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

ALTERNATIVE PRODUCTS
Beko Washing Machine
Code: WMB81431LW
£269.99
Zanussi Washing Machine
Code: ZWH6130P
£269.99
Blomberg Washing Machine
Code: WNF6221
£299.99

You may also like
Jack & Jones
JAMIE - Polo shirt - orange
£21.00
Free delivery & returns

You may also like

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.

Customers Who Bought This Item Also Bought

- Star Wars Trilogy Episodes I-III (Blu-ray + DVD)
  Hayden Christansen
  ★★★★★ 2,042
  Blu-ray $34.96 Prime

- Star Wars: The Force Awakens (Blu-ray/DVD/Digital HD)
  Harrison Ford
  ★★★★★ 10,002
  Blu-ray $24.41 Prime

- Star Wars: Episode I - The Phantom Menace (Widescreen Edition)
  Ewan McGregor
  ★★★★★ 3,633
  DVD $53.24 Prime

- Harry Potter: Complete 8-Film Collection [Blu-ray]
  Daniel Radcliffe
  ★★★★★ 6,945
  Blu-ray $65.00 Prime

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?

![Image from Youtube.com](Source: Youtube.com)
Design Space Examples

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

http://www.configworks.gmbh.online.de - VIBE - the virtual adviser for the Wörthbad Villach spa reso...

Mr Jannach, how do you feel right now? What would you like to improve if it were possible?

- I feel quite tired and would like to recharge my batteries
- I would like to improve my fitness.
- I would like to lose some weight and be slimmer.
- I often feel tense and sometimes have problems with my back.
- I would like to do something about my appearance and my image.
- I feel perfectly healthy and would simply like to relax for a few days.

Think about what you’d really like and I’ll see what I can come up with for you.
Design Space Examples

- It could even be a dialog
Design Space Examples

- It could even be a dialog

I am happy to have found autumn packages for you, as you wished. If you want more suggestions for a specific date, you’ll have to use the detailed advice option (more questions).

We have a whole range at the Wambad-Villach spa resort to suit your request: Leisure and activities programme & Long walks. Ask about them.

Our comprehensive supporting programme of cultural events (Carinthian Summer Music Festival, Villach Carnival, exhibitions at the Wambad culture club, Jazz Over Villach, etc.) all year round and attractions in the vicinity will round off your stay at the

Do you want to feel fit and healthy? Our sports and activities programmes respond to your wishes.
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
  - ...

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 component

Recommendation list

<table>
<thead>
<tr>
<th>item</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>i1</td>
<td>0.9</td>
</tr>
<tr>
<td>i2</td>
<td>1</td>
</tr>
<tr>
<td>i3</td>
<td>0.3</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Recommendation Paradigms

Recommendations are often personalized

User profile & contextual parameters

Recommendation component

Recommendation list

Interacting with Recommender Systems — Background — Dietmar Jannach, Ingrid Nunes, Michael Jugovac
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

Knowledge-based:
"Tell me what fits based on my needs"
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-Neigbors (kNN)

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

<table>
<thead>
<tr>
<th></th>
<th>Item1</th>
<th>Item2</th>
<th>Item3</th>
<th>Item4</th>
<th>Item5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>?</td>
</tr>
<tr>
<td>User1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>User2</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>User3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>User4</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
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?

<table>
<thead>
<tr>
<th></th>
<th>Item1</th>
<th>Item2</th>
<th>Item3</th>
<th>Item4</th>
<th>Item5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>?</td>
</tr>
<tr>
<td>User1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>User2</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>User3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>User4</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Similarity Computation (kNN)

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

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

<table>
<thead>
<tr>
<th></th>
<th>Item1</th>
<th>Item2</th>
<th>Item3</th>
<th>Item4</th>
<th>Item5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>?</td>
</tr>
<tr>
<td>User1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>User2</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>User3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>User4</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Rating Prediction (kNN)

- A common prediction function

\[ \text{pred}(a, p) = \bar{r}_a + \frac{\sum_{b \in N} \text{sim}(a, b) \times (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} \text{sim}(a, b)} \]

- 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

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

\[ M_k = U_k \times \Sigma_k \times V_k^T \]

