Use of clustering for consideration set modeling in recommender systems

Emil Mirzayev
Université Côte d'Azur
emil.mirzayev@skema.edu

Zakaria Babutsitdze
SKEMA Business School
zakaria.babutsidze@skema.edu

William Rand
NC State University
wmrand@ncsu.edu

Thierry Delahaye
Amadeus S.A.S.
thierry.delahaye@amadeus.com

Abstract

The cold-start problem has become a significant challenge in recommender systems. To solve this problem, most approaches use various user-side data and combine them with item-side information in their systems design. However, when such user data is not available, those methods become unfeasible. We provide a novel recommender system design approach which is based on two-stage decision heuristics. By utilizing only the item-side characteristics we first identify the structure of the final choice set and then generate it using stochastic and deterministic approaches.

1. Introduction

With the rise of the Internet, interaction with recommender systems has become a common part of human activity. When there are many options to choose from, recommender systems save consumers time and effort by matching them with items [1]. Making successful recommendations requires knowledge of demand-side factors, such as consumer taste, historical interactions, purchasing power, and socio-demographic characteristics, along with supply-side factors, such as item characteristics.

Because recommender systems are online services implemented by the providers, supply-side information is generally available at all times. However, this is not always the case with demand-side information. On most occasions users are not identified, either because it is not feasible, or because interaction with the system does not require them to identify themselves. This lack of information is referred to as the cold-start problem [2]. Some services, for example Netflix, solve this problem by providing general suggestions until they can gather enough information about the user. Others, such as Goodreads, explicitly survey the new user to solicit such information.

With the current regulations and users' awareness of security and privacy on the internet [3], systems face continuous cold-start problems [4]. In such cases, using supply-side contextual information becomes crucial [2]. One way of utilizing such information is using random utility models. However, such models rely on the notion of perfectly rational consumers having well-defined preferences [5]. Hence, they are not able to account for context-dependent preferences [6]. Under such circumstances, clustering-based approaches can be used to enrich the context in recommender systems. Previous research [5] has demonstrated that such an approach is flexible enough to be extended over imperfectly rational, or context-dependent consumer behavior.

In this paper, we incorporate insights from the decision literature into recommender system design using the dataset of European flight choices [7]. We argue that the choice process occurs in two sequential stages: consumers first identify the small subset of the choices that they would "consider" and then make a choice from that subset. Combining findings in marketing, management and consumer behavior, and using clustering to quantify the contextual information of the choice set, we propose a user-side, two-step "consider-then-choose" [8, 9] approach to recommender system design to tackle the continuous cold-start problem.

2. Theoretical background

2.1. Choice heuristics

Previous literature in psychology and economics has suggested that individuals tend to use various decision heuristics to reduce the cognitive load during the decision-making process [10, 11, 12]. Because consumers tend to behave as *satisficers* rather than *maximizers*, they do not perform an evaluation of all the alternatives available to them but stop as soon as they find an option which has overall better attributes and satisfies their needs [13]. For simplicity, let's consider the case of buying flight tickets. There are N tickets, and they each have k attributes, which can include price, duration, time of the day of the flight, number

URI: https://hdl.handle.net/10125/71135
978-0-9981331-4-0

of connections, and so on. A consumer has a single objective function

$$O = O(T_1, T_2, \ldots, T_N), \tag{1}$$

where T is the linear transformation of the flights' attributes Z^k with some random parameters θ. We can rewrite equation (1) as

$$O = O\left(\sum_k \theta_k Z_1^k, \ldots, \sum_k \theta_k Z_j^k, \ldots, \sum_k \theta_k Z_N^k\right), \tag{2}$$

which would allow a consumer to explicitly compare the marginal contribution of each attribute to to the maximization of the the objective function [14]. However, most of the time, choice attributes do not require or allow for such trade-off calculations, as some of them are too valuable, or there are too many attributes to consider. For example, a consumer flying for business purposes may value the flight duration more than a budget traveler, who would value price above everything else. In such cases, even when consumers are perfectly rational and are well-informed, when they face options that simultaneously differ across many attributes, or it is difficult to calculate such trade-offs, they use various heuristic approaches [15].

Heuristics are mathematical formulas describing different rule-based decision steps taken by individuals to reduce their potential decision effort [16]. One can distinguish several types of heuristics-based approaches: lexicographic rule [10], conjunctive/disjunctive [17], elimination by aspects [18], and so on.

