Understanding Purchase Behavior through Personality-driven Traces

Xiaotong Liu
IBM Research-Almaden
San Jose, California, USA
xiaotong.liu@ibm.com

Rama Akkiraju
IBM Research-Almaden
San Jose, California, USA
akkiraju@us.ibm.com

Anbang Xu
IBM Research-Almaden
San Jose, California, USA
anbangxu@us.ibm.com

Vibha Sinha
IBM Research-Almaden
San Jose, California, USA
vibha.sinha@us.ibm.com

 CHI'17 Extended Abstracts, May 06-11, 2017, Denver, CO, USA
ACM 978-1-4503-4656-6/17/05. http://dx.doi.org/10.1145/3027063.3053168

Abstract
We present a computational approach to understanding users' purchase behavior through personality analytics. We model purchase behavior as interactions between personality traits, consumption preferences and product attributes, and represent such interactions as digital traces. We model all possible digital traces using a likelihood trace graph, and determine the likelihood of an edge based on the associations between adjacent pairs of personality attributes, consumption preferences, and product attributes, respectively. In our approach, users' personality traits are inferred from their written texts, and are correlated to a canonical set of consumption preferences. The consumption preferences are then mapped to product descriptions based on their semantic similarity. We demonstrate the effectiveness and usefulness of our approach through a case study on a real-world purchase data.

Author Keywords
Purchase behavior; personality analytics; consumption preference; digital traces.

Introduction
Psychology and user behavior analysis disciplines tell us that people's personality traits [5], their intrinsic needs [7, 10] and values [17] are known to play a significant role in their purchasing decisions. Marketers are highly interested

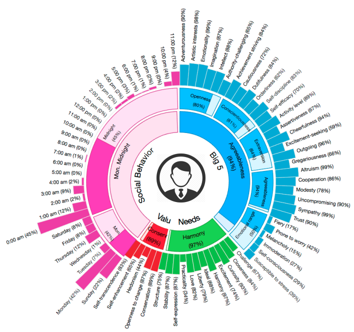


Figure 1: An illustrative visualization of personality traits [11].

in knowing these innate attributes of people that drive their preferences, behaviors and decisions. These insights enable them to design marketing campaigns that appeal to different people differently.

The relationship between personality and purchase behavior has been studied across a variety of domains. For example, users' personality traits have been found to play an important role in establishing personal food purchases [4], and users' overall environmental consciousness has a positive impact on green purchasing decisions [16]. However, existing studies lack a computational framework for understanding and explaining the effects of personality traits on users' purchase behavior. So far marketers mostly correlate personality traits to purchase behavior in specific domains. However, this approach is not scalable. Mapping individual user attributes with individual product attributes is expensive and time-consuming, and would be a never-ending task as there are way too many products to map with.

More importantly, even if such mappings can be made for specific products, marketers are often left without any means to understand the observed behaviors. For instance, what if we knew that people who are *modest*, *conscientious*, *idealistic* and who value *helping others* are likely to prefer low-emission vehicles because they *prefer environmentally friendly products*? This not only explains the motivations and drivers of users' purchase behavior toward low-emission cars but also enables marketers to now consider recommending other environmentally friendly product offers such as organic produce, fair-trade goods, eco-friendly kitchen supplies, etc. Therefore, we argue that introducing an abstract layer of consumption preferences, makes the interaction between personality and purchase behavior more explanatory and generalizable.

In this work, we present a computational approach to understanding users' purchase behavior through the consumption preferences of individuals derived from personality analytics. To understand how personality traits affect purchase behavior, and to explain such effects through consumption preferences, we model purchase behavior as interactions between personality traits, consumption preferences and product attributes, and represent such interactions as *digital traces*. We model all possible digital traces using a likelihood trace graph, and determine the likelihood of an edge based on the associations between adjacent pairs of personality attributes, consumption preferences, and product attributes, respectively. In our approach, users' personality traits are inferred from their written texts, and are correlated to a canonical set of consumption preferences. The consumption preferences are then mapped to product descriptions based on their semantic similarity. We demonstrate the effectiveness and usefulness of our approach through a case study on a real-world purchase data. The study shows that this computational approach is feasible to determine and explain users' purchase behavior based on the personality-driven traces via a canonical consumption preference model.