<table>
<thead>
<tr>
<th>( U_k )</th>
<th>Dim1</th>
<th>Dim2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>0.47</td>
<td>-0.30</td>
</tr>
<tr>
<td>Bob</td>
<td>-0.44</td>
<td>0.23</td>
</tr>
<tr>
<td>Mary</td>
<td>0.70</td>
<td>-0.06</td>
</tr>
<tr>
<td>Sue</td>
<td>0.31</td>
<td>0.93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( V_k^T )</th>
<th>Terminator</th>
<th>Die Hard</th>
<th>Twins</th>
<th>Eat, Pray, Love</th>
<th>Pretty Woman</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dim1</td>
<td>-0.44</td>
<td>-0.57</td>
<td>0.06</td>
<td>0.38</td>
<td>0.57</td>
</tr>
<tr>
<td>Dim2</td>
<td>0.58</td>
<td>-0.66</td>
<td>0.26</td>
<td>0.18</td>
<td>-0.36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \Sigma_k )</th>
<th>Dim1</th>
<th>Dim2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dim1</td>
<td>5.63</td>
<td>0</td>
</tr>
<tr>
<td>Dim2</td>
<td>0</td>
<td>3.23</td>
</tr>
</tbody>
</table>

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

\[ = 3 + 0.84 = 3.84 \]
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

<table>
<thead>
<tr>
<th>Title</th>
<th>Genre</th>
<th>Author</th>
<th>Type</th>
<th>Price</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Lace Reader</td>
<td>Fiction, Mystery</td>
<td>Brunonia Barry</td>
<td>Hardcover</td>
<td>49.90</td>
<td>American contemporary fiction, detective, historical</td>
</tr>
<tr>
<td>Into the Fire</td>
<td>Romance, Suspense</td>
<td>Suzanne Brockmann</td>
<td>Hardcover</td>
<td>45.90</td>
<td>American fiction, murder, neo-Nazism</td>
</tr>
</tbody>
</table>
Recommendation Approach

- Represent items and users in the same way

<table>
<thead>
<tr>
<th>Title</th>
<th>Genre</th>
<th>Author</th>
<th>Type</th>
<th>Price</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Lace Reader</td>
<td>Fiction, Mystery</td>
<td>Brunonia Barry</td>
<td>Hardcover</td>
<td>49.90</td>
<td>American contemporary fiction, detective, historical</td>
</tr>
<tr>
<td>Into the Fire</td>
<td>Romance, Suspense</td>
<td>Suzanne Brockmann</td>
<td>Hardcover</td>
<td>45.90</td>
<td>American fiction, murder, neo-Nazism</td>
</tr>
</tbody>
</table>

- 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
    \[
    2 \times \frac{|keywords(b_i) \cap keywords(b_j)|}{|keywords(b_i)| + |keywords(b_j)|}
    \]
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

I am happy to have found autumn packages for you, as you wished. If you want more suggestions for a specific date, you’ll have to use the detailed advice option (more questions).

- We have a whole range at the Warmbad-Villach spa resort to suit your request Leisure and activities programme & Long walks. Ask about them.

- Our comprehensive supporting programme of cultural events (Carinthian Summer Music Festival, Villach Carnival, exhibitions at the Warmbad culture club, Jazz Over Villach, etc.) all year round and attractions in the vicinity will round off your stay at the

- Do you want to feel fit and healthy? Our sports and activities programmes respond to your wishes.
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

Part I (now)

Preference Elicitation
- Ratings & Likings
- Personality Quizzes

Result Presentation and Feedback
- Result List Design
- Proactivity
- Persuasion
- Explanations

Part II

User

Recommender System

Interacting with Recommender Systems — Part I — Dietmar Jannach, Ingrid Nunes, Michael Jugovac
Interacting with Recommender Systems

Part I: Interactions during Preference Elicitation
Exercise

What methods do you prefer to input your preferences?

1. ★★★★★
2. ★★★★★★
3. Like
4. - +
5. ✔
6. Select
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):

Feedback Scale

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

![Feedback Scale Image](source: [1])

Feedback Scale – Considerations

- Number of options
  - ★★★★★ ★★★★★★☆ vs. 👍 🎥 vs. 👍 Like

- 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?

808,103 views

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]

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

![Preference Forms Diagram]

Source: [netflix.com](http://netflix.com)
Source: [news.google.com](http://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
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]

Compound Critiques

Critiquing — Open Questions

- On mobile?
- Weighting criteria differently?
- Integrate preferences into long-term user models?
- Natural language?
- How to evaluate offline?