The lexicographic rule is the simplest deterministic rule in the heuristic approach. Here, individuals choose the alternative which has the highest value of the feature they desire. If there are several options with equal values, individuals compare those options based on the second most valued feature. This loop continues until there is one option remaining. For example, an individual searching for flight tickets from Paris to New York will have different options varying in time, price, number of connections, baggage allowance, transfer time, and so on. Attributes of the choices are first ranked based on their importance to the consumer: cheaper than 600 Euros, checked and carry-on baggage included, one layover, maximum transfer time of four hours. Then, a filtering stage occurs. After filtering on the price, if there are multiple options remaining, the consumer will switch to baggage allowance, connections and transfer time. As soon as the choice set contains only one option, the search stops.

The conjunctive and disjunctive heuristic approaches are related [17]. In the conjunctive rule, consumers first establish the list of features they consider relevant to the choice problem. Then, they establish various thresholds on those features. If an alternative passes all of those thresholds, it is chosen. In contrast, in the disjunctive approach, an option which exceeds threshold on at least one of the features is chosen [17]. The results of the lexicographic approach and conjunctive approach appear to be similar. The only difference is that, instead of evaluating options based on the first aspect, then the second and so on, in conjunctive approach the consumer evaluates options based on all aspects simultaneously. In some cases consumers might be willing to use a subset-conjunctive approach which generalizes both the conjunctive and disjunctive approaches [19]. It allows some variation in the desired aspects. For example, if the consumer valued 4 aspects as mentioned in the example above, he or she might be willing to accept an option which satisfies 3 of those 4 aspects. This approach is particularly useful when there are time constraints or a fully conjunctive rule would result in no choice [20].

Elimination by aspects is another heuristic approach which has been proposed in the literature [18]. The basic setup of this approach is that an individual chooses one attribute, and eliminates options based on this attribute and repeats this procedure for other attributes if necessary until the remaining options do not share common attributes anymore. Then, as a last step, the final option is chosen according to Luce's choice axiom [21] which states that the probability of selecting one option over the others in a choice set is not affected by presence or absence of other options. Most of the results on this topic [22, 23, 24, 25] indicate that the use of multi-phase heuristic processes can increase the accuracy of the estimation and result in improved interpretability of the models. Elimination by aspects is considered a heuristic method with stochastic rules because of the nature of the comparisons an individual makes, and because the selection process is not based on the relative importance of the features [26]. Such models are also hard to apply successfully because they require tremendous numbers of parameters to be estimated [22]. Although these models can theoretically capture the essence of the two-stage choice process, they are not able to identify the results of separate stages [9].

2.2. Two-stage choice

During the choice process consumers usually face a large number of options [27]. Evaluating that many options drastically increases cognitive load during the decision process and so, to reduce this load, consumers first select a small subset during an initial consideration stage [28] and then make their choice from that subset in the final stage [11, 23, 28]. Firstly, this allows users to remove unrealistic options from thorough consideration. Secondly, because the choice set is much smaller in the final stage, users are able to invest more cognitive effort to analyse individual options more carefully [23]. Also, the decision strategies used in the two stages differ considerably and are therefore not interchangeable. The main reason for that is the cognitive costs of the decision rules should not outweigh their potential benefits during each stage [29].

In the information processing literature the small subsets that consumers make their final decisions from are called consideration sets. There are several definitions of a consideration set. [30] defines a consideration set as a "set of alternatives that are goal-satisfying and accessible to a consumer on a particular occasion". [15] refers to it as a "set of options that receive a significant amount of consideration during the decision making process". In marketing, however, scholars generalize these definitions and refer to consideration sets as a "subset of alternatives surviving the initial screening phase" [31].

Despite the fact that consumers may not always use such a two-stage process to screen products [20], the use of consideration sets is justified because they represent the choice process more realistically and they explain consumer behavior better [32]. Potentially up to 80% of the decision process uncertainty can be resolved if we determine the consideration set correctly [33].

For an empirical study of consideration set formation, one can elicit information on consideration sets in multiple ways [34, 35, 36]. However, for modeling purposes the literature discusses two main ways of consideration set formation: deterministic [17] and stochastic [37, 38]. While stochastic modeling makes all potential sets possible by attaching non-zero choice probability to each of them, deterministic approaches may render some outcomes impossible [26]. Because we can not know which answer is the best and the decision-maker, or consumer, is the final arbiter of the "correct" choice [19], the use of either of these two approaches must consider the choice environment, time frame, future value (or loss) associated with the correct and incorrect choice, and so on [39]. For example, let's consider the flight booking case again and suppose that the consumer lives far from the airport and can reach it only in the afternoon. Consequently, all tickets with departure time before midday would not be considered at all. When forming a consideration set in this case, one must not only consider the characteristics of available options, but also the characteristics of the consumer and the choice environment. While using a purely stochastic approach might yield sets which include some options that consumer would indeed consider, there will also be options which will have zero probability for consideration. In contrast, applying some deterministic rules derived from this particular choice environment, such as the departure time, in the consideration set formation, will exclude those options completely.

