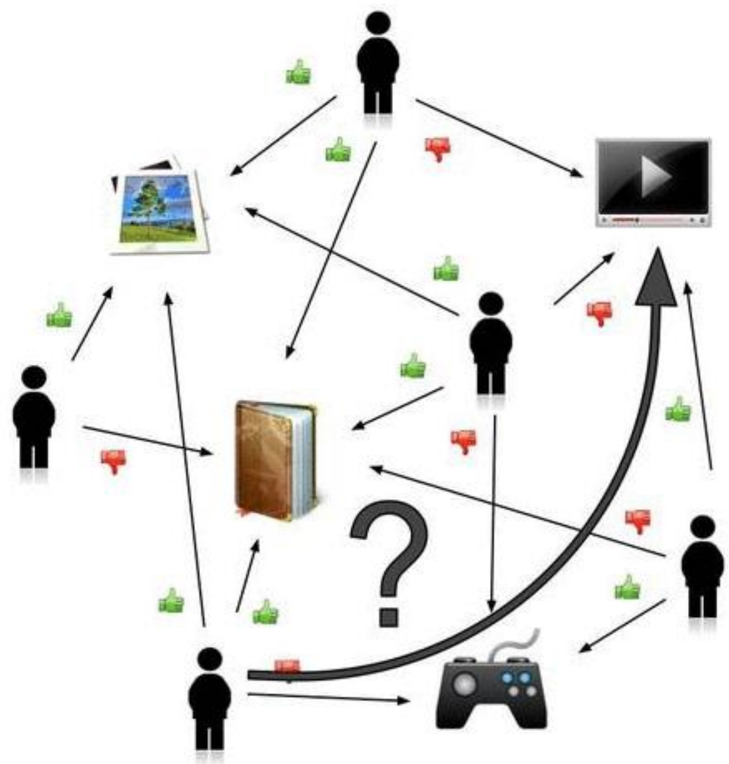
Explaining Collaborative Filtering Recommendations

김연아

2021.06.02

Automated Collaborative Filtering

- **Automated collaborative filtering systems**
  - Predict a person's affinity for items or information by connecting that person's recorded interests with the recorded interests of a community of people and sharing ratings between **like-minded persons**