Background

A person's personality traits are individuals' characteristic patterns of thoughts, feelings, and behavior. Personality traits persist for long periods of time and have profound impact on many areas of our life. In this work, we are exploring the use of personality traits to enrich the understanding of users' purchase behavior. In particular, we focus on studying three types of personality traits: personality characteristics, fundamental needs, and human values (shown in Figure 1). **Personality characteristics** are derived from the Big-five personality traits model [5], which captures the personality characteristics of individuals and how they engage with the world across five primary dimensions: Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism (also known as Emotional Range). **Needs** describe which aspects of a product will resonate with a person: Excitement, Harmony, Curiosity, Ideal, Closeness, Self-expression, Liberty, Love, Practicality, Stability, Challenge, and Structure [7, 10]. **Values** describe motivating factors that influence a person's decision making [17]. We measure them along five dimensions: Self-transcendence / Helping others, Conservation / Tradition, Hedonism / Taking pleasure in life, Self-enhancement / Achieving success, and Open to change / Excitement.

Personality-driven Traces

To understand how personality traits affect purchase behavior, as well as explain such effects through consumption preferences, we model purchase behavior as interactions between personality traits, consumption preferences and product attributes, and represent such interactions as *digital traces*. A digital trace starts from a user's personality trait, goes through a correlated consumption preference, and ends with a relevant product attribute. For example, a digital trace can connect the personality trait *conscientiousness* to the consumption preference *environmentally friendly*, which is connected to the product attribute *low-emission*. Such traces can explain the motivations and drivers of users' purchase behavior: people who are *conscientious* are likely to prefer low emissions vehicles because they *prefer environmentally friendly products*. As personality is the driving factor, a digital trace is also called a *personality-driven trace*.

We model all possible digital traces using a likelihood trace graph called LTGraph. As shown in Figure 2, nodes are represented as aligned rectangles and edges are drawn as curves connecting the corresponding nodes. The width of a curve encodes the likelihood of an edge, which is computed

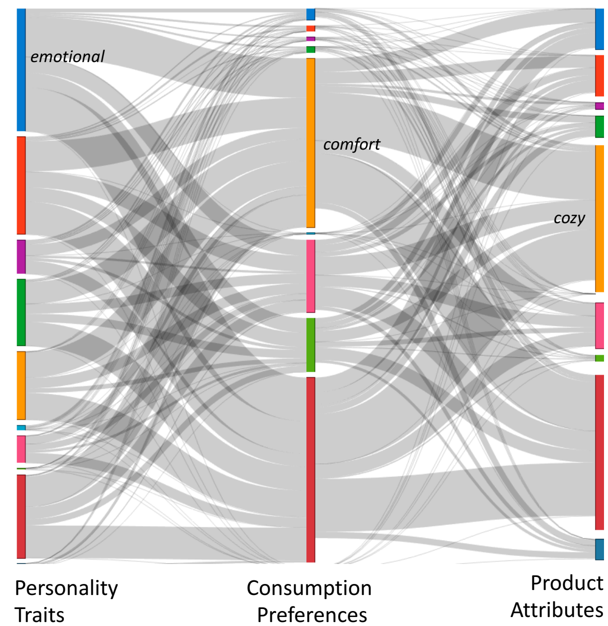


Figure 2: An illustrative visualization of the digital traces in a Likelihood Trace Graph (LTGraph). Nodes are represented as aligned rectangles and edges are drawn as curves connecting the corresponding nodes. For example, a digital trace connects the trait *emotional*, the preference *comfort* and the product attribute *cozy*.

as the association between the connected personality attribute and consumption preference, or the association between the connected consumption preference and product attribute. For instance, Figure 2 shows how a digital trace connects the trait *innovative*, the preference *artistic* and the product attribute *expressive*.

Given users' written texts and product descriptions as input, we now describe how we compute the digital traces in a LTGraph.

Inferring Personality Traits
Recent research in psycholinguistics has shown it is possible to automatically infer personal traits from one's linguistic footprints such as tweets and blogs [18, 3]. In our work, we used a similar computational approach. Specifically, given written texts authored by a user, we compute the word counts of different psychologically-meaningful word categories defined in the Linguistic Inquiry and Word Count (LIWC) dictionary [15]. The LIWC counts were then used to build prediction models to correlate one's word usage with one's ground truth personality traits obtained via a prior psychometric survey. A regression model that is built using this approach is then used to automatically infer a given user's personal traits from that user's written text. More details on our model and the accuracy measurements are available in [11].

