RecoXplainer: An Extensible Toolkit for Explainable Recommender Systems

Ludovik **Coba**, Roberto **Confalonieri**, Markus **Zanker**

MQ2 Tutorial at AAAI-21

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

Resources

- Web page and slides
  http://www.inf.unibz.it/~rconfalonieri/aaai21/
- Repository
  https://github.com/ludovikcoba/recoxplainer

Outline

- Part I: An introduction to Explainable Recommender Systems
- Part II: RecoXplainer

Outline

- An introduction to Explainable Recommender Systems
  - Explainability/Explainable AI
  - Recommender Systems
  - Explanations in Recommender Systems

Outline

- RecoXplainer
  - Overview of the Toolkit
  - Model-based Explanations
  - Post-hoc Explanations
  - Evaluation of Explanations
  - Hands-on Session

Part I
An Introduction to Explainable Recommender Systems

Why Explainability?

AI is now used in many high-stakes decision making applications (credit, employment, admission, sentencing).


Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

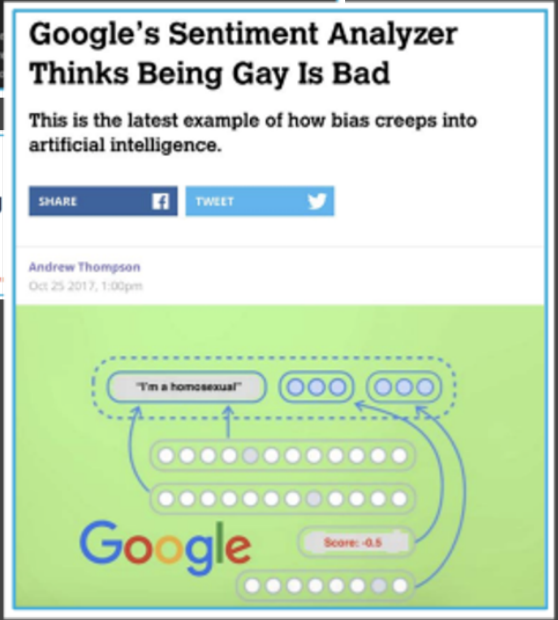
ON A SPRING AFTERNOON IN 2014, Brisha Borden was running late to pick up her god-sister from school when she spotted an unlocked kid's blue Huffy bicycle and a silver Razor scooter. Borden and a friend grabbed the bike and scooter and tried to ride them down the street in the Fort Lauderdale suburb of Coral Springs.

Just as the 18-year-old girls were realizing they were too big for the tiny conveyances — which belonged to a 6-year-old boy — a woman came running after them saying, "That's my kid's stuff." Borden and her friend immediately dropped the bike and scooter and walked away.


BUSINESS NEWS OCTOBER 9, 2018 / 11:12 PM / 12 DAYS AGO

Amazon scraps secret AI recruiting showed bias against women

Jeffrey Dastin


Google's Sentiment Analyzer Thinks Being Gay Is Bad

This is the latest example of how bias creeps into artificial intelligence.

SHARE TWEET

Andrew Thompson
Oct 25 2017, 1:00pm

Most current methods lack "explainability"

89% of consumers say…
“technology companies need to be more transparent”

Technology companies need to comply with GDPR - “Right to Explanation”

Explainability across industries

ADVERTISING

Reputational risk of placing adverts alongside content that doesn’t ‘fit’ the brand

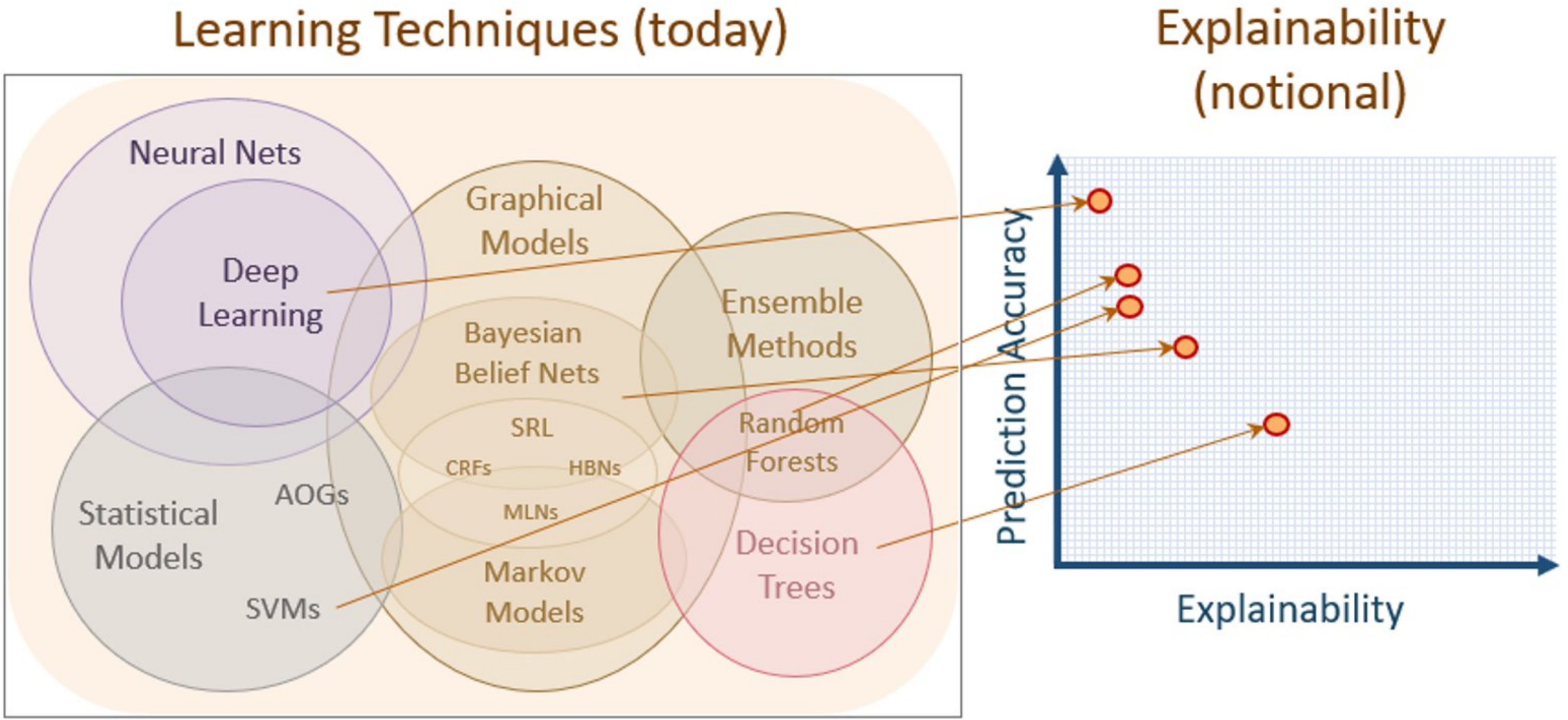
HEALTH

49% of physicians in the US are anxious or uncomfortable with AI

FINANCE

Explaining why automated decision-making rejects loan applications

The most effective algorithms are the hardest to explain

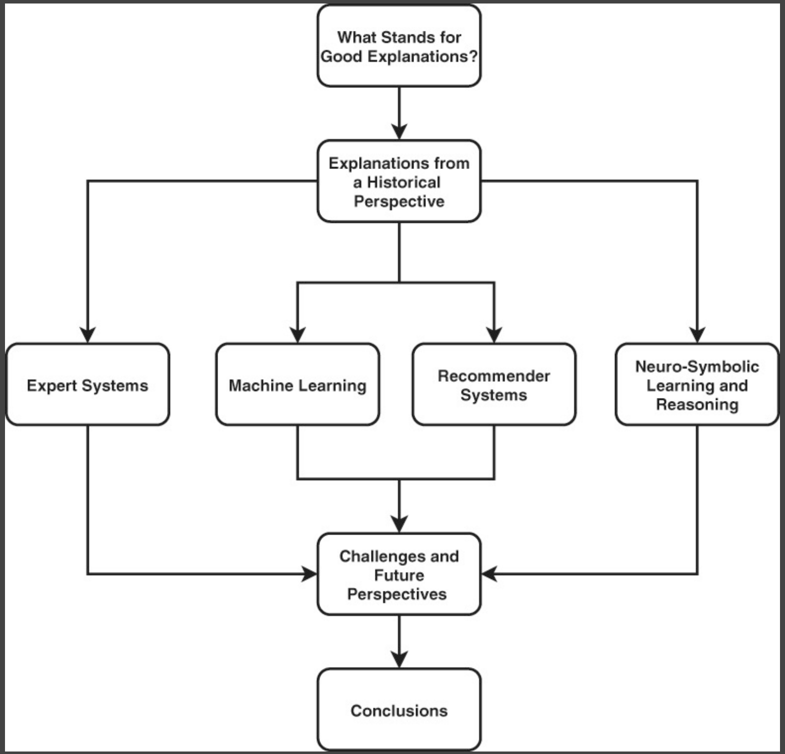


DARPA. Explainable AI - program, 2016. https://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf

Why Explainability is a challenge?

