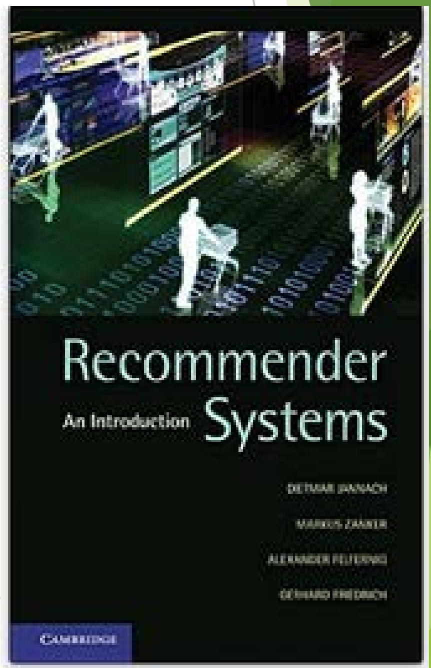
Recommender Systems: More than algorithms

Dietmar Jannach
TU Dortmund, Germany
dietmar.jannach@tu-dortmund.de

RecSys Summer School, Bozen-Bolzano, August 2017

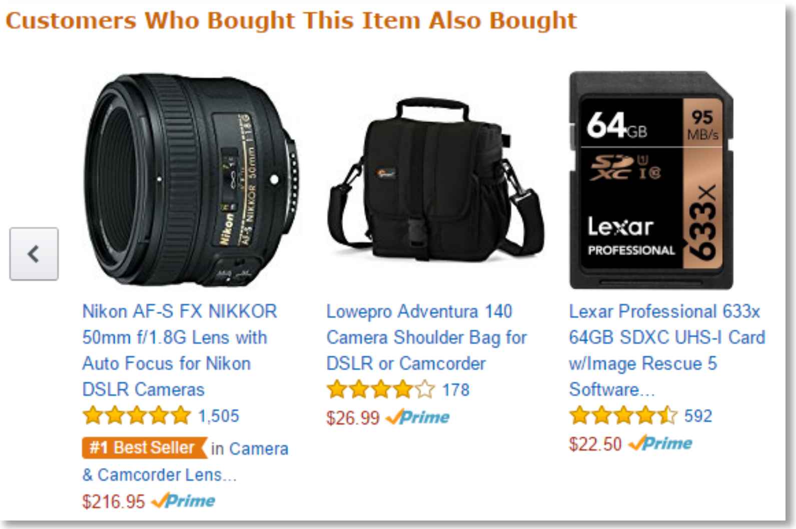
About me

- Professor of Computer Science at TU Dortmund, Germany
- Co-founder of a tech company selling interactive selling solutions (2003-2008)
- Research interests
  - Recommender Systems
    - E-Commerce applications, business value of recommenders
    - Interactive advisory systems
  - Artificial Intelligence
    - Model-based Diagnosis, Constraints
  - Software Engineering
    - Debugging of Spreadsheets



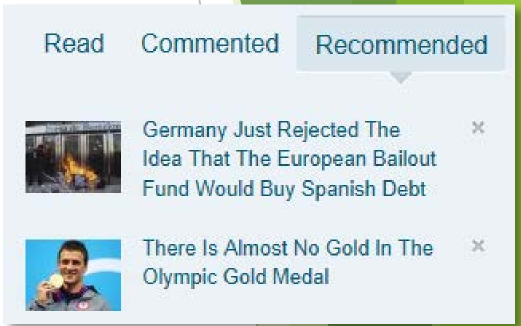
Recommender Systems

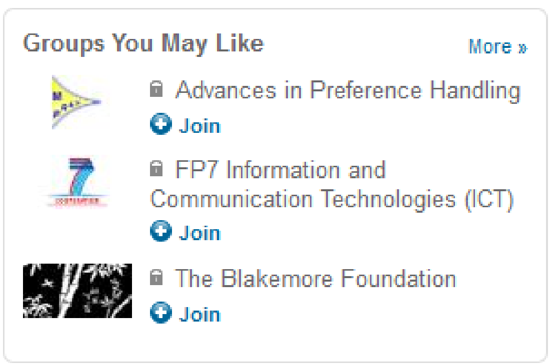
- Automated recommendations
  - A pervasive part of our online user experience
  - Explicitly recommend us shopping items, movies, music, news, friends, jobs, groups or people to follow, restaurants, hotels…
  - Less obvious: Silently select and rank the items
    - News feeds, ads (in some sense)



Recommender Systems

- Once you see them, they are everywhere





This summer school

- Mostly focuses on a computer science oriented perspective, e.g.,
  - Algorithms to process past user actions to predict the relevance of an item for a certain user
  - Information sources that be fed into our algorithms
  - Ways to make our algorithms scale to millions of users and items
  - Methods of measuring that our algorithms are better than previous ones
  - UI mechanisms to acquire user preferences and to present the recommendations
- Algorithm help us predict to which extent an item is generally relevant for an individual user or a group of users
  - Which is a central question

But there is more

- Recommender systems research is not only about algorithms and application design
- Automated recommendations have an effect on recommendation consumers and providers
  - They change the consumer behavior
  - They have an effect on the business
- Some challenges
  - These effects are not always easily predictable
  - There might be conflicting goals
  - Some effects are based on psychological effects
  - And may depend on a variety of other factors, including user trust or website credibility

In this talk

- We will briefly review the history of recommender systems
- We will outline challenges when adopting a purely algorithmic-oriented research perspective
- We sketch a purpose-oriented framework for the design and evaluation of recommender systems

---

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

Jannach, D. and Adomavicius, G.: "*Recommendations with a Purpose*". In: Proceedings of the 10th ACM Conference on Recommender Systems (RecSys 2016). Boston, Massachusetts, USA, 2016, pp. 7-10

A bit of history

- Recommender systems have their roots in various fields
  - e.g., Information Retrieval, Machine Learning, Human Computer Interaction
- Their design can furthermore be influenced by insights from more distant fields
  - e.g., Consumer behavior, psychology, marketing
- Typical goals:
  - Avoid information overload (filtering)
  - Active promotion of content
- Personalization often as a central concept

Information Filtering roots

- Information Filtering
  - Systems that filter incoming streams of information in a personalized way
  - Dates back to the late 1960s
  - Early systems use explicitly stated preferences regarding topics or keywords
  - Later on, automated content analysis and user profiling
- Today:
  - “Content-based Filtering” recommender techniques
  - Personalized Information Retrieval

Leveraging the opinions of others

- 1982: ACM president complained about email junk
  - Envisioned a set of "trusted authorities" that assess the quality of the messages
- 1987: Information Lens
  - Based on manual filters, but could also specify people whose opinions they value
- 1992: Tapestry – "Collaborative Filtering"
  - Continued Information Lens ideas, introduced idea of considering ratings, but still a manual process
- 1994: GroupLens and others
  - System automatically predicted ratings of users, based on "matrix filling" (completion) setup

Collaborative Filtering booms

- The Matrix Completion problem
  - Became established as a standard way of operationalizing research
  - Problem of predicting missing ratings
  - Evaluate algorithms
    - Prediction error, rank measures

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

Collaborative Filtering (CF) booms

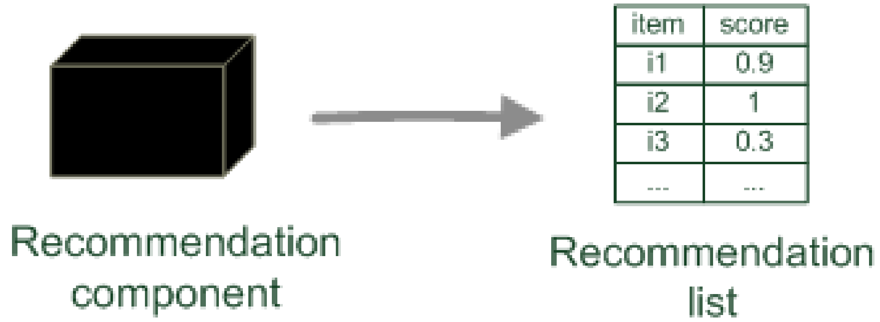
- 1998:
  - Dimensionality reduction for CF, clustering
  - Collaborative/Content-based Hybrids
- 1999: It works in e-commerce!
  - First reports on successful applications in practice (e-commerce, music, video)
- 2000:
  - Item-to-item collaborative filtering
- 2003: Amazon.com
  - Report on the successful use of recommendations at Amazon.com using item-to-item filtering
  - Today, many algorithms, many non-personalized ones as well

The Netflix Prize

- Netflix announced a 1 million $ prize in 2006
  - For beating their system by 10% in terms of the prediction error
  - Provided at that time huge dataset
- Effects
  - Further boosted research on the matrix completion problem
- Contest ended in 2009, some winning ingredients
  - Matrix factorization (not using exact SVD)
  - Ensemble methods
- Today
  - Collaborative Filtering as a standard method in industry

