Rule-based recommender systems with applications

RuleML 2023-10-25

Dr. Len Feremans

Prof. dr. Bart Goethals

Adrem Data Lab, University of Antwerp, Belgium

Overview

- Introduction
- Work in progress

Introduction

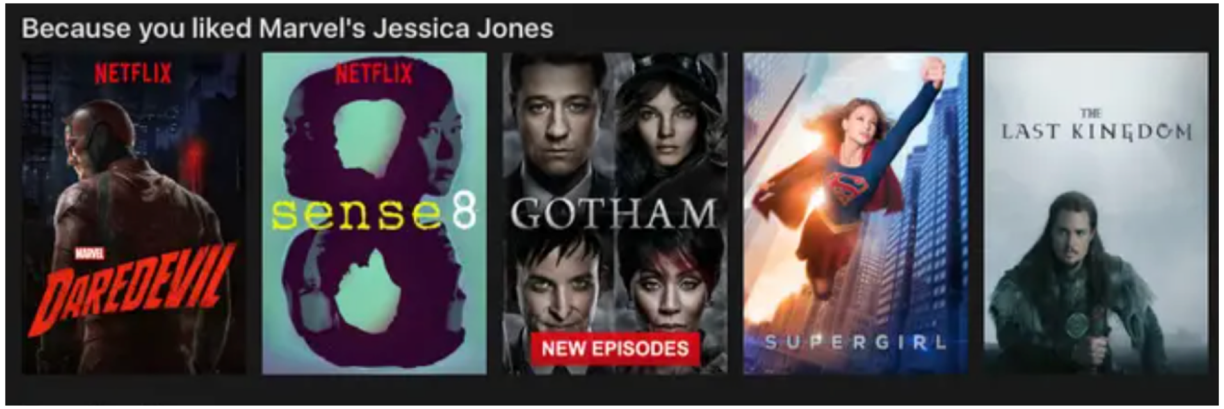
- Interpretable rule-based recommender systems with applications

Introduction

- Recommender systems are vital in e-commerce, news, media, but also in healthcare

- For each user we have a history of interactions with **items**
- Rank **top-k items** most relevant to **each user**

- **Item-based collaborative filtering [1]** results in *"because you like X we recommend Y"*



