INTERNATIONAL DESIGN CONFERENCE – DESIGN 2022
https://doi.org/10.1017/dsd.2022.000

INFORMATION RETRIEVAL AND SURVEY DESIGN FOR TWO-STAGE CUSTOMER PREFERENCE MODELING

[Authors will be inserted automatically]

[Institutions will be inserted automatically]
[Corresponding author data will be inserted automatically]

Abstract

[Abstract will be inserted automatically]
[---]
[---]
[---]
[---]
[---]

[Keywords will be inserted automatically]

1. Introduction

Designing a customer-favored product is critical to a company's success in a competitive market. Companies are particularly interested in what factors influence customer (one who purchases or receives a product or indentd to do so) purchase behaviors and their relative importance. In the past decades, customer preference modeling has been a primary research method to answer these questions in both marketing science (Pescher and Spann, 2014; Stankevich, 2017) and the engineering design community. For example, customer preference modeling can provide designers with insights into identifying customer-preferred product features and how customers make tradeoffs among multiple attributes (Pescher and Spann, 2014; Sha et al., 2017). Furthermore, research shows that a customer's decision-making process typically involves two stages during which the customer first forms a consideration set and second makes the final choice using different criteria (Shocker *et al.*, 1991). The interest in customer preference modeling has primarily focused on two aspects: 1) to understand how product attributes influence customers' decision making. For example, attempts have been made to model the impact of product design attributes on customer considerations and choices using customer-product network modeling (Bi et al., 2021; Cui et al., 2020; Sha et al., 2017; Wang et al., 2015). 2) To understand the role of social influence in customers' decision-making (Argo, 2020), for example, using the data on customer-preferred product attributes before and after peer effects (Narayan et al., 2011) and demographic data from customers' social neighbors (Aral and Walker, 2010; Campbell and Lee, 1991). However, one major gap in current literature is that the impact of social influence and product attributes on customer purchase decisions are investigated separately. This is attributed to the limitations of data in two aspects. First, customers' social network data and the attribute data of their considered and purchased products are not collected simultaneously. Therefore, synthetic social network data has to be created when studying the social influence on customers' choices (Wang *et al.*, 2016). Second, many datasets came from private sectors. Since those data often embed customer preferences, it is of high

commercial value to enterprises, thereby cannot be shared publicly. Consequently, such limitations have affected the reproducibility and repeatability of many existing models (Anon, 2013).
To overcome these limitations, researchers must settle for the second-best to explore obtainable data sources, such as online product reviews, social media, and public customer survey data. Regarding the online review data, the reviews are typically generated by customers who have purchased the products (Lee and Bradlow, 2011), accessible via online stores' websites. Social media data are referred to the online content that customers or experts post on social network platforms such as Twitter or YouTube (Tuarob and Tucker, 2015). However, both types of data have minimal customer demographics, so customer reviews can not be associated with, yet, essential to customer preference modeling. Public customer survey data often includes a few products selected from a large pool of available products and can only support modeling studies with constrained information (Bao et al., 2020; Barnard et al., 2016).
This study aims to develop a systematic approach that combines information retrieval and survey design in support of data collection for customer preference modeling that can address the limitations above. Specifically, we have made the following **contributions**: 1) we created a tool that can extract critical product features from customer reviews, integrating web scraping, text mining, and rule-based semi-supervised learning. 2) We developed a web-based survey platform that supports interactive information retrieving and virtual online shopping. 3) In the survey design, data quality assurance mechanisms, such as customer memory tests and attention check questions, were created and added. 4) The survey supports collecting customers' social network data and their preferences in a unified framework. 5) We designed the survey to support the data collection of both customers' considerations and choices. Thus, the data collected can be used in multi-stage choice modeling to study customers' consideration-then-choice behaviors. Our approach is demonstrated in the customer preference modeling of vacuum cleaners. To benefit a broader community, both the product and customer survey datasets will be made publicly available for researchers interested in customer preference modeling.

2. The Framework of the Proposed Approach

Figure 1 depicts an overview of the proposed information retrieval and survey design approach for two-stage customer preference modeling. It consists of four major modules and two outputs. Next, we provide the description of each module, and the details for the vacuum cleaner preference survey case study are presented in Section 3.

2.1. Module 1: Product Database Establishment

The main goal in Module 1 is to create a product database with basic product information, such as product model names and product attributes. This database acts as an input that is linked directly to the subsequent survey design modules. We first design a well-formatted SQL database. Then, a web scraping tool is used to collect product information, e.g., product image and attributes, etc., and customer review data from major electronic retailers and department stores, e.g., Amazon, BestBuy, and Walmart. Next, utilizing text mining technology (e.g., a two-fold rules-based model (TF-RBM (Rana and Cheah, 2017)), we extract all the product attributes from scrapped customer reviews and allocate quantitative importance scores to each identified attribute based on its frequency of occurrence within the scraped reviews (Rai, 2012). The final list of critical attributes is determined by the rank of their importance scores and expert input. Finally, all of the collected data is organized and saved in the SQL database.

