

A Design Space for Explainable Ranking and Ranking Models

I. Al Hazwani^{1,2} , J. Schmid¹ , M. Sachdeva¹  and J. Bernard^{1,2} 

¹University of Zurich, Switzerland

²Digital Society Initiative, Zürich, Switzerland

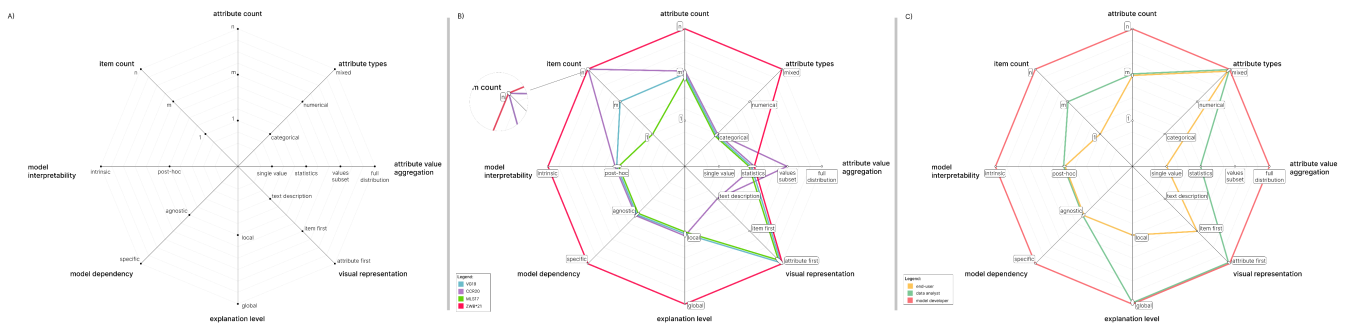


Figure 1: A) Design space for explainable rankings and ranking models composed of 8 dimensions. B) Mapping of four pioneer explainers, according to their explanation characteristics. C) Mapping of three different users groups, exposing different requirements to explainers.

Abstract

Item ranking systems support users in multi-criteria decision-making tasks. Users need to trust rankings and ranking algorithms to reflect user preferences nicely while avoiding systematic errors and biases. However, today only few approaches help end users, model developers, and analysts to explain rankings. We report on the study of explanation approaches from the perspectives of recommender systems, explainable AI, and visualization research and propose the first cross-domain design space for explainers of item rankings. In addition, we leverage the descriptive power of the design space to characterize a) existing explainers and b) three main user groups involved in ranking explanation tasks. The generative power of the design space is a means for future designers and developers to create more target-oriented solutions in this only weakly exploited space.

1. Introduction

Despite the usefulness of item rankings for multi-criteria decision-making tasks, in the real-world end-users often cannot control the ranking creation process by themselves. Existing work reveals that rankings can a) have inherent biases [Cas19], or b) are not human-centered [LV21]. However, only few pioneering works have been presented for explaining item rankings [PP20, ZWB*21]. *Explainable Artificial Intelligence* (XAI) approaches reveal computations behind black-box models [DR20, AB18] such as recommender systems (RS) that generate item rankings. In parallel to explainable RS research, there has been interest from the visualization (VIS) community to assess and analyze item rankings [AL20, LLZ*21, XMT*20] to capture users' feedback [YSJ19, CLK*14], or to support decisions of explanatory chart usage [VHS*17, QRD*20]. In designing explainable item ranking systems, knowledge from different research domains such as XAI, RS, and VIS needs to integrate, with the need for a common language of description. Moreover, different stakeholders with different explainability needs exist, posing the question of how to effectively design explainable ranking systems. In summary, we identify a lack of structure that

would a) allow for a joint description of existing solutions and b) guide designers towards novel ranking explanation solutions.

Our primary contribution is a design space of XAI approaches for explaining ranking results and ranking models. The design space is the result of a careful reflection of related work in the three fields XAI, RS and VIS, as well as a synthesis of their inherent concepts and characteristics. To validate the descriptive power [Bea04] of the design space, we map existing explainers into the space, and analyze commonalities and differences across approaches. Our second contribution is the characterization of three main user groups primarily involved in the creation and use of rankings. Using the design space, we discuss main characteristics and design requirements of user groups to ease the abstraction of design targets for future explainers.

2. A Design Space for Explainable Rankings

We present a design space for explainable rankings and ranking algorithms, as a result of systematic literature reviews, reflections, as well as a synthesis of the fields XAI, RS, and VIS research.

