Recommender Systems
Research Challenges

Francesco Ricci
Free University of Bozen-Bolzano
fricci@unibz.it

unibz
Fakultät für Informatik
Facoltà di Scienze e Tecnologie informatiche
Faculty of Computer Science

iif 2017

Content

- Recommender systems motivations
- Recommender system
- Critical Assumptions
- Preference modeling
- Choice modeling
- System dynamics
- Group dynamics

Context

Choice

Dynamics

Explosion of Choice

- A trip to a **local supermarket**:
  - 85 different varieties and brands of crackers.
  - 285 varieties of cookies.
  - 165 varieties of “juice drinks”
  - 75 iced teas
  - 275 varieties of cereal
  - 120 different pasta sauces
  - 80 different pain relievers
  - 40 options for toothpaste
  - 95 varieties of snacks (chips, pretzels, etc.)
  - 61 varieties of sun tan oil and sunblock
  - 360 types of shampoo, conditioner, gel, and mousse.
  - 90 different cold remedies and decongestants.
  - 230 soups, including 29 different chicken soups
  - 175 different salad dressings and if none of them suited, 15 extra-virgin olive oils and 42 vinegars and make one’s own

Choice and Well-Being

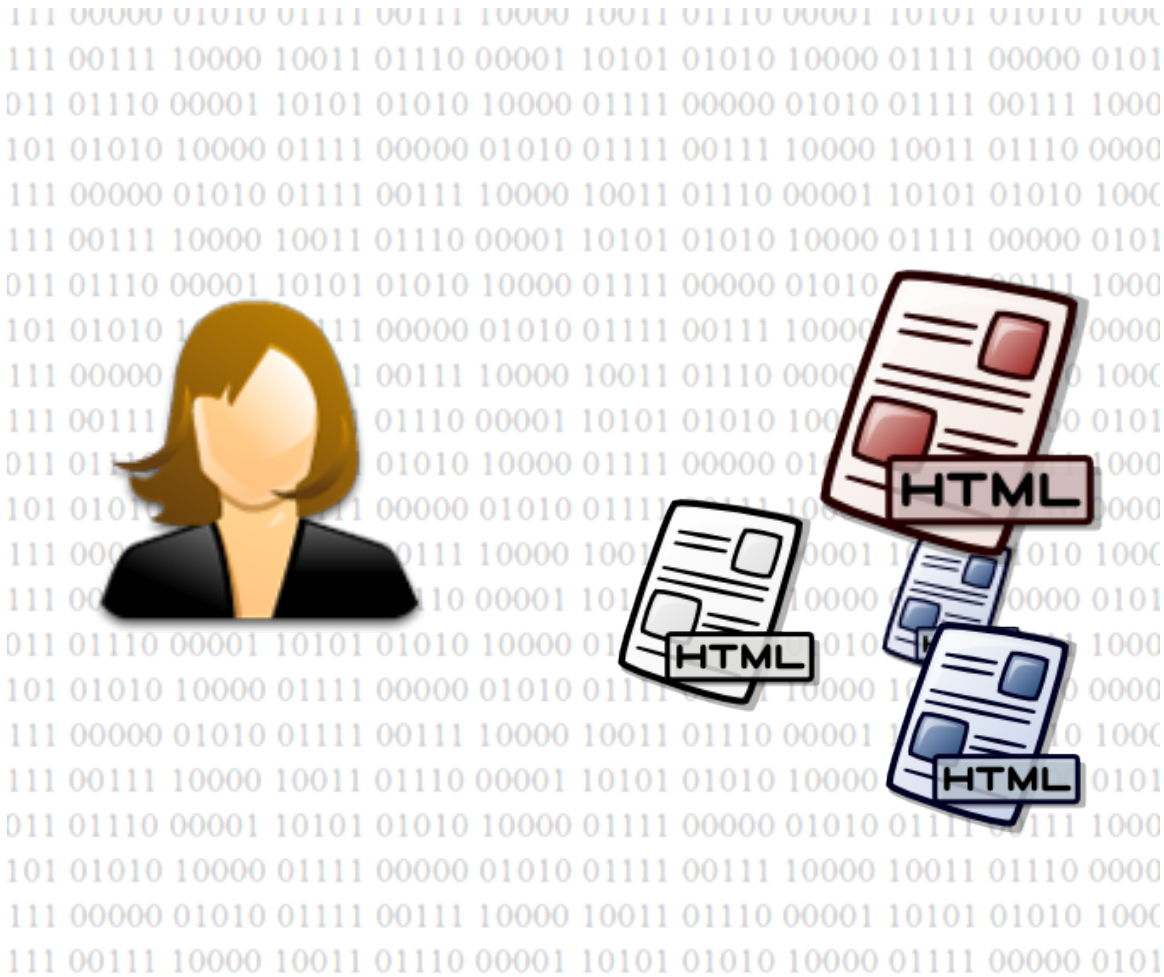
- We have **more choice**, more freedom, autonomy, and self determination
- Increased choice **should improve well-being:**
  - *added options can only make us better off: those who care will benefit, and those who do not care can always ignore the added options*
- Various assessment of well-being have shown that **increased affluence** have accompanied by **decreased well-being**.

Successful Queries are the Minority

Engines:	Total(%)	Organic(%)	PPC(%)
Google	19.41%	18.67%	0.74%
Yahoo	29.59%	29.59%	0.00%
Bing	15.06%	15.06%	0.00%
AOL	40.48%	40.48%	0.00%
ASK	50.98%	50.98%	0.00%
Other	50.66%	50.66%	0.00%

Source: http://www.keyworddiscovery.com/

Queries will disappear



Leverage multiple signals to get rid of queries

Recommender Systems

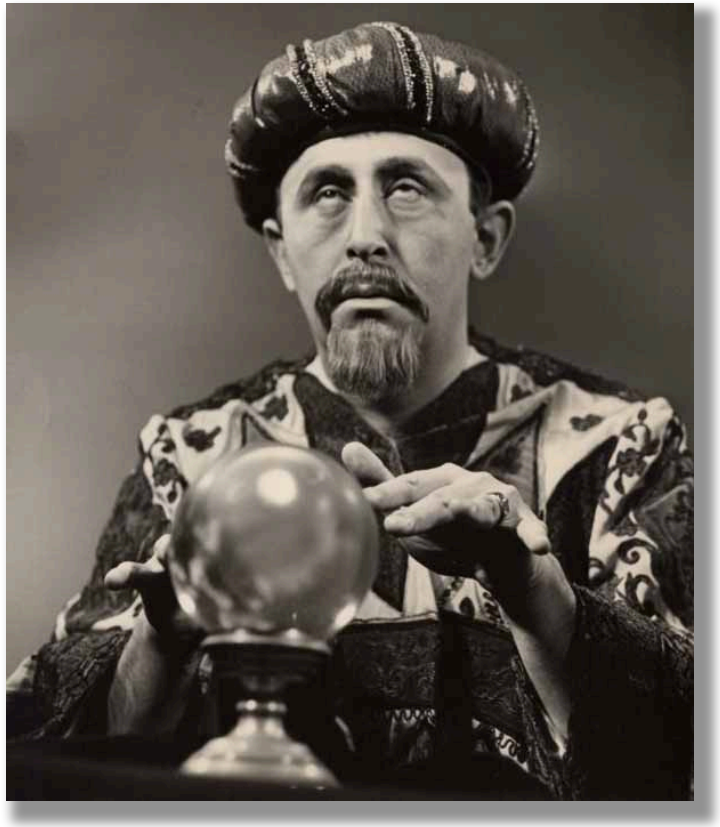
Amazon.it

170 engineers in Amazon are dedicated to the recommender system

Movie Recommendation – YouTube

Recommendations account for about 60% of all video clicks from the home page.

1. Preference Elicitation



2. Predicting

3. Selecting and presenting the recommendations

Classical Recommendation Model

Two types of entities: **Users and Items**

1. A **background knowledge**:
   - A set of **ratings – preferences** - is a map
     - *r: Users x Items → [0,1] U {?}*
   - A set of "features" of the Users and/or Items
2. A method for **predicting** the r function on (user, item) pairs where it is unknown

$$r^*(u, i) = \text{Average}_{\text{su is similar to u}} \{r(su, i)\}$$

3. A method for **selecting** the items to recommend:
   - Recommend to u the item $i^* = \arg\max_{i \in \text{Items}} \{r^*(u,i)\}$

G. Adomavicius, A. Tuzhilin: Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Trans. Knowl. Data Eng. 17(6): 734-749 (2005)

Movie rating data

Training data

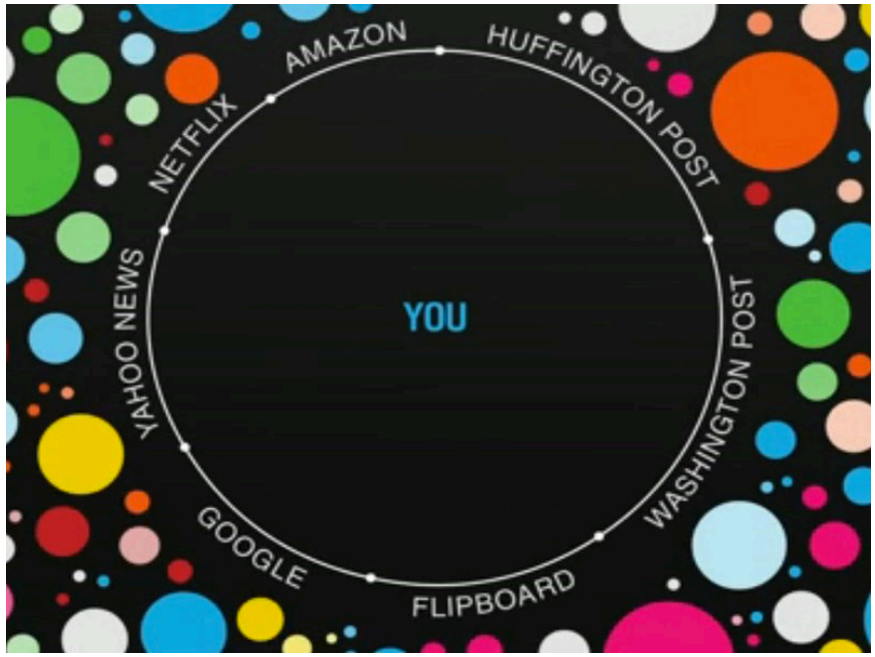
user	movie	date	score
1	21	5/7/02	1
1	213	8/2/04	5
2	345	3/6/01	4
2	123	5/1/05	4
2	768	7/15/02	3
3	76	1/22/01	5
4	45	8/3/00	4
5	568	9/10/05	1
5	342	3/5/03	2
5	234	12/28/00	2
6	76	8/11/02	5
6	56	6/15/03	4

