Similarity-based Decision Models*

Balázs Kovács

University of Lugano

July 5, 2010

THIS PAPER IS VERY MUCH IN THE DEVELOPMENTAL PHASE. PLEASE TREAT IT ACCORDINGLY. DO NOT CIRCULATE.

Abstract

This paper introduces similarity-based models of decision making for decision situations in which data on the attributes of options are not available. We propose that an actor, to evaluate a given option, can rely on the utility similar actors derive from this option, or on the utility she derived from similar options. We propose two models to assess the similarity of actors and options. We analyze the proposed models through numerical analyses, and illustrate their workings on a dataset in which reviewers rate restaurants. We find that the proposed decision making models provide reasonable prediction even if information is not available on the attributes of the options and/or the preferences of the actors. Finally, we discuss the findings and the implications to other domains of decision making.

"In reality, all arguments from experience are founded on the similarity which we discover among natural objects, and by which we are induced to expect effects similar to those which we have found to follow from such objects. (...) From causes which appear similar we expect similar effects. This is the sum of all our experimental conclusions."

David Hume: An Enquiry Concerning Human Understanding, 2004: 21 (1748)

*This paper benefitted from discussions with Jerker Denrell. All remaining errors are my own. Corresponding address: USI Institute of Management, Via Buffi 13, Lugano, Switzerland 6900. Email: kovacsb@usi.ch

1 Introduction

Imagine a situation in which an actor wants to evaluate an option, but she does not know the attributes of this option. All she can observe is how others evaluate this option, and how they have evaluated other options. What can such an actor do? In this paper, we outline decision making models that rely on the similarity between actors and the similarity between options. We propose two approaches the actor can follow: evaluate the option based on the utility other actors that are similar to her derive from this option; or evaluate the option based on the utility she derived from similar options.

The difficulty of similarity-based decision making, naturally, lies in how the actor is to assess the similarity of other actors to her and how to assess the similarity of other options to the focal one. While previous similarity-based decision and learning theories relied on attribute-based similarities (Shepard, 1987; Gilboa and Schmeidler, 2001, 2003; Gilboa et al., 2006; Gayer et al., 2007; Gilboa et al., 2009), attributes are not available to our actor[1]. The models we propose below overcome this difficulty by inferring the required similarities from earlier data on the focal actor's and others' decisions and observed utilities. We outline two approaches to similarity. The first approach proceeds as follows. The actor observes the extent to which other actors' evaluations correspond to her evaluations, and based on this she postulates a similarity ordering among the other actors. Then, she observes the evaluations of the other actors in the focal situation, and weighting the other actors' evaluations with the actors' similarities to her, she evaluates the option. Or, in another version of the first approach, based on others' previous evaluations, she evaluates how similar the focal option is to other options and weighs her previous evaluations with these similarities.

Consider the following example, which shall serve as the empirical illustration in this paper. You consider visiting a restaurant, but you do not know whether you would like it or not. Imagine that you do not know the attributes of the restaurant. All you know is how other people have rated this and other restaurants. Should you visit this restaurant? The model we outlined above would suggest you look up ratings on restaurants you have visited and see whether other reviewers' judgments agree with yours. Thus you can judge which reviewers have similar taste to yours

[1] One might think that attribute-similarity-based models can be applied to such options as well, treating other actors as attributes. This is only partly so, and the significant difference between the present model and attribute-based models shall become apparent when we introduce the second similarity-based decision model.

(you can do this even if you are not actually aware of your own taste), and see how they rated the restaurant you are considering. Or, based on other reviewers' ratings you can estimate the similarity of the focal restaurant to other restaurants you have previously visited, and decide based on whether you had liked restaurants similar to this. As we shall demonstrate later in the paper, these decision rules give a good estimate of whether you should visit the restaurant or not.

The second approach of similarity-based decision making builds on a generalized model of similarity, recently introduced by Kovács (2010). The generalized model, while it keeps the idea: do things that actors similar to you like, it builds on a generalized approach to similarity, and calls actors similar if they derive similar utilities from similar options, and calls options similar if similar actors derive similar utilities from them. To follow the restaurant-example: two restaurants are similar if similar people rate them similarly; and, concurrently, people are similar if they rate similar restaurants similarly. We believe that this second approach to similarity is more congruent with human similarity perception (Landauer and Dumais, 1997; Kovács, 2010), and we expect this approach to perform well especially in conditions of sparse data (where direct similarities are hard or impossible to obtain).

This paper explores the properties of these similarity-based decision models. The structure of the paper is as follows. First, we review the relevant literature of psychology, computer science, machine learning and economics, and we introduce our models for similarity-based decision making. Second, we investigate the behavior of the proposed models in simulated situations. Third, we illustrate and compare the proposed models on an online dataset of restaurants, reviewers, and reviews. In this setting, we investigate how the proposed similarity-based decision making models predict the ratings of reviewers. First we analyze the data as cross-sectional data, then we utilize the timing of the reviews and test whether the predictive power of the similarity-based models increase over time (Ram, 1993; Gonzalez et al., 2003). Finally, we conclude and outline avenues for further research.

2 Similarity, induction, decisions and inference

The models presented in this paper build on former concepts in philosophy, psychology, machine learning, and economics. Below we shortly review the relevant literatures, and indicate how our models differ from those discussed in the literature.

As the quote from Hume demonstrates, similarity-based argumentation and knowledge representation have been around since at least the 17th century. Hume (Hume, 2004 (1748), Locke (Locke, 1690) and others asserted that the basis of human concept formation and decision making is the perceived similarity between objects and other concepts.

Similarity also has been a central concept in cognitive psychology (Tversky, 1977; Medin et al., 1993). The interest in similarity-based reasoning (also called analogical reasoning) is supported by a sizable literature in psychology, and numerous studies demonstrate that similarity plays a central role in human inference making and learning. For example, studies show the importance of prior examples and similarity in learning how to use a text editor on the computer (Ross, 1984), recalling chess boards (Simon and Gobet, 1996), in diagnosis for car mechanics (Lancaster and Kolodner, 1987), and the explanation of strange events (Read and Cesa, 1991). Or, for an example from a different domain, Weinreb (2005) illustrates how analogical reasoning is applied in law. Shepard (1987) and Medin et al. (1995) summarize the role of similarity in learning and the structure of mental representations.

Building on the findings that similarity and analogies form the basis of human reasoning and decision making, scholars in machine learning and artificial intelligence developed the case-based reasoning paradigm (Riesbeck and Schank, 1989; Leake, 1996; Leake et al., 1997; McGinty and Wilson, 2009). This paradigm models the analogical reasoning as follows: recall cases that are similar to the focal case, and follow the action that you would follow in the similar case. Case-based reasoning has been applied to various fields, such as education (Leake et al., 1997), social network search, forest-management, or oil-drilling (for an introduction to recent developments, see McGinty and Wilson, 2009).

The first axiomatized version of the case-based decision theory have been pioneered by Itzhak Gilboa and David Schmeidler, who, together with co-authors, have developed a fruitful research field (for an overview on this line of research, see Gilboa and Schmeidler, 1995, 2001). In a series of papers, Gilboa and Schmeidler developed formal models of similarity-based decision making, they provided an axiomatization (Gilboa and Schmeidler, 2001, 2003; Billot et al., 2005 — more on this later), applied these models to statistical analysis, and used it on empirical data (Gayer et al., 2007). This paper builds on their work.