When there are not many options to consider, options have few attributes, or final choice utility is not evenly distributed among the attributes, the consideration sets may be modelled via simple deterministic rules, because there is not much cognitive load and the decision rules are relatively simple [19, 40] . When forming consideration sets, it is also important to consider their size. It is very difficult to decide an optimal size based on the choice environment and individual processing capabilities [14].

With the current progress in computer science, mathematics and behavioral economics, recommender systems are ideal tools to solve this information overload problem and provide users with the most relevant consideration sets [41].

2.3. Recommender systems

Recommender systems (RS) have been an important part of our daily lives thanks to the rise of the Internet. RS are software tools and/or algorithms which match users to items [42]. One example is Netflix, which recommends a movie similar to the one the user just watched. The general purpose of any RS is to help users who do not have sufficient knowledge or experience, or the capacity to evaluate the item pool fully.

We distinguish between personalized and general recommender systems. Personalized RS may suggest different items to different users or user groups. General RS in turn, are usually directed towards the general public and might be relevant only to some part of it, for example, Billboard Hot 100, IMDB Top 250, or the front page of New York Times [43].

When RS face new users or new items, they may fail to provide personalized content due to the sparsity of information [44]. Because such RS mainly utilize historical interactions of similar users on similar items, and their ratings, facing a new entity about which it has no information makes it impossible to generate recommendations. This problem is referred to in the literature as the cold-start problem [2] and is considered a key challenge in RS design [45].

The literature distinguishes three main cold-start settings [45]: a) recommending existing items for new users (user-side), b) recommending new items for existing users (item-side), c) recommending new items for new users (user- and item-side). However, when trying to address this problem scholars have mainly focused on settings in which the challenge was to recommend new items to existing users [46].

Recently, some progress was made in solving the user-side cold-start problem after the introduction of contextual information into recommender systems. As a result of this effort, Context Aware Recommender systems were introduced [47]. In this approach the context refers to the time of the choice, the location or socio-demographic characteristics of the decision-maker etc. Some approaches have been very successful by combining contextual information with collaborative filtering [48, 49, 50]. Utilizing baseline information for new users [51] and using social network data [52] have also been proven to overcome the cold-start problem to a certain extent.

In practice however, these cold-start problems often transform into continuous cold-start problems [53]. This happens when:

1. The user stays "inactive" for a long period before the initial interaction
2. The user's interactions have a significant time window
3. The user creates a "one-time" account
4. It is not possible or permitted to track users, or (under GDPR) the user has requested their personal information to be removed from the system.

In the case of the continuous cold-start problem, the solutions suggested in the literature discussed above are not feasible. The first reason is that users generally do not need to create an account for interacting with some services, for example, watching videos on YouTube, searching for items on Amazon, or looking for airline tickets. Because of this, systems commonly treat different sessions by the same user as being by new users. Secondly, due to rising awareness of internet security and privacy, people tend to use incognito mode when they make searches [3] which disables most of the tracking, and user identification.

Our approach addresses the user-side continuous cold-start problem, which has not been thoroughly researched before. By utilizing only characteristics at item and search level we propose a novel RS design which is able to tackle the information sparsity. First, we use clustering [54] to quantify the contextual information both on the individual and the search level and we cluster empirically similar items together. Then, we use a hypergeometric sampling technique to generate the structure of the final choice set, meaning how many options from each cluster should be in the final choice set. Because our goal is to provide the design of the RS, the final stage of the choice set generation will consist of applying both stochastic [37, 38] and deterministic rules [19, 40, 17].

Variable	mean	st.dev	min	max
Price	647.12	1105.12	59.55	16997
Duration (in minutes)	518.98	555.04	70	2715
# of flights	2.94	0.95	2	6
# of airlines	1.25	0.45	1	5
Days before departure	32.36	38.03	0	340
Domestic travel	0.49	0.49	0	1
Intercontinental travel	0.06	0.23	0	1

Table 1. Summary statistics of main variables

3. Methodology

3.1. Data

We used a combined dataset of airline booking details matched with real-time queries. Because users do not have to create an account for searching and buying tickets, there is no user-side data available to the system. It makes this dataset ideal for our purposes. Booking details were extracted from the MIDT (Marketing Information Data Types) database and were dated between December 2013 – June 2014 [55]. It had information on booking details including but not limited to, price, duration, time, days before departure, booking office ID. By matching this data with the real time queries run by users on Amadeus S.A.S on office ID, passenger details, search timestamps, origin-destination pairs and other query data, we were able to detect options delivered to users and which options were chosen. It was not possible to know if a specific option was seen by the user. Yet, considering that the average time spent on finding a suitable flight option is around 3.5 hours [56], we assumed that most of these options were seen by users.

