Who Predicts Better? – Results from an Online Study Comparing Humans and an Online Recommender System

Vinod Krishnan, Pradeep Kumar Narayanashetty, Mukesh Nathan,
Richard T. Davies, and Joseph A. Konstan
GroupLens Research
Department of Computer Science and Engineering
University of Minnesota-Twin Cities
+1-612-625-4002
{vinod,pnshetty,mukesh,rdavies,konstan}@cs.umn.edu

ABSTRACT
Algorithmic recommender systems attempt to predict which items a target user will like based on information about the user's prior preferences and the preferences of a larger community. After more than a decade of widespread use, researchers and system users still debate whether such "impersonal" recommender systems actually perform as well as human recommenders. We compare the performance of MovieLens algorithmic predictions with the recommendations made, based on the same user profiles, by active MovieLens users. We found that algorithmic collaborative filtering outperformed humans on average, though some individuals outperformed the system substantially and humans on average outperformed the system on certain prediction tasks.

Categories and Subject Descriptors
H.5.3 [**Information Systems**]: Information interfaces and presentation – *collaborative computing, computer-supported cooperative work, evaluation/methodology*

General Terms
Human Factors.

Keywords
Recommender systems, predictions, human recommenders, MovieLens, MAE, recommender evaluation.

1. INTRODUCTION
Recommender systems try to help individuals find good choices from among an overwhelming set of alternatives. Collaborative Filtering systems apply technological solutions to create the sort of social recommendations and predictions that humans have always indulged in. The earliest collaborative systems like Tapestry simply routed recommendations from one human to another [10]. Automated Collaborative Filtering (CF) [6,7,14,16] aggregated opinions of a large set of users to recommend or predict for a target user, giving more weight to the opinions of those who had previously exhibited similar tastes. This change was significant, as it removed the "explicit recommendation" and changed to a model of mining user preferences (often implicit in user behavior) to generate recommendations for others. Since then, many variations have been developed to suit diverse needs. One is the Item-based Collaborative Filtering, based on item relations rather than user relations that traditional CF algorithms follow. [Sarwar *et.al* 2001] suggest that Item-based algorithms perform better and give better quality results, in addition to reduced computational effort [15]. Yet another is the Content-boosted Collaborative Filtering discussed in [13]. This approach overcomes the problem of rating sparsity and the first-rater problem, as [Melville *et.al* 2002] and [Balabanovic *et.al* 1997] argue [12, 1].

RecSys'08, October 23–25, 2008, Lausanne, Switzerland.

2. RELATED WORK AND MOTIVATION
CF algorithms are used today in some commercial systems like Amazon.com that employs the item-based collaborative filtering to generate recommendations for its customers based on their purchase histories [8]. Chen and McLeod list other instances of deployed CF algorithms such as Launch music on Yahoo!, Cinemax.com, Moviecritic, TV Recommender, Video Guide and the suggestion box, and CDnow.com [2]. Recommender systems have become popular, in part as a way of coping with the fact that too often we lack access to expert human recommenders. We value recommendations at a site like Amazon.com because we don't believe we can often go to a bookstore and find a bookseller who knows both our tastes and enough about books to recommend well to us. (Indeed, those of us who are lucky to have such human recommenders often treasure them). At the same time, we technologists believe, at least a little, that our technology may provide a solution even better than a human recommender. After all, we argue, we are building recommendations from millions of opinions presented by tens of thousands or even millions of users, and that we can scientifically select the data that best matches the question we're asking for you. How can a mere human compete?

Well, we have some sense as to how a mere human competes. Humans may lack the total quantity of data our systems possess, but they are extremely good at processing a variety of heterogeneous data, including gestalt-like patterns. A human may *know* that a particular movie is a dud (to most people) even as a recommender system is confused since the only people who rated were the few who liked it. More to the point, a human can integrate information about popularity, genre, patterns of like/dislike, actors, and much more in ways we're still struggling to replicate with our algorithms and systems. Indeed, the work of Luis von Ahn [9] builds upon these human skills that computers have found hard to replicate.

Furthermore, Sinha *et.al* showed that friends of users consistently provided recommendations that the users preferred over those of algorithmic recommender systems, though users found recommendations generated by the algorithms to be useful and new [17]. This approach is limited in the sense that users were aware of the recommendations that their friends provided, which causes a bias in their preferences.

Thus, we feel one of the best ways to measure how well we're doing, and also to identify ways to improve algorithmic recommender systems is to compare ourselves with human recommenders. This paper reports on a study in which we ask humans to predict the scores that other individuals – people they know only through their movie taste profiles – have given to movies. In the spirit of open inquiry, we present our goals as open research questions rather than formal hypotheses that anticipate one outcome or another.

RQ1: On average, how well do humans predict movie tastes compared with a collaborative filtering recommender system?

RQ2: Are there specific types of user profiles or movies for which humans predict particularly better? Or particularly worse?