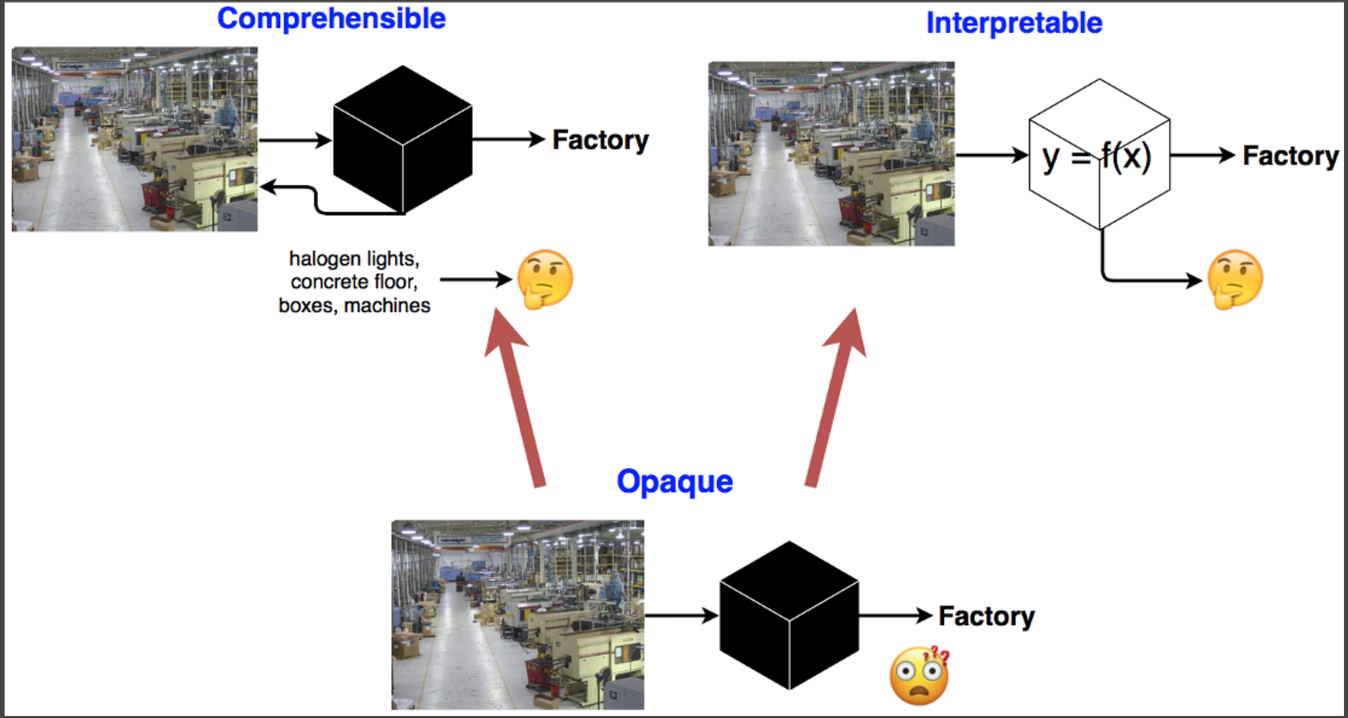
Explainability

- Different **notions**
- Different **requirements**
- **Plethora** of approaches!



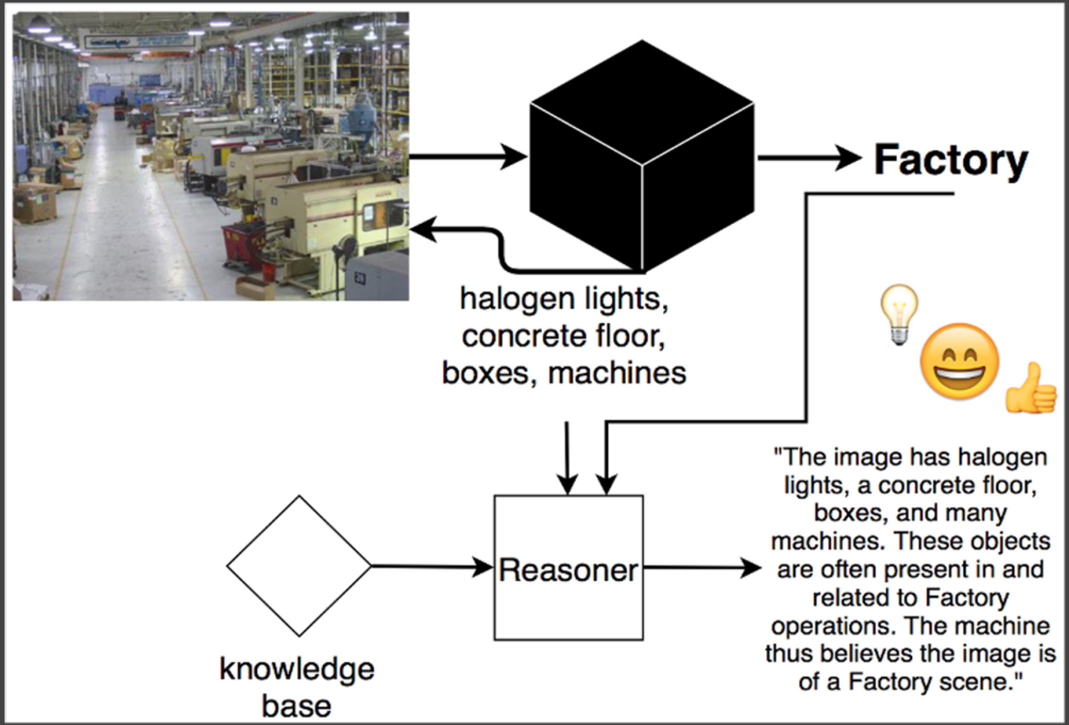
R. Confalonieri, L. Coba, B. Wagner, and T. R. Besold. A historical perspective of explainable artificial intelligence. WIREs Data Mining and Knowledge Discovery, 11(1), 2021. doi: https://doi.org/10.1002/widm.1391

Explainability - Notions



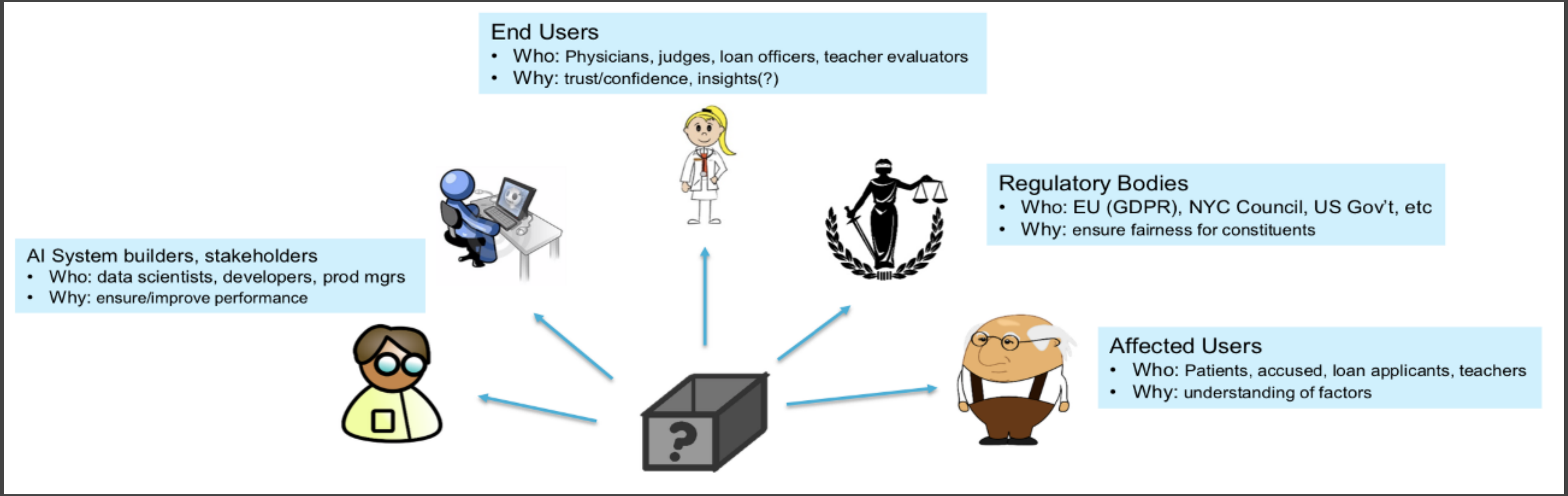
Doran, D., Schulz, S., & Besold, T. R. (2017). What Does Explainable AI Really Mean? A New Conceptualization of Perspectives. 1st International Workshop on Comprehensibility and Explanation in AI and ML Colocated with AI*IA 2017 (Vol. 2071).

