# An Economic Model of User Rating in an Online Recommender System

F. Maxwell Harper[1], Xin Li[2], Yan Chen[2], and Joseph A. Konstan[1]
CommunityLab*

[1] University of Minnesota, Minneapolis, MN, 55406 USA
{harper, konstan}@cs.umn.edu
[2] University of Michigan, Ann Arbor, MI, 48109 USA
{xinl, yanchen}@umich.edu

**Abstract.** Economic modeling provides a formal mechanism to understand user incentives and behavior in online systems. In this paper we describe the process of building a parameterized economic model of user-contributed ratings in an online movie recommender system. We constructed a theoretical model to formalize our initial understanding of the system, and collected survey and behavioral data to calibrate an empirical model. This model explains 34% of the variation in user rating behavior. We found that while economic modeling in this domain requires an initial understanding of user behavior and access to an uncommonly broad set of user survey and behavioral data, it returns significant formal understanding of the activity being modeled.

## 1 Introduction

Designers of online communities struggle with the challenge of eliciting participation from their members. Butler [4] found that 50% of social, hobby, and work mailing lists had no traffic over a 122 day period. Under-contribution is a problem even in communities that do survive; in a majority of active mailing lists, fewer then 50% of subscribers posted even a single message in a four month period [4].

Recommender systems built on collaborative filtering [14] are particularly vulnerable to the problem of undercontribution. If users do not contribute ratings to the community, especially for new and rarely-rated items, the system loses its ability to produce recommendations–its main purpose for existence.

We have been conducting research on how to increase the number of ratings contributed to MovieLens [8, 13], a movie recommendation web site. In this paper we report on our activity using economic modeling to build a parameterized model of the motivations underlying user rating behavior. We model factors that affect users' willingness to rate movies, such as the desire to view accurate movie recommendations and the time and effort needed to rate movies. We

* CommunityLab is a collaborative project of the University of Minnesota, University of Michigan, and Carnegie Mellon University. http://www.communitylab.org/

L. Ardissono, P. Brna, and A. Mitrovic (Eds.): UM 2005, LNAI 3538, pp. 307–316, 2005.

believe that economic modeling will guide future site development by providing us with insights into user motivations, predictions about user behavior, and opportunities to personalize the site to match user goals.

As far as we know, this is the first use of economic analysis to build models of user incentives and behavior in an adaptive web site. Accordingly, we discuss in detail the modeling process and the ways in which such a model can be applied. We discuss why economic modeling in this domain is difficult, and what benefits modelers can expect.

## 2 Related Work

Informal economic analysis has been used to inform the design and analysis of computer-supported cooperative work (CSCW) applications. Grudin, for example, assessed relative costs and benefits of using digital voice versus text in a groupware application [7]. More formal economic modeling has been applied to develop probabilistic models of human interruptibility [9].

The design of online trust and reputation management systems has been heavily influenced by economic theories. For example, Keser used experimental economics to demonstrate the effects of reputation management systems such as eBay's on marketplace success [10]. Friedman and Resnick examined the theoretical effects on trust and reputation based on enforcing costly or permanent pseudonyms in online communities [5].

More broadly, economic theories have informed research investigating the design of e-commerce and auction sites. Bakos examined the theoretical implications of the accessibility of product descriptions and pricing information in e-commerce Web sites, including cases where sellers have intentionally introduced difficulties into the search process to increase revenue [2].

To our knowledge, economic analysis has not been used to inform the design of user adaptive Web sites. Adaptive hypermedia researchers have traditionally used a variety of approaches to explicitly or implicitly gather data about users' knowledge, goals, background, or preferences [3]. This work extends these approaches by modeling incentives and behavior in the language of economics.

## 3 An Economic Model of MovieLens Users

The primary purpose of economic modeling is to generate insights into complex problems [15] and to make predictions about rational agents' behavior. In a canonical economic model, agents act in order to maximize their objectives, subject to constraints. To create an effective model, we make simplifying assumptions while still keeping the essential features of the real world situation.

