Causal Feature Selection for Individual Characteristics Prediction

Tao Ding\(^\ast\), Cheng Zhang\(^\dagger\), and Maarten Bos\(^\dagger\)

\(^\ast\)Department of Information Systems
University of Maryland, Baltimore County
taoding01@umbc.edu

\(^\dagger\)Disney Research,
Pittsburgh, USA
{cheng.zhang,mbos}@disneyresearch.com

Abstract

People can be characterized by their demographic information and personality traits. Characterizing people accurately can help predict their preferences, and aid recommendations and advertising. A growing number of studies infer peoples characteristics from behavioral data. However, context factors make behavioral data noisy, making these data harder to use for predictive analytics. In this paper, we demonstrate how to employ causal identification on feature selection and how to predict individuals’ characteristics based on these selected features. We use visitors’ choice data from a large theme park, combined with personality measurements, to investigate the causal relationship between visitors characteristics and their choices in the park. We demonstrate the benefit of feature selection based on causal identification in a supervised prediction task for individual characteristics. Based on our evaluation, our models that trained with features selected based on causal identification outperformed existing methods.

Introduction

Understanding an individual’s characteristics is useful for many real life applications. The term individual characteristics refers to individual differences in characteristic patterns of thinking (Widiger 2011), and includes both demographics and personality. Knowing someone’s individual characteristics can help to understand that person’s preferences, which has important applications ranging from health care (Smith and Spiro III 2002; Giota and Kleftaras 2013) to marketing (Spence et al. 1997; Odekerken-Schröder et al. 2003; Jamal and Goode 2001; Lin 2002; Wang et al. 2013; Yang et al. 2015; Ding and Pan 2016), health (Booth-Kewley and Vickers 1994; Smith and Spiro III 2002; Giota and Kleftaras 2013) and public politics (Greenstein 1992; Caprara et al. 2006; Chirumbolo and Leone 2010). Individual characteristics - such as gender and personality - are relatively stable over time. E.g. the people who display higher score of a particular personality trait continue to display higher score of this personality trait over time, when compared to other individuals (Caspi et al. 2005). This means that it is practical, and relatively stable over time, to predict a person’s behavior and preferences from individual characteristics.

Personality information can be arduous to obtain. A traditional approach to measuring personality requires participants to take a psychological test (e.g. filling out questionnaires), which is time-consuming and difficult to scale up. It is therefore desirable to circumvent this testing process and instead directly predict individual characteristics based on readily available observational data.

A growing proportion of human activities, such as consuming entertainment, shopping, and social interactions, are now mediated by digital services and devices. These digital footprints of consumer behaviors can be recorded and analyzed. Understandably, there is an interest in automatically predicting individual characteristics including age, gender, income and personality traits from these digital footprints. In this work, we focus on predicting individual characteristics from real-life experiences in theme parks.

This task is challenging because behavioral data are noisy. Noisiness can be caused by different factors, such as that while people move in groups, we are trying to pre-
dict personality at an individual level (Elliot et al. 2012; Zhang et al. 2015). In our study we use behavioral data, such as which facilities a person visited during their stay in the theme park, to predict traits such as openness or agreeableness. Imagine a person is visiting a theme park with his/her child, as shown in Figure 1, and our goal is to predict the parent’s traits. The activities of the parent are influenced by the fact that the child is with them. For example, a visit to a toy shop may not be the result of the parent’s traits, but of the context factor (the child). The signal of the parent’s personality is partially drowned out by the noise of the context. This makes it challenging to distinguish individual differences from group behaviors. Identifying which behaviors are directly caused by a person’s characteristics is essential.

In the paper, we propose a new machine learning pipeline to predict individual characteristics. We present a study on 3294 visitor profiles of a large theme park. We first employ causal identification to identify individual behavioral features. Second, we use these features to build predictive models to predict specific personality traits. Our contribution is threefold:

1. This is the first study that investigates the relationship between someone’s location history (defined as visits of facilities in a theme park) and their personality.
2. We take advantage of causal identification methods for feature selection and build a prediction model to predict individual characteristics. Our results demonstrate improvement over baseline models, such as models that use all features or models that use traditional feature selection methods.
3. Compared to other machine learning methods, causal identification is able to obtain interpretable results in a principled way and provide more insights in behavior.