Sources of Error

1. Model/Process Errors

Does not match the user's requirements

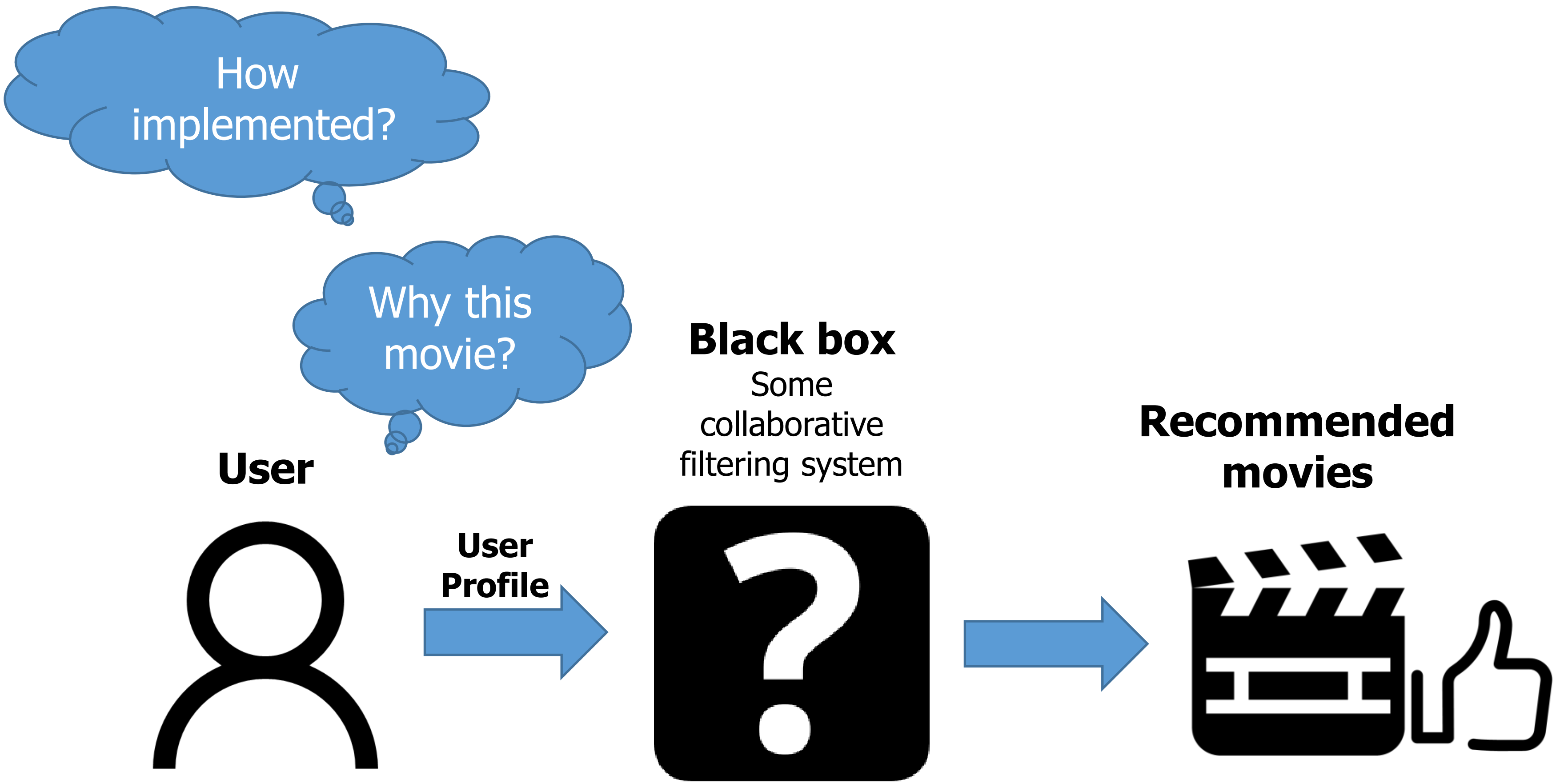
2. Data Errors

Inadequacies of the data used in the computation of recommendations

- Not enough data
- Poor or bad data
- High variance data

→ Need explanations to handle errors that come with a recommendation!

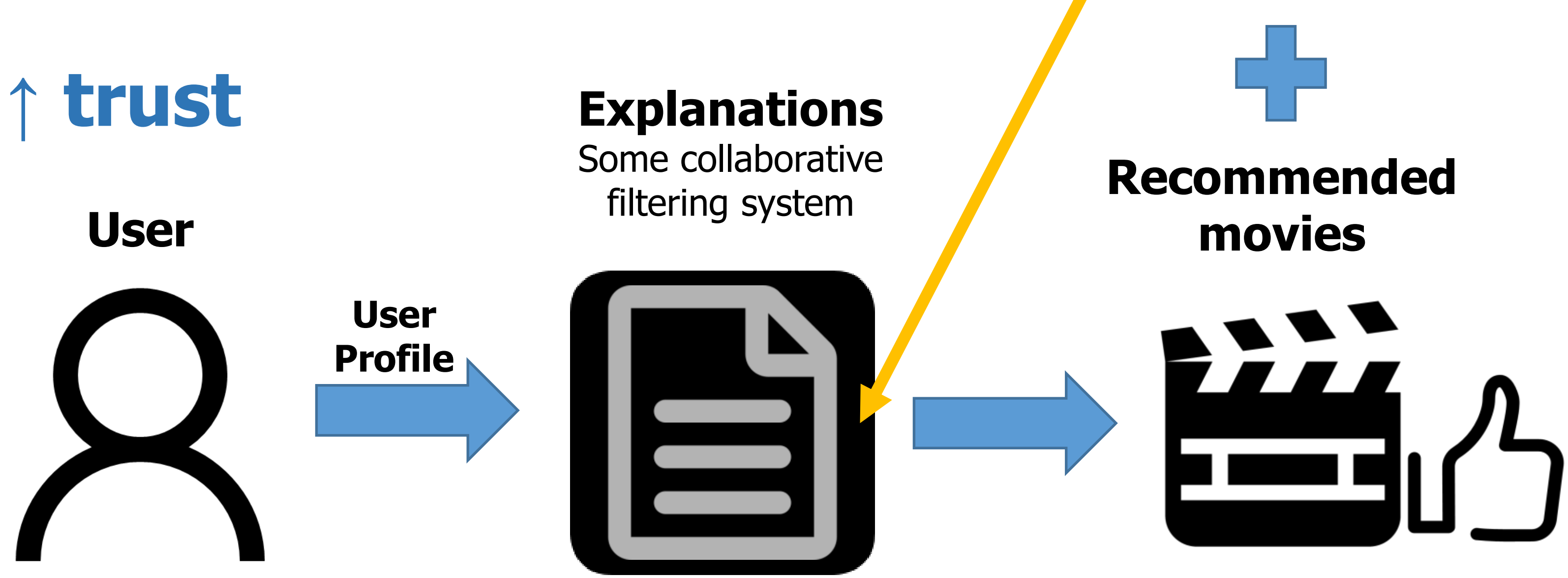
Can you trust?



Can you trust?

