ASSOCIATION FOR CONSUMER RESEARCH

Association for Consumer Research, University of Minnesota Duluth, 115 Chester Park, 31 West College Street Duluth, MN 55812

Predicting Consumer Brand Recall and Choice Using Large-Scale Text Corpora

Zhihao Zhang, University of California Berkeley, USA

Aniruddha Nrusimha, University of California Berkeley, USA

Ming Hsu, University of California Berkeley, USA

We present a novel approach to predict core aspects of consumer memory by leveraging advances in machine learning (ML) and natural language processing (NLP). Specifically, we predict the likelihood that consumers will recall specific brands within a product category using word embeddings models trained on large scale text corpora.

[to cite]:

Zhihao Zhang, Aniruddha Nrusimha, and Ming Hsu (2018) ,"Predicting Consumer Brand Recall and Choice Using Large-Scale Text Corpora", in NA - Advances in Consumer Research Volume 46, eds. Andrew Gershoff, Robert Kozinets, and Tiffany White, Duluth, MN : Association for Consumer Research, Pages: 204-207.

[url]:

http://www.acrwebsite.org/volumes/2412464/volumes/v46/NA-46

Big Data Approaches to Consumer Behavior

Chair: Christopher Olivola, Carnegie Mellon University, USA

Paper #1: Testing Theories of Goal Progress Within Online Learning

Joy Lu, University of Pennsylvania, USA
Eric T. Bradlow, University of Pennsylvania, USA
J. Wesley Hutchinson, University of Pennsylvania, USA

Paper #2: Semantic Processes in Memory-Based Consumer Decision Making

Sudeep Bhatia, University of Pennsylvania, USA

Paper #3: Predicting Consumer Brand Recall and Choice Using Large-Scale Text Corpora

Zhihao Zhang, University of California, Berkeley, USA
Aniruddha Nrusimha, University of California, Berkeley, USA
Ming Hsu, University of California, Berkeley, USA

Paper #4: Data-Driven Computational Brand Perception

Sudeep Bhatia, University of Pennsylvania, USA
Christopher Olivola, Carnegie Mellon University, USA

SESSION OVERVIEW

The past decade has witnessed the explosion of 'Big Data', and with it, many novel opportunities to study and model human cognition and behavior. Large-scale datasets of human activity "mined" from the Internet, in particular, have proven to be extremely useful for studying various aspects of human cognition. The papers in this session demonstrate that this approach can also be useful for studying consumer behavior. Specifically, they showcase various ways in which Big Data, combined with novel computational methods, can be used to derive insights about various aspects of consumer cognition and behavior.

The first paper in this session, by Lu, Bradlow, and Hutchinson, develops and tests a model of goal progress that captures consumptions decisions related to online content, with parameters that can be mapped to specific theories from consumer psychology. They apply this model to large dataset of Coursera.com learners, and find that their consumption patterns are consistent with theories of goal gradient and resource slack. Moreover, they show that the model allows them to predict changes in consumption behavior when content release shifts from weekly installments (drip-release) to on-demand (all-at-once).

The next three papers utilize large scale text corpora drawn from the Internet, combined with machine learning and natural language processing techniques, to model and predict various aspects of consumer cognition and decision making.

First, Bhatia demonstrates that this approach can be used to predict the items that come to consumers' minds when they need to construct choice sets from memory. Moreover, he shows the approach can be applied to a number of specific consumption domains (e.g., food shopping, gift-giving ,etc.), and that it successfully predicts the responses of participants across six preregistered experiments.

Next, Zhang, Nrusimha, and Hsu demonstrate that this approach can be used to predict the likelihood that consumers will recall specific brands (e.g., Coke) within a product category (e.g., soft drink). They verify the accuracy of this approach by comparing its predictions to the responses of a large sample of participants.

Finally, Bhatia and Olivola demonstrate that this approach can be used to predict how entire product categories and individual brands are perceived along various trait dimensions by consumers. Specifically, they show that, across a wide range of product categories (e.g., beauty products, cars, etc.), across specific brands within each category (e.g., for cars: Mazda, Hyundai, etc.), and across evaluative trait dimensions (e.g., competence, excitement, etc.), their model of brand trait perceptions is consistently well correlated with the judgments of participants in studies of brand perceptions.

