Understanding Customer Choices to Improve Recommendations in the Air Travel Industry

Alejandro Mottini
Amadeus SAS
Sophia Antipolis, France

Alix Lhéritier
Amadeus SAS
Sophia Antipolis, France

Rodrigo Acuna-Agost
Amadeus SAS
Sophia Antipolis, France

Maria A. Zuluaga
Amadeus SAS
Sophia Antipolis, France

ABSTRACT

Recommender systems aim at suggesting relevant items to users to support them in various decision-making processes, on the basis of available information on items or users. In the latter, the customer's interests and tastes can be learnt and expressed using historical browsing data, purchase histories, and even other non-traditional data sources such as social networks. Despite its proven success in the on-line retailing industry, in electronic commerce and, even tourism, recommender systems have been less popular in flight itinerary selection processes. This could be partially explained by the fact that customers' interests are only expressed as a flight search request. As a result, this problem has been historically tackled using classical Discrete Choice Modelling techniques and, more recently, through the use of data-driven approaches such as Machine and Deep Learning techniques. At Amadeus, we are interested in the use of choice models with recommender systems for the problem of airline itinerary selection. This work presents a benchmark on three family of methods to identify which is the most suitable for the problem we tackle.

KEYWORDS

Choice Modeling; Choice-based Recommendations; Air Travel Industry

1 INTRODUCTION

In the recent years, recommender systems (RecSys) have proven invaluable for solving problems in the on-line retail industry and e-commerce[15]. While tourism has not been the exception to this success [3], with applications covering almost every area of the travel and hospitality industry [14], RecSys have been less popular on the airline itinerary decision-making process. This can be explained by two factors. On one hand, the available information about users and items is not as rich as for most RecSys in tourism. In the traveller's flight itinerary choice problem, *i.e.* the task of selecting a flight given a proposed list of itinerary recommendations, the user's interests are only expressed as a flight search request, user sessions are usually anonymous and there is no user history in the travel provider's databases. Therefore, classical RecSys algorithms cannot be applied directly.

On the other hand, RecSys techniques suffer from a lack of theoretical understanding of the underlying behavioural process that led to a particular choice [6] by seeing the decision-making process as a *black box* [7]. Collaborative and content-based methods recommend items based on similarities among users or items but, cannot provide further insight. In the flight industry, it is key to understanding passenger behaviour and their flight itinerary preferences. Players in the sector use this knowledge to adapt their offers to market conditions and customer needs, thus having an impact on airline's revenue management and price optimisation systems [4].

To tackle the flight itinerary choice problem and overcome these limitations, the airline industry has historically resorted to Discrete Choice Modeling (CM). Due to its good performance, efficiency and ease of interpretation, the Multinomial Logit model (MNL) [11], a specific CM technique is the most popular approach for the flight itinerary choice problem. In spite of its numerous advantages, CM also presents some weaknesses. For instance, MNL only considers linear combinations of the input features, limiting its predictive capability and requiring expert knowledge to perform feature engineering. Also, they lack the flexibility to handle collinear attributes and correlations between options and it is difficult to model individual's heterogeneities. These shortcomings might be overly restrictive or affect performance [12]. As an example, industrial applications require to develop different models for distinct markets. In the case of the flight itinerary choice prediction problem, this involves estimating models at a city-pair level [5] and/or customer demographic segments [19].

In an effort to cope with CM limitations, recently machine learning and deep learning techniques have been proposed. These algorithms can more easily model non-linear relationships and handle correlated features, and have more modelling power which allows to predict choices on an individual level, thus improving the prediction performance.

Inspired by the work from Chaptini [6], at Amadeus we are working towards the use of CM with recommender systems for the problem of airline itinerary selection. Combining the two approaches should leverage the strengths of both, leading to robust and scalable, but more interpretable models. In this first work, we seek to explore, evaluate and compare three different CM models which can be used as the predictive back-bone of a choice-based RecSys framework. In the remainder of this paper, first we present the theoretical background of CM and demonstrate why CM can be seen as a RecSys problem. Then, we present our experimental setup by describing the data, the evaluated algorithms and the performance measures.