Benefits of Explanations: transparency

- Justification
- User involvement
- Education
- Acceptance



Research questions

- **About the use of explanations with Automated Collaborative Filtering (ACF) systems**

1. What **models and techniques** are effective in supporting explanation in an ACF system?

2. Can explanation facilities **increase the acceptance** of automated collaborative filtering systems?

3. Can explanation facilities **increase the filtering performance** of ACF system users?

Building A Model of Explanations

- **White Box Conceptual Model**

Two components that we need to explain → **the process & the data**

(1) User enters profile of ratings

- knowing how their actions have effected their recommendations
- What kinds of preference information were used in a given explanation

(2) ACF system locates people with similar profiles (neighbors)

- Assure that the system has identified the correct set of neighbors for the user's current context of need

(3) Neighbors' ratings are combined to form recommendations

- Many of the symptoms of weak predictions can be discovered with good explanations
- Knowing exactly how each of user's neighbors rated the item being explained
- Knowing the distribution of ratings for the item being recommended

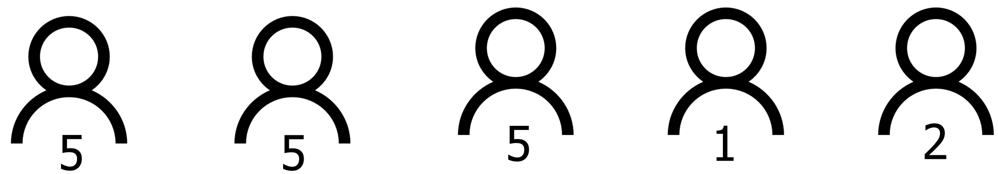
Building A Model of Explanations

- **White Box Conceptual Model**

1. Oh new movie, not well known, or has bad publicity

Explanations

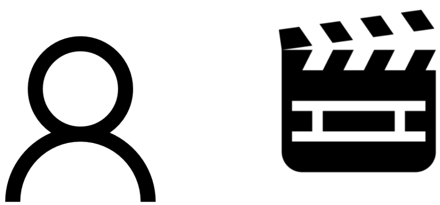
Only 5 neighbors



Share minor interests

Share popular interest

User



2. Oh! Movie is probably **not-well-known**

3. Can **trust** the system

(3) Neighbors' ratings are combined to form recommendations

- Many of the symptoms of weak predictions can be discovered with good explanations
- Knowing exactly how each of user's neighbors rated the item being explained
- Knowing the **distribution of ratings** for the item being recommended

Building A Model of Explanations

- **Black Box Model**
  - Do not wish to reveal the process!

  → instead, provide some sort of justification to trust the recommendation

  - Use the past overall performance of the recommender as justification

  → Ex) our system has 80% accuracy !

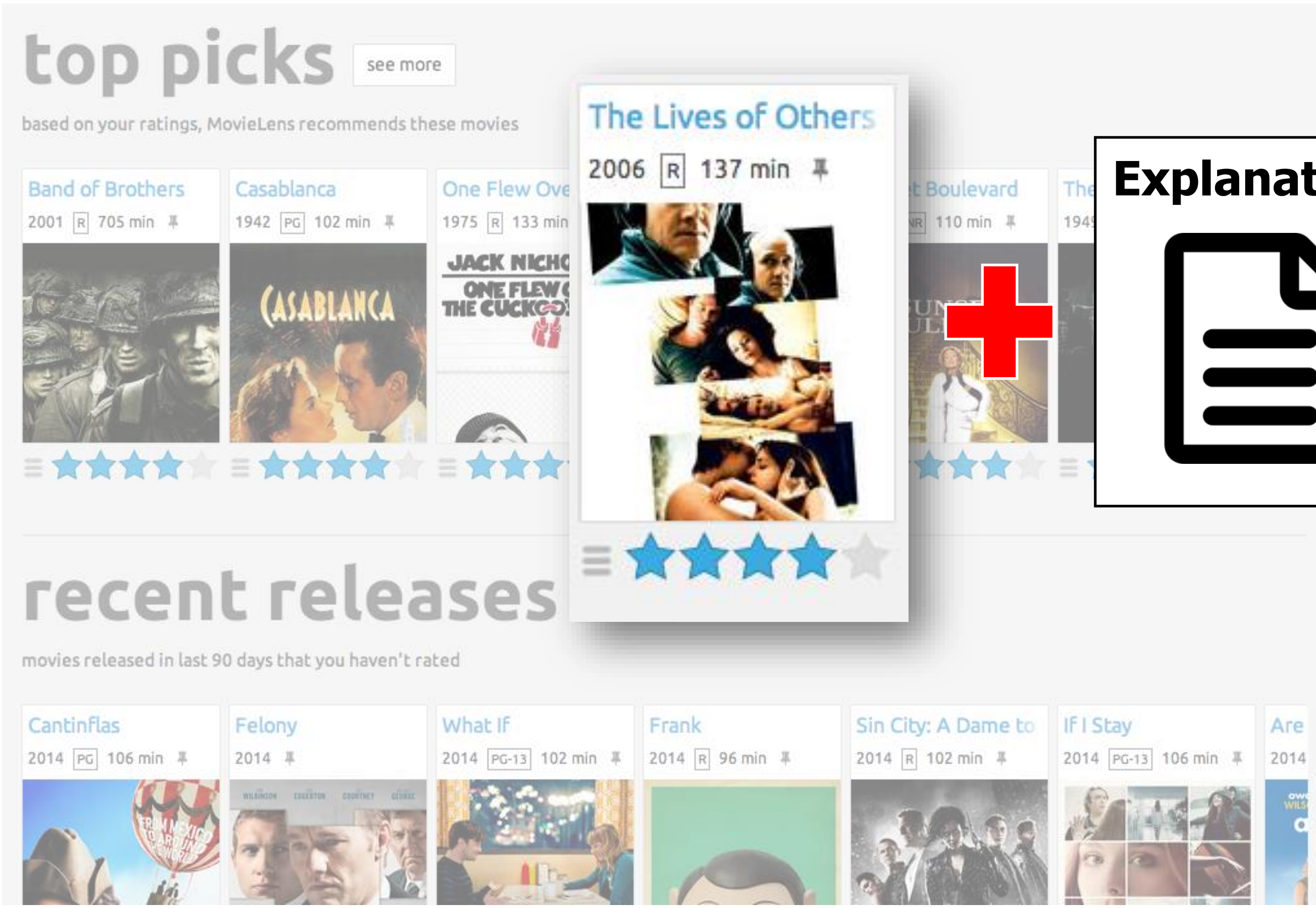
- **Misinformed Conceptual Models**
  - Some users will form incorrect conceptual models of the ACF systems that they are using to filter information

