A Schema-oriented Product Clustering Method Using Online Product Reviews

Research-in-Progress

Zhuolan Bao
School of Business
Faculty of Business and Economics
The University of Hong Kong
Hong Kong
mbao@connect.hku.hk

Michael Chau
School of Business
Faculty of Business and Economics
The University of Hong Kong
Hong Kong
mchau@business.hku.hk

Abstract

In online markets, with the convenient and extensive information search, conventional classification methods cannot afford a precise understanding of products. This research draws on comprehension research and posits that the perceptual schema used by consumers to comprehend product information varies for different products. As product reviews are a major source of product-related information, we use product review perception to derive the perceptual schemas. In the paper, we present our three-step method in detail and use it to generate preliminary product clusters. As an exploration of product classification, this research contributes in several perspectives. First, our generated clusters help understand consumer behaviors towards different products. Second, we provide schema prototypes which depict consumers' perceptual sets towards different target products, contributing to both research and practices of online markets. Third, instead of a top-down approach of classifying products, our bottom-up method provides insights of using and mining the value of online textual content.

Keywords: Product clustering, schema theory, online product reviews, helpfulness perception

Introduction

Product classification is important for people to understand different types of products. Looking at whether the product's quality can be inspected or determined prior to purchase, traditional product classification divides products into two broad types: experience products and search products (Nelson 1970). With the prosperity of online markets, nowadays more and more products are available online. In the new context, the boundary between products' search and experience attributes is getting blurry due to less search cost and diverse product innovation (Huang et al. 2009; Klein 1998). As such, the binary classification for products, such as experience vs. search products is often insufficient and current product classification methods could not afford a precise understanding of product features.

For example, jewelry was classified as experience goods according to Nelson (1970). But nowadays, the detailed specification of jewelry, image or 3-D product presentation and thorough reviews provided by prior buyers together present various facets of the jewelry, dramatically reducing the search costs of the dominant attributes. It is hard to put jewelry into either search or experience products type.

In this study we draw on comprehension research and offer a product clustering method. In comprehension research, there is a debate on whether meaning of a piece of text is conventionalized to the textual content or interpreted based on the readers' knowledge (Mick 1992). Since the act of reading reviews is usually performed deliberately by potential consumers, we examine how they comprehend product reviews from the perspective of subjective comprehension.

Schema theory posits that people interpret, infer and expect information based on their generic knowledge structure (Graesser and Nakamura 1984). Based on this theory, we propose that for different product categories, people construe different perceptual schemas to comprehend the product information. To further identify the perceptual schema, we make use of consumer reviews displayed in online markets to extract textual features for comprehension of product information.

In this research, we propose a three-step clustering method which enables an exploratory approach to identifying schemas and clustering products accordingly. First, we structuralize the textual features in each piece of review. Second, we identify the focal features for each product category. Third, we cluster product categories based on their collections of textual features. With this method, we aim to contribute to a fine-grained understanding of products as well as improving the current methods to classify product types in online markets. Also, the results of our clustering outcome could offer insights on both theory development and practical implementation.

The remaining part of the paper is arranged as follow. In the next section, we introduce the current research on product classification. Then we review literature of comprehension research and discuss the usage of online product review perception. In the methodology part, we propose the mechanism of our clustering process, describe our new method in detail and show a preliminary result with empirical review data from Amazon. In the last section, we discuss our potential contributions as well as limitations of this study.

Theoretical Background

Product Classification

There are various ways to classify products. A physical product incorporates both function and form (Alexander 1964). The function defines "how the product works", which can be evaluated by the product specifications or involved techniques, while the form represents "how the product looks", which can be evaluated by the perception of consumers (Rafaeli and Vilnai-Yavetz 2004).

One way to classify products is by the product's utilitarian and hedonic benefits (Chitturi et al. 2008; Dhar and Wertenbroch 2000; Hirschman and Holbrook 1982). Another way commonly used by marketing and IS researchers is from the perspective of the economics of information and advertising. Nelson (1970) classified product attributes into two general types, search and experience attributes, which is based on the extent to which consumers can evaluate the products' attributes prior to purchase. If the attributes could be inspected prior to purchase, the product is called a search product. And if the attributes could not be inspected before purchase but can be evaluated after using or experiencing it, the product is called an experience product. In their classification scheme, they involved a variety of products. Examples of experience goods are Automobiles, Appliances, Bicycles, Jewelry, and Musical Instruments and examples of search goods are Garden Implements, Sporting Goods, Cameras, and Hobbies and Games. Soon later, Darby and Karni (1973) proposed an additional category, credence attribute/product, which is unable for consumers to evaluate even after use, such as car repair and organic food (Liebermann and Flint-Goor 1996). Norton and Norton Jr (1988) extended the classification of Nelson (1970) by defining durable and non-durable experience products. As more frequent purchase of non-durable products can result in more intensive processing of the relevant information, experience attributes can be gradually changed to search attributes.

