

Choice-based recommender systems

Paula Saavedra
CITIUS
University of Santiago de
Compostela
Santiago de Compostela,
Spain
paula.saavedra@usc.es

Rosa Crujeiras
School of Mathematics
University of Santiago de
Compostela
Santiago de Compostela,
Spain
rosa.crujeiras@usc.es

Pablo Barreiro
CITIUS
University of Santiago de
Compostela
Santiago de Compostela,
Spain
pablobv70@gmail.com

María Loureiro
School of Business
University of Santiago de
Compostela
Santiago de Compostela,
Spain
maria.loureiro@usc.es

Roi Durán
CITIUS
University of Santiago de
Compostela
Santiago de Compostela,
Spain
roiduram@gmail.com

Eduardo Sánchez Vila
CITIUS
University of Santiago de
Compostela
Santiago de Compostela,
Spain
eduardo.sanchez.vila@usc.es

ABSTRACT

Choice-based models are proposed to overcome some of the limitations found in traditional rating-based strategies. The new approach is grounded on decision-making paradigms, such as choice and utility theories. Specifically, random utility models were applied in a recommendation problem. Prediction accuracy was compared with state-of-art rating-based algorithms in a gastronomy dataset. The results show the superior performance of choice-based models, which may suggest that real choices could bring more predictive power than ratings.

CCS Concepts

•Information systems → Collaborative filtering; Social recommendation;

Keywords

Choice models; Random Utility Models; Logit probabilities; Tourism

1. INTRODUCTION

Recommender systems are personalization tools aimed at suggesting relevant items on the basis of available information on items as well as decision-makers [5]. Broadly speaking, recommenders can be classified in two different categories. Content-based recommenders generate a profile for each decision-maker by considering items experienced in the past. The profile typically represents the preferences of the decision-maker, i.e the taste of the decision-maker on each

item's attributes [2]. These preferences can be used to predict the utility of any given item by comparing them with the values of item's attributes. Collaborative recommenders, on the other hand, take advantage of previous ratings provided by the available decision-makers to predict the utility of any given user-item pair [6]. This approach has been widely adopted as it removes the burden of knowing and managing item attributes as well as their corresponding values.

Many algorithms and models have been proposed under the collaborative paradigm. Among them, two families have gained major attraction: neighborhood algorithms and latent factor models. The neighborhood approach was the first to implement to collaborative concept and became the reference model in this research area [9, 4]. The method consists on representing vectors of ratings on either the decision-maker or item space. The distance between any pair of these vectors determine the similarity between either the decision-makers or the items that these vectors represent. Individuals with similar rating's vectors are considered to possess similar tastes or preferences, while items are considered to have similar attributes. The latent factor strategy, in turn, attempts to explain ratings by means of characterizing both users and items with a limited set of factors. Factors are considered unknown variables that can be inferred from the ratings declared by the users. The inference or learning problem can be solved with factorization techniques. The classical factorization method is called Singular Value Decomposition (SVD) and was applied successfully to identify and reduce the number of relevant factors [10]. However, the method requires complete knowledge of the rating matrix and fill-in methods to populate sparse rating matrix come at a cost of inaccurate factor learning. Recently, new factorization techniques have been successfully developed that are capable of learning the factors from sparse rating matrices [7]. Each rating is explained by means of two vectors whose dimensions correspond with the set of latent factors. The first vector represents the item in terms of its degree of possession of each factor, while the second vector represent the decision-maker on the basis of her preference on each factor. These item and decision-maker vectors constitute a pair of

matrices whose values have to be inferred. The learning problem is solved by means of minimizing the regularized error on the set of known ratings.

Despite the success of current recommender systems, the experience with state-of-art approaches reveal some important limitations. First, the degree of performance of a recommender algorithm depends on the specific issues of the problem at hand. Therefore, heuristic models and trial-and-error methodologies are often used to look for the best solution for any given situation. The problem may be approached in a more theoretical and consistent way if recommenders were considered as agents predicting the decisions taken by decision-makers. Under this scope, the first limitation could be stated as follows: *(L1) Current state-of-art approaches are mostly based on heuristic models rather than decision-making theories.* Second, some popular paradigms assume a direct relationship between preferences and ratings: (1) the neighborhood approach considers that decision-makers with similar ratings on a set of items will have similar preferences, and (2) factorization techniques assume that ratings can be the result of a product between item’s latent factors and decision-maker preferences about that factors. In these paradigms unobserved preferences are usually inferred from observed ratings. The issue here comes from the fact that ratings could be mostly explained by variables different to preferences. The quality of the item, the user-item context, and in general any factor involved during the process of experiencing the item, they all could provide more explanatory power about ratings than preferences do. Therefore, the second limitation could be described as follows: *(L2) Preferences are usually derived from ratings without any supporting evidence about the relationship between these variables.*

This work proposes choice-based recommender systems to overcome these limitations. The concept is grounded on choice and utility theory, where real choices replace ratings as the key data to learn the decision-maker’s preferences as well as to make recommendations. The proposed models are then evaluated in the tourism domain with a gastronomy dataset that includes both choices and ratings. In what follows, the choice-based models are presented, the methods are described, and the models evaluated and compared against state-of-art rating-based algorithms. The discussion will comment on the results and highlight the major contributions of the paper.