**Background and Related Works**

We aim to predict individual characteristics from real-life theme park visits data with causal identification for feature selection. Therefore we first explain why individual characteristics are interesting and important. Then we summarize related work on individual characteristics prediction using various type of data. Finally we discuss research on causal identification and causal inference.

**Individual characteristics** We define individual characteristics as both demographic information and personality traits. In the context of our application, demographic information includes age, income, and the number of kids included in a theme park visit; personality traits are described using the well established Big5 personality model (Goldberg 1993). The Big5 personality model (Goldberg 1993) represents personality with scores on five personality traits. These traits are openness, conscientiousness, extraversion, agreeableness and neuroticism. Table 1 shows a description of the traits. The Big5 model is widely used to represent a person’s personality.

| Personality | Description |
|-------------|-------------|
| Openness    | A person’s level of intellectual curiosity, creativity and preference for novelty and variety. |
| Conscientiousness | A person’s tendency to be organized and dependable, show self-discipline, act dutifully, and prefer planned rather than spontaneous behavior. |
| Extraversion | A person’s energy, positivity, assertiveness, sociability, talkativeness, and tendency to seek stimulation from the company of others. |
| Agreeableness | A person’s tendency to be compassionate and cooperative. Also a measure of one’s trustiness, helpfulness, and well-tempered nature. |
| Neuroticism  | A person’s tendency to experience unpleasant emotions easily, and have low emotional stability and impulse control. |

One example study showed that age, level of education, and income among 250 hotel restaurant customers was correlated with complaint behavior (Sujithamrak and Lam 2005). Another study revealed a relationship between smoking behavior and demographic variables (Moody 1980). Besides demographics, personality traits also predict behavior. One study found that smokers score higher on openness to experience and lower on conscientiousness, a personality trait related to a tendency to show self-discipline, act dutifully, and aim for achievement (Campbell et al. 2014). A second study examined the link between personality and alcohol consumption and found that alcohol use is correlated positively with sociability and extraversion (Cook et al. 1998). Another study showed a link between personal environments and personality (Gosling et al. 2002). Yet another study (Fennis and Pruyn 2007) found that a brand’s personality profile may carry over and affect perceptions of the personality of the brand’s owner. These studies are often carried out using questionnaires, surveys, or interviews. These methods are time-consuming, and a more efficient way to predict traits would be useful.

**Machine Learning for Individual Characteristics Prediction** Using machine learning methods to predict individual characteristics from behaviors has increasingly gained attention. Ideally, various types of real-life behavior data would be used to predict individual characteristics. However, existing research is mainly limited to the use of social network data to predict individual characteristics.

There have been several studies to show that users’ digital footprints on social networks can be used to infer their demographics and personalities (Farnadi et al. 2013;
Mislove et al. 2011; Golbeck et al. 2011; Kosinski et al. 2013). When there is a large volume of data available especially for high dimensional data feature learning is needed to reduce data dimensionality (Kosinski et al. 2015; Pennebaker et al. 2007; Kosinski et al. 2013; Schwartz et al. 2013).

An alternative to this approach is supervised learning. Studies often focus on recognizing individual characteristics from a limited amount of noisy annotated data, using univariate and multivariate regression formulations. Supervised learning can be used either on the raw data or on extracted features. These features can be extracted using representation learning methods. In supervised learning, feature selection can improve the accuracy of the learning algorithm. There are existing methods to identify relevant features and remove irrelevant and redundant features before training a model. In a regression model, the common method is to apply L1 regularization that adds a penalty \( \alpha \sum_{i=1}^{n} |w_i| \) to the loss function (L1-norm), which forces weak features to have zeros as coefficients. This is also called LASSO regression (Tibshirani 1996). This inherently creates feature selection. Another method is correlation based feature selection (CFS) which filters the features by performing feature selection based on correlation analysis (Yu and Liu 2003). However, both these methods do not provide interpretable results and lack principled justification.

**Causal Identification** A fundamental assumption for research on individual characteristics prediction is the causal relationship between the individual characteristics and behaviors. Additionally, prediction is only possible if the target is the cause of the data (effect) (Schölkopf et al. 2012). Most feature selection methods do not attempt to uncover cause-effect relationships between feature and target (Guyon et al. 2007). A correlated feature may be predictive of an outcome without intervention. Causal identification (Spirtes et al. 2000) can identify features with direct effects caused by target variables which contain predictive information of the target. Additionally, the result from causal identification - a graph indicating causal relationships - is highly interpretable. In individual characteristics identification, causal identification is able to determine which behaviors are caused by individual characteristics and which behaviors are caused by the context. In this paper, we apply different causal identification methods to identify direct effects caused by personal traits. Using causal identification methods for feature selection leads to better predictive performance in our application.