2.2. Module 2: Purchase Memory Test

When taking a survey, the amount of detailed product information (e.g., the model name) a participant could memorize depends on how long the product was purchased. This leads to the idea of creating Module 2 to account for the memory bias across different participants. Therefore, to ensure the data quality, a purchase memory survey test, e.g., whether the customers who purchased a vacuum cleaner in the past one month, three months, or six months can remember their choices, is designed prior to the formal survey study. Once the memory test result is obtained, we use the test result to determine the type of survey, revealed or stated. In ***the revealed study***, only the participants who actually purchased the product will be eligible to take the survey, and the data will be used for model revealed preference.

Whereas in ***the stated study***, the participants are required to complete the survey based on a virtual online shopping experience.

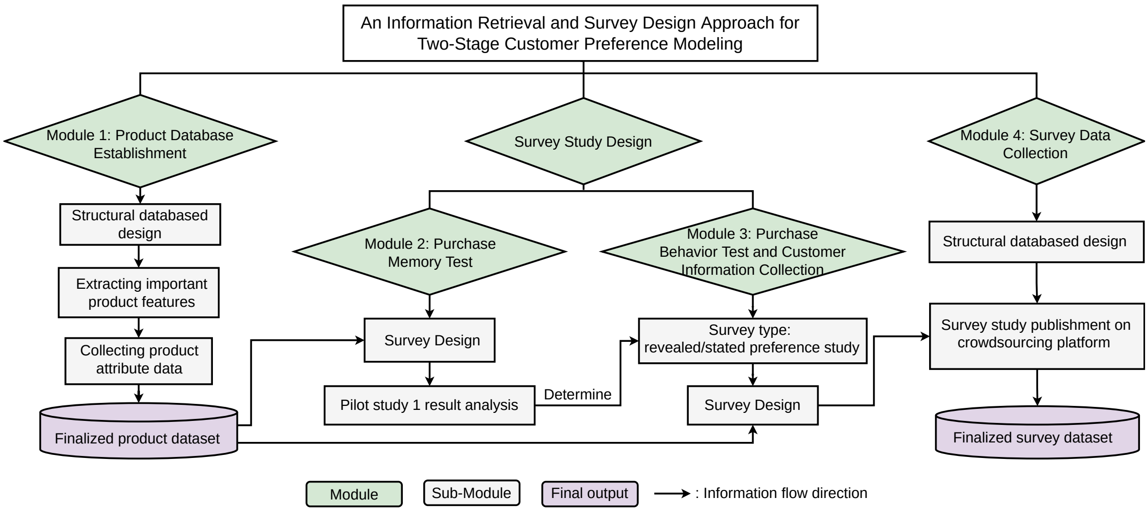


Figure 1. The general framework of the proposed approach

2.3. Module 3: Purchase Behaviour Test and Customer Information Collection

Module 3 focuses on the questionnaire design of the customer preference survey. We divide our questionnaire into three major parts to ensure that the collected data can support both the social influence and the consideration-then-choice behavior analyses. **Part One** is to collect participants' historical consideration and choice data, including the type of product they considered, the exact model they eventually purchased, and the top-rated attributes (features) that influenced their choice-making. In **Part Two**, we design questions to collect participants' social network data. This includes both of their general social networks (GSN) as well as the product-specific social networks (PSN) (Campbell and Lee, 1991). The GSN is a natural social relation network that captures the people with whom respondents communicate about important issues in their daily lives, such as their spouse, parents and close friends. The PSN refers to the people with whom respondents have discussed product purchases, such as their coworkers who have endorsed their purchase, and they may or may not be from respondents' GSN. A persons' PSN has the potential influence on their choice behaviors. **Part Three** focuses on gathering participants' personal information and user preferences. This includes their demographics, the general preferences for household appliances, and product usage context. We use a variety of strategies to guarantee data quality (Bernard, 2013). These strategies are: 1) developing a product searching system to reduce participants' manual workload, thus improving the information retrieval accuracy; 2) setting attention check questions; 3) conducting both internal and external pilot studies; 4) implementing phase-in data collection and adjustment; and 5) incorporating experts' inputs and feedback from multiple disciplines including engineering design, social science, and psychological science.

