Interacting with Recommender Systems

Dietmar Jannach, Ingrid Nunes, Michael Jugovac

About the Presenters

- Dietmar Jannach
  - TU Dortmund, Germany
- Ingrid Nunes
  - UFRGS, Porto Alegre, Brazil
- Michael Jugovac
  - TU Dortmund, Germany

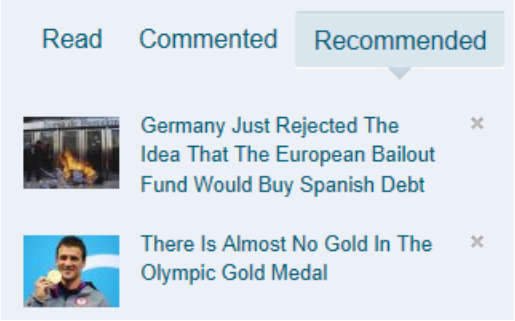
About you

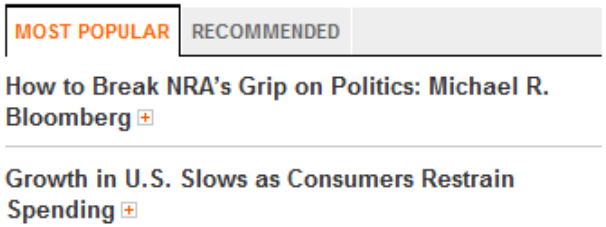
Agenda

- 09:00 – 09:45
  - Introduction & Background
- 09:45 – 10:14
  - Interacting recommender systems – A review (Part I)
- 10:15 – 10:45
  - Coffee break
- 10:45 – 11:15
  - Interacting recommender systems – A review (Part II)
- 11:15 – 12:15
  - Explanations in recommender systems
  - Discussion

Recommender Systems

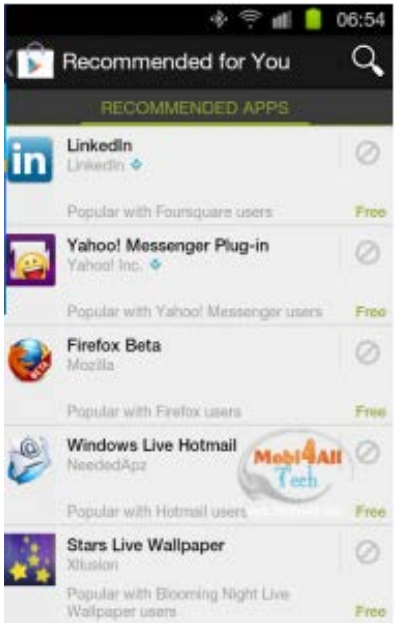
- Application areas

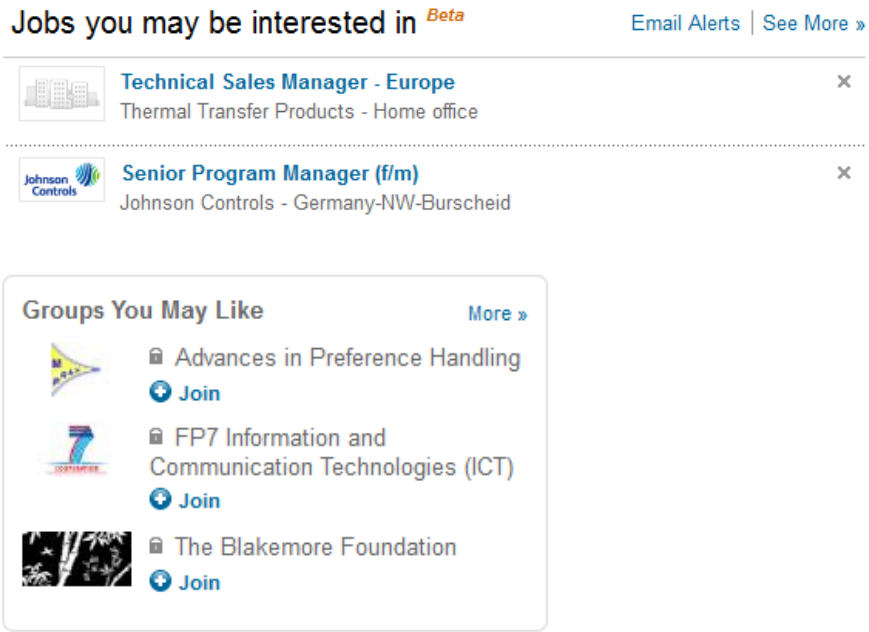


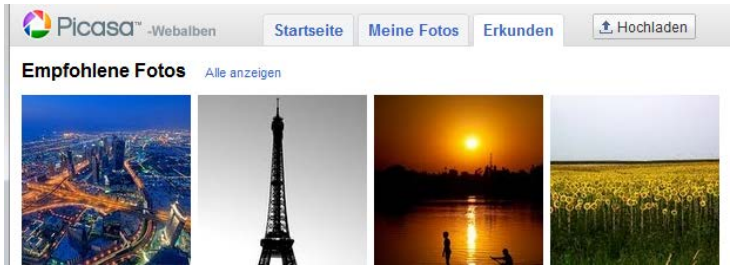


Sources: Amazon.com, Tripadvisor.com, LinkedIn.com, Forbes.com

In the Social Web







Sources: Google Play, LinkedIn.com, Picasa.com

Recommender Systems (RS)

- A pervasive part of our daily user experience
  - Customers who bought … also bought
  - You may also like
  - Similar
  - Recommended / Suggested for you
  - People you may know
  - Who to follow
  - Others also liked
  - Jobs you may be interested in
  - Groups that you may like
  - Trending in your area

User Interaction with RS

- Which user interaction?
  - The system monitors what I do.
  - And then shows me stuff.
  - Which I can click on.

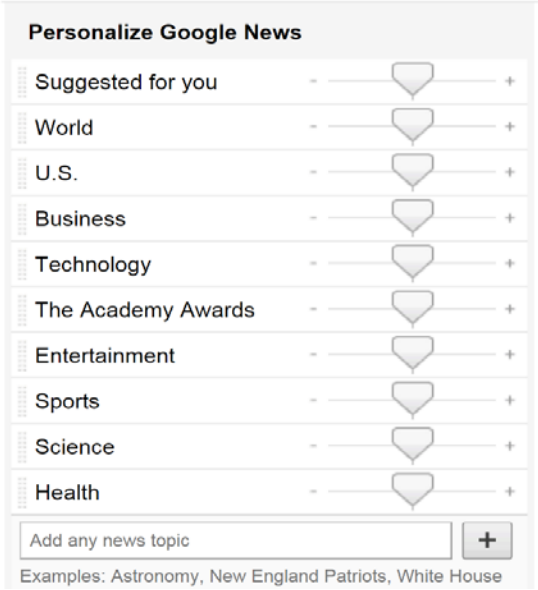
Source: Amazon.com

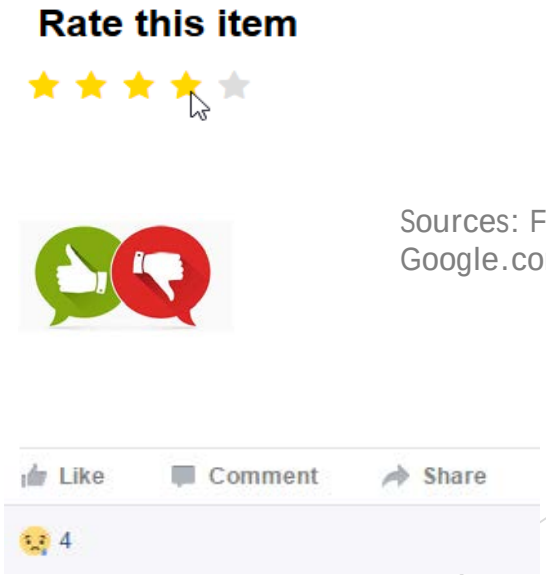
Exercise

- Think of possible ways of interacting with a recommender system

Design Space Examples

- Telling the system explicitly what you like
  - Global settings
  - Ratings
    - But how many options? How many categories?





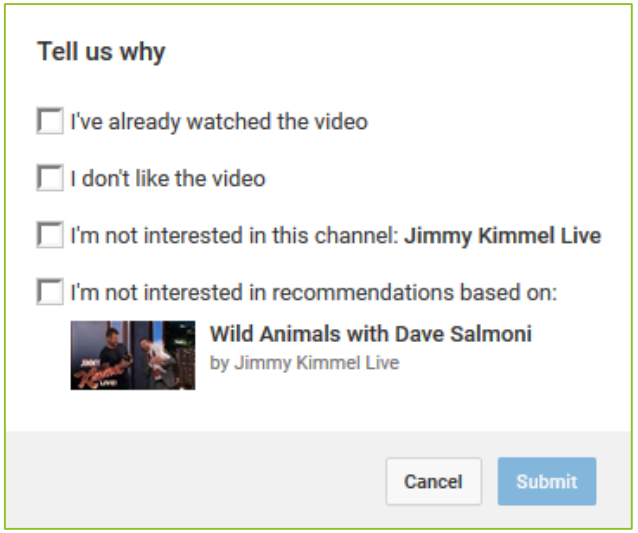
Sources: Facebook.com, Google.com

Design Space Examples

- Different ways of showing you recommendations
  - But how many items to show?
    - One item only?
    - Two, three, or more items?
  - More than one list?
    - How many lists?
    - In which order?
  - Where on the screen?
    - At bottom, on the side, on top?
  - When to recommend?
    - Always? Upon request? As notification?

Design Space Examples

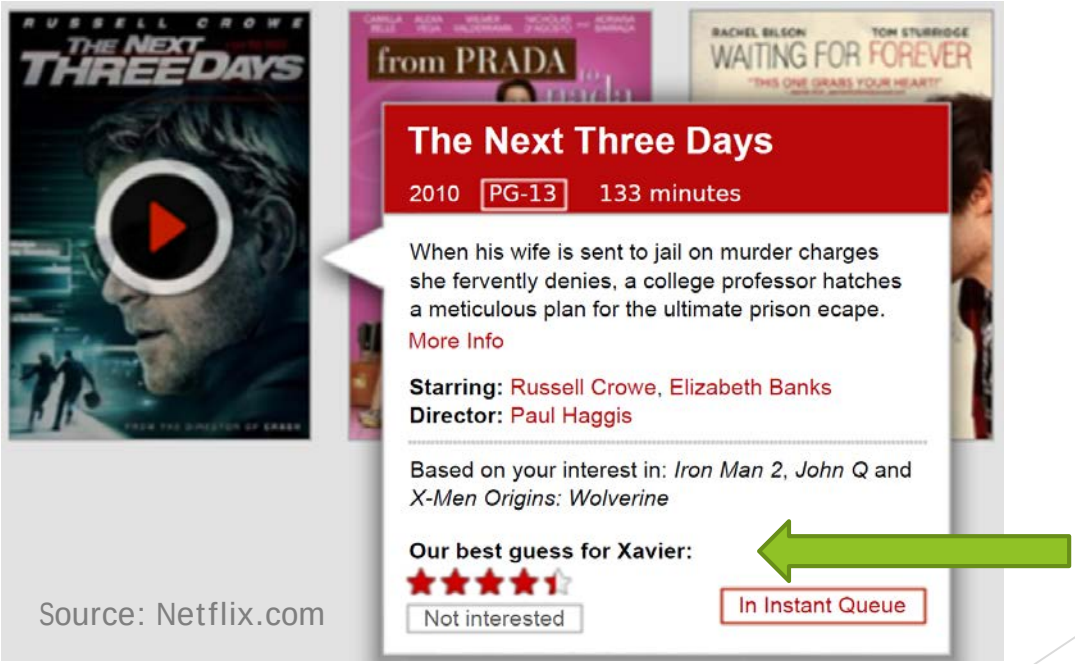
- What to display (in addition to a nice picture)?
  - Just the product/item info?
  - Feedback options?
    - Like /Dislike?
    - Or more?



Source: Youtube.com

Design Space Examples

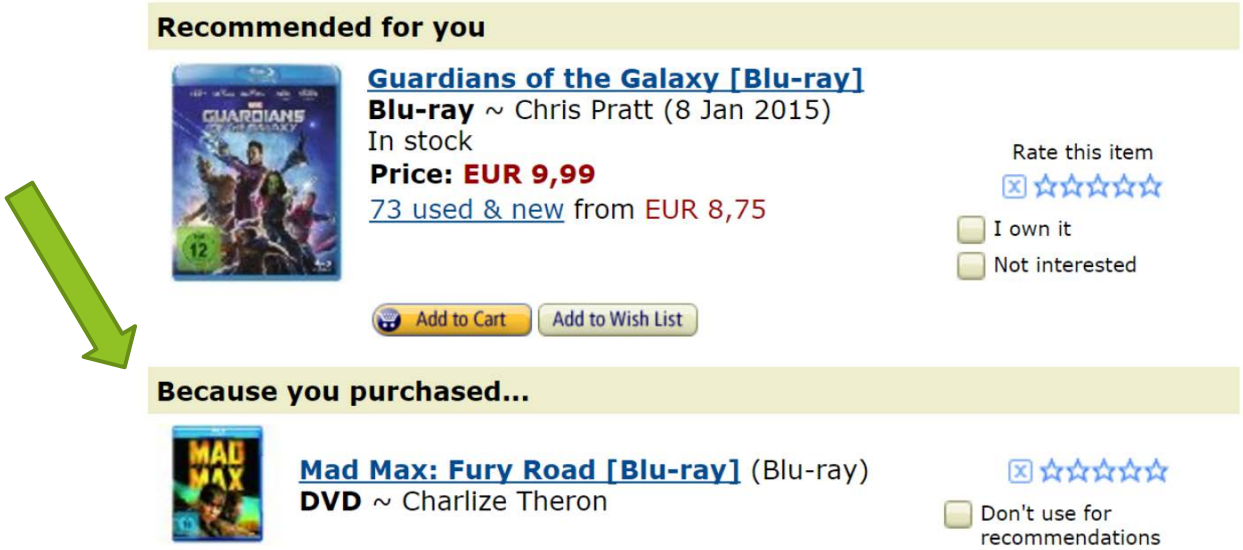
- What to display (in addition to a nice picture)?
  - Maybe some explanation, but which one?
  - A predicted rating?



Source: Netflix.com

Design Space Examples

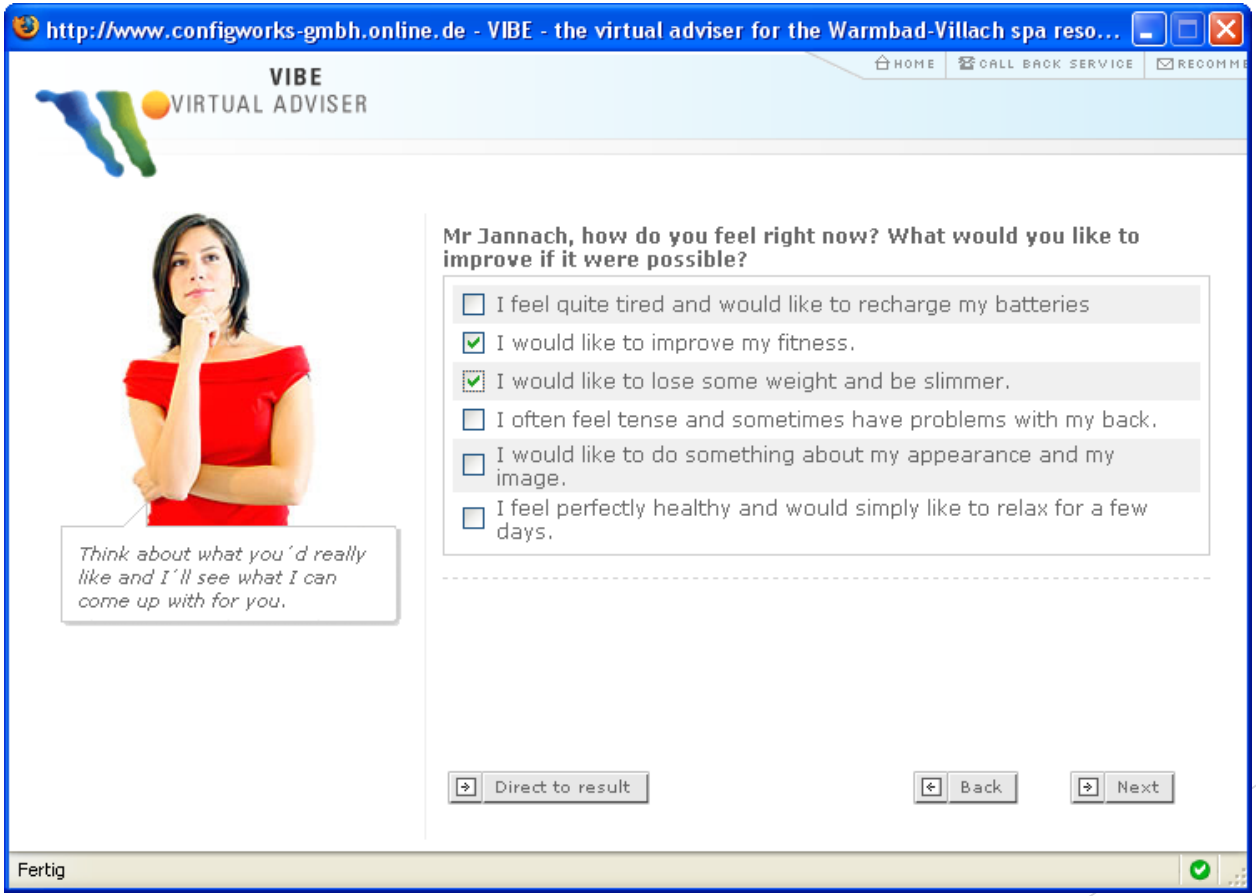
- What to display (in addition to a nice picture)?
  - Maybe some explanation, but which one?
  - Or our logic to recommend this?



Source: Amazon.com

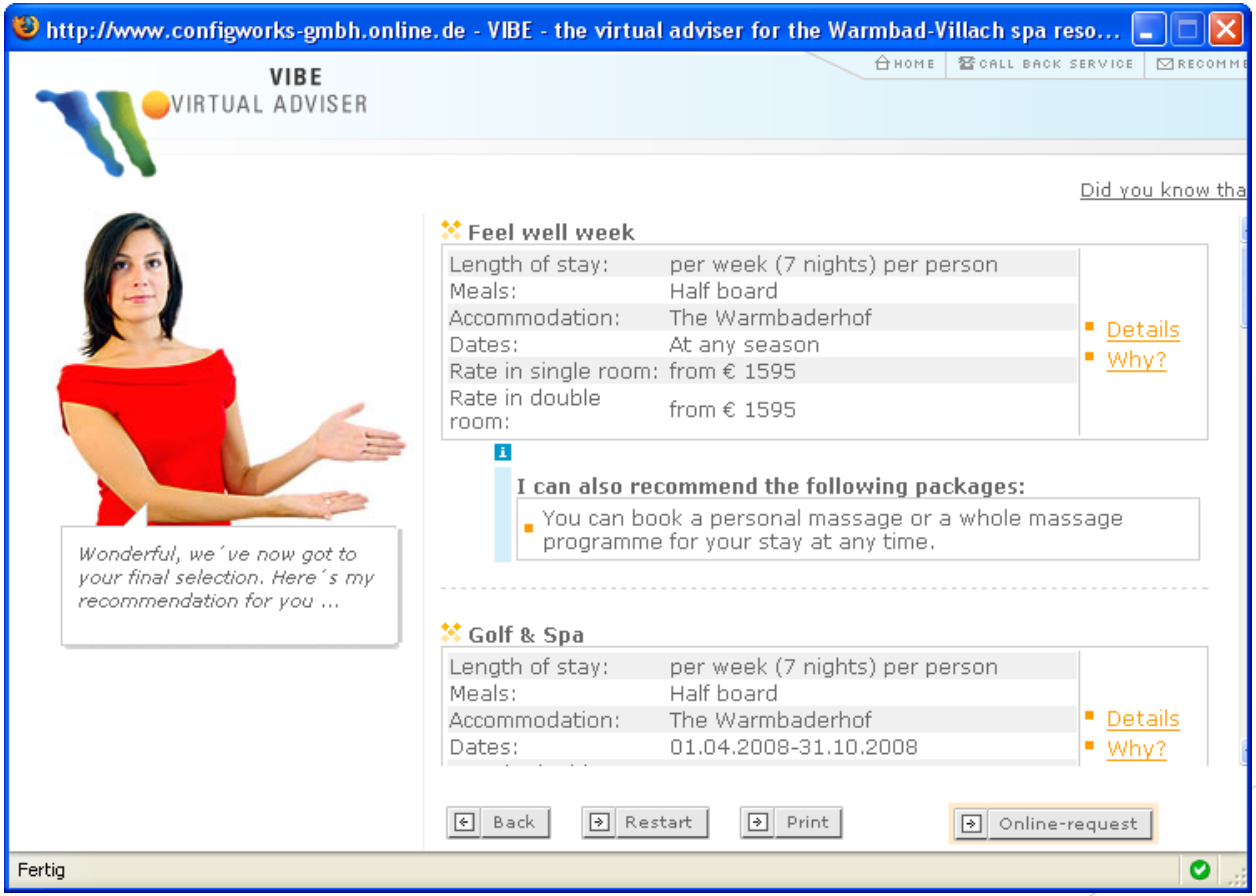
Design Space Examples

- It could even be a dialog



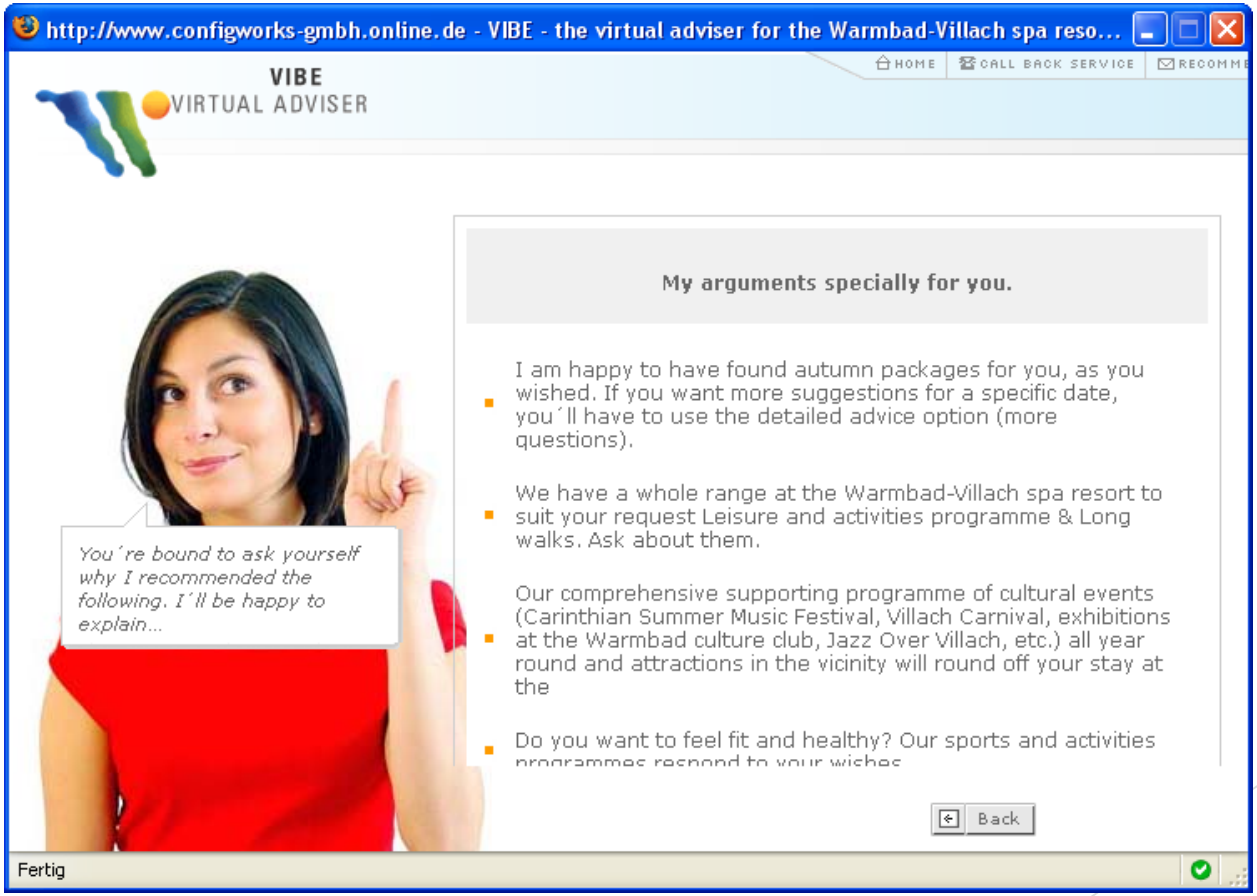
Design Space Examples

- It could even be a dialog



Design Space Examples

- It could even be a dialog



Recommendation Approaches

- Which kind of interactions you can support depends on the underlying algorithm(s)
  - If you show preference predictions, you must predict ratings and not just rank items
  - If you use a complex prediction algorithm, providing explanations can be challenging
  - If you support feedback on recommendations, you should be able to immediately consider it
  - If you want to interact with users with a chatbot, many more challenges will emerge
- Basic background knowledge and terminology of RS will be presented next

Background on Recommenders

- Topics
  - Why using recommender systems at all?
  - What are the basic technical approaches?
    - Collaborative filtering
    - Content-based filtering
    - Knowledge-based approaches
    - Hybrids

Why Using Recommenders?