Test data

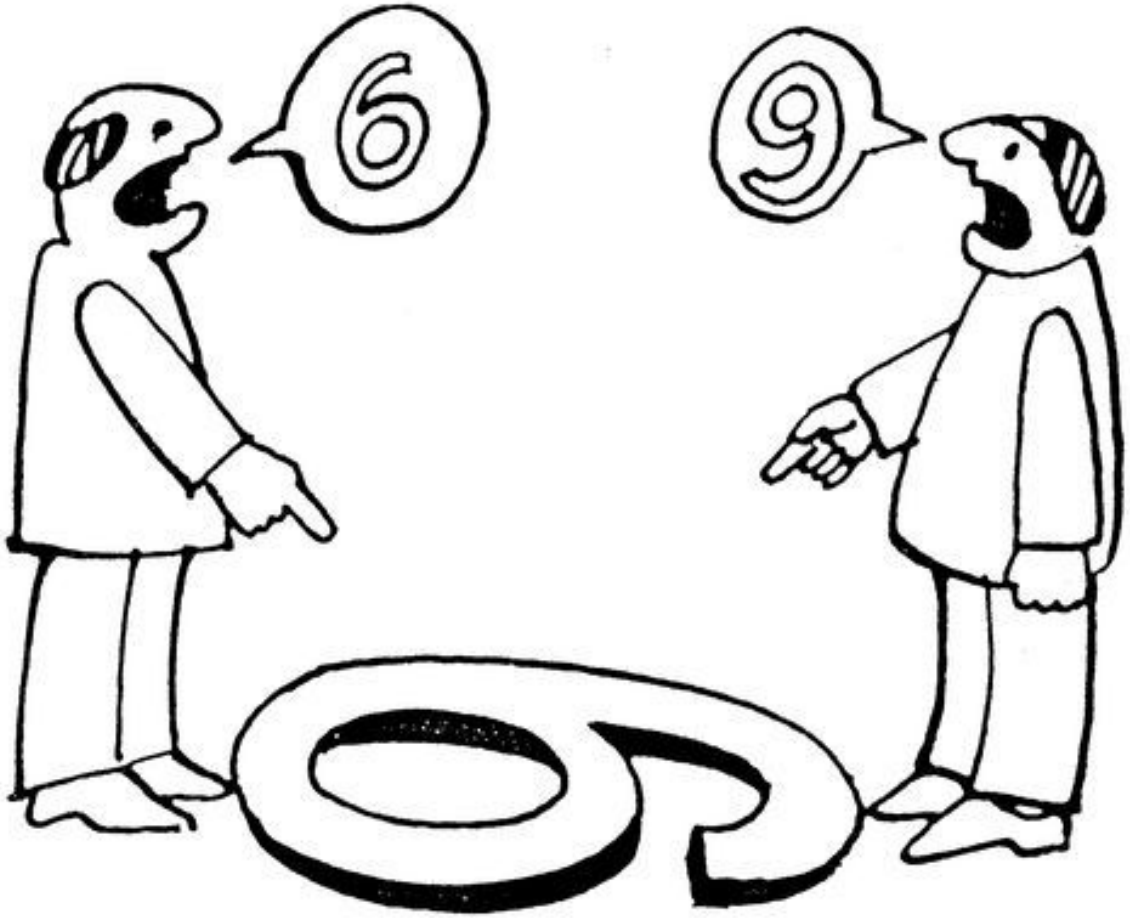
user	movie	date	score
1	62	1/6/05	?
1	96	9/13/04	?
2	7	8/18/05	?
2	3	11/22/05	?
3	47	6/13/02	?
3	15	8/12/01	?
4	41	9/1/00	?
4	28	8/27/05	?
5	93	4/4/05	?
5	74	7/16/03	?
6	69	2/14/04	?
6	83	10/3/03	?

Problems and Issues

- Cold Start (new user and new item)
- Filter Bubble
- How much to personalize
- How to contextualize
- Learning to interact and proactivity
- Recommendations for Groups
- Scalability and big data
- Privacy and security
- Diversity and serendipity
- Stream based recommendations

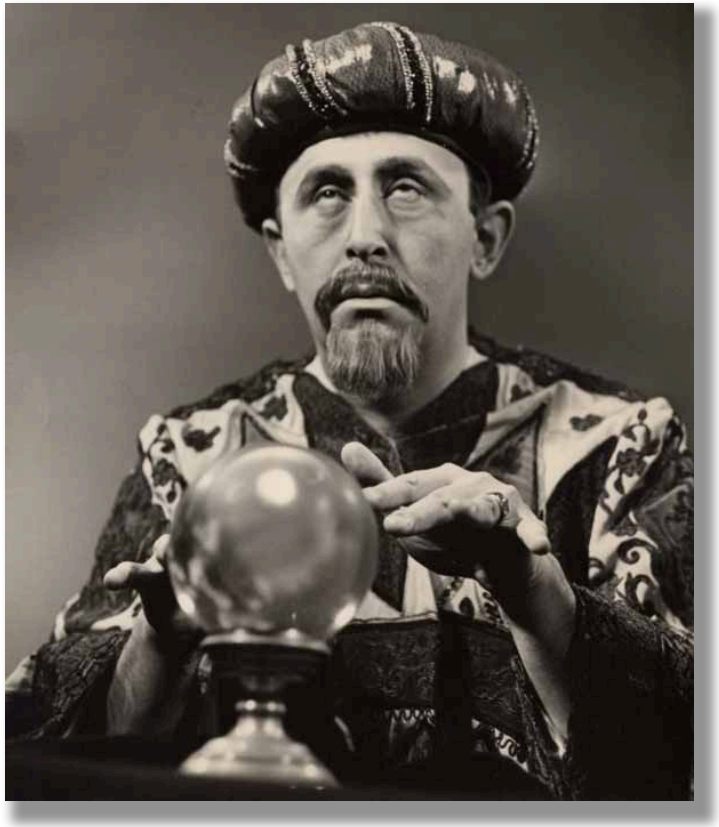


Critical Assumptions



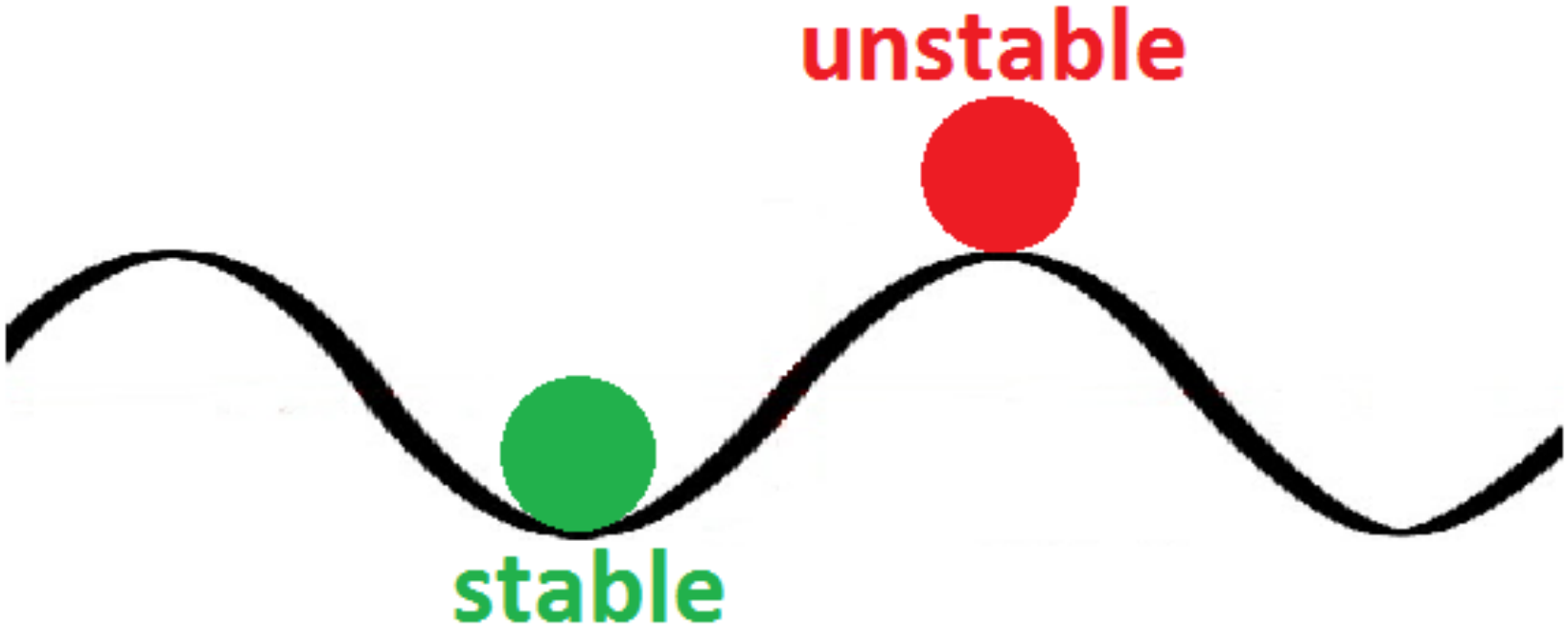
Predictability

- **Predictability:** observing users' behavior the system can build a concise algorithmic model of what they like