2.4. Module 4: Survey Data Collection

Module 4 is associated with two tasks: 1) designing a well-formatted and structured database that is advantageous for later data utilization, and 2) launching the survey on a crowdsourcing platform. The reputation of the crowdsourcing platform is essential because it directly influences the quality of the participants we can recruit. A platform with quality assurance mechanisms such as an AI-drive fraud detection system is always beneficial for us to collect high-quality data. Some popular platforms include MTurk, Prolific, and Cint. Once the data is collected, it is automatically saved in the SQL database.

3. Case Study: Household Vacuum Cleaner

In this study, we choose household vacuum cleaner as our case for several reasons: **1)** it is a common household appliance with heterogeneous categories (e.g., upright, canister, robotic, etc.) and multiple competitors (e.g., Dyson, Shark, etc.) in the market; **2)** it has a large market size with customers who have heterogeneous preferences on vacuum cleaners based on their demographics and usage context, and **3)** its design attributes (features) play an important role in influencing customers' choices (Harmer *et al.*, 2019), so the study on customer preference modeling shed light on design for market systems.

3.1. Vacuum Cleaner Data Collection and Attribute Extraction

We scraped vacuum cleaner information and built the product database using the web crawling technique (Beautiful soup and selenium in Python). The household vacuum cleaners had been scrapped from the mainstream online shopping platforms in the US market (Amazon, Wayfair, Best Buy, Home Depot, and Walmart). Meanwhile, by scrapping the structured website, we collected the product information (product title, customer rating, SKU (stock-keeping unit)), features (list price, product dimension, weight, manufacture, brand, color, capacity, etc.), product description, and customer reviews. This study focused on five primary categories of vacuum cleaners - upright, canister, stick, handheld, and robotic vacuum cleaners. Data cleaning was performed to merge data from different sources, remove duplicated models and noises, and perform text mining to identify missing feature values. In the end, 1170 products with 26 features were collected in our final dataset.

In addition, we extracted product features from online customer reviews to determine the most important (most frequently mentioned) features to be included in the survey questions. We scrapped 60,000 reviews from Amazon (200 reviews for each product) and used a rule-based semi-supervised learning model for extracting features and sentiment/opinion associated with those features. For example, some feature-opinion pairs extracted from the reviews include "strong suction," "heavy weight", "annoying cord," and "loud noise." After obtaining candidate features from the opinion mining, unrelated features were pruned, and the rest features were ranked based on their frequencies in customer reviews. In the end, we identified 22 important product features based on the opinion mining results, including attributes such as price, product type, floor surface recommendation, suitable for pet hair, suction power, noise, power source, bag or bagless, cord or cordless, battery charge time, HEPA filter, warranty, brand, color, weight, dimensions, power, capacity, *navigation system, voice control, remote controls (robotic vacuum cleaner specific attributes)* and overall customer ratings.

3.2. Customer Purchase Memory Test

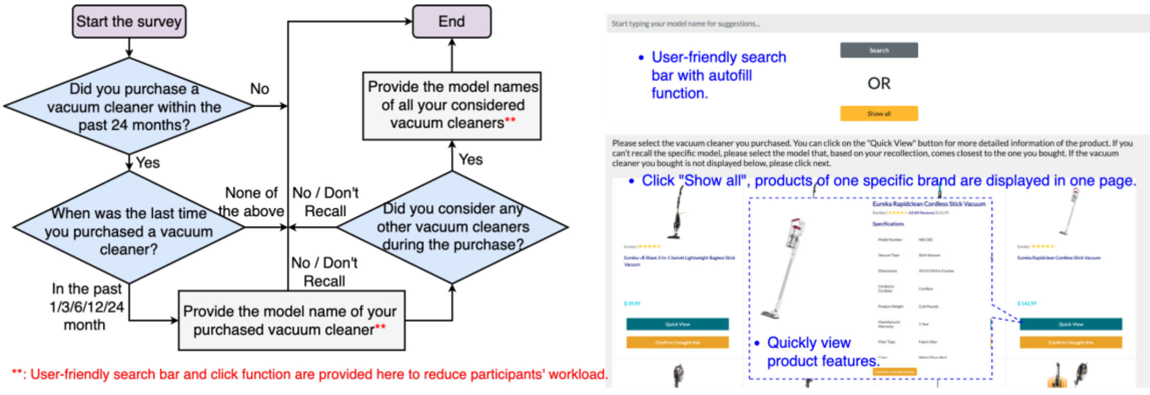
A pilot study was conducted to assess customers' abilities to remember their vacuum cleaner purchase decision-making over the past one month, three months, six months, twelve months, and 24 months to determine the appropriate threshold in soliciting participants. We firstly built a survey web for the test. The survey design logic and web interface examples are shown in Figure 2. To reduce participants' workload, a simulated online shopping system with features such as a user-interactive search bar and product preview was developed. As shown in Figure 2, we collected 30 samples for each period separately. Then, using those 30 samples, we calculated the proportion of participants who can recall the specific models they considered and purchased. Normally, if the ratio is greater than 50%, we consider customers' memory within that time period to be reliable.