Despite this large body of literature, there has been relatively less attention given to the similarity function itself. In most approaches (with some important exceptions that we discuss later), the similarity assessment is based on attributes, and similarity is calculated as an inverse function of a distance function in a, usually metric, space. These approaches, however, do not apply to cases in which data on attributes are not available, or to cases in which the relative importance of the attributes is hard to assess. We, therefore, explore an approach that does not assume that actors have access to the attributes of the options. In this sense, the models in this paper are extensions to case-based decision making inasmuch as they apply in situations where attribute-data is not available.

3 Similarity-based decision models

To illustrate the intuition, let us go back to the decision situation we described in the Introduction: You are considering visiting a restaurant, but you do not know whether you would like it. A similarity-based decision making model would suggest you visit the restaurant if actors who have similar taste to yours have liked it. To assess the similarity of others, you just look at your past experiences of you two going to the same restaurant (not necessarily the same time), and see whether your evaluations on the restaurants coincide. If other actors liked the restaurants you like, and disliked the restaurants you dislike, then you can extrapolate from their experiences of the focal restaurant. This reasoning is illustrated on Figure 1.

The previous argument rests on a regularity assumption that similar people have similar tastes and that people follow their taste in making choices. Or, as Leake (1996) puts it in the language of case-based reasoning: "similar problems have similar solutions." This assumption is not too restrictive, and holds if (as will be assumed in the modeling part of this paper) (1) options' attributes are located in a multidimensional space with (any kind of) metric properties; (2) actors have ideal points in this multidimensional space; and (3) an actor's utility derived from the option is a monotone negative function of the distance between her ideal point and the option's attributes. It is important to note that in our model actors do not have to know their own ideal points or the attributes of the options, we just have to assume that these attributes and preferences exist and satisfy the regularity condition.

Now we turn to introduce the formal model. We have N actors ($N \in \mathbb{N}$), and M options ($M \in \mathbb{N}$). Let D denote the $N \times M$ matrix, which contains the observed utilities the actors derive

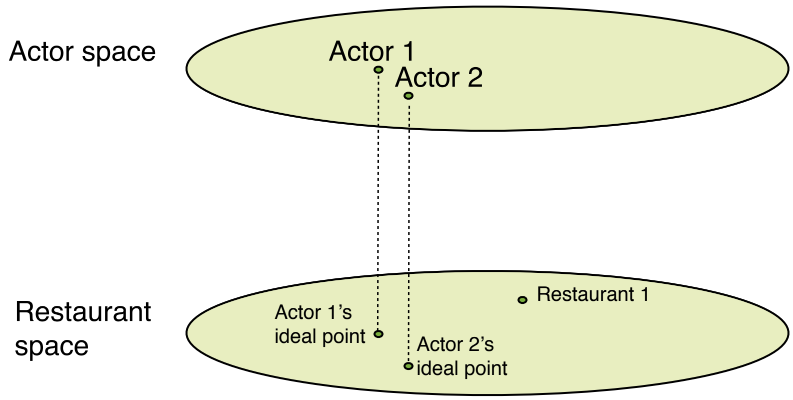

Figure 1: Illustration of the general idea of similarity-based inference and the regularity condition: similar actors have similar ideal points.

from the options[2]. In the restaurant example we discuss below, the cells of D can represent the ratings the reviewers gave to the restaurants. From D, we shall derive two similarity matrices. The first, SIM_A, is a $N \times N$ matrix, which contains the pairwise similarity of actors; and SIM_O, an $M \times M$ matrix, which contains the pairwise similarity of options.

Our model postulates that actors derive a utility level similar to what similar actors have derived from this option. We postulate that the utilities of other actors are weighted by their similarities of other actors to the focal actor (Gilboa and Schmeidler, 2001). Formally, if $Pred(i,j)$ denotes the predicted value of actor i's utility from option j, then

$$Pred(i,j) = \frac{\sum_{q=1..N,\, q\neq i} SIM_A(i,q) \times D(q,j)}{\sum_{q=1..N,\, q\neq i} SIM_A(i,q)}, \tag{1}$$

where $SIM_A(i,q)$ denotes the similarity of actors i and q, and $D(q,j)$ denotes actor q's utility from option j.

Alternatively, we can calculate the similarity of options, and using the data on how much the actor liked the other options, we can predict how much she would like the focal option. In this version, Equation (1) becomes

$$Pred(i,j) = \frac{\sum_{q=1..M,\, q\neq j} SIM_O(j,q) \times D(q,i)}{\sum_{q=1..M,\, q\neq j} SIM_O(j,q)}. \tag{2}$$

As to which of these two directions should be used, as we demonstrate later this depends on the

[2]We assume that the utilities are observable to every actor.

structure of the data.

Note that when the similarity values are all the same, the above formulas equal the simple averaging of the previous values. That is, to predict the utility the actor would derive from the option, these formulas average the utilities other actors derive from this option, or average the utilities the actor derives from other options. The similarity values thus give a weighting over other actors and options.

As Gilboa and Schmeidler (2003) and Gilboa et al. (2009) discuss, these averaging formulas are part of the family of kernel-based estimators in statistics (Akaike, 1954; Silverman, 1986), in which the similarity function serves as the kernel. In this sense, actors are assumed to follow a non-parametric statistical estimator.

Obviously, Equations (1) and (2) are not the only functional forms in which previous experiences and utilities can be weighted and combined for prediction (for example, the equation does not have to be additive or linear). To choose from the possible similarity-weighing functions, Gilboa and Schmeidler (2001, 2003) and Billot et al. (2005) introduced an axiomatic treatment of similarity-based decision making. They investigate what similarity-weighting functions satisfy some basic assumptions of choice ordering. Their main axiom is the combination axiom, which states that if a ranking of options holds in two separate databases, then it needs to hold in the union of the databases as well. Based on the combination axiom and some other axioms, Gilboa, Schmeidler, and colleagues have proven (Billot et al., 2005; Gilboa et al., 2006) that these axioms hold if and only if a similarity function satisfies Equations (1) and (2).

3.1 Assessing the similarity of actors and options

Given Equations (1) and (2), the question now is how to assess the similarity of actors and the similarity of options. Strangely, the approaches to similarity assessment have been less of a focus in the artificial intelligence and case-based decision making literatures. Usually, it had been assumed that we know how to assess similarity, and by this one usually meant attribute-based similarity with some kind of metrics and weightings.

Recently, Gilboa and colleagues have started to explore the estimation of the similarity function (Gilboa et al., 2006; Gayer et al., 2007; Gilboa et al., 2009). Their approach is as follows. First, they compile a dataset with the attributes of options, and assume a weighted-Euclidean distance between the options, with unknown weights. Then, relying on the prediction equation (their equivalent of Equation (1)), they estimate what weighting of the attributes is most likely to result in a similarity function that would predict the observed choices or utilities (Maximum Likelihood estimates). Following this approach, Gayer et al. (2007) investigate how rule-based and similarity-based models of reasoning predict the housing and rent pricing in a student website in Tel Aviv. They find that the similarity-based-reasoning model outperforms the rule-based model in predicting the rental prices. These research is important because this is the first (an to our knowledge only) empirical study on case-based reasoning.

The approach we take in this paper to operationalize similarity is somewhat different from Gayer et al. (2007), because in our models actors do not have access to attribute data, only the observed utilities of other actors.