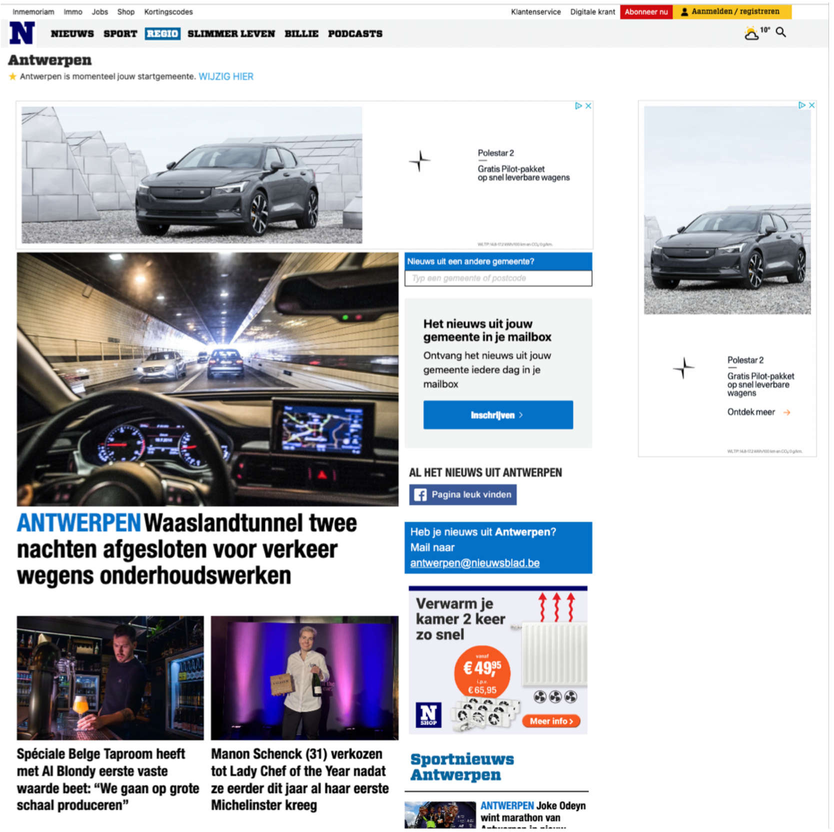
Introduction

- Large-scale **news** (or e-commerce) recommendations [2] in collaboration with Froomle (www.froomle.ai/)

- **Health-care:** Making personalized warnings for Intensive-Care Unit patients.

KU LEUVEN

FROOMLE



Interpretable recommender systems

- **Digital Service Act** (article 27) explicitly mentions recommender systems transparency [3]

- *"Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*" by Cynthia Rudin (2019) [4]:
  - Myth that there is necessarily a **trade-off between accuracy and interpretability**
  - Post-hoc explanations are **not faithful** to what the original model computes

- **Independent evaluation** studies show rule-based baselines are often competitive with deep-learning based methods [5,6]

INDEX DISEACTTPRO TRAINING DSA TRAINING DSA ARTICLES DSA LINKS CYBER RISK GMBH IMPRESSUM

The final text of the Digital Services Act (DSA)

Article 27, Recommender system transparency - the Digital Services Act (DSA)

1. Providers of online platforms that use recommender systems shall set out in their terms and conditions, in plain and intelligible language, the main parameters used in their recommender systems, as well as any options for the recipients of the service to modify or influence those main parameters.

2. The main parameters referred to in paragraph 1 shall explain why certain information is suggested to the recipient of the service. They shall include, at least:

Interpretable recommender systems

- **Concerns** that recommender systems are:
  - **Biased**
    - Towards sensitive attribute of user
    - Towards popular (or category) of items
  - **Unfair**
    - E.g. job recommendations target gender or race
  - **Diversity** and **filter bubbles**
    - All recommendation are similar
    - E.g. because you like "#101 climate change is fake" we recommend …

Best recommender system?

- Ideally **recommender** has:
  - High accuracy
  - Low runtime
  - High coverage*
  - Low popularity bias*
  - High diversity*
  - Calibrated*

Metric	**Description**	**Reference**
Coverage	Number of distinct items appearing in any top-k recommendation, or aggregate diversity	Adomavicius and Kwon (2011a)
Popularity bias	Average popularity of recommended items	Abdollahpouri et al. (2019)
Diversity	Average dissimilarity of all pairs of items in the recommendation set	Zhang and Hurley (2008)
Calibration	Difference in distribution between training and recommended items attributes measured using Kullback-Leibler divergence. For instance, we consider item category, if available, as a common attribute to calibrate in the domain of e-commerce, news or movie recommendations	Steck (2018); Vrijenhoek et al. (2022)

Table 2: Secondary evaluation metrics

- **And is interpretable:**
  - Global: Investigate model (or option to inspect rules)
  - Local: Explain top-1,2,..n prediction for user X

Current research

Research questions

1. Which **existing rule-based methods** are the most accurate and when? Are recommendations diverse? Are they interpretable?
2. How do we create the **next generation rule-based recommenders** with higher top-k predictive accuracy, fewer rules and more diversity?
3. Do **personalized warning** for ICU patients decrease alarm fatigue?