  Examples

  - users making decisions based on content characteristics
  - Too complex computational model that is hard to explain

Experiment 1 – Investigating the model

Movielens.org



Explanations

Explain
what?
How?

Experiment 1 – Investigating the model

- **Let's design an explanation interface to an ACF system**
  - What exactly do we explain and in what manner? (research question 1)

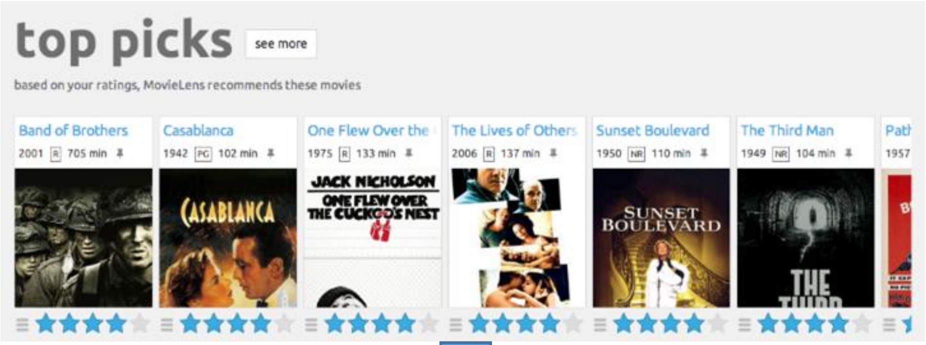
- **What is a successful explanation interface?**
  - Convince users to **purchase** a recommended product!
  - Help users to **identify** predictions that have weak justification!
  - Make users willing to use ACF system!

- **Let's figure it out with this experiment**

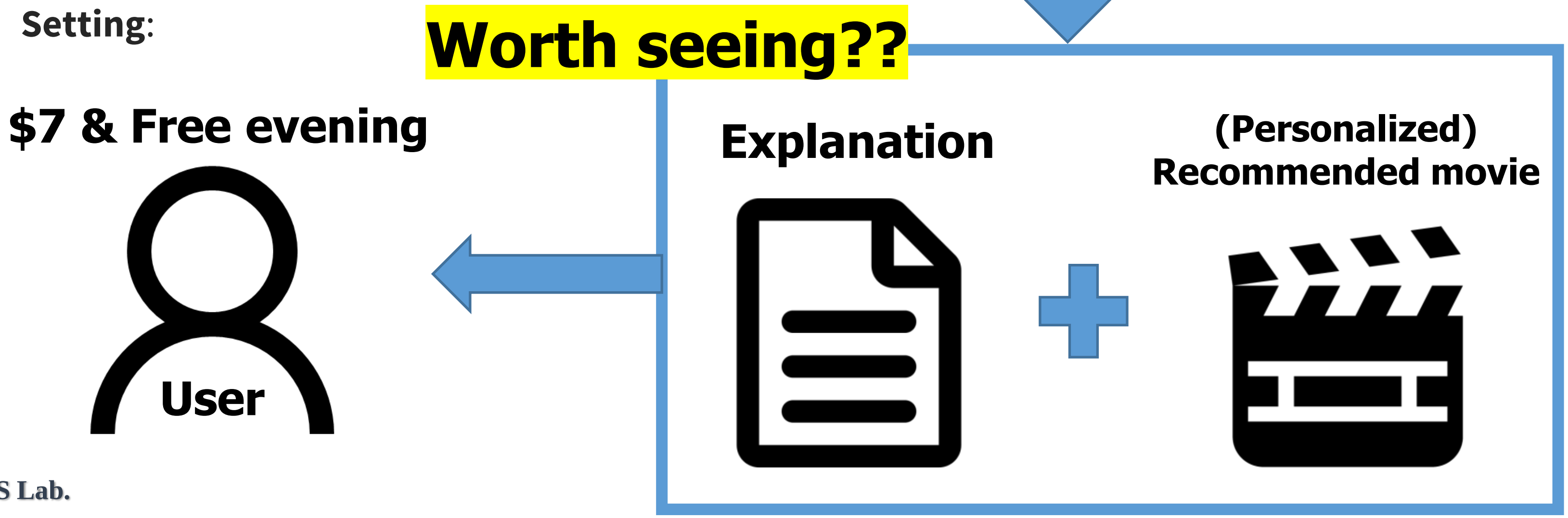
  → **measure how users of an ACF system respond to different explanations**