2.1. Abstraction of Three Influential Research Fields

XAI : Multiple surveys around explainability have been published providing support from both the model-driven [AS22, MZR21, ?] and the human-driven perspective [LV21, AS22]. Relevant to our work are approaches for the explanation of rankings, differing in how many items and attributes are considered [VG19, CCR20, MLS17], and as to whether access to the model architecture is supported [ZWB*21]. Finally, XAI differentiates between model-agnostic and model-specific approaches, the latter in the need of incorporating interpretability constraints within the inherent structure and learning mechanisms underlying the black-model [Rai20].

Recommender Systems: Interactive RS enable transparency and insight on output rankings [HPV16] using a model-agnostic approach. Approaches differ in their focus on using a) content-based techniques [BVG*15], b) collaborative filtering-based techniques [GOB*10], and c) both [BOH12]. Additionally, approaches differ in visualization techniques used for model interpretability.

VIS Research: VIS solutions for the analysis of rankings span from simple forms [KVD*18, ZSHL18] to complex visualization systems [SS04, KR17]. RankVis [PP20] is a pioneer explanation approach that also enables developers to interpret the (learning-to-rank) model, by showing attribute distributions of item subsets. Similar to other VIS approaches, RankVis is attribute-centered, enabling data analysts understand how single [SS04] or multiple [GLG*13, PSTW*17, WDC*18, CL04] focused attributes contribute to the item ranking. In contrast, item-centered attributes show attributes for focused items, e.g., in the context of sensitivity analysis and/or ranking comparisons [GLG*13, XMT*20].

2.2. A Design Space for Explainers of Item Rankings

The synthesis of abstractions in *XAI*, *RS*, and *VIS* reveals eight dominating characteristics that form the dimensions of the design space as shown in Figure 1A. Each dimension contains different levels, always with increasing complexity from the center to the outer, useful to a) describe existing explainers and b) generate uncharted room for yet other types of explainers.

Item Count: number of explained items, from l , to m , to n .

Attribute Count: attributes considered, from l , to m , to n .

Attribute Type: attribute types supported by the explainer for explaining the ranking, from *categorical*, *numerical*, to *mixed*.

Visual Representation: entails the type of information that is preferred, ranging from *text description*, to *item-first*, to *attribute first*.

Attribute Aggregation: characterizes the level of detail used to show distributions of attribute values, from *single value*, to *descriptive statistics*, to *values subset*, to *full distribution*.

Explanation Level: explains how a ranking model works, either *local* if it explains decisions for single instances, or *global* if it explains the decision-making mechanism of the entire ranking model.

Model Dependency: differentiates whether access to the inherent characteristics of a model architecture is needed to generate an explanation (*model-specific*), or if solely the input and output solves the explanation task (*model-agnostic*).

Model Interpretability: an XAI method can be implemented on two different levels: *intrinsic* if the ML model itself provides explanations such as decision trees or logistic regression models, or *post-hoc* in case a helper model is used to mimic the behavior.

3. Leveraging the Descriptive Power of the Design Space

3.1. Mapping of Existing Explainers

We validate the design space by mapping the four existing XAI approaches for rankings into the space, using the notion of radar charts in combination with metro maps [NMPR07], as Figure 1B shows. It becomes apparent that three approaches have similar characteristics, using a local level of explanation, an agnostic model dependency, and a post-hoc model interpretability [VG19, CCR20, MLS17]. In contrast, the work by Zhuang et al. [ZWB*21] uses intrinsic model interpretability and a global explanation level and is by far the most sophisticated XAI solution observed. What also stands out is that many (combinations of) levels in the design space are weakly populated, given the low number of existing approaches, leaving much room for future implementations. Future developments may be inspired by a more systematic consideration of user groups, as elaborated in the next section.

3.2. Characterization and Mapping of User Groups

For the design of explainable item ranking systems, it is essential to take a human-centered approach and align the explanations with user knowledge and user needs [LV21]. Based on existing works and the study of real-world ranking applications, we identify three different user groups with considerable differences, as the design space shows (cf. Figure 1B). With *end users*, *data analysts*, and *model developers*, we characterize user groups by their order of required XAI complexity, from simplistic to most sophisticated.

End Users make decisions based on a given ranking, in professional environments and in personal-life situations alike. This user group is in line with the notion of *non-experts* as often used in visualization research [HKPC18], as well as with *AI novices*, as coined in XAI [MZR21]. End users want to understand why an item is located in a specific position and increase their trust in the system, e.g., to make an informed decision in a shopping experience.

Data Analysts are interested in the (visual) analysis of multiple items at once, in contrast to end users. Analysts' needs include a model-based explainability using visualization, to draw general conclusions about model characteristics, as well as items and attributes. Data analysts align with *domain experts* in VIS and *data experts* in XAI [MZR21]. Moreover, a related user group that gained attention in VIS is *data journalists* [Her18, KBM21], e.g., those interested in unveiling biases in public ranking models.