For some menus [1] we did not have a complete menu because the system had truncated very large choice sets. Also, some of the queries appeared to have incomplete information regarding one or more query attributes. We solved these issues by simply deleting those parts of the data from our analysis. As a result, our dataset consisted of 7,163 choice sessions with 368,735 options. Table 1 gives descriptive information about our main variables.

[1] Previous literature [7, 55] refers to choice sets as menus. Hereafter, these terms will be used interchangeably.

3.2. Clustering

Clustering is used to divide data into different groups where empirically similar elements belong to the same group and dissimilar ones are assigned to different groups [54]. By using clustering, we aimed to identify options which were similar in their context. We used two mainstream clustering algorithms: Affinity propagation (AP) and KMeans (KM).

AP identifies elements as nodes within a network and finds ones that are exemplars ("cluster centers") via recursively passing messages between all data points [57]. A big advantage of the algorithm is that the modeler does not need to set the optimal number or size of the clusters. This comes at a price of the complexity of $O(N^2)$ where N is the number of elements in the set.

The KM clustering algorithm is more popular than AP because of its low complexity [58]. Despite being computationally simple, it requires the optimal number of clusters (K) to be pre-set by the modeler. As in our setting the optimal number of clusters is not known *ex ante*, this requirement adds a layer of complexity to the calculations. We used the silhouette score [59] which enabled us to find an optimal K mathematically while minimizing the complexity added to the algorithm. The silhouette score is calculated by comparing every element to its neighbours within the same cluster and also to those elements outside of its cluster and has values between -1 and +1. A higher score means each object is well-matched to its cluster. A number of distance metrics can be used to calculate this score; we used Euclidean distance.

To quantify the context and clusters we created two variables that captured the characteristics of clusters: relative cluster size and relative cluster dispersion. The first accounts for the normalized number of options within that cluster. The second is derived using

$$\frac{\sum_{i=1}^{m} (x_i - \mu_k)^2}{\sum_{i=1}^{N} (x_i - \mu_M)^2},$$

where N is the number of options within the menu, m is the number of options within the cluster, μ_k is the centroid of cluster k to which x_i belongs and μ_M is the mass center of the menu. [5] discusses the clustering more thoroughly.

Before applying clustering we eliminated the possible scale effects of some dependent variables by using z-score normalization.

3.3. Two-stage choice

Our modeling process consisted of two stages. In the first stage, we modelled the structure of the final choice set and determined how many elements of each cluster should present in the consideration set. Next, we used stochastic and simple deterministic rules to select options following the structure obtained during the first stage.

First, the attractiveness measure of clusters within the menu was calculated. We defined the attractiveness measure as the probability of the given cluster to contain an actual choice and calculated it using the traditional multivariate logistic model [60]. Utilizing both descriptive information of options within clusters and the aforementioned cluster level characteristics as our covariates, we estimated those probabilities for every cluster in the menu according to

$$a_k = Pr(Y = 1|X_k) = \frac{exp(\beta X_k)}{1 + exp(\beta X_k)}, \quad (3)$$

where a_k is the attractiveness measure, X_k is the feature vector of the cluster k and β is a vector of coefficients.

Using cluster level characteristics allowed us to embed the contextual information of options within a cluster into our model. Then, using this metric we determined the structure of the final choice set via hypergeometric sampling.

Let N be the number of options within the menu which belong to k unique clusters and $m_i \in M$ be the number of options that belong to cluster i, so that $\sum_{i=1}^{k} m_i = N$. If we sample n random options from that menu without replacement we get a set $J = \{j_1, j_2, j_3, \ldots, j_k\}$ which follows the hypergeometric distribution and the probability of getting such vector J is determined by

$$P(j_1, j_2, \ldots, j_k) = P(J) = \frac{\binom{m_1}{j_1}\binom{m_2}{j_2}\ldots\binom{m_k}{j_k}}{\binom{N}{n}}, \quad (4)$$

where j_k is the number of elements belonging to cluster k in our sample.

However, using M and N does not allow us to quantify the menu context in terms of its clusters, which was our goal. One way of avoiding this limitation is to use the attractiveness measure instead of M in our sampling. Yet, because the attractiveness measures are in the range of zero to one, it was impossible to use them directly in our sampling. So, we defined the *attractiveness score* of a cluster as $s_k = a_k * 1e6$, where a_k is the attractiveness measure of a cluster k. The constant $1e6$ was chosen to account for the smallest differences between two almost identical a_k.