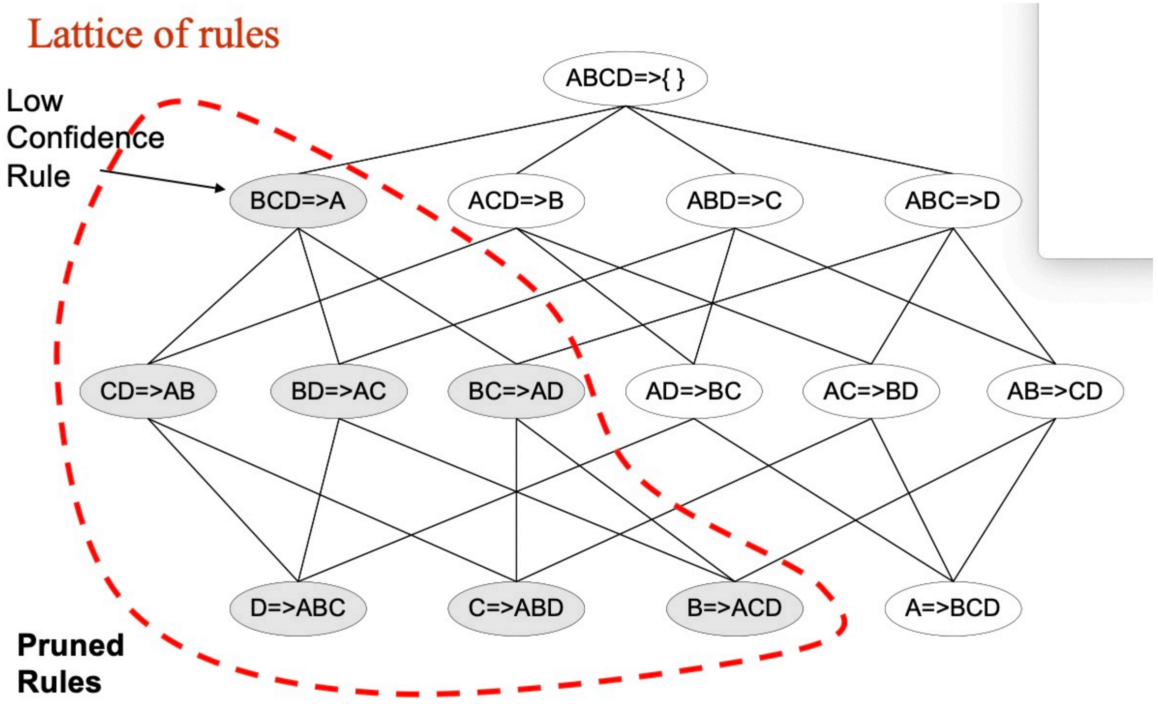
Independent evaluation of rule-based methods

- **Independent evaluation of rule-based recommender algorithms [5,6]**
- Set of public datasets
- Metrics:
  - Accuracy
  - Efficiency
  - Model complexity
  - Interpretability
  - Secondary evaluation metrics

Category	Name	Properties	Applications	Reference
Pairwise	Simple association rules	Prediction based on the last interaction	E-commerce, music	Ludewig and Jannach (2018)
Pairwise	Item-based collaborative filtering	Several confidence and similarity measures. Confidence of matching rules averaged	E-commerce, music, movies, news, ...	Deshpande and Karypis (2004)
Pairwise	Pairwise association rules	Rules ranked on both confidence and support	Dietary recommendations	Osadchiy et al. (2019)
Higher-order	Higher-order item-based collaborative filtering	Confidence of matching rules averaged	E-commerce, music, movies, news, ...	Deshpande and Karypis (2004)
Higher-order	Higher-order rules	Predictions based on last window	Web personalisation	Mobasher et al. (2001)
Higher-order	Weighted Association Rule	Items weighted by recency, frequency and dwell time	Web personalisation	Forsati et al. (2009)
Higher-order	Neighborhood-restricted mining and weighted association rules	Different minimal support and confidence thresholds and mining of rules restricted to similar users	Movies	Gedikli and Jannach (2010)
Sequential	Sequential rules	Pairwise rules weighted by duration of occurrences and prediction based on the last interaction	E-commerce, music	Kamehkhosh et al. (2017)
Sequential	Closes sequential patterns	Closed sequential patterns	Synthetic	Niranjan et al. (2010)
Sequential	Personalized Sequential Pattern Mining	Confidence of rules personalised based on similar users	Web personalisation, e-books	Yap et al. (2012)
Regression	Sparse linear methods	Learn weights of pairwise associations	E-commerce, Movies	Ning and Karypis (2011)
Regression	Higher-order sparse linear methods	Learn weights of pairwise and higher-order associations	E-commerce, social tagging	Christakopoulou and Karypis (2014)

