Analysis of the Axiomatic Foundations of Collaborative Filtering

David M. Pennock

University of Michigan
Artificial Intelligence Lab
1101 Beal Ave
Ann Arbor, MI 48109-2110
dpennock@umich.edu

Eric Horvitz

Microsoft Research
One Microsoft Way
Redmond, WA 98052-6399
horvitz@microsoft.com

Abstract

The growth of Internet commerce has stimulated the use of collaborative filtering (CF) algorithms as recommender systems. Such systems leverage knowledge about the behavior of multiple users to recommend items of interest to individual users. CF methods have been harnessed to make recommendations about such items as web pages, movies, books, and toys. Researchers have proposed several variations of the technology. We take the perspective of CF as a methodology for combining preferences. The preferences predicted for the end user is some function of all of the known preferences for everyone in a database. Social Choice theorists, concerned with the properties of voting methods, have been investigating preference aggregation for decades. At the heart of this body of work is Arrow's result demonstrating the impossibility of combining preferences in a way that satisfies several desirable and innocuous-looking properties. We show that researchers working on CF algorithms often make similar assumptions. We elucidate these assumptions and extend results from Social Choice theory to CF methods. We show that only very restrictive CF functions are consistent with desirable aggregation properties. Finally, we discuss practical implications of these results.

Introduction

The goal of collaborative filtering (CF) is to predict the preferences of one user, referred to as the *active* user, based on the preferences of a group of users. For example, given the active user's ratings for several movies and a database of other users' ratings, the system predicts how the active user would rate unseen movies. The key idea is that the active user will prefer those items that like-minded people prefer, or even that dissimilar people don't prefer. The effectiveness of any CF algorithm is ultimately predicated on the underlying assumption that human preferences are correlated—if they were not, then informed prediction would be impossible. There does not seem to be a single, obvious way to predict preferences, nor to evaluate effectiveness, and many different algorithms and evaluation criteria have been proposed and tested. Most comparisons to date have been empirical or qualitative in nature [Billsus and Pazzani, 1998; Breese *et al.*, 1998; Konstan and Herlocker, 1997; Resnick and Varian, 1997; Resnick *et al.*, 1994; Shardanand and Maes, 1995], though some worst-case performance bounds have been derived [Freund *et al.*, 1998; Nakamura and Abe, 1998] and some general principles have been advocated [Freund *et al.*, 1998]. Initial methods were statistical, though several researchers have recently cast CF as a machine learning problem [Billsus and Pazzani, 1998; Freund *et al.*, 1998; Nakamura and Abe 1998].

We take instead an axiomatic approach, informed by results from Social Choice theory. First, we identify several properties that a CF algorithm might ideally posses, and describe how existing CF implementations obey subsets of these conditions. We show that, under the full set of conditions, only one prediction strategy is possible: The ratings of the active user are derived solely from the ratings of only one other user. This is called the *nearest neighbor* approach [Freund *et al.*, 1998]. The analysis mirrors Arrow's celebrated Impossibility Theorem, which shows that the only voting mechanism that obeys a similar set of properties is a dictatorship [Arrow, 1967; Arrow, 1963]. Under slightly weaker demands, we show that the only possible form for the prediction function is a weighted average of the users' ratings. We also provide a second, separate axiomatization that again admits only the weighted average. The weighted average method is used in practice in many CF applications [Breese *et al.*, 1998; Resnick *et al.*, 1994; Shardanand and Maes, 1995]. One contribution of this paper is to provide a formal justification for it. Stated another way, we identify a set of properties, one of which must be violated by any non-weighted-average CF method. On a broader level, this paper proposes a new connection between theoretical results in Social Choice theory and in CF, providing a new perspective on the task. This angle of attack could lead to other fruitful links between the two areas of study, including a category of CF algorithms based on voting mechanisms. The next section covers background on CF and Social Choice theory. The remaining sections present, in turn, the three axiomatizations, and discuss the practical implications of our analysis.

Background

In this section, we briefly survey previous research in collaborative filtering, describe our formal CF framework, and present relevant background material on utility theory and Social Choice theory.