Explainability - Notions



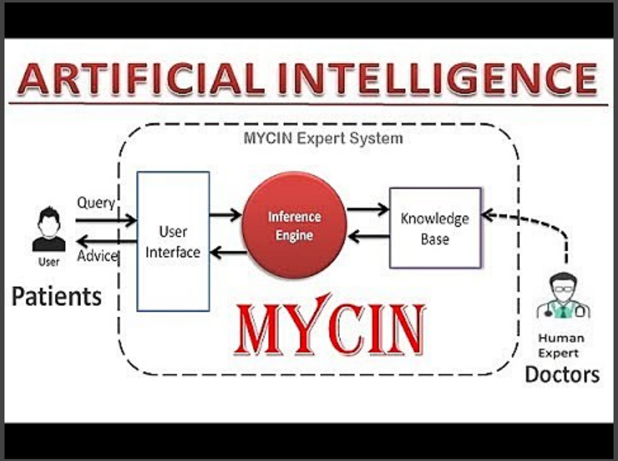
Doran, D., Schulz, S., & Besold, T. R. (2017). What Does Explainable AI Really Mean? A New Conceptualization of Perspectives. 1st International Workshop on Comprehensibility and Explanation in AI and ML Colocated with AI*IA 2017 (Vol. 2071).

Meaningful explanations depend on the stakeholder!



XAI – Expert Systems

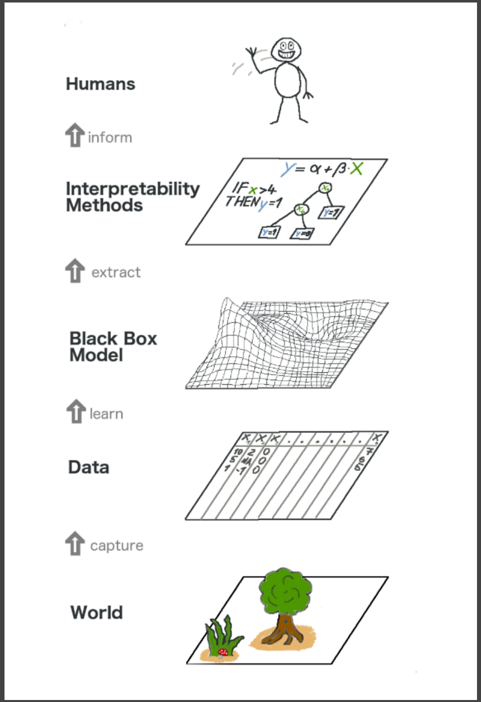
- **Explainable by design**
- Explanations as reasoning traces of decision making process



PREMISE:	(AND (SAME CNTXT GRAM GRAMNEG)
	(SAME CNTXT MORPH ROD)
	(SAME CNTXT AIR ANAEROBIC))
ACTION:	(CONCLUDE CNTXT IDENTITY BACTEROIDES TALLY .6)
IF:	(1) The gram stain of the organism is gramneg,
	(2) The morphology of the organism is rod, and
	(3) The aerobicity of the organism is anaerobic
THEN:	There is suggestive evidence (.6) that
	The identity of the organism is bacteroides

XAI – Machine Learning

- <u>**Post-hoc explanations**</u>
- Classified by **scope** and **model**
  - Local vs Global
  - Specific vs Agnostic



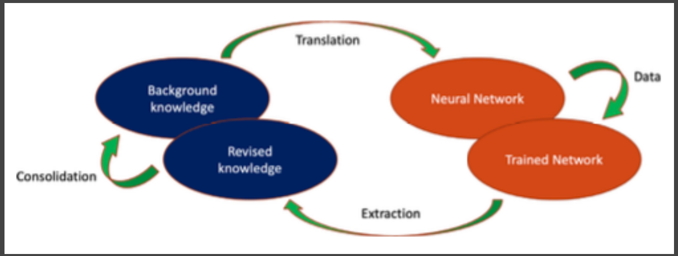
https://www.analyticsvidhya.com/blog/2019/08/decoding-black-box-step-by-step-guide-interpretable-machine-learning-models-python/

XAI – Recommender Systems

- **White-box vs black-box vs model-based**
- Explanations are **goal-oriented** and depend on the **stakeholders**:
  - Persuasive, Trustworthy
  - Efficient, Effective, Satisfying
  - Transparent, Scrutable
- More on this to follow

XAI – Neuro-Symbolic LR

- **Explanations as knowledge extraction**
- Symbolic and connectionist methods
  - Representation
  - Extraction
  - Reasoning
  - Learning



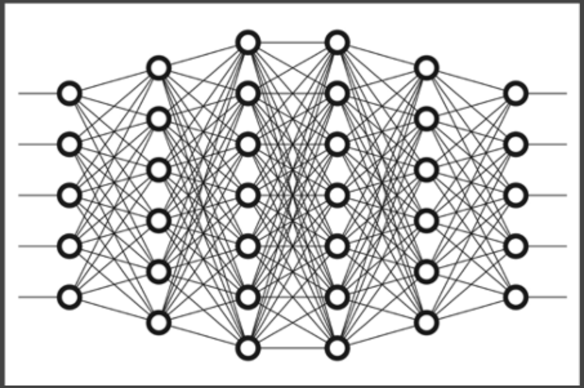
A Neuro-symbolic Example

- **Trepan**: a knowledge extraction algorithm
- Extracts decision tree as rule-like representation describing global model learned by ANN

Craven, M. W., & Shavlik, J. W. (1995). Extracting tree-structured representations of trained networks. In Neural Information Processing Systems (pp. 24–30). Cambridge, MA: MIT Press.

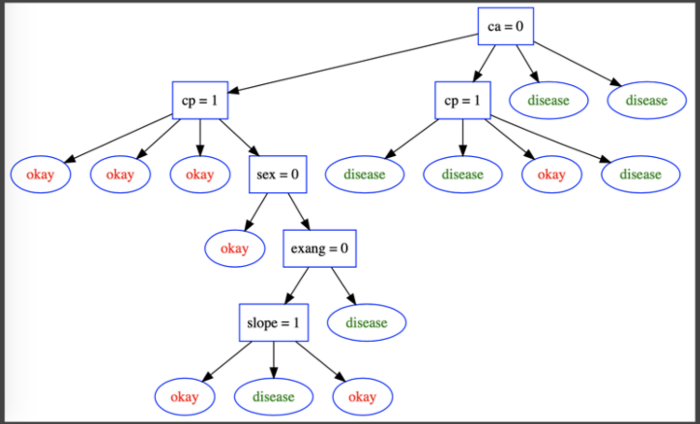
A Neuro-symbolic Example

Oracle (Trained ANN)



Trepan

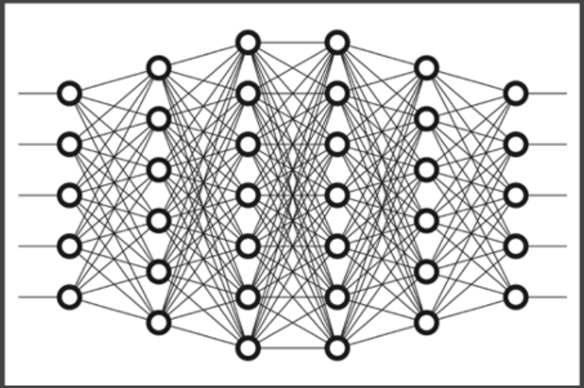
Explanation



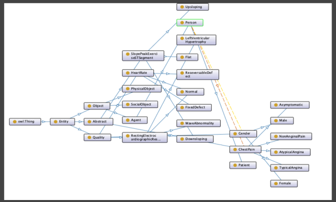
Craven, M. W., & Shavlik, J. W. (1995). Extracting tree-structured representations of trained networks. In Neural Information Processing Systems (pp. 24–30). Cambridge, MA: MIT Press.

Trepan Reloaded

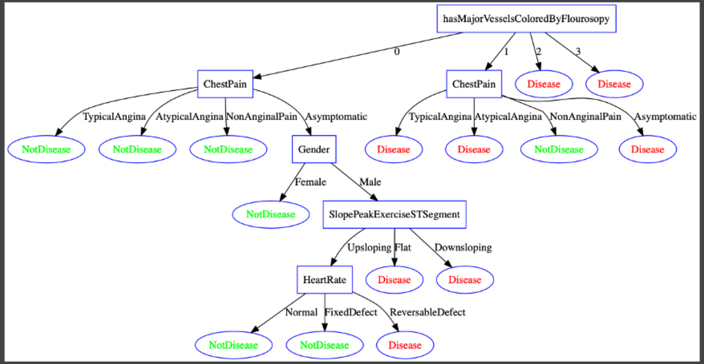
Oracle
(Trained ANN)



Trepan
Reloaded



Knowledge-aware
Explanation



Confalonieri, R., Weyde, T., Besold, T. R., & del Prado Martín, F. M. (2020). Trepan Reloaded: A Knowledge-driven Approach to Explaining Black-box Models. Proceedings of ECAI2020. https://doi.org/10.3233/FAIA200378

Explainable AI (XAI)

- What stands for a (good) explanation?