RQ3: Do humans predict better (or worse) for users that have tastes similar to them?

RQ4: What information do human predictors use? Do better predictions take more time?

3. EXPERIMENTAL DESIGN

3.1 Profile Description and Prediction Set

Subjects were presented with *profiles* of ratings from MovieLens users. MovieLens[1] is the movie recommender system maintained by GroupLens Research[2]. Fig.1 is a screenshot of a part of such a profile. The profiles contained 30 movies: - titles, years, genres and the user's ratings.[3] An 'info' link would display additional information including stars, director and system-wide rating data; one additional link led to that movie's description on IMDB[4]. Upon studying the profile, the human subjects were asked to predict ratings for 10 movies that were rated by the same user. We refer to the set of movies for which the user predicts ratings, as the *prediction set*. From what they inferred about the profile, subjects predicted ratings for the prediction set. Ratings could be given in half-star increments, starting from ½ up to 5.

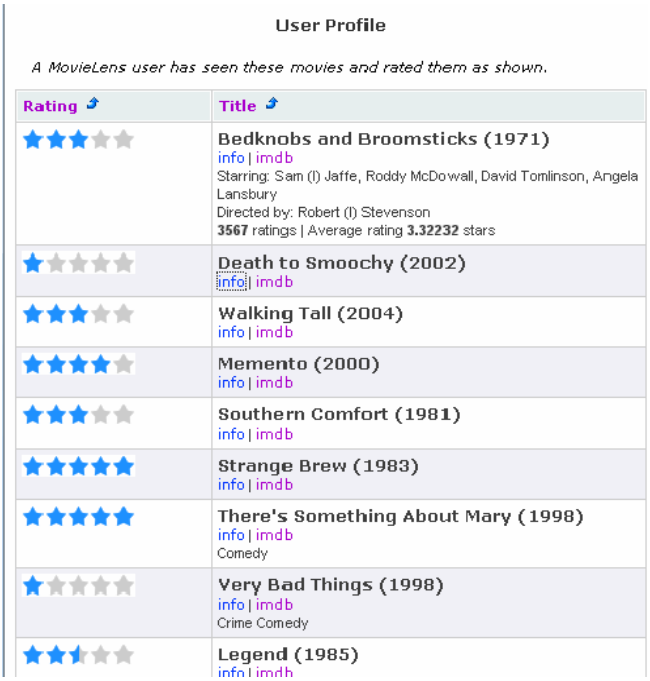

Figure 1. Snapshot of profile presented to subjects

3.2 Selecting the Profile

MovieLens has more than 120,000 registered users' profiles. Of these, only 1237 users have rated more than 1000 movies each, where each movie that each user had rated, received more than 100 ratings. We decided to draw our profiles from this set so that we could conduct later experiments that make available more ratings or explore recommendations rather than just prediction. We wanted to include in our studies a diverse set of users, in particular a mix of both users with mainstream tastes and ones with more eclectic tastes. For each of the 1237 users, we computed the proportion p of their ratings that deviated by 1.5 units from the movie's average rating. These proportions ranged from 0.0026 to 0.5390, with a mean of 0.1925. These extremes evince users at the low end who have tastes very much in line with the MovieLens mainstream, and users at the high end who rate movies quite differently from the mainstream. To obtain a sample of users who are representative of the diversity of this scale, we selected a stratified sample of users as shown in Table 1. Within each bucket, profiles were randomly selected.

Table. 1 Number of users present in each interval of proportion and number of users chosen randomly to represent their interval.

Criterion	# of qualified user profiles	# of user profiles selected
$0 \leq p < 0.1$	217	3
$0.1 \leq p < 0.2$	510	5
$0.2 \leq p < 0.3$	342	4
$0.3 \leq p < 0.4$	117	2
$0.4 \leq p$	51	1

Some movies that these users had rated did not have sufficient information (missing average rating, for example). After screening

[1] http://movielens.umn.edu/
[2] http://grouplens.org/
[3] We shall refer to the subjects as 'subjects' and the MovieLens user from whom the profile was drawn as the 'user'.
[4] (Internet Movie DataBase) http://imdb.com/

them for missing information, we randomly chose 40 movies for each user's profile and from these, randomly assigned 30 movies for the profile and the remaining 10 for the prediction set. To minimize subject burden, each subject would see a single user profile and would predict the ratings of five of the movies in the prediction set.

3.3 The User Interface

Fig. 2 shows the final profile that subjects viewed. The profile and the prediction set were placed adjacent to enable subjects to refer to the information in the profile even as they predicted. Once the subjects predicted, they clicked the "Done" button. The top right screen corner explained the rating scale with just 1 star meaning "Awful" to all 5 stars meaning "Must See". These are the same text anchors used in MovieLens. Prior to viewing this page, every subject also viewed a brief introductory page that summarized the experiment with the tone "*View-Predict-Win!*". The main study provided a monetary incentive for the best 3 performers. Profiles were randomly assigned to subjects. [Crutchfield *et.al*, 1958] showed that people are more eager to make a good impression among others when their performances were being compared to that of others or made public [4]. We believe that both the potential reward and the knowledge that they are being compared will encourage subjects to perform their best.