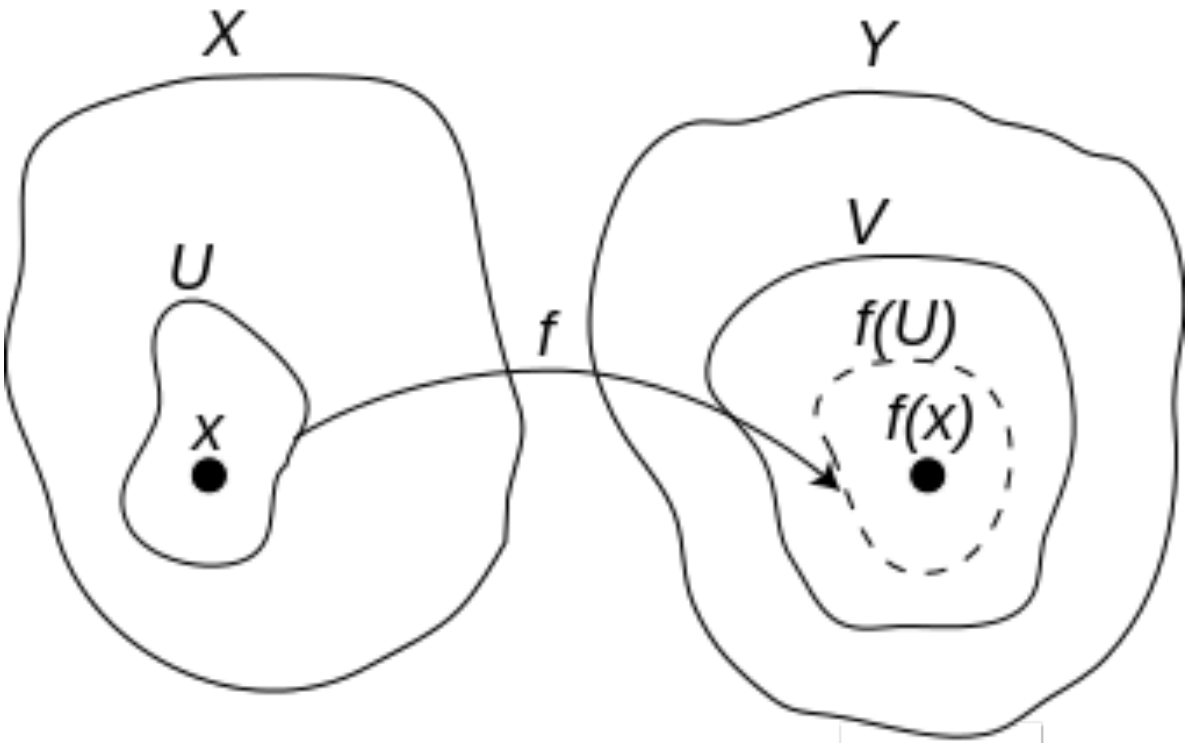
Stability of User Preferences

- User **preferences** are supposed to be rather **stable –** models are built by using historical data



Continuity

- User preference function is "**continuous**": there exist a notion of item-to-item similarity such that similar items generate similar reactions in a user



FALSE

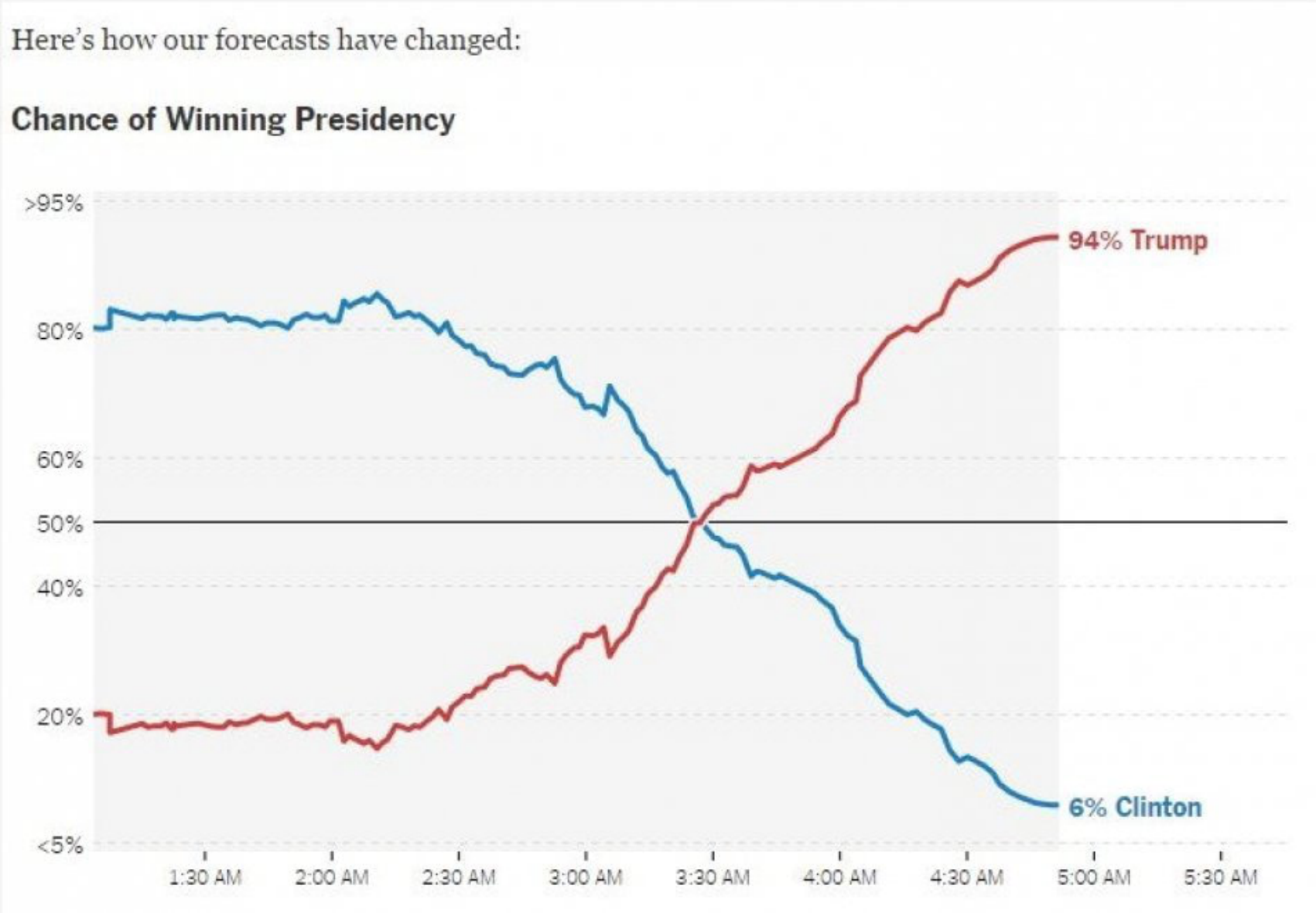
Violation of stability and continuity

- *Today I shave with an electric razor while last month I was shaving with a disposable razor*
- *I went to sea places for the last 3 summers but next year I will hike in the mountains*
- *I like Pustertal but I do not like Vinshgau*

Pustertal

Vinshgau

Predicting user behaviour is hard



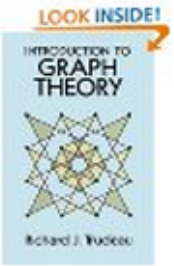
Preferences

Ratings (recommendations)

amazon.com

Help | Close window

Recommended for You



Introduction to Graph Theory (Dover Books on Mathematics)
by Richard J. Trudeau (February 9, 1994)
In Stock
List Price: ~~$14.95~~
Price: $3.99
59 used & new from $3.26

Add to Cart | Add to Wish List

Rate this item
☒ ☆☆☆☆☆
I own it
Not interested

Because you purchased...

Patterns of Software: Tales from the Software Community (Hardcover)
by Richard P. Gabriel (Author)

☒ ☆☆☆☆☆
This was a gift
Don't use for recommendations

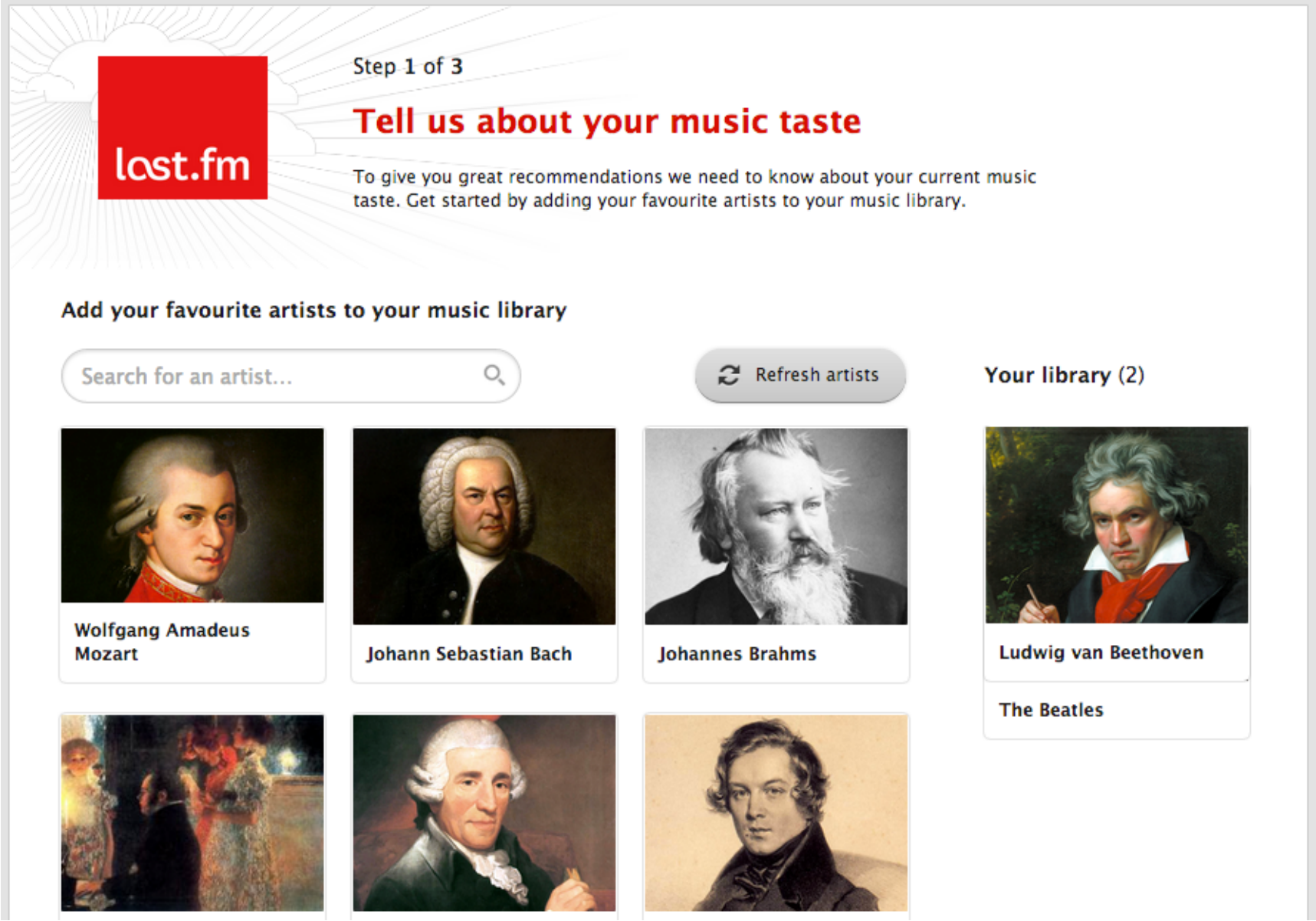
Machine Learning (Hardcover)
by Tom M. Mitchell (Author)