In sum, the papers presented in this session showcase highly novel approaches to studying various aspects of consumer cognition and behavior. In particular, marketing researchers and practitioners can use the growing availability of 'Big Data', in combination with recently developed computational methods, to greatly advance consumer research. This session will appeal to marketing researchers interested in learning about cutting-edge approaches to studying consumer behavior.

Testing Theories of Goal Progress Within Online Learning

EXTENDED ABSTRACT

Online education has experienced rapid growth and change within the last decade. Firms like Coursera and edX offer both free and monetized content, and partner with universities to offer specializations and degrees. The setting of online education has unique features such as scheduled content and learning assessments that make it an appropriate testing ground for theories of how consumers progress towards goals.

We examine the behavior of 508 individuals engaged in two courses offered by Wharton Online through Coursera in 2015: Introduction to Marketing ("Marketing") and Introduction to Operations Management ("Operations"), each consisting of four weeks of lecture videos and quizzes. We selected for individuals who had paid for both course certificates when the two courses were offered during the same four weeks. We build a model that captures decisions about whether to consume Marketing or Operations, whether the content is a lecture or a quiz, and when to take breaks of different lengths. On average, we observe 75.4 (SD = 45.2) of these decisions per person. The parameters of our model can be mapped to specific theories from consumer psychology, which provides an underlying behavioral interpretation for the patterns we find in the data.

The key feature of our model is the "Goal Progress" construct that we embed into the utility specifications of choosing Marketing and Operations, with each individual's goal assumed to be to visit all available lectures and quizzes within a given course. For example, let $G_M[j]$ represent "progress" or the percentage of available content in Marketing visited by choice j. The Goal Progress construct for Marketing is given by the expression $\beta_{1M}G_M[j] + \beta_{2M}G_M[j]^2$. Note that $G_M[j]$ varies as individuals progress towards their goal by visiting more content, as well as when more content is released each week, which resets the goal by increasing the total amount of available content.

Because the Goal Progress construct has both linear and quadratic components, utility can flexibly vary with progress in different ways. Individuals may experience fatigue or satiation and exhibit decreasing utility (Inman 2001; Nelson and Meyvis 2008). In contrast, goal gradient theory (Kivetz, Urminsky, and Zheng 2006) predicts an acceleration of effort as individuals approach completion of their goal. The stuck-in-the-middle effect expands upon this by suggesting that individuals may also be motivated to move away from the starting line, resulting in a dip in utility near the middle (Bonezzi, Brendl,

and De Angelis 2011). Finally, individuals may exhibit a peak early on with a decline towards the end because they overestimated their future slack for time resources (Zauberman and Lynch 2005).

To estimate the parameters of our model (e.g., β_{1M} and β_{2M}), we use a hierarchical Bayes procedure, which allows us to account for unobserved heterogeneity across individuals (Gelman et al. 2014). We obtain a set of parameters for each individual that comes from a multivariate normal distribution for the population. Thus, for both Marketing and Operations, we can use the estimated parameters from the Goal Progress construct to classify each individual as a specific type of learner for each course. In Marketing, the linear and quadratic coefficients have population means of β_{1M} = 0.67 (CI = [0.57, 0.78] and β_{2M} = -0.36 (CI = [-0.60, -0.09]), respectively, which corresponds to a small goal gradient effect. In contrast, for Operations, the linear and quadratic coefficients have population means of β_{1O} = 0.49 (CI = [0.33, 0.72]) and β_{2O} = -4.09 (CI = [-4.48, -3.69]), respectively, which corresponds to a strong resource slack effect. Compared to Marketing, the Operations content was more technical and built up across weeks, which may have resulted in attrition as learners found the material increasing in difficulty.

Since media firms are increasingly making content available "on-demand" (i.e., available all at once instead of in weekly installments), we use the parameter estimates of our model to conduct a counterfactual simulation where all course content was available starting on the first day. Our model predicts a sharp shift in activity towards the first weeks of the course, since individuals now have one overarching goal to complete all of the content, rather than waiting for the goal to be reset each week. To empirically verify our predictions, we took advantage of a policy change where the Coursera platform transitioned the courses from weekly to on-demand release in 2016. Using a new dataset from after this policy change, consisting of 1,907 individuals who matched the criteria of our original sample, we find that our predictions are consistent with observed changes in learner behavior.