These efforts in this product classification are important for marketing strategies. The classification of search, experience, and credence products is widely adopted by consumer behavior research, such as advertising reliance (Ford et al. 1990), consumer information search (Klein 1998), information perception (Huang et al. 2009; Mudambi and Schuff 2010) and so on.

However, owing to the Internet, products are more easily customized to small groups' preferences (Rigby 2011), and more diverse and recreational products have appeared in online markets (Zhou et al. 2007). Thus, both utilitarian and hedonic value of products have changed and diversified, hence the top-down classification by product functions and features becomes difficult.

Furthermore, online search costs for both sellers and buyers drop dramatically (Bakos 1997). As shown in the jewelry example, it has been hard to draw a line between search products and experience products.

Therefore, since the assumption that every product category can be neatly classified under one particular type does not hold in online markets, the current classification methods for products are too reductive. A better way to classify is to assess the perceptual impact of products from the consumer's perspective. For example, consumers purchasing music instrument and cameras might share the same perceptual sets towards the product information of the two. They might care about product materials, the product appearance, the biological perception (hearing or visual perception) of the products, and even their willingness to share with their friends. In this way, instead of examining products' various functions or forms, a bottom-up research on consumer perception towards products is needed in product classification.

In online markets, people learn and evaluate products through various product-related information. Product review is a main source for potential consumers to understand products. Hong et al. (2014) proposed an approach to measure product attributes based on product reviews, but they only consider quantitative rating information, leaving out the potential large value of review texts. To investigate consumer perceptions during comprehension of product information, in the following part, we draw on comprehension research to investigate consumer perceptions towards product reviews.

Research on Comprehension

In studying reader understanding, there are two competing theoretic perspectives, schema theory and text-centered theory. Strawson (1971) also noted the two perspectives as "communication intention" and "formal semantics" in his linguistic research.

Text-centered theory states that meaning is conventionalized in the sentences of a text (Strawson 1971). Thus, the text itself is autonomous and readers are passive recipient of the text (Gere 1980).

On the other hand, schema theory posits that people interpret, infer and expect information based on their generic knowledge structure (Graesser and Nakamura 1984). Schema theory was originally developed by Anderson (1978) in the field of educational psychologist. A schema represents generic knowledge, which is the common truth of a class of things, events, or situations (Anderson 1978). And schema can be viewed as consisting of a set of expectations under certain situation, which guide the information interpretation (Faris and Smeltzer 1997). The reading either does or does not confirm the readers' expectation (Gere 1980). Therefore, a reader's mental representation is an emergent product of the interaction between text-based information and pre-existing knowledge (Whitney 1987).

Similarly, to study the message comprehension in advertising research, scholars conceptualized objective and subjective comprehension as two basic orientations (Mick 1992).

Objective comprehension holds the view that people grasp or extract the advertisement messages in which the meanings are considered intrinsic and intended by the advertiser. Therefore, the focus of the objective comprehension is the reader's ability to extract the meaning in an advertisement.

Subjective comprehension holds that the most important meanings embedded in a text is derived from the recipient with a specific processing context, regardless of whether those meanings were intended by the advertiser (Mick 1992). Therefore, results of subjective comprehension are more open-ended comparing to the objective comprehension.

There are also researchers proposing multiple levels of comprehension (Greenwald and Leavitt 1984; Maclnnis and Jaworski 1989; Mick 1992). In their works, comprehension levels are determined by different attention involvement. Although the concept of involvement was defined from different perspectives, agreement was reached that involvement approximately associates with personal relevance or importance (Mick 1992). Subjective comprehension is the basic assumption in the stream of work, in which lower levels correspond with less usage of perceptual knowledge whereas higher levels correspond with more usage of knowledge. This research stream later converged in the information processing model (McGuire 1978) and rooted the persuasion research (Jacoby and Hoyer 1989; Mick 1992).

Evidence of schema theory and subjective comprehension has been found in various places, such as the match-up effect of spokesperson characteristics and product attributes (Lynch and Schuler 1994), importance of schemas in business communication (Faris and Smeltzer 1997; Suchan and Dulek 1988), congruity proposition of product reviews and product types (Huang et al. 2013) and congruity of advertisement content and format (Bishop et al. 2015).

As review reading is actively done by potential consumers, the purposeful act is consistent with the origination of subjective comprehension. Therefore, we regard subjective comprehension more suitable for sense-making of product reviews. Therefore, we posit that different product stimulates different knowledge set in mind. We put forward our proposition:

Proposition: *Potential consumers comprehend product information with different perceptual schemas for different products.*

To explore the perceptual schemas, we take advantage of product reviews and examine the ways people comprehend online review information.

