Learning What Others Like:
Preference Learning as a Mixed Multinomial Logit Model

Natalia Vélez (nvelez@stanford.edu)
450 Serra Mall, Building 01-420
Stanford, CA 94305 USA

Abstract

People flexibly draw generalizations from others' choices. For example, when one learns that a friend likes a set of movies within a particular genre, one might also infer that said friend would also like novel movies with similar features. The present project uses a Church implementation of mixed multinomial logit (MML) models to capture how people learn and generalize from other people's preferences. Experiment 1 demonstrates that the model can accurately recover a dummy target's preferences from a sequence of observations. Experiment 2 compares model performance to trial-by-trial human performance in a game where participants had to learn someone else's preferences from a sequence of pairwise choices. Here, the model outperformed humans. Finally, we will discuss how to refine the model by including an extra inferential step: preference learning might not only involve learning someone's preferences along a certain set of dimensions, but also discovering the dimensions themselves.

Keywords: preferences; social cognition; learning

Introduction

In everyday social interactions, we are not only observing others' actions, but also finding useful patterns in their behavior. The ability to see these patterns is particularly important when we are choosing on others' behalf: that is, when we are using their past actions to decide how to help, what to teach, or what to give in the future.

For example, consider someone who is buying a Christmas present from a friend. She might have some information about what her friend likes from the choices he has made before—such as what books he reads, what he wears, or what video games he plays. But, in most cases, it would be unusual for her to buy something that her friend already owns and already likes. Instead, she has to use what she knows about her friend to predict what he will like best among a set of novel objects at the store.

MML models have been used extensively in economics to formalize and describe similar decisions, in which one agent is making discrete choices among a set of options (McFadden & Train, 2000). To unpack the problem confronted by the holiday shopper, we assume that her friend's preferences are: (1) stable over time, (2) defined over properties of objects (and can therefore generalize to new objects with similar properties), and (3) stronger for some features than for others.

MML models assert that, when people are deciding among a set of options, their choice is probabilistically related to the subjective utility, or attractiveness, of each option. Specifically, the model is exponentially more likely to select an option as its utility increases (Luce, 1977). Each option i is represented as a binary vector of its features (x_i) and a vector of weights (β_a) corresponding to the agent's preferences for individual features. The utility of option i (u_i) is the weighted sum of its features ($\beta_a x_i$), scaled by a free "temperature" parameter T. Taken together, the probability of an individual choosing option i from a suite of options J is defined as:

$$P(c = i|\boldsymbol{X}, \boldsymbol{\beta_a}) = \frac{\exp(\beta_a \boldsymbol{x}_i/T)}{\sum_j \exp(\beta_a \boldsymbol{x}_j/T)}$$

In cognitive science, MML models have been used to capture how children use other people's preferences to guide their own choices (Lucas et al., 2014). Other people's preferences are a powerful source of information for children, who are themselves navigating the world trying to decide what snacks to eat, which toys to play with, and whom to befriend. Two-year-olds use information about the overlap between their preferences and other people's preferences to guide their own decisions: when given the choice between two mystery toys endorsed by different people, toddlers tend to choose the toy that was recommended by someone more similar to them (Fawcett & Markson, 2010). Lucas et al. (2014) framed the child's choice as two inferences: first, an inference about each actor's preferences and, second, an inference about the features of the mystery toys. Given some information about the child's preferences, these two inferences provided a prediction about the child's choice. The predictions of the MML model closely matched children's actual choices—indeed, they predicted children's responses in several important studies, including Repacholi & Gopnik (1997) and Kushnir, Xu & Wellman (2010).

The experiments described below explore how people learn others' preferences by observing their choices. By creating a MML cognitive model, we reduced the problem of learning someone else's preferences, essentially, to inferring the weights for each feature. Experiment 1 used simulated data to test how well the model recovered preference weights from a sequence of choices, as well as how different settings of the temperature parameter affected model behavior. In Experiment 2, participants predicted someone else's choices among pairs of stimuli and received feedback on the other person's preferences. Participants' trial-by-trial accuracy was compared to model performance. Finally, we will discuss future refinements to the model.

