only in E-(5) and N-(5), whereas all the other significant correlations were found when examined for each day of the week and time of day.

Among BIG5 traits, a stronger correlation exists for O, for which (1) ~ (5) are significantly correlated, and coefficients are larger than for the other traits ($|r|$ of O is 0.22 on average, 0.04 larger than the average for the other traits). We interpret this as follows; car usage behavior is affected by O more strongly than by the other traits, especially by facets of O including “intellect” (ability to employ skills, knowledge and experience) and “curiosity” (tendency to

engage in a process of trial and error). Specifically, for those with a high O score, we consider that the trait affects car usage behavior as follows:

- Drive less frequently, especially on weekday mornings (O- (1)), and lower periodicity for drive count (O- (4)): reflecting a high degree of intellect, participant changes means of transportation including use of own car depending on the degree of traffic congestion. Since traffic congestion occurs more frequently on weekday mornings, the participant is more likely to select public transport instead of driving own car.
- Larger variance of distance and hours of driving on weekday mornings (O- (2), (3)): reflecting a high degree of curiosity, participant changes route from day to day when commuting by car in order to find a better route.

Table 2. Correlation of BIG5 traits with car usage behavior

	O	C	E	A	N
(1)Count	WD-M, avg: -.24 WE-M, avg: -.20 WE-M, var: -.18 WD-N, var: -.21	-	WD-M, var: -.19	-	-
(2)Distance	WD-M, var: .23 WD-A, var: .22	WD-A, avg: .17	WE-A, var: .17	WE-N, var:0.17	-
(3)Hours	WD-M, var: .24	WD-A, avg: .21	-	-	-
(4)Periodicity	WD, Count: -.27 WD, Hours: -.19	-	-	-	-
(5)Distance Rate	WD, >100km: .22	-	AD, <10km: -.17	-	AD, <10km: .18

AD/WD/WE: all day/weekday/weekend, M/A/N: morning/afternoon/night

Note: Coefficients greater than $|0.2|$ are shown in bold

Driving Operation Behavior. We extracted the following driving operation data for each driving situation: right/left turn - data collected from a road section 100m long on either side of the center of an intersection at which a car makes a right/left turn (the intersection is detected by matching the car's location with a map), very slow/slow speed- data collected when car speed is in the range of 4~20/20~40 km/h and stays in that range for over 30/60 seconds, reverse - data collected when gear shift position is in reverse for more than 15 seconds.

To calculate correlation coefficients, we utilized the same approach as described by Enev et al. (2016) for feature extraction; namely, we divided driving operation data into overlapping sliding windows and calculated the following features for each window: average, median, quartile, maximum, minimum and standard deviation. Afterward, we calculated correlation coefficients between BIG5 traits scores and features of windows. Several different values were tested for window size (3~60 seconds) and window overlapping percentage (10-30%), and we adopted the value which outputs the maximum coefficient (in absolute value) with BIG5 traits scores.

Table 3 shows the results ($p < 0.01$). For comparison, the table also presents the correlation with all driving operation behavior regardless of driving situations ("ALL" row in Table 3). Based on these results, it was found that BIG5 traits do not affect driving operation behavior equally regardless of the driving situation, but affect behavior to a different degree depending on the driving situation. When examining driving operation regardless of the driving situation, no BIG5 trait except for C correlates significantly (refer to "ALL" row). On the other hand, if we examine particular driving situations, all the traits correlate significantly with the features extracted from driving operation. In addition, the extent to which each trait affects driving operation varies by

driving situations. For example, driving operation behavior in relation to right/left turn is strongly affected by O compared to the other traits, whereas the behavior at very slow speed is strongly affected by E and A.

For driving situations with which BIG5 traits have a relatively strong correlation ($|r| > 0.2$), we consider BIG5 traits affect driving operation behavior as follows:

- E and A - Very slow/slow speed: In this driving situation, following a vehicle in front is the most likely assumption (e.g., driving in a traffic jam). Since both E and A include facets which are related to a person's attitude toward others (e.g., "assertiveness" facet of E and "sympathy" facet of A), we consider that they affect how a driver controls the distance of and fluctuations in the inter-vehicle gap.
- O - Right/left turn: Since a driver needs to pay attention to many objects in this situation such as traffic signals, crossing pedestrians, oncoming vehicles and bikes/motorcycles passing by, we consider that the "intellect" facet of O affects how smoothly a driver makes a turn.