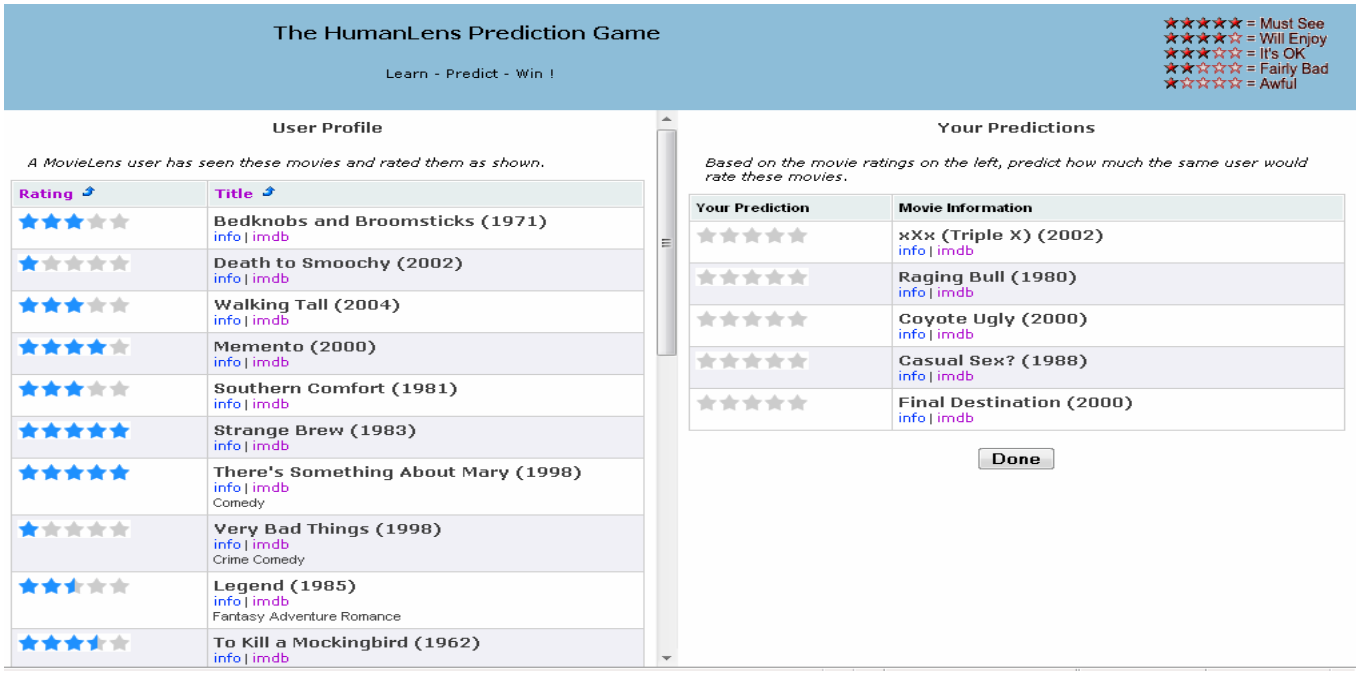

Figure 2. The user interface comprising the profile and the prediction set together in a single page.

3.4 How MovieLens Recommends

We used MovieLens itself as the algorithmic recommender system against which we compared subject predictions. MovieLens has a database with more than 120,000 user profiles, and more than 15 million ratings of over 10,000 movies. The system uses much of this information when it predicts ratings for a user-movie pair. MovieLens uses an open source collaborative filtering recommender engine, known as MultiLens[5], which supports several algorithms. In the live system, it uses the Item-Item algorithm based on the *Cosine Similarity Model* that represents similarity between items using cosine similarity between the vectors of ratings of those items by users [15]. Based on this similarity, MultiLens builds a similarity model incorporating all MovieLens users and their ratings. The MultiLens then uses the model to make predictions.

For this study, we re-built an instance of MovieLens' similarity model excluding all information about the 14 users whose ratings were used to describe the profiles that subjects saw.

3.5 Subjects

After a pilot study conducted locally, we recruited users online, focusing on those who had extensive MovieLens profiles (over 100 movies) since we feel these users are likely movie fans and have a good chance of being familiar with many of the movies in the profiles. (Also, these users would have ratings data that would allow us to explore the relationship between their tastes and their performance.) Subjects were recruited by e-mail, and only subjects who provided us with an e-mail address and agreed to receive invitations to research studies were contacted. In total, we emailed 200 MovieLens members and 50 participated in the study. We do not require MovieLens members to provide demographic information, so we do not know the gender balance of the pool. We do know that participating subjects had rated an average of 1450 movies during their membership period with MovieLens.