Stimuli

In the experiments described below, human and model observers learned a target's movie preferences. However, using real movies for such a task poses two problems: first, participant might be more familiar with some movies than others; second, there is a large space of features, such as genre, mood, casting, director, budget, critical acclaim, etc., that might influence people's actual movie preferences. Thus, to keep the features dictating movie preferences fairly simple, we created a set of fake movies (Figure 1) that vary along three dimensions: valence (positive or negative), setting (historical or futuristic), and genre (action or romance). Each movie was written be positive *or* negative, historical *or* futuristic, and action-packed *or* romantic, but not both. These stimuli impose structure on movie preferences, and they mitigate the influence of extraneous factors, such as critical acclaim or prior knowledge. All movies were normed on Amazon Mechanical Turk (n = 90, data not shown) to ensure that they were indeed novel and that they adhered to the intended categories.

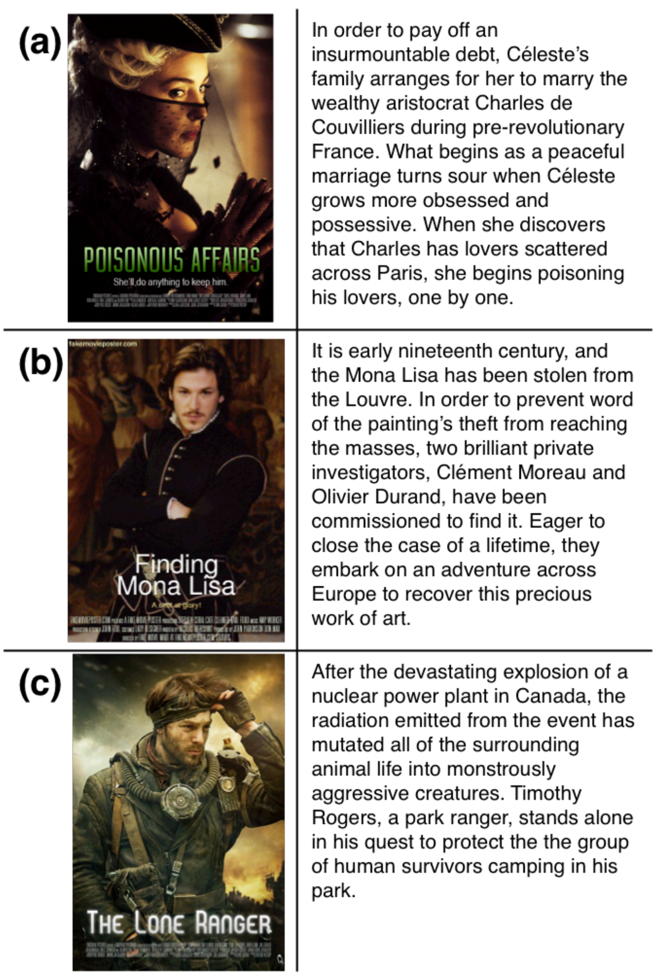

Figure 1: Example movie posters and synopses, representing the (a) negative historical romance, (b) positive historical action, and (c) negative futuristic action categories, respectively.

Experiment 1

Experiment 1 was intended as the simplest test of the model. First, using responses simulated from a dummy "target" with predefined weights, we tested whether the Church implementation of MML models could recover preference weights from a sequence of choices. Second, we tested the model under different settings of the temperature parameter to observe how this free parameter might affect the model's behavior.

Methods[1]

Responses were simulated using a softmax decision rule with pre-defined "target" weights over all possible combinations of conditions. Each feature was coded in binary vectors of -1 or 1; for example, positive films had a value of 1 on the first item of the feature vector, while negative films had a value of -1 on the first item. Similarly, weights ranged from -1 to 1. Therefore, if a target had a strong preference for negative films over positive films, the first item of their preference vector would be -1. This coding scheme exploits the binary nature of the stimuli to simplify the estimation process: the model estimated three, rather than six, weights. The precise sign and ordering of the features was chosen arbitrarily, but was used consistently in both Experiments 1 and 2.

For Experiment 1, the "target" preference weights were (.5, 1, 0). Under the coding scheme described above, these weights corresponded to: moderate preference for positive over negative films, strong preference for futuristic over historical films, and indifference between romance and action films, respectively. These were selected to obtain the most information about the model's ability to recover weights of different magnitudes.

To recover the weights, the model sampled preference weights according to the probability of the observed sequence of responses given the sampled weights. The mean of each weight was taken as the model's "best" estimate of the weight. This process was repeated for settings of the temperature parameter ranging from 0.1 to 10.

Results & Discussion

Figures 2 shows the posterior distribution of the weights for different settings of the temperature parameter, *T*. As the temperature parameter increases, the distributions become "flatter": that is, the distributions cover a wider range of possible weights and tend to center around 0.

Figure 3 compares the model's best estimate of the weights (i.e., the mean sampled value) for every setting of the parameter that was tested in Experiment 1. The model's best estimate of the weights shifts with changes in the temperature parameter: in general, the magnitude of the estimated weights tended to decrease as the temperature parameter increased.

[1] The Church code used to generate this model is documented on GitHub.

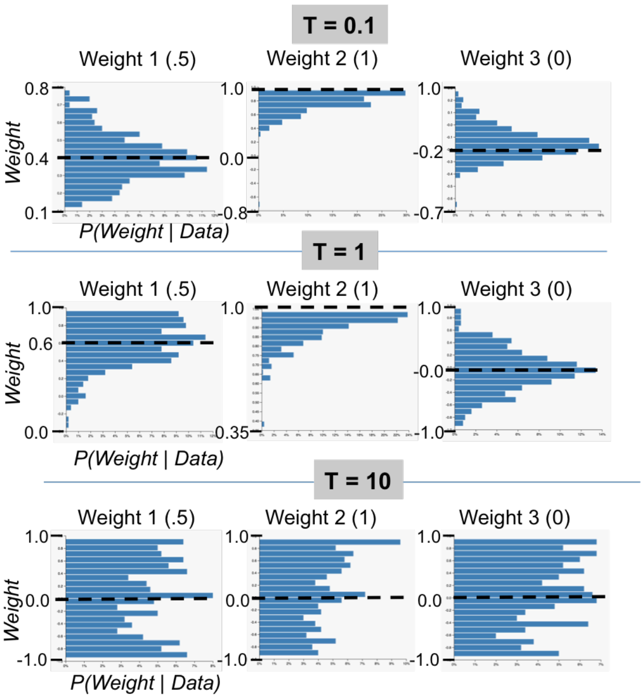

Figure 2: Posterior distribution of weights for different settings of T. Dashed lines mark the mean estimated weight. The correct weights are (.5, 1, 0).

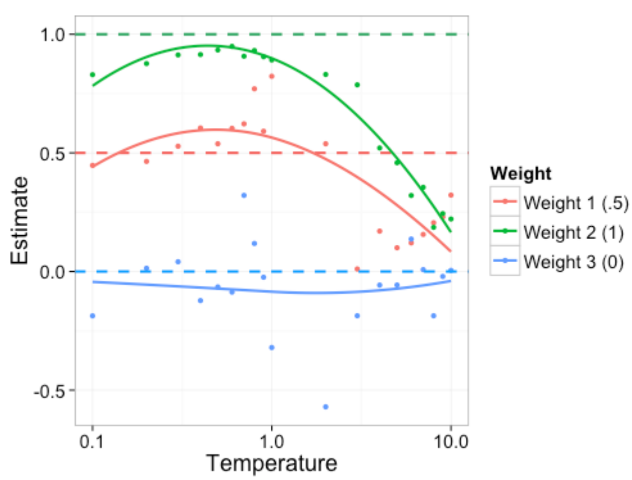

Figure 3: Weights recovered by the model from a sequence of observed choices, given different settings of the temperature parameter.