**Dataset**

The data for our study were collected from visitors of a large theme park resort. The resort has over 30 hotels, more than 100 restaurants and hundreds of attractions. We assume individual characteristics affect the choices visitors make. We asked visitors’ permission to use their choice and location history. From visitors who agreed to this we retrieved the locations they visited, including restaurants, stands and kiosks, attractions&rides, stores and other entertainment. Participating visitors also allowed us access to data such as their park entry time, hotel check-in time, purchases, and other metadata such as length of stay, trip cost, and which of the parks they visited first (see Table 2).

Visitors who participated in our study were asked to fill out a questionnaire, in which they were asked about the previously mentioned experiences at the resort, as well as demographical questions and the Ten-Item Personality Inventory (TIPI). The TIPI was developed as a shortened version of a 50-item personality scale and is often used to measure Big5 personality traits. The TIPI has 10 items, with

| Item                  | Data Type | Representation     |
|-----------------------|-----------|-------------------|
| Gender                | Categorical | 0: male 1: female |
| Age                   | Categorical | 17-78             |
| Total visits          | Numerical  | 0-99              |
| Vacation days         | Numerical  | 0-99              |
| Income                | Numerical  | 1-14              |
| Trip cost             | Numerical  | 0-9999            |
| # of children         | Numerical  | 0-9               |
| Hotel choice          | Categorical | 0: No 1: Yes      |
| How the trip was booked | Categorical | 13 categories     |
| Dining plan           | Categorical | 0: No 1: Yes      |
| First park visited    | Categorical | 4 different theme parks |
two items measuring each of the personality traits. The TIPI uses a 7-point Likert scale from 1 (“Strongly Agree”) to 7 (“Strongly Disagree”). For every participant, we average the scores of the items associated with the personality to achieve five personality scores.

A total of 3997 visitors participated. After filtering out the responses with missing values, we collected the individual characteristics of 3293 participants. Figure 3 show the distribution of demographics. In the remaining data 31% of are males and 69% are females. The average age is 44 years old. The average number of children in the group is 1.2. Figure 2 show the distribution of big5 personalities. The range of the scores are between [1,7]. Our participants experienced a total of 505 different things in the park.

The objective of the study is to measure individual differences based on visitors’ experiences and choices at the theme park resort. As mentioned, studies that analyze people’s digital footprints generally have an advantage in that they are more likely to reflect the choices of the individual filling out the personality survey (such as Facebook likes, publishing/commenting etc.). Most people do not visit the theme parks alone (the average group size is larger than 3 in our dataset). We have to assume that in many cases parents make decisions based on their children’s preferences. That makes it difficult to link individual characteristics to the choices made by the parents. In the following section, we employ causal inference to capture effects caused by individual characteristics in order to reduce others’ impacts in decision making.

Causal Identification

Causal identification aims at inferring causal relationships from observational data (Pearl 2009; Imbens and Rubin 2015). Many observed correlations in observational data are mediated through unobserved confounding variables. The goal of causal identification is to distinguish between such mediated correlations and truly causal relationships.

The basis of our causal identification analysis is a Bayesian network. This is a directed acyclic graph (DAG) in which nodes represent variables and arrows between the nodes represent the direction of causation between the nodes. The problem is that we only observe the nodes; these are our observations (location data, metadata, and the Big5 scores). We want to infer arrows that best explain the observed correlations. This is done in such a way that the number of causal arrows between the nodes is, in some sense, minimal to be consistent with the observed correlations. Having identified the causal structure, we exclude all predictor variables which are not causally associated with the target variables of interest. This is the central idea of how we propose to select features.

There are two main classes of algorithms to learn causal DAGs from observational data: constraint-based and score-based ones. Constraint-based methods use independence and dependence constraints obtained from statistical tests to narrow down the candidate graphs that explain the data. Our first proposed method relies on using the PC algorithm (Spirtes and Meek 1995) for feature selection; it belongs to the first class. The algorithm begins by learning an undirected Bayesian network that explains the variations by running conditional independence tests. The second phase consists of orienting the Bayesian network, thereby avoiding cycles and v-structures. We use a significant level of 0.05 in the independent test.