Deriving Correlations Between Traits and Preferences
We make use of a consumption preference model, proposed by our team at IBM Watson [1], to learn a user's likelihood to prefer different products, services, and activities from their personality traits. The model was built on a taxonomy of 51 consumption preferences in 8 categories: *shopping*, *movie*, *music*, *reading and learning*, *health and activity*, *volunteering*, *environmental concern*, and *entrepreneurship*. The consumption preference predictions were shown to performed at least 9% better than random [1].

Deriving Semantic Similarity between Preferences and Product Attributes
Marketers are adept at crafting messages carefully and meticulously to bring out the salient aspects of the products that they would like to highlight to their specific target audience. These product descriptions often not only accentuate the product features but also emphasize the symbolic aspects of the product so as to appeal to the target audience. We use these product descriptions and product features to find relevant associations with the canonical consumption preferences.

For each production description, a bag of words is constructed using text analysis techniques such as word tokenization, abbreviation expansion and parts of speech tagging. Since the taxonomy of consumption preference model [1] consists of canonical preference keywords (e.g., *comfort*, *style*, *quality* in the *shopping* category), the problem of correlating preferences to product attributes is transformed to the problem of matching a set of preference keywords with a set of words representing product features. A straightforward solution is to use *n*-grams [2] to calculate the number of shared sequences of words between two word set, and consider word sets with larger shared words to be similar. However, experimentally we found few product descriptions shared common words with preference keywords (e.g., no more than 5% of 1000 clothing descriptions contain at least one of the words *comfort*, *style* and *quality* in the *shopping* category. For more details on the dataset from which we derived this result, please see the Case Study section below).

Therefore, to solve this matching problem, we use a word embedding method, the *word2vec* neural network language model [14, 6], to represent words in a continuous feature space. Each dimension of the word2vec embedding represents a latent feature of the word, which can capture useful syntactic and semantic properties. In our approach, we first convert each word in each of the product description space and the consumption preference space to a feature vector using word2vec, and then compute the average of fea-

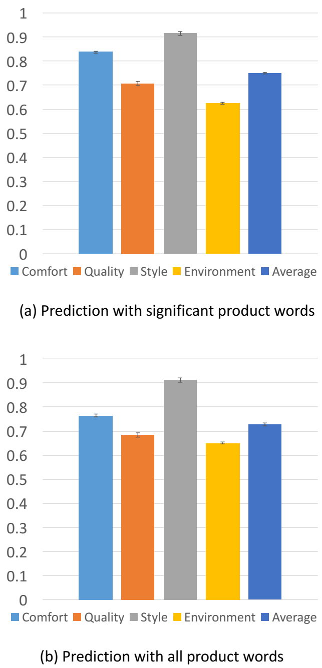


Figure 3: Overall prediction ranks of purchases of the clothing products. (a) Prediction with only significant words in product descriptions. (b) Prediction with all extracted words in product descriptions.

ture vectors for both preference word set and product word set. In this sense, the cosine similarity of the two average vectors approximates the semantic similarity between the corresponding preference and product description.

Case Study

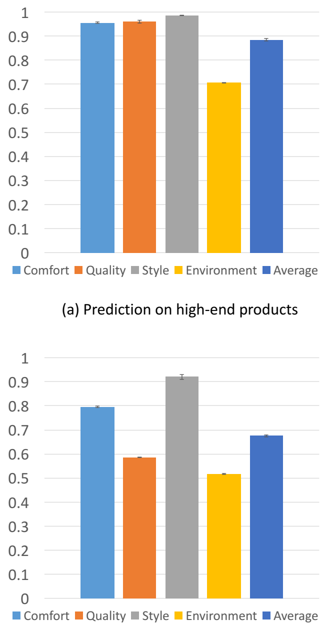
To evaluate the effectiveness and usefulness of our approach, we perform a case study on an Amazon product data set. We are interested in answering two questions in this study: (1) Is it feasible to use personality-driven traces for product recommendation? (2) Is it valid to use personality-driven traces to explain user behavior when considering additional factors?