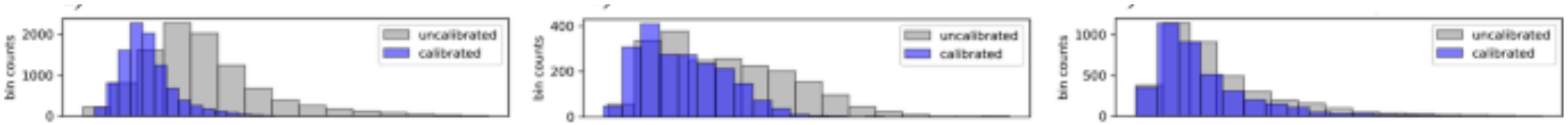
Independent evaluation of rule-based methods

- Challenges:
  - **Efficiency**
    - Pairwise, e.g. A → B
    - Higher-order, e.g. $A_1,A_2,A_3 \rightarrow B$
  - **Analysis of performance** on different domains and datasets?
  - Measuring **interpretability**?
  - **Evaluation scenario?**
    - With an incorrect scenario there is a large gap between offline and online performance [7]



Rule-based recommender algorithms

- OIRI: Optimising Item-based Recommendations towards Interpretability
- Main Idea:
  - Optimisation of the **item-based similarity matrix,** or equivalently, set of pair-wise association rules
  - **Objective 1: Maximal Accuracy**
    - Minimize L2 loss on training test
  - **Objective 2: Fewer rules**
    - Minimize L1 norm of matrix B
  - **Objective 3: Recommendations calibrated** over different popularity classes
    - Minimize KL divergence



  - **Objective 4: High accuracy for long-tail items**
    - Find local association rules
  - **Iterative algorithm** that **minimises multi-objective loss**

Rule-based recommender algorithms

- Challenges:
  - Recommender datasets have **millions** of users and **thousand** of items and interaction matrix is **sparse**
  - We have **implicit**, **noisy** feedback
  - Extremely **skewed distribution**
- Efficiency :
  - Enumerate possible rules of size k: **Exponential** problem
  - Finding best subset of rules: Also exponential problem
- Prior work:
  - In **recommender systems** on calibration, fairness, diversity, etc. [10,11]
  - In **rule-based optimalisation** for classification that consider multiple objectives, i.e. complexity/accuracy trade-off [12, 13, 14, 15]

Personalizing warnings in Intensive Care Unit

- Collaboration with Michela Venturini and Celine Vens at KU Leuven

- About 1400 patients monitored over days/weeks
- **Philips Monitor records alarms** and **vital signs**
- Existing alarms based on **threshold**
- Up to 90% of alarms **are useless** leading to **alarm fatigue** and patient discomfort
- **Annotation** is extremely **expensive**

Method

- **Discover interesting rules:**
  - Validated by healthcare professional
  - Different types of alarms, e.g. cardiac arrest but also malfunctioning of equipment
- **Features**:
  - Discover **multi-variate patterns**
    - e.g. **if** high heartrate **and** low blood-pressure → warn Cardiac arrest
  - Consider **mixed-type** features comprising both alarms and vital signs
  - Focus on **contrastive patterns and rules**, i.e. a rule has a high F1 score towards a certain alarm
  - Look for **time-aware** pattern occurrences, i.e. near an alarm event of a certain type
  - Local to **cluster of patients**

Method

- Technical:
  - We use prefixspan for **pattern mining**
  - We use k-means for **clustering**
  - Segment longitudinal data into **windows**
  - Symbolic Aggregate Approximation (SAX) and summary statistics for **vital signs**
  - Combine features with existing **interpretable classifiers**, i.e. random forest, decision trees, linear model etc.