### 3.1 A Theoretical Model

In a typical session in MovieLens, users spend time rating movies and viewing movie recommendations. Their activity can be modeled mathematically as follows: let $x_i$ be the number of movies user $i$ has rated, and $X_{-i} = \sum_{j \neq i} x_j$ be the total number of ratings from all other users in MovieLens. Based on survey data and our understanding of user behavior from interactions with users, a user's benefit from using MovieLens comes from three sources:

– Recommendation quality, $Q_i(x_i, X_{-i})$: Users enjoy viewing useful movie recommendations. Based on the characteristics of the MovieLens recommendation algorithm, we assume this function is concave in both its components. That is, it increases along with each rating count, but at a decreasing rate.
– Rating fun, $f_i(x_i)$: Users enjoy expressing opinions about movies. We assume that $f'(x_i) > 0$, and $f''(x_i) \leq 0$. Again, rating more movies brings more enjoyment, but at a decreasing rate.
– Non-rating fun, $h_i$: Users enjoy activities such as searching for movies and reading information about movies.

We further assume that there is cost associated with rating. The cost function of rating movies, $c_i(x_i)$, represents the amount of time that user $i$ needs to rate $x_i$ movies. Assume that $c_i(x_i)$ is convex, i.e., the marginal cost is positive, $c_i^{'}(x_i) > 0$, and $c_i^{''}(x_i) \geq 0$ for all $i \in N$. Thus, the marginal cost of rating either remains constant or increases with the number of ratings. This fits with our experience that users rate popular, easily remembered movies first.

For analytical tractability, we assume that various components of the utility function are additively separable. Let $\gamma_i$ denote the marginal benefit to recommendation quality of rating one movie. We can represent user $i$'s utility, $\pi_i$, as

$$\pi_i = \gamma_i Q_i(x_i, X_{-i}) + f_i(x_i) + h_i - c_i(x_i). \tag{1}$$

To calibrate this model with survey and behavioral data, we now parameterize various components of the utility function, and solve for the optimal number of ratings. Based on features of the recommendation algorithm, we assume that $Q_i(x_i, X_{-i}) = \min(\bar{R}, X_{-i}^{\alpha} x_i^{\beta_i})$. This is a Cobb-Douglas production function[1] with an upper bound, $\bar{R}$. The upper bound is included to represent the fact that average recommendation quality has a theoretical limit. $\alpha \in [0, 1]$ measures the impact of system-wide ratings on recommendation quality. $\beta_i \in [0, 1]$ measures user $i$'s taste in movies. A higher $\beta_i$ indicates that a user has rare taste, while a lower $\beta_i$ indicates that a user has mainstream taste.

Furthermore, assume that both the rating fun function and the cost function are linear such that $f_i(x_i) = f_i x_i$ and $c_i(x_i) = c_i x_i$, respectively. While neither function is necessarily linear in general, we do not have enough data to estimate their shape.

Under these assumptions, we consider two cases. In the first case, when $\min(\bar{R}, X_{-i}^{\alpha} x_i^{\beta_i}) = \bar{R}$, a user is getting the best possible recommendation quality from MovieLens. In this case, a user will continue to rate movies as long as the marginal fun is greater than or equal to the marginal cost.

In the second case, when $\min(\bar{R}, X_{-i}^{\alpha} x_i^{\beta_i}) = X_{-i}^{\alpha} x_i^{\beta_i}$ and $\alpha \to 0$, a user's recommendation quality will improve via ratings, but will not improve due to

[1] The Cobb-Douglas production is one of the most commonly used production functions in economics [12].

others contributing ratings to the system - a simplifying assumption possible in MovieLens where there is a large stock of total ratings. In this case, we solve Equation (1) for the optimal number of ratings, $x_i^*$:

$$x_i^* = \left(\frac{\beta_i \gamma_i}{c_i - f_i}\right)^{\frac{1}{1-\beta_i}}. \tag{2}$$

Taking a log transformation of Equation (2), we get

$$\ln x_i^* = \frac{1}{1-\beta_i}[\ln \beta_i + \ln \gamma_i - \ln(c_i - f_i)]. \tag{3}$$

As a sanity check, we change various parameters and see if the optimal number of ratings is moving in the right direction. An increase in marginal cost leads to a decrease in rating quantity, while an increase in marginal fun or marginal benefit from recommendation quality leads to an increase in rating quantity. When $\frac{1-\beta_i}{\beta_i} + \ln \beta_i + \ln \gamma_i - \ln(c_i - f_i) > 0$, an increase in $\beta_i$ also leads to more rating, indicating that, other things being equal, a user with rare taste will rate more movies than one with mainstream taste. These results are consistent with our intuition.