Table 1. The sample size of the purchase memory test

	In the past one month	In the past three months	In the past six months	In the past 12 months	In the past 24 months
# people who have purchased a vacuum cleaner	32	34*	32	35	8

*: This number has excluded the number of people who have purchased a vacuum cleaner in the past one month. A similar explanation applies to the other three periods (in the past six months/12 months/24months).

The survey was conducted on the Cint platform from December 18 to December 21, 2020. Table 1 summarizes the actual collected sample size for the test. Because there were far fewer samples, the 24-month scenario was neglected in the proportion calculation. According to Figure 3, approximately 62% of customers who purchased a vacuum cleaner in the past three months can remember their purchases and considerations, satisfying the 50% threshold. However, if we only focus on the customers who purchased vacuum cleaners within the past three months, we may not be able to collect enough samples for our following-up formal survey. Thus, we made a tradeoff by extending the period to the past six months because it has a high ratio of recall for purchase (75%); meanwhile, the ratio of recall for both purchase and consideration (43.75%) is still acceptable. So, in the formal study, only the customers who purchased the vacuum cleaner in the past six months were eligible to participate in the survey.



(a) Survey questionnaire design for customer purchase memory test (b) Survey web platform design for customer purchase memory test

Figure 2. Survey questionnaire and web platform design for customer purchase memory test

Figure 3. The ratio of participants who can recall the purchased or considered vacuum cleaners

3.3. Vacuum Cleaner Customer Survey Questionnaire Design

The customer purchase behavior test, as introduced in Section 2.3, consisted of three major parts. **Part One** employed the same simulated online shopping system to alleviate respondents' workload. Furthermore, participants can rank the product features that influence their decision-making by dragging them from a list to the corresponding text boxes. The list contained all of the attributes identified by the feature selection algorithm introduced in Section 3.1. In **Part Two**, participants were asked to provide at least one and up to five individuals' information in their general social networks (GSN) as well as all the ones with whom they had discussed the vacuum cleaner purchase . These individuals' demographic information and their contact frequencies with the respondents were also recorded. **Part Three** collected respondents' personal information and attributes, such as their own stated product preference.

To ensure the data quality, aside from the strategies mentioned in Section 2.3, other strategies employed include 1) setting filtering questions, e.g., did you purchase a vacuum cleaner in the past six months, so that only satisfied respondents can participate in this survey; 2) organizing questions by placing important questions first and less important questions last; 3) making questions mandatory to avoid missing data, i.e., participants could proceed the next stage of survey only after answering all the required questions. Lastly, similar to the purchase memory test, an associated survey website of the purchase behavior test was designed.

3.4. Survey Data Collection

We employed the Cint platform to launch our survey due to its established reputation. Additionally, we developed an SQL database on pgAdmin with a fine-tuned columns sequence to ensure that all the respondents' answers could be structurally saved. Note that this database had been configured to communicate effectively with the survey website. To acquire more results, the survey was distributed to different groups, such as those who had recently purchased a vacuum cleaner or those who are interested in home decoration and home appliances. The survey was conducted over a two-month period, from April 25 to June 25, 2021, and a total of 1011 responses were received, with a completion rate of 15.35%. Meanwhile, to mirror the real market, a quota sampling technique (Sudman, 1966) was used to match the age distribution of the US census.

4. Data Utility and Quality

In this section, after cleaning and processing the raw data, we assessed the utility and quality of our data by performing a descriptive analysis on customers' two-stage decision-making processes and social network influence. We also constructed the unidimensional co-consideration and choice networks using our survey data for visualization, which shows the potential of our survey data in supporting customer preference modeling and engineering design.

4.1. Descriptive Analysis of Customer-Product Data in Two-Stage Decision-Making

Survey Respondent Demographics and Usage Context From the demographic data, the average profile of respondents is male (56.87%), Caucasian (74.88%), 35-54 (29.48%), married (63.11%), retired (11.51%), with a bachelor's degree (36.80%), and with an annual household income of $40k - $70k (24.53%). The majority of respondents live in their own homes (76.55%), live in a single house (80.12%), have 6-10 rooms (55.59%), have stairs (65.18%). have multiple types of floors (70.43%), clean their home every week (62.31%), and have at least one pet (80.51%). Table 2 is a list of major usage contexts of survey respondents.