☒ ☆☆☆☆☆
This was a gift
Don't use for recommendations

Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning) (Hardcover)
by Richard S. Sutton (Author), Andrew G. Barto (Author)

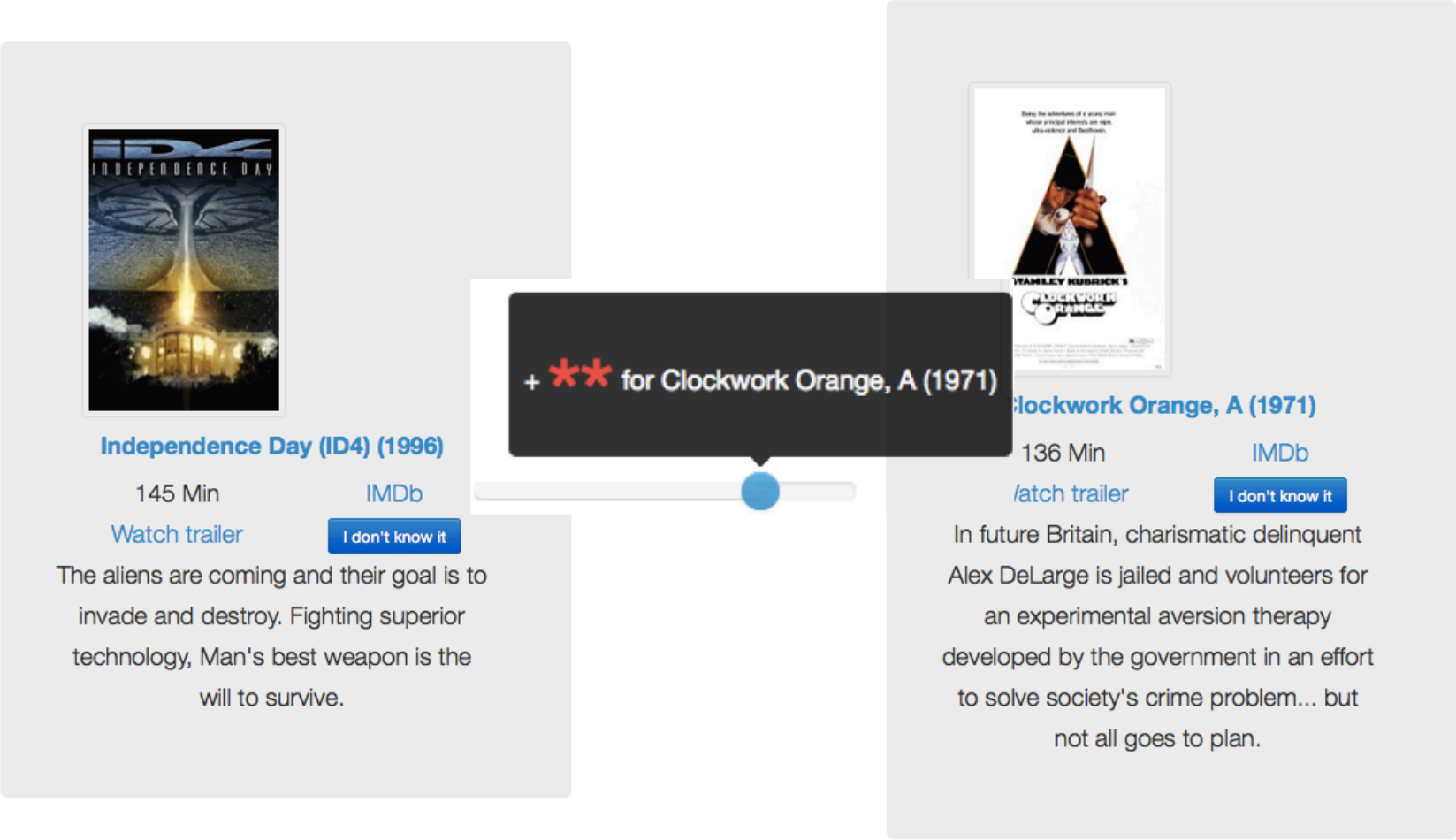
☒ ☆☆☆☆☆
This was a gift
Don't use for recommendations

Likes


last.fm
Step 1 of 3
Tell us about your music taste
To give you great recommendations we need to know about your current music taste. Get started by adding your favourite artists to your music library.
Add your favourite artists to your music library
Search for an artist...
Refresh artists
Your library (2)
Wolfgang Amadeus Mozart
Johann Sebastian Bach
Johannes Brahms
Ludwig van Beethoven
The Beatles

Likes

Pairwise Preferences



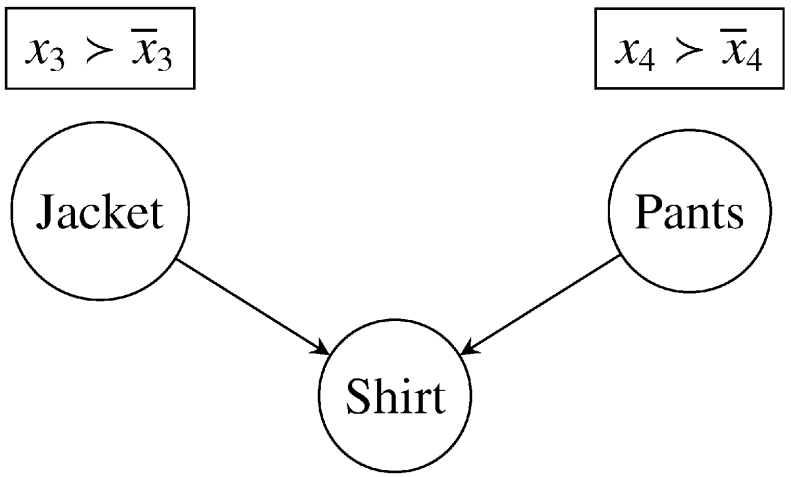
Pairwise-Based Recsys

- System that uses pairwise preferences for eliciting user preferences makes users **more aware of their choice options**
- A system variant based on pairwise preferences **outperformed a rating-based variant** in terms of recommendation accuracy measured by nDCG and precision
- **Nearest-neighbor** approaches are effective, but the user-to-user similarity must be computed with specific metrics (e.g. Goodman Kruskal gamma correlation)

- L. Blédaité, F. Ricci: Pairwise Preferences Elicitation and Exploitation for Conversational Collaborative Filtering. HT 2015: 231-236
- S. Kalloori, F. Ricci, M. Tkalcic: Pairwise Preferences Based Matrix Factorization and Nearest Neighbor Recommendation Techniques. RecSys 2016: 143-146

CP-Network

Variable	Literals
Coat	trench (x_1) and parka ($\overline{x}_1$)
Hat	large (x_2) and small ($\overline{x}_2$)
Jacket	black (x_3) and white ($\overline{x}_3$)
Pants	black (x_4) and white ($\overline{x}_4$)
Shirt	red (x_5) and white ($\overline{x}_5$)
Shoes	heels (x_6) and flats ($\overline{x}_6$)



$$
\begin{array}{l}
x_3 \wedge x_4 : \; x_5 \succ \overline{x}_5 \\
\overline{x}_3 \wedge \overline{x}_4 : \; x_5 \succ \overline{x}_5 \\
x_3 \wedge \overline{x}_4 : \; \overline{x}_5 \succ x_5 \\
\overline{x}_3 \wedge x_4 : \; \overline{x}_5 \succ x_5
\end{array}
$$

Frédéric Koriche, Bruno Zanuttini: Learning conditional preference networks. Artif. Intell. 174(11): 685-703 (2010)

Choice Modeling

The recommender is an agent that can take decision on behalf of the user (for the user)

Decision Making

- A **decision maker** DM selects a single alternative (or action) $a \in A$
- An **outcome** (or consequence) $x \in X$ of the chosen action depends on the **state of the word** $s \in S$
- **Consequence function**:

$$c: A \times S \rightarrow X$$

- **User preferences** are expressed by a value or **utility function** – desirability of outcomes:

$$v: X \rightarrow \mathbb{R}$$

- **Goal:** *select the action $a \in A$ that leads to the best outcome*

D. Brazunas, Computational Approaches to Preference Elicitation, Tech Rep University of Toronto, 2006

Preferences under certainty

- The state $s \in S$ is **known** – one action leads to one outcome
- Preferences over outcomes determines the **optimal action** (recommendation):
  - **Rational agent** selects the action with the most preferred outcome
- **Weak preference** over $X \ni x, y$
  - Binary relation $x \succcurlyeq y$
  - Comparability: $\forall x, y \in X, x \succcurlyeq y \vee y \succcurlyeq x$
  - Transitivity: $\forall x, y, z \in X, x \succcurlyeq y \wedge y \succcurlyeq z \Longrightarrow x \succcurlyeq z$
- Weak preferences can be **represented** (when X is finite) by an ordinal value function: $v: X \rightarrow \mathbb{R}$ that agrees with the ordering $\succcurlyeq$, i.e.:

$$v(x) \geq v(y) \Leftrightarrow x \succcurlyeq y$$

Example – one user - certainty

- **Actions** = *{swim, run}*
- **States** = *Contexts = {sun, rain}*
- **Outcomes** *X = Contexts x Items = {(swim, sun), (swim, rain), (run, sun), (run, rain)}*
- **Preferences in context**:
  - *v(swim, sun) = 3, v(swim, rain) = 4, v(run, sun) = 5, v(run, rain) = 1*
- *Context is know*
  - If it is *sun* then recommend: *run*
  - If it is *rain* then recommend: *swim*

Recommender

- **If** the **context is know**
- **And** we **know** – or we can fully predict - the **preferences** of the user *u* over the space of outcomes *X* (items in context) - either as pairwise comparisons or as an ordinal function (rating):

$$r: U \times I \times C \rightarrow R$$

- *Then we can predict the user choice*

$$i^* = \arg\max_{i \in Items} \{r(u, i, c)\}$$

- **Unfeasible!**
  - We do not fully know the relevant context
  - It is too hard to accurately predict the preferences in the current user context.

G. Adomavicius, A. Tuzhilin: Context-Aware Recommender Systems. Recommender Systems Handbook 2015: 191-226

Preferences under uncertainty

- Consequences of actions are **uncertain**
- **Lottery**: $<x, p, x'>$, x occurs with probability p or x' with probability $(1-p)$
- Rational decision makers are assumed to have **complete and transitive preferences ranking** $\succcurlyeq$ over a set of lotteries L
- If the weak preference relation $\succcurlyeq$ over lotteries is (1) complete, (2) transitive, (3) continuity, (4) independence, then there is an **expected (or linear) utility function** $u: L \rightarrow \mathbb{R}$ which represents $\succcurlyeq$
  - $u(l) \geq u(l') \Leftrightarrow l \succcurlyeq l'$
  - $u(<l, p, l'>) = p\, u(l) + (1-p)\, u(l'), \ \forall l, l' \in L, p \in [0,1]$
  - $u(l)=u(<p_1, x_1; \ldots p_n, x_n>) = p_1\, u(x_1) + \ldots + p_n\, u(x_n)$

Example – one user - uncertainty

- $A = \{swim, run\}$
- $S = C = \{sun, rain\}$
- $X = C \times I = \{(swim, sun), (swim, rain), (run, sun), (run, rain)\}$
- Preferences: $v(swim, sun) = 3$, $v(swim, rain) = 4$, $v(run, sun) = 5$, $v(run, rain) = 1$
- $p(sun) = 0.8$, $p(rain)=0.2$
- Choice is determined by **expected utility**
  - $v(swim) = 3 * 0.8 + 4 * 0.2 = 3.2$
  - $v(run) = 5 * 0.8 + 1 * 0.2 = 4.2$
  - Recommend: *run*

Preference Knowledge

- The system knowledge of the user preferences is not only incomplete but it is also **largely inaccurate**

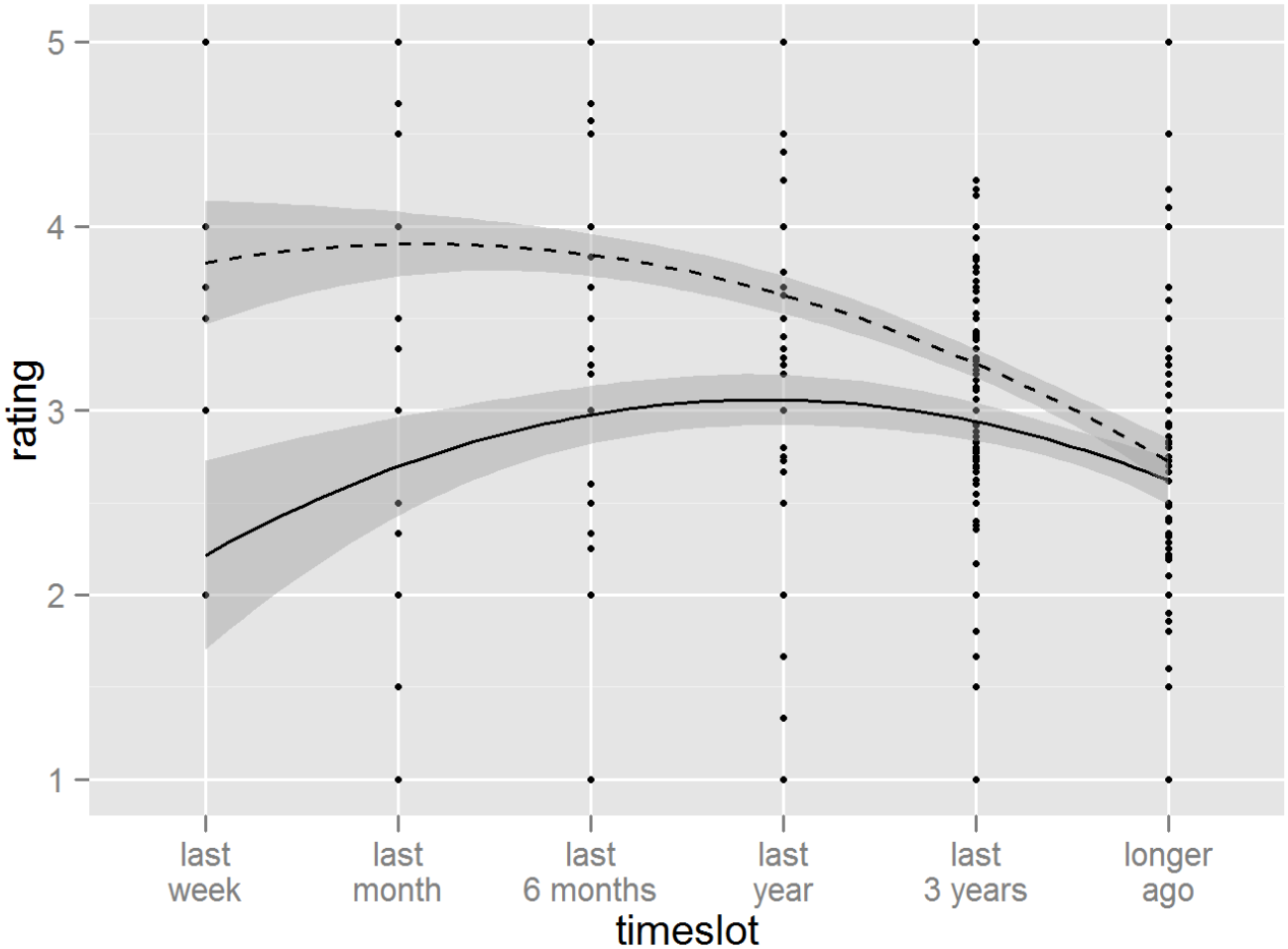
Remembering

- D. Kahneman (nobel prize): what we remember about an experience is determined by (**peak-end rule**)
  - *How the experience felt when it was at its peak (best or worst)*
  - *How it felt when it ended*
- We rely on this summary later to remind how the experience felt and decide whether to have that experience again
- *So how well do we rate or compare?*
  - It is doubtful that we prefer an experience to another very similar just because the first ended better.

Remembering the Stars?

The movies were split based on the average rating in the first timeslot

Over time ratings regress to the middle of the scale.

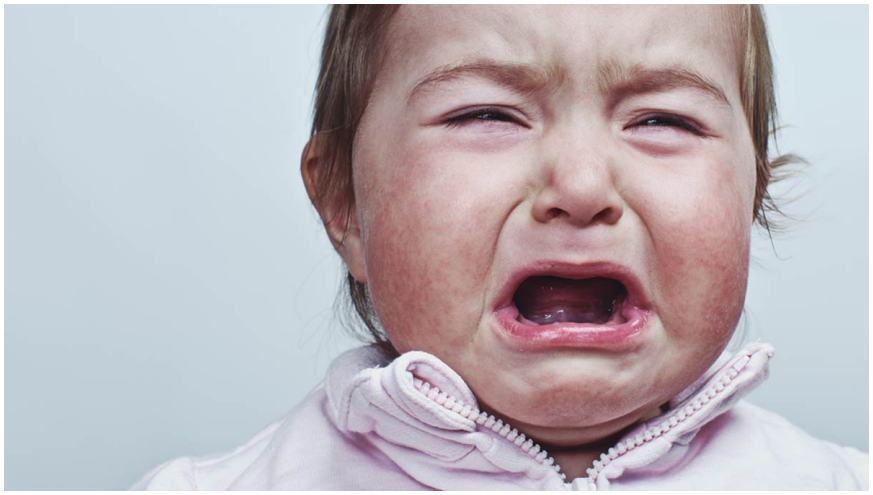


Rating as function of time past after watching a movie. *Dashed line for initially high rated movies, solid line for initially low rated movies.*

D. G. F. M. Bollen, M. P. Graus, M. C. Willemsen: Remembering the stars?: effect of time on preference retrieval from memory. RecSys 2012: 217-220

Summing Up – so far

- Preferences are **context dependent**
- It is practically **impossible to know/predict preferences** in all the potentially relevant contexts
- Preferences judgements acquired **after the experience** of the item are **unreliable**
- Preferences acquired for experiences we had **some time ago** are **not reliable at all**.



Irrelevant Context

- It is hard to say what is really irrelevant

Attraction Effect

- Alternative options:
  - You could get access to all our **web content** for $59,
  - A subscription to the **print edition** for $125,
  - Or a **combined print and web** subscription, also for $125.
- D. Ariely surveyed students about which option they preferred
  - Predictably, nobody chose print subscription alone;
  - 84% opted for the **combination deal**,
  - and 16% for the web subscription.

Ariely, Dan. Predictably Irrational: The Hidden Forces That Shape Our Decisions. New York: Harper Perennial, 2010.

Without Attraction

- Alternative options:
  - You could get access to all our **web content** for $59,
  - Or a **combined print and web** subscription, also for $125.

- D. Ariely surveyed again students about which option they preferred
  - 32% wanted the print subscription (vs 84% in the previous experiment)
  - while 68% preferred to go **web-only** (vs 16% in the previous experiment).