Expert Systems

Accuracy
Adaptability
Comprehensibility

Machine Learning

Accuracy
Fidelity
Causality

Recommender Systems

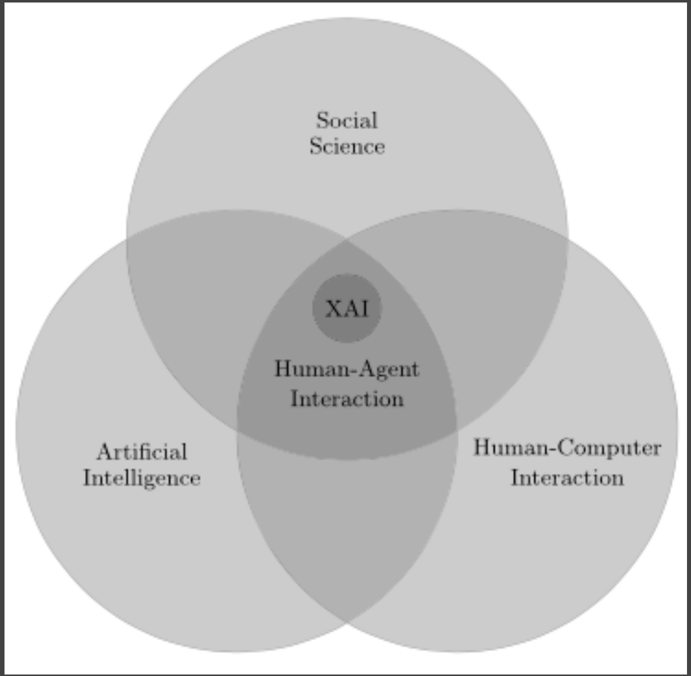
Persuasiveness
Trustworthiness
Efficiency
Effectiveness
Transparency
Scrutability

Neuro-symbolic Learning and Reasoning

Accuracy
Fidelity
Consistency
Comprehensibility

XAI - Human-agent Interaction

- Current approaches suffer from “the inmates running the asylum” phenomenon
- **Human-understandable explanations** are:
  - Contrastive
  - Social
  - Selected



T. Miller. Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267:1–38, 2019. doi: https://doi.org/10.1016/j.artint.2018.07.007

Human-centric explanations

- Causal
- Contrastive
- Social
- Selective
- Transparent
- Privacy-preserving
- Semantic

R. Confalonieri, L. Coba, B. Wagner, and T. R. Besold. A historical perspective of explainable artificial intelligence. WIREs Data Mining and Knowledge Discovery, 11(1), 2021. doi: https://doi.org/10.1002/widm.1391

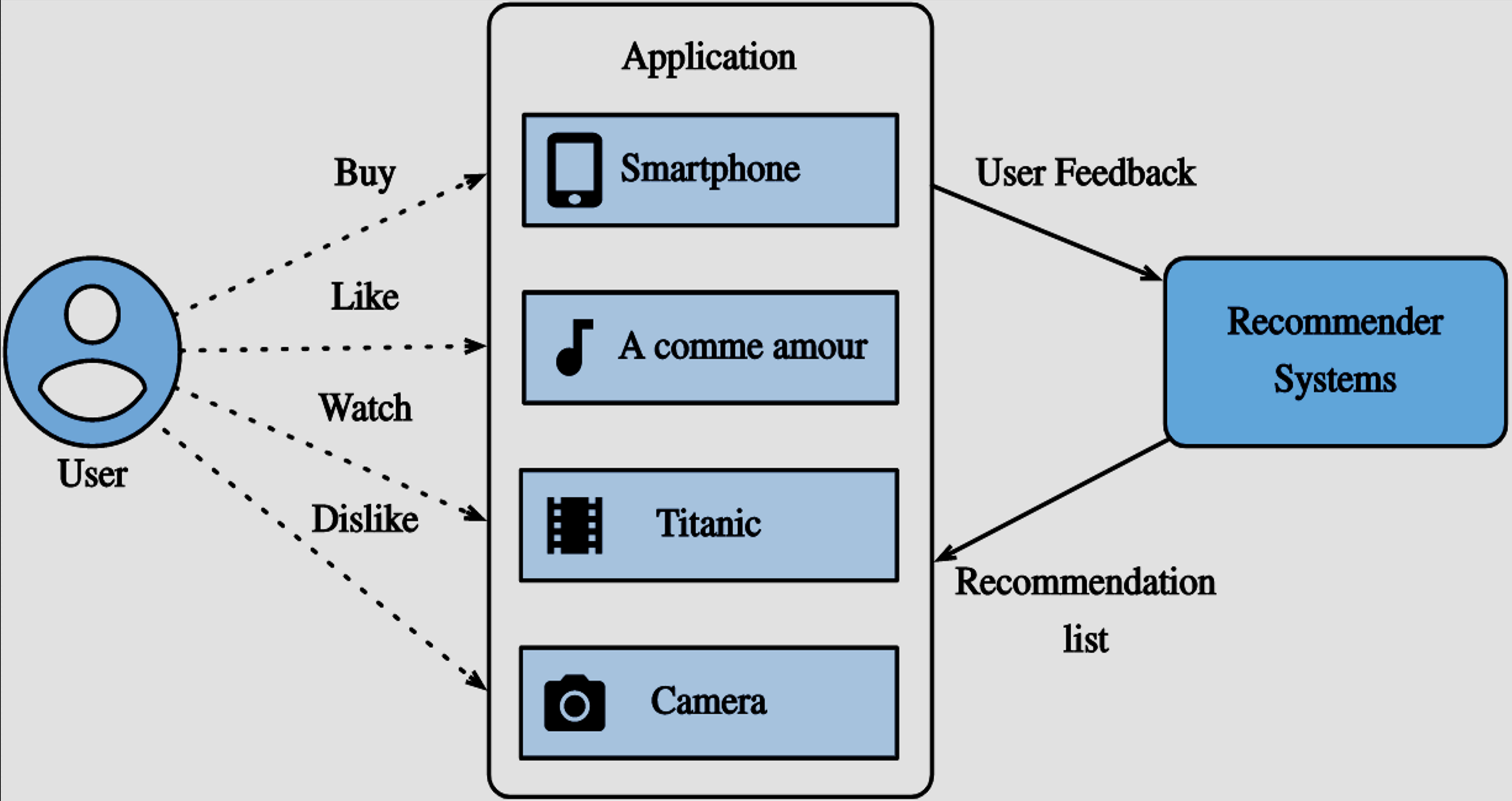
What is a Recommender System?

Problem domain

- Recommendation systems (RecSys) help to match users with items
  - Ease information overload
  - Sales assistance (guidance, advisory, persuasion,…)
- Different system designs / paradigms
  - Based on availability of exploitable data
  - Implicit and explicit user feedback
  - Domain characteristics

RecSys are software agents that elicit the interests and preferences of individual consumers […] and make recommendations accordingly. They [..] support and improve the quality of the decisions consumers make [..] online.