Semantic Processes in Memory-Based Consumer Decision Making

EXTENDED ABSTRACT

Many common decisions do not involve fixed, exogenous sets of choice items. Rather, consumers must construct the choice sets by themselves, typically through the use of memory processes (see Alba and Hutchinson 1987; Lynch and Srull 1982). The key role of memory in generating choice sets in consumer decision-making raises a number of important questions at the intersection of memory and consumer research. From a theoretical perspective: What are the mechanisms that determine the items that are retrieved by consumers when exogenous choice sets are not provided? How do these mechanisms relate to core memory processes known to play a role in non-preferential choice tasks, and do these memory processes facilitate or hinder efficient memory retrieval for consumer decision-making? Practically, can the mechanisms at play in memory-based decision-making be tested? The set of retrieved choice items in everyday consumer decisions is completely unconstrained –any choice item can come to mind, and the items that do come to mind often lack a clear category structure. So how can the relationship between the various retrieved items, and between these items and other relevant variables (such as choice context), be quantified?

I attempt to address these questions using existing insights on human memory combined with novel techniques from machine learning and data science. The task of retrieving feasible sets of choice items from memory has similarities to well-studied cognitive tasks, such as free-recall and free-association. Thus it is likely that both the mechanisms that guide retrieval in these tasks, as well as the effects generated by these mechanisms, carry over to the domain of consumer decision-making. For example, as with free-recall and free-association, the generation of memory-based choice sets may involve associative activation processes. This would cause memory-based choice sets to display semantic clustering, with retrieved items increasing the retrieval probability of other semantically related items (Bousfield and Sedgewick 1944; Gruenewald and Lockhead 1980; Romney et al. 1993). For this reason, we would also expect retrieved items to depend on contextual cues, such as choice context, with items semantically related to these cues being more likely to be retrieved (Hare et al. 2009; Moss et al. 1995; Nelson et al. 2004).

Moreover, it may also be possible to apply novel methodological tools used to study memory processes in free-recall and free-association tasks to the domain of consumer choice. For example, recent work has shown that new techniques from machine learning and data science, such as semantic space models, are able to quantify the semantic similarity between items, and between items and contextual cues (Hills et al. 2012; Howard and Kahana 2002; see also Bhatia 2017). These models possess representations for a very large set of objects and concepts, implying that they can also be used to measure the semantic relationships at play when consumers must generate choice sets from memory.

I tested the applicability of semantic space models for studying semantic clustering effects and context effects in memory-based consumer choice in six preregistered experiments. In Experiments 1A-1C participants were shown a description of a consumer choice setting and asked to list any 20 items that came to their mind as they considered making their decisions. Specifically, the settings were: food choice (Experiment 1A), vacation choice (Experiment 1B), and purchasing a gift for someone else (Experiment 1C). After these items were listed, participants were taken to a second screen on which they rated each of their 20 items in terms of desirability, on a scale from -3 to +3.

I use word2vec, a well-known semantic space model, to quantify the semantic distance between each pair of listed choice items for each participant (Mikolov et al. 2013). I subsequently analyzed semantic clustering using the path analysis method proposed by Romney et al. (1993). This method involves measuring the total semantic distance (in word2vec space) between each pair of adjacently listed items in a participant's list, and comparing this distance to a random path on the same list. Using this method I found that listed paths were significantly shorter than random paths effects in all three settings, demonstrating strong semantic clustering. Listed paths were also shorter than the hypothetical paths that would have been generated had items been listed in order of desirability. This indicates that semantic clustering leads to some degree of suboptimality in memory retrieval, with undesirable items being retrieved significantly earlier than had decision makers been able to retrieve items strictly in order of desirability.