Irrelevant context does matter

- Modeling the alternative options as context

$$r: U \times I \times C \rightarrow R$$

- With the dominated option
  - r(u, web, (print, print+web)) = 4
  - r(u, print, (web, print+web)) = 0
  - **r(u, print+web, (web, print)) = 5**
- Without the dominated option
  - r(u, web, (print+web)) = 4
  - **r(u, print+web, (web)) = 3**

Context space explodes: we must consider even apparently irrelevant context wen estimating preferences.

Preferences and Choice

- The previous example can also be explained by saying that
  - *Preferences do not completely determine user choice*
  - *Users are not maximizing (expected) utility*
  - *More complex choice models are needed*

Choice Model

- A model of choice gives the **probability** of **choosing** an item *i* from a set of choices *X*: *p(i|X)*
- If *i* is represented by a feature vector v_i the **multinomial logit model** (MLM) state that:

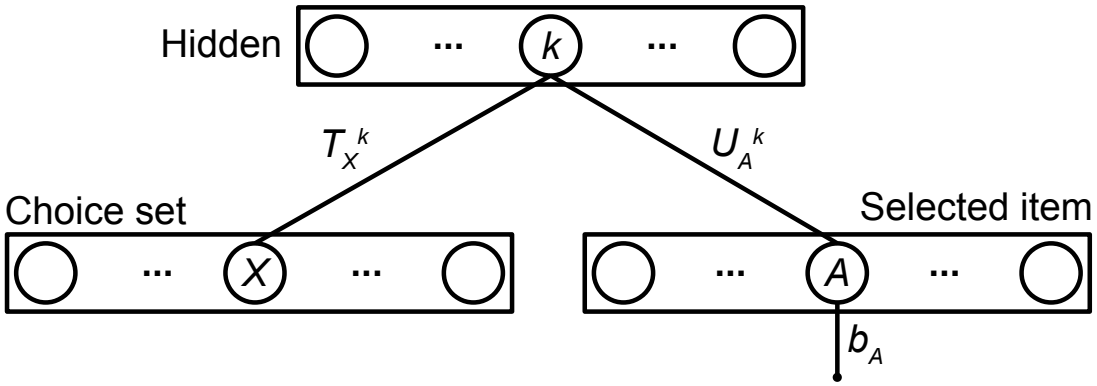
$$p(i\,|X) = \frac{\exp(w^T v_i)}{\sum_{j \in X} \exp(w^T v_j)}$$

- *w* is a vector of weights and $w^T v_i$ is the attractiveness of *i* (modelled by v_i)
- $w^T v_i = r(u,i)$ – assuming *w* is the vector modeling *u*
- This is a step ahead from the assumption that *u* will choose the item *i* that maximizes *r(u,i)*.

T. Osogami, Human choice and good choice, in The role and importance of mathematics in innovation, Springer, 2017.

Restricted Boltzmann Machine

- MLM choice model cannot explain ”attraction” since the ratio of *p(i|X)* and *p(j|X)* does not change if we remove an item *k* from the choice set *X*
- In a restricted Boltzmann machine the attractiveness of an item depends on the attractiveness of the other items



T. Osogami, M. Otsuka: Restricted Boltzmann machines modeling human choice. NIPS 2014: 73-81

System Dynamics

Collaborative-Based Filtering

- A collection of n users U and a collection of m items I
- A $n \times m$ matrix of ratings r_{ui}, with $r_{ui} = ?$ if user u did not rate item i
- Prediction for user *u* and item *j* is computed as

$$r_{uj}^{*} = r_u + K \sum_{v \in N_j(u)} w_{uv} (r_{vj} - r_v)$$

A set of neighbours of u that have rated j

- Where, r_u is the average rating of user u, K is a normalization factor such that the absolute values of w_{uv} sum to 1, and

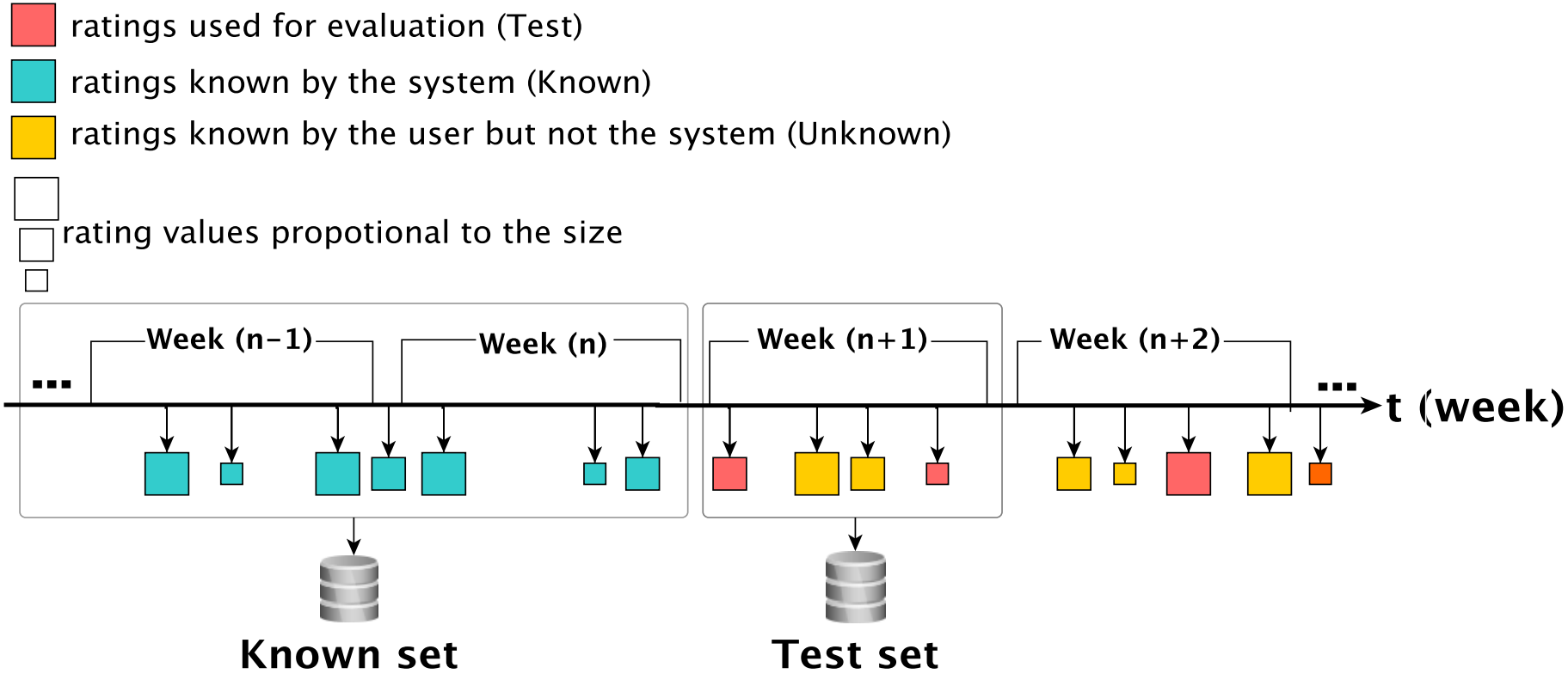
$$w_{uv} = \frac{\sum_{j \in I_{uv}} (r_{uj} - r_u)(r_{vj} - r_v)}{\sqrt{\sum_{j \in I_{uv}} (r_{uj} - r_u)^2 \sum_{j \in I_{uv}} (r_{vj} - r_v)^2}}$$

Pearson Correlation of users u and v

Preference Elicitation

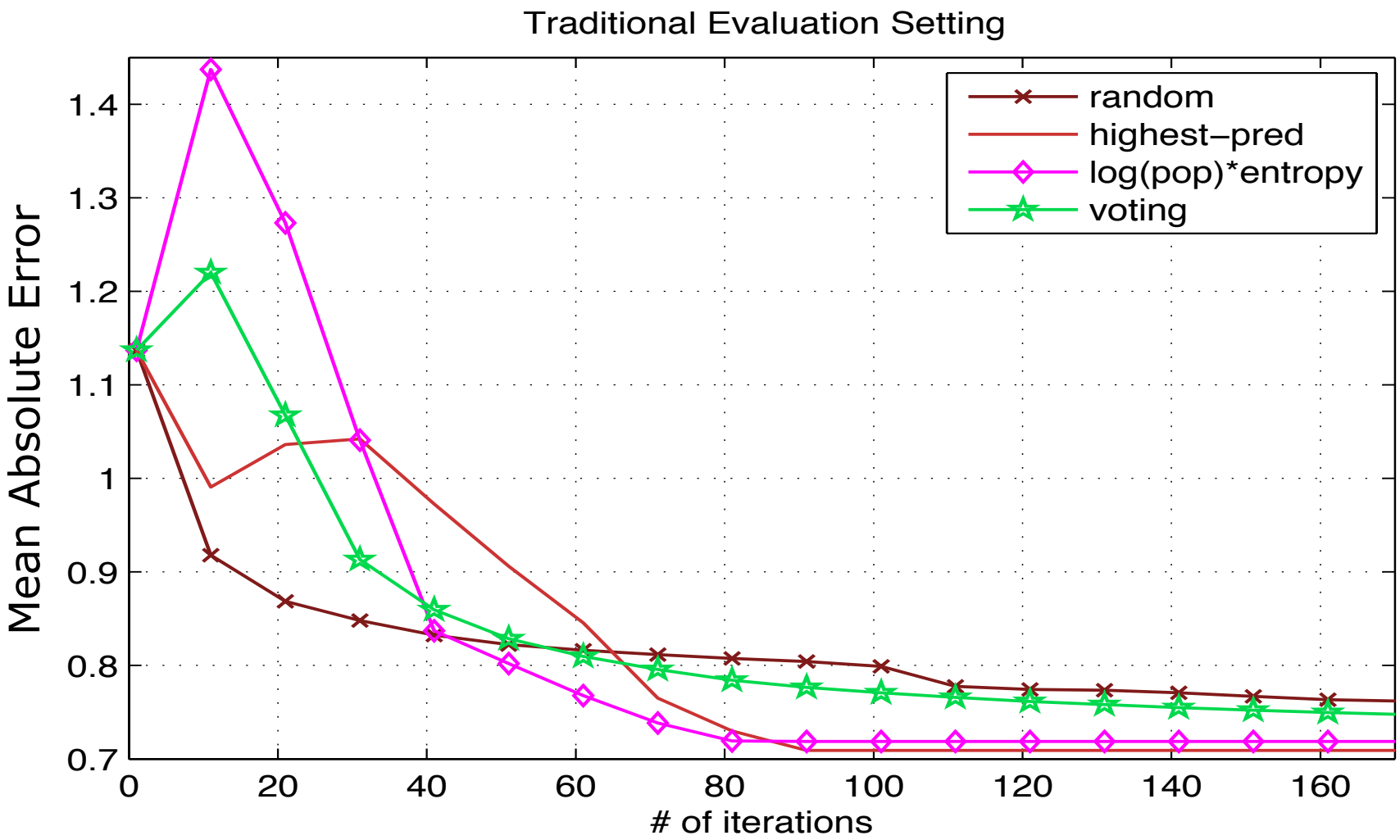
- We will **never** have **complete knowledge** of user preferences
- Preferences and their elicitation are **dynamic**
- Users elicit preferences under a variety of stimuli
  - The recommender
  - The experienced items
  - Reactions to other exposed preferences
- Is the recommender **performance** influenced by the **preference elicitation process**?
- Should a recommender system also (partially) **control** this process?