Xiao and Benbasat, E-commerce product recommendation agents: Use, characteristics, and impact, MIS Quarterly 31 (2007), no. 1, 137–209


Application
Smartphone
A comme amour
Titanic
Camera
Buy
Like
Watch
Dislike
User
User Feedback
Recommender Systems
Recommendation list

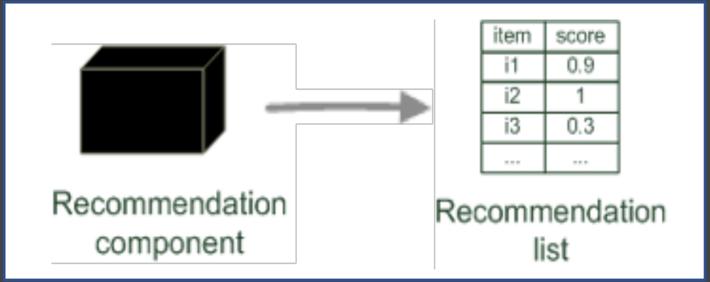
Recommender Systems

- Recommender Systems (RecSys) as a **function**
- Input
  - User model (e.g. ratings, preferences, demographics, situational context)
  - Items (with or without description of item characteristics)
- Output
  - Relevance score. Used for ranking

D. Jannach et al., Recommender Systems – An Introduction, Cambridge University Press, 2011

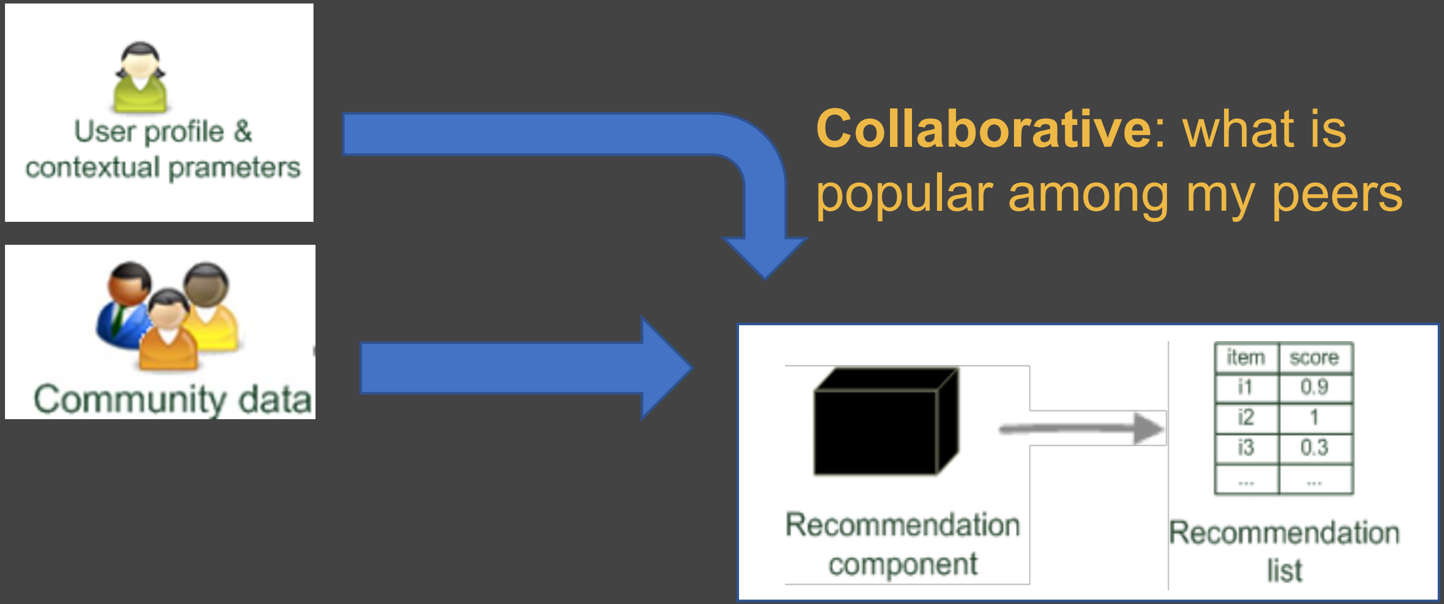
Paradigms of RecSys

Personalized
recommendations



D. Jannach et al., Recommender Systems – An Introduction, Cambridge University Press, 2011

Paradigms of RecSys

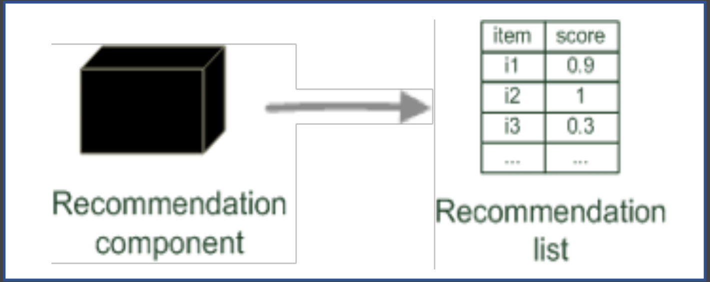


Collaborative: what is popular among my peers

D. Jannach et al., Recommender Systems – An Introduction, Cambridge University Press, 2011

Paradigms of RecSys

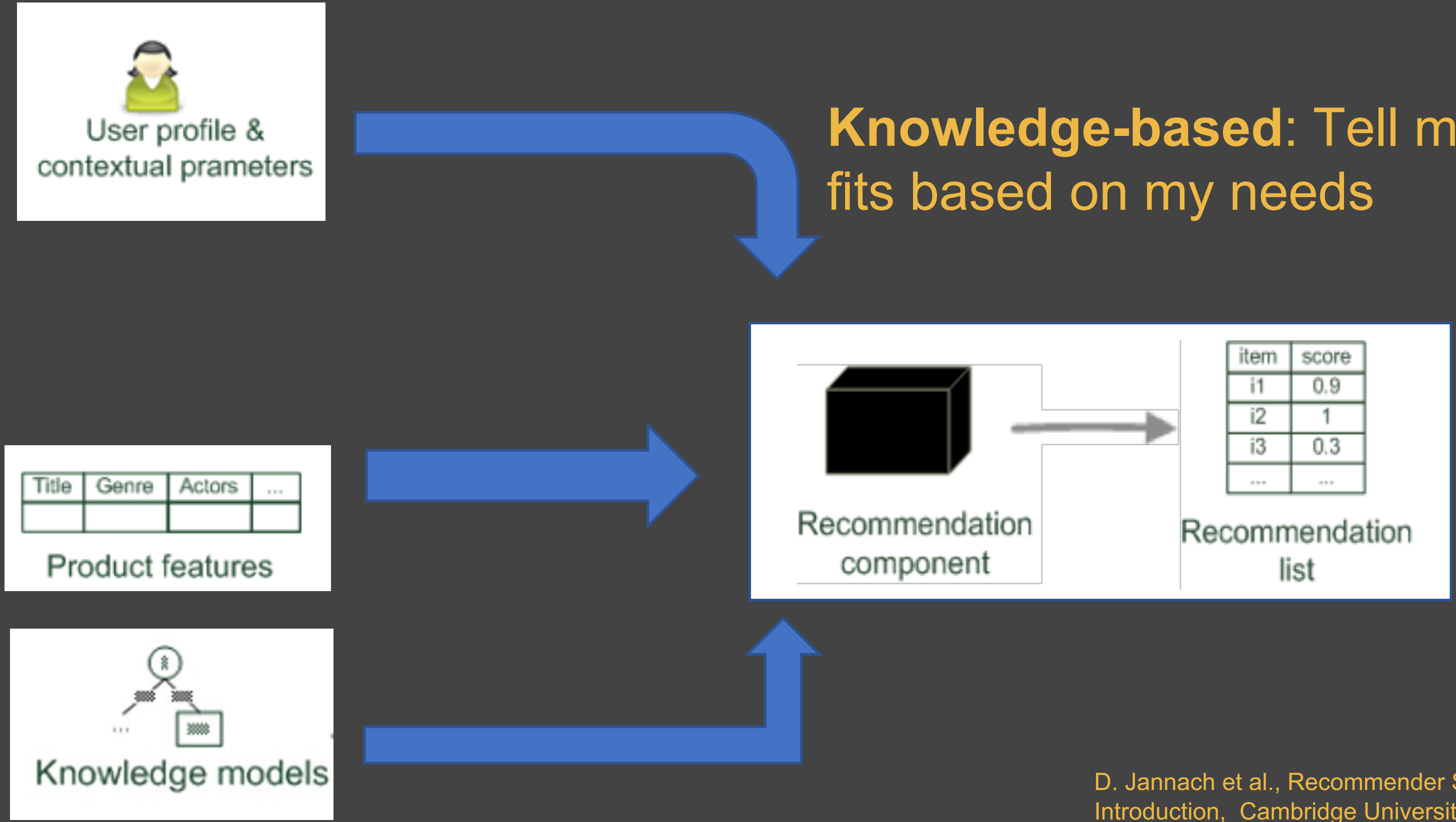
Content-based: show me more of what I liked