Score-based methods, on the other hand, provide a metric of confidence in the entire output model. The algorithm we use in this paper is Fast Greedy Equivalence Search (FGES), which greedily searches over Bayesian network structures (Ramsey 2015), and outputs the highest scoring model it finds. The score which the algorithm uses is the Bayesian Information Criterion (BIC) (Raftery 1995):

\[
BIC = 2 \cdot \ln P(data|\theta, M) - c \cdot k \cdot \ln(n).
\]

Above, \(M\) denotes the Bayesian network, \(\theta\) denotes its parameters, \(k\) is the total number of parameters, and \(n\) the number of data points. The constant \(c\) is a free parameter that we can tune. This constant penalizes large numbers of parameters and thus determines the complexity of the network; we chose \(c = 0.1\) to obtain a comparable number of features between PC and FGES. The FGES performs a forward stepping search in which edges are added between nodes in order to increase the BIC until no single edge addition increases the score. Then, it performs a backward stepping search in which unnecessary edges are removed.

Both the PC and FGES algorithms are used in this paper for feature selection and compared against the LASSO approach in our experimental section. The L1 term \(1\) is set as 0.01 to obtain a comparable number of features.

Predictive Evaluation

We implemented different prediction models to infer individual characteristics from location history and metadata. The experiments are designed to answer the following questions: a) Can the visitors’ choices infer individual characteristics? b) Is there any benefit of performing feature selection according to causal identification? c) Is metadata informative to predict individual characteristics? To answer (a), we denote the visitors’ choice history as a binary representation of a fixed size vector, 1 represents that the visitor visited a place. We use LASSO (Least Absolute Shrinkage and Selection Operator)(Tibshirani 1996) linear regression to perform predictive tasks for continuous outcomes like age, income.

\(^1\)We tried different values of \(c = 1, 0.1, 0.01\), without noticing a significant change in predictive performance.
Figure 4: Stability of Feature Selection for Big5 Personality.

(a) Agreeableness  (b) Conscientiousness  (c) Extraversion  (d) Neuroticism  (e) Openness

Table 3: Root mean square error (RMSE) and coefficient of determination ($R^2$) results for the personality prediction using visitors’ choices and metadata

| Approach | Agr $R^2$ | RMSE | Cons $R^2$ | RMSE | Extr $R^2$ | RMSE | Neu $R^2$ | RMSE | Open $R^2$ | RMSE |
|----------|-----------|------|------------|------|------------|------|-----------|------|------------|------|
| Visitors’ choices |
| LASSO    | -0.090    | 0.171| -0.085     | 0.175| -0.087     | 0.242| -0.121    | 0.192| -0.065     | 0.171|
| PC       | 1.143     | 0.170| 1.802      | 0.174| 1.975      | 0.240| 1.090     | 0.190| 1.954      | 0.169|
| FGES     | 1.770     | 0.170| 1.502      | 0.174| 2.368      | 0.239| 1.088     | 0.190| 1.646      | 0.169|
| Visitors’ choices + Metadata |
| LASSO    | 5.942     | 0.166| 1.152      | 0.174| -0.338     | 0.242| 2.109     | 0.189| 1.047      | 0.170|
| PC       | 6.208     | 0.166| 3.625      | 0.172| 2.248      | 0.239| 2.959     | 0.189| 2.388      | 0.169|
| FGES     | 6.951     | 0.165| 3.525      | 0.172| 2.482      | 0.239| 2.995     | 0.189| 2.251      | 0.169|

The five personality are Extraversion(Extr), Agreeableness(Agr), Conscientiousness(Cons), Neuroticism(Neu), and Openness(Open). LASSO: $\alpha = 0.01$. PC: $p = 0.05$. FGES: $c = 0.1$.

the number of children and personality. LASSO is a regression method that automatically performs both feature selection and regression. All results are based on 10-fold cross validation to avoid overfitting. To answer (b), we run PC and FGES to search for causal explanations on training data for each fold. We train predictive models by using effects of features caused by specific characteristics and predict individual characteristics on the test set. To answer (c), we add metadata in models trained in (a) and (b) to see whether the performance improves.