Alternative Elicitation Techniques

Side-By-Side Comparison, Personality Quizzes, Picture-Based, ...
Alternative Elicitation Techniques: Side-By-Side Comparison of Sets

Source: [1]

Alternative Elicitation Techniques: Picture-Based

Alternative Elicitation Techniques: Tag-Based

Source: movielens.com
Alternative Elicitation Techniques: Personality Quizzes

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

Customers Who Bought This Item Also Bought

<table>
<thead>
<tr>
<th>List label</th>
<th>Item description</th>
<th>Community rating</th>
<th>Highlighting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nikon AF-S FX NIKKOR 50mm f/1.8G Lens with Auto Focus for Nikon DSLR Cameras</td>
<td>★★★★★ 1.505</td>
<td>$216.95 ✨Prime</td>
</tr>
<tr>
<td></td>
<td>Lowepro Adventura 140 Camera Shoulder Bag for DSLR or Camcorder</td>
<td>★★★★☆ 178</td>
<td>$26.99 ✨Prime</td>
</tr>
<tr>
<td></td>
<td>Lexar Professional 633x 64GB SDXC UHS-I Card w/Image Rescue 5 Software...</td>
<td>★★★★☆ 592</td>
<td>$22.50 ✨Prime</td>
</tr>
</tbody>
</table>

Source: amazon.com

Number of options
List Design — Additional Considerations: Placement

Source: amazon.com
List Design — Additional Considerations: Placement

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

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 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]

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


- Familiar user experience with RS as added benefit

- No filtering -> higher user confidence?
Diagrams and Graphs

- **Use diagrams** to communicate item relations

![Diagram](image)

- **Easier for users or additional effort? Fun?**

---

Diagrams and Graphs

- **Use graphs** to communicate item relations

  ![Diagrams](source)

  Source: [1]

- Easier for users or additional effort? Fun?

---

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]

Recommendations in 2D space

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

Source: [1]

Recommendations in 3D space

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

Source: [1]

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

  **Recommended for you**

  **Guardians of the Galaxy [Blu-ray]**
  *Blu-ray* ~ Chris Pratt (8 Jan 2015)
  In stock
  **Price:** EUR 9,99
  73 used & new from EUR 8,75

  **Rate this item**
  ✖️ ★★★★★

  □ I own it
  □ Not interested

  ![Add to Cart] ![Add to Wish List]

  **Because you purchased...**

  **Mad Max: Fury Road [Blu-ray]** (Blu-ray)
  DVD ~ Charlize Theron

  ![Rate this item]

  □ Don't use for recommendations

  ![Add to Cart] ![Add to Wish List]

  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:

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

Source: [1]

User Feedback — Influencing the Strategy

- Let the user take full control

- Can the user understand these options?

Source: [1]

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
Decisive Input Values

- If F1 is *very medium* AND F2 is *high* then likely class 1.

Preference Match

Content-based image retrieval at the end of the early years


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

**Explanation Content**

**Decision Inference Process (1/2)**

- **Inference Trace**

  ![Inference Trace Diagram]

- **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.


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.


Knowledge about Similar Alternatives

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

1. You clicked on DVD **Ocean’s Eleven**

   That DVD has 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**. Inferred energy: 0.140, group: age 13, length: 2.
- **Ocean’s Eleven** and DVD **Ocean’s Twelve** are performed by the same Person **(Brad Pitt)**. Inferred energy: 0.140, group: age 13, length: 2.

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

Knowledge about Peers

---


Knowledge about the Community

Relationship between Knowledge Objects

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:

- **adventure** ★★★★★
- **fun** ★★★★★
- **action** ★★★★★
- **sequel** ★★★★★
- **egypt** ★★★☆☆
- **comedy** ★★★★★
- **brendan fraser** ★★★★★

- **Pros and Cons**

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.

Explanation Content

Statistics

- Decisive Input Values: 29
- Preference Match: 7
- Feature Importance Analysis: 2
- Suitability Estimate: 2
- Inference Trace: 104
- Inference and Domain Knowledge: 30
- Inference Method Side-outcomes: 7
- Self-reflective Statistics: 4
- Knowledge about Peers: 11
- Knowledge about Similar Alternatives: 7
- Relationship between Knowledge Objects: 6
- Background Data: 3
- Knowledge about the Community: 2
- Decisive Features: 25
- Pros and Cons: 17
- Feature-based Domination: 1
- Irrelevant Features: 1
Explanation Content
Additional Observations