Product Reviews

Product review is a peer-generated product evaluation posted on company or third party websites, and the reviews have the potential to facilitate purchase decision-making process (Mudambi and Schuff 2010). There is consensus that online reviews are closely associated with seller trustworthiness (Ba and Pavlou 2002), consumer uncertainty reduction (Dimoka et al. 2012), product sales (Chevalier and Mayzlin 2006; Zhu and Zhang 2010) and firms' dynamic pricing strategy (Yu et al. 2015). Reading product reviews is a major means of understanding products without physically experience the products and the act of reading is intentional.

With the adoption of helpfulness voting mechanism, readers of reviews are able to vote for the helpfulness perception of each review. Value of the voting mechanism is multifaceted. Marketers and consumers both benefit from the helpfulness votes. For example, sellers and marketers need to know what aspects of reviews are most informative from consumers' perspective and hence promote products (Ghose and Ipeirotis 2011). Consumers could also be facilitated in product information search and interpretation by using helpful votes (Chen and Lurie 2013; Kuan et al. 2015).

In this research, we take advantage of the value in helpfulness perception from another perspective – deriving the perceptual schema of review comprehension.

Helpful reviews are those facilitating consumers' purchase decision process (Mudambi and Schuff 2010; Pan and Zhang 2011; Yin et al. 2014). According to the past studies, the helpfulness perception is consistent with the notion of information diagnosticity perception (Mudambi and Schuff 2010), which was defined as the degree to which it facilitates a consumer assign a product into one cognitive category (Hoch and Deighton 1989). In other words, a review's helpfulness also reflects the extent to which it facilitates the consumer to diagnose and understand the products. Therefore, review helpfulness perception, as a representative of consumers' cognitive response towards the product review, could be useful for extracting consumers' perceptual schema of product information comprehension.

In studying the determinants, most prior work focuses on the quantitative measurements of review information, such as review rating, review length and review age (Connors et al. 2011; Mudambi and Schuff 2010; Schindler and Bickart 2012; Bao & Chau 2015; 2016). Studies on reviews' textual determinants of helpfulness perception include readability, subjectivity, writing styles as well as discrete emotions, while they either used data of single category, or acknowledged the distinguished effects in review features for different products (Ghose and Ipeirotis 2011; Korfiatis et al. 2012; Liu et al. 2008; Yin et al. 2014).

The framework of the clustering method is shown in Figure 1. In this study, in order to produce product clusters, we use text-mining methods to explore the textual determinants of helpfulness perception based on and beyond these studies on various textual elements. After extracting the content features that are important for consumers to perceive different products, we apply our method to cluster products and identify perceptual schemas.

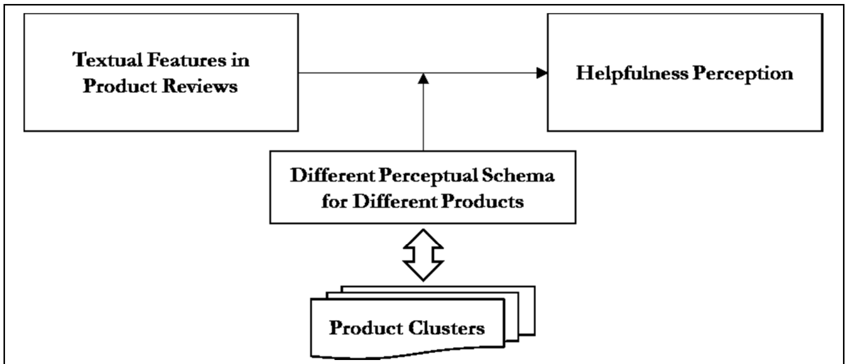

Figure 1. Framework of Schema-Oriented Product Clustering Method

Methodology

Data

The data we will use are Amazon review data, which are collected by the Stanford Network Analysis Project[1] (McAuley and Leskovec 2013). Since Amazon.com is one of the largest online market worldwide, we believe our clustering results based on Amazon reviews will be more easily generalized to other online markets. Before we run the steps described above, we first clean and filter the data with respect to our research goal. The raw data spans over thirteen years, including reviews up to March 2013. But since the voting mechanism was adopted in 2008, we discard the products that were launched before November 2008. Table 1 shows the data of twenty-seven product categories we used for our model.

Table 1. Product Categories in the Data Set

Categories	# of Reviews	# of Products	Categories	# of Reviews	# of Products
Amazon Instant Video	291,273	14,908	Movie	956,292	54,921
Arts	6,851	985	Music	130,969	40,977
Automotive	68,361	19,507	Musical Instruments	28,201	4,585
Baby	731	22	Office Products	37,233	4,110
Beauty	88,811	8,868	Patio	71,210	6,484
Books	477,644	87,667	Pets & Supplies	62,471	4,715
Cellphone & Accessories	5,779	740	Shoes	5,044	466
Clothing & Accessories	8,445	1,098	Software	6,909	1,379
Electronics	182,532	12,952	Sports & Outdoors	101,981	11,518
Gourmet & Food	39,105	5,692	Tools & Home Improvement	113,363	15,616
Health	123,909	12,251	Toys & Games	90,402	11,837
Home & Kitchen	231,472	16,167	Video Games	76,615	7,986
Industrial & Scientific	41,840	6,115	Watches	14,909	1,117
Jewelry	8,691	2,739	**TOTAL**	**3,271,043**	**355,422**

Table 1. Product Categories in the Data Set

[1] http://snap.stanford.edu/index.html

Model and Methods

Since each product type could construe a specific schema and stimulate corresponding perceptual set (Huang et al. 2014; Suchan and Dulek 1988), in this part, we describe the process of extracting schemas from reviews and using them to cluster products.

There are three steps in the process, as shown in Table 2.

Table 2. Clustering Process		
Step	Process	Methods / tools to be used
Step 1	Textual features structuralization	Linguistic Inquiry and Word Count
Step 2	Feature selection	Regression model
Step 3	Product clustering	k-means algorithm

Table 2. Clustering Process

In step one, we first structuralize the textual features of the review content using a dictionary provided by the Linguistic Inquiry and Word Count (LIWC). The reason we use LIWC is that the dictionary has been utilized and considered a robustness tools for analysing texts in prior IS research (Berger and Milkman 2012; Goes et al. 2014; Yin et al. 2014). LIWC was developed by Pennebaker et al. (2007) and designed to calculate the degree to which people use different categories of words across a wide array of texts.

For each piece of text, LIWC counts the number of words that appear in each of the pre-defined word categories. In particular, we focus on six broad psychological dimensions, including social processes, affective processes, cognitive processes, perceptual processes, biological processes and relativity. Those dimensions are further divided into 39 granular psychological dimensions[2]. Take the dimension of social processes for an example, there are three sub-dimensions in it, namely family, friends and humans. In each sub-dimension, there is a list of words. If any word in a piece of review is included in the list, it will be counted under the particular dimension. Similarly, we also look at seven personal concern dimensions, including concerns of work, achievement, leisure, home, money, religion and death. For each review, we calculate the count and percentage (divided by the total word count) of each word category in LIWC.

Next, by using helpfulness perception, we select the feature variables for each product category. Different from usual feature reduction methods such as principal components analysis (PCA), we use regression analysis to explore the association between textual content features and consumer helpfulness perception towards review information and identify the focal content features. The regression we use follows the approach of Kuan et al. (2015) by adopting the Heckman two stage model to analyse all the content features in each product category. In their work, they posited two stages of information processing, an attention stage and a comprehension stage. They examine review's number of voting from the perspective of attention stage, and examine review's helpfulness perception from the perspective of comprehension stage. As the second stage is consistent with our purpose of understanding schemas for product information comprehension, we use the following specifications:

$$Voting_{\ i} = X_i{}'\,\alpha_i + C_i{}'\,\delta_i + \varepsilon$$

$$Helpfulness_i|(Voting_i > k) = X_i{}'\,\beta_i + C_i{}'\,\gamma_i + \varepsilon$$

X_i is the vector of content feature variables for each review and C_i is the vector of control variables. Control variables include quantitative features of a product review, such as review rating, review length, review age, reviewer expertise, product price and so on. They are used to eliminate the influence of other determinants in information comprehension.

We run the above regression model for each product category, identify features that affect the helpfulness perception with appropriate confidence interval. Our dataset is relatively large. To eliminate the effect of different sample sizes, we adopt the method provided by Lin et al. (2013) to validate the coefficients of Heckman model. We first draw a *coefficient/p-value/sample-size (CPS) chart* for each feature variable of each product category using ordinary least squares regression. Example is presented in Figure 2. The

2 For more details, please refer to the LIWC 2007 Manual (http://www.liwc.net/LIWC2007LanguageManual.pdf).

figure shows the CPS plot for affective processes dimension of products of Books. The p-value drops and remains zero beyond some point. To avoid attributing importance to the p-value beyond a certain sample size, we do not include the affective processes dimension for our next step.

The features remained are used as input for our next step of product clustering methods. For the features with positive coefficients, since they influence the helpfulness perception positively, we regard them as favored content in consumers' information perception schema. And for those with negative coefficients, we regard them as un-favored content in perception schema.

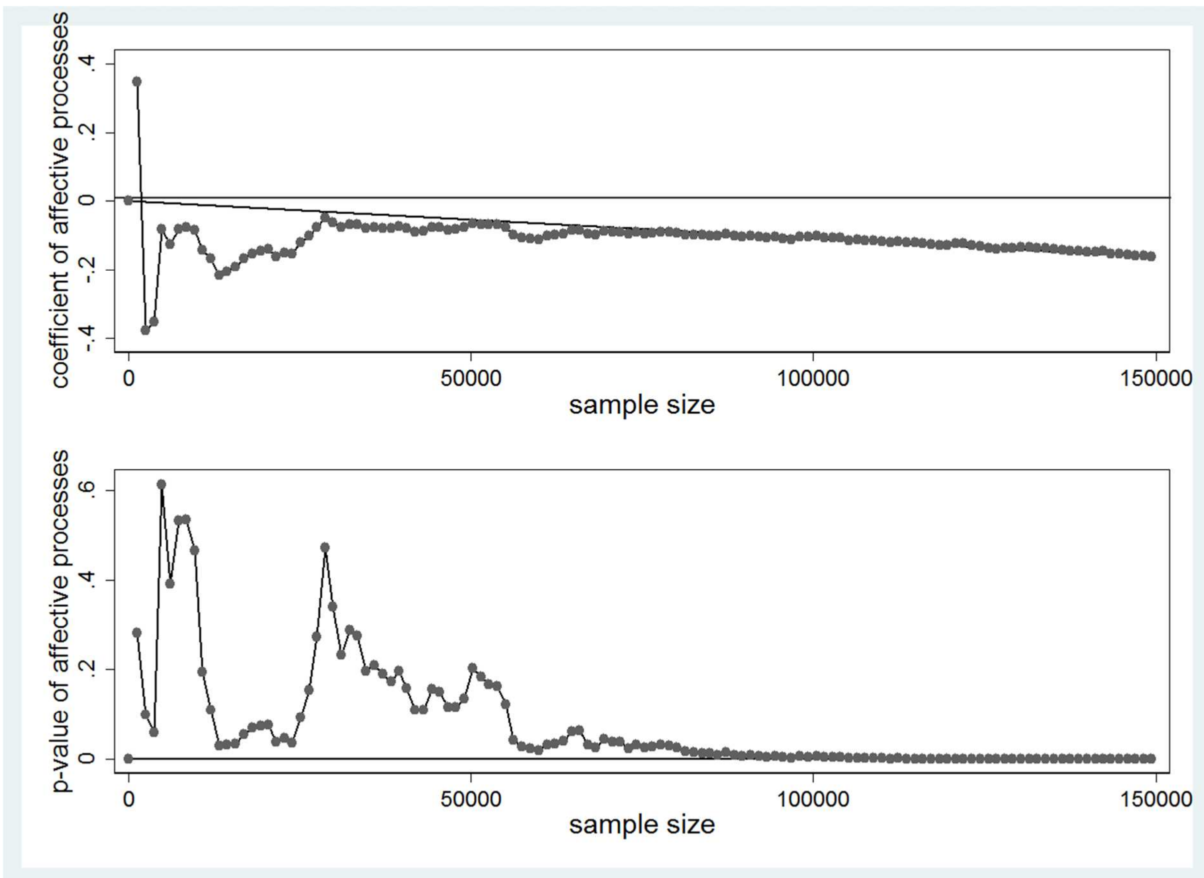

Figure 2. CPS Plot of Social Processes Variable for the Product Category of Books[3]

In the third step, we use the selected features to conduct cluster analysis and generate our product clusters. Our clustering method is intended to be conducted on a large data set and supposed to have high stability, so we use non-hierarchical clustering algorithm to cluster the product categories. K-means is a popular non-hierarchical approach to forming k clusters, where k is a pre-specified desired number of clusters.

The k-means algorithm modifies the partition to reduce the sum of distances of each record from its cluster centroid. The improvement is done repeatedly until the improvement is very small (Huang 1998). To identify the number of initial partitions, we will try a few different values for k and compare the resulting clusters (Shmueli et al. 2007). Here is a preliminary result when k is set to 4.