2. CHOICE MODELS

2.1 Recommendation as a choice problem

The recommendation problem can be described as an optimization problem which consists on (1) estimating the utility of each item $a \in A$, the available item set, for any given decision-maker c , and (2) choosing the item a' that maximizes $U(c, a)$, the decision-maker utility on any item a [1]:

$$a' = \arg \max_{a \in A} U(c, a) \quad (1)$$

It is worth noting that this problem is conceptually the same as the one faced by the Rational Choice Theory, which aims at explaining economic behaviour under choice situations [11]. The theory states that a decision-maker will maximize her utility after satisfying some budget constraints. More formally, the decision-maker will choose alternative a' from a choice set A according to the following rule:

$$CR(A, \succeq) = \{a' \in A \mid a' \succeq a, \forall a \in A\} \quad (2)$$

where CR stands for "choice rule" and the \succeq operator denotes the relationship "preferred to, or at least as preferred as". Basically, it means that the chosen alternative will be the one from which the decision-maker shows a higher preference. The preference operator needs to be quantified to allow a numerical comparison between the alternatives.

The utility theory comes to the rescue to solve this problem. One of the axioms of this theory states that it is possible to define a utility function such that:

$$a \succeq b \iff U(a) \geq U(b). \quad (3)$$

And then, the choice rule in equation 2 can be represented in terms of the utility function and a numerical operator:

$$CR(A, \geq) = \{a' \in A \mid U(a') \geq U(a), \forall a \in A\}. \quad (4)$$

It is now clear that the new choice rule is mathematically equivalent to the recommendation problem described in equation 1:

$$a' = \arg \max_{a \in A} U(c, a) \iff CR(A, \geq) = \{a' \in A \mid U(a') \geq U(a), \forall a \in A\}. \quad (5)$$

As the recommendation problem can be understood as a choice prediction problem, then the powerful models and techniques developed in this field can be naturally applied to generate recommendations.

2.2 Choice models with random utility

The choice rule models how decision-makers take their decisions. However, the problem of predicting such decisions is a different task. In real problems the researcher does not have access to all the factors and variables that decision-makers include to estimate utilities. For a concrete individual c_n , the researcher only knows some attributes of the alternatives, labeled x_j for all a_j alternatives with $j \in \{1, \dots, J\}$, and some attributes of the decision-maker, labeled z_n . A function that relates these observed factors to the decision-maker’s utility can be specified. This function is denoted by $V_{nj} = V(x_j, z_n)$ and it is often called representative utility. It usually depends on parameters that are unknown and, therefore, they must be estimated.

Since there are aspects of utility that the researcher does not or cannot observe, $V_{nj} \neq U_{nj}$. Therefore, the utility can be decomposed as:

$$U_{nj} = V_{nj} + \epsilon_{nj} \quad (6)$$

where ϵ_{nj} captures the unknown factors that modify the utility and are not included in V_{nj} . This decomposition is fully general, since ϵ_{nj} is defined as simply the difference between true utility U_{nj} and the part of utility that the researcher captures in V_{nj} . Given its definition, the characteristics of ϵ_{nj} , such as its distribution, depend critically on the researcher’s specification of V_{nj} . The researcher does not know ϵ_{nj} for all j and therefore these terms are considered random variables that allow the researcher to make probabilistic statements about the decision-maker’s choice. The models derived under this assumptions are called random utility models (RUM) [8].

Now, the choice rule of equation 4, which is deterministic under the decision-maker perspective, becomes probabilistic

under the perspective of the researcher. Then the rule for a decision-maker c_n choosing alternative a_i is:

$$CR(A, \geq) = \{a_i \in A \mid \mathbb{P}_i \geq \mathbb{P}_j, \forall a_j \in A\} \quad (7)$$

and the probability \mathbb{P}_i is estimated as follows:

$$\mathbb{P}(U_{ni} > U_{nj} \text{ for all } j \neq i) = \mathbb{P}(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \text{ for all } j \neq i). \quad (8)$$

If the joint density of $\epsilon_n = (\epsilon_{n1}, \dots, \epsilon_{nj})$ is denoted by f , this cumulative probability can be rewritten as:

$$\mathbb{P}_{ni} = \int_{\epsilon} \mathbb{I}(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \text{ for all } j \neq i) f(\epsilon_n) d\epsilon_n \quad (9)$$

where \mathbb{I} is the indicator function, equaling 1 when the term in parentheses is true and 0 otherwise. This is a multidimensional integral over the density of the unobserved portion of utility, $f(\epsilon_n)$. Different choice models are obtained from different specifications of this density, that is, from different assumptions about the distribution of the unobserved portion of utility. In addition, the choice of the density determines whether the integral takes a closed form or not [12].

