Recommender Systems: Explanations, Attacks, Context

Radek Pelánek

2014

Explanations of Recommendations

- recommendations: selection (ranked list) of items
- explanations: (some) reasons for the choice

Goals of Providing Explanations

Why explanations?

Goals of Providing Explanations

Why explanations?

- transparency, trustworthiness, validity, satisfaction (users are more likely to use the system)
- persuasiveness (users are more likely to follow recommendations)
- effectiveness, efficiency (users can make better/faster decisions)
- education (users understand better the behaviour of the system, may use it in better ways)

Examples of Explanations

- knowledge-based recommenders
  - "Because you, as a customer, told us that simple handling of car is important to you, we included a special sensor system in our offer that will help you park your car easily."
  - algorithms based on CSP representation

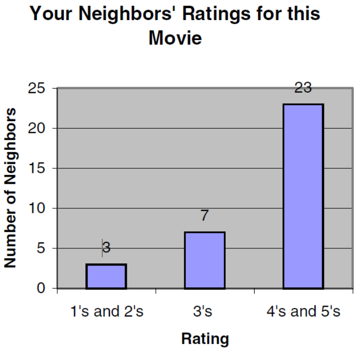
Examples of Explanations

- knowledge-based recommenders
  - "Because you, as a customer, told us that simple handling of car is important to you, we included a special sensor system in our offer that will help you park your car easily."
  - algorithms based on CSP representation
- recommendations based on item-similarity
  - "Because you watched X we recommend Y"

Explanations – Collaborative Filtering

Your Neighbors' Ratings for this Movie

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



Explaining Collaborative Filtering Recommendations, Herlocker, Konstan, Riedl

Explanations – Collaborative Filtering

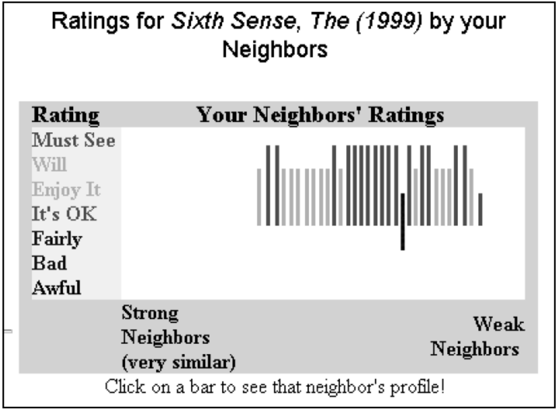


Figure 4. A screen explaining the recommendation for the movie "The Sixth Sense." Each bar represents a rating of a neighbor. Upwardly trending bars are positive ratings, while downward trending ones are negative. The x-axis represents similarity to the user.

Explaining Collaborative Filtering Recommendations, Herlocker, Konstan, Riedl

Explanations – Comparison

#		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

Table 1. Mean response of users to each explanation interface, based on a scale of one to seven. Explanations 11 and 12 represent the base case of no additional information. Shaded rows indicate explanations with a mean response significantly different from the base cases (two-tailed $\alpha = 0.05$).

Explaining Collaborative Filtering Recommendations, Herlocker, Konstan, Riedl

Attacks on Recommender System

- Why?
- What type of recommender systems?
- How?
- Countermeasures?

Attacks

susceptible to attacks: collaborative filtering

reasons for attack:

- make the system worse (unusable)
- influence rating (recommendations) of a particular item
    - *push attacks* – improve rating of "my" items
    - *nuke attacks* – decrease rating of "opponent's" items

Example

Items

Users	1	2	3	4	5	6	7
a	+	−		+	+		+
b	−	+	+	−	−		−
c	+	−	+		−	−	−
d	−	+	+	−			
e	−		−	−	−		−
f	+	−	+	+	+		+
g		−	+	+	−	−	+
h	+	−	+	+	+		?
i	+	−	+		−	−	−
j	−	+	+	−			−
k	−		−	−	−		−
l	+	−	+	+	+		−
m		−	+	+	−	−	−

Authentic profiles (users *a*–*g*)

Target profile (user *h*)

Attack profiles (users *i*–*m*)

Fig. 2 Simplified system database showing authentic user profiles and a number of attack profiles inserted. In this example, user *h* is seeking a prediction for item 7, which is the subject of a product nuke attack.

Robust collaborative recommendation, Burke, O'Mahony, Hurley

Types of Attacks

more knowledge about system $\rightarrow$ more efficient attack

random attack
: generate profiles with random values (preferably with some typical ratings)

average attack
: effective attack on memory-based systems (average ratings $\rightarrow$ many neighbors)

bandwagon attack
: high rating for "blockbusters", random values for others

segment attack
: insert ratings only for items from specific segment

special nuke attacks
: love/hate attack, reverse bandwagon

Example

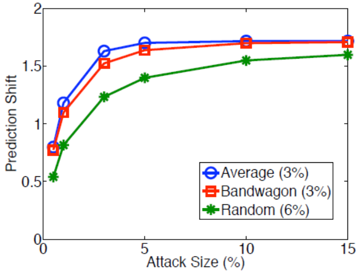


Fig. 3 Prediction shift (left) and hit ratio (right) for product push attacks mounted against the user-based collaborative recommendation algorithm. Hit ratio results relate to a 10% attack size.

Robust collaborative recommendation, Burke, O'Mahony, Hurley

Countermeasures

- more robust techniques: model based techniques, additional information
- increasing injection costs: Captcha, limited number of accounts for single IP address
- automated attack detection

Attacks and Educational Systems

- cheating $\sim$ false rating
  example: Problem Solving Tutor, Binary crossword
- gaming the system – using hints as solutions

can have similar consequences as attacks

Context Aware Recommendations

context:

- **physical** – location, time
- **environmental** – weather, light, sound
- **personal** – health, mood, schedule, activity
- **social** – who is in room, group activity
- **system** – network traffic, status of printers

Contextualization

- pre- post- filtering
- model based
  - multidimensionality: user $\times$ item $\times$ time $\times \ldots$
  - tensor factorization

Context – Applications

- tourism, visitor guides
- museum guides
- home computing and entertainment