2 BACKGROUND

In this section, first we provide a brief background on classical discrete choice modelling theory and then show how it is equivalent to the recommendation problem.

2.1 Discrete Choice Models.

CM defines four basic components: 1) the decision-maker, 2) the alternatives, 3) the attributes, and 4) the decision rules [2]. Formally stated, a decision-maker $i \in I$ chooses from a choice set A_i composed of J_i alternatives, with with $j \in \{1, \ldots, J_i\}$ the index of the j^{th} alternative. For the sake of simplicity and without loss of generality, we will refer to the number of alternatives simply as J, although decision-makers might not be faced with the same set and/or number of alternatives. The decision-maker i obtains an utility U_{ij} from each j and chooses alternative $\hat{j}$ if and only if:

$$U_{i,\hat{j}} \geq U_{ij};\ \forall\, j \in A_i. \tag{1}$$

The utility function is unknown and not observable. However, as it is possible to determine the attributes $\mathbf{x}_{ij}$ perceived by decision-maker i for each j, as well as S_i the vector of characteristics of i, there exists a function $V(\cdot)$ which relates the observed features to the decision-maker's utility:

$$V_{ij} = V(\mathbf{X}_{ij}), \tag{2}$$

where V_{ij} is referred to as the representative utility and $\mathbf{X}_{ij} = h(\mathbf{x}_{ij}, \mathrm{S}_i)$, a simplified representation of $\mathbf{x}_{ij}$ and S_i through the use of any appropriate vector valued function h. V_{ij} is generally a linear combination of the features. For example, if an airline is trying to predict which itinerary a user will choose, a very simple model could be:

$$V_{ij} = a * price_{ij} + b * tripDuration_{ij}$$

with a, b parameters of the model to be estimated, and which are commonly refered to as β.

Since there are aspects of the utility function that cannot be observed, $V_{ij} \neq U_{ij}$. To reflect uncertainty, the utility can be modelled as a random variable,

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \tag{3}$$

where ε_{ij} is a random variable that captures the unknown factors that affect U_{ij}. As U_{ij} is now a random variable, the decision rule needs to be expressed as the probability that decision-maker i chooses the k^{th} alternative:

$$P(k|A_i) = P(U_{ik} \geq U_{ij};\ \ \forall j \in A_i). \tag{4}$$

By replacing U_{ij} accordingly:

$$P(k|A_i) = P(V_{ik} - V_{ij} \geq \varepsilon_{ij} - \varepsilon_{ik};\ \ \forall j \in A_i). \tag{5}$$

Different assumptions about the random term ε_{ij} and the deterministic term V_{ij} produce specific models.

2.2 Choice-based Recommender Systems.

Given a set A_i of J available items presented to a user i, the recommender problem can be seen as an optimisation task that first estimates the utility of each item $j \in A_i$, and then chooses the item $\hat{j}$ that maximizes an utility function $U(i, j)$, representing the user's utility on any item j [1]:

$$\hat{j} = \underset{j \in A_i}{\arg\max}\ U(i, j). \tag{6}$$

Conceptually this is the same optimisation problem as that one formulated by choice theory [2], and described previously in this section. Equation (6) is equivalent to choosing the alternative with the highest utility for a decision-maker, in choice modelling theory. More formally:

$$\hat{j} = \underset{j \in A_i}{\arg\max}\ U(i, j) \Leftrightarrow U_{i\hat{j}} \geq U_{ij};\ \forall\, j \in A_i, \tag{7}$$

which implies that the recommendation problem can be seen as a choice prediction problem. Therefore, the models and techniques developed in CM can be applied to RecSys.

3 MATERIALS AND METHODS

3.1 Data

Experiments were conducted on real datasets of flight search logs and bookings from MIDT, an Amadeus database containing bookings from over 93000 travel agencies.

Bookings are stored using Personal Name Records (PNR), which are created at reservation time by airlines or other air travel providers, and are then stored in the airline's or Global Distribution System (GDS) data centers. PNRs contain the travel itinerary of the passenger, personal and payment information, and/or additional ancillary services sold with the ticket. As these only contain information about the purchased ticket (final choice), and not about the alternatives considered before the purchase, we must also consider flight search logs. These contain both itinerary requests (origin, destination and dates), and the different alternatives presented to the passenger.