3.1.1 Correlation

In situations in which attribute-data is not available, a common approach to measure similarity among actors is to take the cosine distance or Pearson-correlation between the row-vectors of D (Widdows, 2004; Kovács, 2010). We also use Pearson-correlation to assess the similarity for the first model. Two notes are due here. First, we choose to model similarity with correlation because (1) correlation is often used in the literature, and (2) because correlation has a generalized version which we need for the second model (Kovács, 2010). Second, we do not claim here that correlation necessarily is the model that humans follow, or that correlation is the optimal similarity measure (for information extraction purposes). These issues we leave for further research.

Equation (3) shows how Pearson-correlation is calculated for the similarity of actors i and j:

$$SIM_A^{corr}(i,j) = \frac{(D_{i,} - \overline{D_{i,}})(D_{j,} - \overline{D_{j,}})^T}{\sqrt{(D_{i,} - \overline{D_{i,}})(D_{i,} - \overline{D_{i,}})^T}\sqrt{(D_{j,} - \overline{D_{j,}})(D_{j,} - \overline{D_{j,}})^T}}, \tag{3}$$

where $D_{i,}$ denotes the ith row of the D matrix, $\overline{D_{j,}}$ denotes the vector composed of the mean of the jth row, and T denotes matrix transposition.

We calculate the similarity of options i and j as the correlation between the column vectors of D:

$$SIM_O^{corr} = \frac{(D_{,i} - \overline{D_{,i}})(D_{,j} - \overline{D_{,j}})^T}{\sqrt{(D_{,i} - \overline{D_{,i}})(D_{,i} - \overline{D_{,i}})^T}\sqrt{(D_{,j} - \overline{D_{,j}})(D_{,j} - \overline{D_{,j}})^T}}, \tag{4}$$

where $D_{,i}$ denotes the ith column of matrix D.

What values can similarity take? In the usual operationalization of similarity (e.g., Batagelj and Bren, 1995), the similarity values are between 0 and 1, 0 denoting perfect dissimilarity, and 1 denoting perfect similarity. Because Gilboa and Schmeidler assume that similarity is between 0 and 1, and because this is assumed in Equations (1) and (2), we shall also operationalize similarity as a function that maps onto $[0, 1]$. However, as correlation and the generalized similarity model operate in the $[-1, 1]$ range (-1 denoting perfect dissimilarity, 0 denoting independence or neutrality, and 1 denoting perfect similarity), we transform these similarity values to the $[0, 1]$ range. The specific rule of transformation we use in this paper is: $SIM_{new}(i, j) = (SIM_{old}(i, j) + 1)/2$, but further investigations demonstrate (results not shown here) that the main insights of the paper hold across alternative transformation rules.

3.1.2 Similarity based on the Generalized Similarity Model of Kovács (2010)

The second approach to assess the similarity of actors and options builds on the generalized similarity model recently introduced by Kovács (2010), which generalizes Pearson-correlation. A main shortcoming of Pearson-correlation is that it does not take the similarities among the options into account when comparing the actors, and does not incorporate the similarities among actors when comparing options. The generalized similarity model of Kovács (2010), building on a basic relationship in linear algebra (the scalar product of vectors x and y in a base space of A is xAy), provides a modified version of Pearson-correlation that incorporates the non-independence of dimensions. The main idea of the generalized measure is to use the option-similarity matrix, SIM_O, as a base space for calculating the actor similarity matrix ("actors are similar if they derive similar utilities from similar options"), and to use the actor-similarity matrix SIM_A as a base space for calculating the option similarity matrix ("options are similar if similar actors derive similar utilities from them").

Formally, if D denotes the original $M \times N$ actor-option utility matrix, SIM_A denotes the $M \times M$ actor-actor similarity matrix, and SIM_O denotes the $N \times N$ option-option similarity matrix, then the following equation describes the similarity of actors i and j (Note the similarity of this formula to the Mahalanobis-distance (Mahalanobis, 1936)):

$$SIM_A^{GSM}(i,j) = \frac{(D_{i,} - \overline{D_{i,}})\, SIM_O\, (D_{j,} - \overline{D_{j,}})^T}{\sqrt{(D_{i,} - \overline{D_{i,}})\, SIM_O\, (D_{i,} - \overline{D_{i,}})^T}\sqrt{(D_{j,} - \overline{D_{j,}})\, SIM_O\, (D_{j,} - \overline{D_{j,}})^T}}. \quad (5)$$

This formula, as shown in Kovács (2010), has the following properties: (1) if two actors have similar values on similar dimensions, their similarity increases; (2) if two actors have dissimilar values on similar dimensions, their similarity decreases; (3) if two actors have similar values on dissimilar dimensions, their similarity decreases; and (4) if actors have dissimilar values along dissimilar dimensions, their similarity increases. Thus, this formula provides a geometric representation for similarity that incorporates the non-independence of dimensions.

Note, however, that as before for the correlation-based similarity, here as well similarity can be viewed from the other direction, resulting in Equation (6):

$$SIM_O^{GSM}(i,j) = \frac{(D_{,i} - \overline{D_{,i}})^T\, SIM_A\, (D_{,j} - \overline{D_{,j}})}{\sqrt{(D_{,i} - \overline{D_{,i}})^T\, SIM_A\, (D_{,i} - \overline{D_{,i}})}\sqrt{(D_{,j} - \overline{D_{,j}})^T\, SIM_A\, (D_{,j} - \overline{D_{,j}})}}. \quad (6)$$

As Kovács (2010) notes in his duality principle, Equations (5) and (6) have to hold simultaneously, and although he does not provide a proof for convergence, he illustrates that convergence is achieved relatively fast in the numerous simulated and empirical settings he studies. Note that simultaneously solving the two equations unifies the two directions of the correlation-based similarity.

Note that while the first approach can be viewed consistent with the attribute-based approaches (each option can be viewed as an attribute when comparing actors; and each actor can be viewed as an attribute when comparing options), the second model makes it clear that these two modes of data are not independent, and actors and options cannot be simply plugged in to attribute-based models (Kovács, 2010). Indeed, as we shall show in the next section, incorporating the interdependence between actor and option similarities we get a predictive model that outperforms the correlation-based models.

4 Solving and evaluating the predictive power of similarity-based decision models

Combining the similarity Equations (5) or (6) (or Equations (3) and (4)) with the prediction Equations (1) and (2), we can predict the utility the actor would derive from a given option. For the similarity-based decision making model to be a valid model of behavior, it needs to have a reasonable predicting power. To assess the predictive power of the similarity-based decision models, we analyze their out-of-sample prediction efficacy (Stone, 1974). That is, we split the data (which is the D matrix in our models) to a construction sample and a validation sample (Stone, 1974). In this paper, we follow an extreme sampling schedule: we go through all actor-option dyads (i.e., each cells of D), and remove these focal observations, and estimate the similarities between actors and options on this sample[3]. Based on these similarities, we predict the utility of the focal actors in the focal option (i.e., the value of the removed observation). We save the squared difference between the observed and the predicted value for all dyads[4], and calculate the average of the out-of-sample prediction error. The lower the average out-of-sample prediction error is, the better the model performs. We compare the average out-of-sample error values to the null-value of out-of-sample error (i.e., what would have been expected based on random predicted values).

To compare the predicted utilities to the underlying utilities, we need to model a data generating process. We assume that options are evaluated along a single dimension, and this dimension is bounded to the [0,1] interval ($O_i \in [0,1]$). Actors have their ideal points along this interval ($A_i \in [0,1]$) (think about, for example, a liberal-conservative policy space (Poole and Rosenthal, 1997)). We also need to choose a model that maps preferences over options to actors. In general, any negative monotonic function between the difference between the ideal points and the options' locations would satisfy the regularity condition mentioned above. Here, we assume that

$$D_{i,j} = 1 - (A_i - O_j)^2, \tag{7}$$

and investigate how the similarity-models behave given this preference function.