Data Processing

The Amazon product data set contains product reviews (ratings, text, helpfulness votes) and metadata (descriptions, category information, price, brand, and image features) from Amazon, spanning May 1996 - July 2014 [13]. Our experimental data set includes 1000 products in the *cloth* category, which are described by at least 100 words. We extracted all Amazon users that have reviewed these products, and obtained 339 reviewers with 772 reviews. To compute the LTGraph from the experimental data set, we aggregated all reviews each Amazon user has as the written texts to infer his or her personality traits. The distributions of the reviewers' personality traits were close to normal distributions. We scoped our analysis to include four consumption preferences in this clothing shopping scenario from our consumption preference model [1]. They include: *comfort*, *style*, *quality* and *environmental concern*. We used the consumption preference model to derive correlations between these preferences and the inferred personality traits. We determined the similarity between preference keywords and product descriptions using word2vec embeddings. As a result, every edge in the LTGraph is associated with a likelihood value.

Personality-driven Traces for Product Recommendation

With the LTGraph, we can predict whether an Amazon user would purchase a product based on the personality-driven traces. The *purchase likelihood* is determined by multiplying the likelihoods along a digital trace. To evaluate the quality of the predictions in the LTGraph, we adopt a well-established rank metric [8] in recommendation systems, which extends the recall metric to take the positions of correct items in a ranked list into account. We consider the reviews as the ground truth of purchases, assuming an Amazon user will review a product only when he or she has purchased the product (Amazon later added a *verified purchase* tag. The dataset we used did not yet have that tag). We sort all purchase predictions based on the purchase likelihood of the traces, and then find out the positions of the purchases in the prediction list as their ranks. The best prediction rank is 1 and the worst is 0. A random prediction has an expected rank of 0.5 on average, which is considered as the baseline. We report the average prediction ranks as well as the ranks per consumption preference. We experimented with two alternative prediction approaches: one only considers significant words in product descriptions when deriving the likelihoods; the other simply uses all words of product descriptions. The significant words are those with high TF-IDF (term frequency-inverse document frequency) values, which reflect how important words are to a document in a corpus. As shown in Figure 3, our approach achieved an average rank of 0.75 (with significant words) and 0.728 (with all words), both of which are far better than random prediction with expected rank of 0.5. This indicates that personality-driven traces are useful in making product recommendations for users.



(a) Prediction on high-end products

(b) Prediction on low-end products

Figure 4: Prediction ranks of purchases of high-end and low-end clothing products. (a) Prediction on the high-end clothing products. (b) Prediction on the low-end clothing products.

Personality-driven Traces for User Behavior Analysis
As a follow-up study, we further investigated the effects of personality traits on purchase behavior when taking into account additional factors. One of such additional factors is price [9, 12]. Price is a situational factor (e.g. people prefer lower priced product in general). This may reduce the effect of product-self congruence (e.g., the effect of personality traits on purchase behavior). In our study, we partition the products into three groups based on their price, and compare the prediction ranks on one group of high-end products and one group of low-end products. In Figure 4, we observed that personality traits have a stronger effect on purchases of high-end products, with an average rank of 0.885, while the effect was not as strong on purchases of low-end products, with an average rank of 0.675 (the difference is significant based on t-test with p-value < 0.05). This shows that personality has less effects on purchases of low-end products, which is consistent with previous works [9, 12]: when price is not a major consideration, people can afford to consider other factors such as comfort and quality.

Discussion and Future Work

The personality-driven traces bring a novel perspective to understanding of purchase behavior. For example, in our study, we found that a user who is *prone to worry*, *dutiful* and *emotional* strongly prefer *comfort* and decided to purchase a tutu, which is described as *“This adorable basic ballerina tutu is perfect for dance recitals. Fairy princess dress up, costume, play and much. Comes individually packaged. Use for a Tinkerbell dress up accessory and watch her flutter excitedly for hours in her tutu. Very soft elastic waist that is trimmed in satin and stretches.”* Such findings can enable marketers to design specific marketing campaigns that appeal to a particular group of users.