Table 2. Summary of key usage contexts of survey respondents

Cleaning frequency	Frequency	Percentage (%)	Number of pets at home	Frequency	Percentage (%)
Every day	343	33.93	0	197	19.49
Every week	630	62.31	1 - 3	731	72.31
Every month	34	3.36	Over 3	93	8.21

Considered and Purchased Vacuum Cleaners We collected information on the vacuum cleaners that respondents had considered and purchased as part of the study. Respondents reported 1011 vacuum cleaners purchased and another 1473 vacuum cleaners considered but not purchased. About 73.49% of respondents saied they considered other vacuum cleaners before making their purchases, while 21.36% saied they considered another vacuum cleaner, 28.19% saied they considered another two vacuum cleaners, 19.99% saied they considered more than three (up to six) vacuum cleaners.

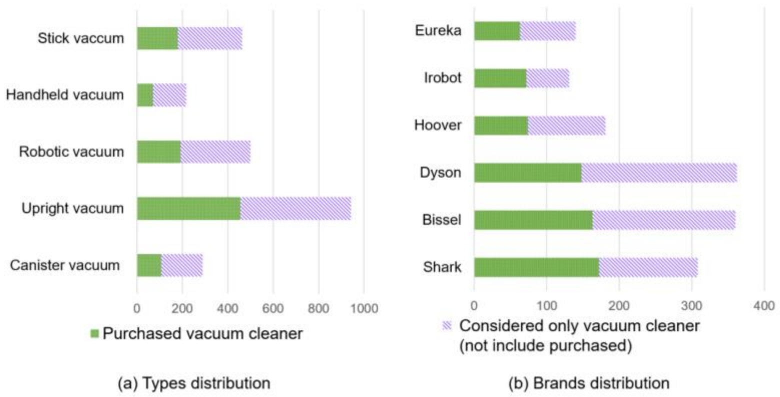


Figure 4. Respondents' considered and purchased vacuum cleaners (a) types distribution and (b) top 6 brands distribution

The majority of vacuum cleaners that respondents have purchased (the solid green bar) and considered (the dashed purple bar) are shown in Figure 4 (a). The total length of each bar indicates the popularity of each type of vacuum cleaner in customers' consideration, while the green bar reveals the popularity in customers' final choices. It's worth noting that upright vacuum cleaners are the most popular at both stages of consideration-then choice. Figure 4 (b) records the most popular brands that have been considered and purchased by respondents. It is noted that Dyson and Bissel are the most popular among respondents in the consideration, but Shark gains more popularity in the choice stage.

The rank of Importance for Product Attributes in Two-Stage Decision-Making We have collected respondents' stated preferences regarding the most important features of vacuum cleaners in their decision-making process. Respondents were asked to pick and rank 3 - 5 of a vacuum cleaner's most important technical features in their consideration stage and choice stage. The importance of the attributes can be obtained by calculating the **weighted sum of customer rankings**, as shown in the following equations:

$$A = \sum(w_i \times c_i), for\ i = 1, 2, 3, 4, 5 \tag{1}$$

where A is the weighted sum, w is the ranking weight, c is the count of the rank, and i is the rank. We assign the ranking weight as 5,4,3,2,1 when the feature is rated as 1st, 2nd, 3rd, 4th, and 5th important. Figure 5 shows the important ranking of vacuum cleaner attributes based on the weighted sum of importance.

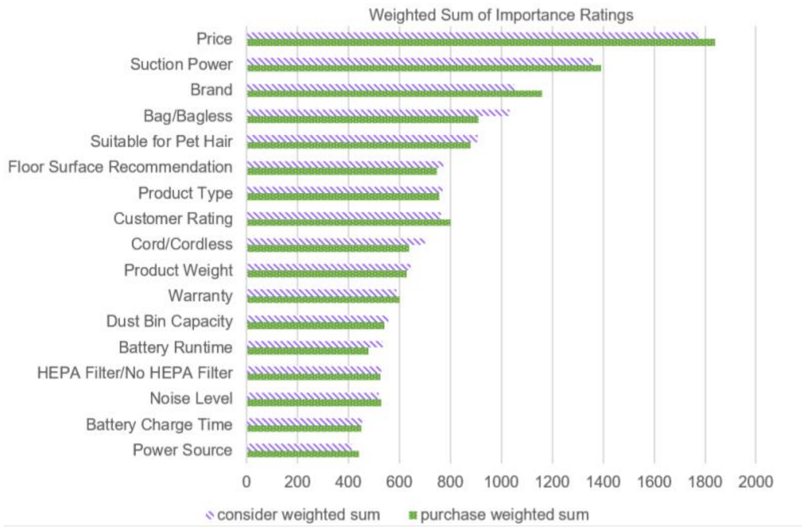


Figure 5. The rank of technical attributes based on the weighted sum of customers' importance rankings in their consideration stage and purchase stage.