- Value for the customer
  - RS helps user find things that are interesting
  - RS helps user narrow down the set of choices
  - RS helps user explore the space of options
  - RS helps user discover new things, entertainment
  - …
- Value for the provider
  - Increased sales, click trough rates, conversion etc.
  - Increased trust and customer loyalty
  - More opportunities for promotion, persuasion
  - More knowledge about customers
  - …

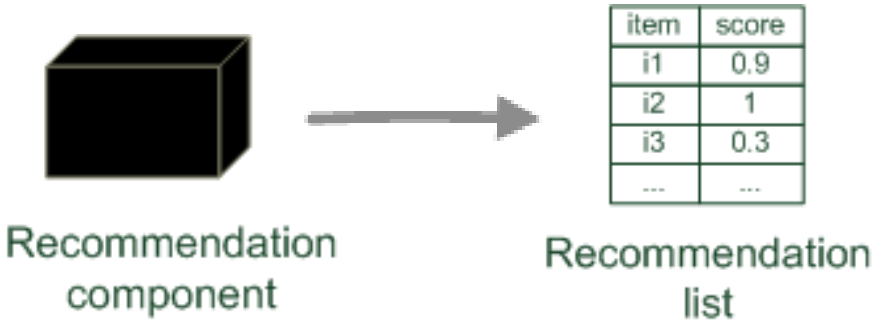
Jannach, D. and Adomavicius, G.: Recommendations with a Purpose, ACM Conference on Recommender Systems, RecSys 2016, Boston, MA, pp. 7-10

A Possible Algorithmic Task

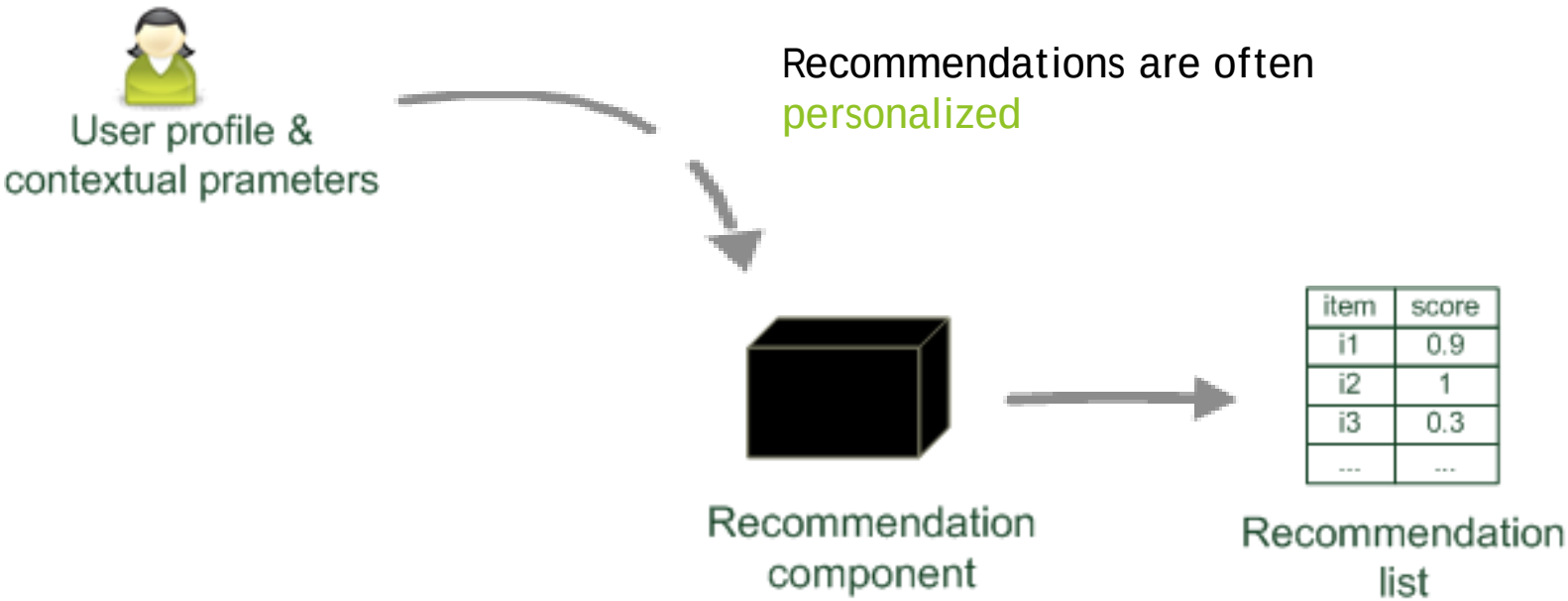
- Given
  - The profile of the "active" user and possibly some situational context
- Compute
  - A relevance (ranking) score for each recommendable item
- The profile …
  - … can include past user ratings (explicit or implicit), demographics and interest scores for item features
- The problem …
  - … is to learn a function that predicts the relevance score for a given (typically unseen) item

Recommendation Paradigms

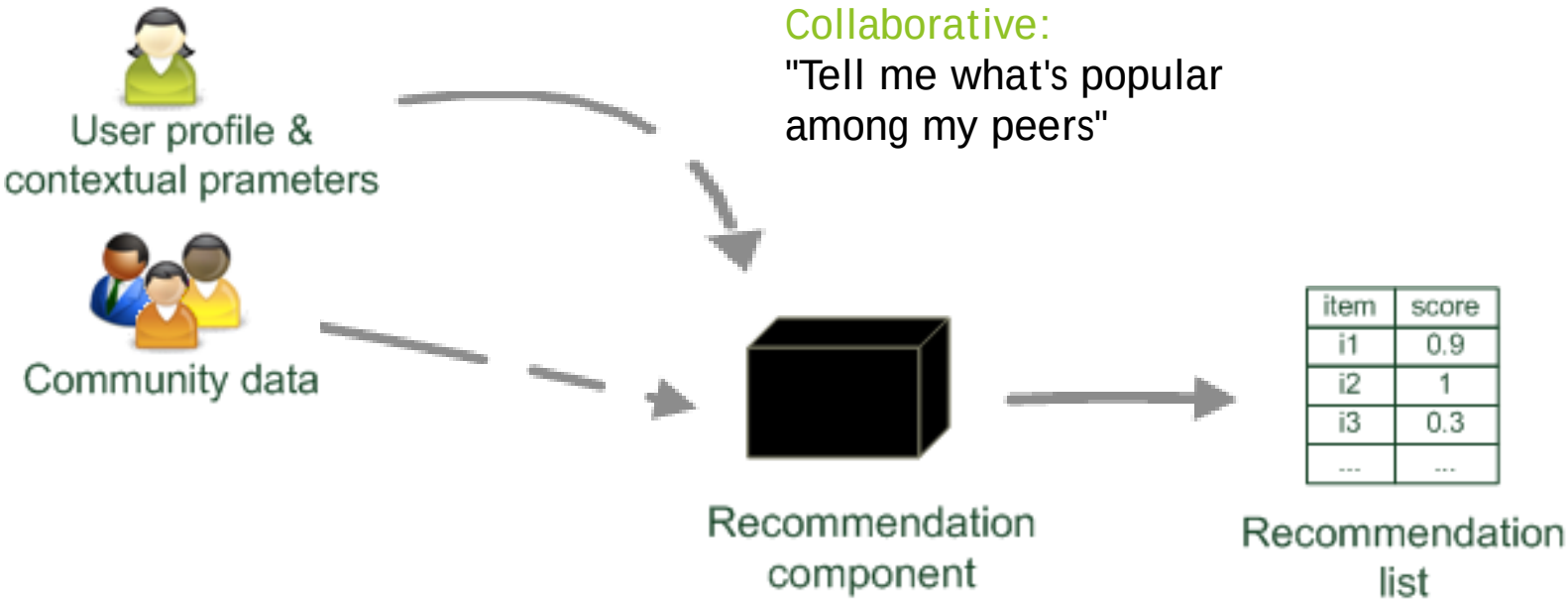
Recommender systems reduce information overload by estimating relevance



Recommendation Paradigms

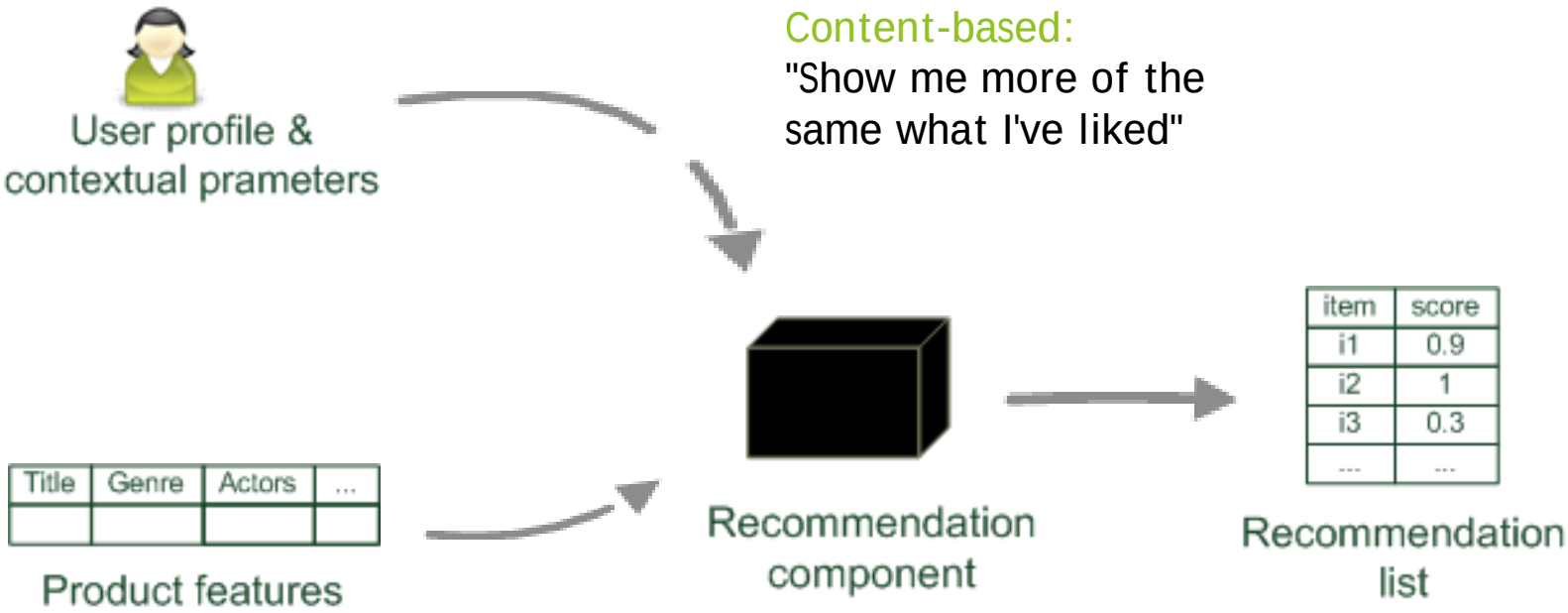


Recommendation Paradigms



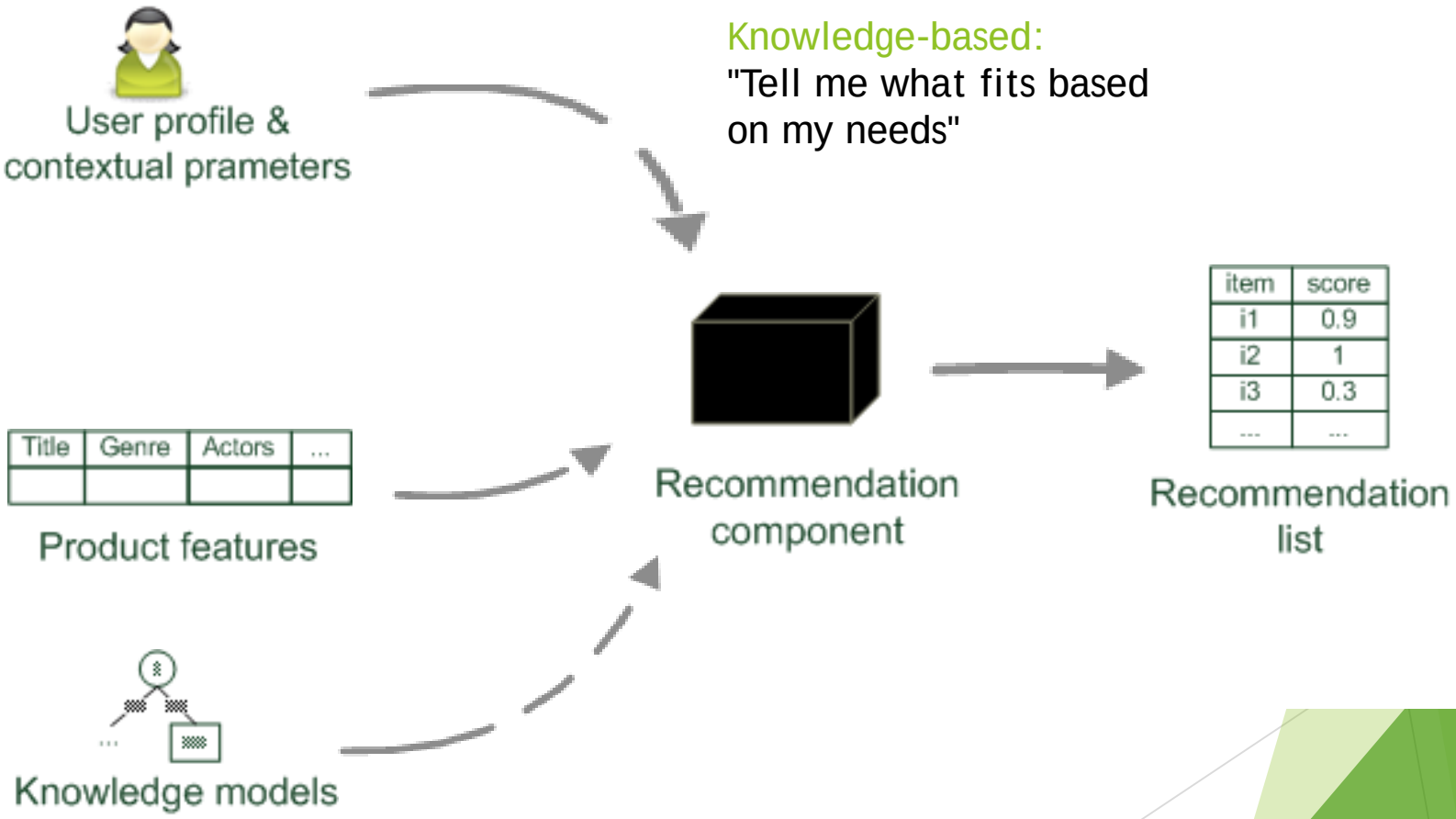
Collaborative:
"Tell me what's popular among my peers"

Recommendation Paradigms

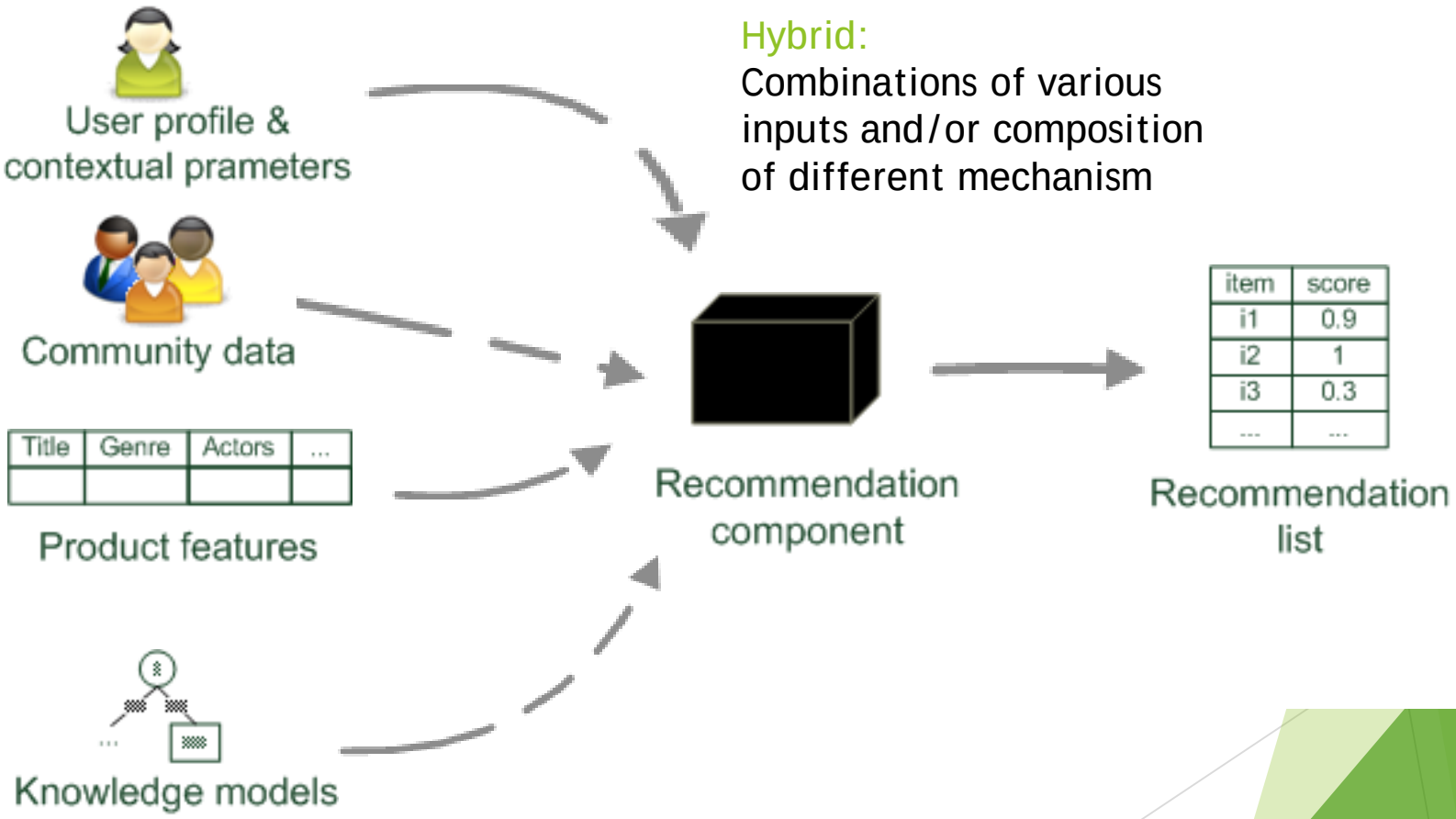


Content-based:
"Show me more of the same what I've liked"

Recommendation Paradigms



Recommendation Paradigms



Hybrid:
Combinations of various inputs and/or composition of different mechanism

Collaborative Filtering

- The most prominent approach to generate recommendations
  - Used by large commercial e-commerce sites
  - Well-understood, various algorithms and variations exist
  - Applicable in many domains
- Approach
  - Use the preference patterns of a community to recommend items
  - One first algorithmic solution
    - Find users that are similar to the current one
    - Recommend what these other users liked

K-Nearest-Neighbors (kNN)

- A common (academic) problem setup
  - Given a matrix of explicit or implicit preferences of users for items
  - Predict the missing cells

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

K-Nearest-Neighbors

- Find k users that are similar to Alice
  - How many neighbors? How to determine similarity?
- User their rating for Item5 to predict Alice's rating
  - How to combine the ratings?

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

Similarity Computation (kNN)

- Using Pearson's correlation coefficient to estimate preference similarity between users

$$sim(a,b) = \frac{\sum_{p \in P}(r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P}(r_{a,p} - \bar{r}_a)^2}\ \sqrt{\sum_{p \in P}(r_{b,p} - \bar{r}_b)^2}}$$

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

sim = 0,85

sim = 0,70

sim = -0,79

Rating Prediction (kNN)

- A common prediction function

$$pred(a,p) = \overline{r_a} + \frac{\sum_{b \in N} sim(a,b) * (r_{b,p} - \overline{r_b})}{\sum_{b \in N} sim(a,b)}$$

  - Takes Alice's average rating into account
  - Takes the similarity degree of the neighbors into account

- Many variations of this prediction scheme exist
  - Including an "item-based" approach, where we look for items that received similar ratings

Generalized Problem

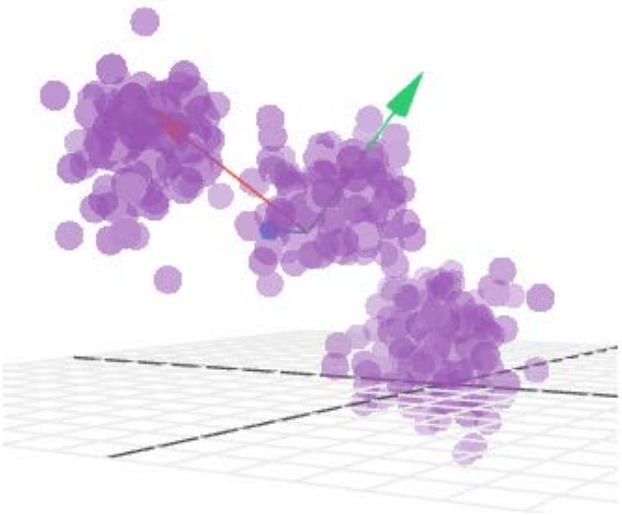
- Rating prediction (and recommendation) reduces to a matrix completion problem
- The general optimization problem:
  - Given a set of noisy rating observations (x,y), learn a function f(x) = y that predicts unknown y values for a given x.
  - Optimization goals: Minimize the prediction error, avoid overfitting to given data
- A large variety of machine learning approaches to estimate the function were proposed

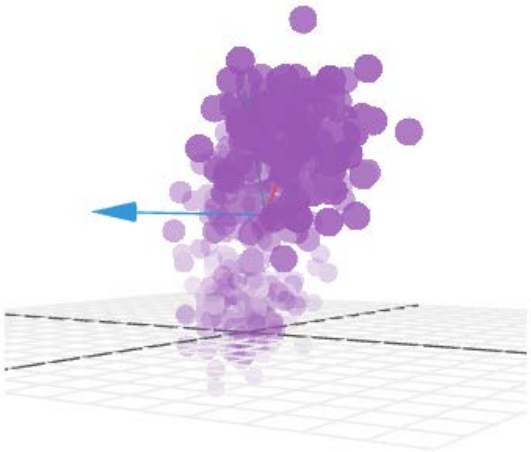
Jannach, D., Resnick, P., Tuzhilin, A., Zanker, M.: Recommender Systems - Beyond Matrix Completion, Communications of the ACM, Vol. 59 (11), 2016 pp. 94-102

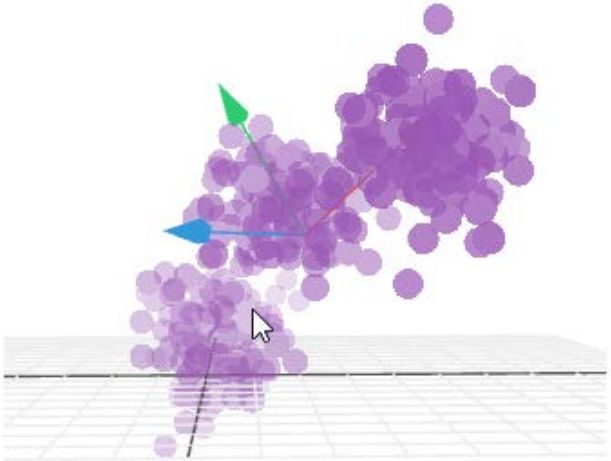
Matrix Factorization (MF)

- Became popular during the Netflix Prize competition
- Adopt ideas of Principal Component Analysis (PCA)
  - Given a large dataset with values for many and possibly correlated variables
  - Convert them to a much smaller set of values of linearly uncorrelated variables (principal components)
  - The first principal component has the largest variance
  - The next component has the highest variance, but has to be orthogonal to the preceding components (i.e., capture a different aspect).

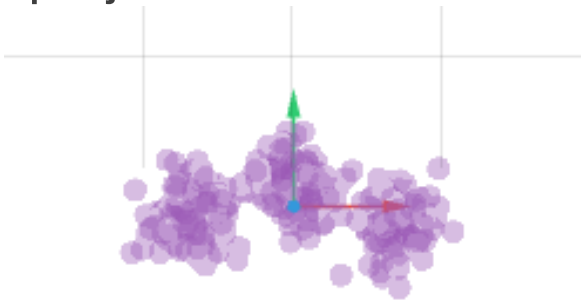
- 3D – Not much information from certain angles







- Not much information lost in 2D projection



Visualization from http://setosa.io/ev/principal-component-analysis/

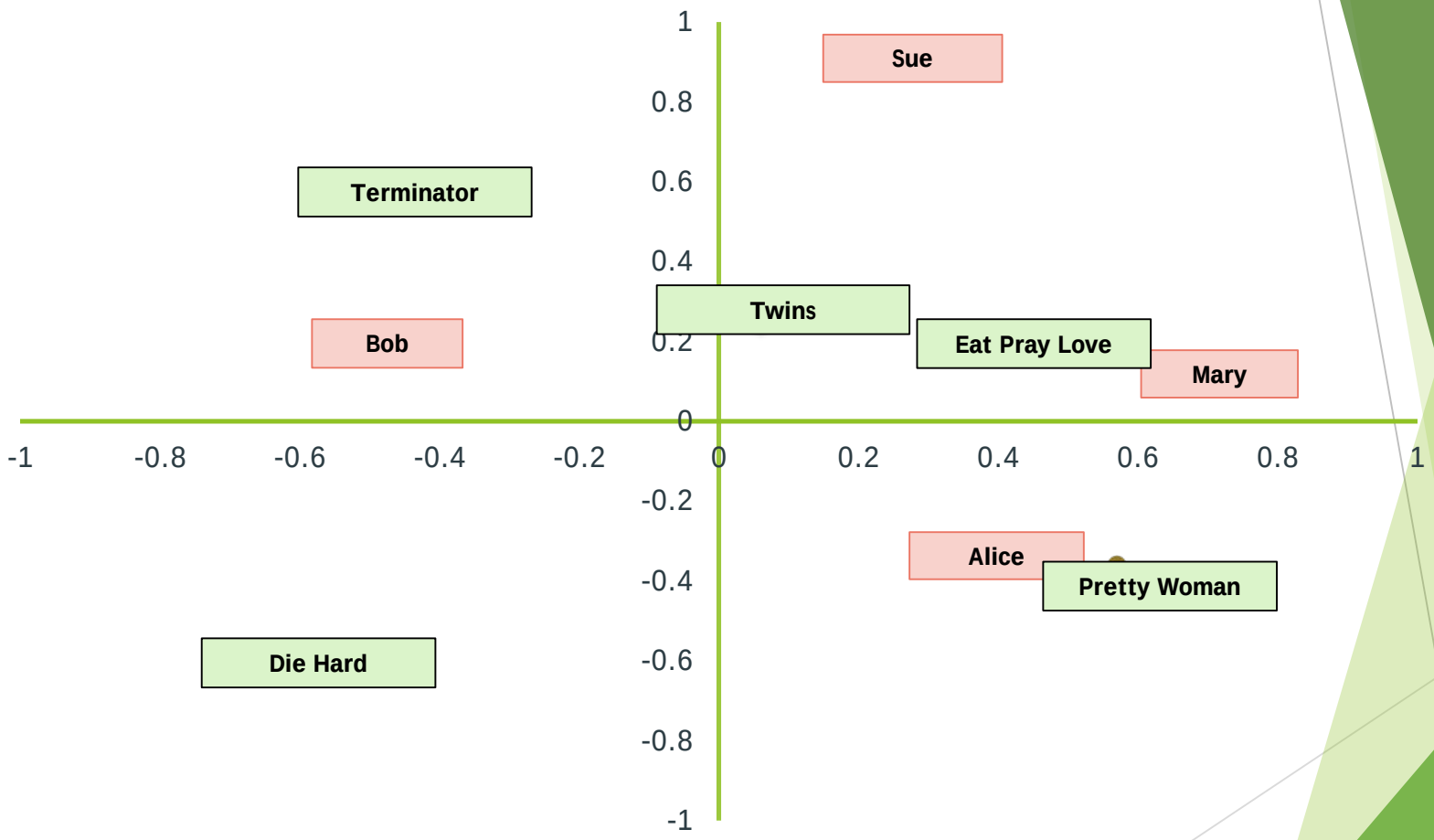
Singular Value Decomposition

- Technical approach – Singular Value Decomposition (SVD)
  - Decompose given matrix M as follows

$$M = U \times \Sigma \times V^T$$

- Only retain the most important signals (largest sing. values)
  - Original matrix is approximated (removes noise, uncovers latent relationships)

Projection into 2D Space



- Problem: We do not know what the dimensions are

Matrix Factorization (MF)

- In recommender systems
  - Project users and items in the same “latent space”
- Technical approach in recommender systems
  - Use only two-matrix decomposition
  - Use an iterative approximation approach, e.g., based on gradient descent
  - Number of latent factors usually set between 50 and 200, to maximize accuracy
  - Combine with additional factors (biases)
- Related to Probabilistic Latent Semantic Analysis
  - Finding latent relationships between concepts in text documents

SVD-based Recommendation

- U and V correspond to latent user and item vectors

$$M_k = U_k \times \Sigma_k \times V_k^T$$

U_k	Dim1	Dim2
Alice	0.47	-0.30
Bob	-0.44	0.23
Mary	0.70	-0.06
Sue	0.31	0.93

V_k^T	Terminator	Die Hard	Twins	Eat Pray Love	Pretty Woman
Dim1	-0.44	-0.57	0.06	0.38	0.57
Dim2	0.58	-0.66	0.26	0.18	-0.36

Σ_k	Dim1	Dim2
Dim1	5.63	0
Dim2	0	3.23

$$\hat{r}_{ui} = \bar{r}_u + U_k(Alice) \times \Sigma_k \times V_k^T(EPL)$$

= 3 + 0.84 = 3.84

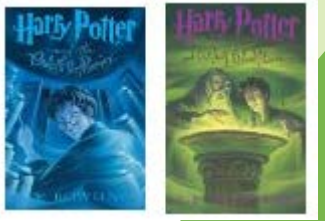
Collaborative Filtering

- Summary
  - Largely used in real-world systems
  - Do not require information about the items or users
  - However, require existence of a user community
  - Many, many algorithms proposed in the literature
    - Including ones that optimize the ranking and not the predictions
    - Most learned “models” (in the machine learning sense) cannot be easily interpreted

Content-based (CB) Filtering

- Again:
  - Determine preferences of user based on past behavior
  - Alternative preference acquisition methods
    - ask the user, look at recently viewed items
- This time, however:
  - Look at what the current user liked (purchased, viewed, …)
  - Estimate the user's preference for certain item features
    - e.g., genre, authors, release date, keywords in the text

Source: Amazon.com



What is the “Content”?