Both data sources are combined into a final dataset containing the alternatives presented to each user and their final choice (Figure 1). The matching process is in itself a challenging problem due to the high volume of data (i.e., around 100 GB of daily search logs) and to the difference in data sources and formats. Moreover, the process cannot be perfectly accurate since there is not a direct link between the two data sources and booking/search times differ. An approximate matching is performed using data fields which are shared between booking and logs (i.e. origin, destination, time and booking agency).

The choice set presented to a user, which we denote a session, contains up to 50 itineraries. The features used for each alternative are summarized in Table 1. The considered dataset contains 33951 sessions split into training/tests sets.

3.2 Algorithms

Methods from three different families of algorithms, classical CM, machine learning- and deep learning-based CM, are explored.

3.2.1 Classical CM. Two classical CM approaches are considered: The Multinomial Logit (MNL) model [11], perhaps the most common CM model, and Latent class choice models (LCM) [8]. McFadden [11] demonstrated that if ε_{ij} is an i.i.d. Gumbel random variable, the probability that a decision-maker i chooses the

Table 1: Feature set classified according to owner (individual or alternative) and data type (numerical, categorical, binary or time).

Owner	Data Type	Features
Individual	Categorical	Origin, destination, office
	Numerical	Days to trip, Trip weekday
	Binary	Stays Saturday, Continental trip, Domestic trip
Alternatives	Categorical	Airline of first flight
	Numerical	Price, Stay length, Trip duration, Connections, Num. of airlines
	DateTime	Arrival time, Departure time

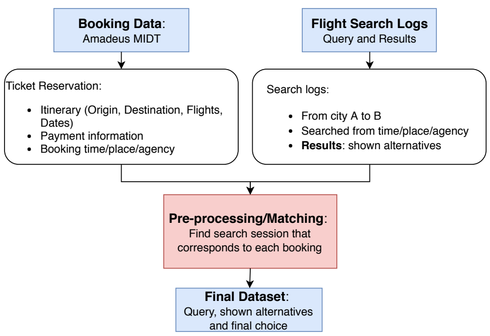

Figure 1: Dataset generation through MIDT bookings and search log matching.

alternative k (Eq. 5), the logit choice probability, is given by:

$$P(k|A_i) = \frac{\exp(V_{ik})}{\sum_{j \in J} \exp(V_{ij})}. \tag{8}$$

LCMs have been proposed to capture unobserved heterogeneity. Under LCM, the probability of choosing an alternative k can be expressed as:

$$P(k|\mathbf{X}_{ij}) = \sum_{q=1}^{Q} P(k|\mathbf{X}_{ij}; \beta_q, A_q) P(q|\mathbf{X}_{ij}; \theta) \tag{9}$$

where Q is the number of latent classes, β_q are the choice model parameters specific to class q, A_q is the choice set specific to class q, θ is an unknown parameter vector, and $\mathbf{X}_{ij}$ the simplified vector representation of attributes of alternatives and characteristics of decision-maker i.

Finally, both MNL and LCM models are optimized using maximum likelihood estimation as they can not be solved in a closed form.

3.2.2 ML. Lheritier *et al.* a have proposed machine-learning based CM (ML) [10] technique which formulates the choice modelling problem as a supervised learning one through the use of Random Forests (RF), a learning algorithm based on an ensemble of decision trees. The training data consists of the set of sample pairs $\mathcal{T} = \{(\mathbf{X}_{ij}, y_{ij})\}$ [1], with y_{ij} the binary indicator of whether decision maker i chooses the j-th alternative. As RF assumes independence of the samples, at training stage, every $\mathbf{X}_{ij}$ is assumed i.i.d., even if they belong to the same decision-maker. At prediction, each unseen alternative $\mathbf{X}_{ij}$ is propagated through the trained forest to obtain the posterior probability of being chosen:

$$P(y_{ij}|\mathbf{X}_{ij}) = \frac{1}{T} \sum_{t=1}^{T} P_{l_t}(y_{ij}(\mathbf{X}_{ij}) = 1) \tag{10}$$

where T denotes the number of trees and $P_{l_t}(\cdot)$ denotes the posterior probability function of a leaf node l in tree t. However, the alternatives associated to an individual's session cannot be treated as independent. There is an inherent dependence among them: only one alternative per session can be selected. To cope with this, the predicted probabilities are considered scores used to rank the alternatives. More formally, the index $\hat{j}$ of the selected alternative $a_{\hat{j}}$ by decision-maker i is:

$$\hat{j} = \underset{1 \le j \le J}{\arg\max}\ P(y_{ij}|\mathbf{X}_{ij})$$

3.2.3 DL. The assessed Deep learning choice modeling (DL) method [13] is based on an encoder-decoder network architecture using a modified pointer-network mechanism [18]. As with ML, the model is trained to predict the chosen alternative using a supervised learning approach. However, DL does not break the i.i.d. assumption among samples, as ML-based CM does. Given the sequential nature of pointer networks, sessions are represented as sequences of itineraries, $Z = \{\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iJ}\}$, which are fed sequentially to the model. The encoder network "encodes" the input into a hidden (encoder) state e. The decoder network will use the encoded information to output a vector u. Finally, a softmax function use the decoder's output to estimate the posterior probability of being chosen for the k^{th} element in the input sequence Z :

$$P(y_k = 1|Z) = \frac{\exp(u_k)}{\sum_{j=1}^{J} \exp(u_j)} \tag{11}$$

with $u_k = d^T W_1 e_k$, the pointer vector to the k^{th} element of Z, e_k the k^{th} encoder state, $d = \tanh(W_2 e_J)$ the decoder, W_1, W_2 learnable parameters and y_k the binary indicator of whether k was chosen ($y_k = 1$) or not. $P(y_k = 1|Z)$ can be interpreted as an estimate of $P(k|A_i)$.

3.3 Performance measurement

We used Top-N accuracy to asses and compare the models. Top-N accuracy evaluates if the user's choice is among the top-N predicted alternatives. It is equivalent to the commonly used top-N error in image classification [16], as it can be formulated in terms of the latter as:

$$accuracy = 1 - error$$

4 RESULTS

Figure 2 presents the Top-N accuracy for MNL, ML and DL methods. Overall, DL presents the highest accuracies across all values of N. These results are confirmed, in more detail, in Table 2 where Top-1, 5 and 15 accuracies are detailed. Top-15 accuracy has a particular importance for ranking flight search recommendations since most websites show approximately 15 results per page.

[1]In the context of RF, $\mathbf{X}_{ij}$ referred to as the feature vector of a sample

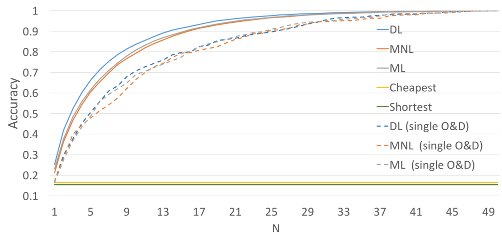

Figure 2: Top-N accuracy using the full data set (solid line) and a subset of the dataset consisting of a single origin/destination (O&D) pair (dashed line).

Table 2: Top-N with $N = 1, 5$ and 15. The best result for each N is presented in bold. Trivial choices of cheapest and shortest flight are included for reference.

Method	Top-1.	Top-5	Top-15
DL	**25.3**	**66.37**	**93.1**
ML	23.1	61.7	92.9
MNL	21.2	60.6	86.4
Cheapest	16.4	16.4	-
Shortest	15.4	15.4	-

To simulate data segmentation, a second experiment was performed in a simplified subset containing a single origin-destination (O&D) pair chosen at random. This resulted in 1617 decision-makers (users) with an associated booking to the O&D. The Top-N accuracy curve (Figure 2 dashed lines) shows how the difference in performance between the methods is less significant w.r.t. that one using the full data set. Despite MNL being the simplest method, results show that, on simpler datasets, it is able to perform as well as more complex methods.