Thus, the temperature parameter acts as a 'gain' on the utility. As the temperature parameter grows smaller (i.e., the gain increases), the softmax decision rule becomes 'greedier'; that is, it is more likely to select the option with the highest utility. Conversely, as the temperature parameter grows smaller, all options become equally likely (Sutton & Barto, 1998). Consistent with this intuition, as the temperature grows larger, the model's inferences of the weights tend to 0: that is, the model infers that the target is indifferent about all three features.

However, it is somewhat inevitable that the model would perform well on this test, given that the sample weights and observed responses were generated using the same random process. A much more powerful test of the model would be to train the model on a set of human-generated responses among one set of movies, and then to cross-validate the learned weights on the same participant's responses among a different set of movies. Such a task is currently being implemented and will be released after changes have been made to the procedure described in Experiment 2 (see below, and the Discussion for an overview of these changes). Unfortunately, the results of this experiment will not be available within the timeframe of the course.[2]

For the time being, in order to avoid the computational costs of modeling weights over different sequences of responses *and* different settings of the temperature parameter, the model described in Experiment 2 used a fixed setting of $T = 0.5$—i.e., one of the "greediest" settings of the temperature parameter. Ideally, future, optimized versions of this model would employ methods from Bayesian data analysis to set the temperature parameter.

Experiment 2

In Experiment 2, participants were shown pairs of movies and asked to predict which movie *someone else* would choose.[3] The MML model was trained using subsets of n trials to compare its predictions to human performance on the (n+1)th trial.

Procedure

US-based participants (n = 40; mean(SD) age: 33(10); 25 male) were recruited from Amazon Mechanical Turk. All participants gave informed consent to participate in the study. However, there was one bit of deception: participants were told that someone else had already taken the HIT and had chosen which movies they would rather watch. In reality, responses were simulated using a similar procedure as that in Experiment 1. All 5 possible sets of target weights were on either extreme of the scale or at indifference, such as (-1, 0, 1). Participants were randomly assigned to each target.

Each HIT was split into 2 phases, during which participants were exposed to 16 of the 32 movies. Movies were pseudo-randomly assigned to each phase to ensure that each phase contained 2 movies in each of the 8 conditions. Each phase was composed of 2 tasks, which were titled "Meet the Movies" and "Choose for Someone Else."

In "Meet the Movies," participants viewed the poster and read the synopsis of each of the 16 movies, then answered a multiple-choice question about the plot. Participants were

[2] However, a prototype of this task can be viewed here.
[3] The task can be viewed on this page, or examined in more detail in this GitHub repository. At the date of submission, all contributions to the linked directory are my own.

not allowed to advance to the next trial unless they had (a) spent at least 5 seconds on the page and (b) answered the question correctly.

In "Choose for Someone Else," participants were shown pairs of movies and had to predict which movie the other person would rather watch. This task consisted of 56 trials, during which participants saw each possible pairing of conditions twice. Movies were pseudo-randomly assigned to each trial, to ensure that each movie was shown exactly seven times. In phase 1, participants received feedback after each trial about their prediction; in phase 2, they received no feedback.

Modeling responses

The model was trained on the simulated responses of a target with preference weights (-1, 1, 0); both the model and a subset of human participants were shown the same sequence of responses generated by this target. In order to simulate human trial-by-trial performance, the model estimated weights after observing n trials and used these weights to predict the target's response on the $(n+1)$th trial. Therefore, "model performance" (below) refers to the proportion of predicted responses on the $(n+1)$th trial that matched the target's response.[4] Because the model ran through two queries—one for the preference weights, and another for the predicted response—it was too costly to compare model and human performance on every trial. Therefore, we used only a subset of the trials, where $n = \{1, 7, 13, 27, 41, 55\}$.