Model Developers differ from other users by their knowledge on AI and RS, and by their focus on model-building. Model developers, called *AI experts* in XAI [MZR21], have an interest in the interpretability of rankings, ranking models, and their behavioral characteristics in particular, e.g., for debugging and refinement purposes. They want detailed explanations (intrinsic model interpretability and specific model dependency) and often rely on RS techniques to create the underlying ranking system [HPV16].

4. Conclusion and Future Work

We proposed a design space for explainable ranking by merging knowledge from XAI, RS, and VIS. We validated our design space by mapping previous work on explainable rankings and user needs. Future work includes a) expanding the design space towards user interaction, b) leveraging the design space to design human-centered XAI ranking systems, and c) cross-cutting the design space with industry human-AI guidelines [WWP*20].

References

- [AB18] ADADI A., BERRADA M.: Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE access* 6 (2018), 52138–52160. 1
- [AL20] AHN Y., LIN Y.-R.: Fairsight: Visual analytics for fairness in decision making. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 26, 1 (2020), 1086–1095. doi:10.1109/TVCG.2019.2934262. 1
- [AS22] ALICIOGLU G., SUN B.: A survey of visual analytics for explainable artificial intelligence methods. *Computers & Graphics* 102 (2022), 502–520. 2
- [Bea04] BEAUDOUIN-LAFON M.: Designing interaction, not interfaces. In *Advanced visual interfaces AVI* (2004), ACM Press, pp. 15–22. doi:10.1145/989863.989865. 1
- [BOH12] BOSTANDJIEV S., O'DONOVAN J., HÖLLERER T.: Tasteweights: a visual interactive hybrid recommender system. In *ACM conference on Recommender systems* (2012), pp. 35–42. 2
- [BVG*15] BRUNS S., VALDEZ A. C., GREVEN C., ZIEFLE M., SCHROEDER U.: What should i read next? a personalized visual publication recommender system. In *Human Interface and the Management of Information* (2015), Springer, pp. 89–100. 2
- [Cas19] CASTILLO C.: Fairness and transparency in ranking. In *ACM SIGIR Forum* (2019), vol. 52, ACM New York, NY, USA, pp. 64–71. 1
- [CCR20] CHOI J., CHOI J., RHEE W.: Interpreting neural ranking models using grad-cam. *arXiv preprint arXiv:2005.05768* (2020). 2
- [CL04] CARENINI G., LOYD J.: Valuecharts: Analyzing linear models expressing preferences and evaluations. Association for Computing Machinery, p. 150–157. doi:10.1145/989863.989885. 2
- [CLK*14] CHOO J., LEE C., KIM H., LEE H., LIU Z., KANNAN R., STOLPER C. D., STASKO J., DRAKE B. L., PARK H.: Visir: Visual analytics for information retrieval and recommendation with large-scale document data. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2014), IEEE, pp. 243–244. 1
- [DR20] DAS A., RAD P.: Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371* (2020). 1
- [GLG*13] GRATZL S., LEX A., GEHLENBORG N., PFISTER H., STREIT M.: Lineup: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 19, 12 (2013), 2277–2286. doi:10.1109/TVCG.2013.173. 2
- [GOB*10] GRETARSSON B., O'DONOVAN J., BOSTANDJIEV S., HALL C., HÖLLERER T.: Smallworlds: visualizing social recommendations. In *Computer Graphics Forum* (2010), vol. 29, Wiley Online Library, pp. 833–842. 2
- [Her18] HERAVI B. R.: Data journalism in 2017: A summary of results from the global data journalism survey. In *Lecture Notes in Computer Science* (2018), vol. 10766, Springer, pp. 107–113. doi:10.1007/978-3-319-78105-1_13. 2
- [HKPC18] HOHMAN F., KAHNG M., PIANTA R., CHAU D. H.: Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 25, 8 (2018), 2674–2693. 2
- [HPV16] HE C., PARRA D., VERBERT K.: Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications* 56 (2016), 9–27. 2
- [KBM21] KASICA S., BERRET C., MUNZNER T.: Table scraps: An actionable framework for multi-table data wrangling from an artifact study of computational journalism. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 27, 2 (2021), 957–966. doi:10.1109/TVCG.2020.3030462. 2