- CB-recommendation techniques were often applied to recommend text documents
  - E.g., web pages or newsgroup message
- The term content however in many applications refers to meta-data
  - E.g., the author of a book, the genre of a movie

Title	Genre	Author	Type	Price	Keywords
The Night of the Gun	Memoir	David Carr	Paperback	29.90	Press and journalism, drug addiction, personal memoirs, New York
The Lace Reader	Fiction, Mystery	Brunonia Barry	Hardcover	49.90	American contemporary fiction, detective, historical
Into the Fire	Romance, Suspense	Suzanne Brockmann	Hardcover	45.90	American fiction, murder, neo-Nazism

Recommendation Approach

- Represent items and users in the same way

Title	Genre	Author	Type	Price	Keywords
The Night of the Gun	Memoir	David Carr	Paperback	29.90	Press and journalism, drug addiction, personal memoirs, New York
The Lace Reader	Fiction, Mystery	Brunonia Barry	Hardcover	49.90	American contemporary fiction, detective, historical
Into the Fire	Romance, Suspense	Suzanne Brockmann	Hardcover	45.90	American fiction, murder, neo-Nazism

Title	Genre	Author	Type	Price	Keywords
…	Fiction	Brunonia, Barry, Ken Follett	Paperback	25.65	Detective, murder, New York

- A simple method
  - Compute the similarity of an unseen item with the user profile based on the keyword overlap (Dice coefficient)
  - Or use and combine multiple metrics

$$\frac{2 \times |keywords(b_i) \cap keywords(b_j)|}{|keywords(b_i)| + |keywords(b_j)|}$$

TF-IDF Encoding

- Simple keyword representation has its problems
  - In particular when automatically extracted:
    - Not every word has similar importance
    - Longer documents have a higher chance to have an overlap with the user profile
- Standard measure: TF-IDF
  - Encodes text documents in multi-dimensional Euclidian space
    - Weighted term vector
  - TF: Measures, how often a term appears
    - Assuming that important terms appear more often
    - Normalization has to be done in order to take document length into account
  - IDF: Aims to reduce the weight of terms that appear in all documents

Recommendation Approach

- Simple method
  - Every item is a vector of terms
  - Take the n most liked items of the user, compute an averaged TF-IDF vector
  - Compare this profile vector with all recommendable items
    - Use the cosine similarity as a similarity (distance) measure
  - Recommend the most similar items

Content-based Filtering

- Summary
  - Additional information about the items can be exploited (but must exist)
  - No large user community required
  - Recommends “more of the same”, no surprises

Knowledge-based Approaches

- Suitable for high-involvement items (cars, mobiles)
  - Cannot be based on few ratings alone, as too many details matter
  - Explicit and detailed preferences of the user
    - “The color of the car should be black"
  - More elaborate interaction mechanisms required
    - Conversational strategy needed

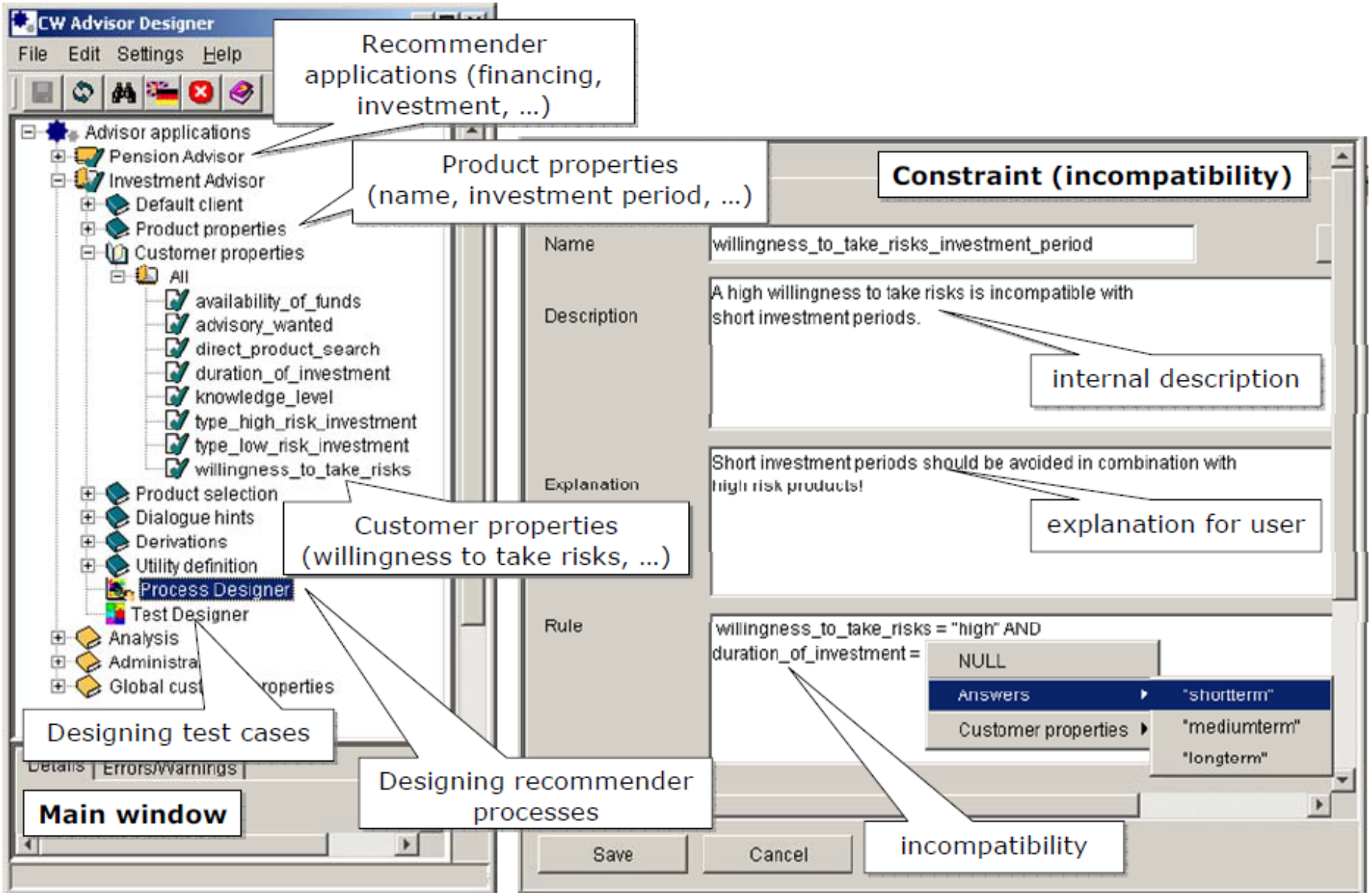
Critiquing

- An intuitive knowledge-based recommendation method
- Navigate the product space by "criticizing" the current solution
- Example:
  - Looking for a restaurant …
- Knowledge types:
  - About items
  - Adaptation step sizes
  - (Similarity functions)

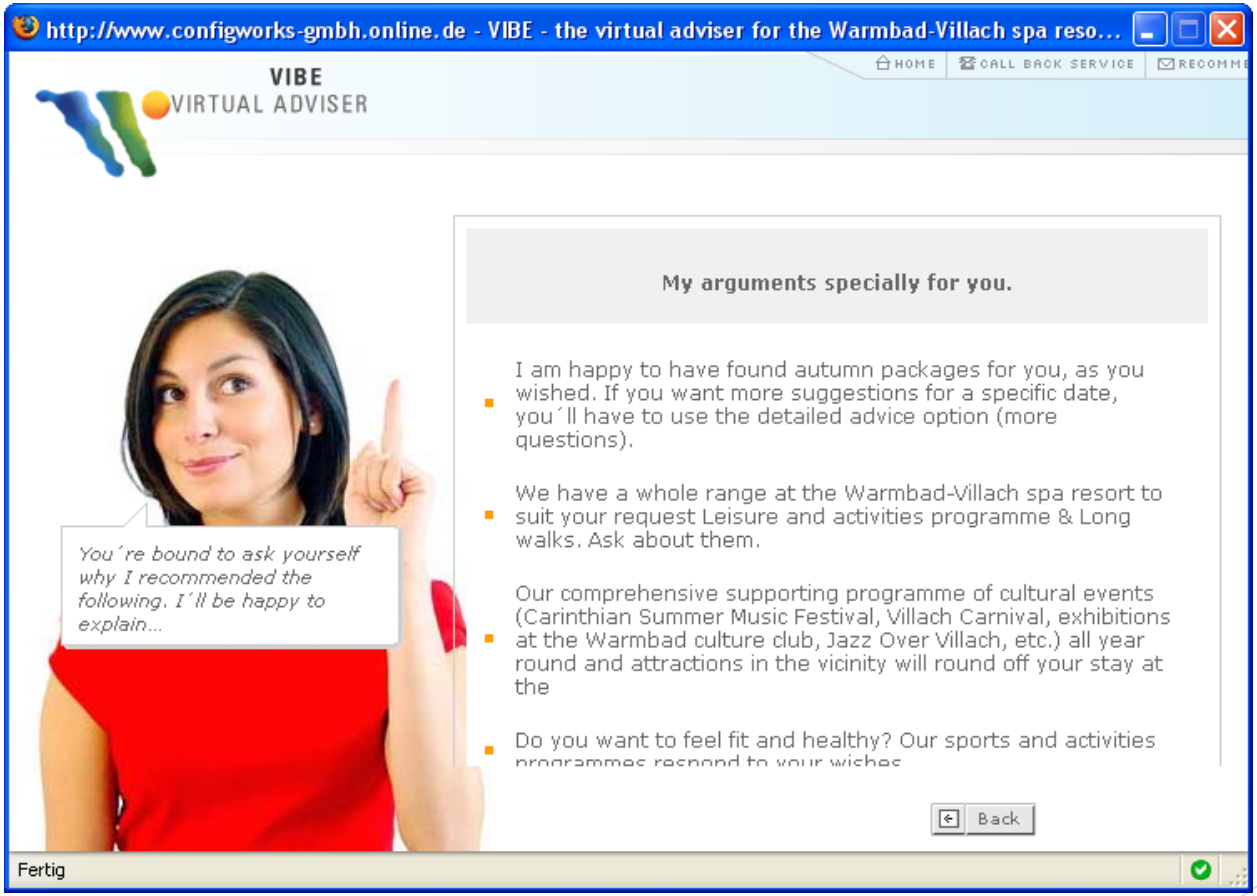
Knowledge-based Approaches

- Selection and ranking of items based on explicit knowledge
  - About user preferences
  - About item characteristics
  - About how to match user preferences with items
    - E.g., If user prefers luxury brands, only recommend items of manufacturers X, Y, and Z.
- Technical approaches
  - Constraints, rules, utility functions, case-based reasoning

Example software: Advisor Suite



Rules used for explanations



Hybrid Approaches

- Collaborative filtering, content-based filtering, knowledge-based recommendation
  - All pieces of information can be relevant in real-world recommendation scenarios, but all may have shortcomings
- Hybrid: Combine two or more approaches
  - Avoid some of the shortcomings
  - Different hybridization designs
    - Monolithic exploiting different features
    - Parallel use of several systems
    - Pipelined invocation of different systems

Summary

- Presented basic approaches to build recommender systems
  - Collaborative filtering, content-based filtering, knowledge-based approaches, hybrids
- Recommender systems operate on the basis of different types of data
  - Ratings
  - Information about items
  - Learned models of different complexities
  - Inference knowledge …
- These aspects determine which types of user interactions can be (easily) supported

Where are we?

09:00 – 09:45

- Introduction & Background

09:45 – 10:14

- Interacting recommender systems – A review (Part I)

10:15 – 10:45

- Coffee break

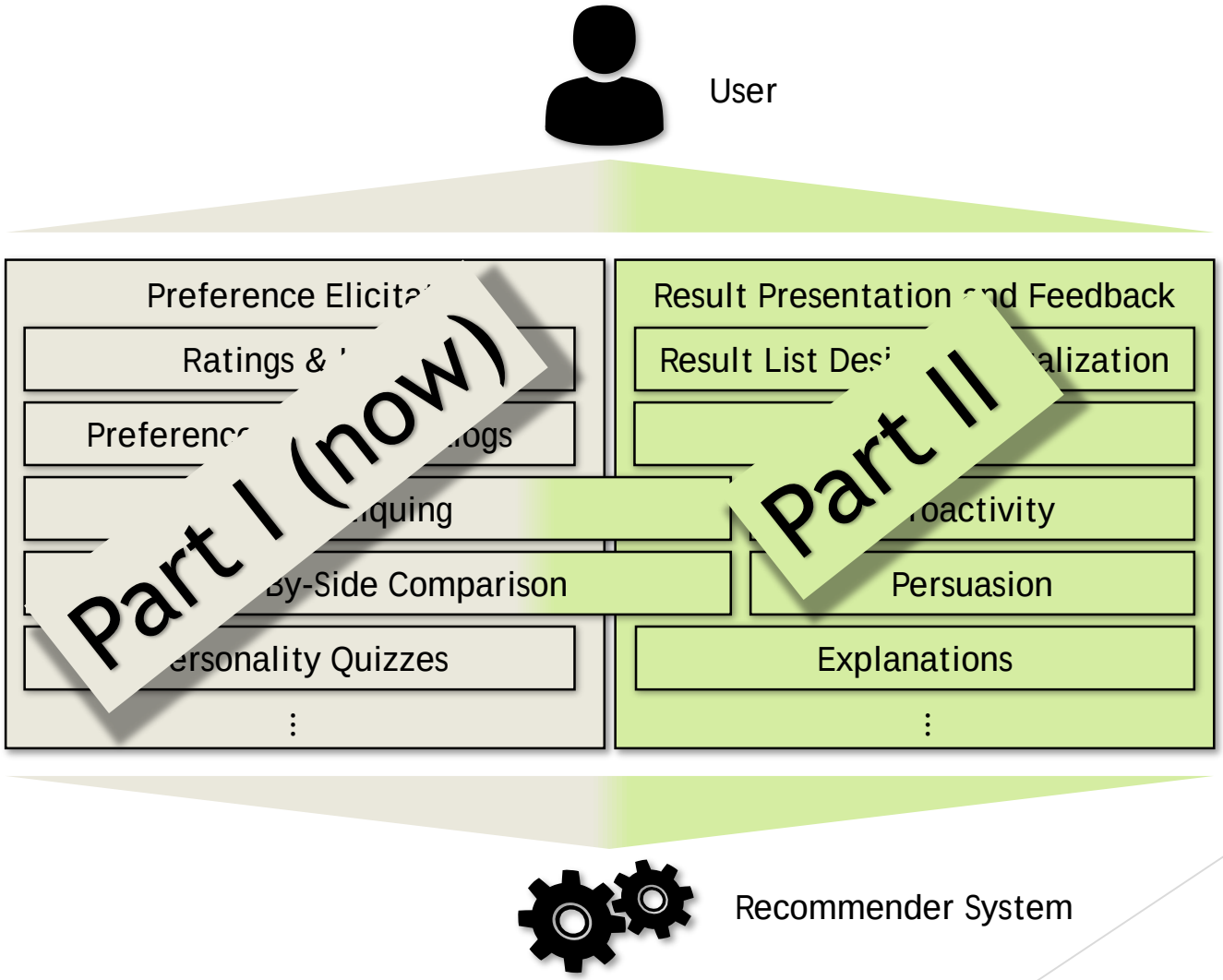
10:45 – 11:15

- Interacting recommender systems – A review (Part II)

11:15 – 12:15

- Explanations in recommender systems
- Discussion

Overview


User
Preference Elicita
Ratings &
Preferenc
ogs
quing
By-Side Comparison
ersonality Quizzes
Result Presentation and Feedback
Result List Des
alization
oactivity
Persuasion
Explanations
Part I (now)
Part II
Recommender System

Interacting with Recommender Systems

Part I: Interactions during Preference Elicitation

Exercise

- What methods do you prefer to input your preferences?

- 1. 
- 2. 
- 3. 

- 4. 
- 5. 
- 6. 

User Preference Elicitation

- Methods to …
  - let users express their tastes
  - acquire a user profile
    or guide users directly to the “right item”

- Requirements
  - Must not be tedious -> keep users interested!
  - As much useful information as possible
    in as few interaction steps as possible
  - Fun, engagement, basis of trust, reliability, …

Topics of This Session

- Explicit ratings and likes
- Preference forms and dialogs
- Critiquing
- Alternative elicitation techniques
  - Side-by-side comparison between items
  - Picture-based/tag-based elicitation
  - Personality quizzes

Explicit Ratings and Likes

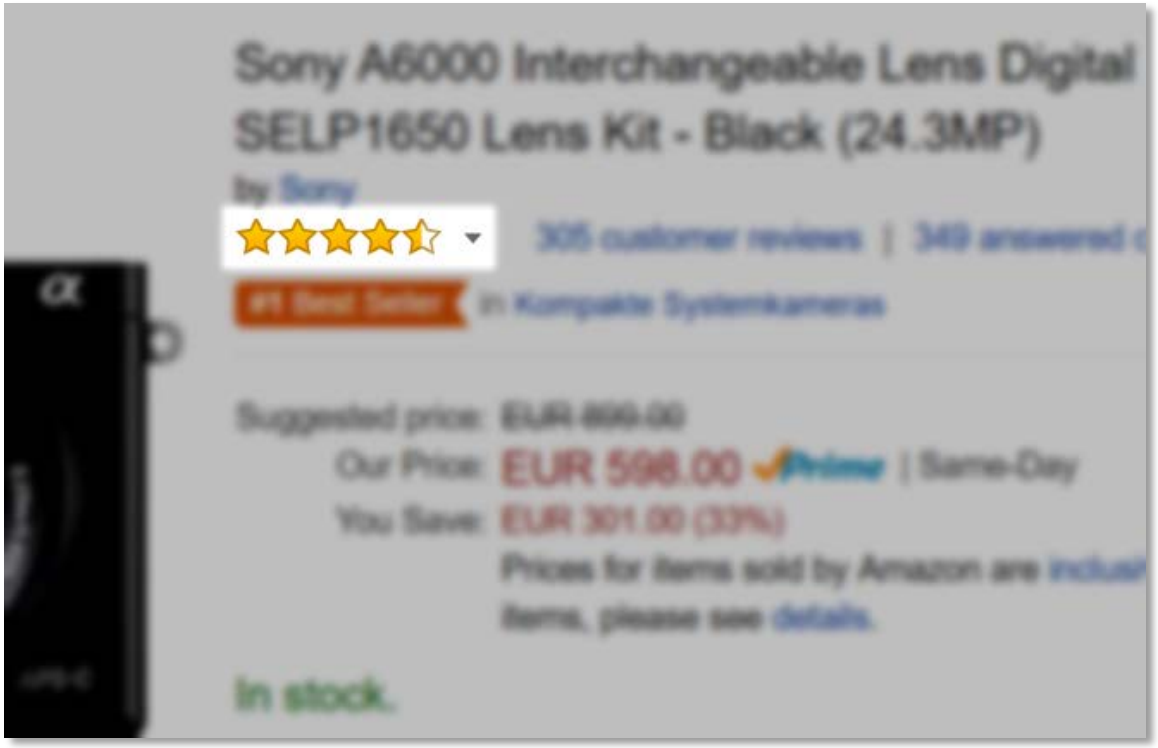
Explicit Ratings and Likes

- Our definition: Deliberate user interactions to tell the system about the preference toward an item

- Many approaches assume that ratings/likes are readily available
- But in reality, many problems arise:
  - Types of user input
  - User effort
  - Data reliability/granularity
  - Cold-start

Feedback Scale

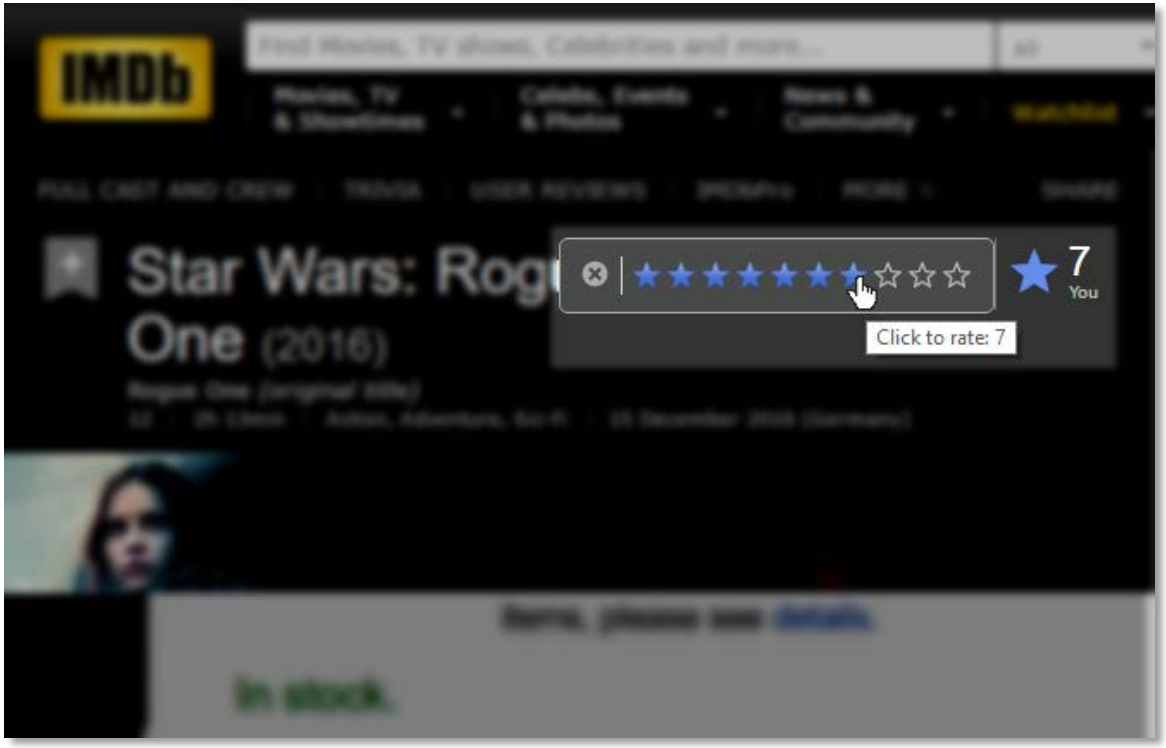
- How to make it as easy as possible to express taste?
- Common approaches:



Source: amazon.com

Feedback Scale

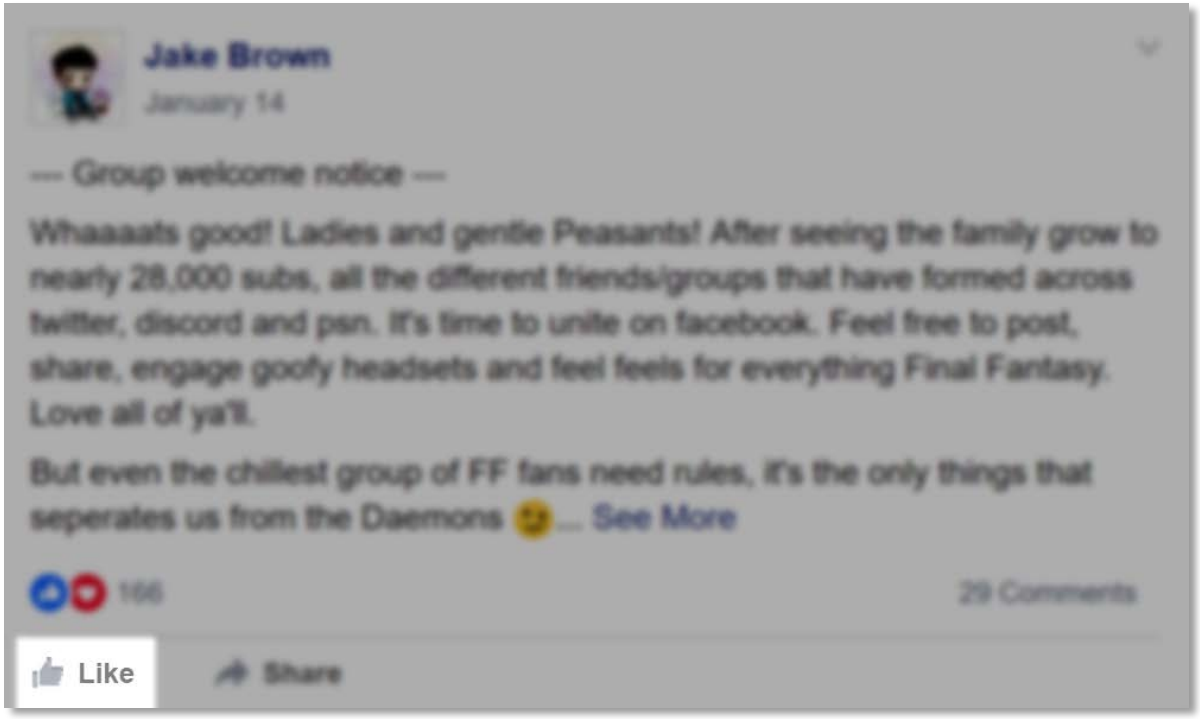
- How to make it as easy as possible express taste?
- Common approaches:



Source: imdb.com

Feedback Scale

- How to make it as easy as possible express taste?
- Common approaches:



Source: facebook.com

Feedback Scale

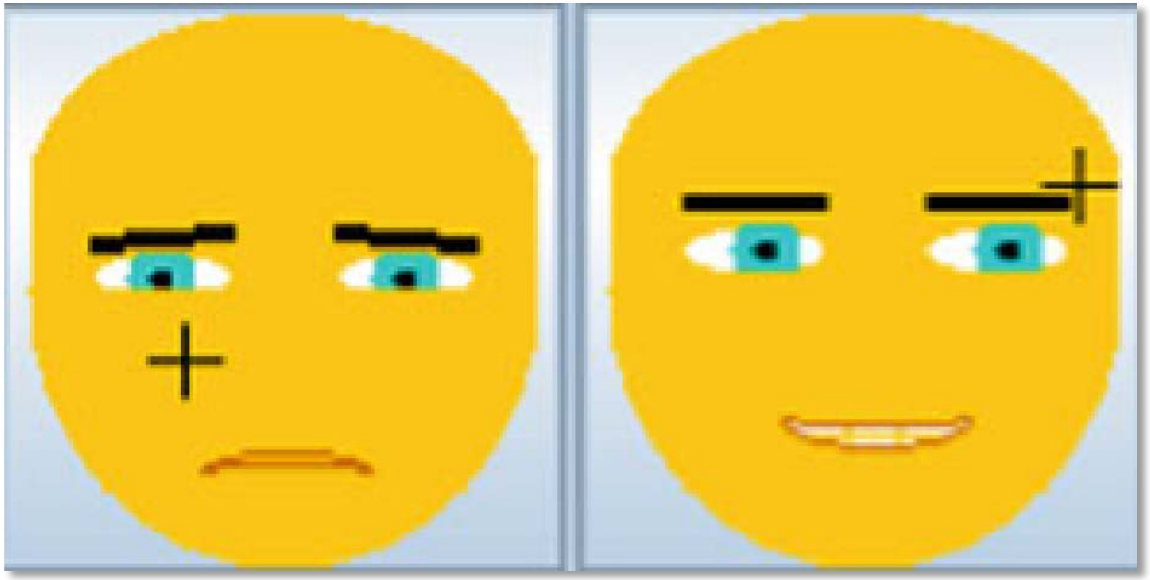
- How to make it as easy as possible express taste?
- **Not-so-common** approaches (from the literature):

Source: [1]

[1] Ken, Theresa Roeder, Dhruv Gupta, and Chris Perkins. 2001. Eigentaste: A Constant Time Collaborative Filtering Algorithm. Information RGoldbergetrieval 4, 2 (2001), 133–151.

Feedback Scale

- How to make it as easy as possible express taste?
- **Not-so-common** approaches (from the literature):



Source: [1]

[1] Alina Pommeranz, Joost Broekens, PascalWiggers,Willem-Paul Brinkman, and Catholijn M. Jonker. 2012. Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process. User Modeling and User-Adapted Interaction 22, 4 (2012), 357–397.

Feedback Scale – Considerations

- Number of options
  -  vs.  vs.  vs. 

- Wording of possible feedback values
  - e.g. “bad … very good” vs. “terrible … excellent”
- Extreme values
  - “(0) … (10)” vs. “(-5) … (5)”
- Presence of a neutral option
  - “(1) (2) **(3)** (4) (5)” vs. “(1) (2) (3) (4)”

- Often not considered in study setups

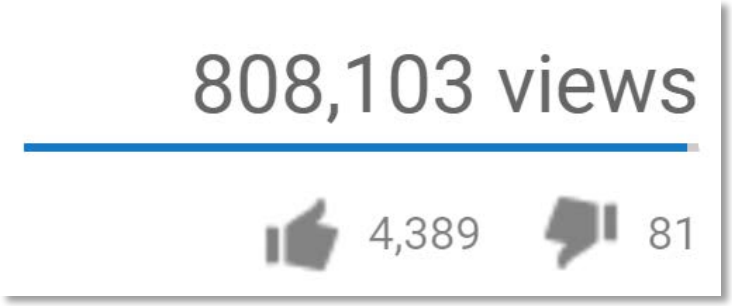
Feedback Scale — Observations

- Finer scales -> sense of control
  - More precision but more effort?
- Continuous sliders -> blurred boundaries -> less effort
- Mobile scenarios: different requirements
  - Reduce functionality? Make everything bigger?

- In the literature: Little consideration about **how** data is acquired, despite pitfalls

Explicit ratings — Reliability

- Even explicit ratings are only reliable to a degree
  - Can users rate consistently (over time)?
  - Feedback directly after consumption or later?
  - Community rating: Help or bias?



Source: youtube.com

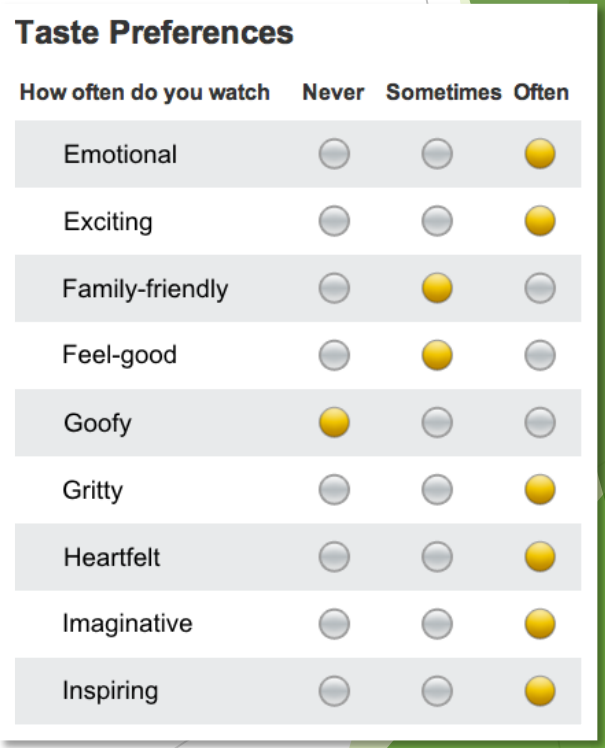
Explicit ratings — Reliability

- Even explicit ratings are only reliable to a degree
  - Can users rate consistently (over time)?
  - Feedback directly after consumption or later?
  - Community rating: Help or bias?
  - Public rating scales
    - Users become critics
    - Users like to rate “good” stuff -> 5 stars dominate

Dealing with Cold-Start Users

- New user -> reduce entry barrier!
- Interaction strategies even more important

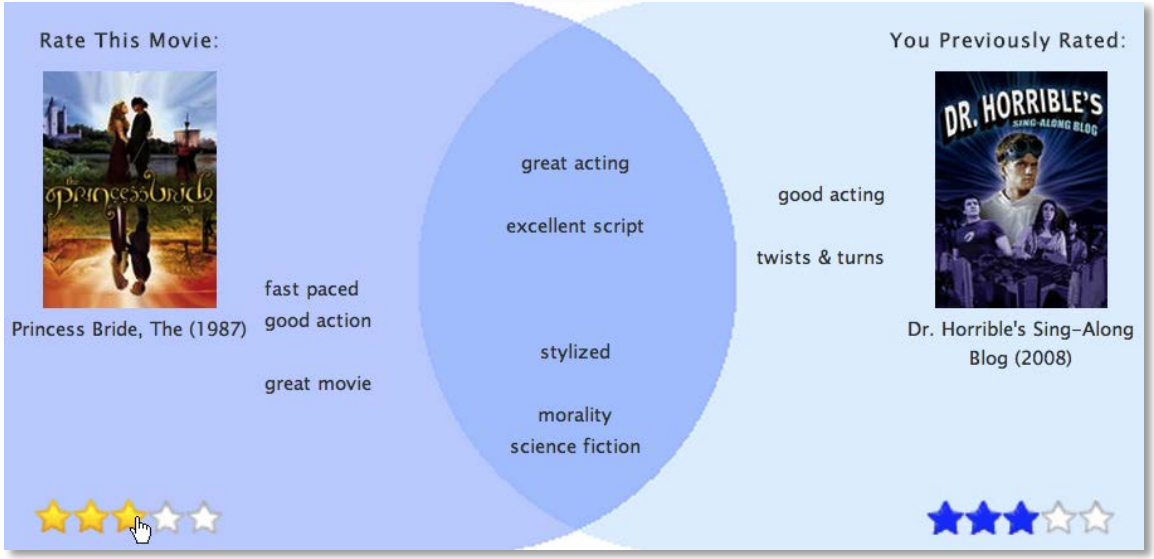
- Simply monitor users: Trust?
- Ask users about their tastes
  - How many questions?
    Trust/accuracy vs. effort
- Force users to rate items upfront
  - Same problem: How many items?



Source: netflix.com

Helping Users to Rate

- How to reduce rating effort for users?
- Help users to understand the item



Source: [1]

[1] Tien T. Nguyen, Daniel Kluver, Ting-YuWang, Pik-Mai Hui, Michael D. Ekstrand, Martijn C.Willemsen, and John Riedl. 2013. Rating support interfaces to improve user experience and recommender accuracy. In P*roceedings of the 7th Conference on Recommender Systems (RecSys '13)*. 149–156.

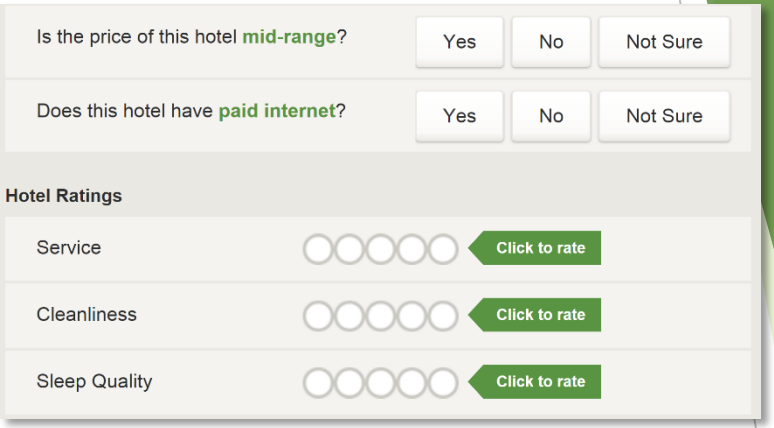
Helping Users to Rate

- How to reduce rating effort for users?

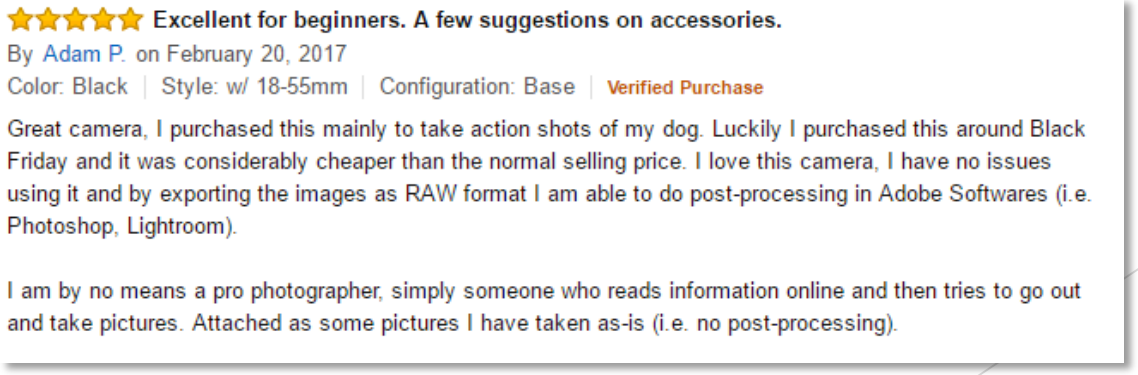
- Help users to understand the item
- Suggest items to rate vs. let users search themselves?
  - Suggestions seem to be easier/more reliable
  - Self-selection seems to lead to higher user satisfaction

Other Explicit Rating Interactions

- Multi-criteria ratings
  - More effort for users?
  - Added benefit? For users? For the system?
- Plain text item reviews
  - Rarely used
  - How to exploit?



Source: tripadvisor.com

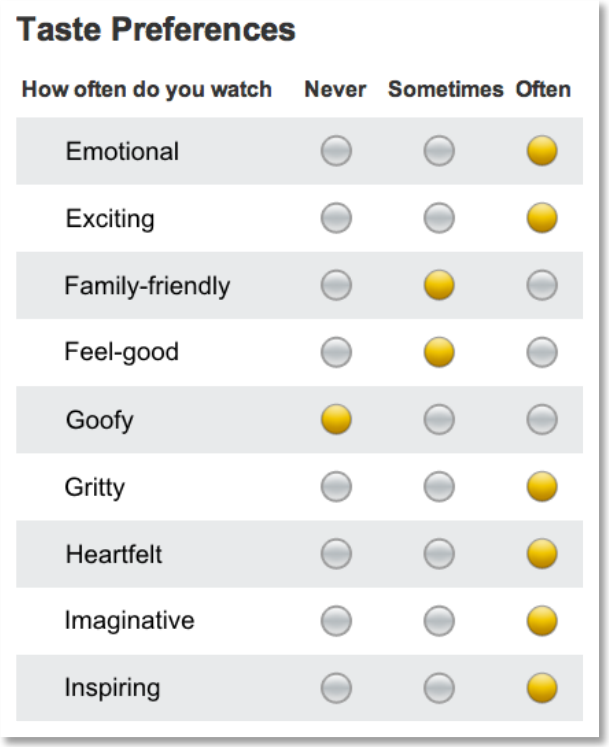


Source: amazon.com

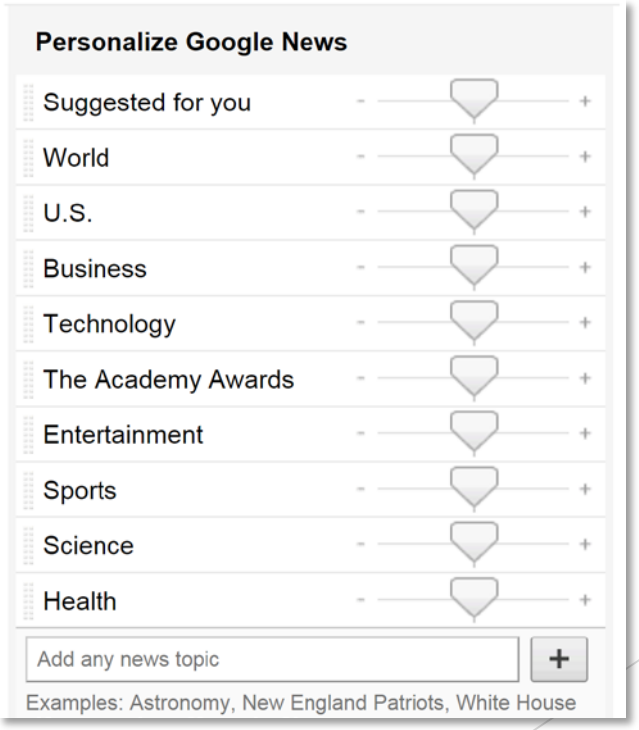
Preference Forms and Dialogs

Preference Forms

- Often used for cold-start users (on streaming sites)
- Sometimes available to the whole user base



Source: netflix.com

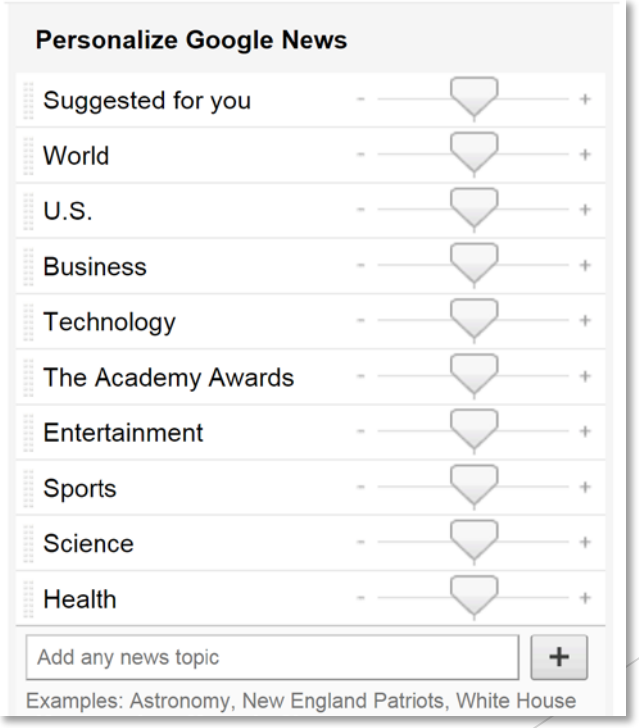


Source: news.google.com

Preference Forms

- Often used for cold-start users (on streaming sites)
- Sometimes available to the whole user base

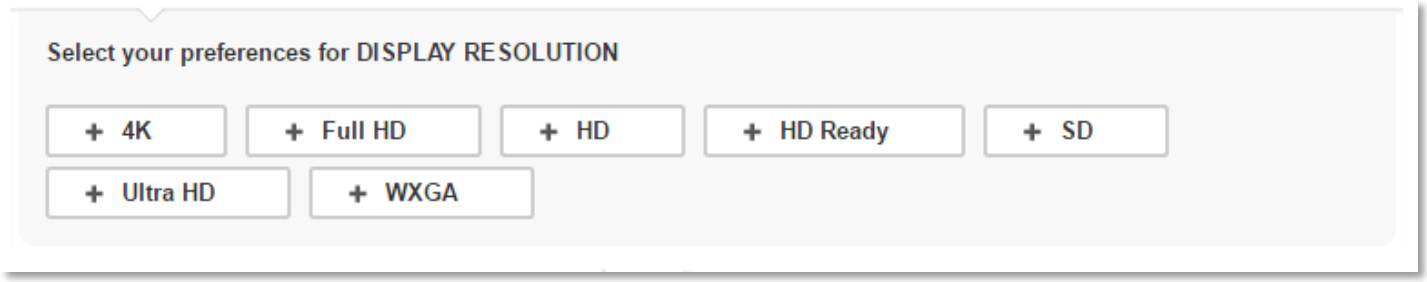
- User in control
- But …
  - Maybe too complex?
  - Will users tweak their profiles over time?
  - What if users mess around too much?
    - Fallbacks?
    - Undo option?



Source: news.google.com

Conversational Recommenders

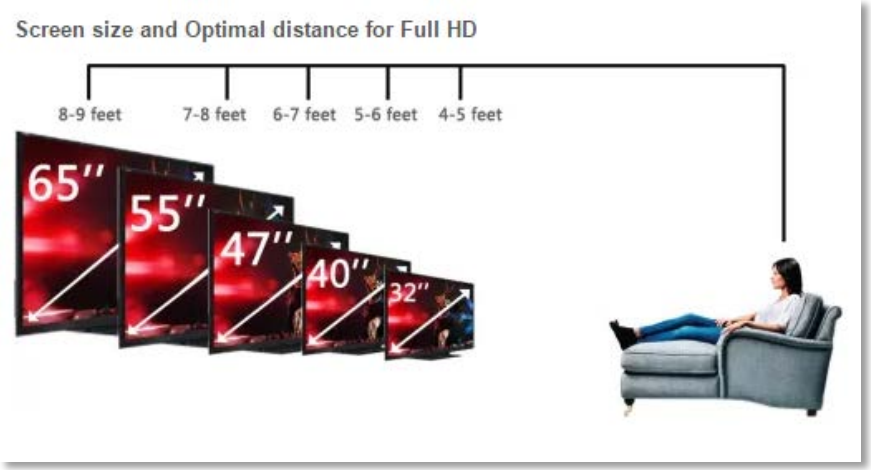
- Mostly for high-involvement products (e.g., cameras, TVs, smartphones, …)
- Idea:
  - Ask the user questions and evaluate answers (step by step)
  - In the end, recommend one (or more) items



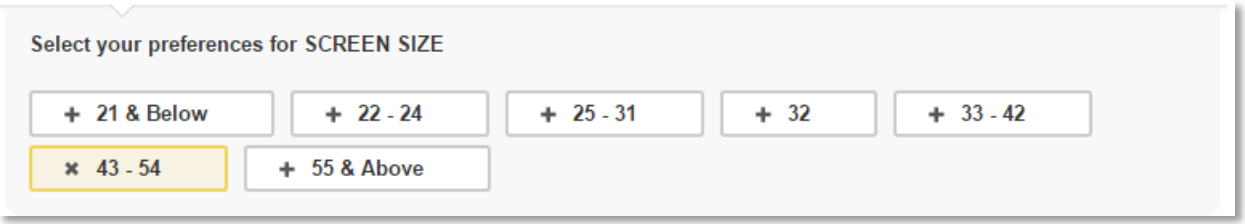
Sources: flipkart.com

Conversational Recommenders

- User guidance possible:



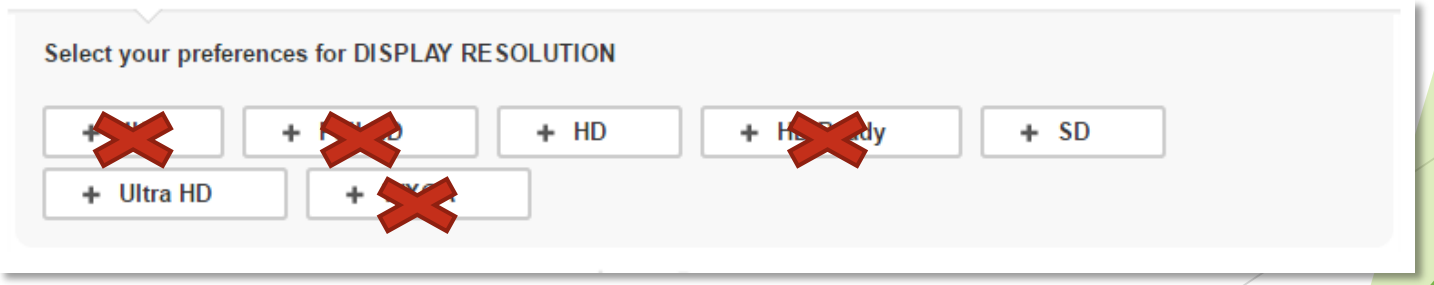
Source: flipkart.com



Source: flipkart.com

Conversational Recommenders

- Problems in comparison to other approaches:
  - Domain-specific knowledge needed
  - Complex implementation/maintenance
  - No long-term profiles
- Open questions:
  - How to deal with conflicting user requirements?
  - One-size-fits-all vs. personalization to expertise?

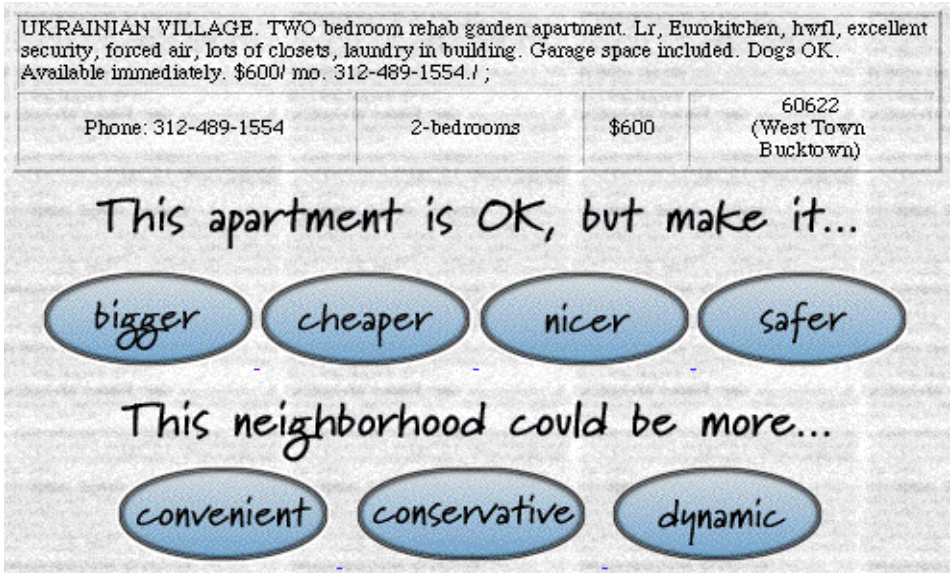


Source: flipkart.com

Critiquing

Critiquing

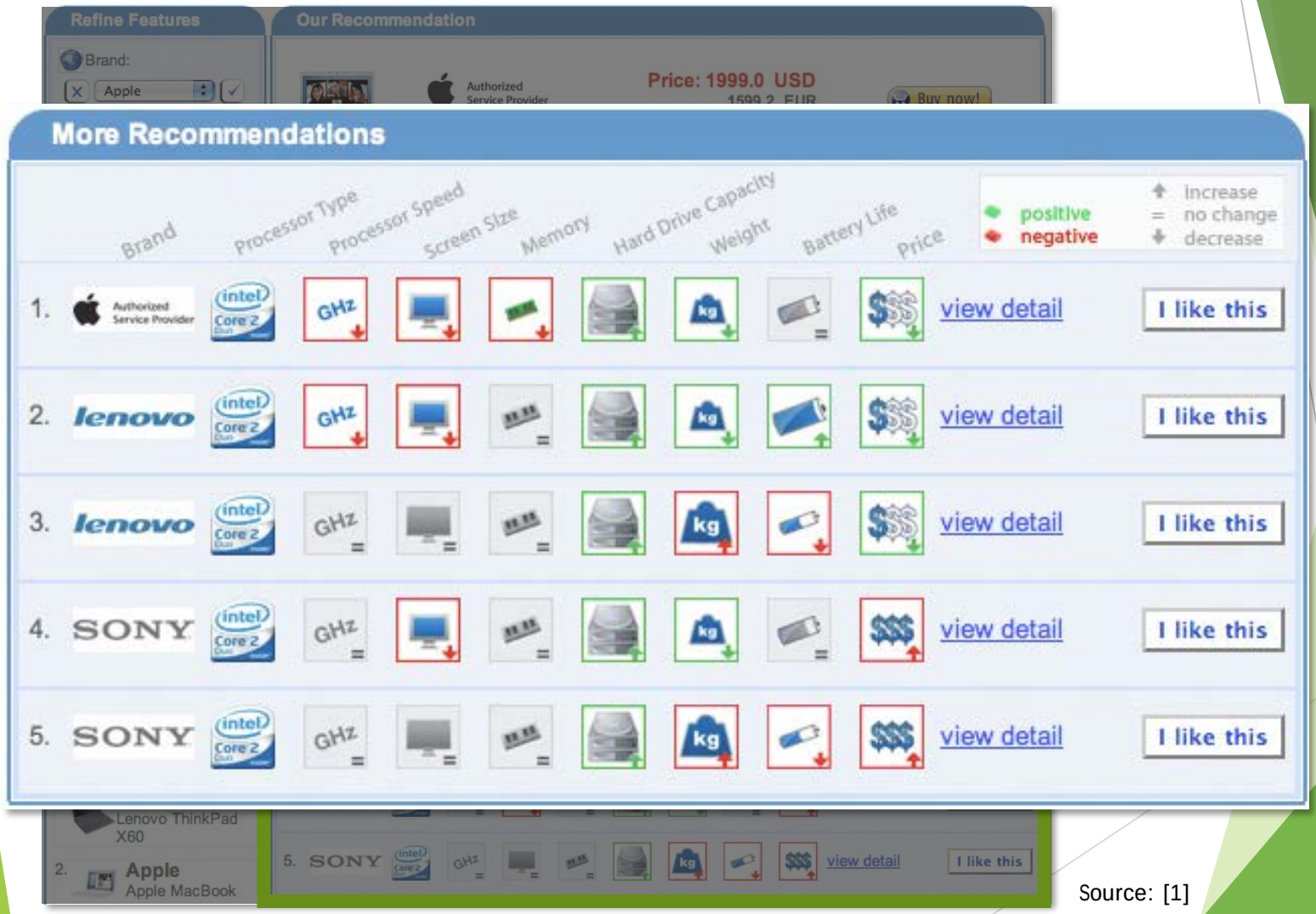
- Similar to conversational recommendation
- But: display a recommendation as soon as possible
- Let user critique the recommendations until a satisfying option is found



Source: [1]

[1] Robin D. Burke, Kristian J. Hammond, and Benjamin C. Young. 1996. Knowledge-Based Navigation of Complex Information Spaces. In *Proceedings of the 13th National Conference on Artificial Intelligence (AAAI '96)*. 462–468.

Compound Critiques



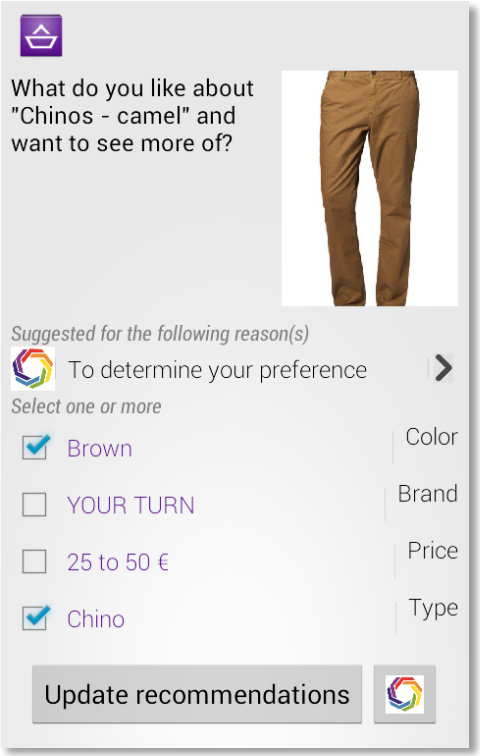
Source: [1]

[1] Jiyong Zhang, Nicolas Jones, and Pearl Pu. 2008. A visual interface for critiquing-based recommender systems. In *Proceedings of the 9th Conference on Electronic Commerce (EC '08)*. ACM, 230–239.

Critiquing – Open Questions

- On mobile?
- Weighting criteria differently?

- Integrate preferences into long-term user models?
- Natural language?
- How to evaluate offline?



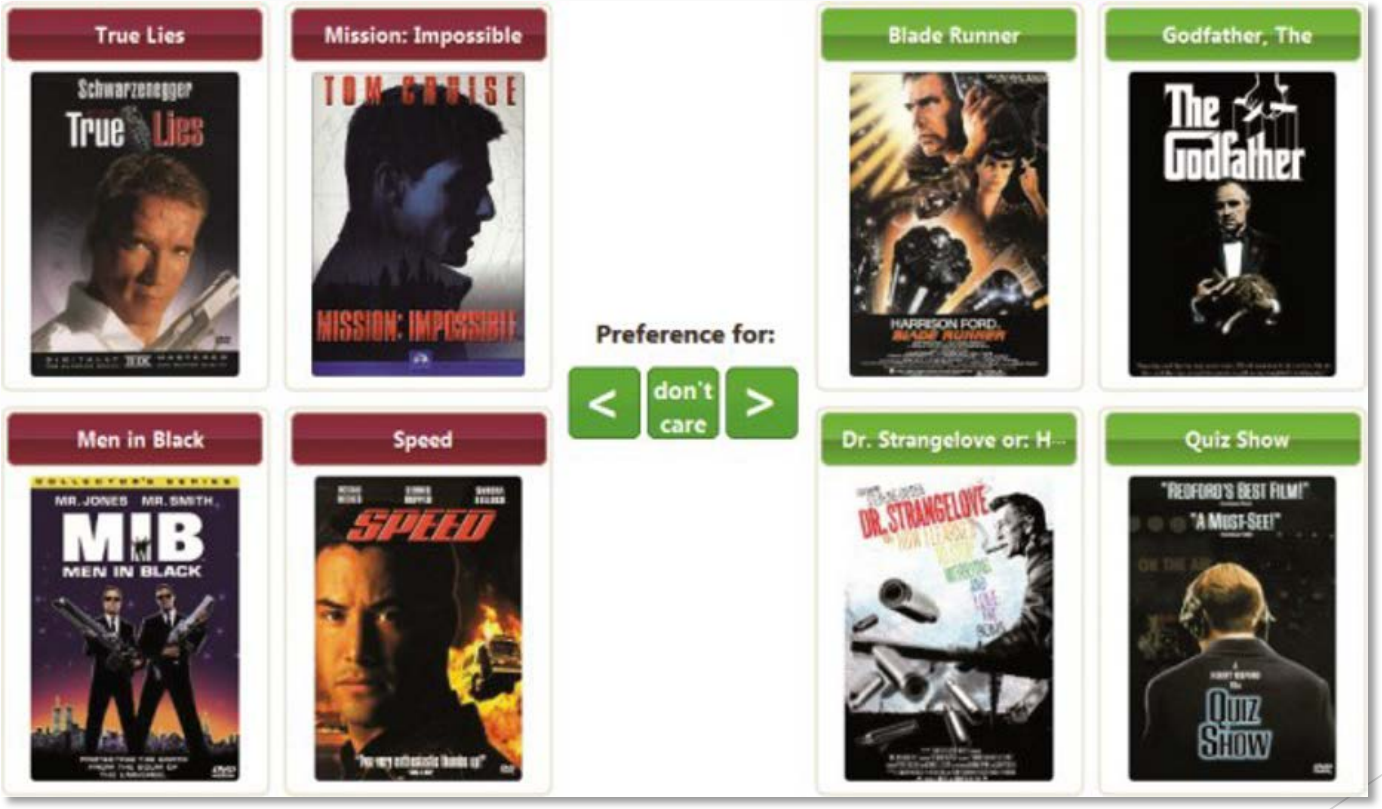
Source: [1]

[1] Béatrice Lamche, Yannick Rödl, Claudius Hauptmann, and Wolfgang Wörndl. 2015. Context-Aware Recommendations for Mobile Shopping. In *Proceedings of the Workshop on Location-Aware Recommendations (LocalRec '15) co-located with the 9th Conference on Recommender Systems (RecSys '15)*. 21–27.

Alternative Elicitation Techniques

Side-By-Side Comparison, Personality Quizzes, Picture-Based, …

Alternative Elicitation Techniques: Side-By-Side Comparison of Sets



Source: [1]

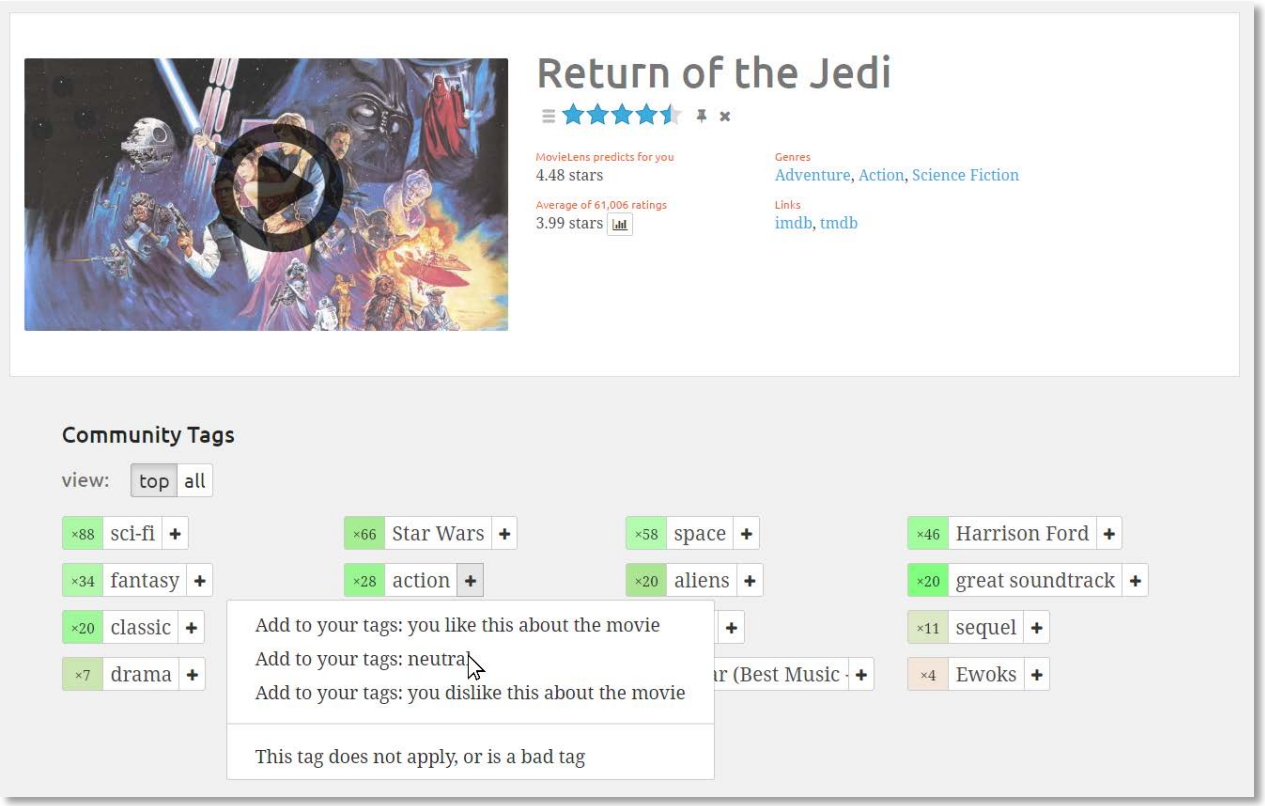
[1] Benedikt Loepp, Tim Hussein, and Jürgen Ziegler. 2014. Choice-based preference elicitation for collaborative filtering recommender systems. In *Proceedings of the 2014 SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. 3085–3094.

Alternative Elicitation Techniques: Picture-Based

Source: [1]

[1] Julia Neidhardt, Leonhard Seyfang, Rainer Schuster, and Hannes Werthner. 2015. A picture-based approach to recommender systems. *Information Technology & Tourism* 15, 1 (2015), 49–69.

Alternative Elicitation Techniques: Tag-Based



Source: movielens.com

Alternative Elicitation Techniques: Personality Quizzes

Source: [1]

[1] Yu Chen and Pearl Pu. 2012. CoFeel: An Interface for Providing Emotional Feedback in Mobile Group Recommender Systems. In *Joint Proceedings of the 1st International Workshop on Recommendation Technologies for Lifesytle Change (LIFESTYLE '12) and the 1st International Workshop on Interfaces for Recommender Systems (InterfaceRS '12)*. 48–55.

Alternative Elicitation Techniques: Open Questions

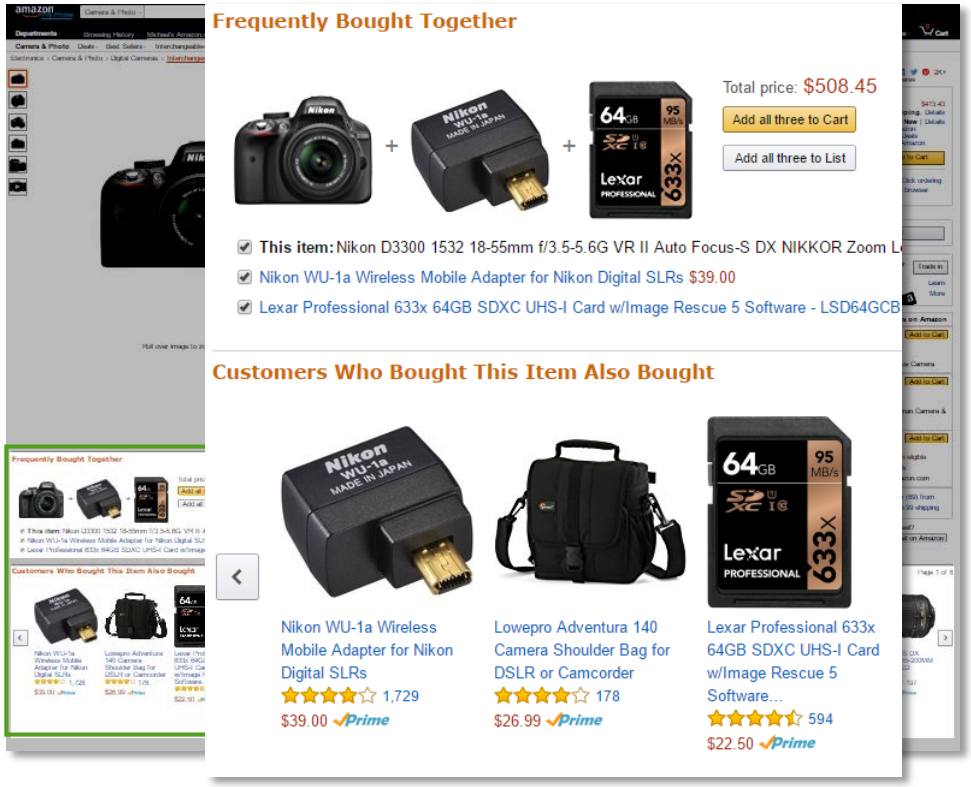
- Interaction complexity
- User acceptance
  - Users have to “learn” something new
- Benefit compared to traditional systems
  - Good fit for some domains?
  - Fun?

Interacting with Recommender Systems

Part II: Interactions during Result Presentation

Exercise

- What could be changed?



Source: amazon.com

Recommendation Result Presentation

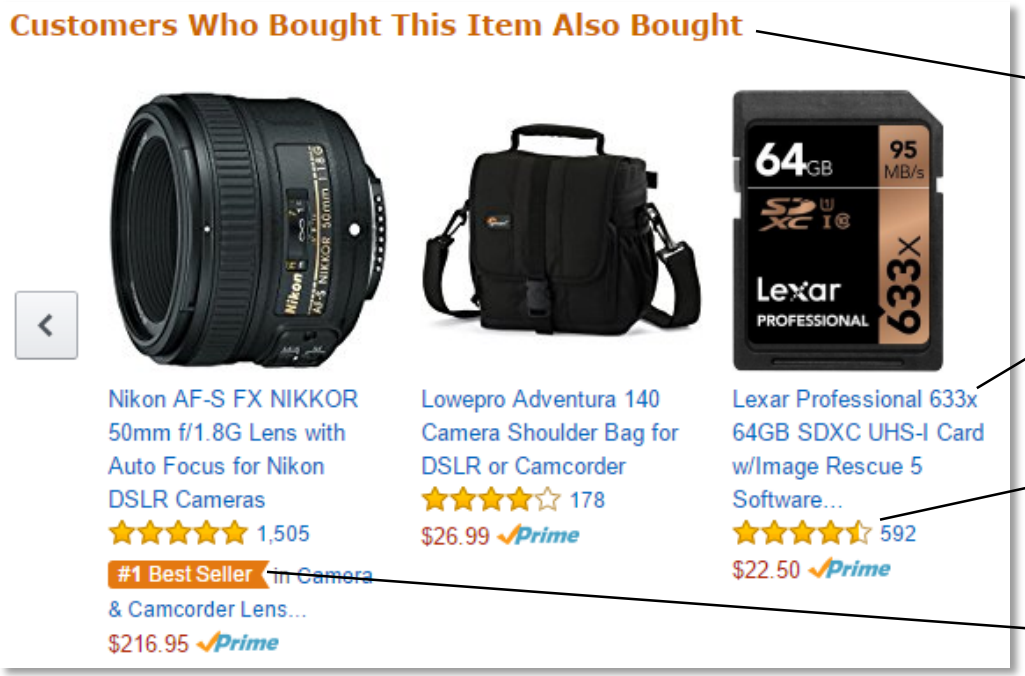
- How present recommendations to the user?
- Common goals (purpose of the RS):
  - User satisfaction/retention
  - Increased purchase rate
- How to achieve this?
  - Reduce user effort/choice overload
  - Make exploration of item space fun/engaging

Topics of This Session

- List design
- Visualization
- User feedback
- Proactivity
- Persuasion

List Design

List Design — Considerations



Source: amazon.com

List label

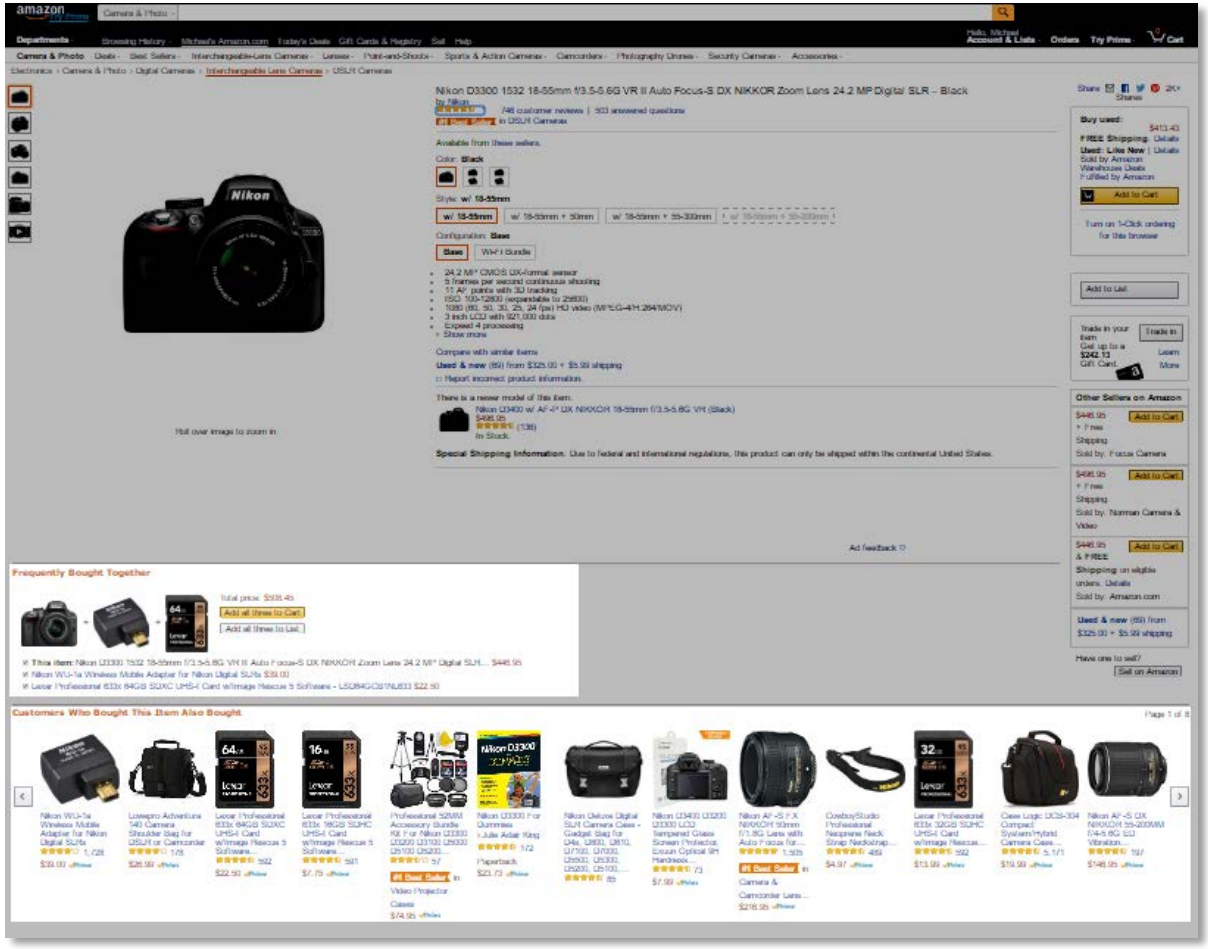
Item description

Community rating

Highlighting

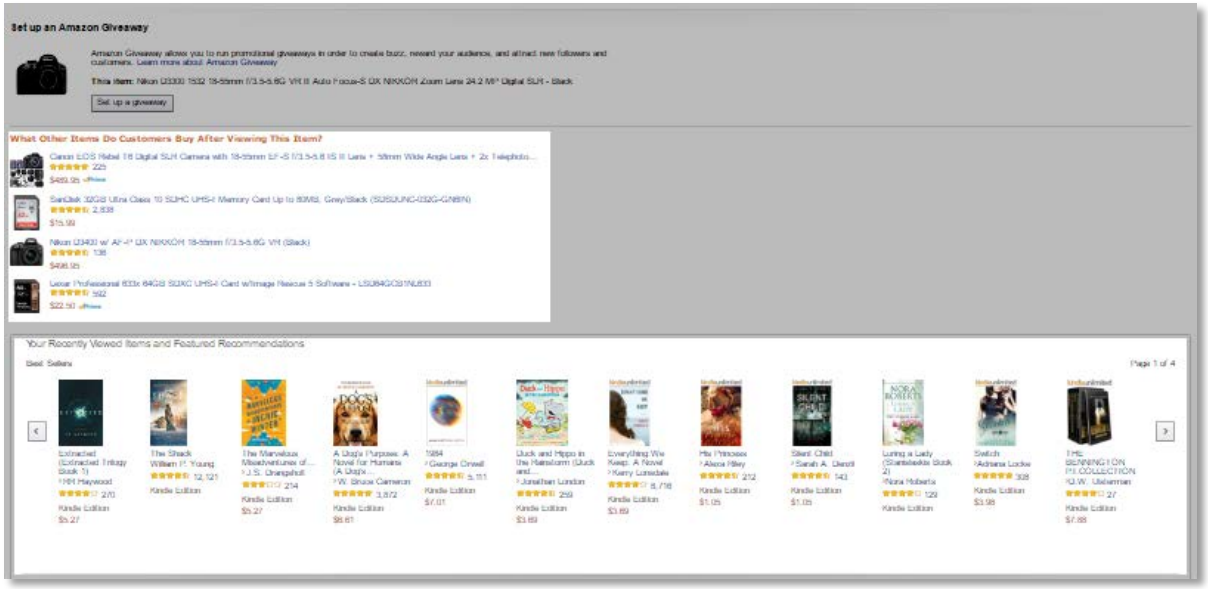
Number of options

List Design – Additional Considerations: Placement



Source: amazon.com

List Design – Additional Considerations: Placement



Source: amazon.com

- How many lists?
- Where?
- Orientation?
- Visible on demand (hover, click, …)?

List Design — Psychological Principles

- Too few vs. too many options
  - Choice overload, reduced sales
- Finding a middle ground can also depend on:
  - User context
  - Item diversity
  - Comparability of items

Source: amazon.com

≠

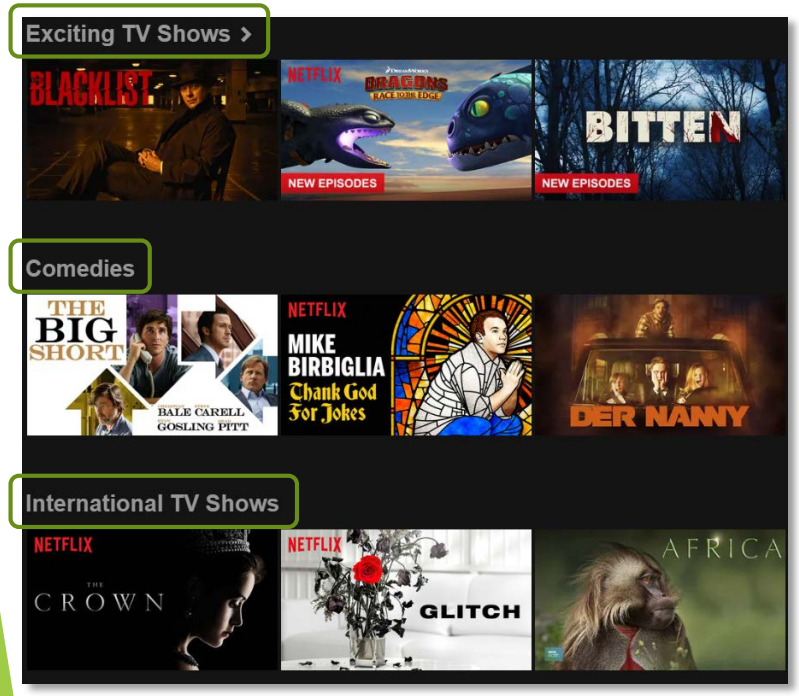
Source: amazon.com

List Design — Psychological Principles

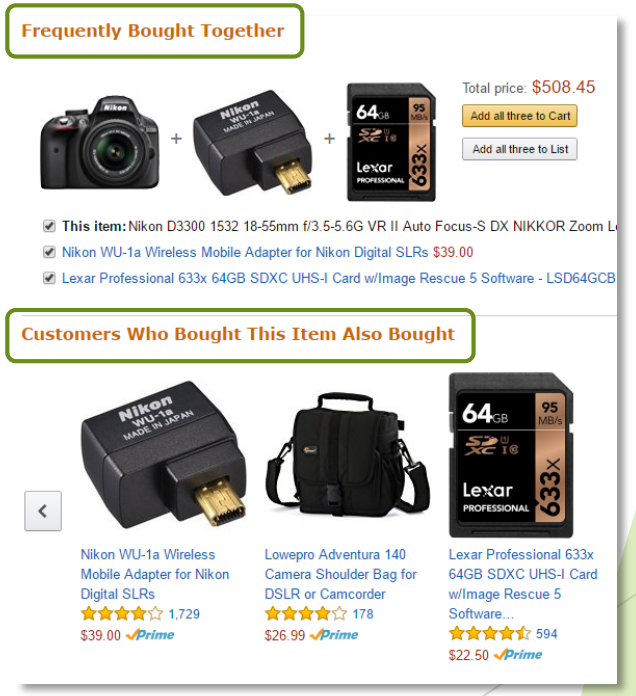
- Too few vs. too many options
  - Choice overload, reduced sales
- Finding a middle ground can also depend on:
  - User context
  - Item diversity
  - Comparability of items
  - List personalization
    - More or less effort/overload?

Clustered Lists

- Group items
  - Semantically
  - Technically



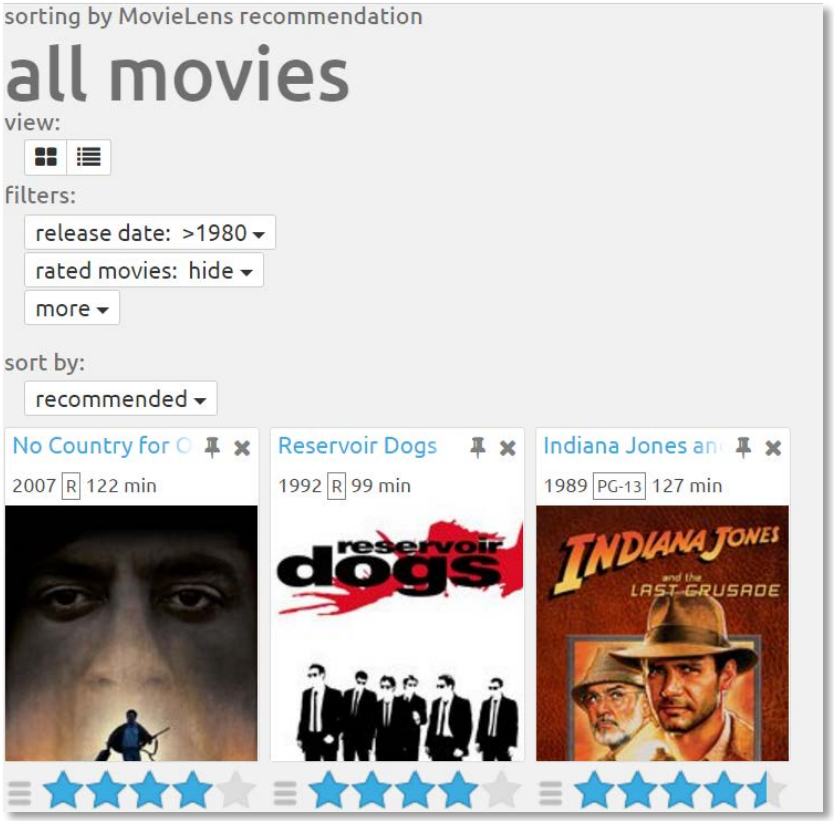
Source: netflix.com



Source: amazon.com

Alternative List Designs

- Sortable/filterable lists



Source: movielens.org

Alternative List Designs

- Sortable/filterable lists
- Alternatives to traditional "lists"
  - Grid
  - Pie/circle[1]

Source: [1]

Source: [1]

[1] Li Chen and Ho Keung Tsoi. 2011. Users' Decision Behavior in Recommender Interfaces: Impact of Layout Design. In *Joint Proceedings of the RecSys 2011 Workshop on Human Decision Making in Recommender Systems (Decisions@RecSys '11) and User-Centric Evaluation of Recommender Systems and Their Interfaces-2 (UCERSTI 2) affiliated with the 5th Conference on Recommender Systems (RecSys '11)*. 21–26.

List Design — Open Questions

- Large number of design options
  - One list vs. multiple (clustered) lists
  - List placement
  - Number of item, description detail
  - List layout
- Considerations:
  - Choice/information overload vs. satisfaction of the user's need for information/exploration
  - Consumer trust vs. persuasion to buy
  - Integration into corporate design
  - Domain specifics
  - …

Visualization

Highlighting

- Instead of presenting a filtered list, highlight items

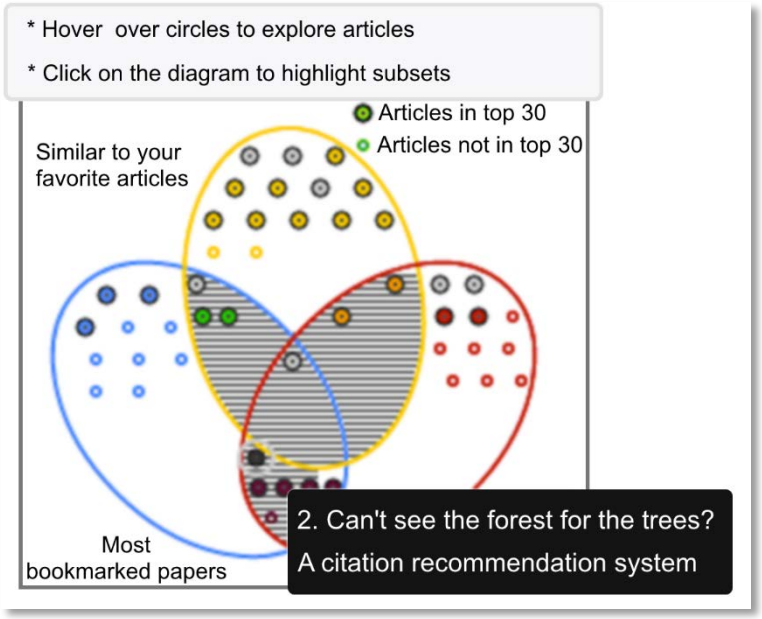
Source: [1]

- Familiar user experience with RS as added benefit
- No filtering -> higher user confidence?

[1] Wesley Waldner and Julita Vassileva. 2014. Emphasize, don't filter!: displaying recommendations in Twitter timelines. In *Proceedings of the 8th Conference on Recommender Systems (RecSys '14)*. 313–316.

Diagrams and Graphs

- Use **diagrams** to communicate item relations



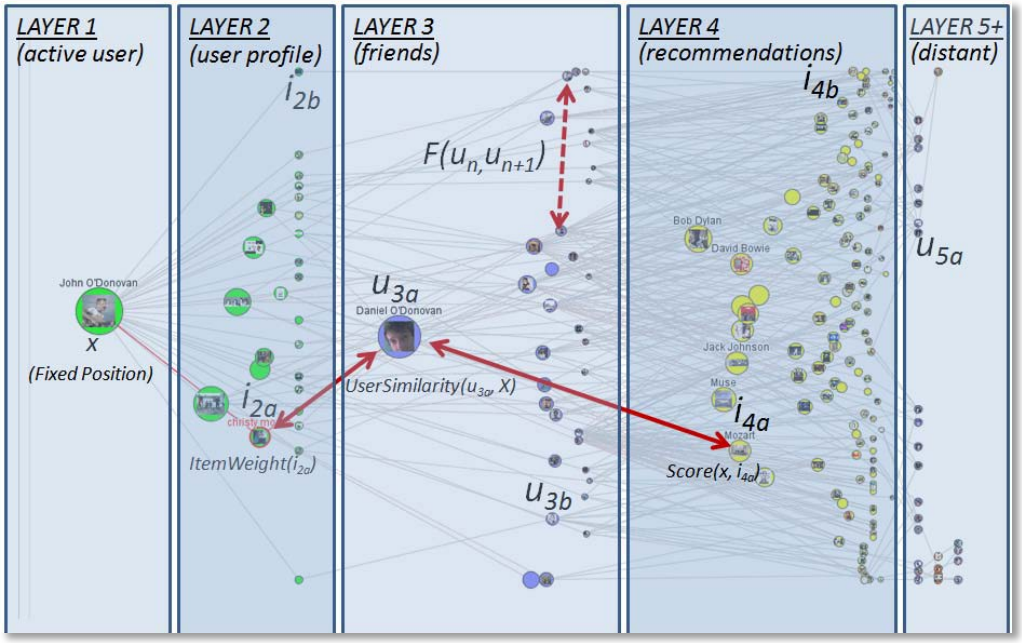
Source: [1]

- Easier for users or additional effort? Fun?

[1] Denis Parra, Peter Brusilovsky, and Christoph Trattner. 2014. See what you want to see: visual user-driven approach for hybrid recommendation. In *Proceedings of the 19th International Conference on Intelligent User Interfaces (IUI '14)*. 235–240.

Diagrams and Graphs

- Use **graphs** to communicate item relations



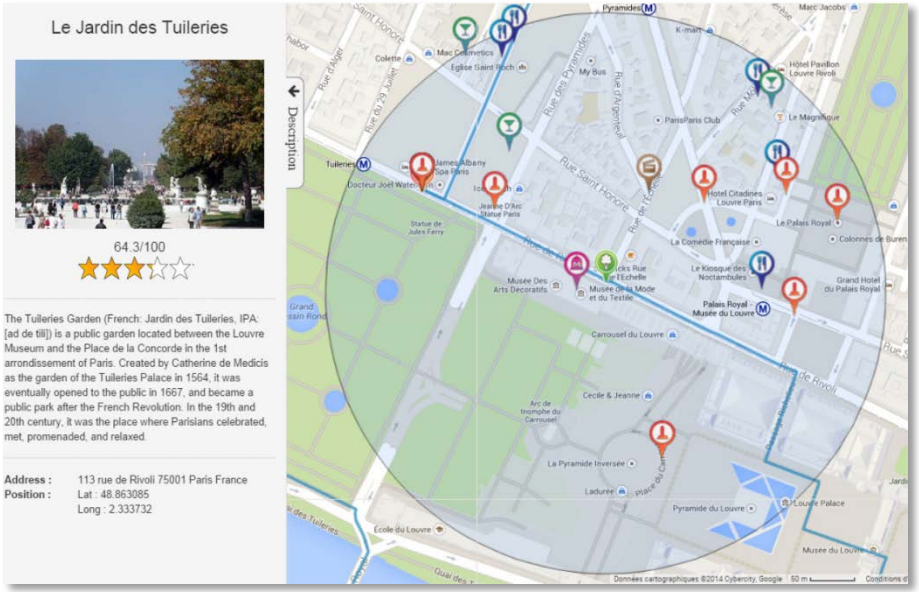
Source: [1]

- Easier for users or additional effort? Fun?

[1] Brynjar Gretarsson, John O'Donovan, Svetlin Bostandjiev, Christopher Hall, and Tobias Höllerer. 2010. Small-Worlds: Visualizing Social Recommendations. *Computer Graphics Forum* 29, 3 (2010), 833–842.

Recommendations on Maps

- Putting recommendations on maps is well-suited for some domains (e.g. POI recommendation)
- Great interaction possibilities (e.g. itinerary recommendations)

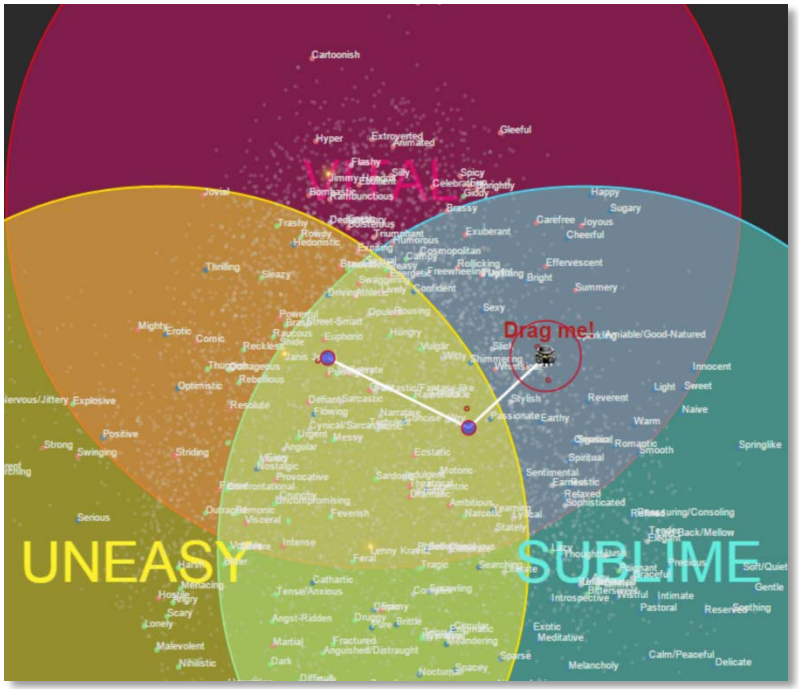


Source: [1]

[1] Alexandre Yahi, Antoine Chassang, Louis Raynaud, Hugo Duthil, and Duen Horng (Polo) Chau. 2015. Aurigo: an Interactive Tour Planner for Personalized Itineraries. In *Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15)*. 275–285.

Recommendations in 2D space

- 2D plot can encompass “hidden feature space”
- Possibility to include user in the plot
  -> interaction potential?

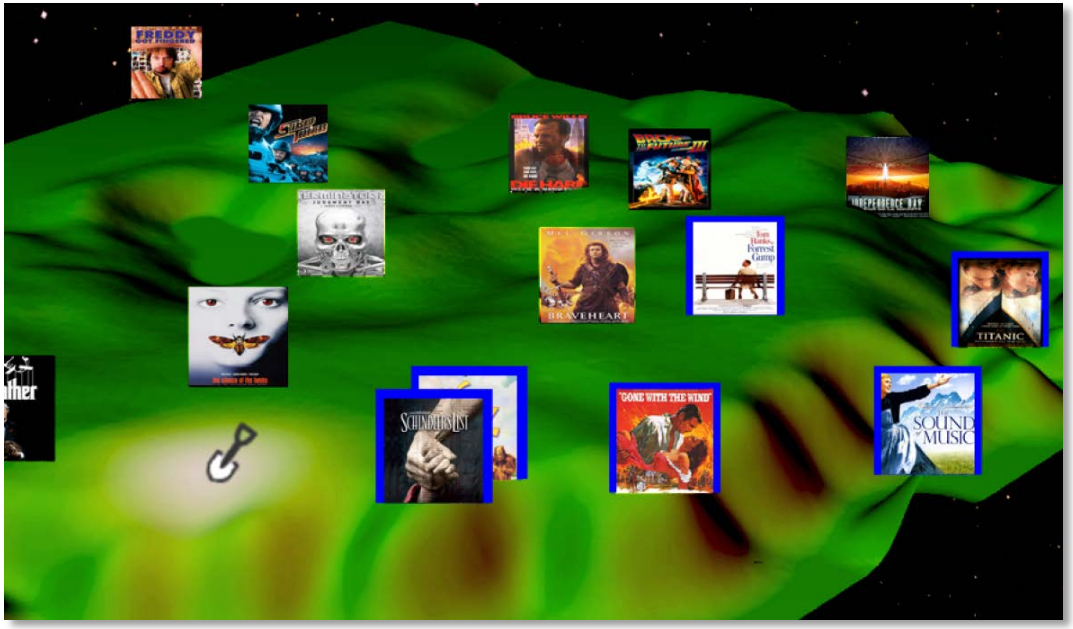


Source: [1]

[1] Ivana Andjelkovic, Denis Parra, and John O’Donovan. 2016. Moodplay: Interactive Mood-based Music Discovery and Recommendation. In *Proceedings of the 24th Conference on User Modeling, Adaptation, and Personalization (UMAP ’16)*. 275–279.

Recommendations in 3D space

- Additional 3rd dimension for extra information (e.g. user profile)



Source: [1]

[1] Johannes Kunkel, Benedikt Loepp, and Jürgen Ziegler. 2015. 3D-Visualisierung zur Eingabe von Präferenzen in Empfehlungssystemen. In *Proceedings of the 2015 Conference on Mensch und Computer 2015*. De Gruyter Oldenbourg, 123–132.

Visualization — Open Questions

- Making exploration *fun* vs. making everyday use *convoluted*
- Extra information vs. extra effort
- Miniscule changes in practice -> huge impact on user experience -> experimentation dangerous? Never touch a running system?

- Exploration of emerging technologies:
  - Gamification
  - Virtual reality, Augmented reality
  - …

User Feedback

User Feedback — General Idea

- So far:
  - Let users express their taste
  - Display (or visualize) recommendations
- What's missing?
  - Let users express their opinion about the recommendations
- Goals:
  - Correct faulty system assumption
  - Improve recommendations over time
  - Help users *feel more in control*

User Feedback – Industry Approaches

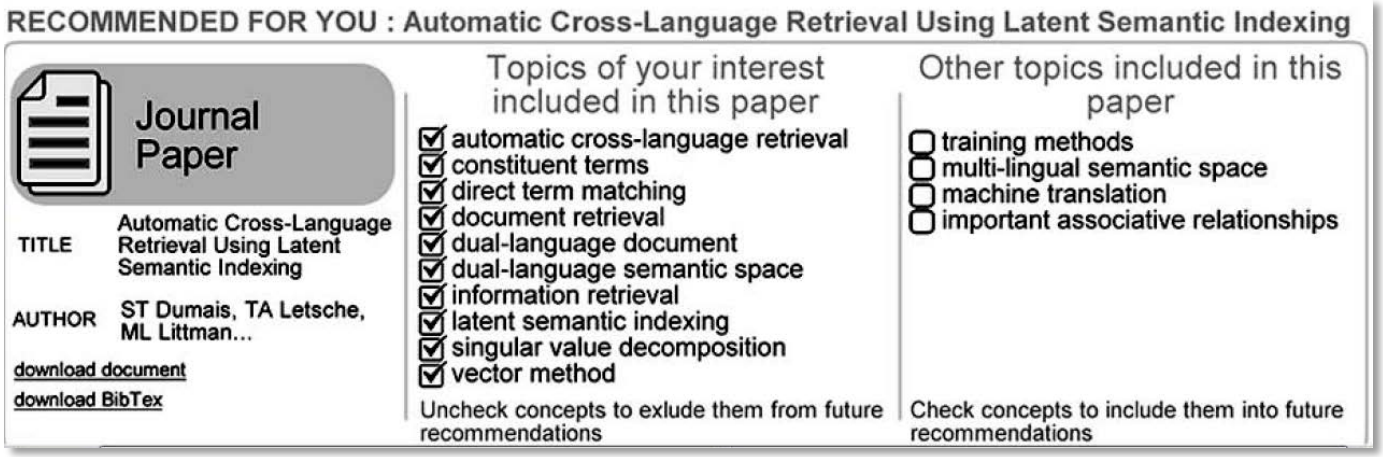
- Well-hidden feature on amazon.com: feedback

Source: amazon.de

- Also available on other platforms (e.g. music streaming "skip"), but often to a lesser extent

User Feedback – Research Approaches

- Similar to Amazon's approach:



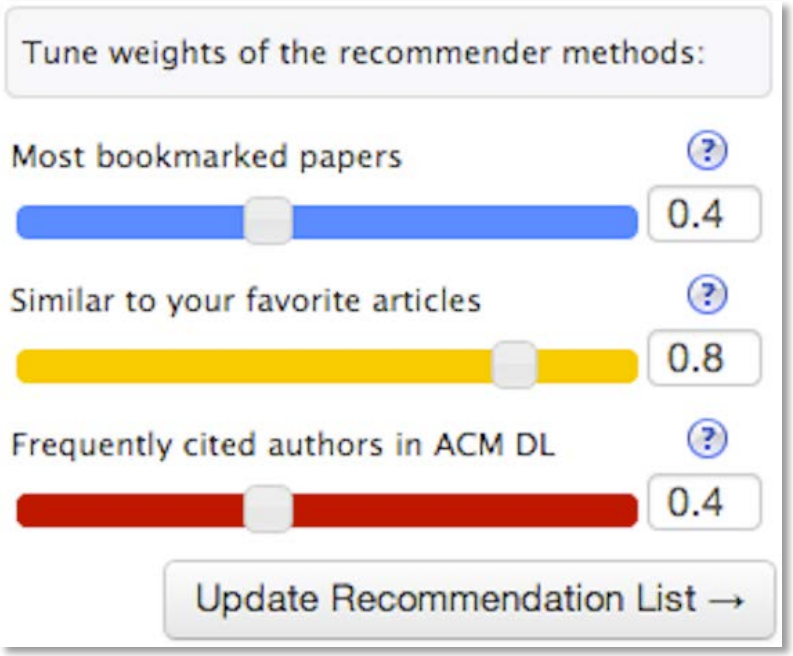
Source: [1]

- Additionally: positive feedback
  - Not only correction of the user profile -> augmentation

[1] Dario De Nart, Felice Ferrara, and Carlo Tasso. 2013. Personalized Access to Scientific Publications: from Recommendation to Explanation. In *Proceedings of the 21st International Conference on User Modeling, Adaptation, and Personalization (UMAP '13)*. 296–301.

User Feedback – Influencing the Strategy

- Let the user take full control



Source: [1]

- Can the user understand these options?

[1] Denis Parra, Peter Brusilovsky, and Christoph Trattner. 2014. See what you want to see: visual user-driven approach for hybrid recommendation. In *Proceedings of the 19th International Conference on Intelligent User Interfaces (IUI '14)*. 235–240.

User Feedback — Open questions

- Best feedback option?
  - Simple skip/dislike actions
  - Complex feedback options with explanations
  - In-depth profile inspection and manipulation
  - Manipulation of the recommendation strategy
- How to weight explicit item preferences (ratings) against feedback

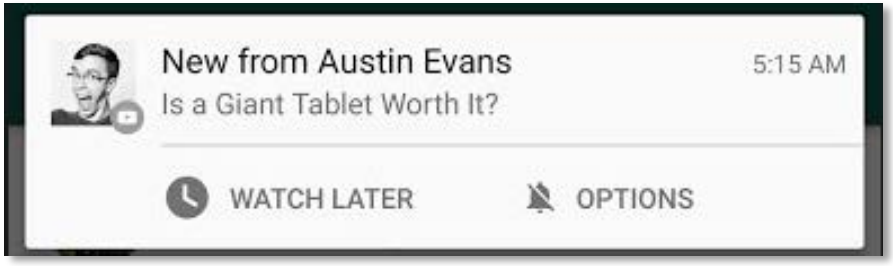
- Do users feel more in control?
- Does it improve system accuracy?

Proactive Recommendations

Proactive Recommendations — General Idea

- So far: *user*-initiated interactions

- Now: *system*-initiated interactions
  - Recommend when the user is not even using the system
    - Email
    - App notification
    - …



Source: YouTube app (Android)

Proactive Recommendations — Open Questions

- How much is too much?
  - Notification frequency
  - Amount of content per notification
  - Explore-exploit
- Best time to notify?
  - Based on user context
  - Based on the user's attention level
- Does it accomplish provider goals?
- Can it have negative effects?

Persuasive Recommender Systems

Persuasive Recommender Systems

- Should the RS *persuade* the user to make the right choice or stay *neutral*?
- What is the right choice?
  - Best for the user -> RS as benevolent advisor
  - Best for the provider -> RS as exploitation device
  - Best/good for both -> win/win
- Ethical concerns?
- In RS research:
  - Persuasion with different psychological effects
    - Primacy, recency, anchoring, framing, decoy effect, etc.
    - Persuasion with explanations

Interacting with Recommender Systems

Part III: Explanations

Part III: Explanations

- A comprehensive overview of what has been done in the field of explanations in **advice-giving systems**
  - Expert systems
  - Decision-support systems
  - Knowledge-based systems
  - Recommender systems
- Insights
  - Explanation taxonomy
  - Open challenges

Terminology

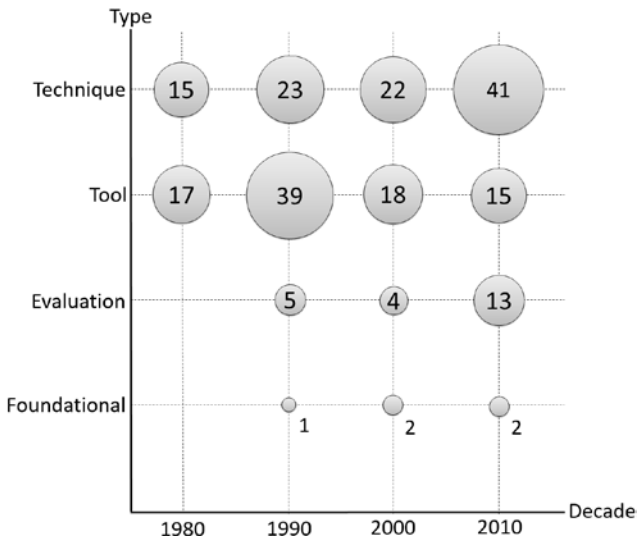
- Advice-giving systems
  - Expert Systems +
  - Decision-support Systems +
  - Knowledge-based Systems +
  - Recommender Systems
- Explanation
  - Information that is presented to the user
- Explanation generation approach
  - Technique or method that takes some data and transforms it into an explanation
- Decision inference method
  - Method that chooses an alternative from a set
    - E.g. decision technique or recommender algorithm
- Decision inference process
  - Execution of the decision inference method (an instance)

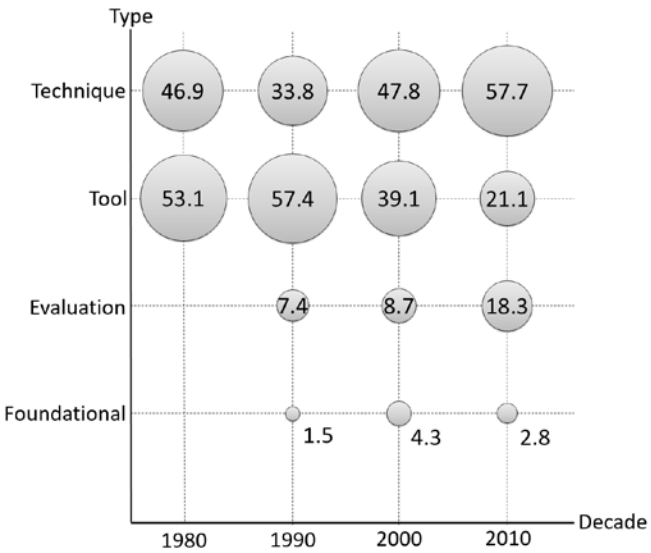
Information Source

- Systematic review of the literature
- Query string: explanations + [system classes]
  - Synonyms: justification, argumentation
- 217 analysed primary studies
  - from 1209
- Digital databases
  - ACM, IEEE, Science Direct, Springer Link
- Categories
  - Techniques (101)
  - Tools (89)
  - Evaluation (22)
  - Foundational Study (5)

Historical Developments

The Rise-Fall-Rise of Explanations





- Publications increased in the recent past years*
- Tools were largely important in the past
- Evaluations received much more attention recently
- Low number of foundational studies

* Papers published before August 12, 2016

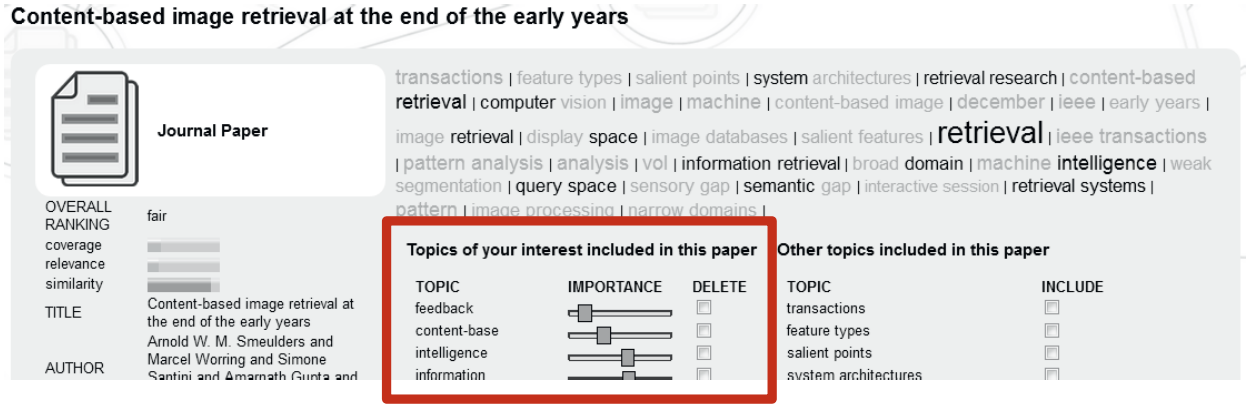
Explanation Styles

Part III: Explanations

Explanation Content

Preferences and Inputs (1/2)

- Decisive Input Values
  - If F1 is *very medium* AND F2 is *high* then likely class 1.
- Preference Match

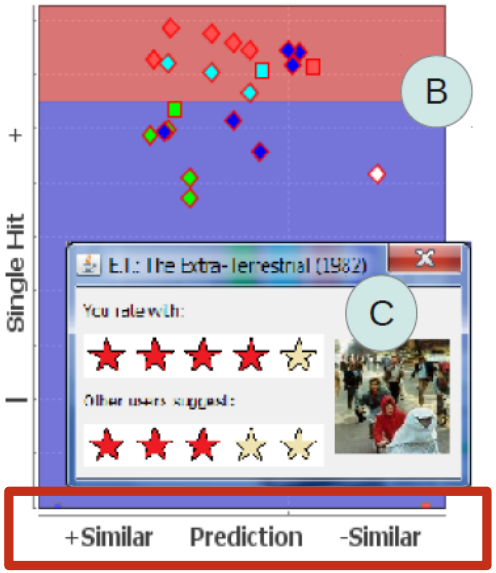


S. Mitra, S. K. Pal, Fuzzy multi-layer perceptron, inferencing and rule generation, IEEE Transactions on Neural Networks 6 (1) (1995) 51–63.
D. D. Nart, C. Tasso, A personalized concept-driven recommender system for scientific libraries, Procedia Computer Science 38 (2014) 84–91.

Explanation Content

Preferences and Inputs (2/2)

- Feature Importance Analysis
  - Even though x is better than y on average, y is preferred to x since y is better than x on the criteria X that **are important** whereas y is worse than x on the criteria Y that **are not important**.
- Suitability Estimate

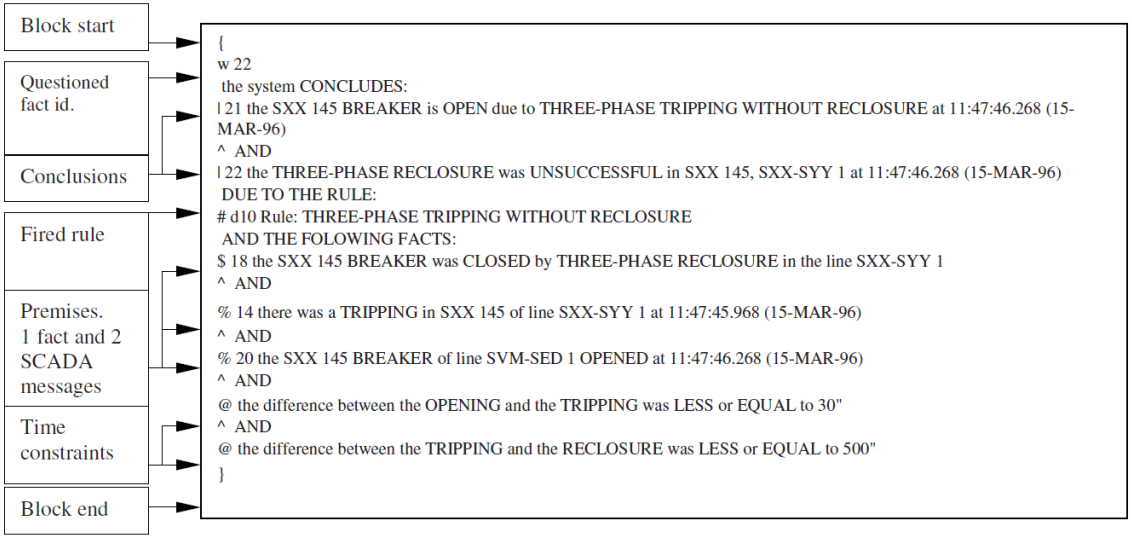


C. Labreuche, A general framework for explaining the results of a multi-attribute preference model, Artificial Intelligence 175 (7) (2011) 1410–1448.
Cleger-Tamayo, J. M. Fernandez-Luna, J. F. Huete, Explaining neighborhood-based recommendations, SIGIR '12, 2012, pp. 1063–1064

Explanation Content

Decision Inference Process (1/2)

- Inference Trace



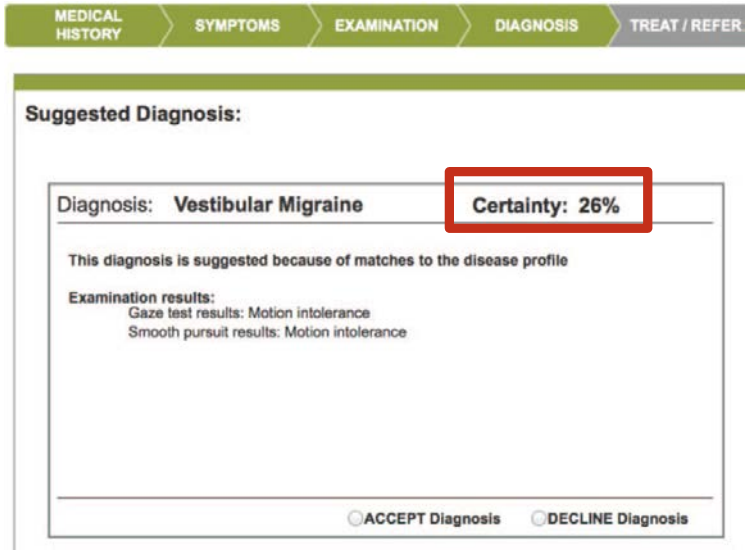
- Inference and Domain Knowledge
  - *What is the goal of the action proposed?*
    It was suggested to switch off the oxygen nozzle in order to reach the following goal(s): keep carbon percentage unchanged.

N. Malheiro, Z. A. Vale, C. Ramos, J. Santos, A. Marques, Enabling Client-Server Explanation Facilities in a Real-Time Expert System, Springer Berlin Heidelberg, 1999, pp. 333–342.
G. Guida, P. Mussio, M. Zanella, User interaction in decision support systems: the role of justification, in: 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation, Vol. 4, 1997, pp. 3215–3220.

Explanation Content

Decision Inference Process (2/2)

- Inference Method Side-outcomes



- Self-reflective Statistics

User: Why did you suggest to view the temple of Hephestos?

Robot: Trust me, I have been correct in most cases (70%) in the past.

A. Bussone, S. Stumpf, D. O'Sullivan, The role of explanations on trust and reliance in clinical decision support systems, ICHI 2015, 2015, pp. 160–169.
D. Vogiatzis, V. Karkaletsis, A cognitive framework for robot guides in art collections, Universal Access in the Information Society 10 (2) (2011) 179–193.

Explanation Content

Background and Complementary Information (1/4)

- Knowledge about Similar Alternatives

Explanation

The DVD **Ocean's Twelve** is recommended to you, because

1. You clicked on DVD **Ocean's Eleven** (**hide**)

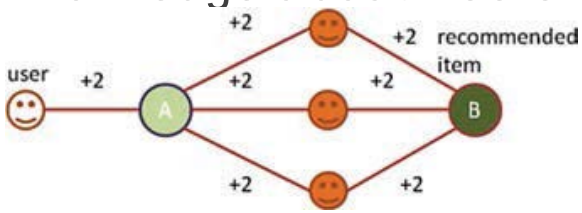
   That DVD has got the following similarities to DVD Ocean's Twelve:
   - Film Genre: Comedy
   - Film Genre: Thriller
   - Person: Bernie Mac
   - Person: Brad Pitt
   - Person: George Clooney
   - Person: Matt Damon
   - Person: Steven Soderbergh
   - Type: Product
   - Type: DVD

   Following argumentation(s) base(s) on endogenous factors, such as your click behavior:

   - **Ocean's Eleven** is performed by Person **Bernie Mac**, just like DVD **Ocean's Twelve**.
     Initiated by an exogenous factor: no. Induced energy: 0,143. Start age: 13. Length: 2.
   - **Ocean's Eleven** and DVD **Ocean's Twelve** are performed by the same Person (**Brad Pitt**).
     Initiated by an exogenous factor: no. Initiated energy: 0,143. Start age: 13. Length: 2.

   SPREADR inferred wrong? In order to get better recommendations you can **give feedback**.

- Knowledge about Peers

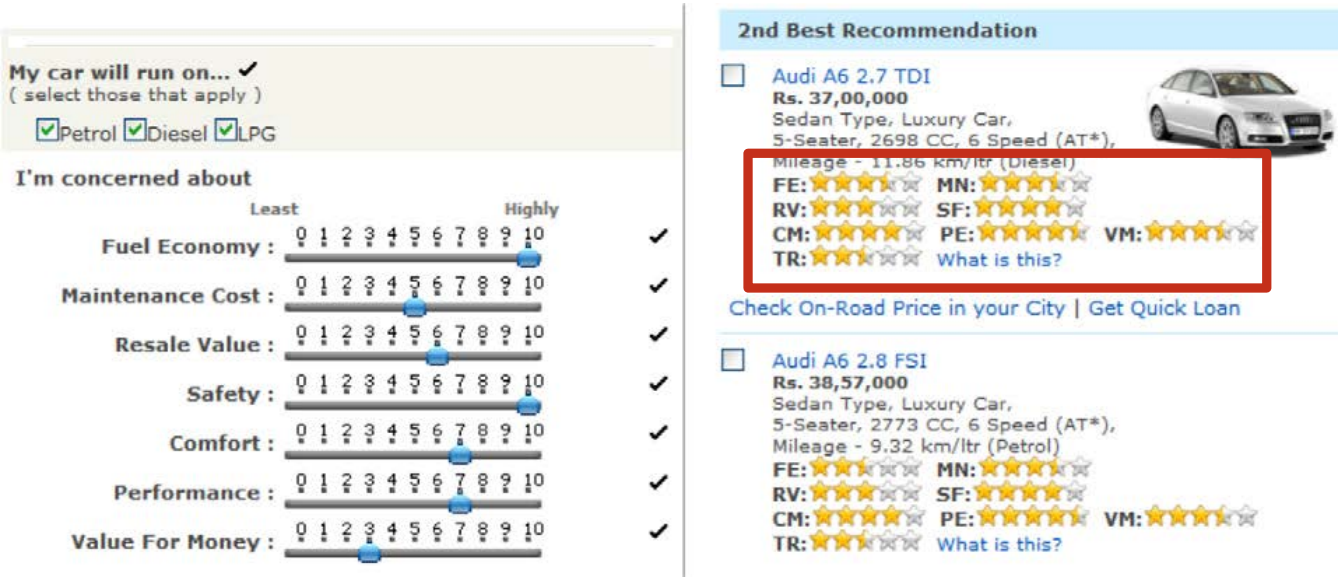


Y.-C. Chen, Y.-S. Lin, Y.-C. Shen, S.-D. Lin, A modified random walk framework for handling negative ratings and generating explanations, ACM Trans. Intell. Syst. Technol. 4 (1) (2013) 12:1–12:21.
T. Hussein, S. Neuhaus, Explanation of spreading activation based recommendations, SEMAIS '10, 2010, pp.24–28.

Explanation Content

Background and Complementary Information (2/4)

- Knowledge about the Community

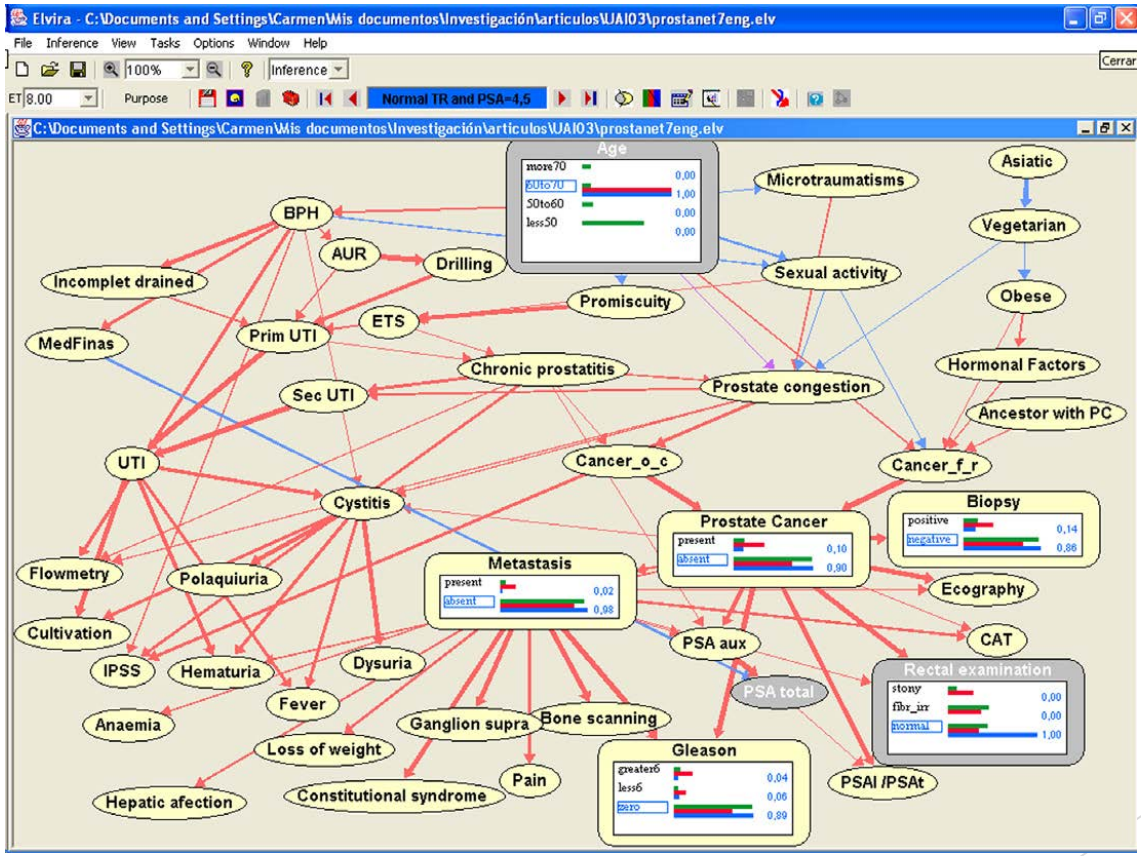


I. Guy, N. Zwerdling, I. Ronen, D. Carmel, E. Uziel, Social media recommendation based on people and tags, SIGIR '10, 2010, pp. 194–201.

Explanation Content

Background and Complementary Information (3/4)

- Relationship between Knowledge Objects

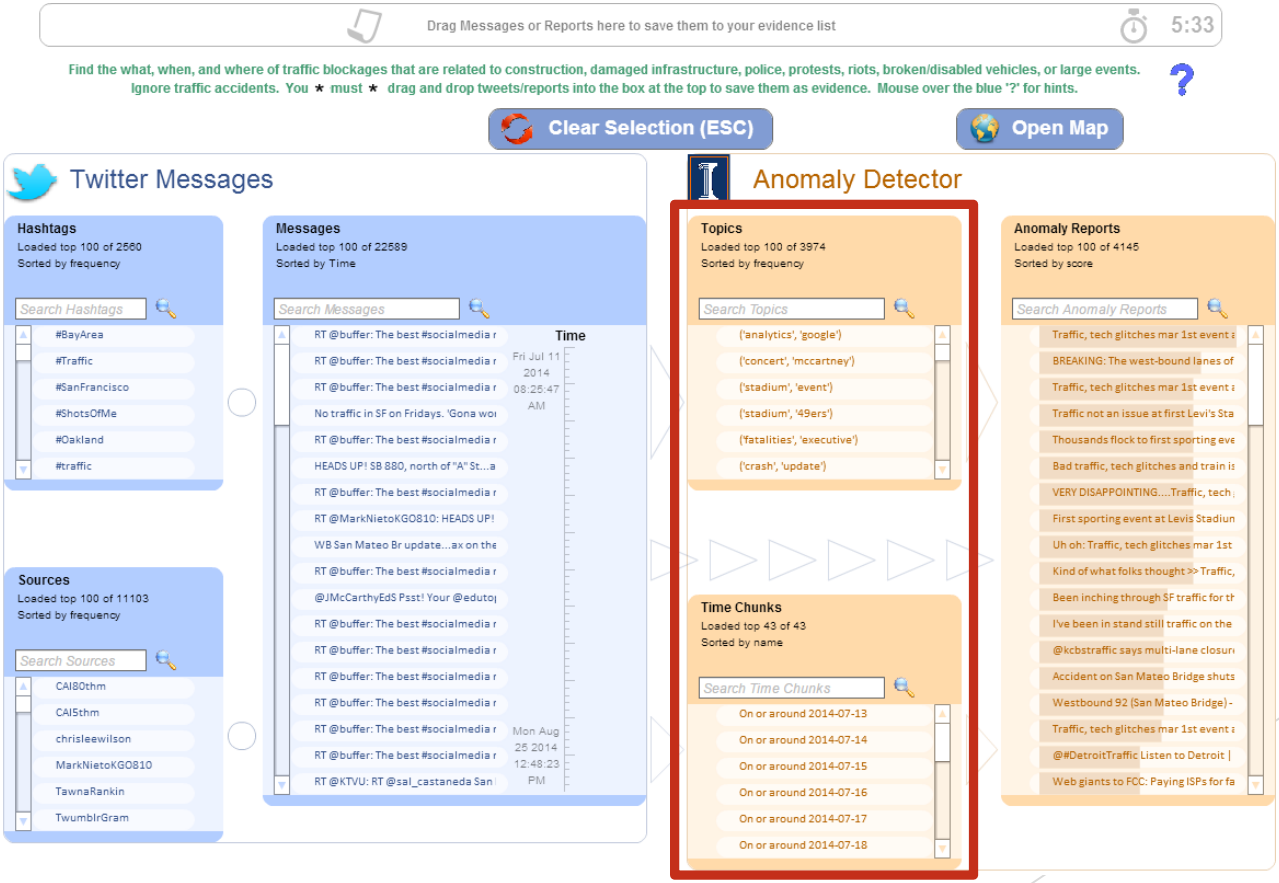


C. Lacave, A. Onisko, F. J. Diez, Use of Elvira's explanation facility for debugging probabilistic expert systems, Knowledge-Based Systems 19 (8) (2006) 730–738.

Explanation Content

Background and Complementary Information (4/4)

- Background Data



J. Schaffer, P. Giridhar, D. Jones, T. Höllerer, T. Abdelzaher, J. O'Donovan, Getting the message?: A study of explanation interfaces for microblog data analysis, IUI '15, 2015, pp. 345–356.

Explanation Content

Alternatives and their Features (1/2)

- Decisive Features

Your prediction is based on how MovieLens thinks you like these aspects of the film:

	Your preference ↓
adventure	★★★★
fun	★★★★
action	★★★★
sequel	★★★★
egypt	★★★½
comedy	★★★½
brendan fraser	★★★½

- Pros and Cons

J. Vig, S. Sen, J. Riedl, Tagsplanations: Explaining recommendations using tags, IUI '09, 2009, pp. 47–56.
K. Muhammad, A. Lawlor, R. Rafter, B. Smyth, Great Explanations: Opinionated Explanations for Recommendations, Springer, 2015, pp. 244–258.

Explanation Content

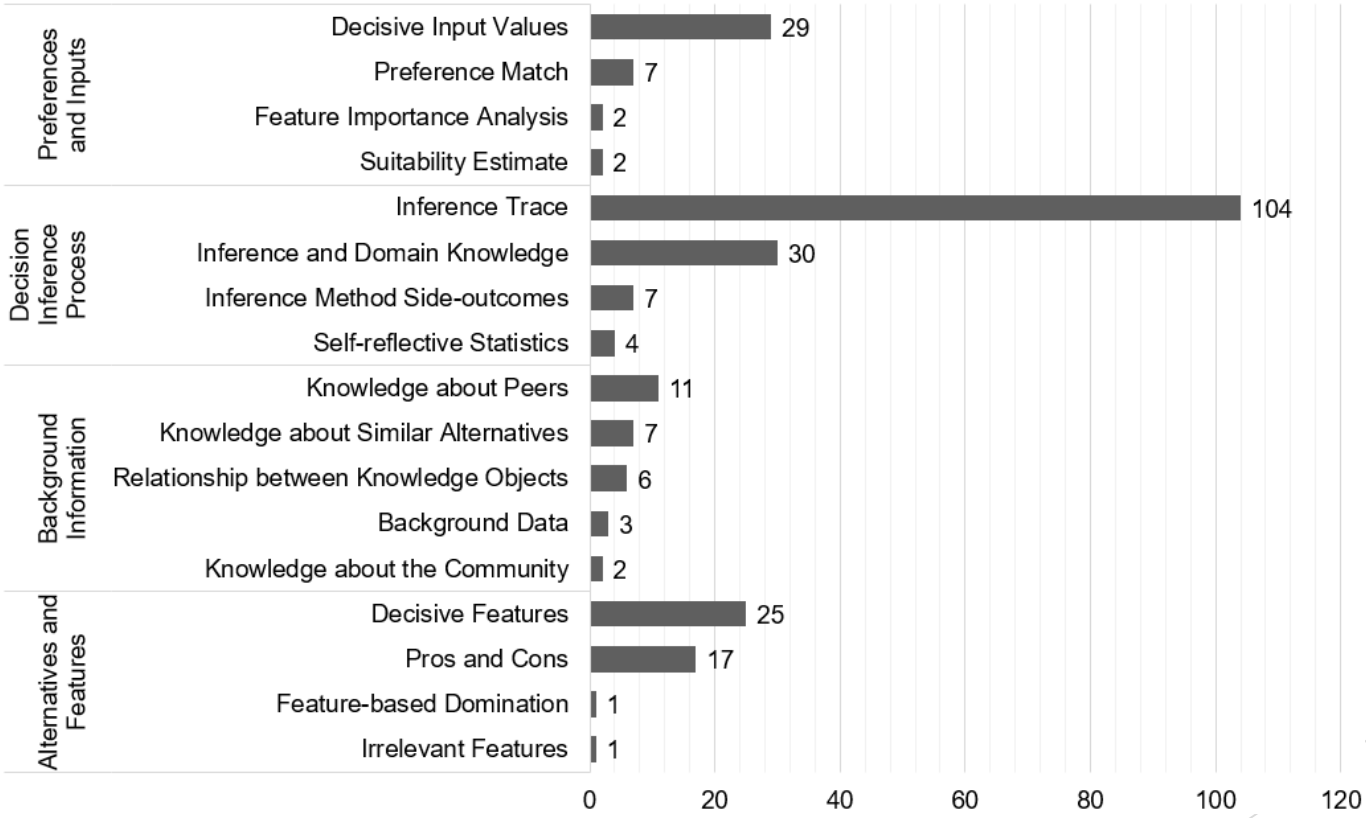
Alternatives and their Features (2/2)

- Feature-based Domination
  - y is preferred to x since y is better than x on **ALL criteria**.
- Irrelevant Features
  - Case 574 differs from your query only in price and is the best case **no matter what transport, duration, or accommodation you prefer**

C. Labreuche, A general framework for explaining the results of a multi-attribute preference model, Artificial Intelligence 175 (7) (2011) 1410–1448.
D. Mcsherry, Explanation in recommender systems, Artificial Intelligence Review 24 (2) (2005) 179–197.

Explanation Content

Statistics



Explanation Content

Additional Observations

- Baselines and Multiple Alternatives
  - Pairwise comparison or comparison with groups

The most popular product

	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard
◉	-	$2'095.00	1.67 GHz	4.5 hour(s)	512 MB	80 GB

We also recommend the following products because

they are cheaper and lighter, but have lower processor speed

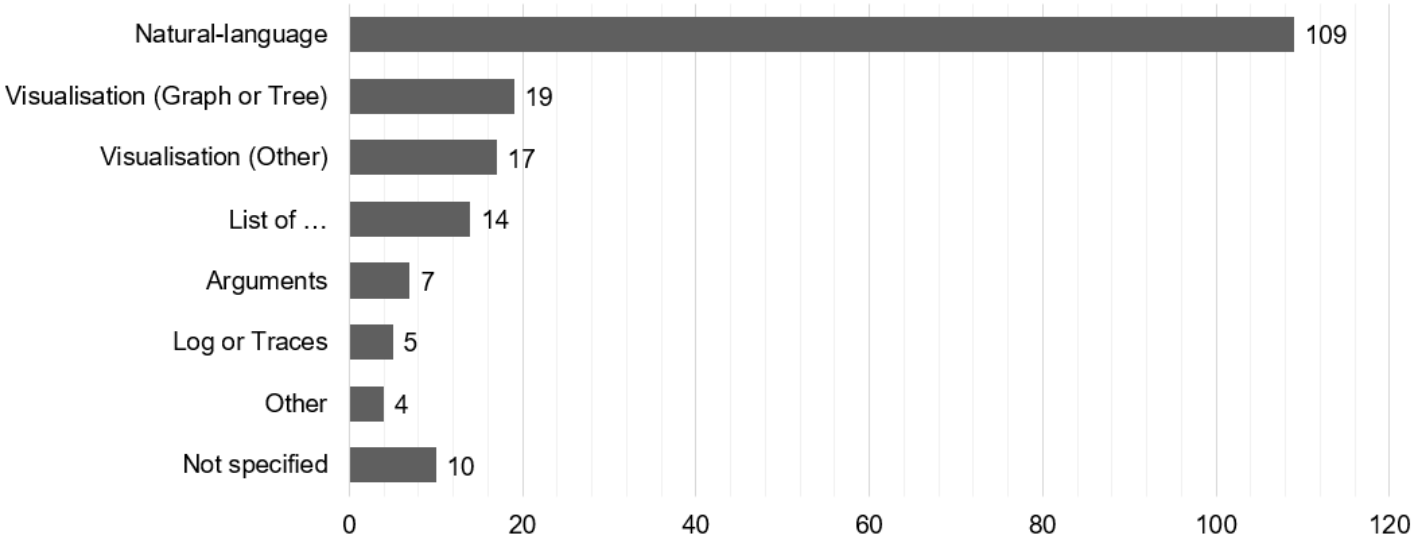
	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard
○	-	$1'499.00	1.5 GHz	5 hour(s)	512 MB	80 GB
○	-	$1'739.99	1.5 GHz	4.5 hour(s)	512 MB	80 GB
○	-	$1'525.99	1.5 GHz	5 hour(s)	512 MB	80 GB
○	-	$1'426.99	1.5 GHz	5 hour(s)	512 MB	60 GB
○	-	$1'929.00	1.2 GHz	4 hour(s)	512 MB	60 GB
○	-	$1'595.00	1 GHz	5.5 hour(s)	512 MB	40 GB

they have higher processor speed and bigger hard drive capacity, but are heavier

	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard
○	-	$1'220.49	1.8 GHz	5 hour(s)	1 GB	100 G
○	-	$2'148.99	2 GHz	4 hour(s)	1 GB	100 G
○	-	$1'379.00	3.3 GHz	2 hour(s)	512 MB	100 G
○	-	$2'235.00	1.8 GHz	2.5 hour(s)	1 GB	100 G
○	-	$2'319.00	1.7 GHz	4.5 hour(s)	512 MB	100 G
○	-	$2'075.00	1.8 GHz	1.67 hour(s)	512 MB	100 G

- Context-tailored Explanations
  - Level of detail tailored to user expertise
- External Sources of Explanation Content
  - Product reviews

Explanation Presentation



- Other
  - Voice
  - Highlighting
  - Query results
  - OWL (Ontology Web Language)

Explanation Generation Approaches

Part III: Explanations

Explanation Generation
Key Observations

- Few approaches provide sophisticated means of generating explanations
  - Use of knowledge data and inference process output

- Exceptions
  - Multi-criteria utility theory (MAUT)
    - Mathematical analysis of attribute weights and values
  - Artificial neural networks (ANN)
    - Rule extraction

Explanation Generation

Key Observations

- 18 (17.8%) approaches are **domain-specific**
  - E.g. domain-specific argument templates

Postulate/Argument structure
Postulate 1: < recommend (Movie, User), good_movie (Movie) $\prec$ avg_rating (Movie) > 3.8 >
Postulate 2: < recommend (Movie, User), likes_by_top genre (Movie, User) $\prec$ top_genre (User, Genre), genre (Movie, Genre) >
Postulate 3: < recommend (Movie, User), likes_by_top actor (Movie, User) $\prec$ top_actor (User, Actor), leads_in (Movie, Actor) >
Postulate 4: < recommend (Movie, User), likes_by_top actor (Movie, User) $\prec$ top_actor (User, Actor), leads_in (Movie, Actor), top_ genre (User, Genre), genre (Movie, Genre) >

Explanation Generation

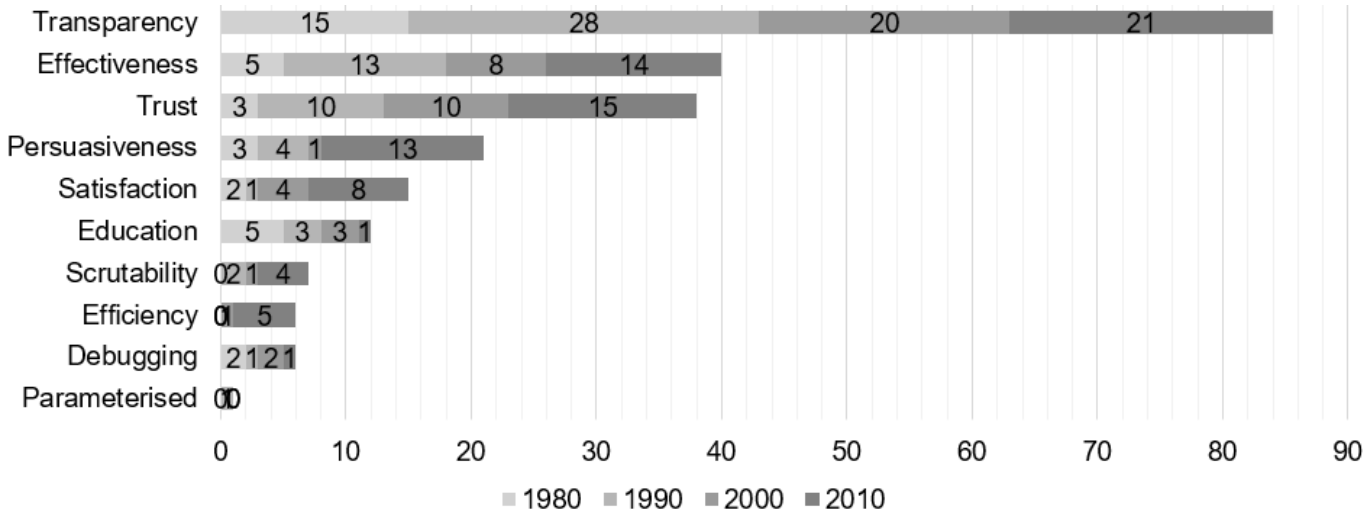
Key Drivers: (Intended) Purpose

Purpose	Description
Transparency	Explain how the system works
Effectiveness	Help users to make good decisions
Trust	Increase users' confidence in the system
Persuasiveness	Convince user to try or buy
Satisfaction	Increase the ease of usability or enjoyment
Education	Allow users to learn something from the system
Scrutability	Allow users to tell the system it is wrong
Efficiency	Help users make decisions faster
Debugging	Allows users to identify that there are defects in the systems

N. Tintarev, J. Masthoff, A survey of explanations in recommender systems, ICDE 2007.
B. G. Buchnan, E. H. Shortliffe, Explanations as a Topic of AI Research, in Rule-based Systems, 1984.