This behaviour explains the motivation behind dataset pre-segmentation often used in classical CM. This is further confirmed by investigating the performance of LCM, as a function of the number of latent classes Q. Figure 3 reports top-1 accuracy of LCM, ML and DL, and demonstrates how it is possible to increase classic CM accuracy in complex data through a good estimation of Q. While MNL reported accuracies lower than ML and DM, LCM can outperform them when Q is estimated correctly. This improvements comes, however, at some cost: LCM requires additional hand engineered features to achieve the segmentation and a good choice of Q.

Although ML and MNL are not as accurate as DL, they have the advantage of having less hyper-parameters to tune. Moreover, they are more interpretable than DL. ML methods based on RF are known for their capacity to provide information on feature importance (Figure 4). This type of information can help to understand the rationale behind the decision-maker's choices, which can be important for some applications in the air travel industry.

5 FINAL REMARKS

RecSys research has so far predominantly focused on optimizing the algorithms used for generating recommendations to increase precision [9]. Precision measures how well the suggested alternatives match a decision-maker's profile based on previous data. While this is an important criterion, its limited assessment of a recommender quality has been criticized for not taking the decision-makers' situational needs into account [9]. Due to their well-known readability, Discrete Choice Modelling appears as a natural alternative to overcome this current limitation of RecSys. However, despite CM being a well-studied problem in various fields of research, literature on its use with recommender systems is very scarce. Existing works have adopted classical CM in combination RecSys [6, 17], while suggesting CM as a promising paradigm in the field of RecSys.

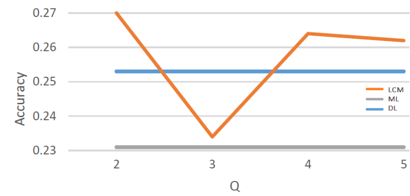

Figure 3: Top-1 accuracy of LCM, , as a function of the number of latent classes Q, compared to ML- and DL-based approaches.

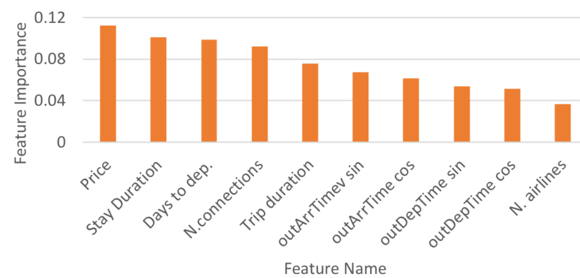

Figure 4: Top 8 feature importance for the ML method.

However, classical choice models tend to suffer from scalability issues as expert knowledge is usually required for model optimisation. ML- and DL-based [10, 13] choice models are non-parametric approaches that overcome this limitation, easing the deployment of choice-based RecSys at large scale. On the down side, model readability can diminish. Although this might not be relevant for some applications, understanding the reasons behind a decision-maker's choice is of high relevance in the air travel industry. ML-based methods appear to be a suitable compromise into readability but, they make strong assumptions on the independence of data that is arguable. Overall, it is possible to say that there is no ideal method and that the selection of one might depend on the specific recommendation application that they target. As a guideline, Table 3 summarises the strengths and pitfalls of the different methods here evaluated when considering choice-based RecSys.

At Amadeus, we work towards the development of informative, readable and interpretable RecSys that suit the needs of the air travel industry. Our hypothesis is that the combination of discrete choice modeling with RecSys can provide improvements to current systems in the air travel industry by keeping readability while improving performance. In that sense, an ML method like the random

Table 3: Advantages and disadvantages of the different families of CM methods.

Method	Advantages	Disadvantages
CM	• Simple and interpretable • Accurate on simple cases	• Feature engineering is required • Limited in handling big data
ML	• Interpretable • Accurate • Suitable for big data • Handles non-linear and latent relationships	• Assumes independence of samples • Feature engineering might be required
DL	• No assumptions on data • Highly accurate • Suitable for big data • Handles non-linear and latent relationships	• Non-interpretable • Many hyper-parameters • Computationally expensive

forests evaluated here represents a good compromise and a promising path to pursue in what we are looking for. On one hand, the method provides information on the relevance of features. On the other one it avoids the limitations of classical CM models. In that sense, although DL approaches have higher accuracy, they are not as advantageous given their limited interpretability.