D. Jannach et al., Recommender Systems – An Introduction, Cambridge University Press, 2011

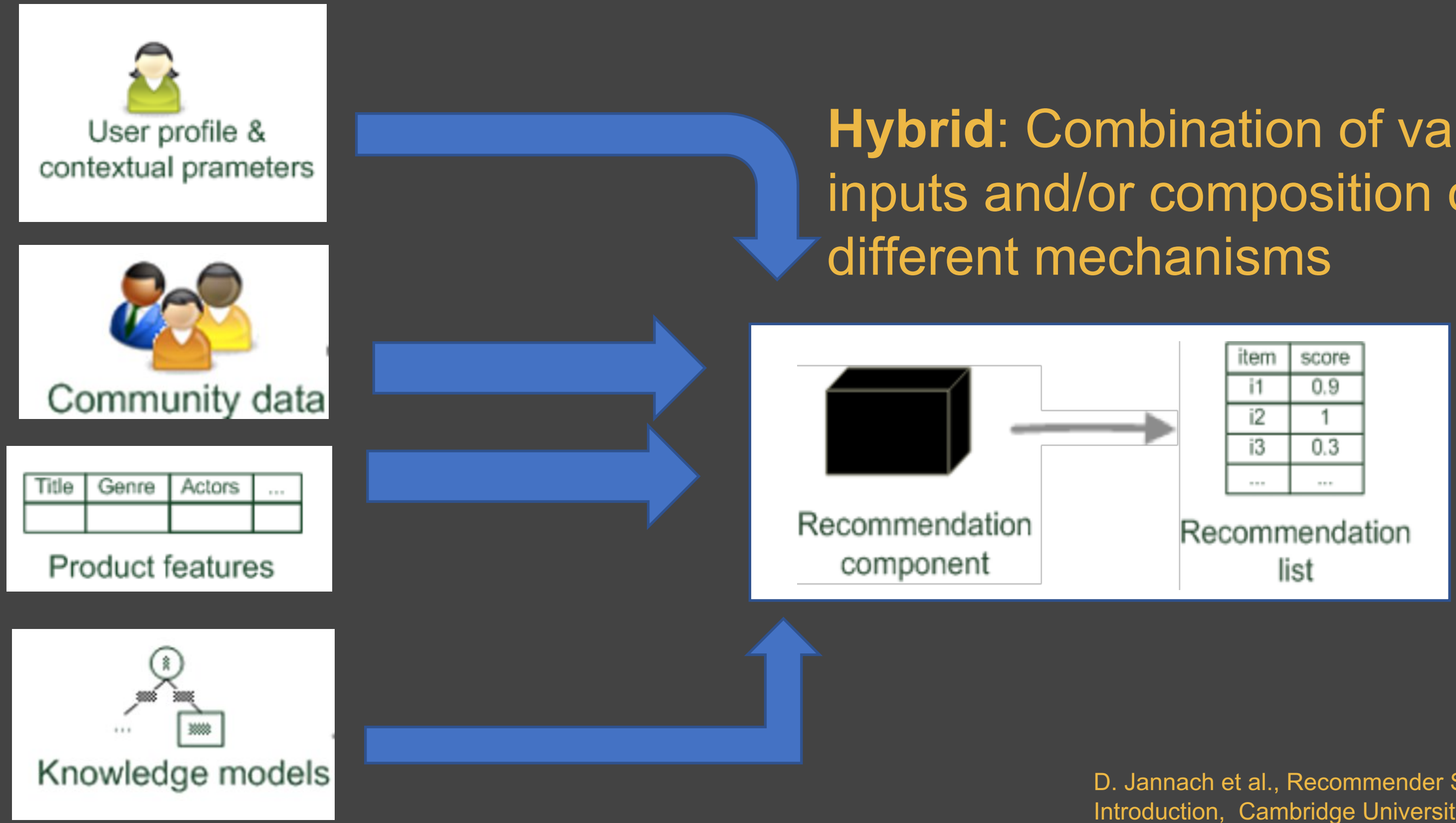
Paradigms of RecSys

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



D. Jannach et al., Recommender Systems – An Introduction, Cambridge University Press, 2011

Paradigms of RecSys



Hybrid: Combination of various inputs and/or composition of different mechanisms

D. Jannach et al., Recommender Systems – An Introduction, Cambridge University Press, 2011

Collaborative Filtering

- Collaborative filtering is the most prominent paradigm
- Approach
  - Use the 'wisdom of the crowd' to recommend items
- Basic idea
  - Users give ratings to catalog items (implicitly or explicitly)
  - Customers, who had similar tastes in the past, will have similar tastes in the future

Collaborative Filtering

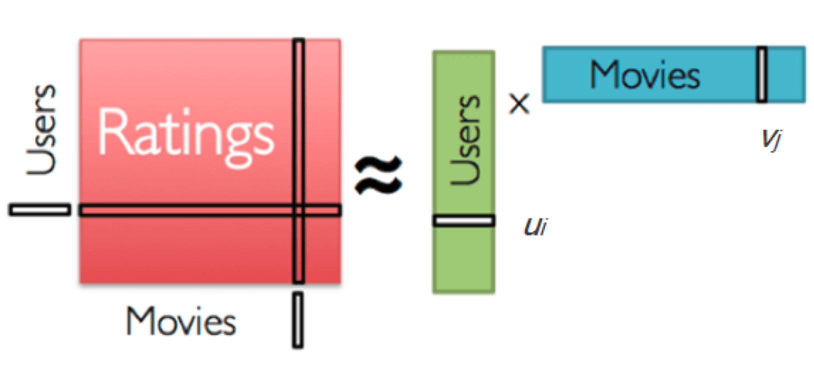
- Input types
  - A matrix of given user-item ratings
  - A sequence user-item interactions
  - Situational context
- Output types
  - A numerical prediction indicating to what degree the current user will like or dislike a certain item
  - A top-N list of recommended items
  - Next item

Memory-based vs model-based

- Memory-based
  - The input is directly used to find neighbors and to make predictions
  - Nearest-Neighbor Methods
  - Scaling problem for real world scenarios
- Model-based
  - Based on a 'model-learning' phase
  - Capture high-level patterns and trends

Algorithms

- Factorization methods
  - Multi-dimensional latent factor space
  - Approximates original rating matrix
- Deep Learning
  - Neural network embeddings



Explanations in Recommender Systems

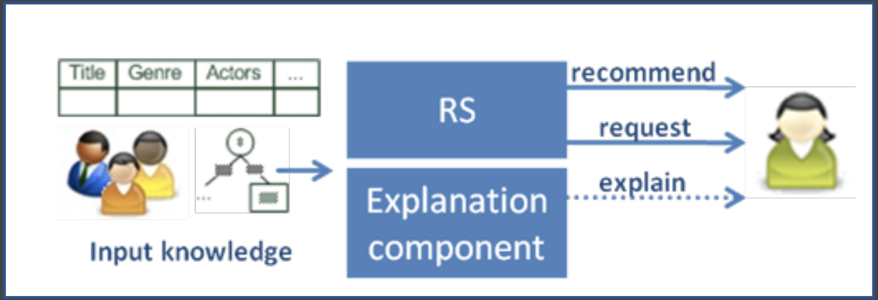
XAI – Recommender Systems

- **Model-based** vs **post-hoc explanations**
- Explanations are **goal-oriented** and depend on the **stakeholders**:
  - A **selling agent** may be interested in promoting particular products
  - A **buying agent** is concerned about making the right buying decision

Friedrich, G.; and Zanker, M. 2011. A Taxonomy for Generating Explanations in Recommender Systems. AI Magazine32(3): 90. ISSN 0738-4602.

Explanations in RecSys

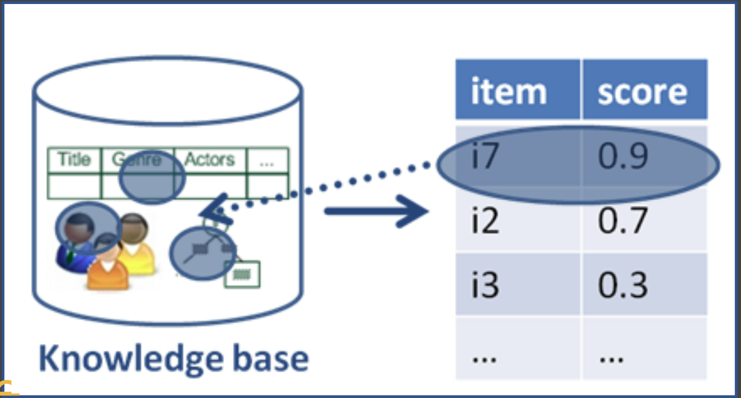
- An explanation in RecSys is additional information to explain the system’s output following some objectives



Friedrich, G.; and Zanker, M. 2011. A Taxonomy for Generating Explanations in Recommender Systems. AI Magazine32(3): 90. ISSN 0738-4602.

Explanations in RecSys

- Form of abductive reasoning
  - Given: $KB \vDash_{RS} i$ (item i is recommended by method RS)
  - Find $KB' \subseteq KB$ s.t. $KB' \vDash_{RS} i$
- Principle of succinctness
  - Find smallest subset of $KB' \subseteq KB$ s.t. $KB' \vDash_{RS} i$
    i.e. for all $KB'' \subset KB'$ holds $KB'' \nvDash_{RS} i$
- But additional filtering
- What is relevant for deduction, might be obvious for humans



Friedrich, G.; and Zanker, M. 2011. A Taxonomy for Generating Explanations in Recommender Systems. AI Magazine32(3): 90. ISSN 0738-4602.

Ultimate Goal

Useful!

- Justify recommendations in a *human-understandable* way

- **But** interpretability is not a goal by itself
- Support the goal of the recommender like improved decision support

Tintarev, N.; and Masthof, J. 2021. Beyond explaining single item recommendations. In Recommender Systems Handbook, to appear.