Explanation Generation

Key Drivers: (Intended) Purpose



- Transparency
  - Motivation: "[the] human user bears the ultimate responsibility for action" Moulin et al. [2002]
  - Means of achieving trust
- Persuasiveness, satisfaction, scrutability, and efficiency received more attention recently

Explanation Generation

Key Drivers: Decision Inference Method

Category	Subcategory	1980	1990	2000	2010	Total	%
Knowledge-based		33	50	28	31	142	68.3%
	Rule-based	*28*	*33*	*11*	*7*	*79*	*55.63%*
	Logic-based	*2*	*3*	*3*	*8*	*16*	*11.27%*
	Multi-Criteria Decision Making	*0*	*1*	*3*	*5*	*9*	*6.34%*
	Constraint-based	*0*	*1*	*1*	*0*	*2*	*1.41%*
	Case-based Reasoning	*0*	*1*	*3*	*2*	*6*	*2.88%*
	Other	*3*	*11*	*7*	*9*	*30*	*21.13%*
Machine Learning		2	13	12	24	51	24.5%
	Feature-based	*2*	*13*	*6*	*11*	*32*	*62.75%*
	Collaborative-filtering	*0*	*0*	*5*	*5*	*10*	*19.61%*
	Hybrid	*0*	*0*	*1*	*8*	*9*	*17.65%*
Mathematical Model		0	2	0	0	2	1.0%
Human-made Decision		0	1	0	1	2	1.0%
Algorithm-independent		0	3	3	5	11	5.3%

Evaluation: Design & Conclusions

Part III: Explanations

Evaluation Design

Presence of an Evaluation

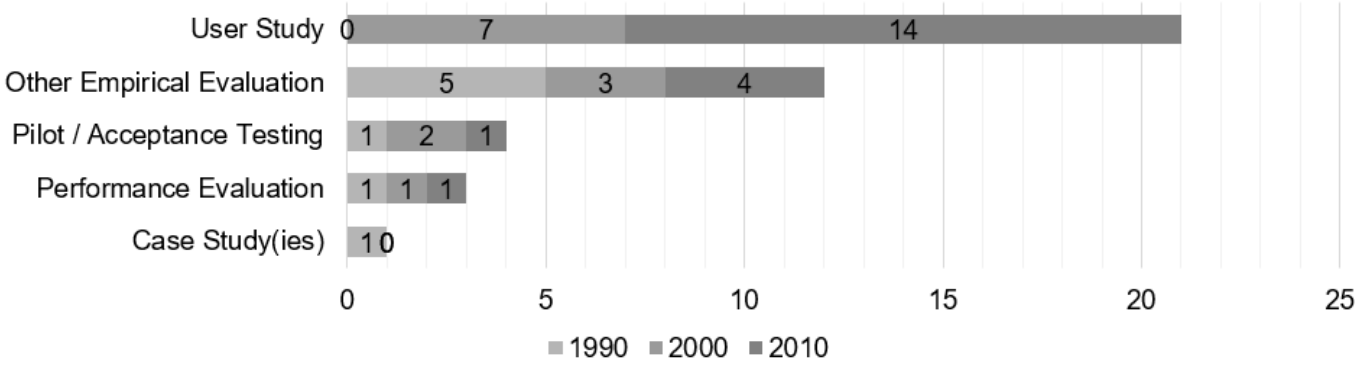
- How many of the explanation generation approaches were evaluated?

	1980	1990	2000	2010	Total
Technique	0%	22%	50%	46%	35%
Tool	0%	8%	6%	7%	6%
Total	**0%**	**13%**	**30%**	**36%**	**21%**

- Past
  - Lower methodological requirements
- Present
  - Still not the majority
  - Some approaches
    - Main contribution is an **algorithm** (explanation is an add-on)

Evaluation Design

Evaluation Types



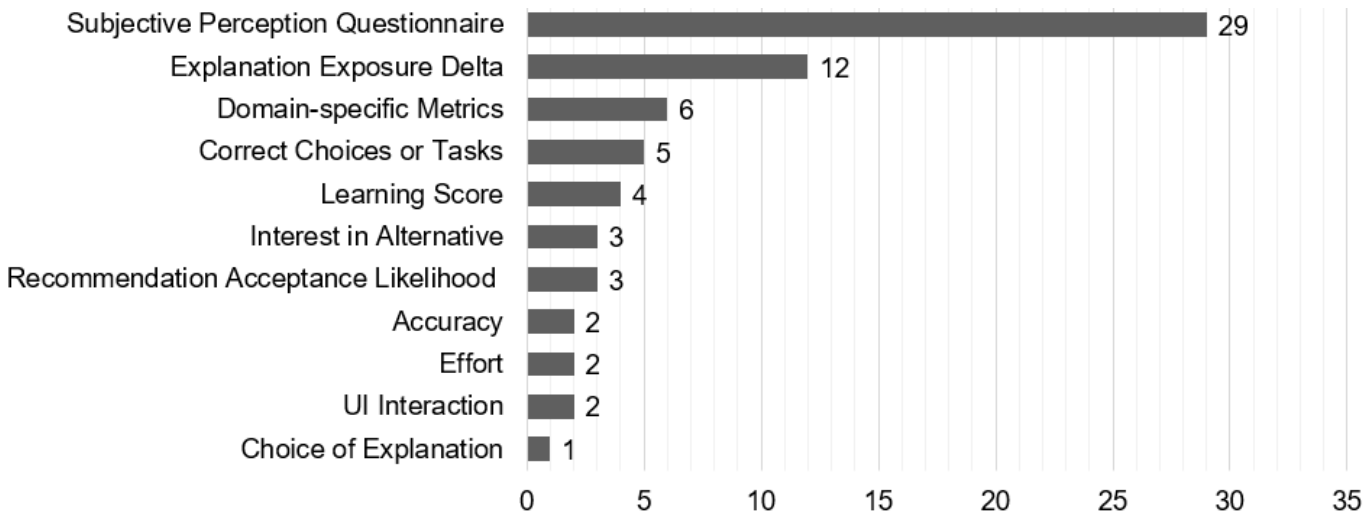
- User study is the most frequent
  - All 22 evaluation studies involve user studies
- Other empirical evaluation
  - Different choices, e.g. explanation coverage
- Most explored domains
  - Media Recommendation, Health, and (e-)Commerce

Evaluation Design
User Studies: Independent Variables

- Single treatment
- With explanations vs. no explanations
- Alternative explanations
  - Most frequent
  - From 2 to 9, exception 21
  - May include no explanations
- Alternative user interfaces/RS versions
  - From 2 to 9
- Other
  - confidence, direction, length, robot ability, …

Evaluation Design

User Studies: Dependent Variables



M. Bilgic and J. Mooney, Explaining recommendations: Satisfaction vs. promotion, in: Proceedings of Beyond Personalization 2005: A Workshop on the Next Stage of Recommender Systems Research at the 2005 International Conference on Intelligent User Interfaces, San Diego, CA, 2005

Evaluation Design

User Studies: Dependent Variables

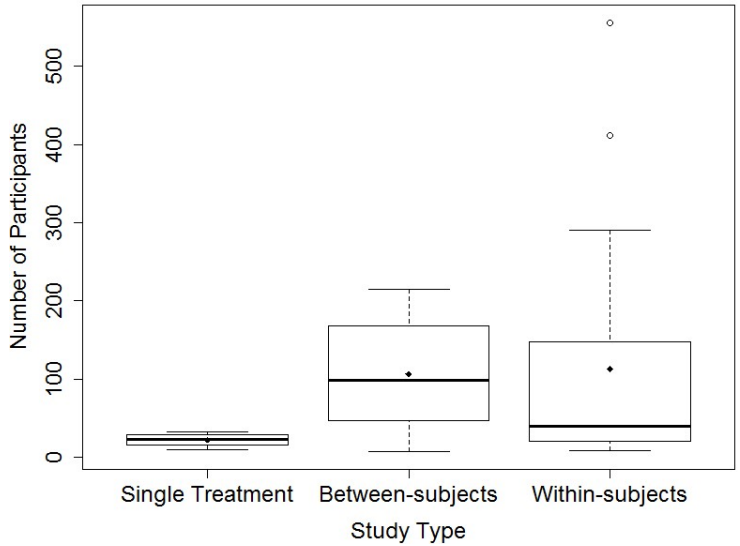
- Subjective perception questionnaire example
  - How much do you think the explanation was helpful for you to make better decisions? → efficiency*
  - How good do you think the explanation was? → satisfaction

*According to Tintarev and Masthof's definition, this should be effectiveness. However, the authors used the term efficiency in the paper: "*W. Hanshi and F. Qiujie and L. Lizhen and S. Wei, A probabilistic rating prediction and explanation inference model for recommender systems, China Communications, 2016, pp. 79-94.*"

Evaluation Design

User Studies: Sample Size

- Largest study: within-subjects (556 participants)
- Highest mean: between-subjects
- Single treatment studies: low number of participants
- Number of factors do not influence sample size



Evaluation Results

User Studies

Purpose	Positive	Neutral	Negative
Effectiveness	++++++	~~~~	
Transparency	++++++	~	
Persuasiveness	++		-
Satisfaction	+++		
Trust	++++	~~	
Usefulness	+++		
Ease of Use	+	~~	
Efficiency	+		- -
Education	+	~~	

Evaluation Results
User Studies

- Effectiveness
  - Divergent results
  - Specific explanation styles are more effective
  - Presence of confounding variables?
- Persuasiveness
  - Social information (e.g. ratings) is more persuasive
- Transparency and most of user-centric purposes
  - Generally positive results
- Ease of Use
  - Extra information decreases usability
  - Organisation of recommendations has a positive effect
- Efficiency
  - Follows ease of use

Evaluation Results

User Studies

- Personalisation
  - +: Satisfaction and Transparency
  - -: Effectiveness and efficiency
- Expertise Levels
  - Experts and novices have different preferences over explanations
- Explanation Direction
  - Negative explanations are more influential than positive explanation
- Explanation Confidence and Length
  - Strongly confident and long explanations are more persuasive

Evaluation Results

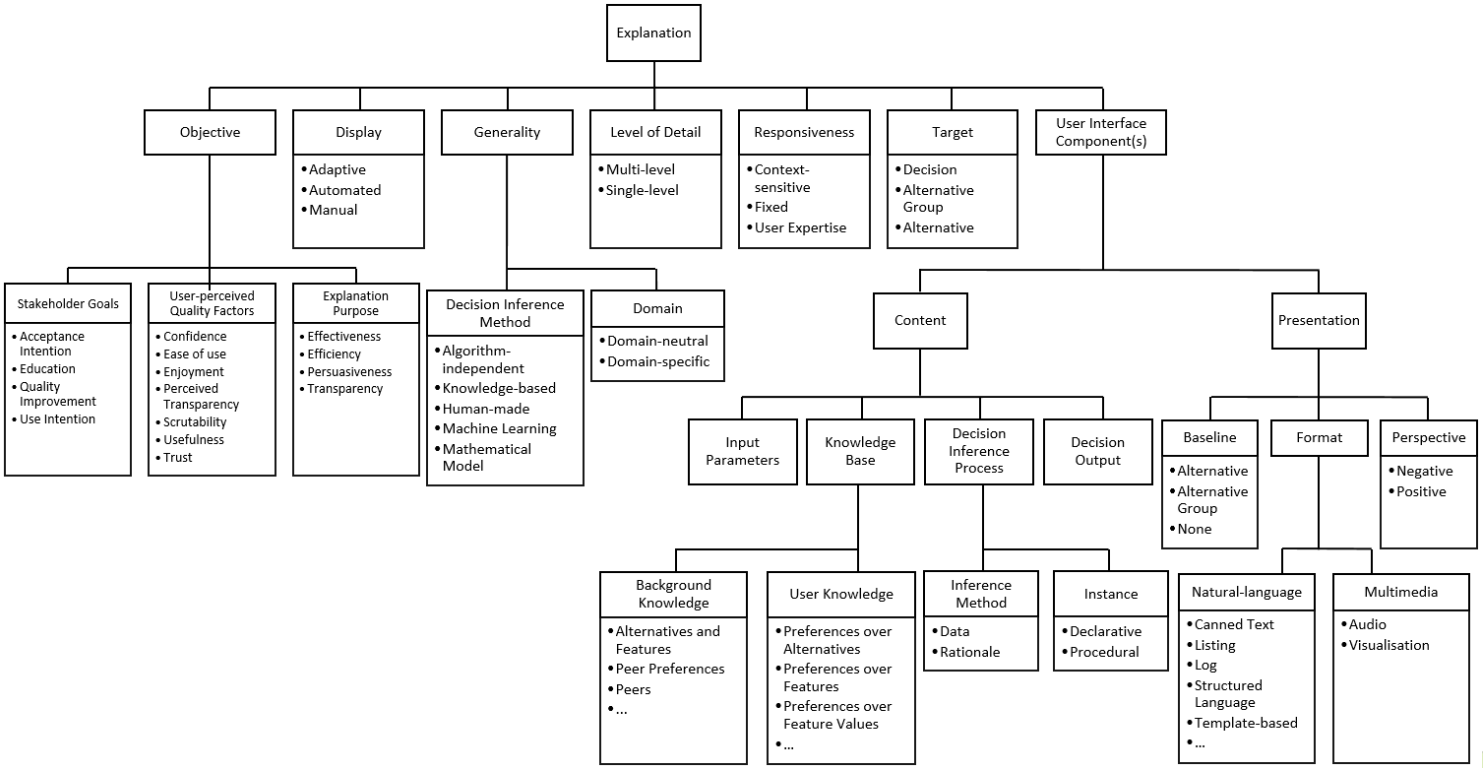
Foundational Studies

- Tintarev and Masthoff (*RecSys 2007*)
  - Explanations must be customised to the user and the context, by selecting features from the suggested alternative accordingly
- Nunes et al. (*UMAP 2012*)
  - Guidelines and patterns
  - Explanations should be concise and focus on the most relevant criteria
- General aspects related to explanations
  - Mental models: improved decisions if users understand ES's reasoning process and provided information
    - *Rook and Donnell, IEEE Trans. on Systems, Man, and Cybernetics, 1993*
  - Cognitive fit: increased KBS acceptance if the KBS explanations match users' internal explanations
    - *Giboney et al., DSS, 2015*
  - Explanations have greater usefulness when decision support systems are used as a cooperative problem solving tool
    - *Gregor, IJHCS, 2001*

A Comprehensive Explanation Taxonomy

Part III: Explanations

Explanation Taxonomy



Explanation Taxonomy

Objective

Stakeholder Goals

- Acceptance Intention
- Education
- Use Intention
- Quality Improvement

User-perceived Quality Factors

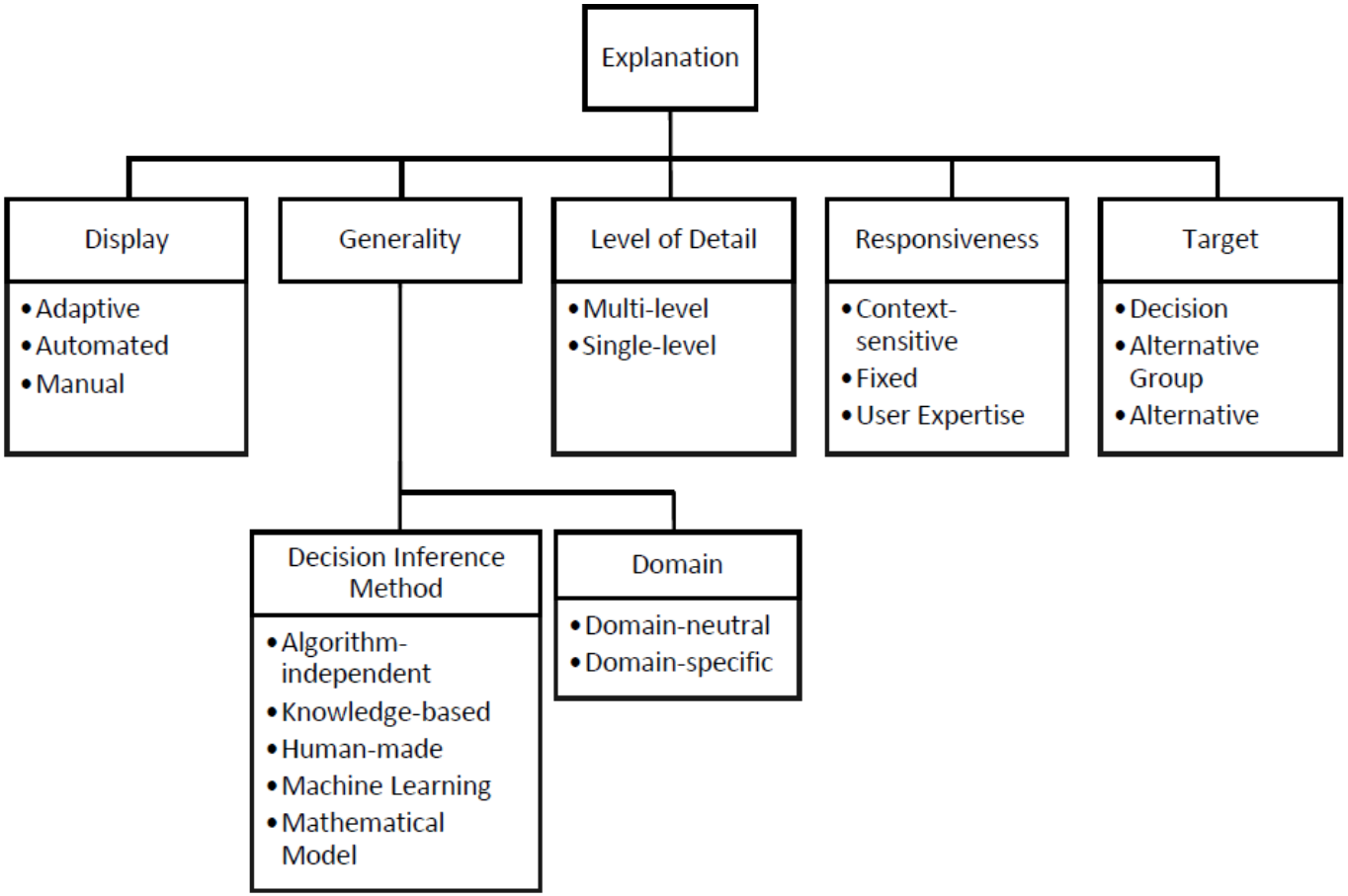
- Confidence
- Ease of Use
- Enjoyment
- Perceived Transparency
- Scrutability
- Usefulness
- Trust

Explanation Purpose

- Effectiveness
- Efficiency
- Persuasiveness
- Transparency

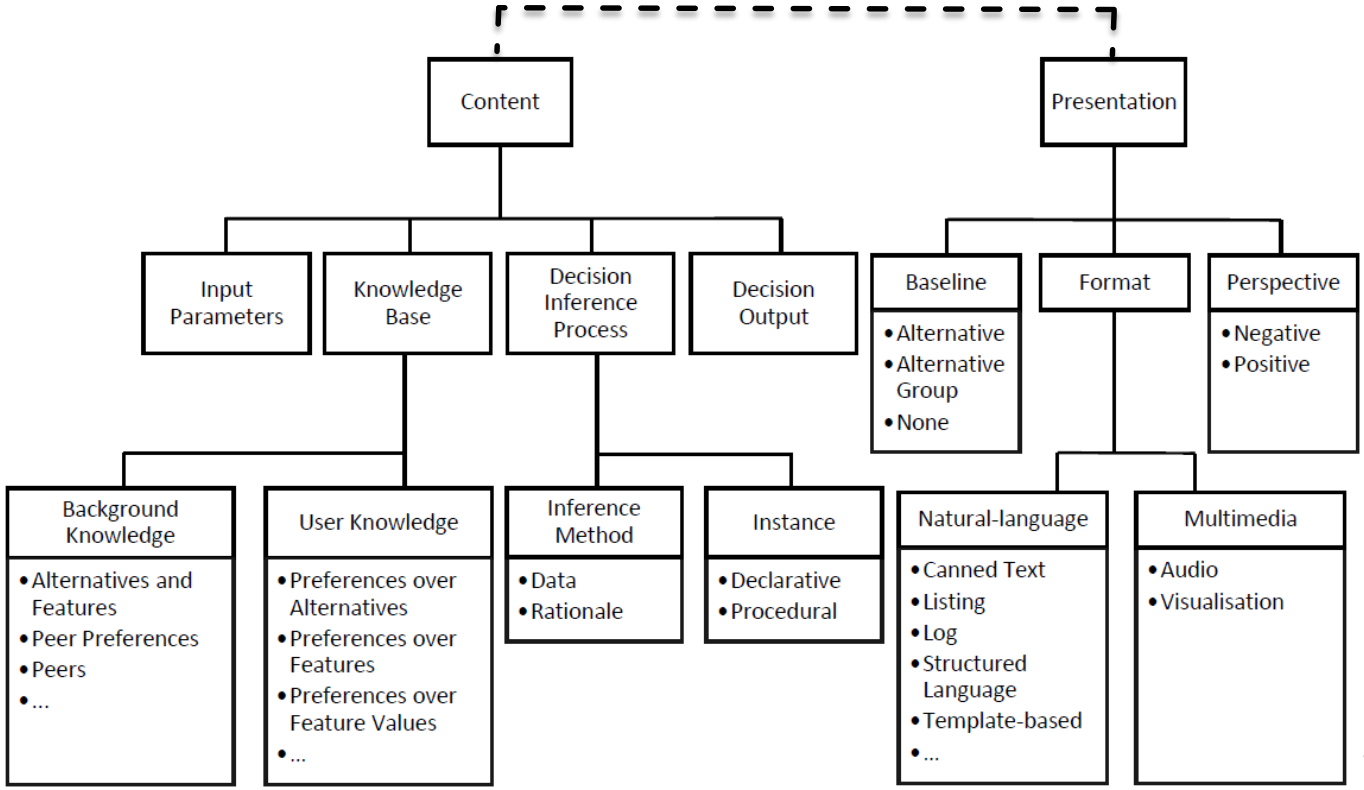
Explanation Taxonomy

General Facets



Explanation Taxonomy

User Interface Component(s)



Open Challenges

- Understanding the relationship among stakeholder goals, user-perceived quality factors, and explanation purposes
- Selecting the right explanation content
  - Inference traces: a rule in the past
  - Today: why an alternative is adequate helpful for users
- Investigating fine-grained details of presentation aspects
  - Length, vocabulary, format
  - Real transparency vs. perceived transparency
- Towards more responsive explanations
  - Only 16 (8%) approaches consider this
- The need for adequate objective evaluation protocols and metrics
  - Objective measurements

Interacting with Recommender Systems

Conclusions and Discussion

Summary

- A variety of advanced user interaction approaches have been proposed over the years
  - Nonetheless, a niche topic in academia
  - Community focuses on algorithms, which is only one component of a recommender system
  - UI design however a relevant success factor in practice
  - “Academic” UI designs often comparably complex

Main Challenges

- Many open questions raised throughout the tutorial
  - How to acquire user preferences
  - How to present results and collect feedback
  - How to explain (with a certain purpose in mind)
- Methodological aspects
  - User studies of various designs as main instrument
  - Nearly no standard protocols or evaluation frameworks exist
- Interdisciplinary approach required
  - Within computer science and outside

Future directions

- Personalized interaction
  - Different UIs for different users
  - User feedback
- Natural language interfaces
  - Voice control, chat bots
- Designing interaction patterns with a purpose in (e.g., persuasion)
- Augmented reality/virtual reality
  - Interact in the real/virtualized world
- More context-aware and adaptive UIs
  - Mobile devices, new sensors

Thank you

- Dietmar Jannach
  - dietmar.jannach@tu-dortmund.de
- Ingrid Nunes
  - ingridnunes@inf.ufrgs.br
- Michael Jugovac
  - michael.jugovac@tu-dortmund.de

- Upcoming paper:
  Jugovac, M., Jannach, D.: Interacting with Recommenders – Overview and Research Directions, ACM Transactions on Intelligent Interactive Systems (TiiS).