The general strategy to calculate the average out-of-sample prediction error of the similarity-

[3]We choose this sampling schedule because this corresponds to the decision situation the actor faces: she can look at all the previous reviews and has to predict her evaluation of the focal option.

[4]The mean out-of-sample prediction error is formally defined as: $\frac{\sum_{i=1..N}\sum_{j=1..M}(Pred(i,j)-D_{i,j})^2}{N \cdot M}$

based decision models involves (1) expressing the utility matrix D in terms of matrices A and O (through Equation (7)); (2) expressing the similarity matrices SIM_A and SIM_O in terms of A and O; (3) predicting the utility of an actor-option dyad through pairwise similarity matrices; (4) taking the integral of step (3) over all possible values of the A and O, and comparing this with the real utility to get the prediction error; and (5) taking the mean of the prediction errors over all cells of D.

Although it is mathematically straightforward to obtain a closed form solution for the mean out-of-sample prediction error, the formula is rather cumbersome even for the simplest 3×2 case[5]. Therefore, here we analyze the mean prediction error numerically. The solid line on Figure 2 shows the mean expected prediction error as a function of the number of options, keeping the number of actors constant at 3. We can see that as the number of options increase, the predictive power of the option-similarity based model increases. This result is in line with our expectations, as more options provide a basis for a more refined similarity measure.

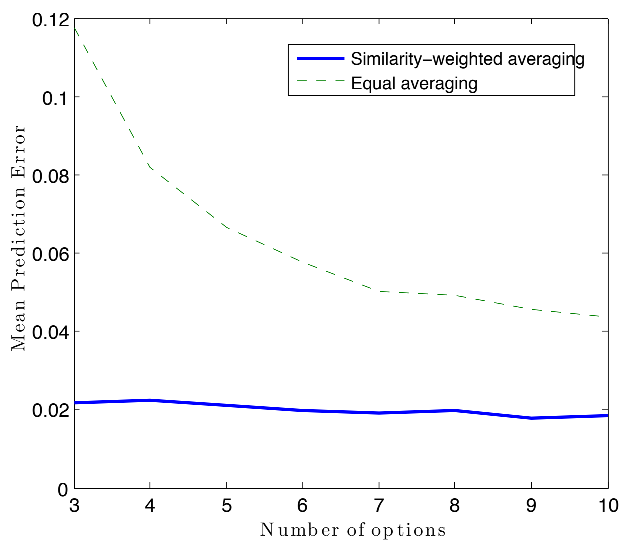

Figure 2: The mean out-of-sample prediction error of the correlation-similarity based decision model.

Next, we compare these results to two null-models. First, the accuracy of the predicted values is compared to the random prediction, i.e., in which all predicted values are picked from the random uniform distribution. The mean prediction error of the random model is 0.2217, independently of

[5]The smallest decision matrix on which the above models and sampling scheme can be executed is of size 3×2, because for a 2×2 matrix, if one observation is removed for validation purposes, then the remaining 3 cells are not enough to estimate either the similarity of actors or the similarity of options.

N and M[6]. Clearly, the similarity-based models are far better than the random baseline, by an order of magnitude. The second null-model implies a simple averaging of the utilities others derive from the focal option (without using their similarity for weighting). As the dotted line on Figure 2 shows, the simple averaging model fares much better than the simple random model, but it is worse than the similarity-based model. In conclusion, the similarity-based decision-making model improves over the random models.

Because the model is symmetric to actors and options, we get qualitatively the same results for the actor-similarity-based models (results not shown here).

4.1 Generalized similarity-based decisions

We now turn to analyze the predictive power of the generalized similarity-based model. As discussed before, we expect the generalized similarity-based models to perform better than the correlation-similarity-based model, as the generalized similarity measure is a more refined measure of similarity. As before, we analyze the effects of the main parameters (such as the number of actors, number of options, and the proportion of missing data) through numerical simulations, because no closed-form solution is known for the generalized similarity model.

Figure 3 compares the mean out-of-sample prediction error for the correlation- and the generalized similarity models, as a function of the number of actors, N and the number of options, M (N and M are set equal in this setup). The figure demonstrates that both models improve as the number of actors and options increase, and shows that the generalized similarity-based model outperforms the correlation-based model.

Next we investigate whether there is any difference in the predictive power of the correlation and the generalized similarity-based models at different levels of underlying utility, $D_{i,j}$. Before going into the analysis, note that due to the construction of the model the distribution of the utility values is skewed toward the high utility values (because both the ideal points of the actors and the locations of the option are drawn from a uniform $[0, 1]$ distribution). Therefore, there exist more observations for high utility values than for low-utility values. Figure 4 shows the mean prediction error for the correlation-based model and the generalized similarity-based model at different values of the underlying utility, $D_{i,j}$. As the figure shows, the mean prediction error

[6]For each cell, the prediction error would be $\int_0^1 \int_0^1 \int_0^1 \left(Pred - \left[1 - (A - O)^2\right]\right)^2 dA\, dO\, dPred$

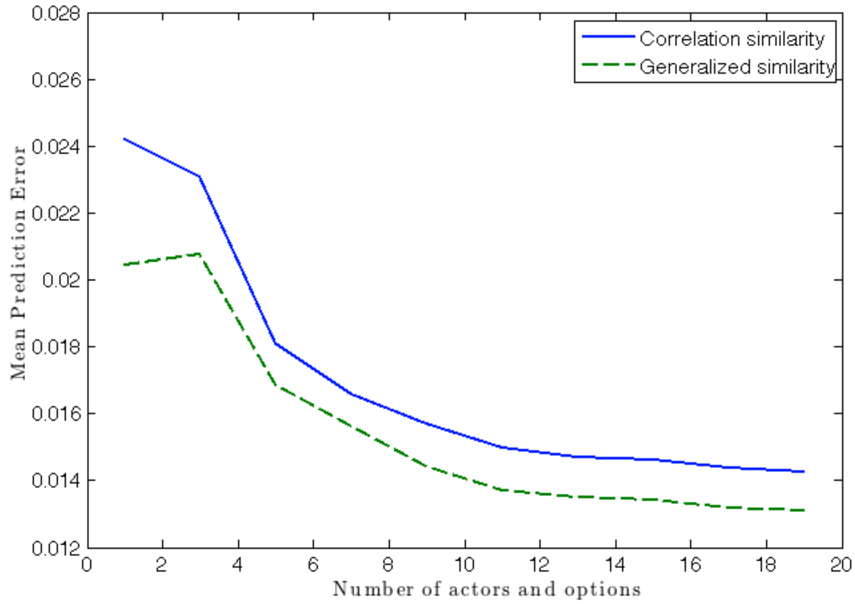

Figure 3: The mean out-of-sample prediction error of the correlation- and generalized similarity based decision models.

decreases with the utility (because there are more observations available at that level). Also, the generalized similarity-based model outperforms the correlation-based model at most levels. Note that the weighted average of these mean prediction errors equal the respective values 0.013 and 0.0141 of Figure 2.

4.2 Similarity and missing observations

We also investigate how the similarity-based models perform when certain observations are missing. By missing we do not necessarily mean that they are missing due to non-response, we just mean that for some reason some observation are not available. For example, if a given person never visited a restaurant or never rated it, then the value of that person-restaurant cell is missing. The issue of the missing value is important because most empirical datasets are sparse (just think about the dataset that contains all pairwise combinations of customers and restaurants in a major city — in such a case most observations are missing because most people have not visited most of the restaurants).