This work represents an initial benchmark that evaluates three families of CM methods in the context of flight itinerary selection/recommmendation. Our future work will focus in the development of a unified framework that can leverage the strengths of the explored CM methods.

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. 2005. 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* 17, 6 (2005), 734–749.

[2] Moshe Ben-Akiva and Michel Bierlaire. 1985. *Discrete choice analysis: theory and application to travel demand.* MIT press, Cambridge, MA.

[3] Joan Borràs, Antonio Moreno, and Aida Valls. 2014. Intelligent tourism recommender systems: A survey. *Expert Systems with Applications* 41, 16 (2014), 7370–7389.

[4] J. Broder and P. Rusmevichientong. 2012. Dynamic pricing under a general parametric choice model. *Operations Research* 60, 4 (2012), 965–980.

[5] Judit G Busquets, Antony D Evans, and Eduardo Alonso. 2016. Predicting Aggregate Air Itinerary Shares Using Discrete Choice Modeling. In *16th AIAA Aviation Technology, Integration, and Operations Conference.* 4076.

[6] Bassam H. Chaptini. 2005. *Use of discrete choice models with recommender systems.* Ph.D. Dissertation. Massachusetts Institute of Technology.

[7] Li Chen, Marco de Gemmis, Alexander Felfernig, Pasquale Lops, Francesco Ricci, and Giovanni Semeraro. 2013. Human Decision Making and Recommender Systems. *ACM Transactions on Interactive Intelligent Systems* 3, 3 (2013), 1–7.

[8] William H Greene and David A Hensher. 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological* 37, 8 (2003), 681–698.

[9] Joseph A. Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction* 22, 1-2 (2012), 101–123.

[10] Alix Lhéritier, Michael Bocamazo, Thierry Delahaye, and Rodrigo Acuna-Agost. 2018. Airline Itinerary Choice Modeling using Machine Learning. *International Journal of Choice Modeling* (2018), https://doi.org/10.1016/j.jocm.2018.02.002.

[11] Daniel McFadden. 1973. Conditional Logit Analysis of Qualitative Choice Behaviour. In *Frontiers in Econometrics*, P. Zarembka (Ed.). Academic Press New York, New York, NY, USA, 105–142.

[12] D. McFadden. 2001. Economic choices. *The American Economic Review* 91, 3 (2001), 351–378.

[13] Alejandro Mottini and Rodrigo Acuna-Agost. 2017. Deep Choice Model Using Pointer Networks for Airline Itinerary Predictions. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, New York, NY, USA, 1575–1583.

[14] Julia Neidhardt, Tsvi Kuflik, and Wolfgang WÃűrndl. 2018. Special section on recommender systems in tourism. *Information Technology & Tourism* 19, 1-4 (2018), 83–85.

[15] Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor (Eds.). 2011. *Recommender Systems Handbook.* Springer Nature, New York.

[16] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115, 3 (2015), 211–252.

[17] Paula Saavedra, Pablo Barreiro, Roi Duran, Rosa Crujeiras, María Loureiro, and Eduardo Sánchez Vila. 2016. Choice-Based Recommender Systems. In *RecTour@ RecSys - Workshop on Recommenders in Tourism held in conjunction with the 10th ACM Conference on Recommender Systems (RecSys).* CEUR Workshop Proceedings, Boston, MA, USA, 38–46.

[18] O. Vinyals, M. Fortunato, and N. Jaitly. 2015. Pointer networks. In *Advances in Neural Information Processing Systems (NIPS 2015).* Curran Associates, Inc., Montreal, Canada, 2692–2700.

[19] V. Warburg, C. Bhat, and T. Adler. 2006. Modeling demographic and unobserved heterogeneity in air passengersâĂŹ sensitivity to service attributes in itinerary choice. *Journal of the Transportation Research Board* 1951 (2006).