To identify causal relationships, we employ a Tetrad (Spirtes et al. 2000)$^2$ analysis to manipulate and individually study the different individual characteristics. Tetrad provides different causal search algorithms to search when there may be unobserved cofounders of measured variables and output graphical representations (Scheines et al. 1998). In our experiment, we run PC and FGES search algorithms. In PC search, we choose a Fisher-z transformation to do independent tests which converts correlations into an almost normally distributed measure. The p-value is set at 0.05. In FGES setting, the penalty discount, which is $c$ in the BIC formula, is set to 0.1 in order to get a similar number of features as the PC algorithm can get from training data.

Stability Test To compare the stability of feature selection methods, we calculated the selection probability for each feature in the Big5 personality prediction task under bootstrapping, following (Mandt et al. 2017). We thereby randomly removed 10 percent of the data and applied our feature selection algorithms on the remaining data. We repeated this procedure 10 times and computed the empirical feature selection probabilities. Figure 4 shows the results, where the features have been ranked according to their selection probability. We see that the top 5-10 features are mostly stable across all methods, with a slightly better performance of the LASSO method on extraversion, conscientiousness and agreeableness, and with a slightly better performance of our proposed causal identification methods on openness and neuroticism. Note that the features selected by the different methods are different. Next, we compare how well the selected features serve in our personality prediction task.

Predictive Performance To evaluate performance, for continuous outcomes, we evaluate the results based on root mean squared error (RMSE) and Co-efficient of Determination($R^2$). RMSE measures the difference between values the model predicted and the observed values. $R^2$ can be described by the following formula:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_{obs}^t - y_{pred}^t)^2}{n}}$$

where $y_{obs}^t$ and $y_{pred}^t$ are the observed and predicted scores for instance $t$, and $n$ is the sample size. $R^2$ is the ratio of

\[http://www.phil.cmu.edu/tetrad/\]
Table 4: Root mean square error (RMSE) and coefficient of determination ($R^2$) results for the demographics using visitors’ choices and metadata

| Approach          | Age $R^2$ | RMSE   | # of Children $R^2$ | RMSE   | Income $R^2$ | RMSE   |
|------------------|-----------|--------|---------------------|--------|--------------|--------|
| Visitors’ choices| LASSO     | 2.294  | 0.146               | -16.53 | 0.153        | -0.352 | 0.256    |
|                  | PC        | 13.675 | 0.137               | 0.097  | 0.142        | 8.160  | 0.245    |
|                  | FGES      | 13.843 | 0.137               | 0.815  | 0.141        | 8.107  | 0.245    |
| Visitors’ choices+Metadata | LASSO | 8.432  | 0.141               | -2.386 | 0.143        | 1.160  | 0.255    |
|                  | PC        | 21.684 | 0.131               | 8.618  | 0.135        | 10.936 | 0.242    |
|                  | FGES      | 21.975 | 0.131               | 9.539  | 0.135        | 10.606 | 0.242    |

LASSO: $\alpha = 0.01$, PC: $p = 0.05$, FGES: $c = 0.1$.

In general, the $R^2$ results of causal identification show the models outperform the constant average baseline. The FGES model for predicting age and the number of children achieved best performance with $R^2 = 13.843$ and $R^2 = 0.815$. The PC model for predicting income achieved best performance with $R^2 = 8.160$.

In summary, there are two findings from predictive evaluation: 1) we prove the effectiveness of the metadata in predicting individual characteristics; 2) the features selected based on causal identification can improve the prediction performance of individual characteristics.

Causal Relationship Analysis

In addition to building models that predict individual characteristics, we are also interested in understanding the causal relationship between a visitor’s characteristics and their choices. The Tetrad method specifies causal relations among the variables via a representation of a directed graph. The edge $X \rightarrow Y$ can be interpreted as $X$ has a direct causal effect on $Y$. We extract causal relationships between personality traits and visitors’ choices with metadata from CBN, which are showed in table 5. Within the 505 locations, agreeableness and neuroticism have direct causal effects only on 5 of the guests’ choices. Agreeableness links to some specific quick food restaurants and some gift shops (we omit actual names of these locations). People who score higher on agreeableness tend to visit popular parks (Park A and Park b) which are visited by most people. People who score higher on neuroticism link to some family style restaurants and dining events, which mean they tend to enjoy the comfort associated with environment. Conscientiousness has direct causal effects on 10 locations, most are facilities located in a specific hotel area of the theme park resort area, which means that conscientious people may be more likely to spend time near their hotel. Openness has direct causal effects on 10 location visits including thrill rides, quick food restaurants, and indoor theatres. Comparing with other personality traits, people who score higher openness link to most thrill rides with big drop, which mean they enjoy new experiences and seek out adventure. Extraversion has direct effects on visits of many different places (13 places, located in 6 different parks).
We demonstrated predictive improvements by using causal identification. Personality inference is complex and can be influenced by factors such as age (Caspi and Silva 1995), income (Raadal et al. 1995), family environment (Hoffman 1991) and cultural differences (Hofstede and McCrae 2004). We did not try to capture all the relevant factors in our model, instead we only used the metadata of a visitors’ profile that they shared with us. However, if one were to analyze a fuller picture of visitors’ choices - such as when they were where - this might show to be a more interpretable causal relationship. Meanwhile, in our model, we only use the item ids of facilities in park without exploring causal relationships between individual characteristics and the facilities’ metadata. Basically any data that are a direct result of the target person’s decision-making could be helpful to predict that person’s characteristics.

## Conclusion

In our study, we focused on three main tasks. (1) We employed causal identification to select features which are most informative of individual characteristics (2) we built personal traits prediction models based on guests’ choices and metadata (3) we employed causal relationship analysis to obtain human interpretable results. Our investigation has shown that models using causal identification significantly outperform the baseline models in all individual characteristics’ prediction, as well as demonstrated the causal relationships between guests’ choices and their personality traits.

## Discussion

We appreciate Michelle Ma for providing the illustration of an example in our paper. We are also grateful to Kun Zhang for his valuable advice and guidance about identifying causal relationship using Tetrad.

## Acknowledgement

Stephanie Booth-Kewley and Ross R Vickers. Associations between major domains of personality and health behavior. *Journal of personality*, 62(3):281–298, 1994.

Stephen Campbell, Lyndsay Henry, Jackie Hammelman, and Maya Pignatore. Personality and smoking behaviour of non-smokers, previous smokers, and habitual smokers. *J Addict Research & Therapy*, 5:191, 2014.

Gian Vittorio Caprara, Shalom Schwartz, Cristina Capanna, Michele Vecchione, and Claudio Barbaranelli. Personality and politics: Values, traits, and political choice. *Political psychology*, 27(1):1–28, 2006.

Avshalom Caspi and Phil A Silva. Temperamental qualities at age three predict personality traits in young adulthood: Longitudinal evidence from a birth cohort. *Child development*, 66(2):486–498, 1995.

Avshalom Caspi, Brent W Roberts, and Rebecca L Shiner. Personality development: Stability and change. *Annu. Rev. Psychol.*, 56:453–484, 2005.

Antonio Chiurambolo and Luigi Leone. Personality and politics: The role of the hexaco model of personality in predicting ideology and voting. *Personality and Individual Differences*, 49(1):43–48, 2010.

Mark Cook, Alison Young, Dean Taylor, and Anthony P Bedford. Personality correlates of alcohol consumption. *Personality and Individual Differences*, 24(5):641–647, 1998.

## Table 5: Identified items that are caused by personality found by PC algorithm. Each [.] present a facility or a meta feature which are used as features. Apart from the personal meta data, only the facility ids are used in our method. Due to confidentiality, we do not list the names of the facilities; we only show the metadata of the facilities to interpret the results. The metadata of the facilities themselves are not used in the analysis. The blue colored features indicate that the causal relationships are mutual.

| Trait | Services/Metadata                                                                 |
|-------|----------------------------------------------------------------------------------|
| Agr   | [Park A: Restaurants, Quick Service], [Park B: Toys, Apparel, Accessories], [Park A: Housewares, Food], [Park A: Gift shop], [Park B: Housewares, Apparel, Accessories], [age], [gender] |
| Cons  | [Park A: Camera, Media, Apparel, Accessories], [Resort A], [Resort B: Spa, Pool Bars], [Resort B: Restaurants, Quick Service], [Resort B: Restaurants, Table Service], [Park: Restaurants, Quick Service], [Park: Restaurants, Quick Service], [Resort C: Apparel, Accessories], [Park F: Camera, Media, Apparel, Accessories], [Park C: Gift shop], [Park D: Apparel, Accessories], [income] |
| Neu   | [Park A: Restaurants, Character dining], [Park A: Restaurants, Table service], [Resort D: Restaurants, Buffet/Family Style], [Park A: Restaurants, Table service], [Park A: Gift shop], [age], [income] |
| Extr  | [Park C: Restaurant, Quick Service ], [Park A: Character showcase, Preschool, Kids], [Park C: Gift shop, Apparel, Accessories], [Park C: Indoor theater], [Park D: Gift shop, Camera, Media], [Resort A], [Resort D: Lounges], [Park A: Apparel, Accessories], [Park D: Restaurant, Character Dining, Family Style], [Park E: Gift shop], [Park A: Art Collection, Gift shop], [Resort E: Health beauty], [Park F: Gift shop, Apparel, Accessories], [gender], [# of children] |
| Open  | [Park A: Thrill ride], [Park G: Thrill ride], [Park A: Thrill ride], [Park D: Indoor theatre], [Park D: Restaurants, Quick Service], [Park A: Restaurants, Quick Service], [Park A: Restaurants, Quick Service], [Park A: Restaurants, Quick Service], [Park A: Gift shop, Apparel, Accessories], [Park A: Slow ride], [Park D: Gift shop, Apparel, Accessories], [age] |
Tao Ding and Shimei Pan. Personalized emphasis framing for persuasive message generation. arXiv preprint arXiv:1607.08898, 2016.

Aronson Elliot, Wilson Timothy D., and Robin M. Akert. Social psychology. Pearson Education Canada, 2012.

Golnoosh Farjadi, Susana Zoghi, Marie-Francine Moens, and Martine De Cock. Recognising personality traits using facebook status updates. In ICWSM’13. AAAI, 2013.

Bob M Fennis and Ad Th H Pruyn. You are what you wear: Brand personality influences on consumer impression formation. Journal of Business Research, 60(6):634–639, 2007.

Kyrilaki G Giota and George Kleftaras. The role of personality and depression in problematic use of social networking sites in greece. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 7(3), 2013.

Jennifer Golbeck, Cristina Robles, and Karen Turner. Predicting personality with social media. In CHI ’11 extended abstracts on human factors in computing systems, pages 253–262. ACM, 2011.

Lewis R Goldberg. The structure of phenotypic personality traits. American psychologist, 48(1):26, 1993.

Samuel D Gosling, Sei Jin Ko, Thomas Mannarelli, and Margaret E Morris. A room with a cue: personality judgments based on offices and bedrooms. Journal of personality and social psychology, 82(3):379, 2002.

Fred I Greenstein. Can personality and politics be studied systematically? Political Psychology, pages 105–128, 1992.

Isabelle Guyon, Constantin Aliferis, and André Elisseeff. Causal feature selection. Computational methods of feature selection, pages 63–82, 2007.

Lois W Hoffman. The influence of the family environment on personality: Accounting for sibling differences. Psychological Bulletin, 110(2):187, 1991.

Geert Hofstede and Robert R McCrae. Personality and culture revisited: Linking traits and dimensions of culture. Cross-cultural research, 38(1):52–88, 2004.

Guido W Imbens and Donald B Rubin. Causal inference in statistics, social, and biomedical sciences. Cambridge University Press, 2015.

Ahmad Jamal and Mark MH Goode. Consumers and brands: a study of the impact of self-image congruence on brand preference and satisfaction. Marketing Intelligence & Planning, 19(7):482–492, 2001.

Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, 110(15):5802–5805, 2013.

Michal Kosinski, Sandra C Matz, Samuel D Gosling, Vesselin Popov, and David Stillwell. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. American Psychologist, 70(6):543, 2015.

Chin-Feng Lin. Segmenting customer brand preference: demographic or psychographic. Journal of Product & Brand Management, 11(4):249–268, 2002.

Stephan Mandt, Florian Wenzel, Shinichi Nakajima, John Cunningham, Christoph Lippert, and Marius Kloft. Sparse probit linear mixed model. Machine Learning, pages 1–22, 2017.

Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J Niels Rosenquist. Understanding the demographics of twitter users. ICWSM, 11:5th, 2011.
Lei Yu and Huan Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *ICML*, pages 856–863, 2003.

Cheng Zhang, John Guiver, Tom Minka, and Yordan Zaykov. Groupbox: A generative model for group recommendation, 2015.