2.3 Standard and mixed logit models

The simplest and most widely used choice model is the standard logit model [8]. It is derived under the assumption that the each unobserved portion of utility ϵ_{nj} is distributed independently, identically extreme value. In this case, f denotes the density for Gumbel distribution:

$$f(\epsilon_{nj}) = e^{-\epsilon_{nj}} e^{-e^{-\epsilon_{nj}}}. \quad (10)$$

Following [8], the logit choice probability that decision-maker c_n chooses alternative i is

$$\mathbb{P}_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}. \quad (11)$$

This model presents a clear interpretation. According to equation 11, if V_{ni} rises, reflecting a matching between the observed attributes of the alternative and the preferences of the decision-maker, with V_{nj} for all $j \neq i$ held constant, \mathbb{P}_{ni} approaches one. And \mathbb{P}_{ni} approaches zero when V_{ni} decreases, since the exponential in the numerator approaches zero as V_{ni} approaches $-\infty$.

The representative utility is usually specified to be linear in the set alternative's attributes: $V_{nj} = \beta_{nj} \cdot x_j$, where x_j is a vector containing, as before, the observed variables of the alternative a_j , and β_{nj} denotes the model coefficients vector which describes the preferences of decision-maker c_n on the attributes of the alternatives a_j . The preferences β_{nj} (model coefficients) are estimated by fitting equation 11 to a dataset of choices. Moreover, since the logit probabilities take a closed form, maximum likelihood procedures are applied for estimation. Concretely, the probability of person c_n choosing the alternative that he was actually observed to choose can be expressed as

$$\prod_i \mathbb{P}_{ni}^{y_{ni}},$$

where $y_{ni} = 1$ if the individual chooses i and zero otherwise. Since $y_{ni} = 0$ for non-chosen alternatives and \mathbb{P}_{ni} raised to the power of zero is 1, this term is simply the probability of the chosen alternative. Assuming that decision-maker's choices are independent, the probability of each individual

choosing the alternative that she was observed actually to choose is

$$L(\beta) = \prod_n \prod_i \mathbb{P}_{ni}^{y_{ni}}$$

where β denotes the vector of all model parameters. Therefore, the log-likelihood function is

$$LL(\beta) = \sum_n \sum_i y_{ni} \log \mathbb{P}_{ni}$$

and the estimator is the value of β that maximizes this function. Importantly, it was proved that the log-likelihood function with these choice probabilities is globally concave in parameters β , which helps in the numerical maximization procedures, see [8] for more details.

A well-known issue of standard logit model deals with capturing the heterogeneity of population [12]. The importance that decision-makers place on each attribute of the possible choices varies, in general, over decision-makers. Although logit model is able to represent the taste variation related to observed characteristics of the decision-maker, it can not represent differences in tastes that can not be linked to observed characteristics. Therefore, if taste variation is at least partly random, a logit model with random parameters should be considered instead. Under this considerations, β is now a vector of random coefficients and these coefficients vary over decision-makers in the population with density g . In most applications that have actually been called mixed logit, g is specified to be continuous. For example, it can be specified to be normal, lognormal, uniform, triangular or, even, gamma. Therefore, this density is a function of parameters θ that represent, in the gaussian case, the mean and covariance of the random coefficient in the population. Then, the choice probabilities can be written as:

$$\mathbb{P}_{ni} = \int \left(\frac{e^{V_{ni}(\beta)}}{\sum_j e^{V_{nj}(\beta)}} \right) g(\beta|\theta) d\beta. \quad (12)$$

Since the previous integral has not a closed form, it must be evaluated numerically through simulation. Once the researcher specifies a distribution g for the coefficients, the parameters θ maximizing the simulated log-likelihood must be estimated. Then, R draws of the coefficients are taken from g and the logit probabilities are computed for every draw. The unconditional probability in equation 12, that is the expected value of the conditional probabilities, is estimated as the average of R probabilities determined previously.

3. METHODS

The performance of choice-based models is compared with a choice of relevant rating-based algorithms from a gastronomic dataset containing the choices of snacks made by a set of decision-makers and their corresponding tapa ratings. The dataset is described in Sections 3.1 and 3.2. Technical details on the two recommendation alternatives considered in this work are briefly presented in Sections 3.3 and 3.4. Finally, the error criteria used to compare them are introduced in Section 3.5.

3.1 Experiment

In the context of the RECTUR project, an experiment was carried out with real users in the context of Santiago(é)Tapas, a gastronomic context that takes place every

year in Santiago de Compostela. In 2011 the fourth edition was held with a total of 56 participating restaurants proposing and elaborating up to three tapas that were sold at a price of 2 euro. The experiment was designed to gather relevant data while preserving the spirit of the contest. Participants were local users as well as Spanish and international tourists. A TapasPassport with the official information about the contest was made available to all participants. It contained: (i) the contest guidelines and other related information to the participants, (ii) restaurants location, (iii) the tapas offered on each restaurant, (iv) an official seal to demonstrate that a participant has visited the minimum number of restaurants required to obtain contest's gifts. Restaurant staff had to sign the TapasPassport to certify that its owners have visited the place.

After consuming a tapa, participants were asked to evaluate their experience. Users had to provide two ratings ranging from 0 to 5: (i) a rating of the tapa, and (ii) a rating of the overall experience (service, place atmosphere, etc.). In addition, they were informed about our research experiment and asked to extend their feedback providing information about the temporal and social context in which the experience took place.

3.2 RECTUR Dataset

The data gathered in the experiment was collected in the RECTUR dataset. It is assumed that the choice of a tapa depends on the user preferences about the levels of tapa attributes, which will in turn depend on the user attributes and context elements. The consumption of a tapa determines a choice from a choice set and will elicit a satisfaction response quantified as a user rating.