- [KR17] KUHLMAN C., RUNDENSTEINER E.: Towards an interactive learn-to-rank system for economic competitiveness understanding. In *Interactive Data Exploration and Analytics Workshop (IDEA)* (2017). 2
- [KVD*18] KUHLMAN C., VANVALKENBURG M., DOHERTY D., NURBEKOVA M., DEVA G., PHYO Z., RUNDENSTEINER E., HARRISON L.: Preference-driven interactive ranking system for personalized decision support. In *Information and Knowledge Management* (2018), p. 1931–1934. doi:10.1145/3269206.3269227. 2
- [LLZ*21] LIU Q., LI Q., ZHU Z., YE T., MA X.: Inspecting the process of bank credit rating via visual analytics. In *2021 IEEE Visualization Conference (VIS)* (2021), pp. 136–140. doi:10.1109/VIS49827.2021.9623312. 1
- [LV21] LIAO Q. V., VARSHNEY K. R.: Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790* (2021). 1, 2
- [MLS17] MUHAMMAD K., LAWLOR A., SMYTH B.: On the pros and cons of explanation-based ranking. In *Conference on Case-Based Reasoning* (2017), Springer, pp. 227–241. 2
- [MZR21] MOHSENI S., ZAREI N., RAGAN E. D.: A multidisciplinary survey and framework for design and evaluation of explainable ai systems. *ACM Transactions on Interactive Intelligent Systems (TiIS)* 11, 3-4 (2021), 1–45. 2
- [NMPR07] NEUMAYER R., MAYER R., PÖLZLBAUER G., RAUBER A.: The metro visualisation of component planes for self-organising maps. In *Conference on Neural Networks (IJCNN)* (2007), IEEE, pp. 2788–2793. doi:10.1109/IJCNN.2007.4371401. 2
- [PP20] PEREIRA M. M., PAULOVICH F. V.: Rankviz: A visualization framework to assist interpretation of learning to rank algorithms. *Computers & Graphics* 93 (2020), 25–38. doi:10.1016/j.cag.2020.09.017. 1, 2
- [PSTW*17] PAJER S., STREIT M., TORSNEY-WEIR T., SPECHTENHAUSER F., MÖLLER T., PIRINGER H.: Weightlifter: Visual weight space exploration for multi-criteria decision making. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 23, 1 (2017), 611–620. doi:10.1109/TVCG.2016.2598589. 2
- [QRD*20] QIAN X., ROSSI R. A., DU F., KIM S., KOH E., MALIK S., LEE T. Y., CHAN J.: ML-based visualization recommendation: Learning to recommend visualizations from data. *arXiv preprint arXiv:2009.12316* (2020). 1
- [Rai20] RAI A.: Explainable ai: From black box to glass box. *Journal of the Academy of Marketing Science* 48, 1 (2020), 137–141. 2
- [SS04] SEO J., SHNEIDERMAN B.: A rank-by-feature framework for unsupervised multidimensional data exploration using low dimensional projections. In *IEEE Symposium on Information Visualization (InfoVis)* (2004), IEEE, pp. 65–72. doi:10.1109/INFVIS.2004.3. 2
- [VG19] VERMA M., GANGULY D.: Lirme: locally interpretable ranking model explanation. In *ACM SIGIR Conference on Research and Development in Information Retrieval* (2019), pp. 1281–1284. 2
- [VHS*17] VARTAK M., HUANG S., SIDDIQUI T., MADDEN S., PARAMESWARAN A.: Towards visualization recommendation systems. *Acm Sigmod Record* 45, 4 (2017), 34–39. 1
- [WDC*18] WALL E., DAS S., CHAWLA R., KALIDINDI B., BROWN E. T., ENDERT A.: Podium: Ranking data using mixed-initiative visual analytics. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 24, 1 (2018), 288–297. doi:10.1109/TVCG.2017.2745078. 2
- [WWP*20] WRIGHT A. P., WANG Z. J., PARK H., GUO G., SPERRLE F., EL-ASSADY M., ENDERT A., KEIM D., CHAU D. H.: A comparative analysis of industry human-ai interaction guidelines. *arXiv preprint arXiv:2010.11761* (2020). 2
- [XMT*20] XIE T., MA Y., TONG H., THAI M. T., MACIEJEWSKI R.: Auditing the sensitivity of graph-based ranking with visual analytics. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 27, 2 (2020), 1459–1469. 1, 2
- [YSJ19] YU T., SHEN Y., JIN H.: A visual dialog augmented interactive recommender system. In *ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (2019), pp. 157–165. 1
- [ZSHL18] ZHOU Z., SHI C., HU M., LIU Y.: Visual ranking of academic influence via paper citation. *Journal of Visual Languages & Computing* 48 (2018), 134–143. doi:10.1016/j.jvlc.2018.08.007. 2
- [ZWB*21] ZHUANG H., WANG X., BENDERSKY M., GRUSHETSKY A., WU Y., MITRICHEV P., STERLING E., BELL N., RAVINA W., QIAN H.: Interpretable ranking with generalized additive models. In *ACM Web Search and Data Mining* (2021), pp. 499–507. 1, 2