In Experiments 2A-2C, I examined the influence of consumer choice context on the retrieval of items. These experiments considered the same choice domains as Experiments 1A-1C but varied the contextual cues given to participants: Breakfast vs. dinner for Experiment 2A, wine tasting vs. camping trip for Experiment 2B, and baby shower vs. Valentine's day gift for Experiment 2C. Each participant was only shown one context and asked to list any 20 items that came to mind. I again used word2vec to quantify semantic distance between pairs of listed items, and found robust semantic clustering effects (replicating Experiments 1A-1C). I also used this method to calculate semantic similarity to the consumer choice context, and

found that participants were much more likely to list items that were semantically related to the context in consideration. The effect of consumer context was most pronounced for items listed earlier.

Overall, these experiments demonstrate that semantic processes play a key role in the formation of memory-based consumer choice sets, by generating both semantic clustering and context dependence. The experiments also show how decisions involving everyday consumer purchases can be studied with the use of semantic space models. These models make it is possible to obtain representations for a very large set of words and concepts. These representations, in turn, provide measures of semantic similarity for nearly any pair of choice items, and between choice items and a wide range of choice contexts, allowing quantitative analyses of item retrieval in unconstrained memory-based consumer decision settings.

Predicting Consumer Brand Recall and Choice Using Large-Scale Text Corpora

EXTENDED ABSTRACT

Consumer memory processes have long been recognized as playing an important role in mediating the effect of marketing actions on consumer behavior (Hoyer and Brown 1990; Lynch 1991; Nedungadi 1990; Posavac, Sanbonmatsu, and Fazio 1997). Brand managers often strive to foster brand awareness, increase brand accessibility, and create favorable brand images. Consumer memory has become an integral component of consumer-based brand equity (Christodoulides and de Chernatony 2010)@, as well as related metrics such as brand salience, brand awareness, and top-of-mind widely used by marketers (Farris 2010).

Despite its importance and significant advances in the theoretical and empirical understanding of the relationship between consumer memory and choice behavior, measures of consumer memory have remained dependent on survey-based self-report methods. This reliance on verbal self-reports places inherent limitations on the ability of researchers and practitioners to scale across product categories and markets, or to integrate with the large amounts of purchase data that are routinely available within firms.

This paper proposes to take a step toward addressing these challenges by leveraging advances in machine learning (ML) and natural language processing (NLP) to draw inferences about core aspects of consumer memory from large text corpora. Specifically, we show that it is possible to develop predictive models of consumer brand recall using a type of models from NLP. A well-established theoretical framework for semantic memory is that concepts (e.g. categories, brands, products) are organized as nodes in an associative network, where the links between any two nodes represent their associations (Collins and Quillian 1969). When one concept (e.g. a category) is processed, the activation will be spread to other concepts (e.g. brands within this category) to the extent that they are closely related to the previous concept (Collins and Loftus 1975). The stronger the association, the higher the probability of the successful retrieval of the latter concepts (Anderson 1983). Recent advances in NLP have made it possible to uncover such semantic relations and associations between words from large text corpora quantitatively, using so called word embeddings models (Mikolov, Sutskever, Chen, Corrado, and Dean 2013). Such models represent words and phrases (‘tokens’) as vectors in high dimensional spaces in a way that the spatial distances between the vectors reflect the semantic relatedness between the tokens. This “word to vector” (or *word2vec*) approach has been shown to improve the performance of machine learning models in applications such as translation and conversation beyond simple co-occurrence statistics. Successful applications of word embeddings also include using the uncovered associations to understand more complicated phenomena, such as stereotypes (Caliskan, Bryson, and Narayanan 2017) and cognitive biases (Bhatia 2017). For example, male words are found to be more associated than female words with mathematics then with arts (Caliskan et al. 2017).

To demonstrate our approach, we examined the extent to which word embeddings trained on large text corpora can be used to make predictions about average brand recall success (e.g., “Coke”) for specific product categories (e.g., “Soft drink”). In three batches of aided recall experiments on Amazon Mechanical Turk (N = 120/batch), we asked participants to type down all brands that came to mind when prompted with a category cue. We included a diverse range of product categories, including both consumer packaged goods (batteries, beer, bottled water, breakfast cereals, chewing gum, fast food, Greek yogurt, orange juice, potato chips, and soft drink), as well as durables (headphones, luxury cars, and laptop computers), and services (auto insurance, gas stations, and hotels). The choice of categories was primarily based on extant literature on brand memory and focused on categories for which memory factors play a crucial role in purchase decisions (Dickson and Sawyer 1986; Lynch 1991). We also avoided categories where most brand names are polysemous, such as airlines and banks (e.g., United, Delta, Chase, Fidelity).

We define the total rate of recall for a certain brand in a given category to be the percentage of participants whose list of brands when prompted with the category included that brand. We used the standard *Google News word2vec* Vectors (3 million English word tokens) trained on the Google News corpus (3 billion running words) (Mikolov et al. 2013). The association between any two such vectors is defined as the cosine distance between them, following practices in existing studies (Bhatia 2017; Mikolov et al. 2013).

We show that mean recall rates are significantly associated with similarity as assessed using cosine distance between brand and category, consistent with the hypothesis that word embeddings capture association strengths. Furthermore, two consistent trends emerged across categories. First, the brands with the highest rate of recall almost always have very high similarities with the categories they belong to in word embeddings. Second, the brands that have low cosine similarities tend to have low recall rate. Interestingly, word embeddings cosine similarity seems to track brand recall in a nonlinear manner, resembling a power function. Few brand names are highly recalled without high cosine similarity to the categories, almost all of them being brands which produce products spanning multiple categories (for example, “Samsung” has relatively low similarity to “television” despite a high recall rate).

By showing a relationship between category-brand associations predicted by word embeddings and empirical recall data, we demonstrate the great potential of applying this novel tool to better understand and quantify brand memory. Using large-scale real-world text corpora, word embeddings provide a high-quality abstraction of the informational environment that the general population is exposed to, from which quantitative inference could be drawn. The availability of off-the-shelf word embeddings models like *word2vec* also makes it a highly accessible and flexible tool for marketers. Further work is needed to improve its performance for multi-category brands, better address individual differences in brand recall, and to link brand memory predictions to purchase behavior.

Data-Driven Computational Brand Perception

EXTENDED ABSTRACT

A key determinant of a brand’s value (and equity) is, of course, how it is *perceived* by consumers --i.e., how it is represented in their

minds. There are numerous pathways through which brand perceptions (and mental representations) influence brand value (Keller 2003, Schmitt 2012). At the most basic, and obvious, level, brands are valued to the extent that they are seen as having desirable characteristics, such as being associated with trustworthy, reliable, and competent firms and products (Keller 2003). These positive associations, in turn, contribute to a brand's intangible qualities, increasing purchase likelihood and customer loyalty (Keller and Lehman 2006). Moreover, companies themselves have corporate brands that reflect how they are perceived by consumers (e.g., Aaker et al. 2010, Bhattacharjee et al. 2017, see also Keller and Lehman 2006). Therefore, an important goal for both firms and marketing researchers, is to be able to measure how individual brands, and product categories, are perceived by consumers (e.g., Aaker 1997).

Although brand perception has long been, and continues to be, a core topic in marketing, the methods for studying it have largely remained the same: to gauge how brands are evaluated on specific dimensions of interest (e.g., masculinity, sophistication, etc.), researchers typically collect responses from human participants (e.g., numerical ratings), then average these to estimate aggregate brand perceptions. This general approach can be costly (both in terms of money and time spent), slow (if one needs to estimate perceptions for a very large set of brands and/or evaluation dimensions), and often relies on relatively small samples of human judges (who are rarely representative of the broader population) to draw inferences about market-level brand perceptions.

We introduce a novel approach to predicting and mapping brand trait perceptions using 'Big Data' mined from the Internet combined with machine learning techniques. Recent developments in computational linguistics have made it possible to uncover human-like associations between a very large range of objects and concepts, including popular brands and various traits. We show such techniques allow us to predict how entire product categories and individual brands are perceived by consumers along various trait dimensions by consumers.

We used a prominent prebuilt set of vector representations in order to predict brand-trait associations. The representations we used were generated by the Word2Vec model, trained using the continuous bag-of-words and skip-gram techniques on a corpus of Google News articles with over 100 billion words tokens. These representations have a vocabulary of 300 million words and phrases, with each word or phrase being defined on 300 dimensions. We computed the association between each brand and trait in our dataset, as assessed by the cosine similarity between their corresponding vector representations.