Findings

- Issue: Real-world data
  - **Messy**, heterogenous mixed-type data
  - **Contextual,** e.g. presence of sensor data most important feature
- **Preliminary reports** highlight interesting findings in clusters
  - Nice patterns for different clusters of patients
    - e.g. healthy/medium/sick or young/old patients towards certain alarms
  - Globally discovered rules are often not that interesting to medical professionals
- Ongoing: Semi-supervised learning
  - Get labels/feedback in collaboration with medical professionals

Take-away points

- Rule-based recommenders are a good option concerning accuracy
- Rule models are intrinsic interpretable models:
  - Legal requirement
  - Essential in healthcare
- Working on **first independent study** of rule-based recommenders
- Working on **novel rule-based recommender systems** that jointly optimises both accuracy, complexity and secondary evaluations metrics
- Working on discovering rules for **personalising** and **smarter alarms** towards **ICU** professionals

Questions?

I am interested in international collaboration/mobility with the RuleML community in the next years!

Please contact me:

- Linked In: https://www.linkedin.com/in/len-feremans-a7b1592/
- E-mail: Len.feremans@uantwerpen.be

References

- [1] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001, April). Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web (pp. 285-295).
- [2] Feremans, L., Verachtert, R., & Goethals, B. (2022). A neighbourhood-based location-and time-aware recommender system. In Workshop on Online Recommender Systems, ACM RecSys 2022, September 23rd 2022, Seattle, WA, USA.
- [3] Digital Service Act: https://www.eu-digital-services-act.com/Digital_Services_Act_Article_27.html#:~:text=1.,2.
- [4] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence, 1(5), 206-215.
- [5] Ludewig, Malte, and Dietmar Jannach. "Evaluation of session-based recommendation algorithms." *User Modeling and User-Adapted Interaction* 28 (2018): 331-390.
- [6] Ferrari Dacrema, M., Boglio, S., Cremonesi, P., & Jannach, D. (2021). A troubling analysis of reproducibility and progress in recommender systems research. *ACM Transactions on Information Systems (TOIS)*, *39*(2), 1-49.
- [7] Michiels, L., Verachtert, R., & Goethals, B. (2022, September). Recpack: An (other) experimentation toolkit for top-n recommendation using implicit feedback data. In *Proceedings of the 16th ACM Conference on Recommender Systems* (pp. 648-651).
- [8] Ning, X., & Karypis, G. (2011, December). Slim: Sparse linear methods for top-n recommender systems. In *2011 IEEE 11th international conference on data mining* (pp. 497-506). IEEE.

References

- [9] Steck, H. (2019, May). Embarrassingly shallow autoencoders for sparse data. In *The World Wide Web Conference* (pp. 3251-3257).
- [10] Abdollahpouri, H., Nazari, Z., Gain, A., Gibson, C., Dimakopoulou, M., Anderton, J., ... & Jebara, T. (2023, February). Calibrated recommendations as a minimum-cost flow problem. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining* (pp. 571-579).
- [11] Steck, H. (2018, September). Calibrated recommendations. In *Proceedings of the 12th ACM conference on recommender systems* (pp. 154-162).
- [12] Boley, M., Teshuva, S., Bodic, P. L., & Webb, G. I. (2021). Better short than greedy: Interpretable models through optimal rule boosting. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)* (pp. 351-359). Society for Industrial and Applied Mathematics.
- [13] Mortier, T., Wydmuch, M., Dembczyński, K., Hüllermeier, E., & Waegeman, W. (2021). Efficient set-valued prediction in multi-class classification. *Data Mining and Knowledge Discovery*, *35*(4), 1435-1469.
- [14] Lakkaraju, H., Bach, S. H., & Leskovec, J. (2016, August). Interpretable decision sets: A joint framework for description and prediction. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1675-1684).
- [15] Dembczynski, K., Jachnik, A., Kotlowski, W., Waegeman, W., & Hüllermeier, E. (2013, May). Optimizing the F-measure in multi-label classification: Plug-in rule approach versus structured loss minimization. In *International conference on machine learning* (pp. 1130-1138). PMLR.