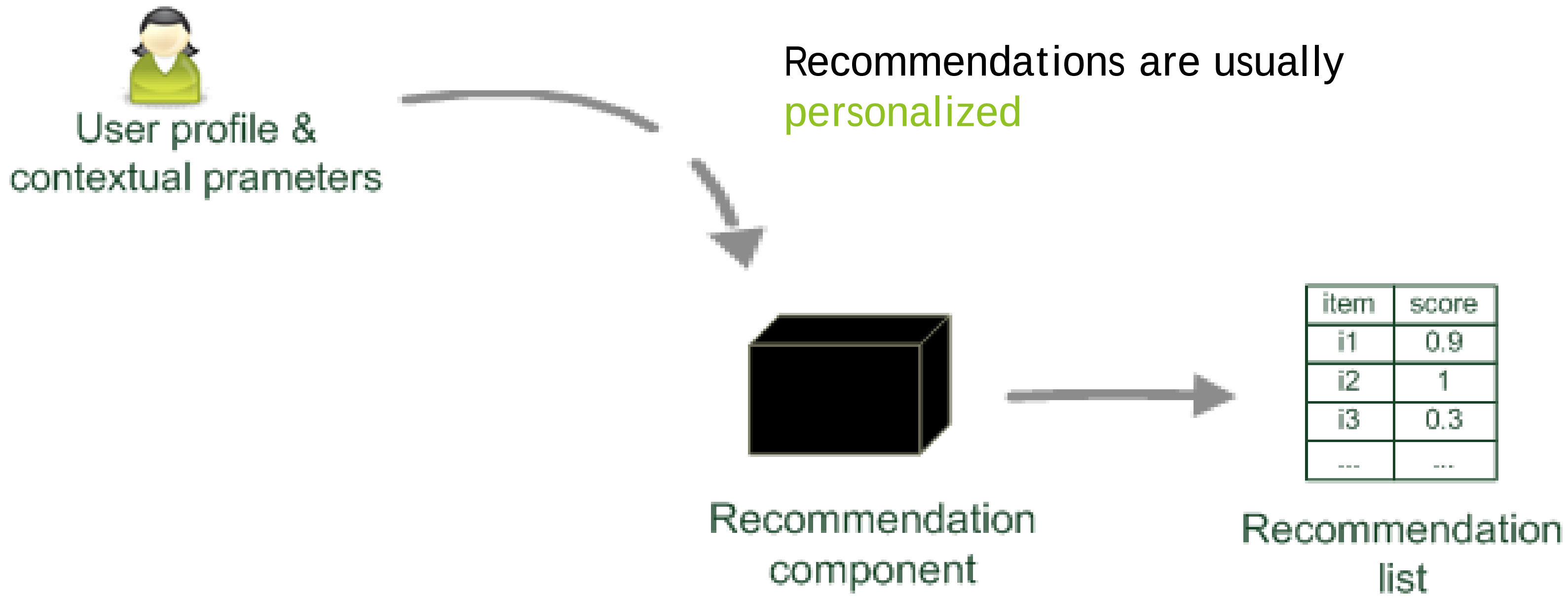
Recommendation Paradigms

Recommender systems reduce information overload by estimating relevance



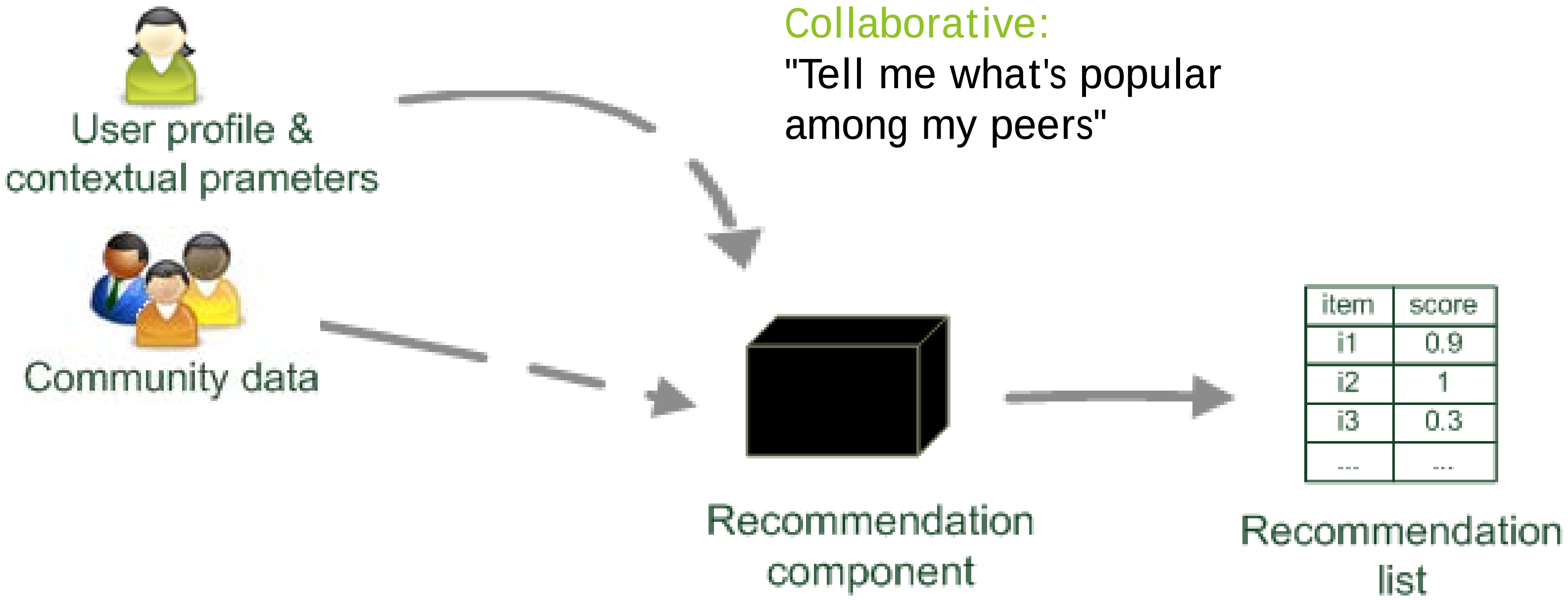
Recommendation Paradigms

Recommendations are usually personalized



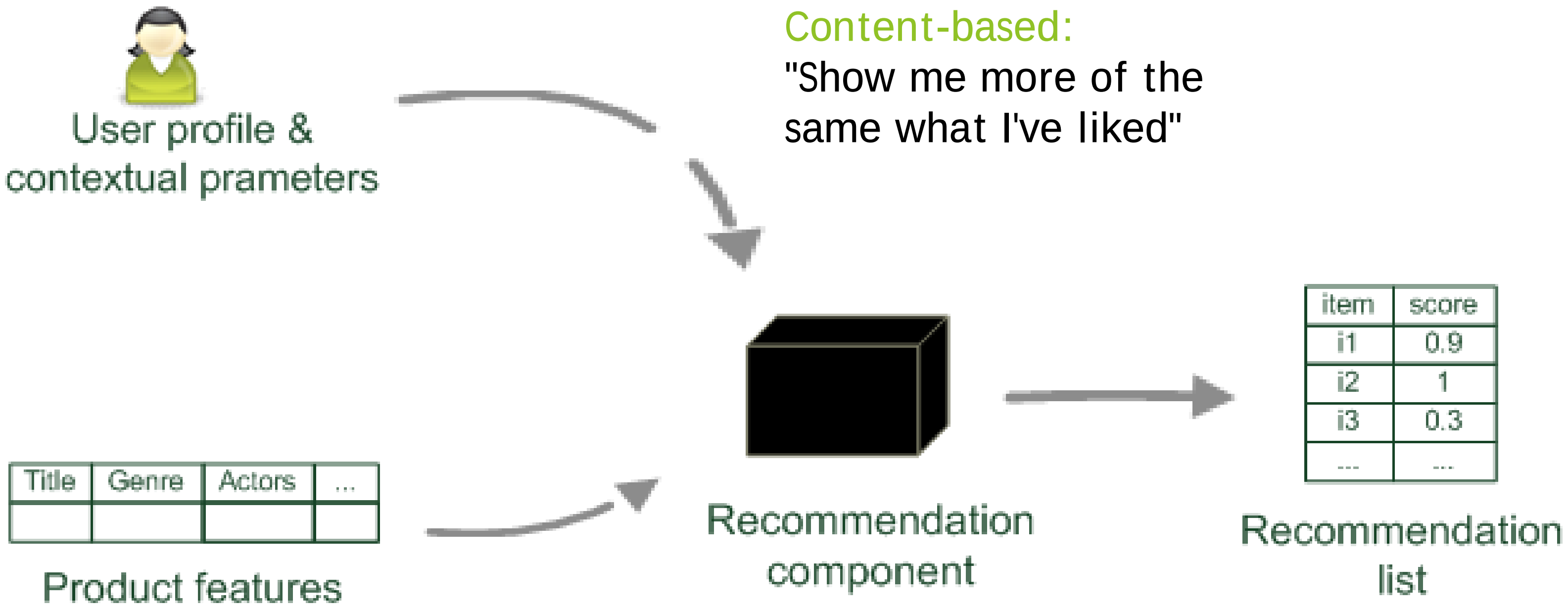
Recommendation Paradigms

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



item	score
i1	0.9
i2	1
i3	0.3
...	...