For each tapa, we gathered the following attributes:

- Choice sets. Different choice sets could be defined for each choice. We acquired information about the following sets:
 - Set of tapas in the same area of the city (outlying, new or old zone).
 - Set of tapas in the same restaurant.
- Tapa attributes. The gathered attributes are:
 - Type: Cheese, egg, fish, meat, vegetable, shellfish and other. The main ingredient defined the type of the tapa.
 - Character: Traditional or daring. Traditional tapas are those that follow popular well-known recipes, while daring tapas are creative and provide innovative recipes.
 - Restaurant. The restaurant that offers the tapa was also categorized in terms of its location (outlying, new or old area), atmosphere and style.
- Rating. The rating provided by each consumer.

3.3 Choice-based models

The standard logit model as well as the mixed logit model assuming Gaussian distribution on the coefficients, both described in Section 2.3, were chosen as basic representatives of the family of random utility choice-based models to be compared with rating-based algorithms. From attributes type and character of each tapa described in Section 3.2, eight

binary variables associated to each alternative (or snack) were generated for fitting these two models. Next, the construction of the variables is briefly described through an example. The choice set associated to the old area contains, as possible choices, the set of tapas distributed in restaurants of this zone. For each one of these snacks, the dichotomous variables cheese, egg, fish, meat, vegetable, shellfish and traditional are generated. According to Figure 2, the main ingredient of t100 is meat. However, this tapa is not traditional. Therefore, only the variable meat will be equal to 1. The rest of variables associated to t100 will take the value zero.

Within the discrete choice framework, the set of alternatives known as the choice set must verify three properties. It has to be finite, exhaustive (the decision-maker always chooses one of the alternatives) and mutually exclusive (the choice of one alternative necessarily implies not choosing any of the other ones). Due to the last property, three different choice subsets were established in this work. They correspond to the three possible restaurant locations (old, new and outlying areas of the city). Therefore, standard and mixed logit models are estimated separately from these three choice subsets that contain only the tapas associated to each zone. This assumption could be less general. For instance, considering the set of tapas of a concrete restaurant would provide a new choice set and, as consequence, a new choice problem.

Estimations results for these six models are shown in Section 4.2. For the same area of the city, standard and mixed logit models present similar estimations for the coefficients. As consequence, only prediction accuracy of the standard logit model was compared with rating-based algorithms.

3.4 Baselines: Rating-based models

The proposed choice-based models were compared with two popular rating-based models: User-based collaborative filtering (UBCF) and matrix factorization (MF). User-based collaborative filtering assumes that individuals with similar preferences will rate items in a similar way. Then, missing ratings for a concrete user c_n could be predicted finding a neighborhood $N(n)$ of similar users and aggregating their ratings to calculate the corresponding prediction. The concept of similarity between users is used for defining this neighborhood given all users within a similarity threshold. In this work, the cosine similarity measure is taken into account and $|N(n)|$ was fixed equal to 25. For an item i and an individual c_n , the ratings predicted, \hat{r}_{ni} , can be written as

$$\hat{r}_{ni} = \frac{1}{|N(n)|} \sum_{j \in N(n)} r_{ji}$$

where $||$ denotes the cardinal of $N(n)$.

Matrix factorization, on the other hand, characterizes both items and users by vectors of factors inferred from item rating patterns. For a given item i and a user c_n , the vector q_i measure the extent to which the item possesses those factors and the vector p_n , the extent of interest the user has in items that are high on the corresponding factors. The dot product $q_i^T p_n$ captures the user's interest in the item's characteristics. This approximates user c_n 's rating of item i , r_{ni} , leading to the estimate