Table 3. Correlation of BIG5 traits with driving operation behavior

Driving Situation	O	C	E	A	N
RT	acc-z, med: -.23	-	-	-	-
LT	acc-z, 1Q: -.20	-	-	-	-
VS	-	-	AP, min: -.21 acc-z, std: -.17	acc-z, avg: -.15	-
SS	acc-x, avg: -.19	-	BP, min: .21	acc-x, max: .21	acc-z, std: -.16
RV	speed, avg: -.19 acc-z, max: -.15	acc-x, med: -.17	AP, med: -.15	acc-x, max: .17	
ALL	-	acc-z, 1Q: .17	-	-	-

AP/BP: Degree to which accelerator pedal/brake pedal depressed, SW: Steering wheel angle

RT: right turn, LT: left turn, VS: very slow speed, SS: slow speed, RV: reverse, ALL: all driving situations

Note: Coefficients greater than $|0.2|$ are shown in bold, and those less than $|0.15|$ are not shown

Predicting BIG5 from Daily Driving Behavior

We evaluated BIG5 prediction accuracy using car usage behavior and driving operation behavior respectively. For each BIG5 trait, we conducted two kinds of binary classification based on the participant's BIG5 trait score, one of which was to classify whether the score is "high"/"not high" and the other was for "low"/"not low" (i.e., 10 different classifications in total). Criteria for "high" and "low" were set at $\mu + \sigma$ and $\mu - \sigma$, respectively (μ : average, σ : standard deviation).

For each classification, we created a group of 16 participants consisting of 8 participants randomly selected from the "high"/"low" group and the other 8 randomly selected from the "not high"/"not low" group, out of which 12 were selected (6 from each) for training and the remaining 4 were tested (2 from each). In a group, we conducted training and testing 25 times by changing the participants being trained and tested so that the same set of participants would not be used for testing or training more than once.

We used a simple machine learning technique, Random Forest, by which training and testing were carried out as follows.

(A) Car usage behavior: For each participant, we calculated the statistics of car usage behavior in the same way as described in the previous subsection and used them as features for training and testing.

(B) Driving operation behavior: Training and testing were conducted for each driving situation. From each participant, we randomly extracted 10 datasets for each driving situation (e.g., 10 left turns, 10 right turns, etc.). Then, we applied a sliding window to datasets and calculated features for each window in the same way as in the previous subsection. To train a classifier, we used features of windows and the score label of a participant (i.e., “high”/“not high” or “low”/“not low”) as the training data. When testing, the classifier calculated the probabilities of score labels for windows, and output the classification result based on the average of their probabilities.

Prediction accuracy was evaluated by ROC-AUC. Table 4 shows the results, where results with an accuracy ROC-AUC < 0.5 (i.e., lower than random prediction) are omitted.

Table 4. BIG5 prediction accuracy in ROC-AUC

Classification Type		O		C		E		A		N	
		H	L	H	L	H	L	H	L	H	L
Car Usage Behavior		0.72	0.61	-	0.64	0.53	-	-	0.85	0.74	0.74
Driving Operation Behavior	RT	0.63	0.67	0.51	0.56	0.53	0.58	0.67	0.51	0.55	0.54
	LT	-	0.67	0.58	0.56	0.53	-	0.60	-	-	0.51
	VS	0.58	0.74	0.57	0.66	0.53	0.62	0.53	-	-	-
	SS	0.50	0.68	-	-	0.63	-	0.57	0.73	-	-
	RV	-	0.62	0.63	0.59	0.51	0.54	-	-	-	-
	ALL	0.51	0.72	0.70	0.64	-	-	-	0.51	0.58	-
Best Accuracy		0.72	0.74	0.70	0.66	0.63	0.62	0.67	0.85	0.74	0.74
Avg		0.73		0.68		0.62		0.76		0.74	

H: “high”/“not high”, L: “low”/“not low”

RT: right turn, LT: left turn, VS: very slow speed, SS: slow speed, RV: reverse, ALL: all driving situations

Note: Results in bold indicate the best accuracy for each trait predicted from driving operation behavior

BIG5 score labels were predicted at an accuracy of ROC-AUC 0.62~0.85 at best (refer to the “Best Accuracy” row in the table). In particular, O, A and N were predicted at a higher accuracy than the other traits (ROC-AUC > 0.7 at “Avg” row). Regarding prediction from driving operation behavior, except for classifications of C-“high”/“not high” and N-“high”/“not high”, prediction accuracies are higher when using driving operation data of a particular driving situation compared to using all the data regardless of driving situation (refer to the results in bold). This result suggests that, for better prediction accuracy, driving operation data need to be extracted from particular driving situations.

CONCLUSION

The contribution of this paper is twofold. First, through investigation into the daily driving behavior of 140 drivers, we verified the potential of daily driving behavior for predicting a driver's BIG5 traits. Second, we evaluated the approach for driving behavior analysis which focuses on particular driving conditions including traffic environment, purpose of driving and driving situation, and confirmed that the approach improves the prediction accuracy. In conclusion, it is our belief that the BIG5 prediction approach studied in this paper has the potential to widen the applicability of BIG5 to the prediction and prevention of unsafe driving.

In comparison with related work, existing research on the relationship between BIG5 traits and driving behavior only focused on unidirectional relations, predicting the risk of unsafe driving behavior from BIG5 traits. To the best of our knowledge, no study has been conducted to examine the opposite directional relation, namely, predicting BIG5 traits from driving behavior, which is covered by the present study for the first time. In addition, the scope of existing research has been focused on unusual unsafe driving behavior, and the relationship between usual daily driving behavior and BIG5 traits had not been well studied until this paper.

There has also been research to predict BIG5 traits from daily behavior other than driving behavior such as activities in SNS (Azucar et al., 2018) and usage patterns of mobile phones (de Montjoye et al., 2013). However, the former approach can only be applied to people who are active in SNS (e.g., upload messages or photos sufficiently enough for prediction). Although the latter approach is widely applicable, its prediction accuracy varies in relation to the BIG5 traits and is relatively low for O, C and A, which are as important as the other traits to predict and prevent unsafe driving. In contrast, our approach predicts O, A and N with relatively high accuracy and has the potential to complement the latter approach in the sense that it can improve the prediction accuracy of O and A when used in combination.

REFERENCES

- Arthur Jr, W., & Graziano, W. G. (1996). The five trait model, conscientiousness, and driving accident involvement. *Journal of Personality*, 64(3), 593-618.
- Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, 124, 150-159.
- Dahlen, Eric R. & White, Roy P. (2006). The Big Five traits, sensation seeking, and driving anger in the prediction of unsafe driving. *Personality and Individual Differences*, 41(5), 903-915.
- Enev, M., Takakuwa, A., Koscher, K., & Kohno, T. (2016). Automobile driver fingerprinting. *Proceedings on Privacy Enhancing Technologies*, 2016(1), 34-50.
- de Montjoye, Y. A., Quoidbach, J., Robic, F., & Pentland, A. S. (2013, April). Predicting personality using novel mobile phone-based metrics. In *International conference on social computing, behavioral-cultural modeling, and prediction* (pp. 48-55). Springer, Berlin, Heidelberg.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated trait structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology*, 66(6), 574.
- Parr, M. N., Ross, L. A., McManus, B., Bishop, H. J., Wittig, S. M. O., & Stavrinou, D. (2016). Differential impact of personality traits on distracted driving behaviors in teens and older adults. *Accident Analysis and Prevention*, 92, 107-112.
- Shimonaka, Y., Nakazato, K., Gondo, Y., & Takayama, M. (1998). Construction and factorial Validity of the Japanese NEO-PI-R. *The Japanese Journal of Personality*, 6, 138-147.
- Widiger, T. A. (2015). *The Oxford Handbook of THE FIVE FACTOR MODEL*. OXFORD UNIVERSITY PRESS.