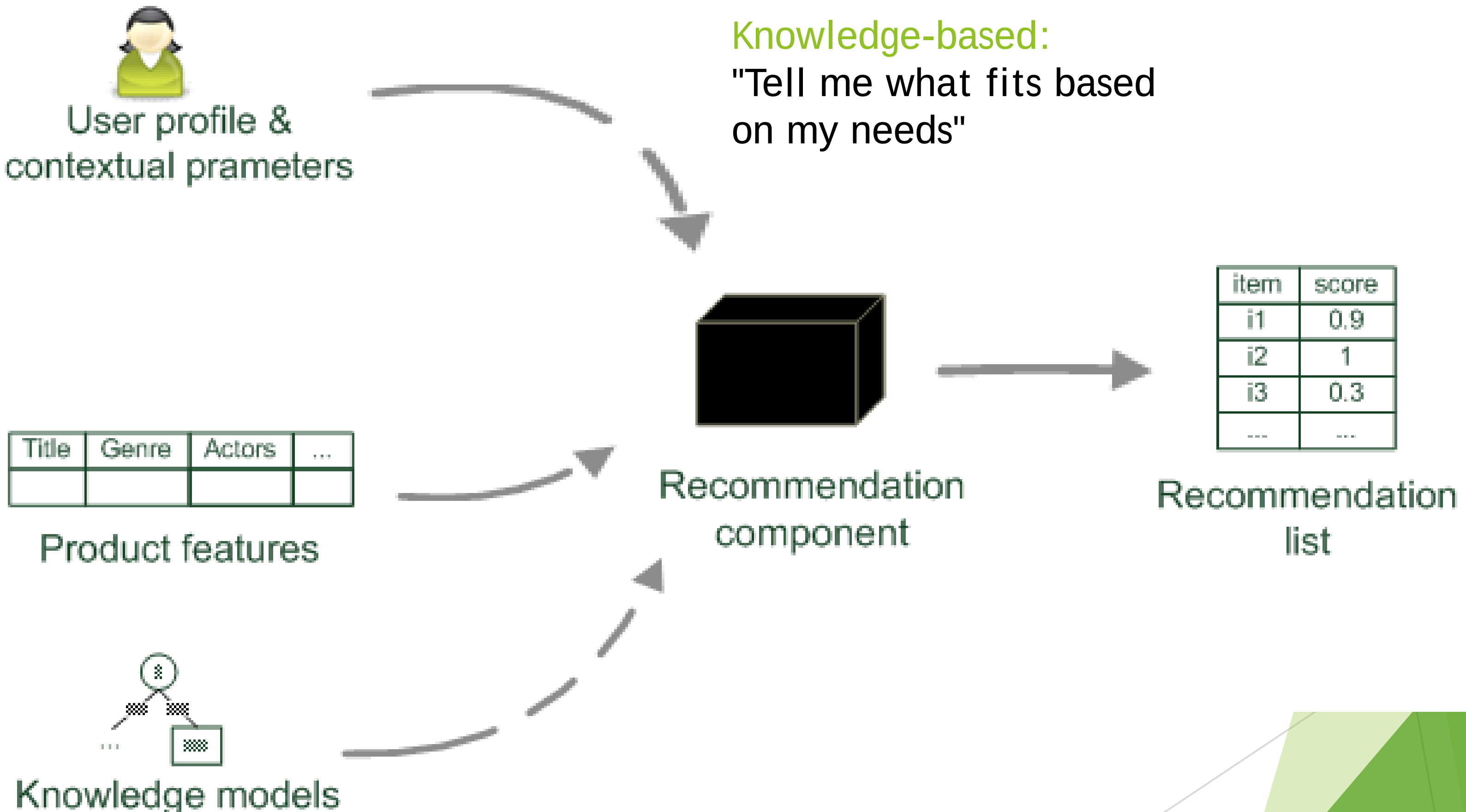
Recommendation Paradigms



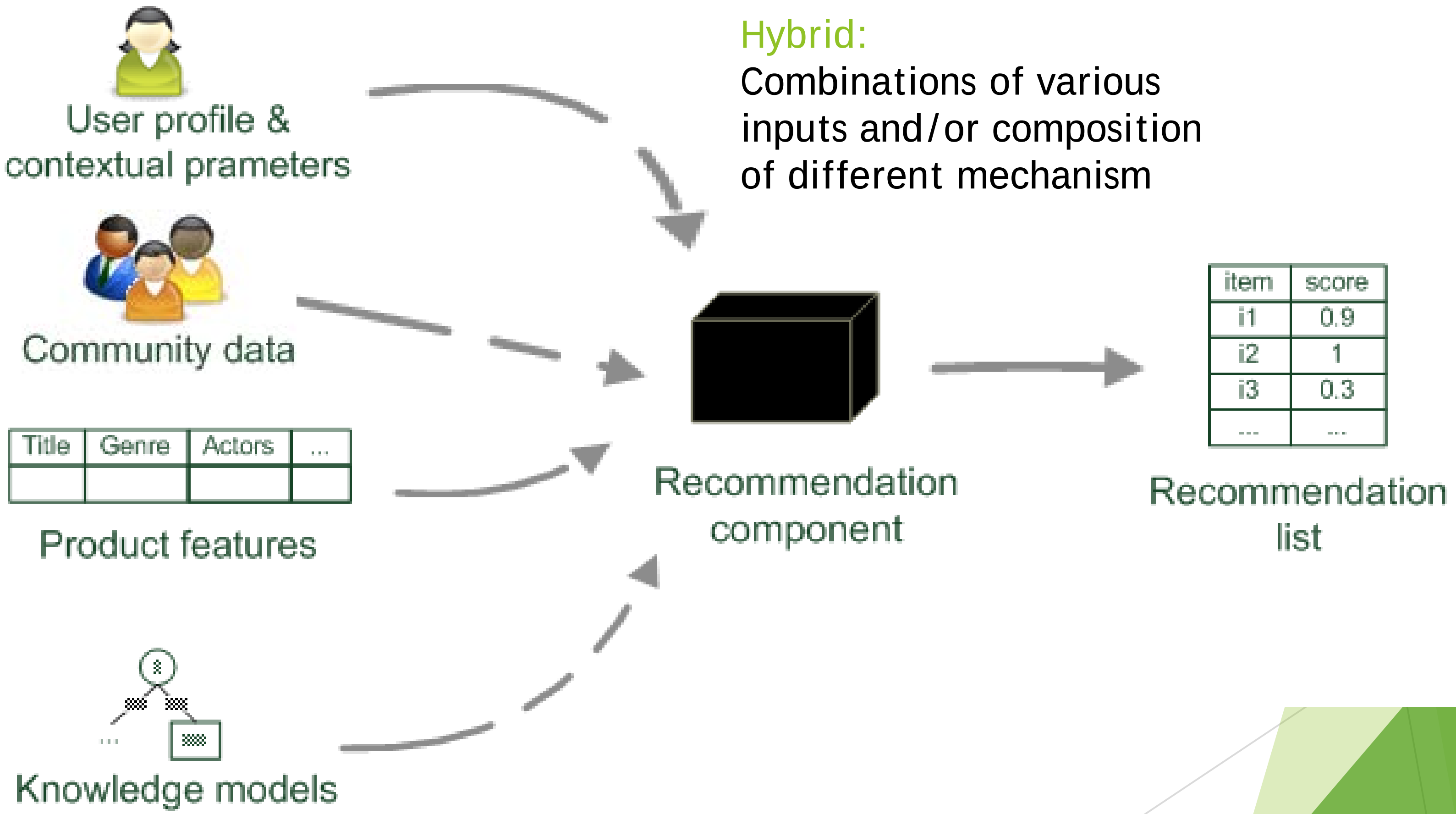
Content-based:
"Show me more of the same what I've liked"

Recommendation Paradigms

Knowledge-based:
"Tell me what fits based on my needs"



Recommendation Paradigms



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

In this talk

- We will briefly review the history of recommender systems
- We will outline challenges when adopting a purely algorithmic-oriented research perspective
- We sketch a purpose-oriented framework for the design and evaluation of recommender systems

Beyond matrix completion

- Research based on the matrix completion problem formulation still predominant today
  - An established and domain-independent problem abstraction
  - Established evaluation procedures
  - Allows for reproducibility, in theory
  - Hundreds of papers published each year
  - Plethora of technical approaches, many of them comparably complex
- The problem formulation, in combination with some surrounding effects, however has its limitations

Beyond Matrix Completion

- Problem setup and data
  - Post-diction is not prediction
  - Benchmark dataset issues
  - Single-interaction assumption
- Accuracy addiction and other components of utility
- Context matters
- What about the user interface?
- What about long-term effects?

Beyond Matrix Completion

- Problem setup and data
  - Post-diction is not prediction
  - Benchmark dataset issues
  - Single-interaction assumption
- Accuracy addiction and other components of utility
- Context matters
- What about the user interface?
- What about long-term effects?

Rating prediction problems

- Rating prediction not often relevant in practice
  - (Relevance or utility prediction, however, are)
  - Item ranking more important
  - Learning-to-rank methods important in recent years
- "Post-diction" is not prediction
  - Users do not rate items at random
  - Algorithm optimization and evaluation procedures however only considers the rated items
  - Models optimized for known ratings might not work best in the real world
    - More and more field tests (A/B tests) done, often confirming this issue

Data set and evaluation issues

- Public (rating) datasets help us compare our methods
  - But how representative is non-contextualized movie recommendation?
  - How do the findings generalize?
  - Ratings as quality assessments or as joy of experience?
- However:
  - Many different accuracy measures and evaluation protocol variants
  - Non-public datasets or data sampling
  - No source code provided

Data set and evaluation issues

- Evaluation focused on finding good items
  - Avoiding “bad” recommendations might also be important
  - Accurately predicting the 1-star or 2-star ratings might be of limited value

Single-interaction assumption

- There is at most one preference signal per user and item
  - In reality there are a lot of implicit signals over time
  - Implicit feedback based algorithms often use the matrix completion setup as well
- Reminding users of known items not in the scope
  - Even though it might be relevant in practice
  - Recent log analysis from e-commerce shop
    - 40% of “successful” recommendations were viewed before by the user
  - Recent field test with reminders
    - Reminding can be helpful also for the business

Lerche, L., Jannach, D. and Ludewig, M.: *"On the Value of Reminders within E-Commerce Recommendations"*. In: Proceedings UMAP 2016. Halifax, Canada, 2016.

What about short-term intents?

- In many domains, users visit a site with a recommender with a very specific intent
  - To purchase something specific, to listen to a special type of music
- Recommending only based on long-term preference models (e.g., using a rating matrix) might be insufficient
  - This particular type of "context" is usually not covered by the matrix completion formulation
- A highly-relevant problem in practice

Problem illustration

- Being able to predict which kinds of things a certain user generally likes, is important
- However, assume you visit your favorite online shop, and here's what you looked at or purchased during the last weeks



- Now, you return to the shop and browse these items

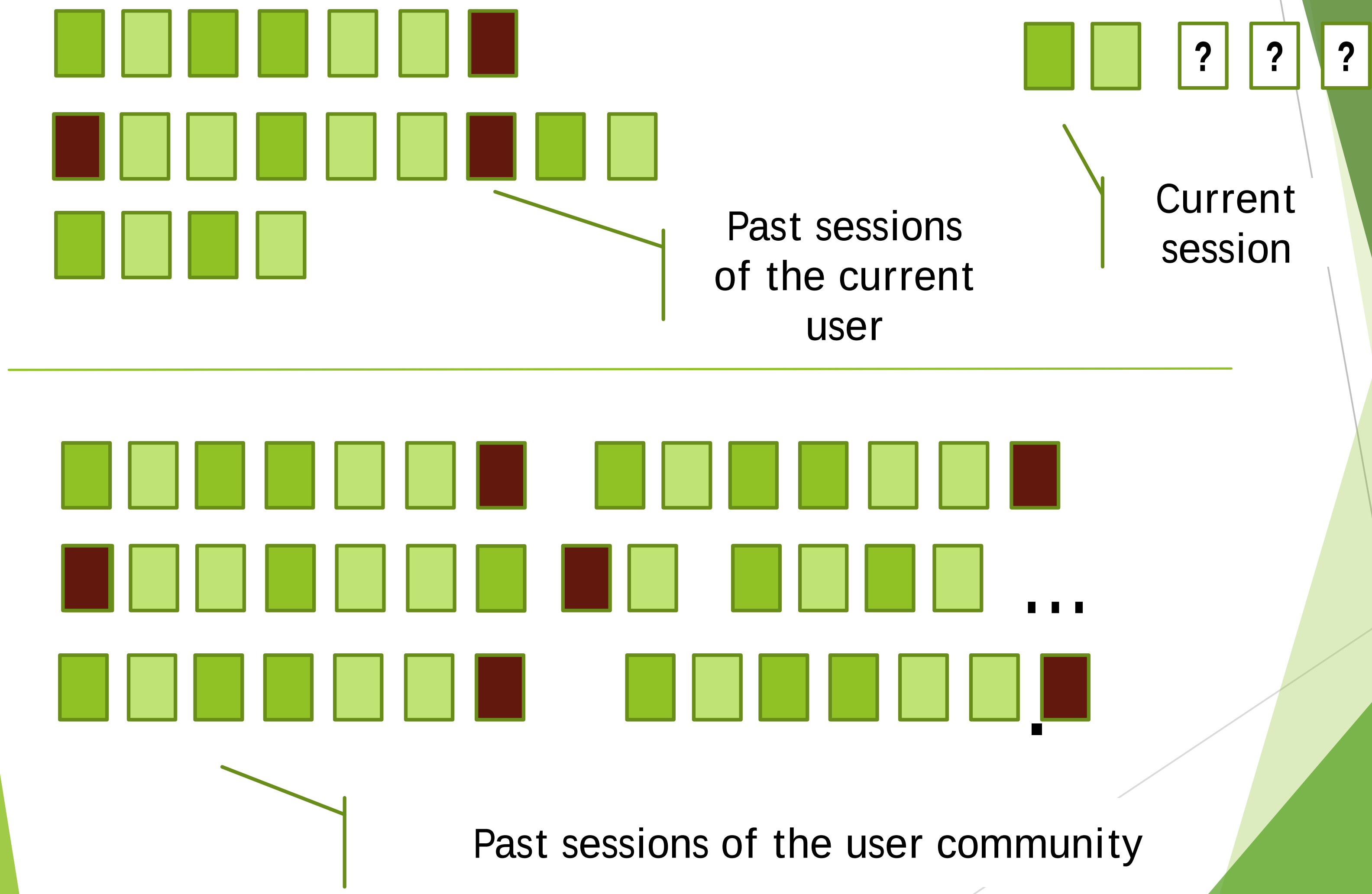


What to recommend?

- Some plausible options
  - Only shoes or only watches?
  - Mostly Nike shoes?
  - Maybe also some T-shirts?

- Using the matrix completion formulation
  - One trains a model based only on past actions
  - The context of the user's current shopping intent is considered only in "context-aware" recommenders
  - Without the context:
    - The algorithm will probably most recommend only T-shirts and trousers
    - Might not be what you expect

General problem abstraction



Long- and short-term models

- What is the relative importance of each model?
- Results of a study using log data from a fashion retailer
  - Trained various baseline models on long-term preferences
  - Applied various re-ranking strategies to adapt to short-term situation, e.g.,
    - Customers who bought ...
    - Prefer items that are similar to the recently viewed ones
    - Prefer items that the user has recently inspected
    - Combinations

Jannach, D., Lerche, L. and Jugovac, M.: *"Adaptation and Evaluation of Recommendations for Short-term Shopping Goals"*. In: Proceedings of the 9th ACM Conference on Recommender Systems (RecSys 2015). Vienna, Austria, 2015, pp. 211-218

Empirical results

- Observations for dense dataset (example)
  - Recall of best baseline method (BPR): 40%
  - Other:
    - Customers who bought … : 49%
    - Just show me what I have seen : 64%
    - Show me similar things : 71%
    - Combining long- and short-term : 73%
- Short-term adaptation is crucial
  - Choice of baseline has an effect
  - Do computationally complex models pay off?
- Reminding is very effective
- Consideration of trends and discounts

Jannach, D., Ludewig, M. and Lerche, L.: *"Session-based Item Recommendation in E-Commerce: On Short-Term Intents, Reminders, Trends, and Discounts"*. User-Modeling and User-Adapted Interaction. Springer, (forthcoming)

Beyond Matrix Completion

- Problem setup and data
  - Post-diction is not prediction
  - Benchmark dataset issues
  - Single-interaction assumption
- Accuracy addiction and other components of utility
- Context matters
- What about the user interface?
- What about long-term effects?

Prediction accuracy addiction

- Prediction and ranking accuracy
  - Predicting the (relative) relevance of unseen items as the main focus of algorithmic works still today
  - However, researchers for many years know that prediction accuracy is often not enough
- Other components of utility
  - Diversity, novelty, and serendipity
    - But how much of it (in a given domain)?
  - Utility for the consumer
    - Recommending the obvious might be accurate but pointless
  - Utility for the provider
    - Business value (see also later)

Jugovac, M., Jannach, D. and Lerche, L.: *"Efficient Optimization of Multiple Recommendation Quality Factors According to Individual User Tendencies"*. Expert Systems With Applications, 2017

Prediction accuracy addiction

- Additional undesired biases can exist
  - Focus mostly on popular items (popularity bias)
    - Leads to high precision and recall, but limited value
    - Can lead to “rich-get-richer-effect
  - Focus on a small set of items (concentration bias)
    - Limited catalog coverage
- Some algorithms lead to similar accuracy values but to largely different recommendations
- Undesired long-term effects

Top 10 lists for the same user

BPR
The Lord of the Rings (2002)
Indiana Jones (Raiders) (1981)
Signs (2002)
Star Wars: Episode I (1999)
Shrek (2001)
Monsters, Inc. (2001)
A Christmas Story (1983)
Rain Man (1988)
Life is beautiful (1997)
Titanic (1997)

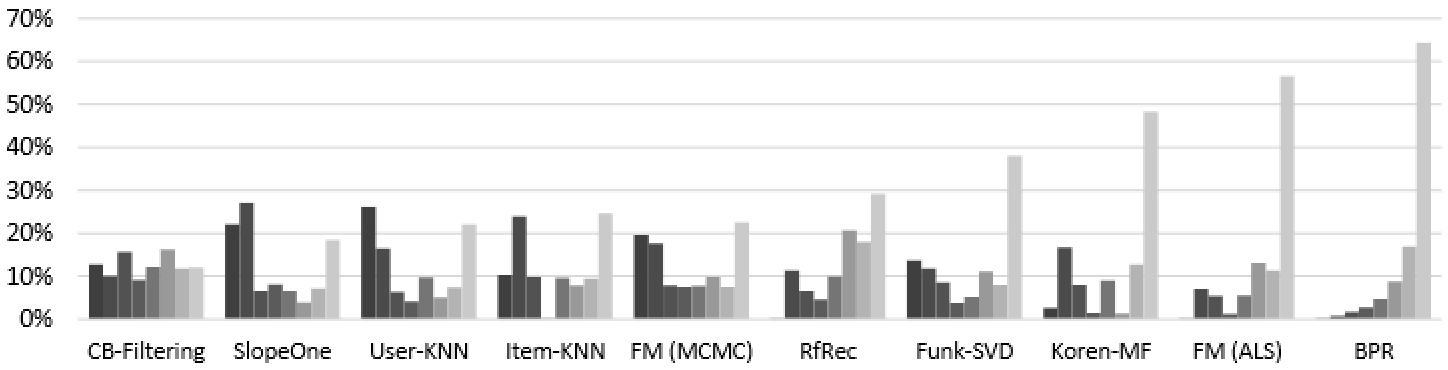
Funk-SVD
Shawshank Redemption (1994)
Schindler's List (1993)
Star Wars (1977)
The Godfather (1972)
Once (2006)
Indiana Jones (Raiders) (1981)
Festen (1998)
The Silence of the Lambs (1991)
The Lives of Others (2006)
The Dark Knight (2008)

Koren-MF
A Clockwork Orange (1971)
The Godfather: Part II (1974)
Leaving Las Vegas (1995)
Annie Hall (1977)
Fargo (1996)
Hoop Dreams (1994)
American Beauty (1999)
Dr. Strangelove (1964)
Memento (2000)
Life of Brian (1979)

Rf-Rec
Shawshank Redemption (1994)
Rear Window (1954)
The Godfather (1972)
The Usual Suspects (1995)
City of God (2002)
Wallace and Gromit (1995)
Shichinin no samurai (1954)
Once Upon a Time in the West (1968)
Schindler's List (1993)
The Third Man (1949)

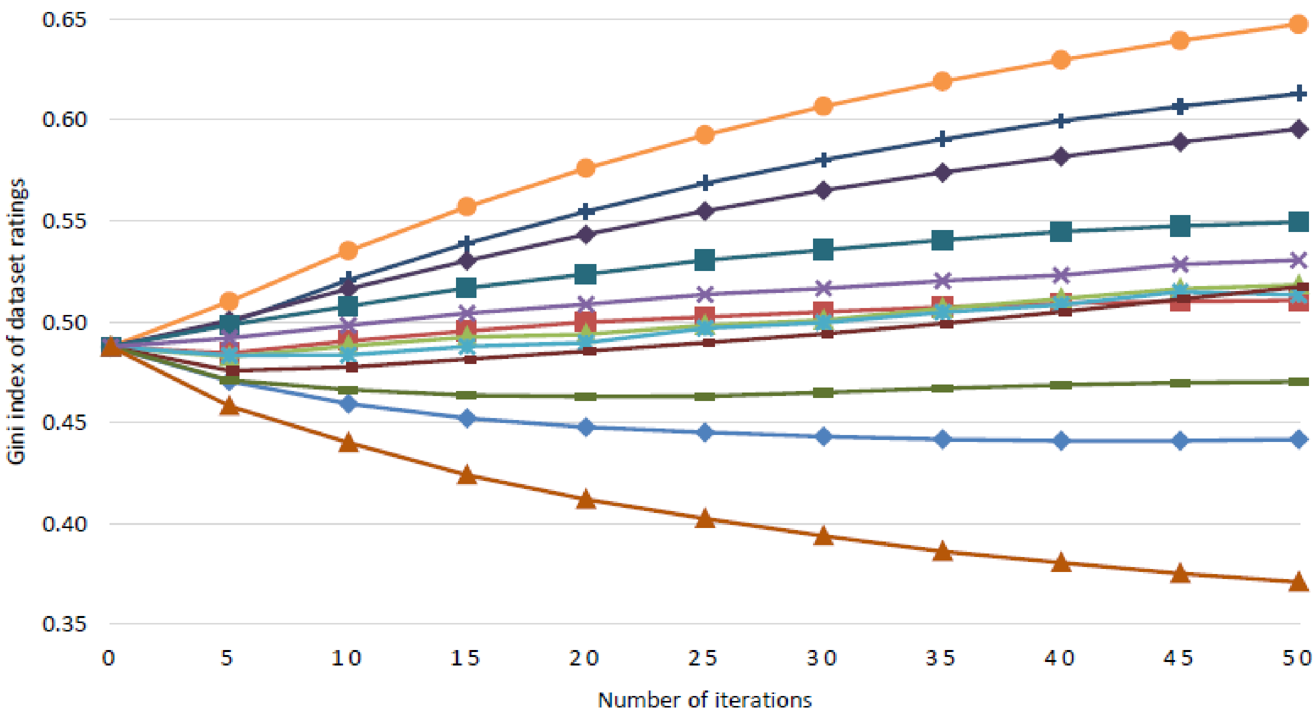
Popularity Biases

- More even distributions indicate that both popular and unpopular items are recommended



- One algorithm’s choices seem to be directly related with the popularity of the items
- Variants of the same algorithm (FM) lead to quite different effects

Rich-get-richer simulation



Jannach, D., Lerche, L., Kamehkhosh, I. and Jugovac, M.: *"What recommenders recommend: an analysis of recommendation biases and possible countermeasures"*. User Modeling and User-Adapted Interaction, Vol. 25(5). Springer Nature, 2015, pp. 427-491.

Beyond Matrix Completion

- Problem setup and data
  - Post-diction is not prediction
  - Benchmark dataset issues
  - Single-interaction assumption
- Accuracy addiction and other components of utility
- Context matters
- What about the user interface?
- What about long-term effects?