The presence of missing observations raises an issue concerning the similarity-based decision models: What should the model predict if similarity data are not available? Such is the case if

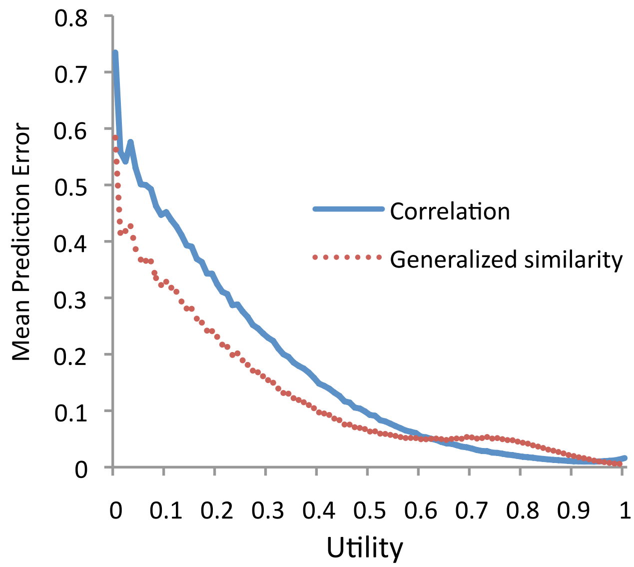

Figure 4: The mean out-of-sample prediction error of the correlation- and generalized similarity based decision models at different levels of the underlying utility, $1-(A_i-O_j)^2$, $N=M=20$.

there is no previous review on the restaurant, or if the reviewer has not reviewed any restaurants before (or, in more general, if there are not enough prior data to estimate the required similarities). In this case, our prediction models regress to the null-models. That is, if the pairwise-similarities cannot be estimated, then the model takes the average of the observed values as predictor. If no previous observation on the focal restaurant or reviewer is available, then our model regresses to the random model.

To investigate how missing data affects the efficiency of similarity-based decisions, we randomly remove a certain p proportion of the observations. We expect that the higher the proportion of missing data is, the worse the precision of the similarity based decision model will be (because there are fewer observations). Figure 5 confirms our expectations: as the proportion of missing observations increases, the mean prediction error increases. Note that the generalized similarity model outperforms the correlation-based models at every level of missing data.

4.2.1 Endogeneity bias

In the above analysis, we investigated the effects of randomly missing observations on the efficiency of the similarity-based decision models. But the missing observations are often not randomly missing. For example, in our setting, restaurant reviews, we would expect that there exist more

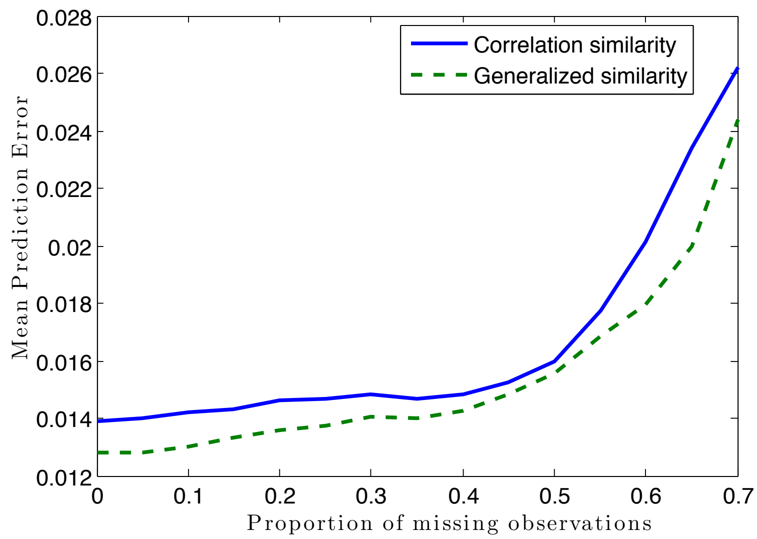

Figure 5: The mean out-of-sample prediction error of the correlation-similarity-based decision model and the generalized similarity-based model, as a function of the proportion of missing observations ($N = 20$; $M = 20$).

higher ratings than lower ratings, because reviewers are more likely to visit restaurants that they expect to like. That is, the missing observations are often a result of an endogenous process in which the reviewer predicts whether she would like the restaurant, and if she predicts that she would like it, she visits the restaurant. Indeed, this expectation is confirmed by the distribution of ratings (see Table 1). This endogeneity may introduce a bias to the efficiency of similarity-based models.

To analyze this bias, we make the probability that an observation is present the function of the predicted utility the actor would derive from the focal option. Specifically, we assume that the probability that an observation is present equals to the utility the actor predicts she would derive from the option (in the above model, regulated by Equation (7), it is ensured that the utility is bound to $[0, 1]$).

Figure 6 shows the relative mean prediction errors of the non-randomly missing (as described in the previous paragraph) and the randomly missing observations (MPE(non-randomly missing observations)/MPE(randomly missing observations)[7]. As the figure shows, if the abovementioned bias is present, the similarity-based models underperform in the low-utility regions, and overper-

[7]To keep the overall level of missing observations the same as in the non-randomly missing case, we randomly removed 16.7% of the observations.

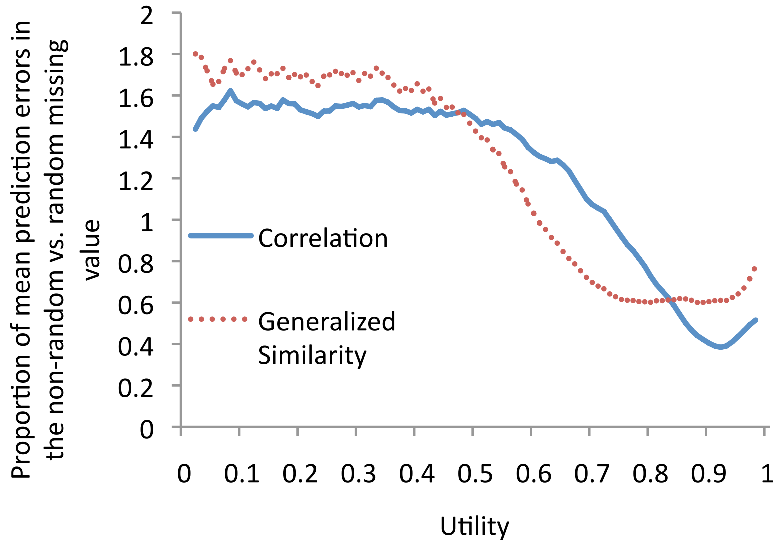

Figure 6: The proportion of the mean prediction errors of the non-randomly missing and the randomly missing observation setups, by utility level ($N = 20$; $M = 20$).

form in the high-utility regions (as opposed to when the bias is not present). Also note that the generalized similarity based model outperforms the correlation model at moderate utility levels (where the majority of the observations are).

4.3 Alternative preference functions

So far we investigated the predictive power of the similarity-based decision models under the assumption that the preference function is a quadratic function of the match between the ideal points of the actors and the location of the options. Here we relax this assumption, and investigate the behavior of the similarity-based models under a more general family of preference functions. First, we assume that the preference function is of the following form:

$$D_{i,j} = 1 - (A_i - O_j)^{\alpha}, \tag{8}$$

where α denotes the weight of the match between actor i's ideal point and the location of option j.