$$\hat{r}_{ni} = q_i^T p_n.$$

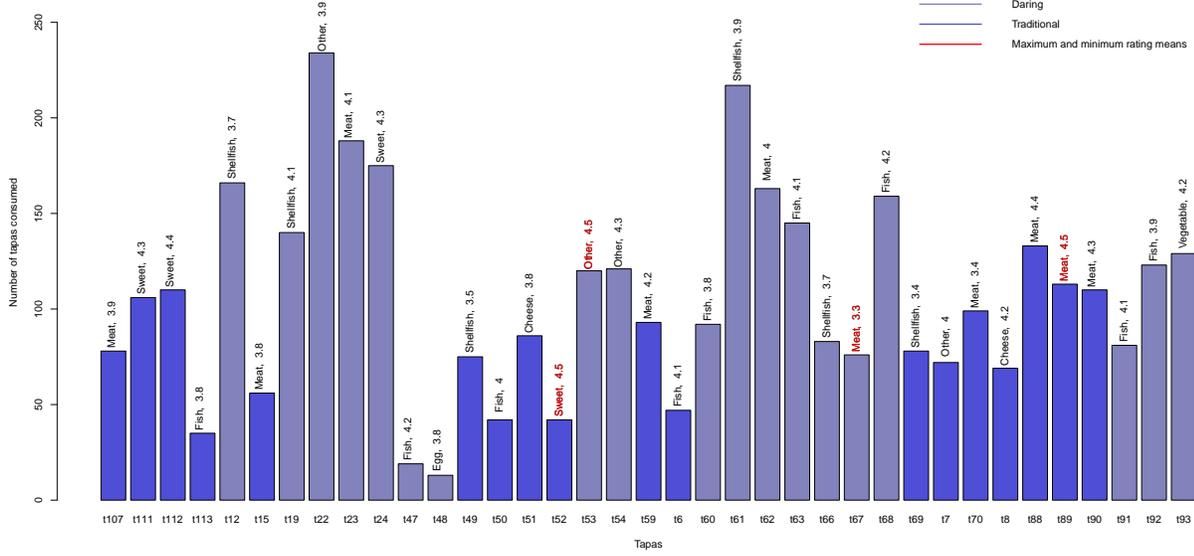


Figure 1: Bar plot for number of different tapas consumed, main ingredient and mean of users' ratings in the new zone of the city.

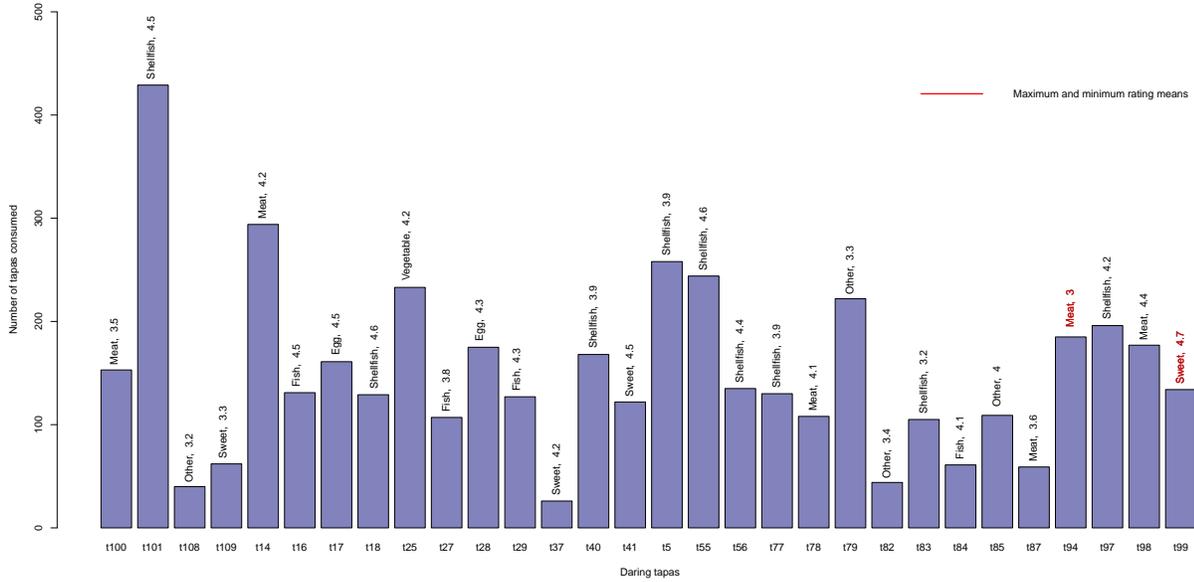


Figure 2: Bar plot for number of different daring tapas consumed, main ingredient and mean of users' ratings in the old zone of the city.

Therefore, the challenge is computing the mapping of each item and user to vectors q_i and p_n . Here, singular value decomposition will be applied factoring the user-item rating matrix that could be sparse. In order to learn the factor vectors (p_n and q_i), the regularized squared error on the set of known ratings is minimized:

$$\min_{q^*, p^*} \sum_{(u, i) \in K} (r_{ni} - q_i^T p_n)^2 + \lambda (\|q_i\|^2 + \|p_n\|^2)$$

where K is the set of the (c_n, i) pairs for which r_{ni} is known,

$\| \cdot \|$ is the Euclidean norm and λ denotes a constant controlling the extent of regularization. In this work, $\lambda = 1.5$.

3.5 Evaluation

Classical ranking error metrics could not be applied mainly because of the lack of information about all the relevant tapas for the decision-maker on any choice situation. Therefore, two error metrics are proposed in order to compare the behaviour of choice-based and rating-based algorithms. The metrics are described considering that only the tapas