Context matters

- Which items are relevant can depend on the context
  - With whom I watch a video, my mood, environmental parameters
- Traditional matrix completion setups do not consider the user context
  - A number of technical approaches developed in recent years
  - Still, a lack of datasets to do research on
- Long-term and short-term interests
  - Users may have a diverse profile, but arrive at the site with a specific shopping intent

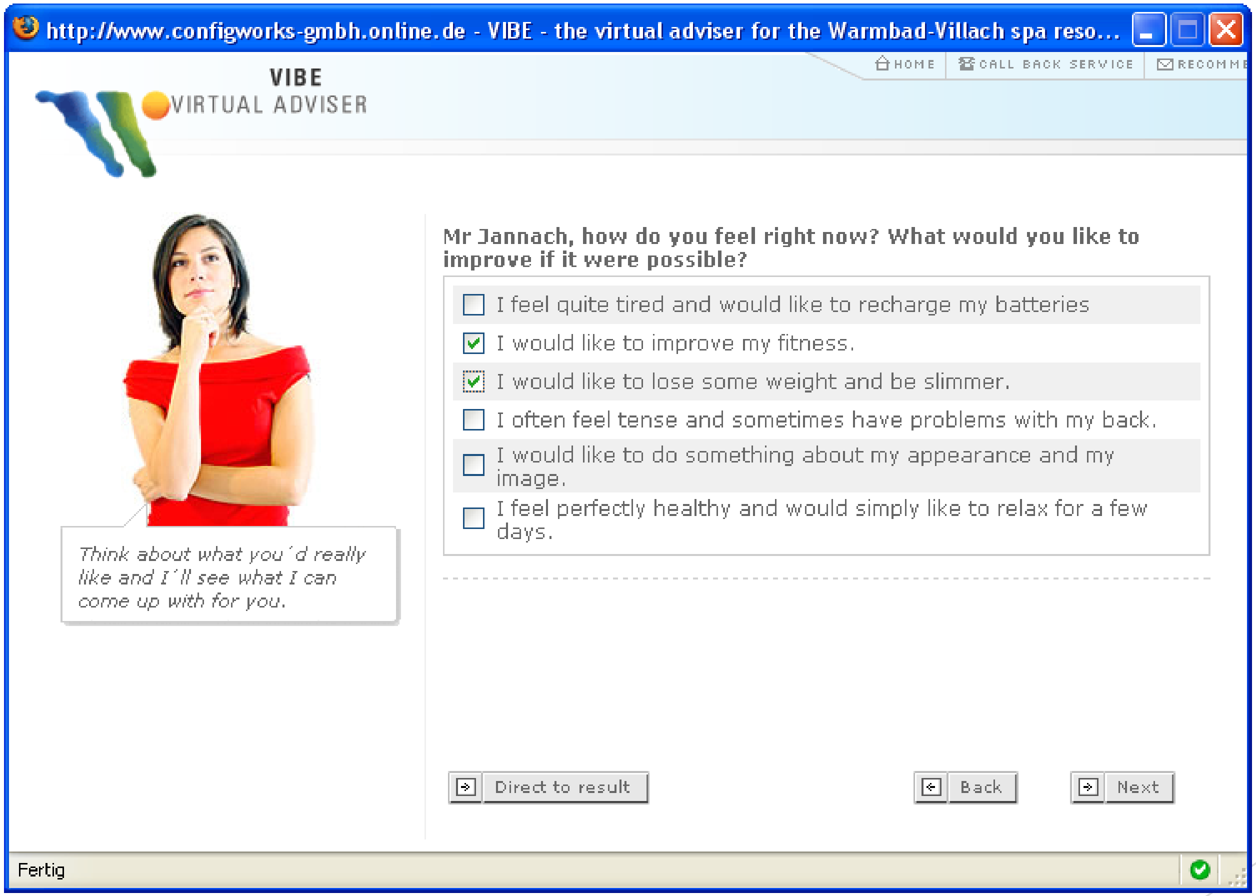
Jannach, D., Lerche, L. and Jugovac, M.: *"Adaptation and Evaluation of Recommendations for Short-term Shopping Goals"*. In: Proceedings RecSys 2015. Vienna, Austria, 2015, pp. 211-218

Interacting with users

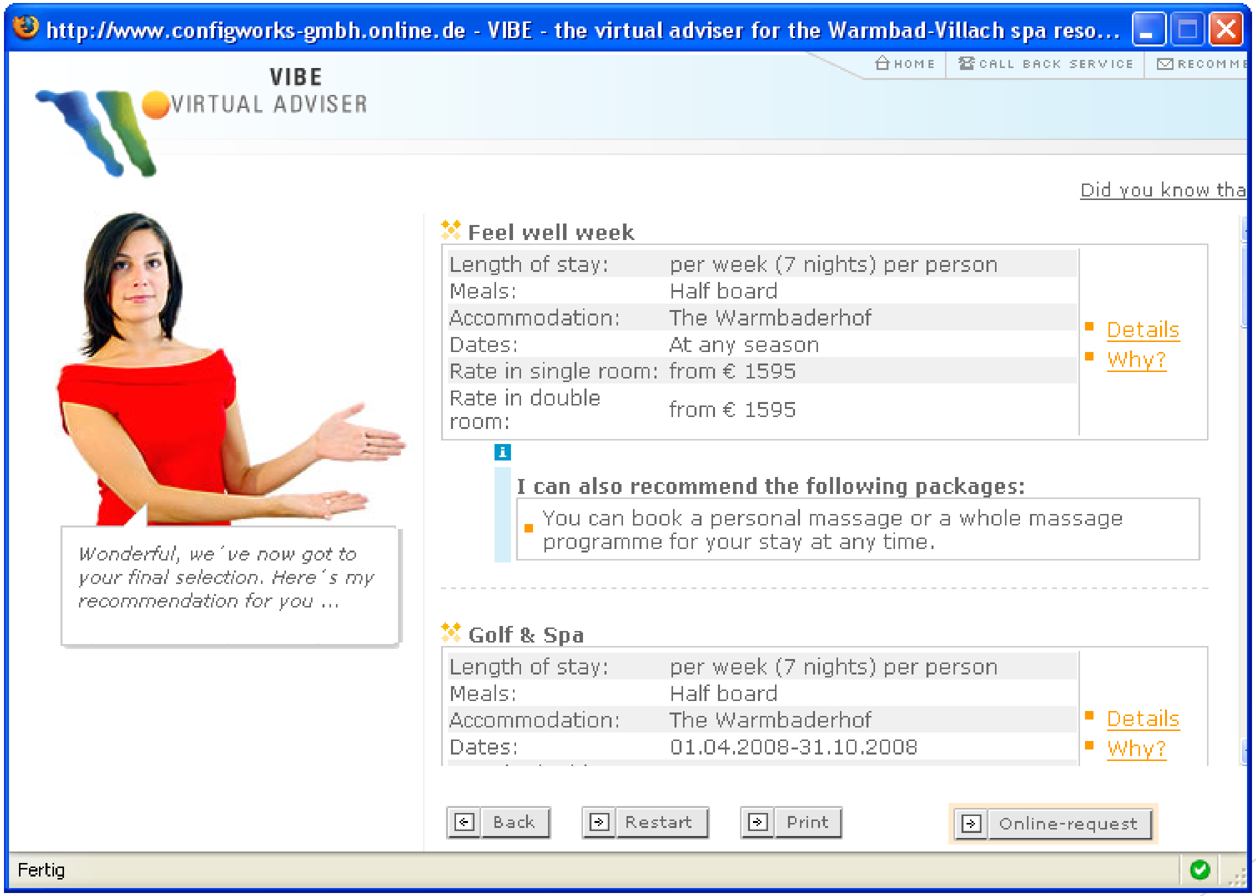
- How the user can interact with the system can significantly impact their effectiveness
- Much less research on UI/UX-related issues than on algorithmic approaches
- Questions, e.g.,:
  - How do we acquire the preferences? How can users correct them?
  - How do we present the results?
  - Are there any ways to convince or persuade a user?
  - How should the system explain its recommendations?
  - When should recommendations be presented?

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

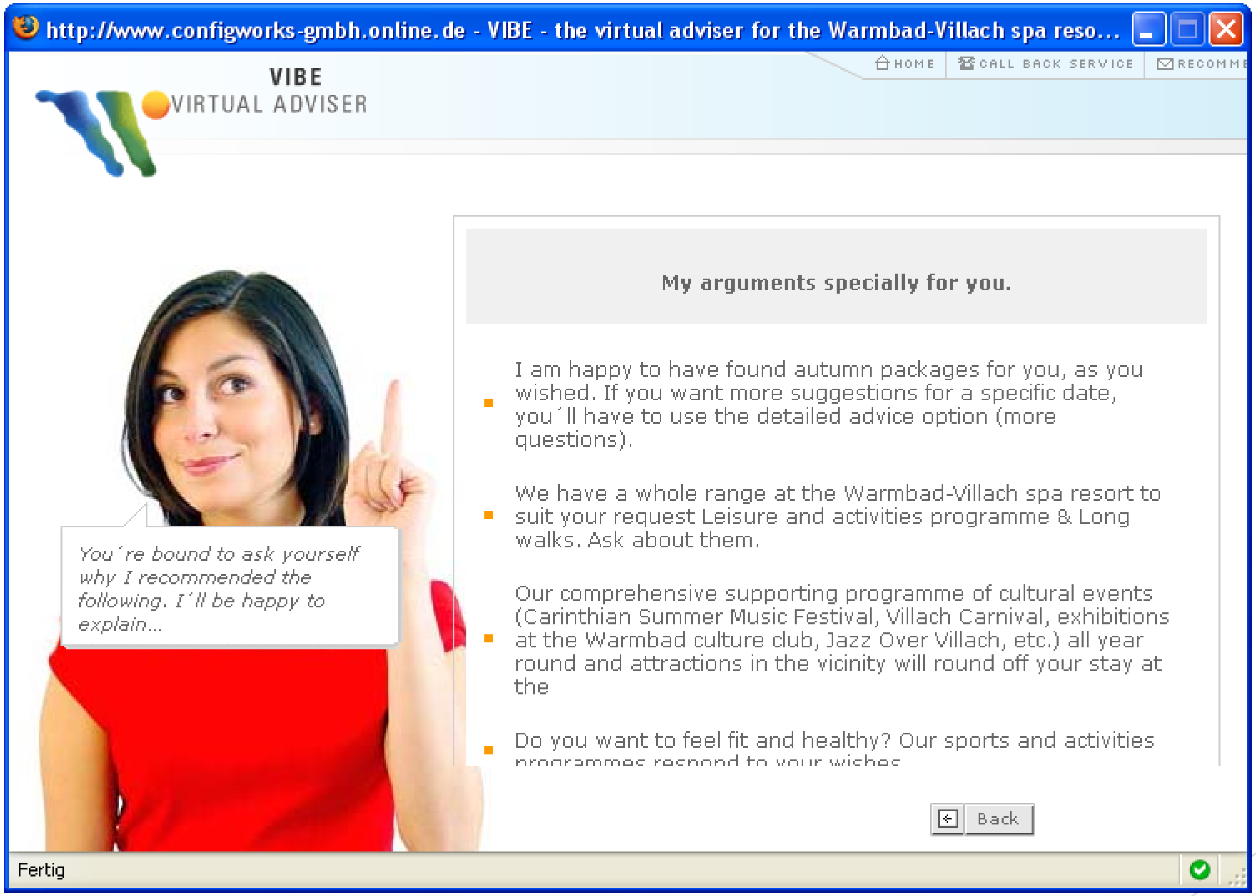
Is this even a recommender?



Is this even recommender?



Is this even a recommender?



User control

- Recommenders are mostly a black box to users
- How do we help users change their profiles and "correct" the system's assumptions?

More research is required

- Preliminary survey among CS students
- Research questions:
  - Do people know about the feedback and control functionality?
  - Do they use it?
  - If not, why not?

- Two-stage study based on questionnaire
  - 75/26 participants
  - 1st stage: “Do you know/use it?”
  - 2nd stage: “Why do you not use it?” (Free-text answers)

Jannach, D., Naveed, S. and Jugovac, M.: *"User Control in Recommender Systems: Overview and Interaction Challenges"*. In: EC-Web 2016. Porto, Portugal, 2016

Outcomes

- 93% say they know there are possibilities to influence recommendations
- 16% are aware of the special page with feedback/control functionality
- 8% have ever used the feedback/control functionality
- Even though
  - 53% said the functionality was clear or very clear, and
  - 24% said it could be guessed

But why not using it?

In the long run

- Trust and loyalty
  - Key targets in the long run from the provider perspective
- Trust needs repeated positive experiences
  - Continuously persuading the user to take a non-optimal decision can be detrimental for the service
  - Balance between provider and consumer benefit must be found
- Explanations maybe a key factor
  - Transparency as an important trust-enabling factor

Nilashi, M., Jannach, D., bin Ibrahim, O., Esfahani, M. D. and Ahmadi, H.: *"Recommendation quality, transparency, and website quality for trust-building in recommendation agents"*. Electronic Commerce Research and Applications, Vol. 19. Elsevier BV, 2016, pp. 70-84

In this talk

- We will briefly review the history of recommender systems
- We will outline challenges when adopting a purely algorithmic-oriented research perspective
- We sketch a purpose-oriented framework for the design and evaluation of recommender systems

A purpose-oriented approach

- A central, but often not-addressed question

  What is a good recommender system?

- Some possible answers
  - One that achieves a low RMSE on historical data?
  - One that produces diverse item lists?
  - One that leads to high click-through-rates?

  ...
- Or:

  One that creates some form of utility

Utility for whom

- 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
  - ...

It all depends

- Recommendations can serve different purposes
  - Whether a recommendation is good nor not depends on the intended purpose and the perspective
  - The purpose can very specific for a domain or application

- In academia:
  - we often abstract from such domain-specifics

In the literature

- Set of abstract, computational tasks
  1. Find (all or some) good items
  2. Predict the relevance of unseen items ("annotate in context")
  3. Recommend sequence
  4. Just browsing

Current research practice

- Operationalization of the research problem
  - Limited set of tasks, mostly relevance prediction
  - Abstract, domain-independent performance measures
- Plus:
  - Standard evaluation schemes
  - Public datasets
- Benefits:
  - Well-defined problem
  - Continuous improvement
  - Comparability & reproducibility

Some dangers

- Do we over-simplify or over-generalize things?
  - High diversity might be good in some domains, but not in others
  - What a "good item" is, depends on the viewpoint and purpose
  - How do we know that our abstract measures reflect either viewpoint?
- Time to re-assess our research practice
  - Re-visit the fundamental goals & tasks of recommenders and how we evaluate such systems
  - Is our approach too narrow – can we cover more than what we are currently do today?

A conceptual framework

- Goals
  - For a structured discussion of goals and purposes
  - To point out areas of future research
  - Consider provider and consumer side
- Structure – 4 layers

Overarching goal of the system, strategic value
Recommendation purpose / Intended utility
System (algorithm) task
Computational metrics

		Consumer's Viewpoint	Provider's Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	"Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, ...	• Change user behavior in desired directions • Create additional demand • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse set of alternatives • Find mix of familiar and relevant unknown items • Find suitable accessories • ...	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .	

		Consumer's Viewpoint	Provider's Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	"Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, …	• Change user behavior in desired directions • Create additional demand • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse set of alternatives • Find mix of familiar and relevant unknown items • Find suitable accessories • …	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .	

		Consumer's Viewpoint	Provider's Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	"Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, …	• Change user behavior in desired directions • Create additional demand • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse set of alternatives • Find mix of familiar and relevant unknown items • Find suitable accessories • …	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .	

		Consumer's Viewpoint	Provider's Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	"Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, …	• Change user behavior in desired directions • Create additional demand • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse set of alternatives • Find mix of familiar and relevant unknown items • Find suitable accessories • …	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .	

		Consumer's Viewpoint	Provider's Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	"Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, …	• Change user behavior in desired directions • Create additional demand • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse set of alternatives • Find mix of familiar and relevant unknown items • Find suitable accessories • …	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .	

		Consumer’s Viewpoint	Provider’s Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	"Organizational Utility": Profit, Revenue, Return on Investment, Growth, **Customer Retention**
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, …	• Change user behavior in desired directions • Create additional demand • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse set of alternatives • Find mix of familiar and relevant unknown items • Find suitable accessories • …	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .	

		Consumer's Viewpoint	Provider's Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	"Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, …	• Change user behavior in desired directions • Create additional demand • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse set of alternatives • Find mix of familiar and relevant unknown items • Find suitable accessories • …	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .	

		Consumer's Viewpoint	Provider's Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	• "Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, ...	• Change user behavior in desired directions • Create additional demand • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse set of alternatives • Find mix of familiar and unknown items • Find suitable accessories • ...	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .	

		Consumer's Viewpoint	Provider's Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	• "Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, ...	• Change user behavior in desired directions • Create additional demand • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse set of alternatives • Find mix of familiar and unknown items • Find suitable accessories • ...	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .?	

		Consumer’s Viewpoint	Provider’s Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	"Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention, Added Customer Value
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, ...	• Change user behavior in desired directions • Help in the decision making process • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse set of alternatives for item of interest • Find mix of familiar and relevant new items • Find suitable accessories • ...	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .	

		Consumer’s Viewpoint	Provider’s Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	"Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention, Added Customer Value
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, …	• Change user behavior in desired directions • Help in the decision making process • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse set of alternatives for item of interest • Find mix of familiar and relevant new items • Find suitable accessories • …	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .	

		Consumer's Viewpoint	Provider's Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	"Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention, Added Customer Value
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, …	• Change user behavior in desired directions • Support customer's decision making process • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse set of alternatives for item of interest • Find mix of familiar and relevant new items • Find suitable accessories • …	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .	

		Consumer's Viewpoint	Provider's Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	"Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention, **Added Customer Value**
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, …	• Change user behavior in desired directions • **Support customer's decision making process** • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • **Create diverse but relevant set of alternatives for item of interest** • Find mix of familiar and relevant new items • Find suitable accessories • …	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .	

		Consumer's Viewpoint	Provider's Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	"Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention, Added Customer Value
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, …	• Change user behavior in desired directions • Support customer's decision making process • Help users discover new artists, directors, genres • Increase activity on the site • . . .
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse but relevant set of alternatives for item of interest • Find mix of familiar and relevant new items • Find suitable accessories • …	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .?	