Experiment 1 – Investigating the model

- **Design**
  - **Survey**!
  - **Test subjects:** volunteer users of the **<u>MovieLens</u>** web-based movie recommender



  - **Setting**:



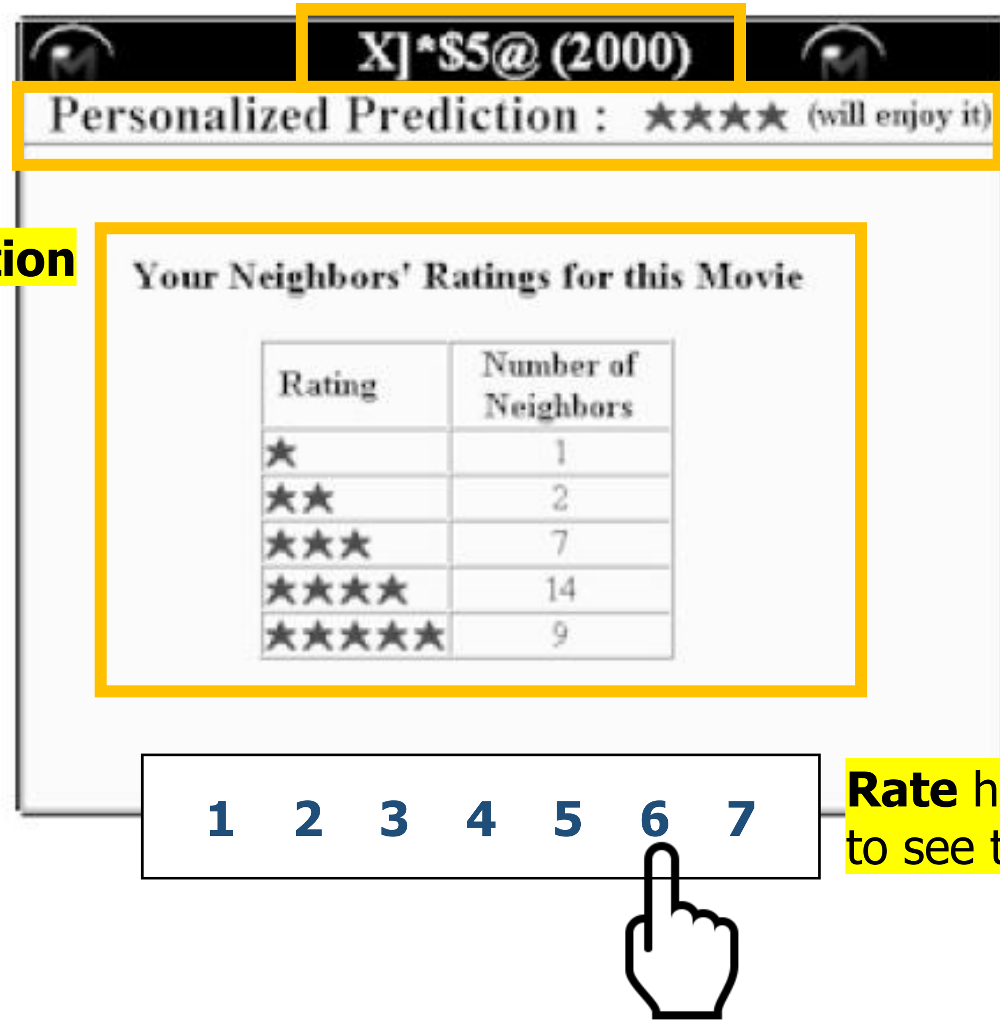
Experiment 1 – Investigating the model

- **21 explanation components for SAME movie**

Encoded title

Predicted rating

Explanation



Rating	Number of Neighbors
★	1
★★	2
★★★	7
★★★★	14
★★★★★	9

Rate how likely you would go to see the movie

Experiment 1 – Investigating the model

- **Results**

Good

#		N	Mean Response	Std Dev
1	Histogram with grouping	76	5.25	1.29
2	Past performance	77	5.19	1.16
3	Neighbor ratings histogram	78	5.09	1.22
4	Table of neighbors ratings	78	4.97	1.29
5	Similarity to other movies rated	77	4.97	1.50
6	Favorite actor or actress	76	4.92	1.73
7	MovieLens percent confidence in prediction	77	4.71	1.02
8	Won awards	76	4.67	1.49
9	Detailed process description	77	4.64	1.40
10	# neighbors	75	4.60	1.29
11	No extra data – focus on system	75	4.53	1.20
12	No extra data – focus on users	78	4.51	1.35
13	MovieLens confidence in prediction	77	4.51	1.20
14	Good profile	77	4.45	1.53
15	Overall percent rated 4+	75	4.37	1.26
16	Complex graph: count, ratings, similarity	74	4.36	1.47
17	Recommended by movie critics	76	4.21	1.47
18	Rating and %agreement of closest neighbor	77	4.21	1.20
19	# neighbors with std. deviation	78	4.19	1.45
20	# neighbors with avg correlation	76	4.08	1.46
21	Overall average rating	77	3.94	1.22