Accordingly, we replaced N with $D = \sum_{i=1}^{k} s_i$. So equation (4) becomes

$$P(j_1, j_2, \ldots, j_k) = P(J) = \frac{\binom{s_1}{j_1}\binom{s_2}{j_2}\ldots\binom{s_k}{j_k}}{\binom{D}{n}}. \quad (5)$$

To save computational time and overcome the sparsity of the vector J, during sampling we used only attractiveness scores of the top n most probable clusters. Because our n was assumed to be relatively small, we were able define beforehand all the possible J vectors such that $j_1 \geq j_2 \geq j_3 \ldots \geq j_k; k \in n$ using integer partitioning.

In order to make sampling results also dependent on M we used M as a constraint for J, so that

$$\forall j \in J, m \in M : j_i \leq m_i.$$

If this condition could not be satisfied for some i then we did the assignment $j_i \leftarrow m_i$ and the remainder $j_i - m_i$ was added to the leftmost possible element of J. Yet, for 108 menus, it was still the case that there was not a valid J complying with these rules. We simply removed those menus from our analysis.

To better understand the approach, let J be $[4, 3, 2, 1, \ldots, 0]$ and M be $[8, 2, 4, 1, \ldots, 6]$. Then, $j_2 \geq m_2$, which violates the constraint above. So, the assignment $j_2 \leftarrow 2$ is made and the remainder 1 is added to leftmost possible element of J. Our result becomes [5,2,2,1,. . . ,0].

Finally, by randomly sampling according to equation 5 one hundred thousand times we picked our most likely J by finding the most repeated sample. Then, we selected the top $j_1, j_2, \ldots, j_k$ options from the top n most probable clusters based on the attractiveness score of options obtained using equation 3.

After identifying the clusters and the number of elements to select from, we applied two methods to generate the final choice set. The first method was stochastic and consisted of randomly selecting elements according to vector J. The second method was deterministic and used the price of the option as a determinant. The cheapest options were selected according to J. As a baseline, we used the same two approaches, but selection was done disregarding J. Therefore, the baseline of the first method was the random selection of one option from every cluster. The baseline of the second method was the selection of the cheapest option from every cluster. Recall that we used two different clustering methods, AP and KM. Hence, we calculate four models per clustering method:

$Model1$. Random selection following J

$Model1b$. Random selection (baseline of model one)

$Model2$. Selection of the cheapest options following J

$Model2b$. Selection of the cheapest options (baseline of model two)

3.4. Performance metrics

To evaluate each model's performance we used accuracy at top-N, which is a commonly used metric not only in classification tasks but also in RS design studies, especially for context-based recommendations [43]. In classification, it measures whether the actual class is in the top N predicted classes of the model. Similarly, in RS design it measures if the chosen option is among the top N suggestions of the system. We conformed with the existing literature and selected accuracy at top-5 and top-10 as our evaluation metrics [61].

Accordingly, $n = 5$ and $n = 10$ were chosen. So, all possible J values were found via integer partitioning of five and ten, which gave us 7 and 42 possible variations accordingly.

4. Results

4.1. Clustering results

The results of clustering methods are clearly different. While AP tended to create fewer but larger clusters (7.62 on average), KM generally identified more clusters (10.29 on average) with relatively smaller sizes. This indicates that both algorithms were able to identify the contextual information but in different ways.

The runtime of these algorithms also differed considerably. Because AP did not need an initial number of clusters, while for KM we had to compute the optimal cluster count in each menu, for the same menu AP converged on average 7.2 times faster. This makes AP more viable for larger choice spaces.

4.2. First stage results

Table 2 gives descriptive information about the structure of the consideration sets for the different clustering methods. We notice the similarities between AP and KM in terms of the average number of clusters present in the choice sets. Despite the different contexts identified by those algorithms in the clustering phase, both algorithms appeared to identify the "important" clusters. We also see that KM resulted in more variance, yet generated less diverse consideration sets in general. On the contrary, AP appeared to be more robust when it came to different choice environment setups and was able to generate consideration sets that were more distinct.

	$n=5$		$n=10$	
	AP	KM	AP	KM
Mean	2.74	2.59	3.62	3.41
Standard deviation	0.84	1.16	0.98	1.61
Minimum	1	1	1	1
Maximum	5	5	7	10

Table 2. Consideration set structure in terms of unique clusters across clustering methods

4.3. Second stage results

We see that both clustering methods were also robust to the selection methods used in the second stage. This indicates that item-side contextual information helps capture the choice environment better and it also provides meaningful insights into the consumer behavior.