Goals for Explanations

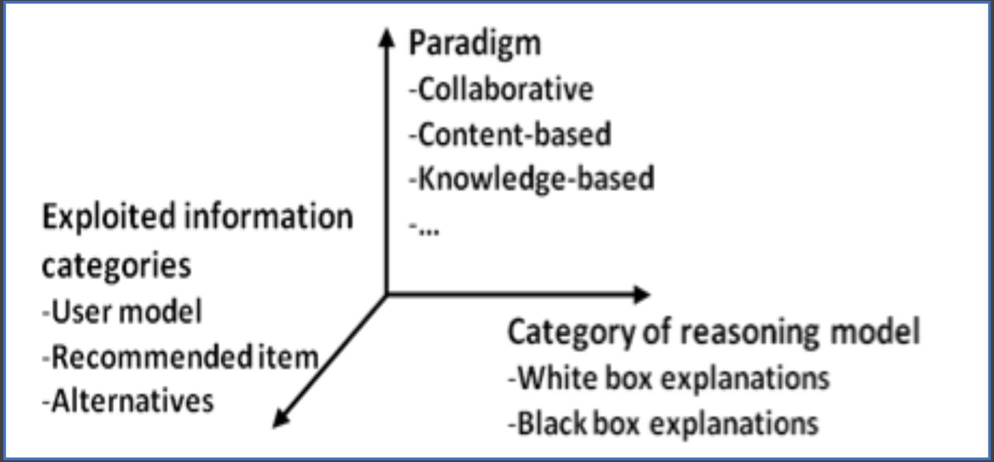
- Transparency
- Validity
- Trustworthiness
- Persuasiveness
- Effectiveness
- Efficiency
- Satisfaction
- Relevance
- Comprehensibility
- Education

Tintarev, N.; and Masthof, J. 2015. Explaining recommen-dations: design and evaluation. In Recommender Systems Handbook, 217–253. Boston, MA: Springer US.

Taxonomy for Explanations

Major design dimensions of current explanation components:

- Category of reasoning model
- Paradigm
- Information categories

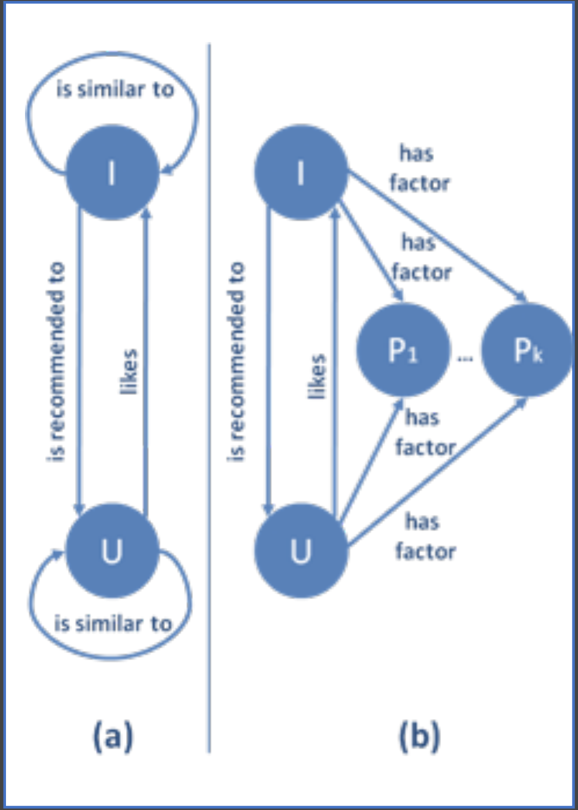


Information categories

- Which information is exploited to derive explanations?
- User model
- Features of the recommended item
- Alternatives

Reasoning paradigm

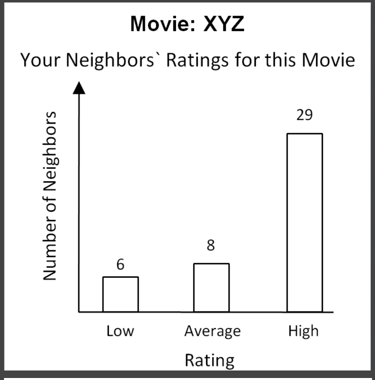
- Classes of objects
  - Users
  - Items
  - Properties
- N-ary relations between them
- Collaborative Filtering
  - Neighborhood based CF (a)
  - Matrix Factorization (b)



Well-known example

- Best-performing explanation interfaces are based on the ratings of neighbors

- Similar neighbors liked the recommended film. The histogram performed better than the table



Movie: XYZ

Personalized Prediction: ****

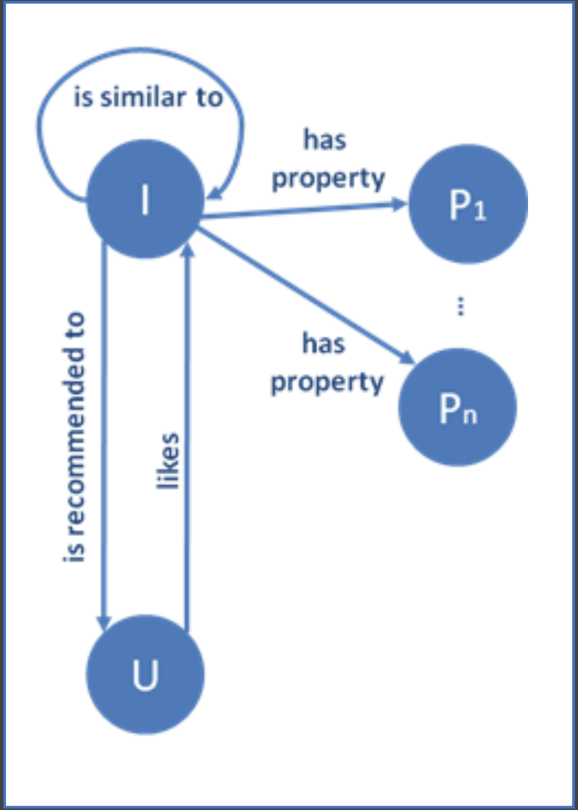
Your Neighbors` Ratings for this Movie

Rating	Number of Neighbors
★	2
★★	4
★★★	8
★★★★	20
★★★★★	9

J. Herlocker. Explaining collaborative filtering recommendations, Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW ’00) (Philadelphia), ACM, 2000, pp. 241–250

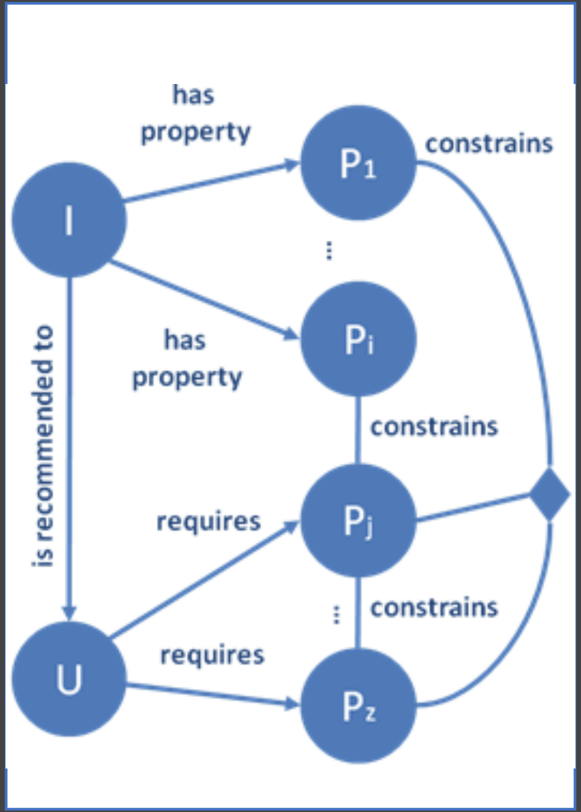
Reasoning paradigm

- Content-based
  - Features/properties characterizing items
  - TF*IDF model
  - Feature-style: explaining based on item features