Base case

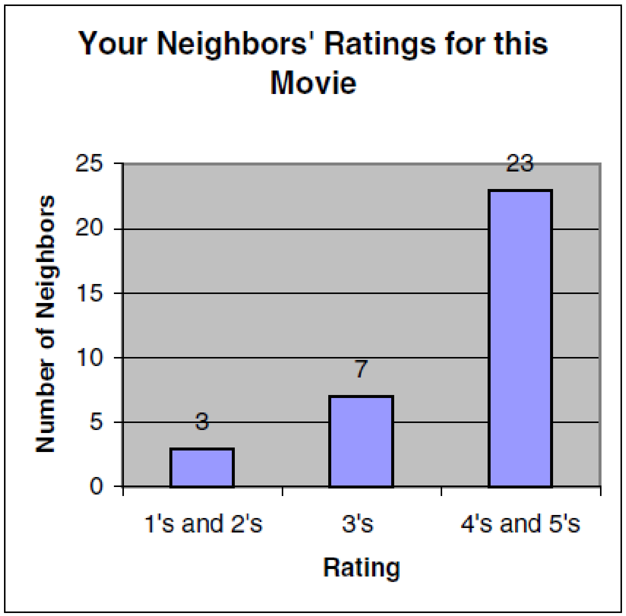
Negatively contributing to the acceptance of the recommendation

= poorly designed explanations **decreases** the effectiveness of a recommender system

Experiment 1 – Investigating the model

- **Analysis**

Binary comparison = Simple = better



Complex

#		N	Mean Response	Std Dev
1	Histogram with grouping	76	5.25	1.29
2	Past performance	77	5.19	1.16
3	Neighbor ratings histogram	78	5.09	1.22
4	Table of neighbors ratings	78	4.97	1.29
5	Similarity to other movies rated	77	4.97	1.50
6	Favorite actor or actress	76	4.92	1.73
7	MovieLens percent confidence in prediction	77	4.71	1.02
8	Won awards	76	4.67	1.49
9	Detailed process description	77	4.64	1.40
10	# neighbors	75	4.60	1.29
11	No extra data – focus on system	75	4.53	1.20
12	No extra data – focus on users	78	4.51	1.35
13	MovieLens confidence in prediction	77	4.51	1.20
14	Good profile	77	4.45	1.53
15	Overall percent rated 4+	75	4.37	1.26
16	Complex graph: count, ratings, similarity	74	4.36	1.47
17	Recommended by movie critics	76	4.21	1.47
18	Rating and %agreement of closest neighbor	77	4.21	1.20
19	# neighbors with std. deviation	78	4.19	1.45
20	# neighbors with avg correlation	76	4.08	1.46
21	Overall average rating	77	3.94	1.22

Experiment 1 – Investigating the model

- **Analysis**

Past performance

"MovieLens has predicted ***correctly for you 80%*** *of the time in the past"*

#		N	Mean Response	Std Dev
1	Histogram with grouping	76	5.25	1.29
2	Past performance	77	5.19	1.16
3	Neighbor ratings histogram	78	5.09	1.22
4	Table of neighbors ratings	78	4.97	1.29
5	Similarity to other movies rated	77	4.97	1.50
6	Favorite actor or actress	76	4.92	1.73
7	MovieLens percent confidence in prediction	77	4.71	1.02
8	Won awards	76	4.67	1.49
9	Detailed process description	77	4.64	1.40
10	# neighbors	75	4.60	1.29
11	No extra data – focus on system	75	4.53	1.20
12	No extra data – focus on users	78	4.51	1.35
13	MovieLens confidence in prediction	77	4.51	1.20
14	Good profile	77	4.45	1.53
15	Overall percent rated 4+	75	4.37	1.26
16	Complex graph: count, ratings, similarity	74	4.36	1.47
17	Recommended by movie critics	76	4.21	1.47
18	Rating and %agreement of closest neighbor	77	4.21	1.20
19	# neighbors with std. deviation	78	4.19	1.45
20	# neighbors with avg correlation	76	4.08	1.46
21	Overall average rating	77	3.94	1.22