Collaborative Filtering Approaches

A variety of collaborative filters or recommender systems have been designed and deployed. The Tapestry system relied on each user to identify like-minded users manually [Goldberg *et al*., 1992]. GroupLens [Resnick *et al*., 1994] and Ringo [Shardanand and Maes, 1995], developed independently, were the first CF algorithms to automate prediction. Both are examples of a more general class we call *similarity-based* approaches. We define this class loosely as including those methods that first compute a matrix of pairwise similarity measures between users (or between titles). A variety of similarity metrics are possible. Resnick *et al*. [1994] employ the Pearson correlation coefficient for this purpose. Shardanand and Maes [1995] test a few measures, including correlation and mean squared difference. Breese *et al*. [1998] propose a metric called vector similarity, based on the vector cosine measure. All of the similarity-based algorithms cited predict the active user's rating as a weighted sum of the others users' ratings, where weights are similarity scores. Yet there is no *a priori* reason why the weighted average should be the aggregation function of choice. Below, we provide two possible axiomatic justifications.

Breese *et al*. [1998] identify a second general class of CF algorithms called *model-based* algorithms. In this approach, an underlying model of user preferences (for example, a Bayesian network model) is first constructed, from which predictions are inferred.

CF technology is in current use in several Internet commerce applications. For example, firefly (http://www.firefly.com), originally a recommender much like GroupLens and Ringo, offers more general personalized services based on individual and community preferences. Alexa (http://www.alexa.com) is a web browser plug-in that recommends related links based in part on other people's web surfing habits.

Formal Description of Task

A CF algorithm recommends items or *titles* to the active user based on the ratings of n others. Denote the set of all titles as T and the rating of user i for title j as $r_i(j)$. The function $r_i: T \rightarrow \Re \cup \{\perp\}$ maps titles to real numbers or to $\perp$, the symbol for "no rating." Denote the vector of all of user i's ratings for all titles as $\boldsymbol{r}_i(T)$, and the vector of all of the active user's ratings as $\boldsymbol{r}_a(T)$. Define $NR \subset T$ to be the subset of titles that the active user has not rated, and thus for which we would like to provide predictions. That is, title j is in the set NR if and only if $r_a(j) = \perp$. Then the subset of titles that the active user *has* rated is $T\text{-}NR$. Define the vector $\boldsymbol{r}_i(S)$ to be all of user i's ratings for any subset of titles $S \subseteq T$, and $\boldsymbol{r}_a(S)$ analogously. Finally, denote the matrix of all users' ratings for all titles simply as $\boldsymbol{r}$.

In general terms, a collaborative filter is a function f that takes as input all ratings for all users, and outputs the predicted ratings for the active user:

$$\boldsymbol{r}_a(NR) \; = \; f(\boldsymbol{r}_1(T), \boldsymbol{r}_2(T), \ldots , \boldsymbol{r}_n(T)) \; = \; f(\boldsymbol{r}) \; , \qquad (1)$$

where the $\boldsymbol{r}_i(T)$'s include the ratings of the active user.

Utility Theory and Social Choice Theory

Social choice theorists are also interested in functions similar to (1), though they are concerned with combining preferences or utilities instead of ratings. Preferences refer to ordinal rankings of outcomes. For example, Alice's preferences might hold that sunny days (*sd*) are better than cloudy days (*cd*), and cloudy days are better than rainy days (*rd*). Utilities, on the other hand, are numeric expressions. Alice's utilities $\boldsymbol{u}$ for the outcomes *sd*, *cd*, and *rd* might be $u(sd) = 10$, $u(cd) = 4$, and $u(rd) = 2$, respectively. If Alice's utilities are such that $u(sd) > u(cd)$, then Alice prefers *sd* to *cd*. Axiomatizations by Savage [1954] and von Neumann and Morgenstern [1953] provide persuasive postulates which imply the existence of utilities, and show that maximizing expected utility is the optimal way to make choices. If two utility functions $\boldsymbol{u}$ and $\boldsymbol{u}'$ are positive linear transformations of one another, then they are considered equivalent, since maximizing expected utility would lead to the same choice in both cases.

Now consider the problem of combining many peoples' preferences into a single expression of societal preference. Arrow proved the startling result that this aggregation task is simply impossible, if the combined preferences are to satisfy a few compelling and rather innocuous-looking properties [Arrow, 1967; Arrow, 1963].[1] This influential result forms the core of a vast literature in Social Choice theory. Sen [1986] provides an excellent survey of this body of work. Researchers have since extended Arrow's theorem to the case of combining utilities. In general, economists argue that the absolute magnitude of utilities are not comparable between individuals, since utilities are invariant under positive affine transformations. In this context, Arrow's theorem on preference aggregation applies to the case of combining utilities as well [Fishburn, 1987; Sen, 1986].

---

[1] Arrow won the Nobel Prize in part for this result, which is ranked at `http://www.northnet.org/clemens/decor/mathhist.htm` as one of seven milestones in mathematical history this century.

Nearest Neighbor Collaborative Filtering

We now describe four conditions on a CF function, argue why they are desirable, and discuss how existing CF implementations adhere to different subsets of them. We then show that the only CF function that satisfies all four properties is the nearest neighbor strategy, in which recommendations to the active user are simply the preferred titles of one single other user.

Property 1 (UNIV) *Universal domain and minimal functionality*. The function $f(\boldsymbol{r})$ is defined over all possible inputs $\boldsymbol{r}$. Moreover, if $r_i(j) \neq \perp$ for all $i \neq a$ and for all $j \in NR$, then $r_a(j) \neq \perp$ for all $j \in NR$.

UNIV simply states that f must always returns some output and, when all users rate all titles in NR, f must return ratings. To our knowledge, all existing CF functions adhere to this property.

Property 2 (UNAM) *Unanimity*. For all $j,k \in NR$, if $r_i(j) > r_i(k)$ for all $i \neq a$, then $r_a(j) > r_a(k)$.

UNAM is often called the weak Pareto property in the Social Choice and Economics literatures. Under this condition, if all users rate j strictly higher than k, then we predict that the active user will prefer j over k.

This property seems natural: If everyone agrees that title j is better than k, including those most similar to the active user, then it hard to justify a reversed prediction. Nevertheless, correlation methods can violate UNAM if, for example, the active user is negatively correlated with all other users. Other similarity-based techniques that use only positive weights, including vector similarity and mean squared difference, do satisfy this property.

Property 3 (IIA) *Independence of Irrelevant Alternatives*. Consider two input ratings matrices, $\boldsymbol{r}$ and $\boldsymbol{r}'$, such that $\boldsymbol{r}(T\text{-}NR) = \mathbf{r}'(T\text{-}NR)$. Furthermore, suppose that $r(j) = r'(j)$ and $r(k) = r'(k)$ for some $j,k \in NR$. That is, $\boldsymbol{r}$ and $\boldsymbol{r}'$ are identical on all ratings of titles that the active user has seen, and on two of the titles, j and k, that the active user has not seen. Then $r_a(j) > r_a(k)$ if and only if $r'_a(j) > r'_a(k)$.

The intuition for IIA is as follows. The ratings $\boldsymbol{r}(T\text{-}NR)$ for those titles that the active user has seen tell us how similar the active user is to each of the other users, and we assume that the ratings $\boldsymbol{r}(NR)$ do not bear upon this similarity measure. This is the assumption made by most similarity-based CF algorithms. Once a similarity score is calculated, it makes sense that the predicted relative ranking between two titles j and k should only depend on the ratings for j and k. For example, if the active user has not rated the movie "Waterworld," then everyone else's opinion of it should have no bearing on whether the active user prefers "Ishtar" to "The Apartment," or vice versa.

IIA lends stability to the system. To see this, suppose that $NR = \{j, k, l\}$, and f predicts the active user's ratings such that $r_a(j) > r_a(k) > r_a(l)$, or title j is most recommended. Now suppose that a new title, m, is added to the database, and that the active user has not rated it. If IIA holds, then the relative ordering among j, k, and l will remain unchanged, and the only task will be to position m somewhere within that order. If, on the other hand, the function does not adhere to IIA, then adding m to the database might upset the previous relative ordering, causing k, or even l, to become the overall most recommended title. Such an effect of presumably irrelevant information seems counterintuitive.

All of the similarity-based CF functions identified here—GroupLens, Ringo, and vector similarity—obey IIA.

Property 4 (SI) *Scale Invariance*. Consider two input ratings matrices, $\boldsymbol{r}$ and $\boldsymbol{r}'$, such that, for all users i and titles j, $r_i(j) = \alpha_i \cdot r'_i(j) + \beta_i$ for any positive constants α_i and any constants β_i. Then $r_a(j) > r_a(k)$ if and only if $r'_a(j) > r'_a(k)$, for all titles $j,k \in NR$.

This property is motivated by the belief, widely accepted by economists [Arrow, 1963; Sen, 1986], that one user's internal scale is not comparable to another user's scale. Suppose that the database contains ratings from 1 to 10. One user might tend to use ratings in the high end of the scale, while another tends to use the low end. Or, the data might even have been gathered from different sources, each of which elicited ratings on a different scale (e.g., one scale is [-10,10], another is [1,100]). We would ideally like to obtain the same results, regardless of how each user reports his or her ratings, as long as his or her mapping from internal utilities to ratings is a positive linear transformation; that is, as long as his or her reported ratings are themselves expressions of utility.

One way to impose SI is to normalize all of the users' ratings to a common scale before applying f. Another way to ensure SI is to constrain f to depend only on the relative rank among titles (the ordinal preferences of users), and *not* on the magnitude of ratings. Freund *et al*. [1998] strongly advocate this approach.

> One important property of [the collaborative filtering] problem is that the most relevant information to be combined represents *relative preferences* rather than *absolute ratings*. In other words, even if the ranking of [titles] is expressed by assigning each [title] a numeric score, we would like to ignore the absolute values of these scores and concentrate only on their relative order.

By ignoring all but relative rank, Freund *et al*.'s algorithm obeys SI. On the other hand, the similarity-based methods violate it.

Different researchers favor one or the other of these four properties; the following proposition shows that only one very restrictive CF function obeys them all.

Proposition 1 (Nearest neighbor). Assuming that $|NR| > 2$, then the only function f of the form (1) that satisfies UNIV, UNAM, IIA, and SI is such that:

$$r_a(j) > r_a(k) \text{ if and only if } r_i(j) > r_i(k) \ ,$$

for all titles $j,k \in NR$, and for one distinguished user i. The choice of user i can depend on the ratings $\boldsymbol{r}(T\text{-}NR)$, but once the "best" i is determined (for example using correlation or vector similarity), his or her ratings for the titles in NR must be fully adopted by the active user.

Proof (Sketch). First, rewrite equation (1) in the following, equivalent, form:

$$\begin{aligned} \boldsymbol{r}_a(NR) &= f(\boldsymbol{r}(T\text{-}NR), \boldsymbol{r}(NR)) \qquad (1) \\ &= g(\boldsymbol{r}(NR)) \ , \end{aligned}$$

where the choice of function g is itself allowed to depend on $\boldsymbol{r}(T\text{-}NR)$. Because f must be defined for all inputs (UNIV), g must be defined for the case when all users, except the active user, have recorded ratings for all titles in NR. With the minimal functionality clause of UNIV, the problem has been cast into the same terms as in the Social Choice literature. It follows, from Sen's [1986] or Robert's [1980] extension of Arrow's theorem [1967; 1963], that g, and therefore f, must be of the nearest neighbor form specified.

Weighted Average Collaborative Filtering

We now examine a slight weakening of the set of properties leading to Proposition 1. Under these new conditions, we find that the only possible CF function is a weighted sum: The active user's predicted rating for each title is a weighted average of the other users' ratings for the same title. Our argument is again based on results from Social Choice theory; we largely follow Fishburn's [1987] explication of work originally due to Roberts [1980].

We replace the SI property with a weaker one:

Property 4* (TI) *Translation Invariance*. Consider two input ratings matrices, $\boldsymbol{r}$ and $\boldsymbol{r}'$, such that, for all users i and titles j, $r_i(j) = \alpha \cdot r'_i(j) + \beta_i$ for any positive constant α, and any constants β_i. Then $r_a(j) > r_a(k)$ if and only if $r'_a(j) > r'_a(k)$, for all titles $j,k \in NR$.

This condition requires that recommendations remain unchanged when all ratings are multiplied by the same constant, and/or when any of the individual ratings are shifted by additive constants. The TI property, like SI, still honors the belief that the absolute rating of one title by one user is not comparable to the absolute rating of another user. Unlike SI, it assumes that the magnitude of ratings differences, $r_i(j) - r_i(k)$ and $r_h(j) - r_h(k)$, *are* comparable between users i and h.

Though they violate SI, the similarity-based methods of GroupLens, Ringo, and vector correlation obey TI.

Proposition 2 (Weighted average). Assuming that $|NR| > 2$, then the only continuous function f of the form (1) that satisfies UNIV, UNAM, IIA, and TI is such that:

$$r_a(j) > r_a(k) \text{ if and only if } \textstyle\sum_i w_i \cdot r_i(j) > \sum_i w_i \cdot r_i(k)$$

for all titles $j,k \in NR$, where all of the w_i are nonnegative, and at least one is positive. The specific weights can depend on the ratings $\boldsymbol{r}(T\text{-}NR)$.

Proof. Follows from Roberts [1980]. •

Proposition 2 does not rule out the nearest neighbor policy, as all but one of the w_i could be zero.

The conditions for Proposition 2 are technically not weaker than those for Proposition 1, as a continuity assumption was added. Alternatively, we could substitute the following stronger property for UNAM [Fishburn, 1987]:

Property 2* (SUNAM) *Strong Unanimity*. For all $j,k \in NR$, if $r_i(j) \geq r_i(k)$ for all $i \neq a$ and $r_h(j) > r_h(k)$ for some $h \neq a$, then $r_a(j) > r_a(k)$.

Weighted Average Collaborative Filtering, … Again

Next, we derive the same conclusion as Proposition 2 working from a different axiomatization. This result is adapted from Harsanyi [1955].

The derivation requires two assumptions.

Property 5 (RRU) *Ratings are utilities*. Each user's rating $\boldsymbol{r}_i(T)$ are a positive linear transformation from his or her utilities. That is, the ratings themselves are expressions of utility.

We also assume that users obey the rationality postulates of expected utility theory [Savage, 1954; von Neumann and Morgenstern, 1953]. For example, if a user's ratings for three titles are such that $r_i(j) > r_i(k) > r_i(l)$, then there is some probability p for which the user would be indifferent between the following two situations: (1) getting title j with probability p or title l with probability $1 - p$, and (2) getting title k for sure.

Property 2 (UnamE)** *Unanimity of Equality*. For all $j,k \in NR$, if $r_i(j) = r_i(k)$ for all $i \neq a$, then $r_a(j) = r_a(k)$.

Proposition 3 (Weighted average, … again). The only function f of the form (1) that satisfies both RRU and UnamE is such that:

$$r_a(j) = \sum_i w_i \cdot r_i(j) ,$$

for all titles $j \in NR$.

Proof. Follows from Harsanyi [1955]. •

Note that this proposition, unlike the previous, admits negative weights.

Implications of the Analysis

We turn to a discussion of the implications of the theoretical limitations highlighted in Propositions 1–3. First, we believe that identifying the connection between CF and Social Choice theory allows CF researchers to leverage a great deal of previous work on preference and utility aggregation. A Social Choice perspective on combining default reasoning rules has yielded valuable insights for that task [Doyle and Wellman, 1991], and similar benefits may accrue for CF. Additionally, weighted versions of any of the many proposed voting schemes [Fishburn, 1973] are immediate candidates for new CF algorithms.

Understanding what is theoretically impossible is an important first step in algorithm design. We believe that the results in this paper may help guide CF development in the future. Though our derivations constrain the type of CF function, they do not contain a recommendation as to how exactly to choose the best neighbor, or how to choose the optimal set of weights. Nonetheless, identifying the functional forms themselves can be of value, by constraining the search among algorithms to one of finding the best instantiation of a particular form.

With regards to real-world applications, CF designers for Internet commerce applications might typically be interested more in the predictive performance of a CF algorithm, rather than in the properties of preference coalescence that it does or does not obey. Yet there is no consensus on how best to measure effectiveness, as evidenced by the proliferation of many proposed evaluation scores. As a result, comparisons among the various algorithms are blurred. Even if a standard, accepted evaluation measure is somehow settled upon, empirical performance can be measured only for a limited number of special cases, whereas the theoretical results apply in all circumstances.

Conclusion

We have illustrated a correspondence between collaborative filtering (CF) and Social Choice theory. Both frameworks center on the goal of combining the preferences (expressed as ratings and utilities, respectively) of a group into a single preference relation. Some of the properties that Social Choice theorists have found to be compelling are also arguably desirable in the context of CF. In particular, universal domain (UNIV) is universally accepted. Unanimity (UNAM) is compelling and common. Most of the other properties have been advocated (at least implicitly) elsewhere in the literature. Similarity-based methods with only positive reinforcement obey UNAM, including vector similarity and mean squared difference. Most other similarity-based techniques obey independence of irrelevant alternatives (IIA) and translation invariance (TI). Freund *et al*. [1998] make the case for scale invariance (SI).

We have identified constraints that a CF designer must live with, if their algorithms are to satisfy sets of these conditions. Along with UNIV and UNAM, IIA and SI imply the nearest neighbor method, while IIA and TI imply the weighted average. A second derivation shows that, if all users' ratings are utilities, and if unanimity of equality holds, then, once again, only the weighted average is available.

Finally, we discussed implications of this analysis, highlighting the fundamental limitations of CF, and identifying a bridge from results and discussion in Social Choice theory to work in CF. This avenue of opportunity includes the implementation of weighted versions of voting mechanisms as potential new CF algorithms.

Acknowledgments

Thanks to Jack Breese for ideas and insights.

References

[Arrow, 1963] Kenneth J. Arrow. *Social Choice and Individual Values*. Yale University Press, second edition, 1963.

[Arrow, 1967] Kenneth J. Arrow. Values and collective decision-making. In *Philosophy, Politics and Society (third series)*, pages 215–232, 1967.

[Billsus and Pazzani, 1998] Daniel Billsus and Michael J. Pazzani. Learning collaborative information filters. In *Proceedings of the Fifteenth International Conference on Machine Learning*, pages 46–54, July 1998.

[Breese *et al*., 1998] John S. Breese, David Heckerman and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence*, pages 43–52, July 1998.

[Doyle and Wellman, 1991] Jon Doyle and Michael P. Wellman. Impediments to universal preference-based default theories. *Artificial Intelligence*, 49: 97–128, 1991.

[Fishburn, 1973] Peter C. Fishburn. *The Theory of Social Choice*. Princeton University Press, Princeton, New Jersey, 1973.

[Fishburn, 1987] Peter C. Fishburn. *Interprofile Conditions and Impossibility*. Harwood Academic Publishers, New York, 1987.

[Freund *et al*., 1998] Yoav Freund, Raj Iyer, Robert E. Schapire, and Yoram Singer. An Efficient boosting algorithm for combining preferences. In *Proceedings of the Fifteenth International Conference on Machine Learning*, pages 170–178, July 1998.

[Goldberg *et al*., 1992] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12): 61–70, December 1992.

[Harsanyi, 1955] John C. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63(4): 309–321, August 1955.

[Konstan and Herlocker, 1997] Joseph A. Konstan, Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R. Gordon, and John Riedl. GroupLens: Applying collaborative filtering to Usenet news. *Communications of the ACM*, 40(3): 77–87, March 1997.

[Nakamura and Abe, 1998] Atsuyoshi Nakamura and Naoki Abe. Collaborative filtering using weighted majority prediction algorithms. In *Proceedings of the Fifteenth International Conference on Machine Learning*, pages 395–403, July 1998.

[Resnick and Varian, 1997] Paul Resnick and Hal R. Varian. Recommender systems. *Communications of the ACM*, 40(3): 56–58, March 1997.

[Resnick *et al*., 1994] Paul Resnick, Neophyts Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. GroupLens: An Open architecture for collaborative filtering of netnews. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pages 175–186, 1994.

[Roberts, 1980] K. W. S. Roberts. Interpersonal comparability and social choice theory. *Review of Economic Studies*, 47: 421–439, 1980.

[Savage, 1954] Leonard J. Savage. *The Foundations of Statistics*. Dover Publications, New York, 1972.

[Sen, 1986] Amartya Sen. Social Choice theory. In *Handbook of Mathematical Economics*, volume 3, Elsevier Science Publishers, 1986.

[Shardanand and Maes, 1995] Upendra Shardanand and Pattie Maes. Social information filtering: Algorithms for automating “word of mouth.” *In Proceedings of Computer Human Interaction*, pages 210–217, May 1995.

[von Neumann and Morgenstern, 1953] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, New Jersey, 1953 (© 1944).