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

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Price</th>
<th>Processor speed</th>
<th>Battery life</th>
<th>Installed memory</th>
<th>Hard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$1,499.99</td>
<td>1.5 GHz</td>
<td>5 hours(s)</td>
<td>8 GB</td>
<td>80 GB</td>
</tr>
<tr>
<td></td>
<td>$1,729.99</td>
<td>1.5 GHz</td>
<td>4.5 hour(s)</td>
<td>8 GB</td>
<td>80 GB</td>
</tr>
<tr>
<td></td>
<td>$1,929.99</td>
<td>1.2 GHz</td>
<td>4 hours(s)</td>
<td>8 GB</td>
<td>80 GB</td>
</tr>
<tr>
<td></td>
<td>$1,799.99</td>
<td>1.0 GHz</td>
<td>5.5 hour(s)</td>
<td>8 GB</td>
<td>80 GB</td>
</tr>
</tbody>
</table>

We also recommend the following products because they are cheaper and lighter, but have lower processor speed:

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Price</th>
<th>Processor speed</th>
<th>Battery life</th>
<th>Installed memory</th>
<th>Hard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$1,222.49</td>
<td>1.0 GHz</td>
<td>5 hours(s)</td>
<td>1 GB</td>
<td>100 GB</td>
</tr>
<tr>
<td></td>
<td>$1,149.99</td>
<td>2.0 GHz</td>
<td>4 hours(s)</td>
<td>1 GB</td>
<td>100 GB</td>
</tr>
<tr>
<td></td>
<td>$1,377.00</td>
<td>2.8 GHz</td>
<td>2 hours(s)</td>
<td>8 GB</td>
<td>100 GB</td>
</tr>
<tr>
<td></td>
<td>$1,233.00</td>
<td>1.6 GHz</td>
<td>2.5 hour(s)</td>
<td>1 GB</td>
<td>100 GB</td>
</tr>
<tr>
<td></td>
<td>$1,233.00</td>
<td>1.7 GHz</td>
<td>4.5 hour(s)</td>
<td>8 GB</td>
<td>100 GB</td>
</tr>
<tr>
<td></td>
<td>$1,233.00</td>
<td>1.8 GHz</td>
<td>3.67 hour(s)</td>
<td>8 GB</td>
<td>100 GB</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Price</th>
<th>Processor speed</th>
<th>Battery life</th>
<th>Installed memory</th>
<th>Hard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$1,222.49</td>
<td>1.0 GHz</td>
<td>5 hours(s)</td>
<td>1 GB</td>
<td>100 GB</td>
</tr>
<tr>
<td></td>
<td>$1,149.99</td>
<td>2.0 GHz</td>
<td>4 hours(s)</td>
<td>1 GB</td>
<td>100 GB</td>
</tr>
<tr>
<td></td>
<td>$1,377.00</td>
<td>2.8 GHz</td>
<td>2 hours(s)</td>
<td>8 GB</td>
<td>100 GB</td>
</tr>
<tr>
<td></td>
<td>$1,233.00</td>
<td>1.6 GHz</td>
<td>2.5 hour(s)</td>
<td>1 GB</td>
<td>100 GB</td>
</tr>
<tr>
<td></td>
<td>$1,233.00</td>
<td>1.7 GHz</td>
<td>4.5 hour(s)</td>
<td>8 GB</td>
<td>100 GB</td>
</tr>
<tr>
<td></td>
<td>$1,233.00</td>
<td>1.8 GHz</td>
<td>3.67 hour(s)</td>
<td>8 GB</td>
<td>100 GB</td>
</tr>
</tbody>
</table>

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

- Natural-Language: 109
- Visualisation (Graph or Tree): 19
- Visualisation (Other): 17
- List of ...: 14
- Arguments: 7
- Log or Traces: 5
- Other: 4
- Not specified: 10