3 Zoomed in to sample size ≤ 150,000. Horizontal dashed line corresponds to $p = 0.01$.

Table 3. Clustering Results

	Amazon Categories	Products' perceptual schema	Favored content in consumers' perceptual schema	Un-favored content in consumers' perceptual schema
1	Automotive, Gourmet & Food, Sports & Outdoors, Toys & Games	Products for home usage; products stimulating affection.	Affective processes and home-related issue	Social processes and leisure-related issues
2	Arts, Electronics, Health, Home & Kitchen, Industrial & Scientific, Office products, Patio, Pets & Supplies, Shoes, Tools & Home Improvement, Watches	Products for personal achievement and leisure; products stimulating biological reactions.	Biological processes, relativity, achievement-, leisure-, and death-related issues	Religion-related issues
3	Baby, Beauty, Cellphone, Clothing & Accessories, Musical Instruments, Video games	Products that are context-specific, such as time and space.	Relativity	
4	Amazon Instant Video, Books, Movie, Music, Software	Products for home usage; products stimulating perception processes.	Perception processes and home-related issues	Cognitive processes, relativity, work- and money-related issues

Table 3. Clustering Results

After the three steps, we will evaluate the performance of the algorithms. We will investigate the cluster validity and cluster stability (Jain and Dubes 1988; Lange et al. 2004). As Jain (2010) noted that there is no best clustering algorithm, our evaluation tries to ensure a robust and feasible cluster result for our clustering method.

With the resulting non-overlapping clusters, we will be able to identify and interpret perceptual schemas for each product type and provide correspondent schema prototypes. Also, insights will be generated by deeper inspection and discussion of information on within-cluster dispersions and between-cluster distances.

Potential Contributions and Limitations

Potential Contributions

Our research offers both theoretical contributions and practical implications. First, by proposing the clustering method, we advance the current measurement of product classifications in online context. As Internet reduces search costs and conventional method could not afford a precise understanding of products, our method of product clustering delineates products from the perspective of consumers' perception, providing new ideas to understand products from the view of consumers. Also, our method brings insights of using and mining value of online information. Comparing to the top-down categorization of products based on functions and features, we take advantage of the helpfulness information towards product information to cluster them in a bottom-up approach, which we believe will shed light upon future studies using big data.

Besides, we will give product clusters based on product review data from Amazon.com. The resulting clusters and schemas could help scholars and marketers from different perspectives. Scholars could utilize the results for further exploring consumer behaviors such as information seeking and interpretation for different types of products. As the schema prototypes depict consumers' perceptual sets towards different target products, they will be beneficial to future exploration and development of schema theory.

For practitioners such as marketers or sellers, they could make use of our clusters to improve their understanding towards potential consumers as well as to optimize their marketing strategies in online environment. The schemas may also guide sellers and platform managers in more powerful presentation. For example, based on the perceptual schema of consumers, sellers could rearrange or redesign their ways of presenting products and hence promote them effectively.

Limitations and Future Work

There are some limitations in our research. First, although we theorize that people have different perceptual schemas towards review texts for different products, our exploration of the schemas is subjected to the effects of text-mining methods. In this paper, we apply LIWC tool to investigate the usage degree of each dimension in LIWC for reviews and extract schemes based on categories developed in the dictionary. However, the choice of these dimensions could be challenged. Thus the extent to which the tools detect and collect schemas accurately can increase the reliability of the results. As an initial exploration of this act, we call for future studies to develop reliable mining tools based on theories to further support the stream of research.

Second, as shown in our preliminary results, our clustering method is applied to data on product category, instead of the data on product. Therefore, to discover the schemas at a finer level, further studies could use our approach for a deeper examination of clustering on product level. Also, with that, future studies could analyze hierarchy of product categories which might provide insights for sellers or platform managers to manage their product categories and sub-categories.

Finally, there are limitations in the feature selection of our clustering method. Currently we only use k-means algorithm, however, other means of feature selection and reduction could be implemented and compared with the current one.

Acknowledgements

This research is supported in part by a HKU Seed Funding for Basic Research grant. We thank Reno Ha of the University of Hong Kong for his help in data processing, and the Stanford Network Analysis Project for making the Amazon data set available for research.

References

Alexander, C. 1964. *Notes on the Synthesis of Form*. Harvard University Press.

Anderson, R. C. 1978. *Schema-Directed Processes in Language Comprehension*. Springer.

Ba, S., and Pavlou, P. A. 2002. "Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior," *MIS Quarterly* (26:3), pp. 243-268.

Bakos, J. Y. 1997. "Reducing Buyer Search Costs: Implications for Electronic Marketplaces," *Management Science* (43:12), pp. 1676-1692.

Bao, Z. and Chau, M. 2015. "In Search of Plagiarism Behaviors: An Empirical Study of Online Reviews," *Proceedings of the Pacific Asia Workshop on Intelligence and Security Informatics*, Ho Chi Minh City, Vietnam, May 19, 2015, Springer, pp. 102-112

Bao, Z. and Chau, M. 2016. "The Effect of Collective Rating on the Perception of Online Reviews," in *Proceedings of the Pacific Asia Conference on Information Systems*, Chiayi, Taiwan, June 27 - July 1, 2016.

Berger, J., and Milkman, K. L. 2012. "What Makes Online Content Viral?," *Journal of Marketing Research* (49:2), pp. 192-205.