Simulating Rating Acquisition



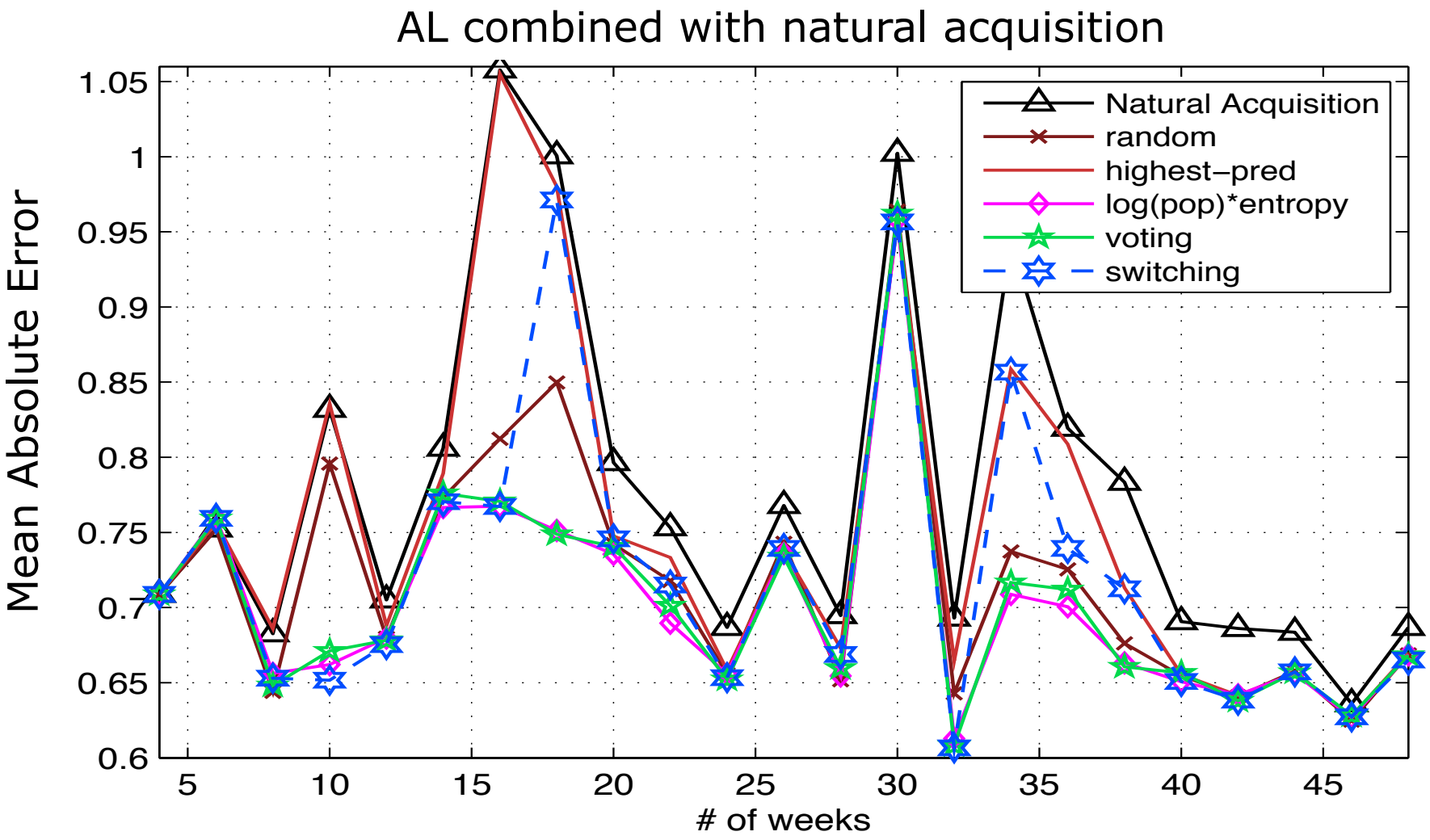
[M. Elahi, F. Ricci, N. Rubens: Active learning strategies for rating elicitation in collaborative filtering: A system-wide perspective. ACM TIST 5(1): 13:1-13:33 (2013)]

Active Learning Strategies



M. Elahi, F. Ricci, N. Rubens: Active learning strategies for rating elicitation in collaborative filtering: A system-wide perspective. ACM TIST 5(1): 13:1-13:33 (2013)

Active Learning and Natural Acquisition



Group (Micro) Dynamics

Group Recommendations

- Recommenders are usually designed to provide recommendations **adapted** to the **preferences** of a **single user**
- In many situations the recommended items are consumed by a **group of users**
  - A travel with friends
  - A movie to watch with the family during Christmas holidays
  - Music to be played in a car for the passengers

Group Recommendation Model

- Items will be experienced by individuals **together** with the other group members: the preference function **depends** on the **group:**

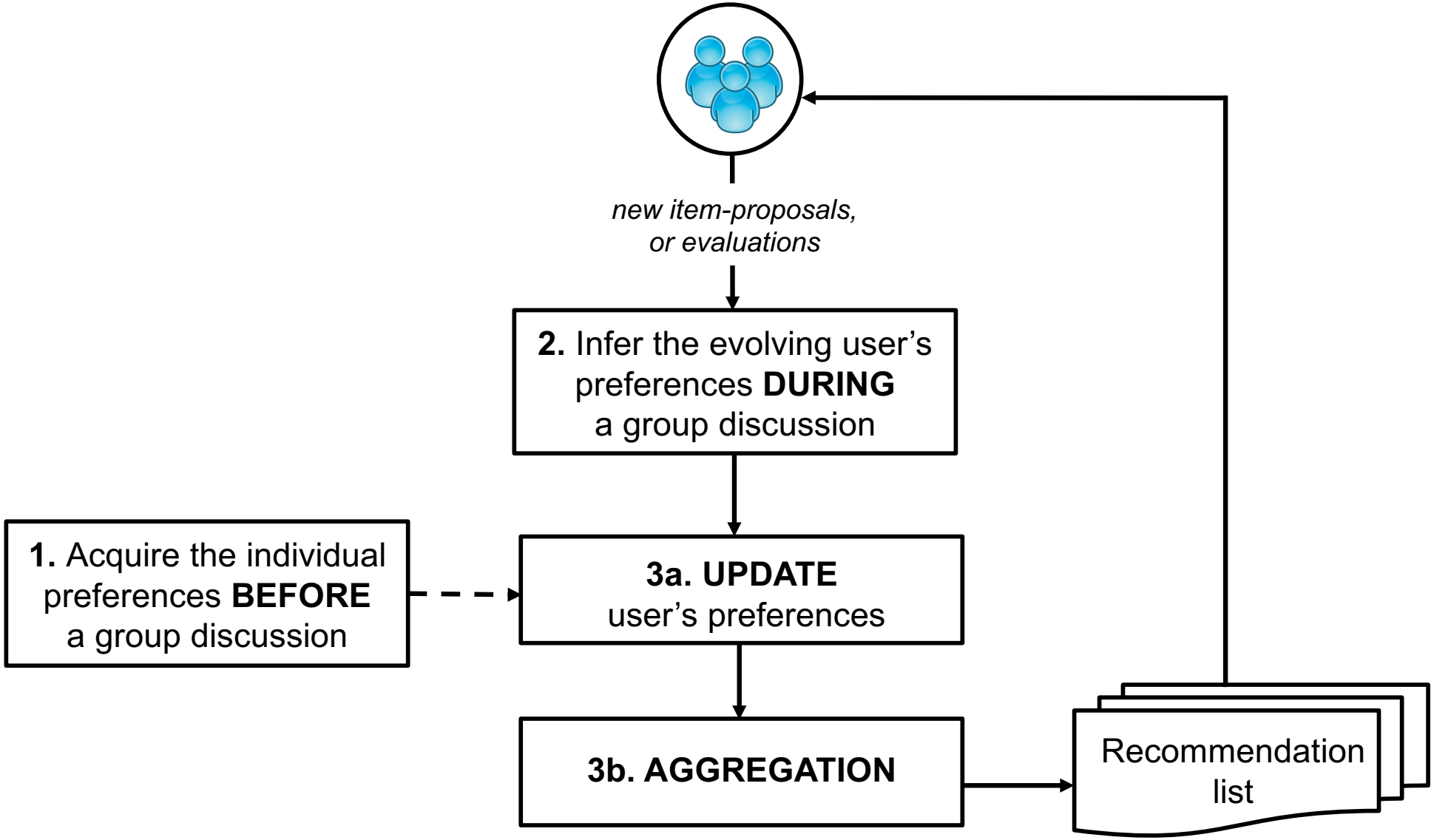
$$r : U \times I \times \wp(U) \rightarrow E$$

- *U* is the set of users, *I* is the set of Items, *P(U)* is the set of subsets of users (groups), *E* is the evaluation space (e.g. the ratings *{?, 1, 2, 3, 4, 5}*) of the rating function *r*
- *In general r(u, i) ≠ r(u, i, g),* for $g \ni u$
- Users are influenced in their evaluation by the group composition (e.g., emotional contagion [Masthoff & Gatt, 2006]).