### 3.2 Data to Calibrate the Model

To calibrate our model, we collected both survey and behavioral data. An online survey consisting of ten multi-part questions was given to 357 users in June and July, 2004[2]. Only users who had logged in at least 3 times and who had rated at least 30 movies were presented with an invitation to participate. The survey was promoted on the MovieLens main page.

The survey focused on understanding users' motivations. Motivations are not only important for understanding user costs and benefits, but can later be used to calibrate reduced models for new users with little history data.

We found that MovieLens users do have differing motivations. 92% of users listed viewing movie recommendations as one of their top-three reasons for using the system. However, we found that people rate movies for a wider variety of reasons: to keep a personal list of movies they've seen, to influence others, and because they find rating itself to be fun. We also found that users perceive that the quality of movie recommendations provided by the system improves over time.

We also gathered historical behavioral data about the volunteers who took the survey. This data includes, for example, information about the use of MovieLens features and the quality of recommendations received. Table 1 summarizes some of the key behavioral variables we used in this study.

The users that we studied were disproportionately "power users", i.e., those users who use the system often and rate a lot of movies. They also tended to be quite happy with MovieLens, based on their survey responses. However, the

[2] See `http://www.grouplens.org/data/mlsurvey0604.html` for a list of survey questions and a summary of responses.

**Table 1.** Selected MovieLens Survey-Taker Behavioral Data

| | Mean | Median | Min. | Max. | Std. Dev. |
|---|---|---|---|---|---|
| # Ratings, User Lifetime | 693.52 | 556 | 41 | 3235 | 525.16 |
| # Ratings, Last 3 Months | 86.67 | 37 | 0 | 1041 | 150.34 |
| Recent Error of Recommendations[a] | 0.54 | 0.50 | 0.13 | 1.57 | 0.22 |
| Unusual Taste[b] | 0.65 | 0.64 | 0.32 | 1.24 | 0.14 |
| Fraction of Ratings that are Rare Movies[c] | 0.07 | 0.05 | 0 | 0.33 | 0.06 |
| Number of Movie Suggestions Contributed | 3.00 | 0 | 0 | 218 | 13.87 |
| # Saved Movie Searches | 3.68 | 3 | 0 | 93 | 5.54 |
| # Sessions, Last 3 Months | 24.61 | 12 | 1 | 299 | 35.99 |
| Fraction of Sessions w/ Ratings, Last 3 Months | 0.60 | 0.60 | 0 | 1 | 0.26 |
| Weeks Since Registration | 33.21 | 34 | 1 | 72 | 22.99 |

[a] Measured by the mean absolute error (MAE) of last 20 ratings.
[b] Measured by the mean distance between a user's ratings and system average ratings.
[c] We define a rare movie as one with fewer than 250 ratings from the 88,000 users in MovieLens. By comparison, the top 100 movies average about 23,500 ratings each.

users in our study did exhibit a great deal of variety in a number of areas such as movie taste, interface customization and usage, and in the average quality of movie recommendations they receive.

### 3.3 Calibration and Results

We now describe an empirical model that we will calibrate using survey and behavioral data. Recall that Equation (3) characterizes the optimal number of ratings for a user. Since the distribution of ratings is skewed, we use a logarithm transformation of the number of ratings as the dependent variable. The main explanatory variables include a user's marginal benefit (MB) from the quality of recommendations, $\gamma_i$, the taste parameter, $\beta_i$, the fun score, $f_i$, and the marginal cost (MC) of providing additional ratings, $c_i$. We also control for other characteristics, $\overrightarrow{Z}$, such as a user's age in MovieLens and how many times a user has used MovieLens recently. Our empirical model is defined as

$$\ln x_i = a_0 + a_1\gamma_i + a_2\beta_i + a_3 f_i + a_4 c_i + \overrightarrow{\Lambda}\overrightarrow{Z} + \varepsilon_i. \tag{4}$$

Calibrating the marginal cost and benefit parameters proved challenging in constructing the empirical model. Our survey questions designed for calibrating these parameters were phrased in terms of money, but many survey takers resisted assigning monetary values to a free web service. As such, many respondents failed to answer these questions, and many responded with $0 values or very unlikely costs. To handle this type of truncated data, we employed a Tobit maximum likelihood approach[3] to predict marginal benefit and cost using other survey and behavioral data.