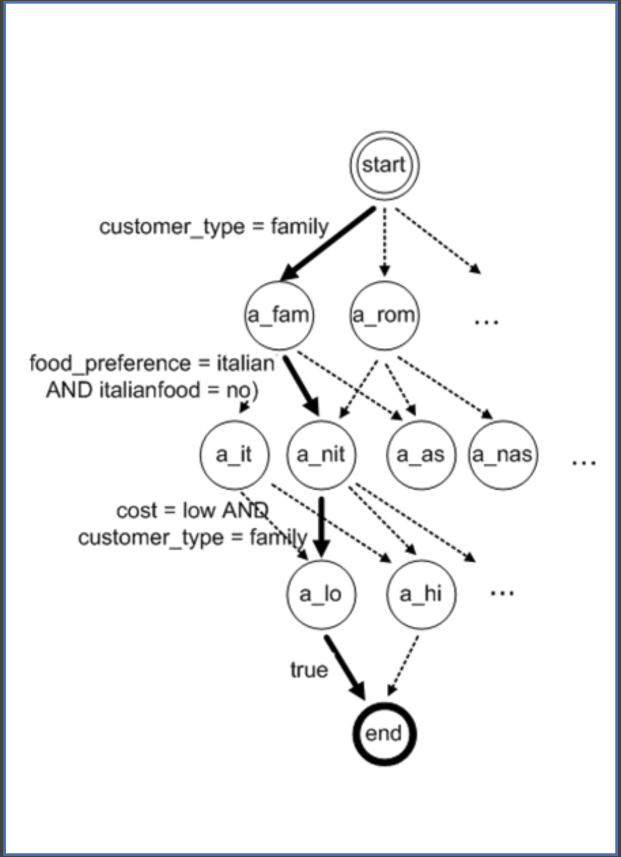
Reasoning paradigm

- Knowledge-based
  - Properties of items
  - User Model
  - Additional mediating domain constraints

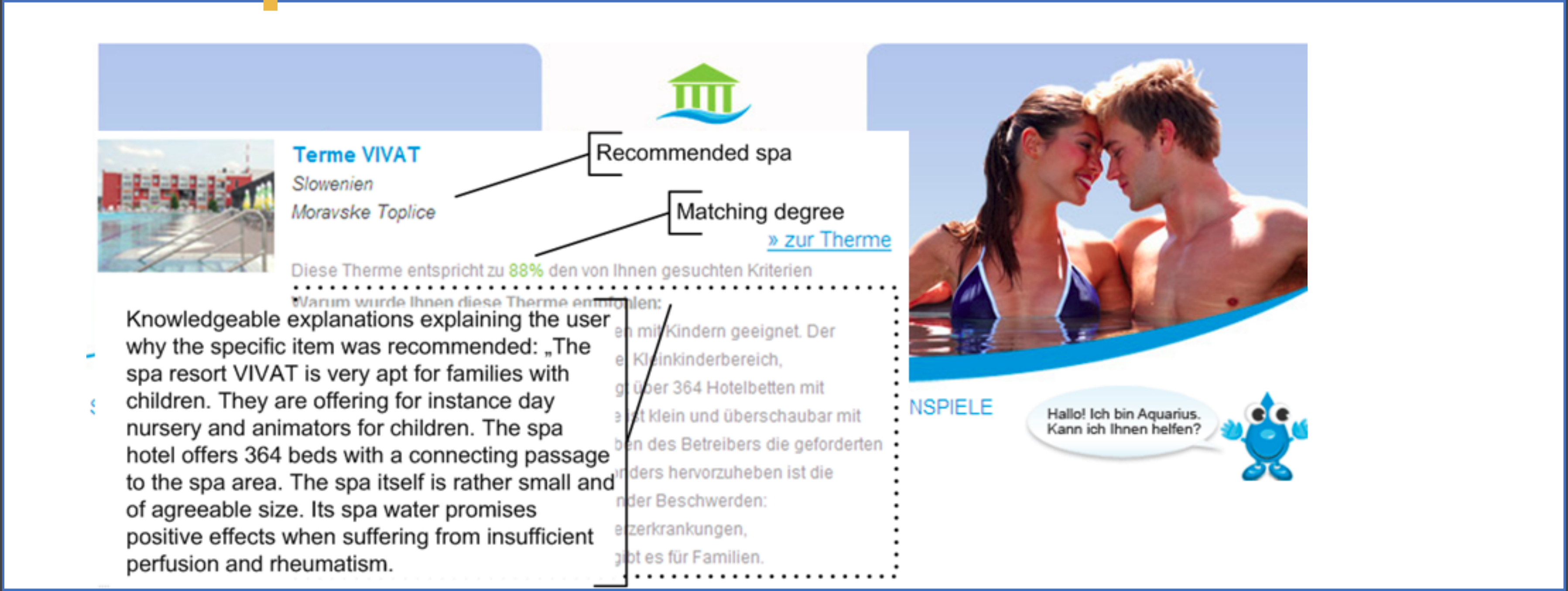


Example

- Layered directed acyclic graph (DAG)
  - U = {customer_type,..}
  - I = {italianfood,..}
  - Nodes represent arguments (canned text)
  - Transition from start to end node not violating domain constraints

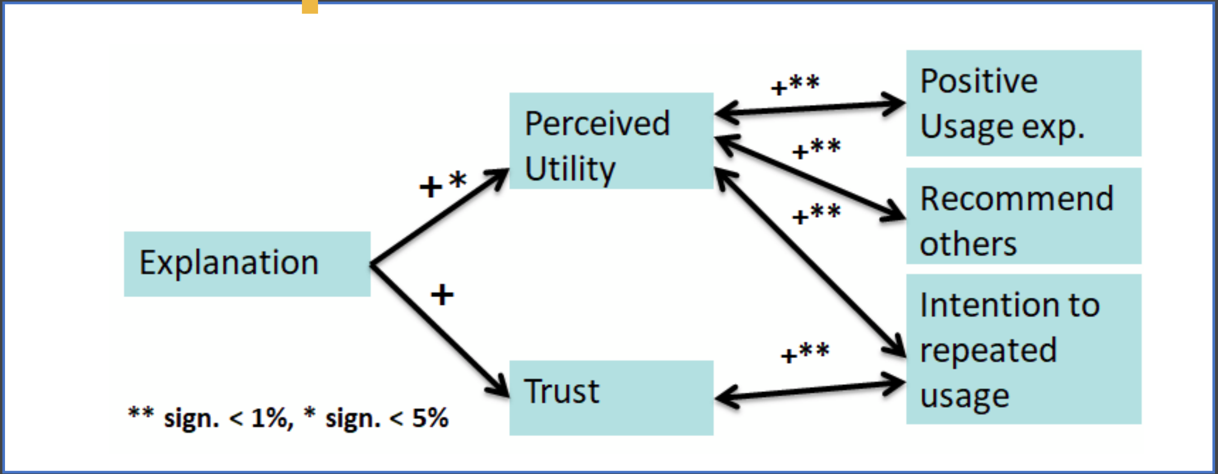


Example



- Search platform for spa resorts

Example



- A/B test: knowledgeable explanations increased perceived utility and intention to use

M. Zanker. The influence of knowledgeable recommendations on users' perception of a recommender system, ACM Conference on Recommender Systems (RecSys ’12), ACM, 2012, pp. 269–272

Category of reasoning model

White-box or explainable-by-design explanations:

- How did the system derive a recommendation

Black-box or post-hoc explanations:

- What justifies the recommendation in the eyes of its recipient

Model-based explanations:

In between the previous two

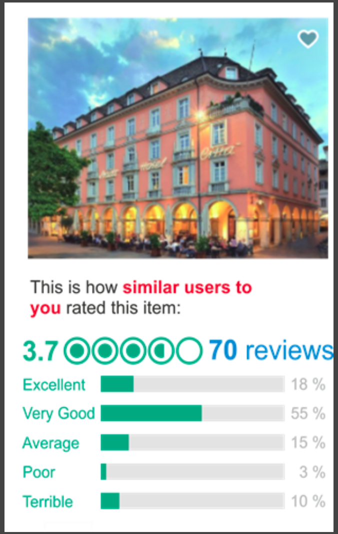
Explanations in CF

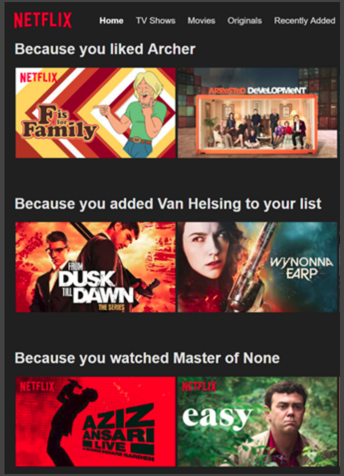
- Explicit recommendation knowledge is not available
- Recommendations based on CF cannot provide arguments as to *why a product is appropriate* for a customer or *why a product does not meet* a customer's requirements
- Post-hoc explanations (see later)

Explanation formats

- User-style
  - It provides explanations based on similar users

- Item-style
  - It is based on choices made by users on similar items





Thank you!
Questions?

Wrapping up

RecoXplainer: a unified, extendable, easy-to-use Python library to develop explainable RecSys

Code available at: https://github.com/ludovikcoba/recoxplainer

Looking for use-cases

Who we are

Dr. Ludovik **Coba**
ludovik.coba@unibz.it

Dr. Roberto **Confalonieri**
rconfalonieri@unibz.it

Prof. Markus **Zanker**
markus.zanker@unibz.it

Thank you!
Questions?