Effects of Groups on User Satisfaction

- **Emotional Contagion**
  - Other users being satisfied may increase a user's satisfaction (and viceversa)
  - Influenced by your personality and the social relationships with the other group members
- **Conformity**
  - *Normative influence:* you want to be part of the group
  - *Informational influence:* opinion changes because you believe the group must be right.

Group Recommendation Model



T. N. Nguyen and F. Ricci: Dynamic Elicitation of User Preferences in a Chat-Based Group Recommender System. SAC 2017

Preference updating

- Users in the group have an initial utility function

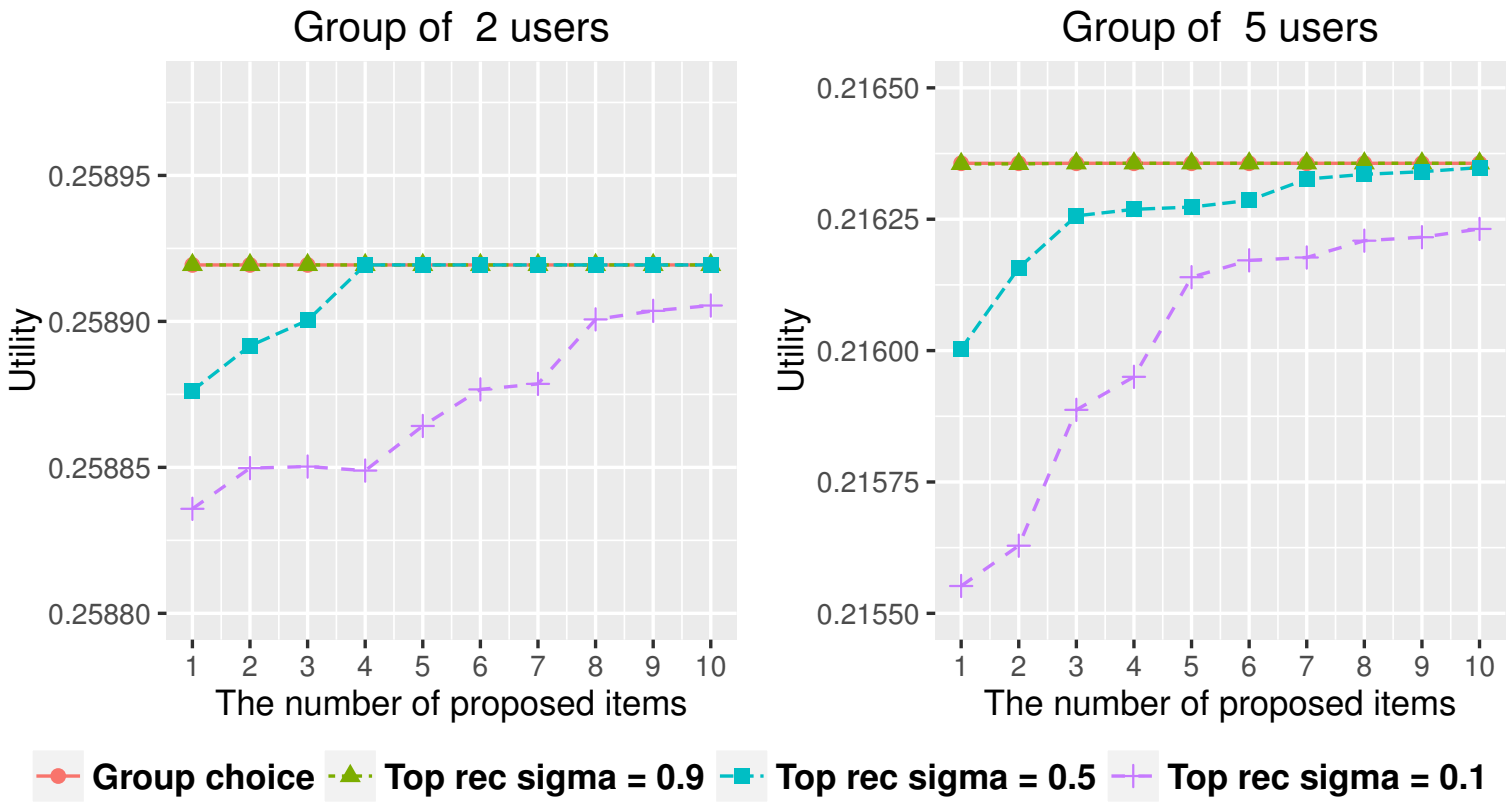
$$U(u,i) = \sum_{j=1}^{n} w_j^{(u)} x_j^{(i)}$$

- $w_j^{(u)}$ are the user weights, $x_j^{(i)}$ are the item features
- When group members interact in a discussion evaluations of discussed items reveal new preference constraints
  - I like item *i* more than item *j: U(u,i) > U(u,j)*
- Search for *U(u, i, g)*, defined by a vector $\boldsymbol{w}^{(u)}{}_g$, that satisfies the constraints expressed during the group discussion
- Combine the two utilities linearly: s $\mathbf{w}^{(u)}$ + (1-s) $\boldsymbol{w}^{(u)}{}_g$

T. N. Nguyen and F. Ricci: Dynamic Elicitation of User Preferences in a Chat-Based Group Recommender System. SAC 2017

Simulations

- Assuming that the group has **no influence** on user preferences



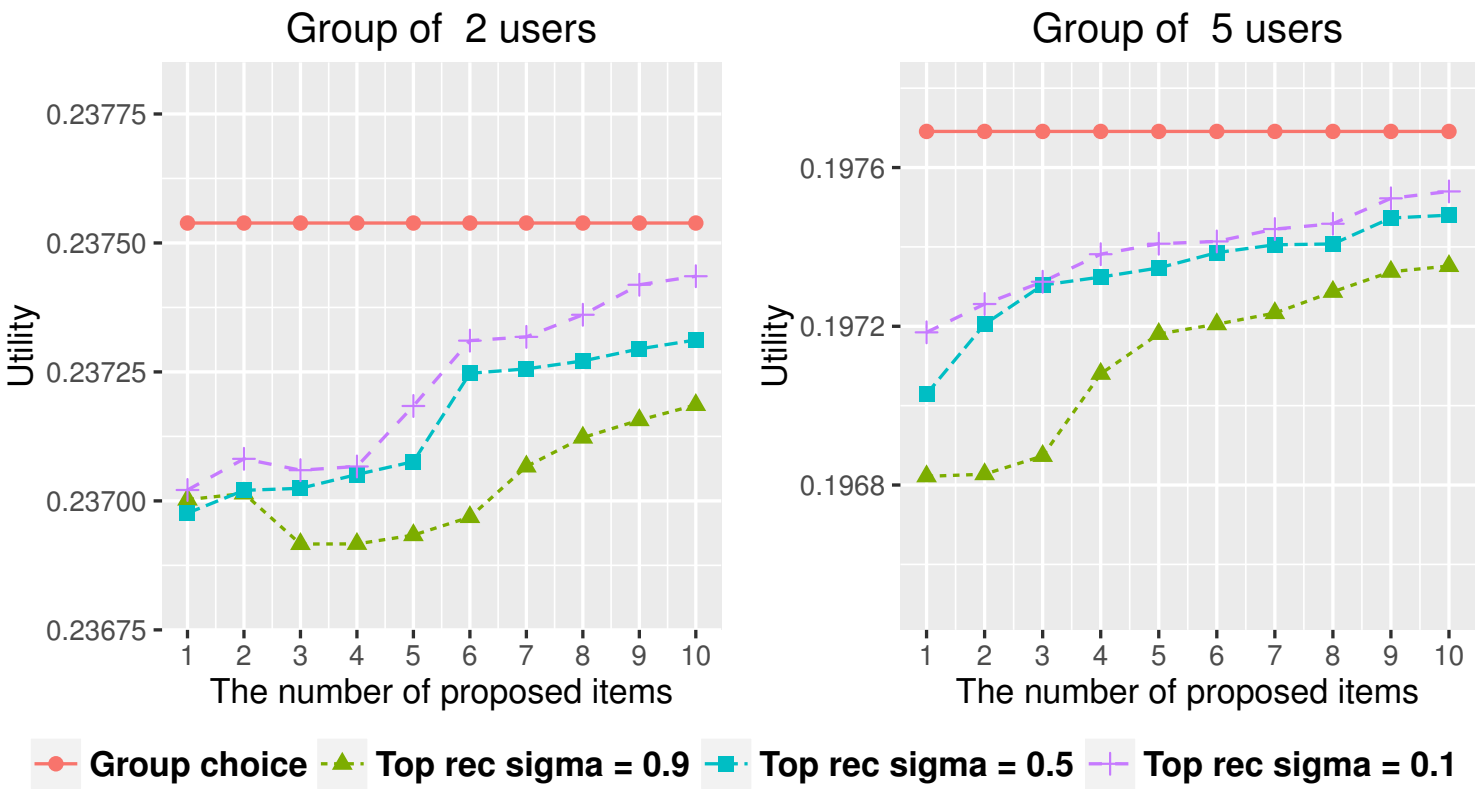
RS weighs more the long-term preferences

Mixture

RS weighs more the short-term preferences

Simulations

- Assuming that the group induces the group members to **differentiate** their preferences



RS weighs more the long-term preferences

Mixture

RS weighs more the short-term preferences

Group Dynamics

- Depending on the **group context** – i.e., the group is converging or diverging – the system must use a **different preference model**

Lesson Learned

- Preferences are **contextual**, **dynamic** and **hard to predict**
- Predicting preferences does not suffice for supporting decision making with recommendations - **choice model**
- **Preference dynamics** is important to monitor to identify better preference elicitation and recommendation techniques
- **Group** recommendations is a challenging domain for testing new technics facing the above mentioned issues.

Thanks

- In particular to my students and collaborators who contributed to develop these ideas:
  - David Massimo
  - Linas Baltrunas
  - Laura Bledaite
  - Marius Kaminskas
  - Marko Gasparic
  - Marko Tkalcic
  - Matthias Braunhofer
  - Mehdi Elahi
  - Saikishore Kalloori
  - Tural Gurbanov
  - Thuy Ngoc Nguyen