Figure 7 shows how the similarity-based models perform under various values of α. First, note that α has a non-monotonic effect on the mean prediction error. The low values of the mean prediction error at low α are due to the fact that when α is very low, the actors are not sensitive to the differences between their ideal points and the value of the options, thus there will be low variance in the values of D, resulting in good predicting power (as all values of D will be close to 0). Second, note that for most values of α the generalized similarity model outperforms the correlation-based model.

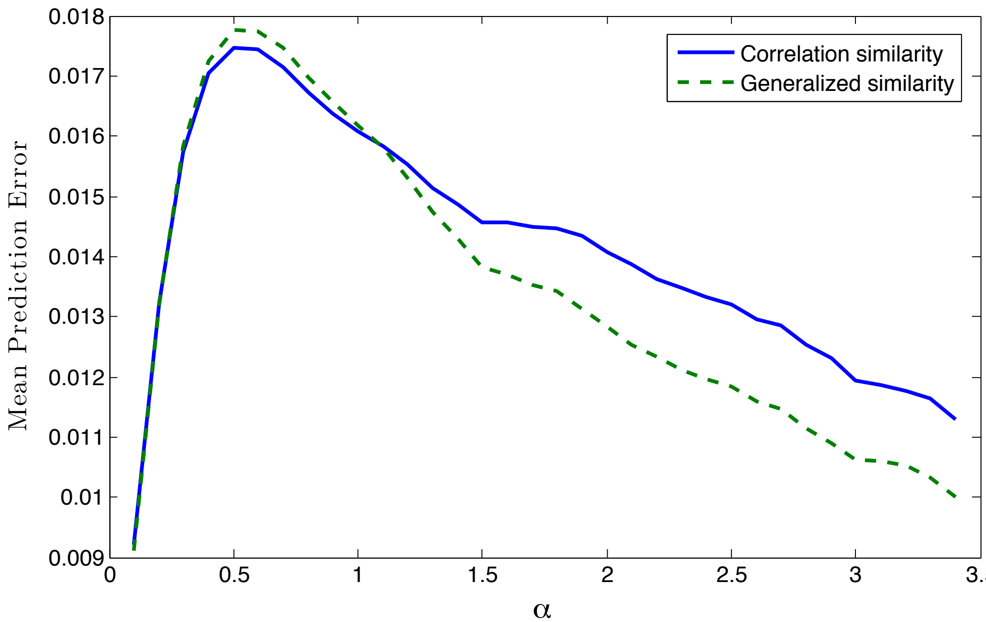

Figure 7: The mean out-of-sample prediction error of the correlation-similarity-based decision model and the generalized similarity-based models, as a function of the alpha parameter ($n = 20$; $m = 20$).

We also analyzed how the predictive power of the similarity-based models changes if the space is not unidimensional. We assume that instead of one there are two dimensions (still bounded to 0 and 1) along which the options and the ideal points are located. Then, we operationalize $D_{i,j}$.as $D_{i,j} = 1 - \sqrt{(A_{i,1} - O_{j,1})^2 + (A_{i,2} - O_{j,2})^2}$, and run the same analyses. The results are similar to the results before, except the main level of the mean of the predicter error is much lower, on average it is the 20% of the values with one dimension. This is the case because the observed utilities better reflect match between actors' ideal points and the locations of options.

Table 1: The distribution of restaurant ratings.

Star rating	Count	Percent
1	20,345	5.7%
2	33,128	9.2%
3	75,186	20.9%
4	139,855	39.0%
5	90,666	25.2%

5 Empirical illustration — Similarity and restaurant choices

To illustrate the proposed similarity models, we analyze consumers' reviews of restaurants, as posted on a review website. Our sample consists of 357,537 reviews written on 2,906 the San Francisco-based restaurants by 50,082 reviewers. The observation period runs from October 2004 through February 2009. Each reviewer can give one to five stars to the restaurants. Table 1 shows the distribution of ratings. Figure 8 shows that the distribution of the natural logarithm of the total number of reviews per reviewer is very skewed: ten reviewers have posted more than 300 reviews, 290 reviewers have posted more than 100 reviews, but most of the reviewers wrote only a few reviews. The mean number of reviews per person is 7.17.

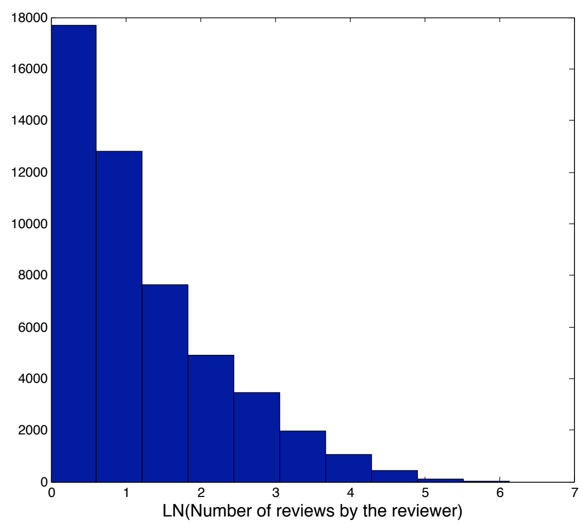

Figure 8: The distribution of the count of reviews per reviewer.

Our aim here is to illustrate the workings of the similarity-based decision models on the restaurant-reviews data. As before, we want to test the validity of the model by showing that the similarity-based decision models improve the prediction of reviewer's liking of the restaurant above the random model and the simple averaging. The data are very sparse, only 0.24% of the possible reviews are written (353,537/(50,082*2,906)), thus we expect the generalized similarity-based model to outperform the correlation-based similarity model.

To assess the performance of the similarity-based models, first we analyze the data as a cross-sectional data, that is, we aggregate all observations. We follow the sampling scheme we used in the modeling section: remove one observation, predict the similarities based on the rest of the data, predict the focal observation based on the similarities, and compare the predicted value to the observed value. Ideally, we would like to repeat the above estimation and comparison for all observations in the dataset. However, as the number of observations is large, and computing the similarities for all of them is computationally very intensive[8], we conduct the comparison only on a random sample of the observations. Specifically, we randomly choose 800 reviewers and 800 restaurants. This sample contains 11,416 reviews (thus, the resulting D matrix is sparse, with 98.21% missing values). We calculate a number of statistics for this sample, and investigate how the similarity-based models predict the ratings.

First, we analyze the reviewer-similarity-based models (based on Equation (1)). The idea here is that the rating can be predicted based on the similarity across reviewers: the reviewer will rate the restaurant similarly as to how similar reviewers have rated it. The respective baseline mean prediction errors are as follows. If we randomly fill out the similarity-matrices, then the resulting prediction's mean prediction error is 4.55. If we take as prediction the mean value of ratings, which is 3.6852, then the resulting mean prediction error is 1.0608. If we take as prediction the mean value of the ratings the reviewers gave to this restaurant, the mean prediction error is 1.0782. The mean prediction error for the correlation model is 0.9523, which improves over the simple averaging model. The generalized similarity model performs the best: its mean prediction error is 0.8902.