3.6 Survey and Logging

After predicting, subjects took a short online survey where they described their experience about the task. Attributes that the survey captured include:

- Sufficiency of information in the profile and prediction set
- Information they primarily relied to predict
- Perceived uniqueness of the profile

[5] http://www.cs.luther.edu/~bmiller/dynahome.php?page=multilens

- How often they used movie information
- Similarity between profile's tastes and theirs
- Confidence in their predictions

Two questions in the survey were re-phrased after the pilot to disambiguate and clarify its meaning. This was done to ensure that the questions reflect what they intend to measure.

4. RESULTS

4.1 Comparing predictions

4.1.1 Overall comparisons

To answer RQ1, we used the Mean Absolute Error (MAE) accuracy metric to compare predictions of MovieLens and the subjects [18]. The MAE is a widely used accuracy metric that computes the mean absolute deviation of a data set from a reference number and averages it across the set. The overall MAE for all the subjects was 0.985, while MovieLens scored better with a score of 0.797 ($t = 9.12$, MovieLens' overall MAE < Subjects' overall MAE; 2-sided p-value << 0.0005). Fig. 3 presents a summary comparison of the MAEs of MovieLens and the subjects. Each point on the x-axis represents a specific subject, showing both the MAE of that subject's predictions and the MAE of MovieLens on the same profile predictions (note that the 14 profiles were randomly assigned to the users, and that MovieLens therefore repeats each profile as different users are assigned to it).

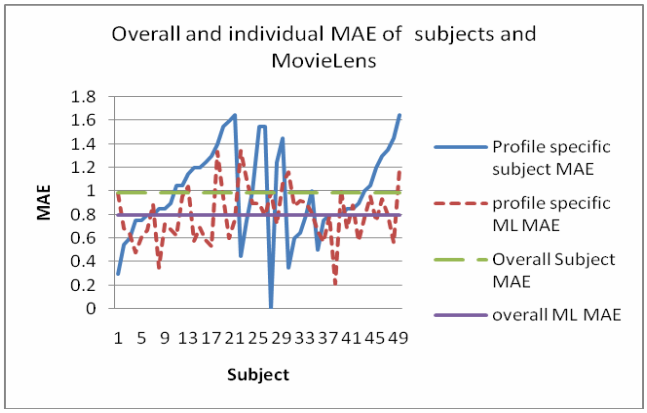

Figure 3. Overall and individual MAE of subjects and MovieLens.

Ordered by the subject MAEs, Fig. 4 shows that MovieLens MAE is largely uncorrelated with subject MAE, and also how the subjects had a wider range of MAEs (both better and worse performance).

Fig. 5 shows the variation of MAEs averaged over profiles ordered by MovieLens' MAE. Interestingly, MovieLens outperformed the average human predictor on only 6 out of 14 profiles, thought it substantially outperformed subjects overall. MovieLens also had a much wider range of prediction quality, suggesting that a small group of humans might be an approach for developing a more effective human recommender. While MovieLens performed badly for certain profiles and performed well for certain other profiles, subject clusters had lesser variations of this kind.

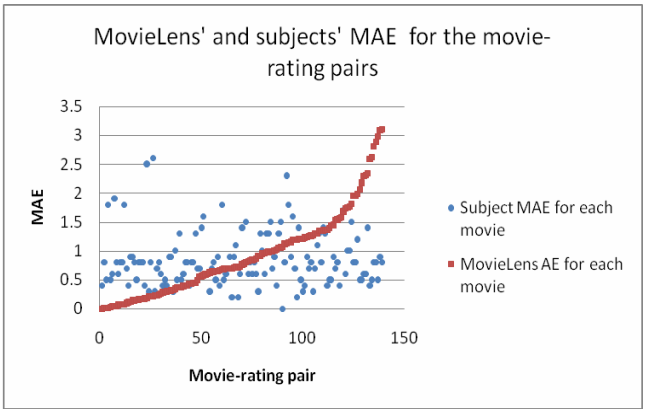

Figure 4. MovieLens AE vs. Subjects' MAE for the 140 movie-rating pairs.

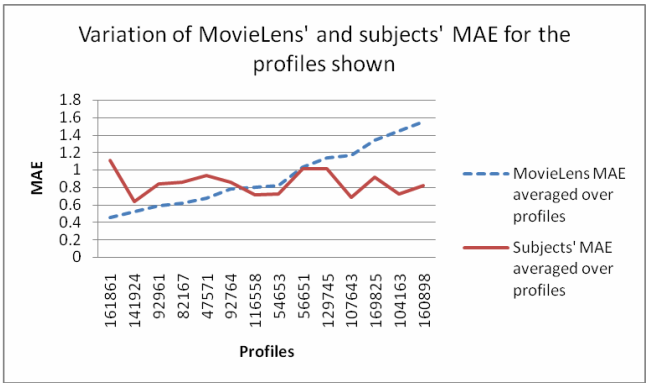

Figure 5. MAEs of Subjects and MovieLens averaged over profiles.