[3] The Tobit model is an econometric model able to handle the case where the dependent variable is zero for a nontrivial fraction of the sample.[16, 6].

**Table 2.** Tobit Analysis: Estimating Marginal Benefit and Marginal Cost

| | (1) Reported MB | (2) Reported MC |
|---|---|---|
| Freq of Picking Movies to Watch | 1.358 (0.515)*** | |
| Freq of Searching for a Particular Movie | 0.198 (0.431) | |
| Freq of Looking Only at 1st Screen of Recs | -0.786 (0.333)** | |
| Freq of Looking at 5+ Screens of Recs | 0.420 (0.447) | |
| # "Hide this Movie" Ratings | 0.001 (0.003) | |
| # Saved Searches | 0.100 (0.053)* | |
| Reported Time Estimate to Rate 10 Movies | | 4.231 (1.076)*** |
| # Ratings/Login, Last 3 Months | | 0.184 (1.066) |
| Constant | -4.246 (1.551)*** | -13.391 (3.880)*** |
| Observations | 339 | 338 |
| Pseudo R-squared | 0.02 | 0.03 |
| Corr(predicted, user reported) | 0.228 | 0.320 |
| p-value | 0.000 | 0.000 |

*Notes:*

1. Standard errors in parentheses.
2. Significant at: * 10-percent level; ** 5-percent level; *** 1-percent level.

The results of the Tobit estimation are presented in Table 2. The strongest indicator of marginal benefit is the frequency of using MovieLens to pick movies to watch. This indicator, along with other data reflecting usage patterns, provide an estimated value for $\gamma_i$ that correlates with the monetized survey responses at 0.228. The estimation of marginal cost is based on survey responses concerning the time required to rate movies along with behavioral data on the number of ratings provided per session. The correlation between our estimated measure of $c_i$ and the monetized value provided in the survey is 0.320. We report later on the strength of these measures in the discussion of the final model.

The taste and fun parameters are constructed more directly. The taste parameter is constructed from how often a user rates rare movies (as defined above, in Table 1) and how different a user's ratings are from movie averages. The fun score is derived from the frequency of using MovieLens to rate just-seen movies, the number of ratings sessions per month, and several other behavioral factors reflecting enjoyment of the rating process.

Table 3 reports the results of the empirical analysis, where we explain individual users' rating behavior in terms of Equation (4). Column (1) shows the explanatory variables used in the analysis. Column (2) shows only behavioral data to demonstrate the relative power of a reduced model without survey data. We focus on column (1) when interpreting the results.

The results are consistent with the theoretical predictions. Marginal cost has the expected significant and negative correlation with the quantity of ratings. Marginal benefit from movie recommendation shows a positive but not statistically significant correlation with number of ratings.

As suggested by survey responses, many MovieLens users consider rating movies to be an entertaining activity. This is reflected by the positive coefficient

**Table 3.** Regression Analysis: Predicting the Quantity of User Ratings

| | Dependent Variable: log(ratings) | |
|---|---|---|
| | (1) | (2) |
| | Behavioral+Survey | Behavioral only |
| MC of Rating 10 Movies, $c_i$ | -0.042 (0.019)** | |
| MB of 10 Recommendations, $\gamma_i$ | 0.028 (0.053) | |
| Fun Score, $f_i$ | 0.353 (0.104)*** | |
| Uncommon Taste, $\beta_i$ | 0.974 (0.284)*** | 0.922 (0.290)*** |
| % Ratings that are Rare Movies, $\beta_i$ | 1.906 (0.379)*** | 2.137 (0.382)*** |
| Altruism Score, $\overrightarrow{Z}$ | -0.053 (0.030)* | |
| Weeks Since Registration, $\overrightarrow{Z}$ | 0.001(0.001)** | |
| Helpful Subject Score, $\overrightarrow{Z}$ | 0.171(0.050)*** | 0.195 (0.050)*** |
| # Logins, Last 3 Months, $\overrightarrow{Z}$ | 0.004 (0.001)*** | 0.006 (0.001)*** |
| % Sessions with Rating Activity, $\overrightarrow{Z}$ | 0.732 (0.144)*** | 0.872 (0.141)*** |
| Recent Error of Recommendations, $\overrightarrow{Z}$ | -0.477 (0.186)** | -0.427(0.191)** |
| Constant, $a_0$ | 4.351 (0.308)*** | 4.606 (0.210)*** |
| Observations | 356 | 356 |
| Adjusted R squared | 0.342 | 0.304 |
| Corr(predict #ratings, actual #ratings) | 0.622 | 0.514 |
| p-value | 0.000 | 0.000 |