Both of our models outperformed their baseline counterparts considerably. Stochastic models performed in general better in KM than AP which is not surprising. The main reason for this is that KM identified smaller clusters and so the chance of randomly selecting a correct option was therefore higher. This difference decreased in cases where the selection was made based on deterministic rules.

The performance of our models using a deterministic rule to make the selections may indicate that consumers use multiple determinants as criteria during the decision-making process, which also complies with previous findings [11, 40]. Table 3 summarizes our second stage results.

5. Conclusion and discussion

We have proposed a novel approach to tackling the user-side continuous cold-start problem in RS design. By using the contextual information of the menu we were able to generate relevant choice sets using a two-step choice modeling approach. Our structural approach to choice set generation proved to be robust not only to selection criteria, be it stochastic or deterministic, but also to the clustering method used. Because in an online environment the calculation time is critically important, using AP as the clustering method appears to be advantageous.

	$n=5$		$n=10$	
	AP	KM	AP	KM
Model 1	0.39	0.40	0.56	0.55
Model 1b	0.21	0.23	0.21	0.25
Model 2	0.49	0.48	0.63	0.62
Model 2b	0.32	0.32	0.32	0.34

Table 3. Top-5 and top-10 accuracy scores across clustering methods

The findings of this work can be implemented by various systems which face continuous cold-start problems. They also help to understand the decision-making process of consumers and hence reduce their search cost by introducing the most relevant alternatives. This also benefits the supply-side via the reduction of the overall time spent by users on the platform.

This work has some limitations. RS design using one-stage simple MNL probabilities would result in 52% and 65% in top-5 and top-10 accuracy, respectively. Such random utility models violate Luce's choice axiom [21], which states that the choice probabilities of options in the choice set must be equally affected by the introduction or removal of a new option. However, one possible way to improve our approach could be the integration those probabilities into our models. Another possible avenue for future research could be using more complex characteristics derived from the choice set along with clustering.

Acknowledgements

The authors acknowledge the support of the following programs operated by the French National Research Agency (Agence Nationale de Recherche): UCA^{JEDI} (ANR-15-IDEX-01) and DS4H (ANR-17-EURE-004). They would like to thank the participants of the SAMET20 workshop, in particular Rodrigo Acuna-Agost, Gijs Overgoor and Alix Lheritier, and anonymous referees for insightful comments.

References

[1] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-based systems*, vol. 46, pp. 109–132, 2013.

[2] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE transactions on knowledge and data engineering*, vol. 17, no. 6, pp. 734–749, 2005.

[3] A. I. Antón, J. B. Earp, and J. D. Young, "How internet users' privacy concerns have evolved since 2002," *IEEE Security & Privacy*, vol. 8, no. 1, pp. 21–27, 2010.

[4] R. K. Wong, V. W. Chu, and T. Hao, "Online role mining for context-aware mobile service recommendation," *Personal and Ubiquitous Computing*, vol. 18, no. 5, pp. 1029–1046, 2014.

[5] Z. Babutsidze, W. Rand, E. Mirzayev, I. Rafai, N. Hanaki, T. Delahaye, and R. Acuna-Agost, "Asymmetric dominance in airfare choice," in *6th International Choice Modelling Conference*, ICMC, 2019.

[6] A. Tversky and S. Sattath, "Preference trees.," *Psychological Review*, vol. 86, no. 6, p. 542, 1979.

[7] A. Lhéritier, M. Bocamazo, T. Delahaye, and R. Acuna-Agost, "Airline itinerary choice modeling using machine learning," *Journal of choice modelling*, vol. 31, pp. 198–209, 2019.

[8] Q. Liu and N. Arora, "Efficient choice designs for a consider-then-choose model," *Marketing Science*, vol. 30, no. 2, pp. 321–338, 2011.

[9] T. J. Gilbride and G. M. Allenby, "A choice model with conjunctive, disjunctive, and compensatory screening rules," *Marketing Science*, vol. 23, no. 3, pp. 391–406, 2004.

[10] Fishburn, "Exceptional paper-lexicographic orders, utilities and decision rules: a survey," *Management Science*, vol. 20, no. 11, pp. 1442–1471, 1974.

[11] J. R. Bettman, "Memory factors in consumer choice: A review," *Journal of Marketing*, vol. 43, no. 2, pp. 37–53, 1979.