Bishop, M. M., Brocato, E. D., and Vijayalakshmi, A. 2015. "The Role of Medium Content and Ad Format Congruity in Influencing Advertising Outcomes," *Journal of Marketing Communications*, pp. 1-14.

Chen, Z., and Lurie, N. H. 2013. "Temporal Contiguity and Negativity Bias in the Impact of Online Word of Mouth," *Journal of Marketing Research* (50:4), pp. 463-476.

Chevalier, J. A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* (43:3), pp. 345-354.

Chitturi, R., Raghunathan, R., and Mahajan, V. 2008. "Delight by Design: The Role of Hedonic Versus Utilitarian Benefits," *Journal of Marketing* (72:3), pp. 48-63.

Connors, L., Mudambi, S. M., and Schuff, D. 2011. "Is It the Review or the Reviewer? A Multi-Method Approach to Determine the Antecedents of Online Review Helpfulness," *44th Hawaii International Conference on System Sciences (HICSS), 2011*: IEEE, pp. 1-10.

Darby, M. R., and Karni, E. 1973. "Free Competition and the Optimal Amount of Fraud," *The Journal of Law & Economics* (16:1), pp. 67-88.

Dhar, R., and Wertenbroch, K. 2000. "Consumer Choice between Hedonic and Utilitarian Goods," *Journal of Marketing Research* (37:1), pp. 60-71.

Dimoka, A., Hong, Y., and Pavlou, P. A. 2012. "On Product Uncertainty in Online Markets: Theory and Evidence," *MIS Quarterly* (36).

Faris, K. A., and Smeltzer, L. R. 1997. "Schema Theory Compared to Text-Centered Theory as an Explanation for the Readers' Understanding of a Business Message," *Journal of Business Communication* (34:1), pp. 7-26.

Ford, G. T., Smith, D. B., and Swasy, J. L. 1990. "Consumer Skepticism of Advertising Claims: Testing Hypotheses from Economics of Information," *Journal of Consumer Research*), pp. 433-441.

Gere, A. R. 1980. "Written Composition: Toward a Theory of Evaluation," *College English* (42:1), pp. 44-58.

Ghose, A., and Ipeirotis, P. G. 2011. "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *Knowledge and Data Engineering, IEEE Transactions on* (23:10), pp. 1498-1512.

Goes, P. B., Lin, M., and Yeung, C.-m. A. 2014. "Popularity Effect" in User-Generated Content: Evidence from Online Product Reviews," *Information Systems Research* (25:2), pp. 222 - 238.

Graesser, A. C., and Nakamura, G. V. 1984. "The Impact of a Schema on Comprehension and Memory," *The Psychology of Learning and Motivation* (16), pp. 59-109.

Greenwald, A. G., and Leavitt, C. 1984. "Audience Involvement in Advertising: Four Levels," *Journal of Consumer Research*, pp. 581-592.

Hirschman, E. C., and Holbrook, M. B. 1982. "Hedonic Consumption: Emerging Concepts, Methods and Propositions," *The Journal of Marketing*, pp. 92-101.

Hoch, S. J., and Deighton, J. 1989. "Managing What Consumers Learn from Experience," *The Journal of Marketing* (53:2), pp. 1-20.

Hong, Y., Chen, P.-y., and Hitt, L. M. 2014. "Measuring Product Type with Dynamics of Online Product Review Variances: A Theoretical Model and the Empirical Applications," *Available at SSRN 2506328*.

Huang, L., Tan, C.-H., Ke, W., and Wei, K.-K. 2013. "Comprehension and Assessment of Product Reviews: A Review-Product Congruity Proposition," *Journal of Management Information Systems* (30:3), pp. 311-343.

Huang, L., Tan, C.-H., Ke, W., and Wei, K.-K. 2014. "Do We Order Product Review Information Display? How?," *Information & Management* (51:7), pp. 883-894.

Huang, P., Lurie, N. H., and Mitra, S. 2009. "Searching for Experience on the Web: An Empirical Examination of Consumer Behavior for Search and Experience Goods," *Journal of Marketing* (73:2), pp. 55-69.

Huang, Z. 1998. "Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values," *Data mining and knowledge discovery* (2:3), pp. 283-304.

Jacoby, J., and Hoyer, W. D. 1989. "The Comprehension/Miscomprehension of Print Communication: Selected Findings," *Journal of Consumer Research* (15:4), pp. 434-443.

Jain, A. K. 2010. "Data Clustering: 50 Years Beyond K-Means," *Pattern Recognition Letters* (31:8), pp. 651-666.

Jain, A. K., and Dubes, R. C. 1988. *Algorithms for Clustering Data*. Prentice-Hall, Inc.