We find that, across a wide range of product categories (e.g., beauty products, cars, etc.), specific brands within each category (e.g., for cars: Mazda, Hyundai, etc.), and evaluative trait dimensions (e.g., competence, excitement, etc.), our model of brand trait perceptions is consistently well correlated with the judgments of participants in previous studies of brand perceptions.

In sum, we propose a totally different approach to estimating the way brands are represented in the minds of consumers. In particular, marketing researchers and practitioners can use the contents of the Internet, in combination with computational methods, to measure and map the representations of brands in the collective consumer mind. This computational, big-data approach presents many advantages, such as allowing researchers to rapidly collect data on a huge range of products across a large set of evaluation dimensions.

REFERENCES

Aaker, Jennifer (1997), "Dimensions of Brand Personality," *Journal of Marketing Research*, 34 (3), 347-356.

Aaker, Jennifer, Kathleen D. Vohs, and Cassie Mogilner (2010), "Nonprofits Are Seen as Warm and For-Profits as Competent: Firm Stereotypes Matter," *Journal of Consumer Research*, 37 (2), 224-237.

Alba, Joseph W., and J. Wesley Hutchinson (1987), "Dimensions of Consumer Expertise," *Journal of Consumer Research*, 13 (4), 411-454.

Anderson, John R. (1983), "A Spreading Activation Theory of Memory," *Journal of Verbal Learning and Verbal Behavior*, 22 (3), 261-295.

Bhattacharjee, Amit, Jason Dana, and Jonathan Baron (2017), "Anti-Profit Beliefs: How People Neglect the Societal Benefits of Profit," *Journal of Personality and Social Psychology*, 113 (5), 671-696.

Bhatia, Sudeep (2017), "Associative Judgment and Vector Space Semantics," *Psychological Review*, 124 (1), 1-20.

Bonezzi, Andrea, C. Miguel Brendl, and Matteo De Angelis (2011), "Stuck in the Middle: The Psychophysics of Goal Pursuit," *Psychological Science*, 22 (5), 607-612.

Bousfield, W. A., and C. H. W. Sedgewick (1944), "An Analysis of Sequences of Restricted Associative Responses," *Journal of General Psychology*, 30 (2), 149-165.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan (2017), "Semantics Derived Automatically From Language Corpora Contain Human-Like Biases," *Science*, 356 (6334), 183-186.

Christodoulides, George, and Leslie de Chernatony (2010), "Consumer-Based Brand Equity Conceptualisation And Measurement: A Literature Review," *International Journal of Market Research*, 52 (1), 43-66.

Collins, Allan M., and Elizabeth F. Loftus (1975), "Spreading Activation Theory of Semantic Processing," *Psychological Review*, 82 (6), 407-428.

Collins, Allan M., and M. Ross Quillian (1969), "Retrieval Time from Semantic Memory," *Journal of Verbal Learning and Verbal Behavior*, 8 (2), 240-247.

Dickson, Peter R., and Sawyer, Alan G. (1986). *Point-Of-Purchase Behavior And Price Perceptions Of Supermarket Shoppers*. Marketing Science Institute.

Farris, Paul W., Neil Bendle, Phillip Pfeifer, and David Reibstein (2010). *Marketing Metrics : The Definitive Guide To Measuring Marketing Performance (2nd ed.)*. Upper Saddle River, N.J.: FT Press.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubi (2014), *Bayesian Data Analysis, 3rd ed.* Boca Raton, FL: CRC Press.

Gruenewald, Paul. J., and Gregory R. Lockhead (1980), "The Free Recall of Category Examples," *Journal of Experimental Psychology: Human Learning and Memory*, 6 (3), 225-240.

Hare, Mary, Michael Jones, Caroline Thomson, Sarah Kelly, and Ken McRae (2009), "Activating Event Knowledge," *Cognition*, 111 (2), 151-167.

Hills, Thomas T., Michael N. Jones, and Peter M. Todd (2012), "Optimal Foraging in Semantic Memory," *Psychological Review*, 119 (2), 431-440.

Howard, Marc W., and Michael J. Kahan (2002), "When Does Semantic Similarity Help Episodic Retrieval?" *Journal of Memory and Language*, 46 (1), 85-98.