Second, we analyze the restaurant-similarity-based models (based on Equation (2)). Again, the idea here is that we can predict the rating the reviewer gives to the focal restaurant based on how she has rated similar restaurants. The baseline mean prediction errors are as follows. If we randomly fill out the similarity-matrices, then the resulting prediction's mean prediction error is 4.55. If we take for prediction the mean value of ratings, which is 3.6852, then the resulting mean prediction error is 1.0608. If we take as prediction the mean value of the ratings the reviewer

[8]The computation speed increases quadratically with N and M. For example, to compute the pairwise similarities for all 50,082 reviewers, we would need to store a $50,082 \times 50,082$ matrix, which would contain about 2.5 billion cells. This itself would be feasible, but repeating this in a large number of times (for all 357,537 reviews) is not.

gave to other restaurants, the mean prediction error is 0.8096. The mean prediction error for the correlation model is 0.875, which is worse than the simple averaging model. The generalized similarity model, however, improves over the simple-averaging model: its mean prediction error is 0.7881.

In summary, on a cross-sectional sample of reviews, we have demonstrated that both similarity-based models outperform the random prediction model and the simple averaging model. We have also seen that the generalized similarity model outperforms the correlation-similarity-based model. We found that in this dataset restaurant-similarity is a better predictor than reviewer-similarity, because due to the structure of the data, from a randomly chosen subset of reviewers and restaurants it is easier to calculate restaurant-similarity than reviewer-similarity.

5.1 Learning about similarities: Similarity and prediction over time

[THIS SECTION IS NOT FINISHED]

In the previoussection, we compared the similarity-based models on a cross-section of reviews. If we want to use the similarity-models as models for human decision making and learning, we need to take the temporal nature of the data into account, as the reviewer can only use review data that is available at the time of the decision situation (Gonzalez et al., 2003; Gayer et al., 2007; Gilboa et al., 2009). Thus, we time-order the reviews (the unit of time is days), and for each review we estimate the similarity matrices SIM_A and SIM_O on the review-data prior to the review. We compare each actual review with the predicted values based on the prior similarity matrices according to the six prediction rules (random; simple averaging; restaurant-correlation-based; reviewer-correlation-based; restaurant general similarity-based; and reviewer general similarity-based). Day 1 in our dataset is October 20th 2004, when the first review was posted. Figure 9 shows the count of reviews in a monthly aggregation.

We expect that the predictive power of the similarity-based decision models increases over time. The reason for this expectation is that as time passes, more and more reviews are added to the dataset, thus the similarity-based models have an increasing number of instances on which to estimate the similarity of restaurants and the similarity of reviews. This expectiation is consistent with previous research on similarity-based decision making (Ram, 1993; Gonzalez et al., 2003). For example, Gonzalez et al. (2003) demonstrates how, as time passes in the experiment and subjects accumulate more instances to learn from, the subject shift from heuristics-based decision making to similarity-based decision making. In our case, because no attributes are available, we expect our learning model to shift from the random baseline model to similarity-based decision.

Figure 9 shows time-series data on the review dataset. Figure 9a shows the number of reviews in each month: the number of reviews per month is increasing over time. Figure 9b shows the number of new reviewers in each months (i.e., number of people who have their first review in that month), which also shows an increasing trend, reflecting the increasing popularity of the review site. The number of restaurants entering the dataset shows a different trajectory: after a warm-up period, most restaurants enter the database between January 2005 and 2006. These temporal patterns are important because at each point in time the size of the learning dataset equals the number of reviewers multiplied by the number of restaurants. To get the proportion of missing data, we have to divide the number of reviews with this nominator (see Figure 11).

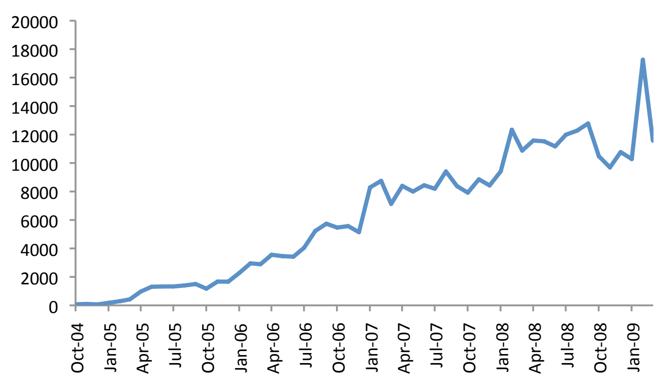

Figure 9: Number of reviews per month.

Figure X shows how the mean prediction error changes over time for the similarity-based models and the baseline models (the random model and the simple aggregation model). Because showing daily data results in too much noise, we aggregate the mean prediction errors by weeks.

6 Discussion

In this paper, we have introduced similarity-based decision models for decision situations in which data on the attributes of options are not available. Building on earlier research in psychology, machine learning and decision theory, we introduced two models that assess the similarity of

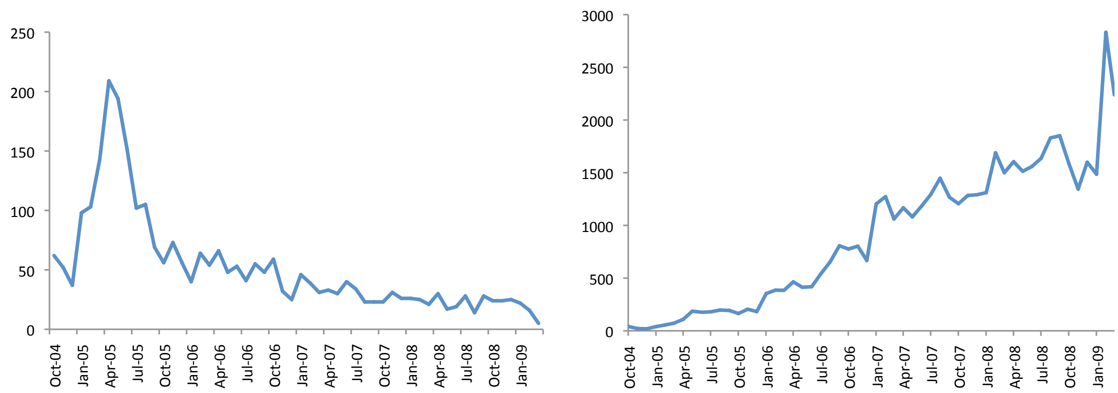

Figure 10: Number of new restaurants reviewed (a); and number of new reviewers (b) per month.

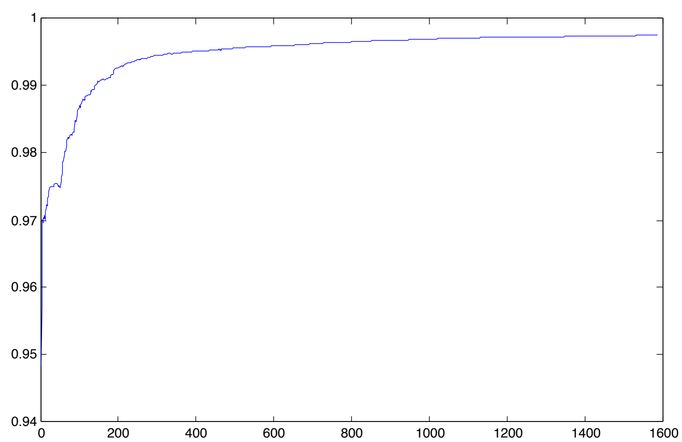

Figure 11: Proportion of empty reviewer-restaurant cells over time.

actors and options based on how similar actors have evaluated the focal option, and how the focal actor has evaluated similar options.