The contribution of this work is three-fold: (1) We take into account the consumption preferences to bridge the explanatory gaps between personality traits and product attributes. It becomes feasible to determine and explain users’ purchase behavior based on the personality-driven traces. With our personality-driven traces approach, we are able to better explain users’ purchase behavior. (2) We provide a computational approach to automatically predict potential purchases from inferred personalities traits of users. This approach could be particularly useful when little knowledge is available about users’ purchase history and demographics, or when conducting surveys is not practical. (3) We describe a case study that shows our personality-driven traces approach is useful for product recommendation and user behavior analysis.

There are several interesting directions for generalizing and extending our current work. First, our approach can be integrated with existing product recommendation systems [19] that use demographic and identity attributes of consumers for making purchase predictions, to provide supplementary interpretation and explanation of the recommendation results. Second, our model is not limited to the selected set of consumption preferences and products, and our current study is an initial evaluation of our approach in one shopping category. We plan to conduct further studies to investigate the interactions between personality and other factors (such as product category and users’ demographics) that would influence purchase behavior, and compare our approach with advanced collaborative filtering methods. Third, it would be interesting to extend our approach to other domains that are beyond binary decisions. Finally, we would like to take a visual analytic approach to explore and understand the patterns of traces in the likelihood trace graph, which can provide more insights into the interactions between personality, preferences and product attributes.

References

[1] BOYETTE, N. H. https://www.ibm.com/blogs/watson/2016/10/leveraging-personality-predict-consumption-preferences, 2016.

[2] CAVNAR, W. B., TRENKLE, J. M., ET AL. N-gram-based text categorization. *Ann Arbor MI 48113*, 2 (1994), 161–175.

[3] CHEN, J., HSIEH, G., MAHMUD, J. U., AND NICHOLS, J. Understanding individuals' personal values from social media word use. In *ACM Conference on Computer-Supported Cooperative Work and Social Computing* (2014).

[4] CHEN, M.-F. Consumer attitudes and purchase intentions in relation to organic foods in taiwan: Moderating effects of food-related personality traits. *Food Quality and preference 18*, 7 (2007), 1008–1021.

[5] COSTA, P. T., AND MCCREA, R. R. *Revised neo personality inventory and neo five-factor inventory*. Psychological Assessment Resources, 1992.

[6] FAST, E., CHEN, B., AND BERNSTEIN, M. Empath: Understanding topic signals in large-scale text. *arXiv preprint arXiv:1602.06979* (2016).

[7] FORD, J. K. *Brands laid bare: Using market research for evidence-based brand management*. John Wiley & Sons, 2005.

[8] JANNACH, D., ZANKER, M., FELFERNIG, A., AND FRIEDRICH, G. *Recommender systems: an introduction*. Cambridge University Press, 2010.

[9] KASSARJIAN, H. H. Personality and consumer behavior: A review. *Journal of marketing Research* (1971), 409–418.

[10] KOTLER, P., AND ARMSTRONG, G. *Principles of marketing*. pearson education, 2010.

[11] MAHMUD, J. https://developer.ibm.com/watson/blog/2015/12/08/pi-machine-learning, 2015.

[12] MALHOTRA, N. K. Self concept and product choice: An integrated perspective. *Journal of Economic Psychology 9*, 1 (1988), 1–28.

[13] MCAULEY, J., PANDEY, R., AND LESKOVEC, J. Inferring networks of substitutable and complementary products. In *SIGKDD* (2015).

[14] MIKOLOV, T., SUTSKEVER, I., CHEN, K., CORRADO, G. S., AND DEAN, J. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (2013), pp. 3111–3119.

[15] PENNEBAKER, J. W., BOYD, R. L., JORDAN, K., AND BLACKBURN, K. The development and psychometric properties of liwc2015. *UT Faculty/Researcher Works* (2015).

[16] SCHLEGELMILCH, B. B., BOHLEN, G. M., AND DIAMANTOPOULOS, A. The link between green purchasing decisions and measures of environmental consciousness. *European journal of marketing 30*, 5 (1996), 35–55.

[17] SCHWARTZ, S. H. Basic human values: Theory, measurement, and applications. *Revue Francaise de Sociologie* (2006).

[18] YARKONI, T. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality 44*, 3 (2010), 363–373.

[19] ZHAO, X. W., GUO, Y., HE, Y., JIANG, H., WU, Y., AND LI, X. We know what you want to buy: a demographic-based system for product recommendation on microblogs. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (2014), ACM, pp. 1935–1944.