We can see the overall trends are consistent in consideration and choice stages, indicating price, suction power, the brand are the more important features in their decision-making process. Besides, there are some discrepancies between the two stages. For example, in the consideration stage, features such as product types, cord/cordless, bag or bagless, and floor surface recommendation are more important, while in the second stage, detailed and technical features such as price, suction power, and customer ratings are more important.

4.2. Social Network Influence Analysis

In our survey, we asked respondents to name the people (up to 5 people) with whom they most frequently discuss important matters, as well as whether they had discussed their vacuum cleaner purchases with those people or anybody else (up to 5 people). In such a way, we investigate the respondents' general social network and vacuum cleaner-specific social network.

General Social Network (GSN). The respondents' general social network consists of people with whom they discuss important things in their daily lives. According to the results, respondents named 2.15

people on average, and the frequencies of naming a certain number of people are presented in Table 3 (the number of people in GSN). Among the people in their GSN, the most frequent relationships are with spouses (24.72%), friends (23.94%), and parents (12.18%). We also looked at the vacuum cleaners owned by the people in the respondents' GSN. It turns out that individuals with the same make and model as the respondents account for 31.99% of the total. 13.14% have the same make but different models, and 7.53% have the same type but different makes and models. The data is a preliminary indicator that GSN has an impact on repsondents' vacuum cleaner purchase.

Vacuum Cleaner-Specific Social Network (VCSN). We further investigate the individuals with whom the respondents have discussed vacuum cleaner purchases. While the respondents talked about their vacuum cleaner purchases with an average of 1.77 people in their GSN, they also stated they had discussed their purchases with an extra 0.42 people on average (the frequencies of the number of people in VCSN are shown in Table 3, GSN&VCSN, and VCSN only). Among the additional persons outside of a respondent's GSN, 19.85% are their friends, 17.49% are their acquaintances, 9.22% are their spouses, 7.57% are their neighbors, and 2.73% are salespersons. According to the survey, people in respondents' vacuum cleaner-specific social network (VCSN) plays a vital role in their consideration and choice stages. For example, in their consideration stage, 42.55% of respondents say their VCSNs are very important (the highest among the five Likert scales). In their choice stage, 43.65% of respondents think their VCSNs are very important.

Table 3. The frequency of different numbers of the people in respondents' GSN and VCSN

# of people in	0	1	2	3	4	5
GSN	0.00%	48.66%	19.68%	13.45%	4.15%	14.04%
GSN&VCSN	8.11%	48.76%	20.08%	11.67%	3.76%	7.62%
VCSN only	73.69%	19.49%	3.07%	1.09%	0.40%	2.27%

The data collected in this survey also includes the demographic information of the people in GSN and VCSN, the frequency of contact (which determines the strength of their relationship), and their personal viewpoints. All of the data we collect will be useful in understanding how social network influence affects customers' vacuum cleaner consideration and purchase decisions in the future work.

4.3. Co-consideration Network and Choice Network construction

One important usage of the customer survey data is to build the customer-product networks based the two-stage (consideration-then-choice) customers' decision-making process. As an illustration, we construct two simplified unidimensional networks, which only consists of the product nodes, to demonstrate the co-consideration and choice relationships among products. The undirected *co-consideration network* in Figure 6 (a) presents vacuum cleaner models as nodes, and the frequencies of two vacuum cleaners being co-cosndiered by customers as link. The directed *choice network* in Figure 6 (b) presents the same set of nodes, while the directed links denotes when two products are co-considered, which product between the two is more likely to be bought by customers.

In the *co-consideration network*, there are 672 unique vacuum cleaner models, while 63 products are isolated (which are not co-considered with others). The model "Dyson Upright Vacuum Cleaner, Ball Multi Floor 2, Yellow" is the most popular vacuum cleaner. It was co-considered with other vacuum cleaners by 46 times. The nodes' average weighted degree is 6.45, which indicates that the vacuum cleaner models in our network are co-considered by 6.45 customers on average. In the *choice network*, there are 72 isolated items among the 672 vacuum cleaner models. The average weighted in-degree of the nodes is 1.79, implying that a vacuum cleaner is initially co-considered with other products before being picked by 1.79 consumers on average. The most popular purchased vacuum cleaner model is the same as the most considered vacuum cleaner model - "Dyson Upright Vacuum Cleaner, Ball Multi Floor 2, Yellow" (selected by 37 times). The product competition relationship can be represented by both the co-consideration network and the choice network, indicating customers' aggregated preferences. Once these networks are constructed, more statistical analysis can follow to analyze and predict customer preferences (Cui et al., 2020; Sha et al., 2017; Wang et al., 2015).