*Notes:*

1. Standard errors in parentheses.
2. Significant at: * 10-percent level; ** 5-percent level; *** 1-percent level.

of the fun score, statistically significant at 1 percent. Both measures of taste are significant and have strong effects, which confirms the theoretical prediction that users with rare tastes tend to rate more.

Our control variables, used to account for user-specific characteristics, also improve the overall predictive power of the model. The percentage of sessions over the last three months that includes rating activity has a strong, significant effect. Recent error of recommendations, measured in terms of a user's mean absolute error (MAE) over the last 20 ratings has a significant negative effect on predicted ratings.

The regression analysis has an adjusted R-squared of 0.342 and the correlation between the predicted and actual number of ratings is 0.622 ($p < 0.001$). As is common practice in the social sciences, we report the adjusted R-squared, which imposes a penalty for adding additional but irrelevant independent variables. In terms of a cross-sectional economic study, this is a strong result.[4]

[4] "In the social science, low R-squared in regression equations are not uncommon, especially for cross-sectional analysis."[16] For cross-section data, such as those we have, a R-squared above 0.2 is usually considered decent. For example, Ashenfelter and Krueger report R-squared in the range of 0.2 and 0.3, with a sample size of 298 [1]. Levitt reports R-squared in the range of 0.06 and 0.37 with a sample size between 1,276 and 4,801 [11].

Table 3 also shows the results of our reduced model, which consists of only behavioral data. The adjusted R-squared is 0.304 and the correlation between the predicted number of ratings and the actual number of ratings is 0.514 ($p < 0.001$), both slightly worse than the full model. However, the advantage of the reduced model is that the necessary data are available without extensive surveying.

We have conducted ten-fold cross-validation for both the full and reduced empirical models. The results are robust and are available upon request.

## 4 Discussion

Achieving an R-squared value of 0.34 implies that we are able to explain a significant portion of individual rating behavior. This has two direct applications. First, we use these results to increase our understanding of user motivations and behavior. We have identified markers of behavior that guide us in further site development. To be specific, before conducting this analysis we believed that a particularly effective way of increasing ratings would be to reveal to users the extent of their effect on others. In view of the results we found, we now believe it may be more effective to focus on increasing the fun and non-prediction personal benefits of rating through better interfaces for rating and making lists, better interfaces for browsing collections of one's own ratings, and increased use of games that engage users in the system. At the same time, we originally were quite skeptical of any "pre-surveys" or other barriers to entering the system, but now see that using them may serve as an indicator of good citizenship and might well lead to increasing the percentage of new users who become high raters.

Second, we can use these results to start thinking about personalized interfaces, to go along with the already-personalized content. Now that we can efficiently fit users into this model, we can choose to emphasize different elements of the system to them. Users who most directly benefit from prediction quality can be given updated information on the quality of their recent predictions and the estimated increase in quality from the next quantum of ratings. Users who are more interested in the fun of rating itself can receive different cues and prompts.

Of course, we must include a very important caveat. We discovered quite early that the users in our sample are very much power users. Before making major site changes that would affect all users we would want to extend this analysis to include a broader range of newer and infrequent users.

**Lessons Learned.** The entire process of conducting this analysis was filled with lessons, many of them the direct result of a first-time collaboration between a pair of computer scientists and a pair of economists. While the process was rewarding, we should warn those attempting it the first time that there is a substantial learning curve. The computer scientists in the team not only had to re-learn the Greek alphabet, but had to learn to formalize years of intuition about user behavior in new ways. This led to challenging but rewarding attempts to operationalize the abstract parameters of the analysis through mixtures of survey and behavioral data. At the same time, the nature of working with the online MovieLens community handicapped the economists on our team. In contrast with most experimental economics work, our users steadfastly resisted attempts to monetize their experience with the system, adding substantial challenges to the task of estimating value and cost.