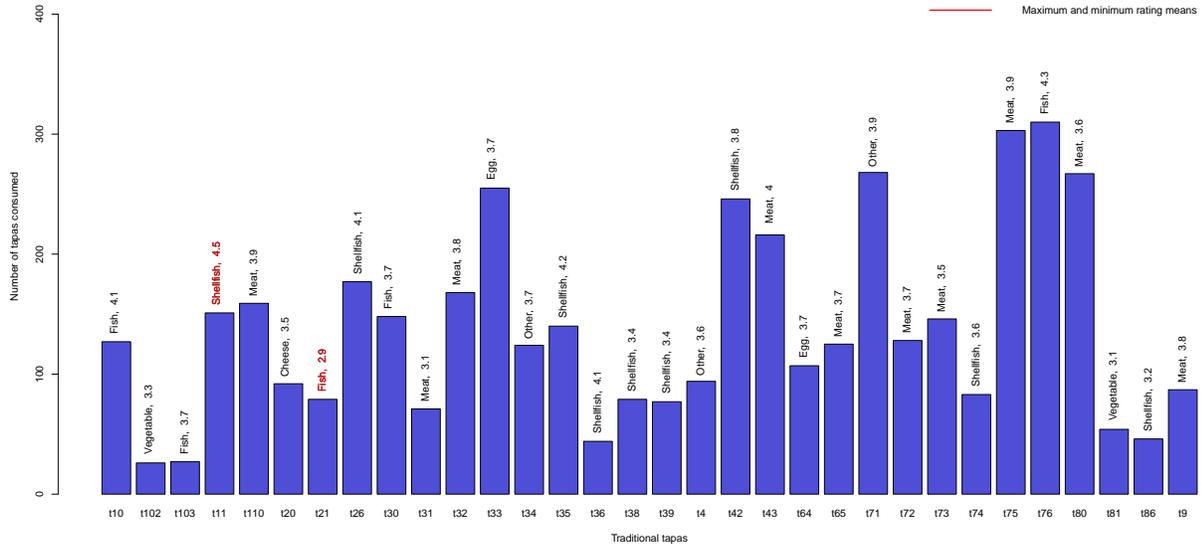


Figure 3: Bar plot for number of different traditional tapas consumed, main ingredient and mean of users' ratings in the old zone of the city.

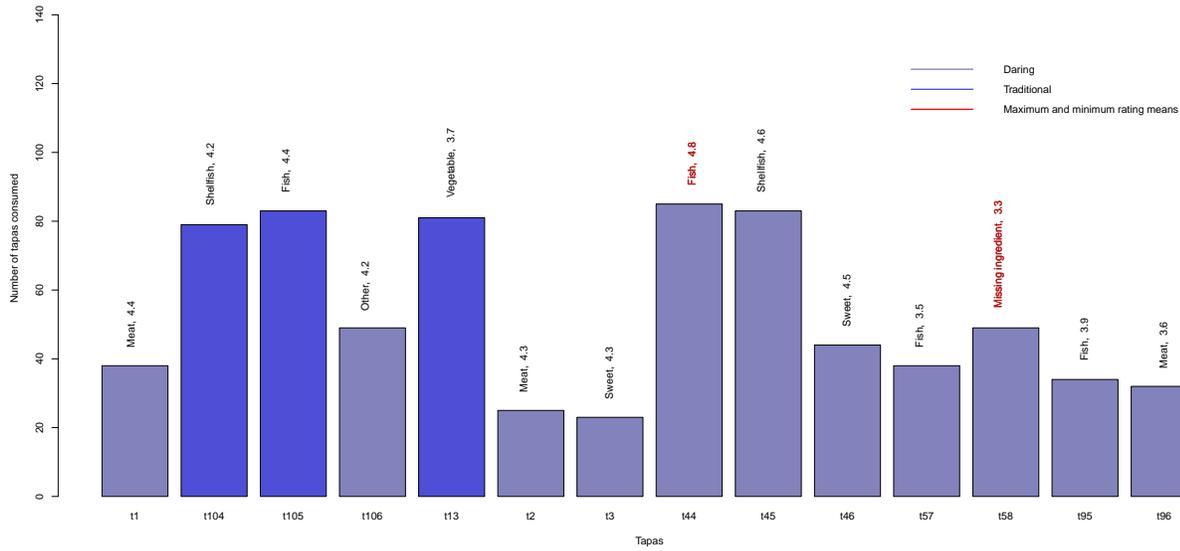


Figure 4: Bar plot for number of different tapas consumed, main ingredient and mean of users' ratings in the outlying zone of the city.

with the highest associated rating or probability is recommended/predicted (top 1). Error I is equal to one if the item predicted does not coincide with the true alternative chosen by the individual and zero otherwise. Therefore, given an individual c_n , the true choice i and the recommended item j is:

$$\text{error I}(c_n, i) = \begin{cases} 1 & : \text{if } i \neq j \\ 0 & : \text{otherwise.} \end{cases}$$

The second measure of error, error II, is equal to the position of the real choice in the ordered list of recommendation minus one. Therefore, if the item recommended is equal to the chosen one then the error is equal to zero. Let $(i_1, \dots, i_k, \dots, i_j)$ be the list of ordered items to be recommended, the error for the user c_n with true choice i can be

written as:

$$\text{error II } (c_n, i) = k - 1 \text{ if } i_k \neq i.$$

For instance, if one decision-maker c_n chose the snack t1 among the snacks (t1, t104, t105, ...) and the prediction (ordered according to the highest ratings or probabilities) is equal to (t105, t104, t1, ...), then error I ($c_n, t1$) = 1. However, error II ($c_n, t1$) = 2.

Error I and error II can be generalized easily if a list of a concrete number of ordered items (in terms of probabilities or ratings) is recommended instead of recommending only one alternative. These two errors are equal to zero if, for an individual c_n , the true choice i belongs to the recommended list of items. Otherwise, error I will take the value one and error II, the position of the true choice i in the ordered list of non-recommended alternatives. In this work, a list of five alternatives will be considered (top 5).

4. RESULTS

4.1 Data description

RECTUR dataset presented in Section 3.2 deals with 5517 individuals, that make one or a sequential choices of one tapa among a set of 113 tapas distributed in Santiago de Compostela. According to comments in Section 3.3, three subsets of the original dataset will be considered distinguishing three different choice contexts or, equivalently, three zones of the city.

Next, the three scenarios will be briefly described.

The total number of tapas consumed in *new area of the city* is 3888. However, the number of different tapas associated to this zone is only 37; 18 of them present a traditional character and 19, a daring character. Furthermore, the number of users in this area is 2030. Then, although most of these individuals had only one snack, some of them took several ones. Figure 1 shows the total number of tapas that users consumed for the 37 possible choices. According to the results, t22 and t61 were the most common choices. However, t47 and t48 were rarely selected. According to the information in Figure 1, only one tapa is made of eggs and vegetables; two tapas has cheese as main ingredient; four tapas are made of a sweet component or other; shellfish is the ingredient of six snacks; meat and fish are the most common components with ten and nine tapas, respectively. Tapa ratings that users gave to one consumed tapa are available too. The values of these ratings are 0, 1, 2, 3, 4 and 5. High values for ratings are associated to a high customer satisfaction. Means of tapa ratings for the 37 tapas in the new area are shown in Figure 1. The lowest means of tapa ratings are associated to t67, t70, t69 and t49. However, all of these means are greater than 3. So, the level of satisfaction tends to be high.

As for the *old zone of the city*, a total of 8948 tapas were consumed. As before, the number of different snacks associated to this zone is only 62; 32 of them present a traditional character and 30, a daring character. Furthermore, the number of users in this area is 3953. As before, although most of these individuals had only one snack, some of them took several ones. Figures 2 and 3 show the total number of daring and traditional tapas that users consumed for the 62 possible choices, respectively. According to the results, t101 was the most common choice. However, t37, t103 and t102 were rarely selected. As regards means of snack ratings, the

lowest ones correspond to t21 and t94. The highest ones, to t11 and t99.

The number of snacks consumed in the *outlying area of the city* is 743. Again, the number of different snacks associated to this zone is smaller. Concretely, it is equal to 14; 3 of them present a traditional character and 11, a daring character. Furthermore, the number of users in this area is 436. Figure 4 shows the total number of daring and traditional tapas that users consumed for the 14 choices. According to the results, t44, t45, t104 and t105 were the most chosen snacks. However, t2 and t3 were rarely selected. The snacks t58 and t44 correspond to the tapas with lowest and highest means of ratings, respectively. The main ingredient of t58 is a missing value. In addition, cheese and egg are not the main component for any snack.

4.2 Choice models fitting

The standard and mixed logit models have been fitted from the three choice sets described in Section 4.1. Due to the price is the same for every snack, the determinants of these choices, x_j , are eight dichotomous alternative specific variables. Seven of them indicate the main component of each tapa: Cheese, egg, fish, meat, shellfish, sweet and vegetable. The eighth variable takes value equal to one when the snack has a traditional character. In addition, for mixed logit model, Gaussian distribution was assumed on the coefficients and $R = 100$ was fixed.

	New zone	Old zone	Outlying zone
Cheese	-0.07	-0.25	
Egg	-2.48	0.31	
Fish	-0.46	-0.02	0.14
Meat	0.06	0.28	-0.44
Shellfish	-0.03	0.21	0.38
Sweet	0.07	-0.46	-0.38
Vegetable	-0.18	-0.17	0.26
Traditional	-0.62	-0.15	0.24
Log-Likelihood:	-13772	-36757	-1913.8

Table 1: Estimation by maximum likelihood of the standard logit model coefficients for different areas of the city. Significant coefficients are in black.

	New zone	Old zone	Outlying zone
Cheese	-0.07	-0.24	
Egg	-2.48	0.31	
Fish	-0.46	-0.01	0.13
Meat	-0.07	0.27	-0.67
Shellfish	-0.03	0.21	0.37
Sweet	-0.003	-0.46	-0.38
Vegetable	-0.18	-0.17	0.26
Traditional	-0.93	-0.09	-0.01
Log-Likelihood:	-13631	-36680	-1897.9

Table 2: Estimation of the means for mixed logit model coefficients assuming normal distribution for different areas of the city. Significant coefficients are in black.

The coefficients obtained are shown for each area of the city in Tables 1 and 2, respectively, and most of them are significant in the three areas of the city. For the mixed logit

Error	Choice model		UBCF		MF	
	Top 1					
	CV ₁	CV ₂	CV ₁	CV ₂	CV ₁	CV ₂
I	0.895	0.876	1	1	1	1
II	5.057	4.741	8.885	9.006	8.868	8.824
Top 5						
I	0.408	0.409	1	1	1	1
II	1.841	1.859	6.262	6.295	6.128	6.115

Table 3: Cross validation predictions errors for standard logit choice model, user-based collaborative filtering and matrix factorization algorithms in the outlying area of the city. Random and leave-one-out cross validation are denoted by CV₁ and CV₂, respectively. In this zone, the number of different tapas to be recommended is 14.

model (Table 2), only the mean estimations of Gaussian distributions are shown. As for the utility, positive coefficients, see egg and meat in Table 1 for the old zone, increase its value. However, negative coefficients, see egg and traditional in Table 1 for the new area, reduce it.