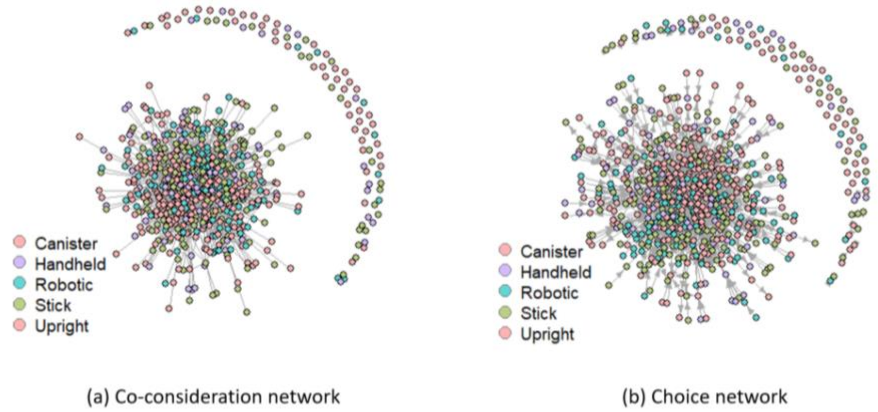


Figure 6. (a) Unidimensional co-consideration network and (b) choice network

5. Conclusion

In this study, we presented a systematic approach that combines information retrieval and survey design in support of data collection for customer preference modeling. This approach supports a systematic design of customer surveys that collect customers' social network and preference data in both consideration and final choice stages. Therefore, the resulting datasets can support the study of a wide range of customer preference models, such as the social influence modeling and consideration-then-choice analysis, which can help product designers understand the feature importance and make crtical design decisions. Another merit of this study is the integration of state-of-the-art information retrieval techniques and survey design guidelines, including web scraping, text mining, SQL data management, and data quality assurance (e.g., purchase memory test). We have demonstrated how the approach works and how the techniques and guidelines are integrated using a case study on household vacuum cleaners. Our approach can be generally applied in collecting data of engineered products that are physical and discrete. We also conducted preliminary data analyses to assess the utility and quality of the obtained data. These data will be made available to the public for broader impact. One limitation of this work lies in the design of the survey questions related to the customers' social influence. The current questions do not adequately ask customers' opinions on how the people in their social network would influence their decision-making. In the future, we will refine those survey questions. In addition, we plan to further demonstrate the utility of the data in customer preference modeling by performing network-based choice analysis and prediction.

Acknowledgment

The authors acknowledge Lada Nuzhna, Fiona, Olga Lew-Kiedrowska, Laith Kassisieh, Neelam Modi for their assistance in data management on Cint and/or the inputs during research meetings. We also greatly acknowledge the funding support from NSF CMMI #2005661 and #2203080.

References

Anon, J. (2013), "Announcement: Reducing our irreproducibility", *Nature*, Vol. 496 No. April, p. 398.

Aral, S. and Walker, D. (2010), "Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks", *SSRN Electronic Journal*, No. October 2021, available at:https://doi.org/10.2139/ssrn.1564856. DOI:10.2139/ssrn.1564856.

Argo, J.J. (2020), "A contemporary review of three types of social influence in consumer psychology", *Consumer Psychology Review*, Vol. 3 No. 1, pp. 126–140. DOI:10.1002/arcp.1059.

Bao, Q., Sinitskaya, E., Gomez, K.J., MacDonald, E.F. and Yang, M.C. (2020), "A human-centered design approach to evaluating factors in residential solar PV adoption: A survey of homeowners in California and Massachusetts", *Renewable Energy*, Elsevier Ltd, Vol. 151, pp. 503–513. DOI:10.1016/j.renene.2019.11.047.

Barnard, A.S., Louviere, J.J., Wei, E. and Zadorin, L. (2016), "Using hypothetical product configurators to measure consumer preferences for nanoparticle size and concentration in sunscreens", *Design Science*, Vol. 2, p. e12. DOI:10.1017/dsj.2016.12.

Bernard, H.R. (2013), *Methods in Human Geography*, edited by Flowerdew, R. and Martin, D.M., Routledge, available at:https://doi.org/10.4324/9781315837277. DOI:10.4324/9781315837277.