Klein, L. R. 1998. "Evaluating the Potential of Interactive Media through a New Lens: Search Versus Experience Goods," *Journal of business research* (41:3), pp. 195-203.

Korfiatis, N., García-Bariocanal, E., and Sánchez-Alonso, S. 2012. "Evaluating Content Quality and Helpfulness of Online Product Reviews: The Interplay of Review Helpfulness Vs. Review Content," *Electronic Commerce Research and Applications* (11:3), pp. 205-217.

Kuan, K. K., Hui, K.-L., Prasarnphanich, P., and Lai, H.-Y. 2015. "What Makes a Review Voted? An Empirical Investigation of Review Voting in Online Review Systems," *Journal of the Association for Information Systems* (16:1), pp. 48-71.

Lange, T., Roth, V., Braun, M. L., and Buhmann, J. M. 2004. "Stability-Based Validation of Clustering Solutions," *Neural Computation* (16:6), pp. 1299-1323.

Liebermann, Y., and Flint-Goor, A. 1996. "Message Strategy by Product-Class Type: A Matching Model," *International Journal of Research in Marketing* (13:3), pp. 237-249.

Lin, M., Lucas Jr, H. C., and Shmueli, G. 2013. "Research Commentary-Too Big to Fail: Large Samples and the P-Value Problem," *Information Systems Research* (24:4), pp. 906-917.

Liu, Y., Huang, X., An, A., and Yu, X. 2008. "Modeling and Predicting the Helpfulness of Online Reviews," *Eighth IEEE International Conference on Data Mining, 2008. ICDM'08.* : IEEE, pp. 443-452.

Lynch, J., and Schuler, D. 1994. "The Matchup Effect of Spokesperson and Product Congruency: A Schema Theory Interpretation," *Psychology & Marketing* (11:5), pp. 417-445.

Maclnnis, D. J., and Jaworski, B. J. 1989. "Information Processing from Advertisements: Toward an Integrative Framework," *Journal of Marketing*, pp. 1-23.

McGuire, W. J. 1978. "An Information-Processing Model of Advertising Effectiveness," *Behavioral and management science in marketing*, pp. 156-180.

Mick, D. G. 1992. "Levels of Subjective Comprehension in Advertising Processing and Their Relations to Ad Perceptions, Attitudes, and Memory," *Journal of Consumer Research*, pp. 411-424.

Mudambi, S. M., and Schuff, D. 2010. "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon. Com," *MIS Quarterly* (34:1), pp. 185-200.

Nelson, P. 1970. "Information and Consumer Behavior," *The Journal of Political Economy*, pp. 311-329.

Norton, S. W., and Norton Jr, W. 1988. "An Economic Perspective on the Information Content of Magazine Advertisements," *International Journal of Advertising* (7:2), pp. 138-148.

Pan, Y., and Zhang, J. Q. 2011. "Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews," *Journal of Retailing* (87:4), pp. 598-612.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. 2007. "Linguistic Inquiry and Word Count: Liwc [Computer Software]," *Austin, TX: liwc. net*.

Rafaeli, A., and Vilnai-Yavetz, I. 2004. "Emotion as a Connection of Physical Artifacts and Organizations," *Organization Science* (15:6), pp. 671-686.

Rigby, D. 2011. "The Future of Shopping," *Harvard Business Review* (89:12), pp. 65-76.

Schindler, R. M., and Bickart, B. 2012. "Perceived Helpfulness of Online Consumer Reviews: The Role of Message Content and Style," *Journal of Consumer Behaviour* (11:3), pp. 234-243.

Shmueli, G., Patel, N. R., and Bruce, P. C. 2007. *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with Xlminer*. John Wiley & Sons.

Strawson, P. F. 1971. *Logico-Linguistic Papers*. Methuen London.

Suchan, J., and Dulek, R. 1988. "Toward a Better Understanding of Reader Analysis," *Journal of Business Communication* (25:2), pp. 29-45.

Whitney, P. 1987. "Psychological Theories of Elaborative Inferences: Implications for Schema-Theoretic Views of Comprehension," *Reading Research Quarterly*, pp. 299-310.

Yin, D., Bond, S., and Zhang, H. 2014. "Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews," *MIS Quarterly (38:2), pp. 539-560*.

Yu, M., Debo, L., and Kapuscinski, R. 2015. "Strategic Waiting for Consumer-Generated Quality Information: Dynamic Pricing of New Experience Goods," *Management Science*.

Zhou, L., Dai, L., and Zhang, D. 2007. "Online Shopping Acceptance Model-a Critical Survey of Consumer Factors in Online Shopping," *Journal of Electronic Commerce Research* (8:1), p. 41.

Zhu, F., and Zhang, X. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing* (74:2), pp. 133-148.