[12] E. J. Johnson, R. J. Meyer, and S. Ghose, "When choice models fail: Compensatory models in negatively correlated environments," *Journal of Marketing Research*, vol. 26, no. 3, pp. 255–270, 1989.

[13] H. A. Simon, "Rational choice and the structure of the environment.," *Psychological review*, vol. 63, no. 2, p. 129, 1956.

[14] J. de Dios Ortúzar and L. G. Willumsen, *Modelling transport*. John wiley & sons, 2011.

[15] J. R. Hauser and B. Wernerfelt, "An evaluation cost model of consideration sets," *Journal of consumer research*, vol. 16, no. 4, pp. 393–408, 1990.

[16] J. R. Bettman, M. F. Luce, and J. W. Payne, "Constructive consumer choice processes," *Journal of consumer research*, vol. 25, no. 3, pp. 187–217, 1998.

[17] C. H. Coombs, "Mathematical models in psychological scaling," *Journal of the American Statistical Association*, vol. 46, no. 256, pp. 480–489, 1951.

[18] A. Tversky, "Elimination by aspects: A theory of choice.," *Psychological review*, vol. 79, no. 4, p. 281, 1972.

[19] J. R. Hauser, "Consideration-set heuristics," *Journal of Business Research*, vol. 67, no. 8, pp. 1688–1699, 2014.

[20] J. R. Hauser, M. Ding, and S. P. Gaskin, "Non-compensatory (and compensatory) models of consideration-set decisions," in *Proceedings of the Sawtooth Software Conference*, vol. 14, pp. 207–232, 2009.

[21] R. D. Luce, *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2012.

[22] R. R. Batsell and J. C. Polking, "A new class of market share models," *Marketing Science*, vol. 4, no. 3, pp. 177–198, 1985.

[23] D. H. Gensch, "A two-stage disaggregate attribute choice model," *Marketing Science*, vol. 6, no. 3, pp. 223–239, 1987.

[24] A. K. Manrai and P. Sinha, "Elimination-by-cutoffs," *Marketing Science*, vol. 8, no. 2, pp. 133–152, 1989.

[25] I. S. Currim, R. J. Meyer, and N. T. Le, "Disaggregate tree-structured modeling of consumer choice data," *Journal of Marketing Research*, vol. 25, no. 3, pp. 253–265, 1988.

[26] A. Aribarg, T. Otter, D. Zantedeschi, G. M. Allenby, T. Bentley, D. J. Curry, M. Dotson, T. Henderson, E. Honka, R. Kohli, *et al.*, "Advancing non-compensatory choice models in marketing," *Customer Needs and Solutions*, vol. 5, no. 1-2, pp. 82–92, 2018.

[27] J. W. Payne, J. R. Bettman, and E. J. Johnson, "Adaptive strategy selection in decision making.," *Journal of experimental psychology: Learning, Memory, and Cognition*, vol. 14, no. 3, p. 534, 1988.

[28] M. Paulssen and R. P. Bagozzi, "A self-regulatory model of consideration set formation," *Psychology & Marketing*, vol. 22, no. 10, pp. 785–812, 2005.

[29] J. R. Bettman, E. J. Johnson, and J. W. Payne, "A componential analysis of cognitive effort in choice," *Organizational behavior and human decision processes*, vol. 45, no. 1, pp. 111–139, 1990.

[30] A. D. Shocker, M. Ben-Akiva, B. Boccara, and P. Nedungadi, "Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions," *Marketing letters*, vol. 2, no. 3, pp. 181–197, 1991.

[31] G. Häubl and V. Trifts, "Consumer decision making in online shopping environments: The effects of interactive decision aids," *Marketing science*, vol. 19, no. 1, pp. 4–21, 2000.

[32] J. L. Horowitz and J. J. Louviere, "What is the role of consideration sets in choice modeling?," *international Journal of Research in Marketing*, vol. 12, no. 1, pp. 39–54, 1995.

[33] J. R. Hauser, "Testing the accuracy, usefulness, and significance of probabilistic choice models: An information-theoretic approach," *Operations Research*, vol. 26, no. 3, pp. 406–421, 1978.

[34] M. Ding, J. R. Hauser, S. Dong, D. Dzyabura, Z. Yang, S. Chenting, and S. P. Gaskin, "Unstructured direct elicitation of decision rules," *Journal of Marketing Research*, vol. 48, no. 1, pp. 116–127, 2011.

[35] S. Gaskin, T. Evgeniou, D. Bailiff, and J. Hauser, "Two-stage models: Identifying non-compensatory heuristics for the consideration set then adaptive polyhedral methods within the consideration set," in *Proceedings of the Sawtooth Software Conference*, vol. 13, pp. 67–83, Citeseer, 2007.