Experiment 1 – Investigating the model

- **Analysis**

Neighbor ratings

#		N	Mean Response	Std Dev
1	Histogram with grouping	76	5.25	1.29
2	Past performance	77	5.19	1.16
3	Neighbor ratings histogram	78	5.09	1.22
4	Table of neighbors ratings	78	4.97	1.29
5	Similarity to other movies rated	77	4.97	1.50
6	Favorite actor or actress	76	4.92	1.73
7	MovieLens percent confidence in prediction	77	4.71	1.02
8	Won awards	76	4.67	1.49
9	Detailed process description	77	4.64	1.40
10	# neighbors	75	4.60	1.29
11	No extra data – focus on system	75	4.53	1.20
12	No extra data – focus on users	78	4.51	1.35
13	MovieLens confidence in prediction	77	4.51	1.20
14	Good profile	77	4.45	1.53
15	Overall percent rated 4+	75	4.37	1.26
16	Complex graph: count, ratings, similarity	74	4.36	1.47
17	Recommended by movie critics	76	4.21	1.47
18	Rating and %agreement of closest neighbor	77	4.21	1.20
19	# neighbors with std. deviation	78	4.19	1.45
20	# neighbors with avg correlation	76	4.08	1.46
21	Overall average rating	77	3.94	1.22

Experiment 2 – Acceptance and Performance

- **With these explanations (ex. Experiment 1)…**
  - Improve acceptance of ACF systems? (research question 2)
  - Improve the filtering performance of users? (research question 3)

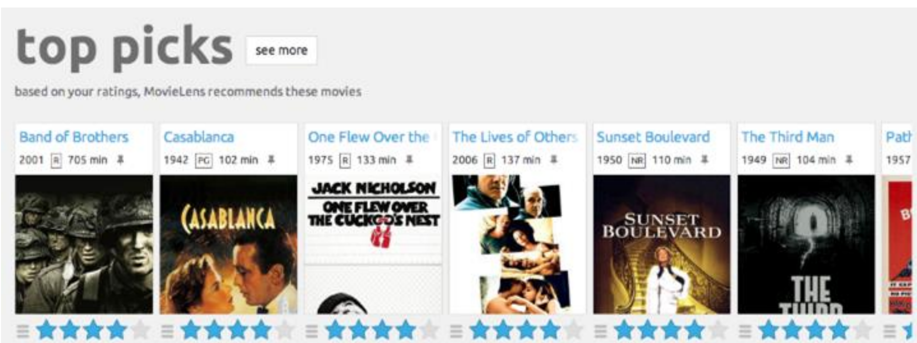
- **Hypotheses**

  Hypothesis 1: adding explanation interfaces to an ACF system will **improve the acceptance** of that system among users.

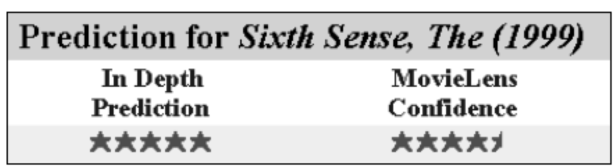
  Hypothesis 2: adding explanation interfaces to an ACF system will **improve the performance** of filtering decisions made by users of the ACF system.

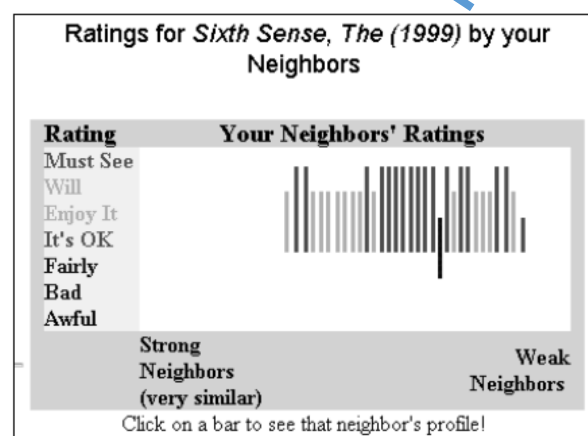
Experiment 2 – Acceptance and Performance

- **Design**
  - **Test subjects:** users of the **MovieLens** web-based ACF movie recommender
  - A month-long study
  - **Control group:** the standard MovieLens interface or with aesthetic changes
  - **Non-control group: a new experimental explanation interface**



Explanations





Experiment 2 – Acceptance and Performance

- **Survey** upon entering and leaving the experiment regarding their impressions of the MovieLens site
  - To figure out how explanation interfaces might affect the acceptance of ACF system

- Which movie did you see?
- Did you go because you thought you would enjoy the movie or did you go for other reasons (such as other viewers)?
- Did you consult MovieLens before going?
- If you consulted MovieLens, what did MovieLens predict?
- How much did MovieLens influence your decision?
- Was the movie worth seeing?
- What would you now rate the movie?

Experiment 2 – Acceptance and Performance

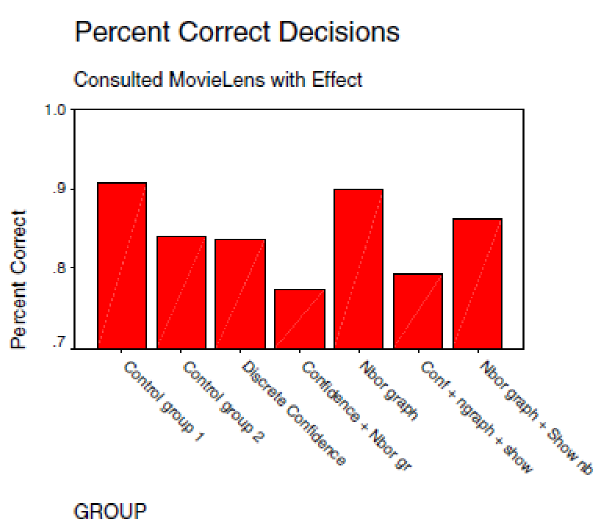
- **Results**
  - 210 users filling out 743 surveys

43% → consulted MovieLens before seeing the movie

82% of them said "MovieLens had some effect on the user's decision to see the movie"

83% of them said "MovieLens was NOT the sole reason for choosing a movie"

  - The filtering performance of each experimental group



No statistically significant difference between experimental groups

Experiment 2 – Acceptance and Performance

- **Results**
  - **Non-control group (a new experimental explanation interface)'s exit survey**
  - 97 participants

86% of them said that they would like to see their explanation interface added to the system

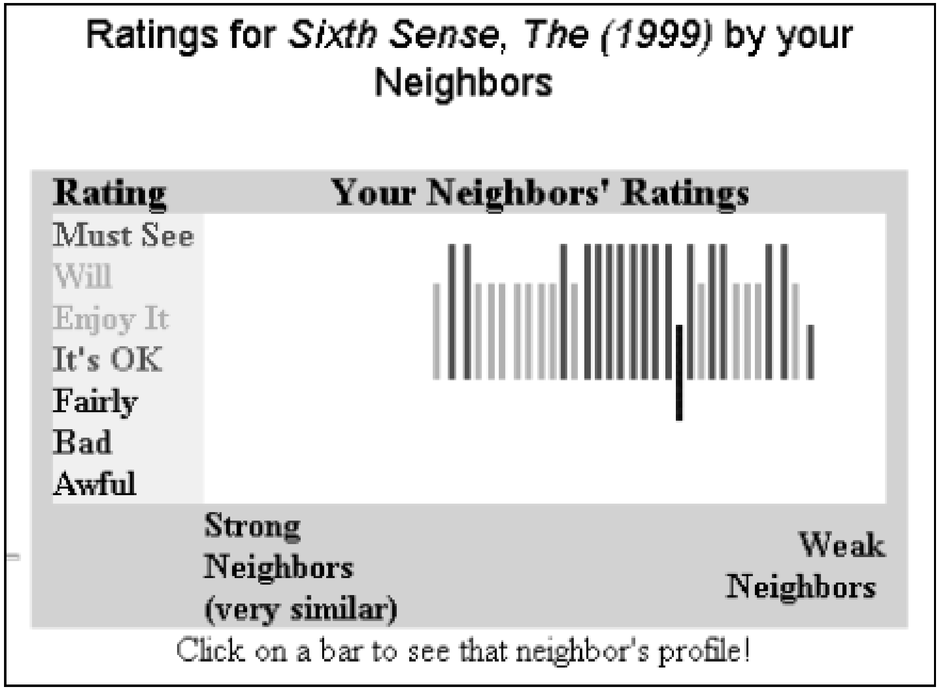
- 60 users' text comments on what they like most & least about the explanation interfaces = almost entirely positive

"It made sense of what seemed to be somewhat senseless numerical ratings"

"I could see the variety of responses to a film which corresponds to what I do with my friends. It helps me see how strongly they felt and the power or range of that diversity which always helps me be prepared for a film which evokes powerful response."

Experiment 2 – Acceptance and Performance

- **More comments & suggestions from participants**
  - View neighbors' rating profiles
  - Bookmark certain users!



- **Some negative comments**
  - Inadequacies in the prediction algorithm and NOT in the explanation interface!

Experiment 2 – Acceptance and Performance

- **Analysis**

Hypothesis 1: adding explanation interfaces to an ACF system will **improve the acceptance** of that system among users.

Users see explanation as a **valuable** component of an ACF system (86% of survey participants)

Hypothesis 2: adding explanation interfaces to an ACF system will **improve the performance** of filtering decisions made by users of the ACF system.

Filtering performance measurements → **inconclusive…**

- Lack of good data
- Large amount of uncontrolled variance (between users)

Conclusion

1. **What models and techniques are effective in supporting explanation in an ACF system?**
   - Simple interface
   - neighbor's information (rating histograms)
   - Past performance of the system

2. **Can explanation facilities increase the acceptance of automated collaborative filtering systems?**
   - Most users value the explanations and would like to see them!

3. **Can explanation facilities increase the filtering performance of ACF system users?**
   - Fail to prove or disprove…
   - Users perform filtering based on many different channels of input, and attempting to isolate the affect of one filtering or decision aid requires well controlled studies