If we were to repeat this effort, we would likely take a more iterative approach to surveying the users. While our survey *design* was appropriate for our modeling task, the effect of user behavior on that design made the modeling much harder. With greater iteration we probably would have been better able to substitute time for money in the overall analysis of cost and value.

Finally, we must address the question of whether economic modeling is a valuable approach for studying and personalizing an interactive web site. While this answer certainly depends upon the details of the site, in general we think it is so long as a sufficient amount of data is available to support the process. Economic models have the nice property of building formality from a base of initial understanding. Unlike a neural network, they don't simply appear from data. But unlike a neural network, they return significant understanding of the population being analyzed.

**Future Work.** We are currently engaged in a series of experiments to learn whether certain laboratory-tested economic theories of collective action apply to the more real-world environment of online communities. Specifically, we plan a set of field studies that look at theories of reciprocity and inequality aversion to determine how user contributions to a collective good are affected by awareness of the contributions of others. Following this work, we plan to explore incentive-personalized interfaces to MovieLens–interfaces that provide the specific cues and information that motivate each particular user to contribute to the system.

## 5 Conclusions

Economic modeling is a formal method for combining initial understanding about a user population with data to refine that understanding. We developed an economic model of rating behavior of MovieLens users, tying that behavior to a number of factors that determine how much the user benefits from ratings–directly and indirectly–and how much effort the user requires to enter those ratings. The process gave us insight into the motivations of our user community, and resulted in a useful model able to explain a substantial percentage of user variation in rating behavior.

**Acknowledgments.** We would like to thank Dan Cosley and Sean McNee for their feedback on early drafts of this paper, and Robert Kraut for his ideas that helped guide our research. This work is supported by a grant from the National Science Foundation (IIS-0324851).

## References


1. Ashenfelter, O., Krueger, A.B.: Estimates of the economic returns to schooling from a new sample of twins. American Economic Review **84** (1994) 1157–73
2. Bakos, J.Y.: Reducing buyer search costs: implications for electronic marketplaces. Manage. Sci. **43** (1997) 1676–1692
3. Brusilovsky, P.: Methods and techniques of adaptive hypermedia. User Modeling and User-Adapted Interaction **6** (1996) 87–129
4. Butler, B.S.: Membership size, communication activity, and sustainability: A resource-based model of online social structures. Info. Sys. Research **12** (2001) 346–362
5. Friedman, E.J., Resnick, P.: The social cost of cheap pseudonyms. Journal of Economics & Management Strategy **10** (2001) 173–199
6. Greene, W.H.: Econometric analysis. 4th edn. Prentice Hall, Upper Saddle River, NJ (2000)
7. Grudin, J.: Why CSCW applications fail: problems in the design and evaluation of organizational interfaces. In: Proceedings of CSCW '88, ACM Press (1988) 85–93
8. Herlocker, J., Konstan, J.A., Riedl, J.: An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. Inf. Retr. **5** (2002) 287–310
9. Horvitz, E., Apacible, J.: Learning and reasoning about interruption. In: Proceedings of ICMI '03, ACM Press (2003) 20–27
10. Keser, C.: Experimental games for the design of reputation management systems. IBM Systems Journal **42** (2003) 498–506
11. Levitt, S.D.: Using electoral cycles in police hiring to estimate the effect of police on crime. American Economic Review **87** (1997) 270–90
12. Mas-Colell, A., Whinston, M.D., Green, J.R.: Microeconomic Theory. Oxford University Press (1995)
13. McNee, S.M., Lam, S.K., Konstan, J.A., Riedl, J.: Interfaces for eliciting new user preferences in recommender systems. In: Proceedings of User Modeling 2003. (2003) 178–187
14. Resnick, P., Varian, H.R.: Recommender systems. Commun. ACM **40** (1997) 56–58
15. Varian, H.R.: How to build an economic model in your spare time. In Szenberg, M., ed.: Passion and Craft, How Economists Work. University of Michigan Press (1995)
16. Wooldridge, J.M.: Introductory Econometrics: A Modern Approach. 2nd edn. South-Western College (2002)