Results: Human performance

Figure 4 shows human accuracy in "Choose For Someone Else," both in phase 1, where participants received feedback on their choices, and in phase 2, where participants received no feedback. Overall, participants' performance seems to improve steadily over the course of phase 1, suggesting that they are indeed learning the target's preferences. However, in phase 2, participants were consistently at chance (more on this result in the Discussion). Thus, while the original plan was to cross-validate responses from phases 1 and 2, we instead compared trial-by-trial performance in phase 1 to model predictions.

Results: Model performance

Figure 5 compares participants' overall accuracy in phase 1 of "Choose for Someone Else" with model predictions. To reiterate, model responses were generated by inferring the weights after observing $(n+1)$ trials, then predicting the target's response on the nth trial.

It is immediately apparent that the model far outperforms human observers: while participants' performance peaked at 70% towards the end of the 56 trials, the model reached 90% accuracy within the first 14 trials. This result suggests that participants have to make additional inferences—about each object's features and about the relevant dimensions—that the model did not account for, which we will explore further below.

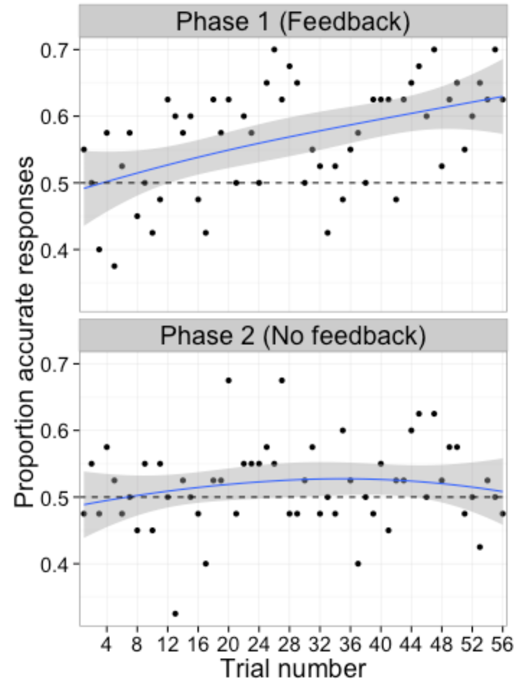

Figure 4: Overall accuracy for each trial in "Choose For Someone Else," represented as the proportion of participants (n = 40) that responded correctly on each trial.

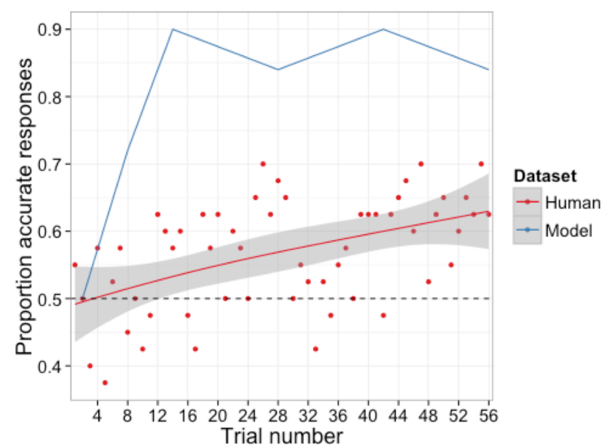

Figure 5: Comparison of model performance and human performance on phase 1 of "Choose for Someone Else."

Discussion

This project consisted of two experiments, both of which explored the potential of MML models to capture people's inferences about other people's choices. Experiment 1 served as a way to explore implementations of MML models in Church and to calibrate different settings of the model's temperature parameter. We found that the model accurately recovered a dummy target's preferences from a sequence of observed choices. Furthermore, the behavior of the model changed with different settings of the temperature

[4] The full documentation of this version of the model is also on GitHub.

parameter, which served as a "gain" on the subjective utility of an option: as the value of the temperature parameter decreased, the model became "greedier," choosing the option with the highest utility with higher probability.

In Experiment 2, participants were shown pairs of movies and asked to predict what someone else would rather watch. In phase 1, participants were given feedback on their predictions after each trial; in phase 2, they were exposed to a new set of movies and received no feedback. Participants' trial-by-trial performance was compared to model performance. The experiment yielded two unexpected results: first, that participants performed at chance in the second phase of the experiment; and, second, that the model vastly outperformed human observers in the first phase. We will explore each of these results in turn.

First, it is puzzling that participants performed at chance in the second phase of the experiment. Even though they seem to be learning the other person's preferences in phase 1, they do not generalize from those preferences to choose accurately on the other person's behalf in phase 2. The most worrying interpretation of this result is that the task is too hard; that is, that the dimensions along which we manipulated our stimuli are not meaningful, or too subtle to be detected by a naïve participant.

Before seriously considering this deeper problem, there are many other superficial factors that could have instead driven our results. Namely, participants may have: (1) not paid attention to the movies during "Meet the Movies"; (2) not been sufficiently motivated, after a long task, to carefully consider the other person's preferences in phase 2; or (3) not understood that the person they were choosing for in phase 2 was in fact the same person they had just seen in phase 1. These will be addressed in future iterations of this task through clearer instructions, attention checks interspersed throughout the tasks, and bonuses for good performance in phase 2.

Second, it is particularly interesting that our model performed *too* well compared to human participants; it suggests that we underestimated how complex inferring someone else's preferences can be to a naïve observer. In particular, our model had at least three significant, and perhaps unfair, advantages over the human observer: it (1) knew exactly what three dimensions we were manipulating, (2) knew the exact features of each movie, and (3) remembered every trial perfectly. The latter two might be incorporated in our model by simply forgetting trials and adding uncertainty about the features of the items.

The former is an interesting problem in itself, even within the highly constrained space of our movies. As an immediate next step, it may be possible to circumvent this problem by simply *telling* participants what the dimensions are. This way, participants are still faced with the interesting, and more tractable, problems of inferring each item's features and inferring the other person's preferences for each feature. However, discovering the feature space of preferences merits future exploration in its own right. One way to extend our current model is to incorporate an additional step before the model infers the weights. Namtely, the model would first infer the dimensions along which these movies vary; this could be implemented using dimensionality reduction to split the movies into dimensions or a Chinese restaurant process to cluster the movies into categories. Next, given the most likely dimensions, the model would infer the target's weights for each dimension. Both of these approaches could be compared to a simpler model that creates a ranked order of the movies, regardless of their features, as the inferences generated by this model should not generalize to set of novel options.

The current project is the first step in a larger, longer effort to understand how people learn other people's preferences, as well as how people generalize from those preferences to choose on others' behalf. The key goals of this project were to implement MML models within a cognitive model in Church, and to compare model predictions to human performance in a preference-learning task. In these respects, the project was successful. Where the project fell short, it provided useful clues into how our approach could be adapted, expanded, and refined in the future. In particular, it seems we may have underestimated just how complex—and fascinating—learning another person's preferences could be.

Acknowledgments

This project is part of an ongoing collaboration with Yuan Chang Leong (PhD student, psychology) and Chelsey Pan (undergraduate research assistant). Both have played a substantial role in developing the stimuli and have contributed their feedback on the experiments described here. The procedure, analysis, modeling, and writing for Experiments 1 and 2 are my own work.

References

Fawcett, C. A., & Markson, L. (2010). Children reason about shared preferences. *Developmental psychology*, *46*(2), 299.s

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, *21*(8), 1134-1140.

Lucas, C. G., Griffiths, T. L., Xu, F., Fawcett, C., Gopnik, A., Kushnir, T., ... & Hu, J. (2014). The child as econometrician: A rational model of preference understanding in children. *PloS one*, *9*(3), e92160.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, *15*(3), 215-233.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of applied Econometrics*, *15*(5), 447-470.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: evidence from 14-and 18-month-olds. *Developmental psychology*, *33*(1), 12.

Sutton, R. S. and Barto A. G. *Reinforcement Learning: An Introduction*. The MIT Press, Cambridge, MA, 1998.