4.3 Choice-based vs rating-based predictions

The behaviour of choice-based and rating-based models for recommending tapas in the three areas of the city was analyzed using random sub-sampling and leave-one-out cross validation from RECTUR dataset.

For random sub-sampling validation, 100 iterations were considered using the 25% of randomly selected individuals as test data for predictions. Therefore, in each iteration and once the 25% of decision-makers was randomly selected, the rest of individuals is used as training data for rating-based algorithms or for fitting the choice model. Then, for each decision-maker in the test data and for each recommendation method, prediction error measures introduced in Section 3.5 can be determined. The procedure for leave-one-out cross validation is similar. In this case, the number of iterations is equal to the number of users and, in each iteration, the test data contains an only decision-maker.

Tables 3, 4, and 5 contain the empirical means of errors described previously for the new, old and outlying areas of the city, respectively. According to results shown in Section 4.2, standard and mixed logit models provide similar estimations for model coefficients. Therefore, only the first choice-based model, the standard logit one, were taken into account to be compared with the rating-based algorithms.

The results show that choice-based models offer a better performance (lower prediction errors) compared with rating-based schemes (UBCF and MF). See, in particular, error II for the top 5 scheme taking into account the different number of tapas recommended in each area of the city. Furthermore, the accuracy of predictions is reduced as long as the choice set increases from the outlying to the old area, which indicates the importance of the choice set and the choice situation.

5. DISCUSSION

The main point of this work is that the recommendation problem can be considered as a choice prediction problem. This is the main difference of our proposal compared with current paradigms in recommender systems that focus on rating prediction. The key aspects of our choice-based mod-

Error	Choice model		UBCF		MF	
	Top 1					
	CV ₁	CV ₂	CV ₁	CV ₂	CV ₁	CV ₂
I	0.955	0.954	1	1	1	1
II	14.552	14.060	25.438	25.475	25.511	25.499
Top 5						
I	0.795	0.789	1	1	1	1
II	10.606	10.481	22.606	22.640	22.658	22.506

Table 4: Cross validation predictions errors for standard logit choice model, user-based collaborative filtering and matrix factorization algorithms in the new area of the city. Random and leave-one-out cross validation are denoted by CV₁ and CV₂, respectively. In this zone, the number of different tapas to be recommended is 37.

Error	Choice model		UBCF		MF	
	Top 1					
	CV ₁	CV ₂	CV ₁	CV ₂	CV ₁	CV ₂
I	0.982	0.987	1	1	1	1
II	27.005	26.921	43.908	43.862	43.843	43.787
Top 5						
I	0.905	0.904	1	1	1	1
II	23.141	23.097	41.013	40.964	40.945	40.970

Table 5: Cross validation predictions errors for standard logit choice model, user-based collaborative filtering and matrix factorization algorithms in the old area of the city. Random and leave-one-out cross validation are denoted by CV₁ and CV₂, respectively. In this zone, the number of different tapas to be recommended is 62.

els are: (1) preferences are learnt from choices, (2) the choice set of each choice situation is included as a relevant variable to both explain and predict future choices, and (3) unobserved factors affecting the decision-making process are captured through random variables. On the basis of these elements the models presented in this paper differ from both collaborative methods, as they infer preferences from ratings, and content-based techniques, as they do not handle the choice set of the items experienced in the past. Recent content-based approaches share the same idea about the utility of user choices to derive preferences but are limited to pairwise rather than full choice set comparisons [3].

With regard to the limitations stated in the introduction, choice models face issue *L1* by building random utility models from solid decision-making theories, and solve issue *L2* by using choices, rather than ratings, to estimate preferences. The drawback of gathering information about the domain (attributes and values) is compensated in two ways: (1) by using more accurate data, choices rather than ratings, and (2) by removing the burden of interrogating decision-makers about their post-experience satisfaction. In summary, choice modelling seems to be a promising paradigm in the field of recommender systems.

Acknowledgments

This research was sponsored by EMALCSA/Coruña Smart City under grant CSC-14-13, the Ministry of Science and Innovation of Spain under grant TIN2014-56633-C3-1-R, and the Ministry of Economy and Competitiveness of Spain under grant MTM2013-41383P.

6. REFERENCES

- [1] T. A. Adomavicius G. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.*, 17(6):734–749, 2005.
- [2] M. Balabanović and Y. Shoham. Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66–72, 1997.
- [3] L. Blédaité and F. Ricci. Pairwise preferences elicitation and exploitation for conversational collaborative filtering. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pages 231–236. ACM, 2015.
- [4] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 43–52. Morgan Kaufmann Publishers Inc., 1998.
- [5] G. M. Burke R., Felfernig A. Recommender systems: An overview. *AI Magazine*, 32(3):13–18, 2011.
- [6] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70, 1992.
- [7] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM, 2008.
- [8] D. McFadden et al. Conditional logit analysis of qualitative choice behavior. 1973.
- [9] V. H. Resnick P. Recommender systems. *Communications of the ACM*, 40(3):56–58, 1997.
- [10] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Application of dimensionality reduction in recommender system—a case study. Technical report, DTIC Document, 2000.
- [11] A. Sen. Rational behaviour. In *Utility and probability*, pages 198–216. Springer, 1990.
- [12] K. E. Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.