Abstract

Personality determination and mapping these to business applications have long since happened by analysing written texts and other social networking attributes of the person. However, consciously given written evidence is a poor indicator of personality. In our project, we make use of cellphone data that measure the latent and repetitive patterns in a person's daily life - like his location coordinates, how he moves, how often he responds to texts, how often he initiates call conversations, what apps he uses most often, etc. With these, we determine what we call an eigenbehaviour. This eigenbehaviour is then mapped to an *eigenpersonality*, which are then compared to the Big 5 personality traits.

Acknowledgement

Our sincere gratitude goes to *Ir. Abhijeet Parmar*, who has been an excellent mentor throughout the course of the project. His excellent sight in envisioning the project and constantly guide us through the processes is truly commendable.
We are ever thankful for his unending support and insightful and invaluable inputs.

1 Introduction

Predicting personality via any means have its challenges and difficulties. A human being's personality has many facets and depths. The term 'personality' itself is so abstract and unmeasurable that any attempts at computing values for it have so far been very mildly successful. Determining what a person is like, from any means whatsoever, is a herculean task.

There are scores of business applications that depend on the personalities of users. In fact, only very few businesses might claim to not care about the end users personalities. However, for the former, having a corpus of user personalities is of great value to them monetarily. Using this corpus they can direct their marketing strategies, bring forth new products, and gauge the demand of their existing products.

1.1 The scene so far

There have been a number of attempts in determining personality. Of the notable ones, one is "Predicting Personality with Social Behaviour" by Jennifer Golbeck at University of Maryland, College Park, in 2013. Another is "Our Twitter Profiles, Our Selves: Predicting Personality with Twitter" by Daniele Quercia at University of Cambridge.

The above use online social networking websites' data to predict personalities. A major breakthrough in social physics research has been engendered by reality mining pioneer Alex Pentland. We refer to his work multiple times in our report. The paper that we gained most help from was "Predicting people personality using novel mobile phone based metrics" by Alex Pentland at MIT Media Lab in 2013.

1.2 The fallacies so far

Though mapping their online web usage to their personalities has its advantages, but the inaccuracies are hard to ignore. For example, one behaves in an entirely different fashion on the web, than they do in real life. They may or may not be so extrovert, they might be very under confident with their sarcasm in real life, or they may be not as social online as they are offline.

The fallacy in the system lies in the assumption that personality can be predicted from consciously submitted data on social networking websites. According to present research limitations, when human beings change the way they interact with others online, automatically their personality changes, which in itself is an implausible condition.

1.3 The novelty

Personality can be predicted when humans are judged for things they do absent mindedly or subconsciously. It also makes sense to use data generated by a person's day to day activities, like his movement patterns, how he responds to conversations, or how often he initiates them.

Data of the kind explained above can be obtained from the one device that is attached to a person all day - mobile phone. One can extract a number of statistics from a user's cellphone in order to predict his personality, provided he permits us to do so. We have obtained cellphone usage data like GPS coordinates of his movements throughout the day, SMS logs, call logs, etc through a mobile application FUNF hosted by MIT Media Lab.

The above methodology has been adopted by Alex Pentland in his work "Predicting people personality using novel mobile phone based metrics" published in 2013. He uses mobile phone data and predicts the Big 5 personality traits for the volunteer. We adopted the same methodology except that we have gone further to determine the Big 5 personality traits upto a much higher resolution. We have thread-bared each trait into its following less abstract components :

Neuroticism : Anxious, self-pity, tense, touchy, unstable, worrying

Extraversion : Active, assertive, energetic, enthusiastic, outgoing, talkative

Openness : Artistic, curious, imaginative, insightful, original, wide interest

Conscientiousness : Efficient, organised, planful, reliable, responsible, thorough

Agreeableness : Appreciative, forgiving, generous, kind, sympathetic

We build eigenbehaviours for each test subject using factor analysis, and we further determine something called the eigenpersonality with recursive learning using Tensor networks. These eigenpersonalities are the Big 5 personality traits for the subject.