Bi, Y., Qiu, Y., Sha, Z., Wang, M., Fu, Y., Contractor, N. and Chen, W. (2021), "Modeling Multi-Year Customers' Considerations and Choices in China's Auto Market Using Two-Stage Bipartite Network Analysis", *Networks and Spatial Economics*, Networks and Spatial Economics, Vol. 21 No. 2, pp. 365–385. DOI:10.1007/s11067-021-09526-9.

Campbell, K.E. and Lee, B.A. (1991), "Name generators in surveys of personal networks", *Social Networks*, Vol. 13 No. 3, pp. 203–221. DOI:10.1016/0378-8733(91)90006-F.

Cui, Y., Ahmed, F., Sha, Z., Wang, L., Fu, Y. and Chen, W. (2020), "A Weighted Network Modeling Approach for Analyzing Product Competition", *Volume 11A: 46th Design Automation Conference (DAC)*, Vol. 11A-2020, American Society of Mechanical Engineers, pp. 1–13. DOI:10.1115/DETC2020-22591.

Harmer, L., Cooper, T., Fisher, T., Salvia, G. and Barr, C. (2019), "Design, Dirt and Disposal: Influences on the maintenance of vacuum cleaners", *Journal of Cleaner Production*, Elsevier Ltd, Vol. 228, pp. 1176–1186. DOI:10.1016/j.jclepro.2019.04.101.

Lee, T.Y. and Bradlow, E.T. (2011), "Automated Marketing Research Using Online Customer Reviews", *Journal of Marketing Research*, Vol. 48 No. 5, pp. 881–894. DOI:10.1509/jmkr.48.5.881.

Narayan, V., Rao, V.R. and Saunders, C. (2011), "How Peer Influence Affects Attribute Preferences: A Bayesian Updating Mechanism", *Marketing Science*, Vol. 30 No. 2, pp. 368–384. DOI:10.1287/mksc.1100.0618.

Pescher, C. and Spann, M. (2014), "Relevance of actors in bridging positions for product-related information diffusion", *Journal of Business Research*, Elsevier Inc., Vol. 67 No. 8, pp. 1630–1637. DOI:10.1016/j.jbusres.2013.09.005.

Rai, R. (2012), "Identifying Key Product Attributes and Their Importance Levels From Online Customer Reviews", *Volume 3: 38th Design Automation Conference, Parts A and B*, Vol. 3, American Society of Mechanical Engineers, pp. 533–540. DOI:10.1115/DETC2012-70493.

Rana, T.A. and Cheah, Y.-N. (2017), "A two-fold rule-based model for aspect extraction", *Expert Systems with Applications*, Elsevier Ltd, Vol. 89, pp. 273–285. DOI:10.1016/j.eswa.2017.07.047.

Sha, Z., Saeger, V., Wang, M., Fu, Y. and Chen, W. (2017), "Analyzing Customer Preference to Product Optional Features in Supporting Product Configuration", *SAE International Journal of Materials and Manufacturing*, Vol. 10 No. 3, pp. 2017-01–0243. DOI:10.4271/2017-01-0243.

Shocker, A.D., Ben-Akiva, M., Boccara, B. and Nedungadi, P. (1991), "Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions", *Marketing Letters*, Vol. 2 No. 3, pp. 181–197. DOI:10.1007/BF02404071.

Stankevich, A. (2017), "Explaining the Consumer Decision-Making Process: Critical Literature Review", *JOURNAL OF INTERNATIONAL BUSINESS RESEARCH AND MARKETING*, Vol. 2 No. 6, pp. 7–14. DOI:10.18775/jibrm.1849-8558.2015.26.3001.

Sudman, S. (1966), "Probability Sampling with Quotas", *Journal of the American Statistical Association*, Vol. 61 No. 315, pp. 749–771. DOI:10.1080/01621459.1966.10480903.

Tuarob, S. and Tucker, C.S. (2015), "Quantifying Product Favorability and Extracting Notable Product Features Using Large Scale Social Media Data", *Journal of Computing and Information Science in Engineering*, Vol. 15 No. 3, pp. 1–12. DOI:10.1115/1.4029562.

Wang, M., Chen, W., Fu, Y. and Yang, Y. (2015), "Analyzing and Predicting Heterogeneous Customer Preferences in China's Auto Market Using Choice Modeling and Network Analysis", *SAE International Journal of Materials and Manufacturing*, Vol. 8 No. 3, pp. 2015-01–0468. DOI:10.4271/2015-01-0468.

Wang, M., Chen, W., Huang, Y., Contractor, N.S. and Fu, Y. (2016), "Modeling customer preferences using multidimensional network analysis in engineering design", *Design Science*, Vol. 2, p. e11. DOI:10.1017/dsj.2016.11.