		Consumer's Viewpoint	Provider's Viewpoint
Strategic Perspective	Overarching Goal	"Personal Utility": Happiness, Satisfaction, Knowledge, Entertainment, Benefit	"Organizational Utility": Profit, Revenue, Return on Investment, Growth, Customer Retention, Added Customer Value
	Recommendation Purpose	• Help users find objects that match the user's long-term preferences • Show alternatives • Help users explore or understand the item space, ...	• Change user behavior in desired directions • Support customer's decision making process • Help users discover new artists, directors, genres • Increase activity on the site • Promote high-margin items
Operational Perspective	System Task	• Annotate in context (i.e., estimate preference of a given item) • Find good items • Create diverse set of alternatives for item of interest with a focus on high-margin items • Find mix of familiar and relevant new items • Find suitable accessories • ...	
	Computational Metric	Predictive accuracy (e.g., RMSE, MAE), classification accuracy (e.g., Precision, Recall, AUC), ranking and top-n accuracy (e.g., rank correlation, MRR, NDCG, etc.), item discoverability (diversity, novelty, or serendipity measures), recommendation biases (e.g., concentration or popularity biases) and blockbuster effects, survey-based user satisfaction scores, business- and domain-specific measures (e.g., conversion rates or click-through-rates), . . .?	

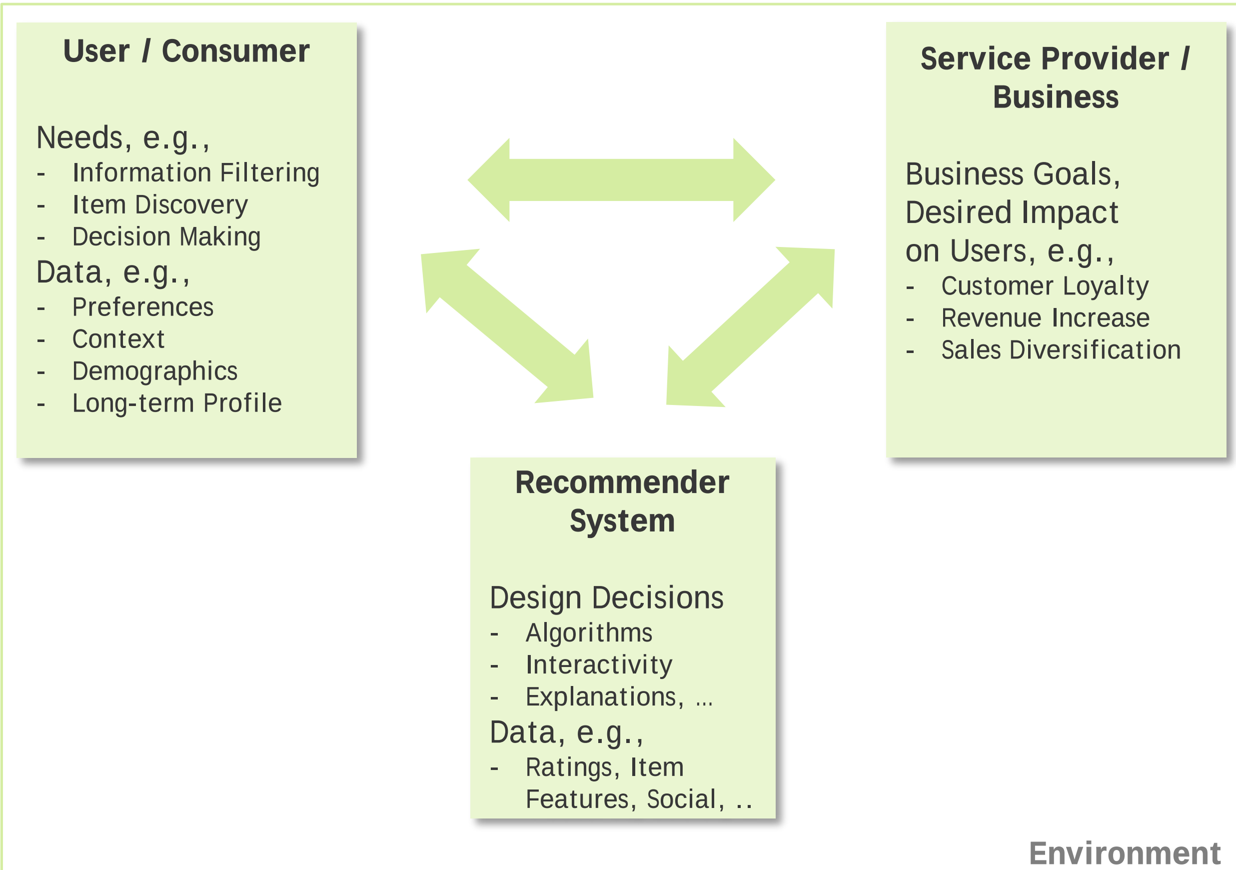
Immediate challenges

- Defining new tasks and metrics
  - Next-item recommendation as a candidate
  - Consider metric-purpose-fit
- Multi-metric evaluations, understanding trade-offs
- Data and protocol issues
- Moving beyond computer science (RECO-nomics)

From Algorithms to Systems

- For a more comprehensive research approach, need to move beyond computer science
  - Often, too much focus on abstract accuracy measures (in machine learning based research)
  - Question of the purpose of the system seldom asked
    - What to measure and where to be good at depends on the purpose
- Research in Information Systems literature
  - Not much visibility in CS literature
  - Accuracy only one of many factors for RS success
- Putting the user back in the loop

A more comprehensive picture



Open issues

- Need to address problem (more often) with an interdisciplinary approach
  - Focus on problems other than algorithms
  - Develop a richer repertoire of research methods
    - Many of them are already out there
- "Standardize" research operationalization of relevant practical problems
  - E.g., next-item recommendation, session-based recommendation, usage of multiple recommendation lists, ….
- Need to better understand real-world implications of research results
  - Do a real-world check, consider specific purposes or our systems, consider the stakeholder's roles
  - Several studies show that the most accurate methods in offline experiments lead to the best user perception or business success

Summary

- Sketched importance of recommenders
- Discussed history of recommender systems
- Outlined challenges of current research practice
  - Recommendation is not (only) a machine learning problem
  - And it is not solved
- Reviewed a conceptual framework to the design and evaluation of recommender systems

- Thank you for your attention

- Contact:
  - dietmar.jannach@tu-dortmund.de

References

1) Jannach, D., Resnick, P., Tuzhilin, A. and Zanker, M.: "**Recommender Systems - Beyond Matrix Completion**". Communications of the ACM, Vol. 59(11). Association for Computing Machinery (ACM), 2016, pp. 94-102

2) Jannach, D. and Adomavicius, G.: "**Recommendations with a Purpose**". In: Proceedings of the 10th ACM Conference on Recommender Systems (RecSys 2016). Boston, Massachusetts, USA, 2016, pp. 7-10

3) Jannach, D., Lerche, L., Kamehkhosh, I. and Jugovac, M.: "**What recommenders recommend: an analysis of recommendation biases and possible countermeasures**". User Modeling and User-Adapted Interaction, Vol. 25(5). Springer Nature, 2015, pp. 427-491.

4) Lerche, L., Jannach, D. and Ludewig, M.: "**On the Value of Reminders within E-Commerce Recommendations**". In: Proceedings UMAP 2016. Halifax, Canada, 2016

5) Jugovac, M., Jannach, D. and Lerche, L.: "**Efficient Optimization of Multiple Recommendation Quality Factors According to Individual User Tendencies**". Expert Systems With Applications, 2017

6) Jannach, D., Lerche, L. and Jugovac, M.: "**Adaptation and Evaluation of Recommendations for Short-term Shopping Goals**". In: Proceedings RecSys 2015. Vienna, Austria, 2015, pp. 211-218

7) Jugovac, M. and Jannach, D.: "**Interacting with Recommenders – Overview and Research Directions**". ACM Transactions on Intelligent Interactive Systems (ACM TiiS). Forthcoming

8) Jannach, D., Naveed, S. and Jugovac, M.: "**User Control in Recommender Systems: Overview and Interaction Challenges**". In: 17th International Conference on Electronic Commerce and Web Technologies (EC-Web 2016). Porto, Portugal, 2016

9) Nilashi, M., Jannach, D., bin Ibrahim, O., Esfahani, M. D. and Ahmadi, H.: "**Recommendation quality, transparency, and website quality for trust-building in recommendation agents**". Electronic Commerce Research and Applications, Vol. 19. Elsevier BV, 2016, pp. 70-84

10) Jannach, D., Ludewig, M. and Lerche, L.: "Session-based Item Recommendation in E-Commerce: On Short-Term Intents, Reminders, Trends, and Discounts". User-Modeling and User-Adapted Interaction. Springer, (forthcoming)