Future work: Due to lack of a regular inflow of mobile data, we have been unable to validate these traits with the ground truth obtained from a credible online API (Apply Magic Sauce API by University of Cambridge). We do propose to validate the model once the mobile application is installed by a sufficient number of users in the future.

2 Methodology

The three major parts of our method were obtaining cellular metrics, analyzing the metrics to get personality of the participant and comparing the results with the personality prediction API.

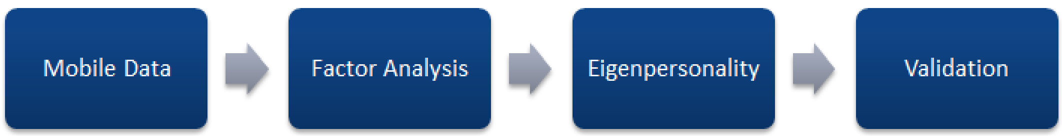

Figure 1: Methodology workflow

2.1 Metrics

Thirty six metrics that were calculated for each volunteer using phone data are as follows:

- Regularity

  Inter event time: The time elapsed between two events is known as inter event time. This considers both the average and variance of the inter-event time of participant?s call, text and call+ text.

  Home Regularity: This metric can be defined as the regularity at which participant is coming back home.

  AR coefficients: The list of all calls and texts made by a participant were converted into a time series which was used as an input to the auto regression model.

- Diversity

  Entropy of contacts: It is the qualitative measure reflecting number of different categories in a given random variable and takes into account the even distribution of basic units.

Contacts to interaction ratio: It is the ratio of overall contacts of the participant to the contacts he interacts with daily.

Number of contacts: Total number of contacts of the participant

- Spatial behavior

  Radius of gyration: It is the smallest circle that covers all the locations a participant has visited in a day.

  Distance traveled: It is the sum of the distance between the consecutive places a participant has been to in a day.

  Entropy of places

- Active behavior

  Response rate: It is the percentage of text participants respond to.

  Response latency: It is the median time taken by participants to answer a text.

  Percent initiated: It is the percent of times the participant has initiated a text or call conversation.

- Basic phone use

  Number of interaction: It is the number of contacts a participant interacts with.

2.2 Analyzing the metrics

Factor Analysis

Factor analysis is a method for investigating whether a number of variables of interest Y_1, Y_2, ... , Y_n are linearly related to a smaller number of unobservable factors F_1, F_2,... , F_k . The covariance matrix obtained from the metrics data of the participant is compared to the covariance matrix generated using the initial loadings known as lambda. The difference between the two covariance matrices is minimized by continuous back propagation of lambda.

Next, eigenvectors of the optimized lambda are calculated. PCA of the eigenvector with the maximum eigenvalue is obtained which is known as *eigenbehavior* of the participant.

Eigenpersonality

Further, a recursive learning using tenser network is applied on the *eigenbehavior* to obtain the *eigenpersonality* of the participant.

2.3 Validation

An online API, Apply Magic Sauce, will be used to validate the results predicted by our system.

Its credibility was initially established by comparing its results with the ground truth (personality survey data) with the accuracy of 74.2 %.

3 Results

We have successfully been able to obtain *eigenpersonalities* for our 9 test subjects. These satisfactorily tell about the person's personality, from their phone usage data alone. This will have a significant use in business applications, wherein just by installing a mobile phone application and a few days of usage, the business can determine personalities of their target users.

As mentioned in the introduction, our future work includes validating the *eigenpersonality* vectors with the Big 5 personality traits predicted by Apply Magic Sauce API, once we have enough test subjects using the mobile application for a number of days continuously.

4 Bibliography

Pentland, Alex, et al. "Predicting people personality using novel mobile phone-based metrics", Social Computing, Behavorial-Cultural Modeling and Prediction, 2013.

Pentland , Alex, et al. "Eigenbehavior: Identifying structure in routine", Springer-Verlag, 2009

http://applymagicsauce.com/ - API by Cambridge University

http://www.outofservice.com/bigfive/ - Online personality test