A particular novelty of this paper is the measure of similarity applied. The models proposed in this paper, instead of relying on attribute-based similarity like previous studies of similarity-based decision theories (Gilboa and Schmeidler, 2001), assess similarity of the actors from previous behavior of actors toward other options. The introduction of the second mode, options, also gives rise to the possibility of assessing the similarity of options, and explore a second angle of similarity-based decision making: actors can predict the utility they would derive from the focal option based on the utility they have derived from similar options. While both of these directions provide similarity-based predictions, the duality of the models calls for a unification, and this is what we provide in this paper with applying the generalized model of similarity of Kovács (2010). Building on the principles: "Actors are similar if they derive similar utilities from similar options," and "options are similar if similar actors derive similar utilities from them," we provide a generalized similarity-based decision model. This model, as we demonstrate, predicts well the behavior in actor in both simulated situations, and in an empirical illustration of restaurant reviewing.

This paper is novel in first applying longitudinal data to similarity-based decision. While Gilboa et al. (2009) emphasizes the temporal nature of similarity-based decision making, to our knowledge this article is the first taking the temporal nature of the data seriously. We find evidence for learning: the similarity-based decision making models increases its efficiency as more data become available over time. Early in the learning process, there are not much data on which similarities can be estimated, thus the similarity-based models provide similar results to the random and the simple averaging baseline models. As more data become available, the similarity-based models improve over the baseline model. The generalized similarity-based models outperform the correlation-similarity-based models through the whole period, but especially in the early phases of the process.

During the article, we avoided addressing the issue whether the similarity-based decision making theory is a normative or a descriptive theory. We, together with Gilboa and Schmeidler (2001), believe that the line between normative and descriptive theories is not always clear, and this is the case for the similarity-based decision making as well. We believe that the similarity-based decision theory can be used both descriptively for prediction (this is what we followed in this paper), but also normatively: actors, to maximize their welfare, could follow a similarity-based decision model (although we would be careful with this latter statement, as there have not been many investigations about whether the proposed models maximize utility. To do this, we would need to investigate what the optimal similarity function is).

Of course, the paper is not lack of limitations. First, we assumed that everybody knows all information. This is a strong assumption (although it applies to our empirical setting). We could introduce local information, or maybe a social network for observation (Jackson, 2008). Second, one might argue that utility is usually not observable, thus our model has limited applicability. We argue, however, that with the digital revolution there are increasing number of settings in which utilities are observable, such as online ratings on books, restaurants, music, movies, etc.

A further way to significantly broaden the applicability of similarity-based decision models would be to connect the preference of actors to their behavior through revealed preference (Varian, 2006). In situations where action (which is more often observable than preference) can be substituted for preferences, one could apply the similarity-based models by saying "Actors that behave similarly in similar situations are similar," and "situations in which similar actors behave similarly are similar." The exploration of this link, however, we leave for further research.

Nevertheless the above limitations and yet unexplored aspects of the proposed model, we believe that the similarity-based decision making model have important implications for and applications in numerous disciplines, including psychology, decision making, marketing, and machine learning.

References

Akaike, Hirotugu. 1954. An approximation to the density function. *Annals of the Institute of Statistical Mathematics* 127–132.

Batagelj, Vladimir, Matevz Bren. 1995. Comparing resemblance measures. *Journal of Classification* **12**(1) 73–90.

Billot, Antoine, Itzhak Gilboa, Dov Samet, David Schmeidler. 2005. Probabilities as similarity-weighted frequencies. *Econometrica* **73** 1125–1136.

Gayer, Gabrielle, Itzhak Gilboa, Offer Lieberman. 2007. Rule-Based and Case-Based Reasoning in Housing Prices. *The B. E. Journal of Theoretical Economics* **7**(1).

Gilboa, Itzhak, Offer Lieberman, David Schmeidler. 2006. Empirical similarity. *The Review of Economics and Statistics* **88**(3) 433–444.

Gilboa, Itzhak, Offer Lieberman, David Schmeidler. 2009. A similarity-based approach to prediction. *Journal of Econometrics* .

Gilboa, Itzhak, David Schmeidler. 1995. Case-based decision theory. *The Quarterly Journal of Economics* **110** 605–639.

Gilboa, Itzhak, David Schmeidler. 2001. *A theory of case-based decisions*. Cambridge Univ Press.

Gilboa, Itzhak, David Schmeidler. 2003. Inductive inference: An axiomatic approach. *Econometrica* **71**(1) 1–26.

Gonzalez, Cleotilde, Javier F. Lerch, Christian Lebiere. 2003. Instance-based learning in dynamic decision making. *Cognitive Science* **27** 591–635.

Hume, David. 2004 (1748). *An Enquiry Concerning Human Understanding*. Mineola, New York: Dover.

Jackson, Matthew O. 2008. *Social and economic networks*. Princeton University Press.

Kovács, Balázs. 2010. A generalized model of relational similarity. *Social Networks* **32** 197–211.

Lancaster, J.S., J.L. Kolodner. 1987. Problem solving in a natural task as a function of experience. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum, 727–736.

Landauer, Thomas, Susan Dumais. 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* **104** 211–40.

Leake, David B. 1996. *Case-based reasoning: Experiences, lessons and future directions*. MIT Press Cambridge, MA, USA.

Leake, David B., Andrew Kinley, David Wilson. 1997. Case-based similarity assessment: Estimating adaptability from experience. *Proceedings of the national conference on artificial intelligence*. 674–679.

Locke, John. 1690. *An essay concerning humane understanding*. Baffet.

Mahalanobis, Prasanta C. 1936. On the generalised distance in statistics. *Proceedings of the National Institute of Science of India* **12** 49–55.

McGinty, L., D.C. Wilson. 2009. Case-Based Reasoning Research and Development. *Lecture Notes in Computer Science* **5650**.

Medin, Douglas L., R. L. Goldstone, A. B. Markman. 1995. Comparison and choice: Relations between similarity processing and decision processing. *Psychonomic Bulletin and Review* **2** 1–19.

Medin, Douglas L., Robert L. Goldstone, Dedre Gentner. 1993. Respects for similarity. *Psychological Review* **100** 254–278.

Poole, Keith T., Howard Rosenthal. 1997. *Congress: A political-economic history of roll call voting*. Oxford University Press, USA.

Ram, A. 1993. Indexing, elaboration, and refinement: Incremental learning of explanatory cases. *Machine Learning* **10** 201–248.

Read, Stephen J., I.L. Cesa. 1991. This reminds me of the time when...: Expectation failures in reminding and explanation. *Journal of Experimental Social Psychology* **27**(1) 1–25.

Riesbeck, Christopher K., Robert C. Schank. 1989. *Inside Case-based Reasoning*. Hillsdale, NJ: Lawrence Erlbaum.

Ross, Brian H. 1984. Remindings and their effects in learning a cognitive skill. *Cognitive Psychology* **16**(3) 371–416.

Shepard, Roger N. 1987. Toward a universal law of generalization for psychological science. *Science* **237** 1317–23.

Silverman, Bernard W. 1986. *Density Estimation for Statistics and Data Analysis*. London and New York: Chapman and Hall.

Simon, Herbert A., F. Gobet. 1996. Templates in chess memory: A mechanism for recalling several boards. *Cognitive Science* **31** 1–40.

Stone, Mervyn. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B* **36** 111–147.

Tversky, Amos. 1977. Features of similarity. *Psychological Review* **84** 327–352.

Varian, Hal R. 2006. Samuelsonian economics and the twenty-first century. chap. Revealed Preference. Oxford University Press, 99–115.

Weinreb, Lloyd L. 2005. *Legal reason: The use of analogy in legal argument*. Cambridge Univ Pr.

Widdows, Dominic. 2004. *Geometry and Meaning*. Center for the Study of Language and Information/SRI.