[36] M. Yee, E. Dahan, J. R. Hauser, and J. Orlin, "Greedoid-based noncompensatory inference," *Marketing Science*, vol. 26, no. 4, pp. 532–549, 2007.

[37] D. McFadden *et al.*, "Conditional logit analysis of qualitative choice behavior," 1973.

[38] G. L. Urban, P. L. Johnson, and J. R. Hauser, "Testing competitive market structures," *Marketing Science*, vol. 3, no. 2, pp. 83–112, 1984.

[39] G. Punj and R. Moore, "Information search and consideration set formation in a web-based store environment," *Journal of Business Research*, vol. 62, no. 6, pp. 644–650, 2009.

[40] B.-K. Lee and W.-N. Lee, "The effect of information overload on consumer choice quality in an on-line environment," *Psychology & Marketing*, vol. 21, no. 3, pp. 159–183, 2004.

[41] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," *arXiv preprint arXiv:1301.7363*, 2013.

[42] T. Mahmood and F. Ricci, "Improving recommender systems with adaptive conversational strategies," in *Proceedings of the 20th ACM conference on Hypertext and hypermedia*, pp. 73–82, 2009.

[43] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, *Recommender Systems Handbook*. Berlin, Heidelberg: Springer-Verlag, 1st ed., 2010.

[44] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, "Facing the cold start problem in recommender systems," *Expert Systems with Applications*, vol. 41, no. 4, pp. 2065–2073, 2014.

[45] S.-T. Park and W. Chu, "Pairwise preference regression for cold-start recommendation," in *Proceedings of the third ACM conference on Recommender systems*, pp. 21–28, 2009.

[46] Z.-K. Zhang, C. Liu, Y.-C. Zhang, and T. Zhou, "Solving the cold-start problem in recommender systems with social tags," *EPL (Europhysics Letters)*, vol. 92, no. 2, p. 28002, 2010.

[47] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," in *Recommender systems handbook*, pp. 217–253, Springer, 2011.

[48] M. Aharon, N. Aizenberg, E. Bortnikov, R. Lempel, R. Adadi, T. Benyamini, L. Levin, R. Roth, and O. Serfaty, "Off-set: one-pass factorization of feature sets for online recommendation in persistent cold start settings," in *Proceedings of the 7th ACM Conference on Recommender Systems*, pp. 375–378, 2013.

[49] S. Bykau, F. Koutrika, and Y. Velegrakis, "Coping with the persistent cold-start problem," *Personalized Access, Profile Management, and Context Awareness in Databases*, 2013.

[50] M. Saveski and A. Mantrach, "Item cold-start recommendations: learning local collective embeddings," in *Proceedings of the 8th ACM Conference on Recommender systems*, pp. 89–96, 2014.

[51] D. Kluver and J. A. Konstan, "Evaluating recommender behavior for new users," in *Proceedings of the 8th ACM Conference on Recommender Systems*, pp. 121–128, 2014.

[52] I. Guy, N. Zwerdling, D. Carmel, I. Ronen, E. Uziel, S. Yogev, and S. Ofek-Koifman, "Personalized recommendation of social software items based on social relations," in *Proceedings of the third ACM conference on Recommender systems*, pp. 53–60, 2009.

[53] J. Kiseleva, A. Tuzhilin, J. Kamps, M. J. Mueller, L. Bernardi, C. Davis, I. Kovacek, M. S. Einarsen, and D. Hiemstra, "Beyond movie recommendations: Solving the continuous cold start problem in e-commercerecommendations," *arXiv preprint arXiv:1607.07904*, 2016.

[54] L. Rokach and O. Maimon, "Clustering methods," in *Data mining and knowledge discovery handbook*, pp. 321–352, Springer, 2005.

[55] A. Mottini and R. Acuna-Agost, "Deep choice model using pointer networks for airline itinerary prediction," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1575–1583, 2017.

[56] S. Corporation, "Europeans spend longer searching for flights than they do experiencing them," 2017.

[57] D. Dueck and B. J. Frey, "Non-metric affinity propagation for unsupervised image categorization," in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8, IEEE, 2007.

[58] S. Lloyd, "Least squares quantization in pcm," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

[59] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[60] M. E. Ben-Akiva, S. R. Lerman, and S. R. Lerman, *Discrete choice analysis: theory and application to travel demand*, vol. 9. MIT press, 1985.

[61] P. Cremonesi, Y. Koren, and R. Turrin, "Performance of recommender algorithms on top-n recommendation tasks," in *Proceedings of the fourth ACM conference on Recommender systems*, pp. 39–46, 2010.