4.1.2 Comparisons based on profiles

Fig. 6 shows the proportion specific variation of MAE of subjects and MovieLens. To observe how MovieLens and the subjects predicted for mainstream and unusual profiles, we labeled users with $0 \leq p < 0.3$ as those with mainstream tastes and those with $0.3 \leq p$ as those with unusual tastes. Having more than 30% of one's ratings deviate by more than 1.5 units on a 5-point scale, we argue is a reasonably eclectic profile. Subjects performed better for eclectic profiles than mainstream profiles, while MovieLens' behavior suggests that it performs better for mainstream profiles than eclectic profiles.

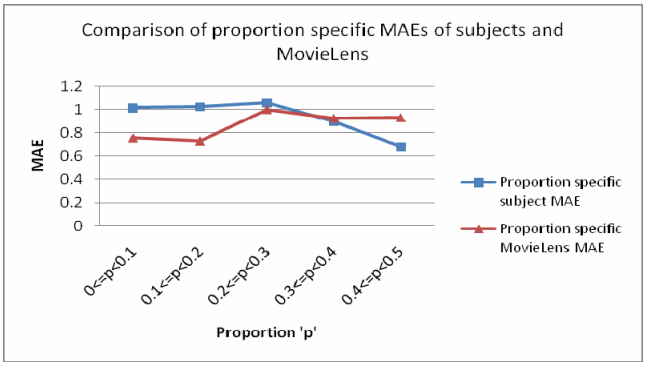

Figure 6. Proportion specific variation of MAEs of subjects and MovieLens

4.2 Factors affecting accuracy

4.2.1 Impact of similarity

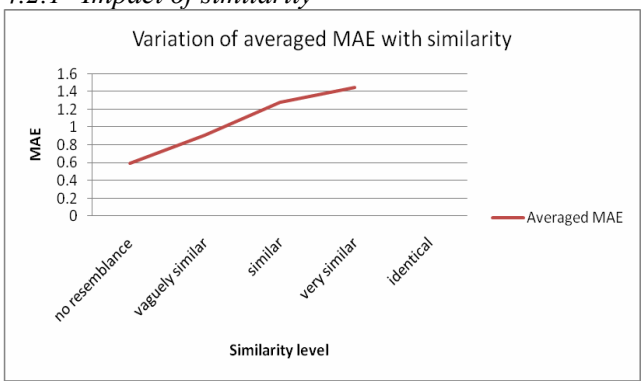

Figure 7. Subjects' MAEs averaged over individual similarity scores.

In the survey, subjects responded when asked about similarity of the profile to either themselves or to someone they know, on a 5-level scale – *no resemblance, vaguely similar, similar, very similar* and *identical*. We studied the relationship between subjects' MAE and the similarity measure they provided us with. We then aggregated all subject MAEs who responded with the same similarity measure. Much to our surprise, we found that subjects' performance worsened with increasing similarity. Fig. 7 demonstrates this observation.

4.2.2 Impact of time

We also studied the variation of MAE with the time spent on a profile. We were unable to detect any definitive trend that governed the relationship between the two factors. There was very weak correlation between time and MAE. This sharply contrasts the behavior of a recommender system whose accuracy steadily improves with time as it gains more information about users' preferences. With more than a million ratings at its disposal, MovieLens would perform more accurately with more opinions.

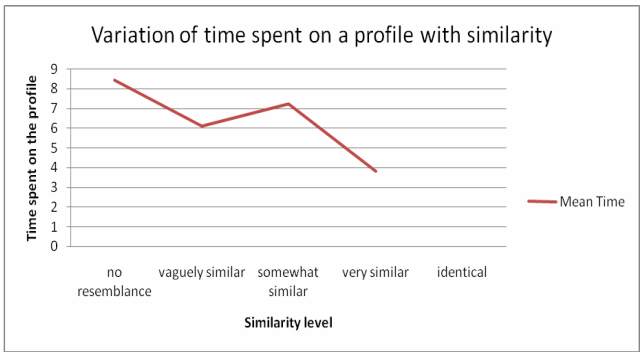

Figure 8. The mean time spent on profiles by subjects ordered by similarity.

Also, subjects on an average spent more time with dissimilar profiles and lesser time with similar ones. Fig. 8 demonstrates this trend. In an attempt to explain the three factors, subjects spent more time with dissimilar profiles and scored a better MAE. Subjects also spent less time with similar profiles, but scored poorly. From our previous observations, subjects perform better with dissimilar and eclectic profiles than, while they do not score well with similar and mainstream profiles.

4.2.3 Impact of confidence

Subjects were asked to respond about their confidence in their predictions. A clear trend visible was that the majority of the subjects (56%) responded 4 on a 5-level scale. Table 2 lists the responses of subjects when asked about their overall confidence in their predictions. In addition, they also responded to their confidence in each prediction they made.

Table 2. Confidence among subjects.

Confidence	Count
Strongly Agree	1 (2%)
Somewhat Agree	28 (56%)
Neutral	13 (26%)
Somewhat disagree	8 (16%)
Strongly Disagree	0

Fig. 9 indicates that subjects' confidence levels were lower for unusual profiles than for mainstream profiles. This is interesting because subjects performed better for unusual profiles (Fig. 8) but were less confident about it, and ironically were more confident where they did not perform well.

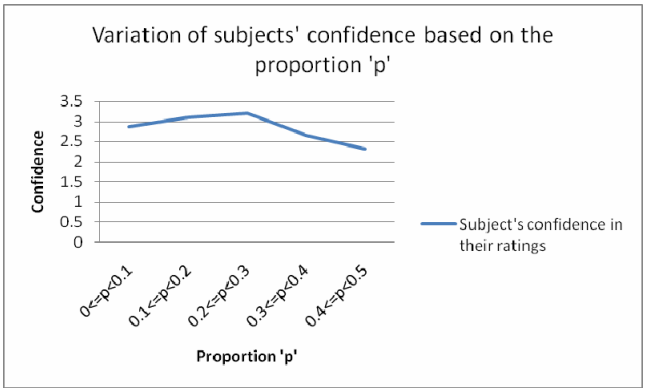

Figure 9. Mean confidence of subjects ordered by proportion

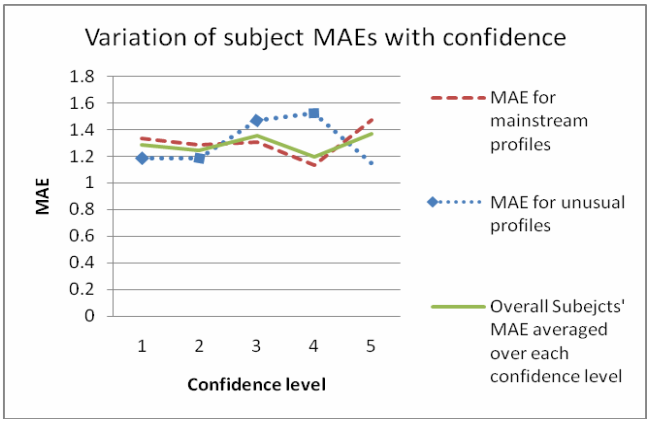

Figure 10. Subjects' variation of MAE with confidence for mainstream and unusual profiles.

To further our analysis, we also looked at how confidence and MAE are related. We averaged the MAEs for each confidence level for mainstream and unusual profiles. Fig.10 suggests that

subjects' confidence showed lesser variance for mainstream profiles, compared to unusual profiles. The trend is decisively strong to assert a relationship between the two. Also, the overall confidence trend of subjects closely resembles their trend in mainstream profiles, reiterating that they were more confident of their ratings for mainstream profiles.

4.3 How people go about predicting

Subjects responded when asked about how often they used movie information (genre, cast, director, year of release, number of ratings, average rating) in the profile. They responded on a 5-level scale – *never, rarely, moderately, frequently* and *always*.

4.3.1 What information subjects used to predict

Subjects were allowed to select multiple choices to indicate what information they used. Those who relied on other information were asked to explain their choices. Their responses had common elements like:

- "*Combination of critical acclaim and my opinion of the movie"*
- *"My own knowledge of movie content and how much the profile liked it"*
- *"Ratings of similar movies in the profile"*
- *"IMDB's average rating and my own knowledge of how well the film was received upon its release"*
- *"I work at a video store!"*
- *"I looked at how the profile rated similar movies"*

As an interesting addition, in our pilot study, a subject looked at photos of the films' premiers, posters and trailers to recognize the actors and 'get the feel of the movies' as she put it. With 27 (54%) responding "Yes" and the remaining 23 responding "No", subjects in the online study were divided when asked if it ever occurred to them to match the profile's preferences with either their own or that of someone whom they know. These responses collectively indicate that subjects use a combination of descriptive data and similarity to understand a profile's preferences to predict.

Table 3. Preferred information that subjects used to predict.

Information used	Count
Extreme ratings in the profile (1 or 5 stars)	33 (66%)
Movie information (cast, director, genre, year of release, plotline from IMDB)	23 (46%)
Average rating, number of ratings the movie received in MovieLens	18 (36%)
Others	18 (36%)

4.3.2 What subjects felt could have helped them predict better

Subjects suggested a variety of other information that they felt could have helped them understand the profile better and predict better, as listed in Table 4. These results reinforced our initial observations from the pilot with a more elaborate distribution of subject choices. In the pilot, nearly two-thirds of the subjects felt that sufficient information was not present, while the online study was again divided in opinion with 28 (53.8%) indicating that there was a dearth of content, while the remaining 24 (46.2%) indicated otherwise. Interestingly, despite the perceived lack of information, 13 subjects predicted with equal or better accuracy (by the MAE metric) than MovieLens.

Table 4. Content that subjects would like to see in a profile that they felt could have helped them predict better.

Content that subjects felt could have helped them better	Count
Age and/or gender of the profile	35 (70%)
More genres and/or more movies in a given genre	30 (60%)
Viewer comments about the movie	26 (52%)
Snippets/summary of movie's plot	20 (40%)
Tags used by MovieLens users to label the movies	12 (26%)
Awards that the movies won	4 (8%)
Others	4 (8%)

4.3.3 *What they inferred about the profile*

Movies were chosen from the mainstream and eclectic choices of the profile's preferences. Subjects provided qualitative responses on what they inferred about the profile that helped them predict. They inferred the profile's preferences in genres, and tended to compare the profile's preferences with their own, be it mainstream or eclectic choices. Subjects also responded observing the variability in preferences in the profile they viewed, highlighting in some cases that profiles rated similar movies differently. In several cases, subjects also remarked whether the profile they viewed had mainstream or eclectic choices.

5. DISCUSSIONS

5.1 Recap of results

MovieLens scored a better overall MAE than subjects, but there were 13 subjects who had a better individual MAE than MovieLens. Subjects had lesser variation in their MAEs for the profiles used in the study. However, they performed better for profiles with high number of deviant ratings than mainstream profiles, antithetic to MovieLens' behavior.

Surprisingly, subjects performed poorly with similar profiles, spending less time on them and performed well with dissimilar profiles spending more time on them. In relation to this, subjects were less confident where they performed well, owing to the dissimilarity with eclectic profiles, while they were more confident where they performed poorly with similar profiles that were mainstream profiles.

Subjects used more content information, and asked for more qualitative content like gender of the user represented by the profile, and summaries of movies, reinforcing that humans value descriptive content.

5.2 Limitations of this study

This study is small in the sense that we are not trying to look for statistical significance on our results to make definitive remarks on the capabilities of humans or CF algorithms. Also, in an attempt to make this a 'fair fight', an inherent inequality in the information presented to humans and the recommender system could have crept in. Subjects had more movie information, while MovieLens had more rating information. Another limitation is the interface – it has not been optimized for a prediction task.

6. CONCLUSIONS

This study compares how humans fare against recommender systems in predicting for a user, given his or her preferences. Though there were special cases who scored a better MAE than MovieLens, the latter is still better overall. We acknowledge that people are highly variable, but interpreting how people predict, especially the efficient ones who are systematically better, is valuable. Another implication is that Collaborative Filtering is stable and evincing this against humans only adds to its credibility. However, MovieLens did not perform its best for eclectic profiles, which now sparks the question of how it can made to predict better in the extremities of its users' preferences.

6.1 Future Work

We view this as the first part of stream of work looking at how humans perform recommendations. Next we hope to look specifically at the problem of recommendation rather than prediction. Comparing such recommendations is more complex, since the conventional methods that look at whether the recommended item is already rated are heavily biased against novel but useful recommendations. We expect to use some human evaluation – invite some users to evaluate subjectively how good the recommendation list is for them. We hope to learn something about where humans outperform current algorithms both to help us improve the algorithms and to help us ascertain where it may be worth involving humans more directly in the recommendation process.

Indeed, comparing recommendations are more complex since there are no standard metrics to compare recommendations from 2 sources. [McNee *et.al* 2006] and [Herlocker *et.al* 2004] discuss newer and more relevant metrics to compare recommendations from different sources [12, 5]. Some recent and widely accepted metrics include trust [3], diversity [19] and serendipity [11]. All of these will affect the manner in which humans will evaluate recommendations and how recommender systems are calibrated to respond to these metrics. This is of interest to the entire recommender systems community since the results will give new insights into why and where recommender systems lag and how the systems can be tweaked to personalize their recommendations to fit users' preferences.

7. ACKNOWLEDGEMENTS

We appreciate that the idea for this project came out of a panel session at the ACM Recommender Systems Conference. We thank the participants who stimulated this research. We would like to thank our research subjects and would like to thank other members of GroupLens Research for their support throughout the project. This work was supported in part by grant IIS – 0324851 from the National Science Foundation.

8. REFERENCES

[1] Balabanovic, M., and Shoham, Y. 1997. Content-Based Collaborative Recommendation. Communications of the ACM, 40(3).

[2] Chen, A. and McLeod, D. 2006. Collaborative Filtering for Information Recommender Systems. In Encyclopedia of E-Commerce, E-Government and Mobile Commerce. Khosrow-Pour, M. Ed. Information Science Reference.

[3] Chen, L. and Pu, P. Trust building in recommender agents. 2007. In the Proceedings of the IEEE 3rd International Workshop on Web Personalization, Recommender Systems and Intelligent User Interfaces (Istanbul, Turkey, April 16-20, 2007), WPRSIUI'07

[4] Crutchfield, R. S., Woodworth, D. G., & Albrecht, R. E.

Perceptual performance and The Effective Person WADC-TN-58-60). Lackland Air Force Base, Texas: Wright Air Development Center, 1958.

[5] Herlocker ,J., Konstan , J., Terveen , L., Riedl, J. 2004. Evaluating Collaborative Filtering Recommender Systems. ACM Transactions on Information Systems, v.22 n.1, (January 2004) 5-53. DOI= http://doi.acm.org/10.1145/963770.963772

[6] Hill,W.,Stead, L.,Rosenstein,M.,and Funas, G. 1995. Recommending and Evaluating Choices in a Virtual Community of Use. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ((Denver, CO, USA, April 14-18, 1995) CHI '95. ACM Press, New York, NY. 194-201. DOI= http://doi.acm.org/10.1145/223904.223929

[7] Konstan, J., Miller, B., Maltz,D., Herlocker, J.,Gordon, L., and Riedl , J. 1997. GroupLens: Applying Collaborative Filtering to Usenet News. Communications of the ACM, 40(3), pp.77-87

[8] Linden, G., Smith, B., and York, J.2003. Amazon.com Recommendations. IEEE Internet Computing 7 no.1, (January-February 2003), 76-80

[9] Luis von Ahn. 2007. Invited Talk: Human Computation. In the Proceedings of ACM Conference on Knowledge Capture and Constraint Programming. (Whistler, Canada, October 28-31, 2007). K-CAP'07. ACM Press, New York, NY. 5-6. DOI=http://doi.acm.org/10.1145/1298406.1298408

[10] Maltz, D., Ehrlich, K. 1995. Pointing the way: Active Collaborative Filtering. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems. (Denver, CO, USA, April 14-18). Chi'95. ACM Press New York, NY. 202-209

[11] McNee, S., Riedl, J., Konstan, J. 2006. Being Accurate is Not Enough: How Accuracy Metrics have hurt Recommender Systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (Montreal, Canada, April 22-27, 2006) CHI'06. ACM Press, New York, NY, 1097-1101. DOI=http://doi.acm.org/10.1145/1125451.1125659

[12] Mcnee, S., Riedl, J., Konstan, J. 2006. Making Recommendations Better: An Analytical Model for Human-Recommender Interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (Montreal, Canada, April 5-10, 2008). CHI'08. ACM Press,

New York, NY. 1103-1108. DOI= http://doi.acm.org/10.1145/1125451.1125660

[13] Melville, P., Mooney, L.R, and Nagarajan, R. 2002. Content-Boosted Collaborative Filtering for Improved Recommendations. In Proceedings of the SIGAI Conference of American Association for Artificial Intelligence. (Edmonton, Canada, July 28-August 1, 2002). AAAI'02. ACM Press, New York, NY. 187-192.

[14] Resnick, P., Iacovoum, N., Suchak, M., Bergstrom, P., and Riedl, J. 1994. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In Proceedings of the ACM Conference on Computer Supported Cooperative Work (Chapel Hill, NC, USA ,October 22-26, 1994) CSCW'94. ACM Press New York, NY. 175-186. DOI= http://doi.acm.org/10.1145/192844.192905

[15] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. 2001. Item-Based Collaborative Filtering Recommendation Algorithms. In Proceedings of the ACM Conference on World Wide Web. (HongKong, May 1-5, 2001) WWW'01. ACM Press, New York, NY. 285-295. DOI=http://doi.acm.org/10.1145/371920.372071

[16] Shardanand, U., and Maes, P. 1995. Social Information Filtering: Algorithms for Automating the 'Word of Mouth'. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (Denver, CO, USA, April 14-18, 1995) CHI '95. ACM Press, New York, NY. 194-201. DOI= http://doi.acm.org/10.1145/223904.223931

[17] Sinha, R. and Swearingen, K. 2001. Comparing Recommendations made by Online Systems and Friends. In Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries (Dublin, Ireland, June 18-20, 2001).

[18] Vozalis, E., and Margaritis, K.G. 2003. Analysis of Recommender System' Algorithms. In proceedings of HERCMA '03. Athens, Greece

[19] Ziegler, C., McNee, S.M., Konstan, J.A. and Lausen, G. 2005. Improving recommendation lists through topic diversification. In Proceedings of The 14th International Conference on World Wide Web. (Chiba, Japan, May 10-14, 2005). WWW'05. . ACM Press, New York, NY. 1103-1108. DOI=http://doi.acm.org/10.1145/1060745.1060754