- 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

<table>
<thead>
<tr>
<th>Postulate/Argument structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postulate 1: (&lt; \text{recommend} (\text{Movie, User}), \text{good_movie} (\text{Movie}) \rightarrow \text{avg_rating} (\text{Movie}) \geq 3.8 ) &gt;</td>
</tr>
<tr>
<td>Postulate 2: (&lt; \text{recommend} (\text{Movie, User}), \text{likes_by_top_genre} (\text{Movie, User}) \rightarrow \text{top_genre} (\text{User, Genre}), \text{genre} (\text{Movie, Genre}) ) &gt;</td>
</tr>
<tr>
<td>Postulate 3: (&lt; \text{recommend} (\text{Movie, User}), \text{likes_by_top_actor} (\text{Movie, User}) \rightarrow \text{top_actor} (\text{User, Actor}), \text{leads_in} (\text{Movie, Actor}) ) &gt;</td>
</tr>
<tr>
<td>Postulate 4: (&lt; \text{recommend} (\text{Movie, User}), \text{likes_by_top_actor} (\text{Movie, User}) \rightarrow \text{top_actor} (\text{User, Actor}), \text{leads_in} (\text{Movie, Actor}), \text{top_genre} (\text{User, Genre}), \text{genre} (\text{Movie, Genre}) ) &gt;</td>
</tr>
</tbody>
</table>
# Explanation Generation

## Key Drivers: (Intended) Purpose

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

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

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>1980</th>
<th>1990</th>
<th>2000</th>
<th>2010</th>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rule-based</td>
<td>28</td>
<td>33</td>
<td>11</td>
<td>7</td>
<td>79</td>
<td>55.63%</td>
</tr>
<tr>
<td></td>
<td>Logic-based</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>16</td>
<td>11.27%</td>
</tr>
<tr>
<td></td>
<td>Multi-Criteria Decision Making</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>6.34%</td>
</tr>
<tr>
<td></td>
<td>Constraint-based</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1.41%</td>
</tr>
<tr>
<td></td>
<td>Case-based Reasoning</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>2.88%</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>3</td>
<td>11</td>
<td>7</td>
<td>9</td>
<td>30</td>
<td>21.13%</td>
</tr>
<tr>
<td>Machine Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feature-based</td>
<td>2</td>
<td>13</td>
<td>6</td>
<td>11</td>
<td>32</td>
<td>62.75%</td>
</tr>
<tr>
<td></td>
<td>Collaborative-filtering</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>19.61%</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>9</td>
<td>17.65%</td>
</tr>
<tr>
<td>Mathematical Model</td>
<td></td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1.0%</td>
</tr>
<tr>
<td>Human-made Decision</td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1.0%</td>
</tr>
<tr>
<td>Algorithm-independent</td>
<td></td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>11</td>
<td>5.3%</td>
</tr>
</tbody>
</table>
Evaluation: Design & Conclusions

Part III: Explanations
Evaluation Design
Presence of an Evaluation

- How many of the explanation generation approaches were evaluated?

<table>
<thead>
<tr>
<th></th>
<th>1980</th>
<th>1990</th>
<th>2000</th>
<th>2010</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technique</td>
<td>0%</td>
<td>22%</td>
<td>50%</td>
<td>46%</td>
<td>35%</td>
</tr>
<tr>
<td>Tool</td>
<td>0%</td>
<td>8%</td>
<td>6%</td>
<td>7%</td>
<td>6%</td>
</tr>
<tr>
<td>Total</td>
<td>0%</td>
<td>13%</td>
<td>30%</td>
<td>36%</td>
<td>21%</td>
</tr>
</tbody>
</table>

- 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

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?
  - How good do you think the explanation was?

*According to Tintarev and Masthof’s definition, this should be effectiveness. However, the authors used the term efficiency in the paper: “W. Hanshui 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

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Positive</th>
<th>Neutral</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effectiveness</td>
<td>+++++++</td>
<td>~~~~~</td>
<td></td>
</tr>
<tr>
<td>Transparency</td>
<td>+++++++</td>
<td>~</td>
<td></td>
</tr>
<tr>
<td>Persuasiveness</td>
<td>++</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>+++</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trust</td>
<td>+++++</td>
<td>~~</td>
<td></td>
</tr>
<tr>
<td>Usefulness</td>
<td>+++</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ease of Use</td>
<td>+</td>
<td>~~</td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>+</td>
<td>~~</td>
<td></td>
</tr>
</tbody>
</table>
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
  - 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

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

- Display
  - Adaptive
  - Automated
  - Manual

- Generality
  - Multi-level
  - Single-level

- Level of Detail
  - Multi-level
  - Single-level

- Responsiveness
  - Context-sensitive
  - Fixed
  - User Expertise

- Target
  - Decision
  - Alternative Group
  - Alternative

- Decision Inference Method
  - Algorithm-independent
  - Knowledge-based
  - Human-made
  - Machine Learning
  - Mathematical Model

- Domain
  - Domain-neutral
  - Domain-specific
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: