Unofficial template Master SSE Thesis

by

Studentname Middlename Lastname

GRADUATION REPORT

Submitted to
Hanze University of Applied Sciences

in partial fulfillment of the requirements
for the degree of

Master Smart Systems Engineering

Confidential document

Groningen

2024

ABSTRACT

Insert your abstract here. Original Hanze format states one can use a maximum of 200 words. Though, it is advised is to discuss with your supervisor(s) if you need more.

Keywords: Keyword, Keyword, Keyword.

Contents

Chapter 1

Literature Review

1.1 Explainable AI In Knowledge Graph Recommender Systesm

Knowledge Graphs (KGs) are pivotal in enhancing the explainability and accuracy of recommender systems. These structured, relational frameworks capture complex interactions among users, items, and their attributes, allowing for more nuanced recommendations coupled with clear, logical explanations. This literature review synthesizes recent advancements in explainable artificial intelligence (XAI) that utilize KGs to illustrate how these technologies not only refine recommendation quality but also enhance user trust and understanding through transparency.

1.1.1 Path-based Modeling

Path-based modeling has emerged as a fundamental innovation in the utilization of KGs for recommender systems. Techniques such as the Knowledge-aware Path Recurrent Network (KPRN) and Path Language Modeling Recommendation (PLM-Rec) illustrate this trend's dynamic nature. KPRN leverages LSTM networks to interpret paths of entities and relationships, emphasizing those connections that are most influential in understanding user preferences. This method enriches the recommendation process by providing a temporal and semantic depth that traditional models lack, allowing for a better prediction of user behavior based on past interactions (Wang et al., 2019). On the other hand, PLM-Rec employs a novel approach by integrating natural language processing techniques to extend KG paths. This model treats paths as sentences, using a language model to dynamically predict and extend these paths within the KG. Such extensions help the system explore new, potentially uncharted areas of the KG, thereby enhancing the system's ability to recommend items that were previously unreachable. This approach addresses the inherent limitations of static KG structures and improves the system's recall capabilities, making it particularly valuable for discovering long-tail items (Geng et al., 2022). Together, these path-based methods signify a shift towards more dynamic and exploratory use of KGs, expanding both the depth and breadth of what recommender systems can achieve.

1.1.2 Integration of User Profiles and Behavior

The integration of user behavioral data into KGs has significantly refined the personalization capabilities of recommender systems. The "Cafe" model by Xian et al. (2020) represents a sophisticated application of this concept, employing a coarse-to-fine strategy where initially broad user profiles help to narrow down and guide the path-finding algorithms in KGs. These profiles are crafted from historical data and are instrumental in focusing the recommendation process on paths most relevant to individual users, thus enhancing both the relevance and personalization of the recommendations. This method mirrors strategies used in other models that combine knowledge-base embeddings (KBE) with collaborative filtering. By embedding user behaviors and item characteristics into a unified representation, these models achieve a granular understanding of user-item relationships. This integration allows for a tailored recommendation experience, where the system's outputs are closely aligned with individual preferences and behaviors, as demonstrated in the work by Ai et al. (2018).

1.1.3 Explainability through Symbolic and Logical Reasoning

The demand for explainability in AI has driven the adoption of models that incorporate transparent, logical reasoning processes. Monotonic GNNs (MGNNs), introduced by Tena et al. (2022), exemplify this trend by ensuring that every transformation within the network adheres to a set of logical rules, akin to traditional rule-based systems. This adherence guarantees that the network's operations are interpretable and justifiable, enhancing user trust by providing comprehensible explanations for the recommendations made. Similarly, the Policy-Guided Path Reasoning (PGPR) model uses reinforcement learning to navigate through the KG, selecting paths that not only lead to relevant recommendations but are also interpretable. This model provides explicit paths that detail the reasoning behind each recommendation, fulfilling the dual requirements of accuracy and transparency in the recommendation process (Xian et al., 2019).

The convergence of these methodologies highlights a crucial trend towards enhancing both the predictive accuracy and the interpretability of KG-based systems. Through the integration of dynamic path exploration, personalized user profile analysis, and logical reasoning, these approaches offer a more profound understanding of the intricacies involved in making recommendations. They collectively emphasize a shift towards recommender systems that are not only effective in their predictions but also provide transparent and understandable explanations, aligning with the growing user demand for transparency and accountability in AI systems.

1.2 Counterfactual Methods in Different Types of Recommender Systems

Counterfactual reasoning in recommender systems has emerged as a pivotal technique within the domain of explainable artificial intelligence (XAI), enhancing both the transparency and fairness of recommendations. By modeling alternative scenarios where specific variables are modified, this approach provides insights into the potential impacts of different data configurations, helping to elucidate the inner workings and dependencies within these systems.

The introduction of the KGCR model (Y. Wei et al., 2023) marks a significant advancement in embedding causal inference within graph-based recommender systems. Utilizing Graph Convolutional Networks, this model enriches user, item, and attribute embeddings, which allow for a more nuanced understanding of user preferences. By constructing a causal graph and applying do-calculus interventions, the KGCR model effectively mitigates biases introduced by previous user interactions, offering a refined approach to understanding how bias influences recommendation outcomes.

In a similar vein, Tran et al. (2021) developed the ACCENT framework, which facilitates the generation of actionable counterfactual explanations in neural recommender systems. This framework leverages extended influence functions to explore how changes in user-item interactions could affect recommendation outputs, significantly enhancing computational efficiency through Fast Influence Analysis. This methodology underscores the minimal adjustments in user behavior that could lead to different recommendations, thereby aiding in the creation of more transparent recommendation mechanisms.

Addressing selection bias, (Liu et al., 2022) implemented counterfactual policy learning to recalibrate recommendation fairness and effectiveness. Their approach utilizes Inverse Propensity Scoring to weigh observed interactions, allowing the system to simulate outcomes under different recommendation policies. By integrating these counterfactual outcomes into the learning process, the model achieves an improved balance, enhancing both the performance and equity of recommendations across various user groups and item categories.

The Prince method (Ghazimatin et al., 2020), emphasizes the importance of trust and understanding in recommendation systems through counterfactual reasoning within heterogeneous information networks. By identifying key user actions and employing Personalized PageRank, Prince efficiently predicts the impact of these actions on recommendation outcomes. This approach not only avoids exhaustive computations but also outperforms traditional heuristic methods in providing understandable and trust-enhancing explanations.

Yang et al. (2021) utilize causal inference through Structural Equation Models (SEMs) to address data sparsity in recommender systems. By generating counterfactual training samples, they enrich the dataset with diverse user responses that are otherwise not observed but plausible. This approach not only enhances the performance of the recommender systems but also strengthens their capacity to handle scenarios marked by data imbalance.

Finally, the Counterfactual Explainable Recommendation (CountER) model (Tan et al., 2021) focuses on identifying minimal attribute changes that could reverse a recommendation decision. Through a structured optimization process, CountER iteratively adjusts item attributes to discover the least extensive yet impactful changes required for altering outcomes. This model utilizes novel metrics to evaluate the necessity and sufficiency of these changes, demonstrating enhanced precision in providing actionable insights into recommendation decisions.

In conclusion, counterfactual reasoning offers a robust framework for enhancing the explainability and fairness of recommender systems by providing a deeper understanding of the implications of various data interactions and policies. These innovative approaches not only clarify the decision-making processes but also foster more equitable and user-centric recommendation practices.

1.3 Application of Counterfactuals for Fairness and Bias Mitigation

Counterfactual reasoning plays a pivotal role in the domain of explainable artificial intelligence (XAI), especially for mitigating biases in automated decision-making systems. This method involves hypothesizing alternative scenarios where key variables are altered, allowing for the exploration of how such changes impact outcomes. This not only uncovers hidden biases but also ensures fairness in AI operations. Broadly applied in various AI frameworks, from graph neural networks to recommender systems, counterfactual reasoning enhances transparency and equity in AI outcomes, establishing it as an essential tool for ethical AI development.

1.3.1 Mitigating Bias Across Different AI Frameworks

The use of counterfactual reasoning in graph-based models like those studied Guo et al. (2023) demonstrates a rigorous approach to maintaining consistency in model predictions across varying sensitive attributes. By implementing Graph Variational Autoencoders (GraphVAE), they not only perturb attributes but also train the network to minimize discrepancies in outputs between the original and counterfactual nodes. This methodology effectively addresses biases at a fundamental level, ensuring the fairness of the model's outcomes. Medda et al. (2024) extend this approach within graph neural network-based recommender systems. Their innovative use of counterfactual reasoning to adjust user-item interactions on a bipartite graph includes strategically adding or removing connections, which serves to simulate various scenarios where demographic disparities can be analyzed and mitigated, ensuring a more equitable distribution of utility among users. The field of recommender systems frequently grapples with biases such as popularity and exposure, which can distort user preferences. T. Wei et al. (2021) dissect these issues through the Model-Agnostic Counterfactual Reasoning (MACR) framework, which explicitly separates the influence of item popularity from actual user preferences. By adjusting input data to simulate a scenario where item popularity is neutralized, MACR provides a recalibrated basis for recommendation, aligning more closely with unbiased user preferences. Meanwhile, Xu et al. (2020) focus on exposure bias by employing a counterfactual approach that involves a minimax adversarial model. This model simulates worst-case scenarios to test the resilience of the recommendation system, ensuring that it can withstand and adapt to a range of user exposure conditions, thus promoting a more fair and balanced recommendation landscape.

1.3.2 Enriching Data and Ensuring Equitable Outcomes

Addressing data sparsity and imbalance, Yang et al. (2021) utilize causal inference via Structural Equation Models (SEMs) to generate counterfactual scenarios that enrich training datasets. This not only addresses the immediate issue of insufficient data but also simulates a broader spectrum of user interactions, which helps in developing a more robust and responsive recommender system. On a more focused level, Chiappa (2019) pioneers the use of Path-Specific Counterfactual Fairness (PSCF) within decision-making processes. This approach manipulates causal pathways, particularly those that might be influenced by sensitive attributes such as race or gender, to ensure that resulting decisions are free from the undue influence of these attributes, thus promoting fairness in critical decision-making contexts.

1.3.3 Leveraging Knowledge Graphs for Fair Recommendations

Expanding the utility of knowledge graphs, Balloccu et al. (2022) integrate counterfactual reasoning within the Policy-Guided Path Reasoning (PGPR) model to optimize recommendation systems. By re-ranking items and explanations based on various fairness-oriented criteria, such as recency, popularity, and diversity, PGPR enhances the quality and equity of recommendations. This approach not only improves the relevance of the recommendations but also significantly increases user trust and satisfaction by ensuring that recommendations cater equitably to diverse user groups.

Chapter 2

Methodology

2.1 Research Design and Methodology

2.2 Research Design

The primary objective of this research is to enhance the explainability of recommendations through counterfactual analysis within an existing knowledge graph-based recommendation system. This study integrates counterfactual analysis into a pre-existing recommender system as described in the literature. This framework's unique aspect is that it is adaptable to any knowledge graph-based recommender system, provided that the system supports path-based recommendations. This method contributes to the broader field of recommendation systems research by providing deeper insights into the causality and reasoning behind recommendations. The theoretical basis for employing counterfactual analysis in recommendation systems lies in its ability to generate "what-if" scenarios, thereby enhancing the transparency and explainability of recommendations. The human cognitive process inherently seeks counterfactual scenarios for enhanced understanding. Relevant literature and theories support this approach by demonstrating how counterfactuals can reveal hidden patterns and dependencies in data, thereby making recommendations more understandable and actionable for users.

2.3 Data Collection and Preparation

For this study, we utilize the implementation of the knowledge graph-based recommender system known as CAFÉ (Coarse-to-Fine Neural Symbolic Reasoning for Explainable Recommendation). CAFÉ integrates knowledge graphs with neural symbolic reasoning to enhance e-commerce recommendations. It operates on a coarse-to-fine principle, initially creating user profiles from historical data that outline general user behaviors. These profiles then guide the system in navigating the KG to generate reasoning paths that lead to specific item recommendations. This method not only improves recommendation accuracy but also provides clear explanations for why items are recommended by tracing the reasoning paths in the KG.

2.4 Knowledge Graph Construction

The knowledge graph is constructed using data from the existing recommender system, which includes entities such as users, products, brands, and words used in reviews, as well as related products that have been bought together, viewed, or also bought. Relationships like 'purchase', 'mentions', 'produced by', and 'described by' are included. This process involves: • Representing the knowledge graph as a dictionary of entities and relationships, and including the reverse of each relationship. • Connecting each user to the words they have mentioned and products they have purchased. Products are linked to the brands they are produced by, the categories they belong to, the words they are described by, and related products.

2.5 Metapaths

Metapaths refer to predefined paths in the knowledge graph that represent sequences of relations connecting different entities (like users and items). These paths are structured to reflect potential behavioral or relational patterns between entities relevant to making recommendations. The reasoning process using these metapaths operates as follows: The system first generates a coarse sketch of user behavior by identifying prominent user-centric patterns from historical data. These patterns, essentially metapaths, highlight typical ways users interact with items or other entities in the knowledge graph. Using the metapaths outlined in the user profiles as a guide, the system then performs fine-grained path reasoning. This involves a path-finding algorithm that navigates through the knowledge graph, starting from a user entity and moving through the specified entities and relations. The goal is to reach item entities that best match the user's profiled behavior. This path reasoning is augmented by neural symbolic reasoning modules, which assist in making decisions at each step of the path (like choosing which relation to follow next based on the current context). This combination of structured metapaths and neural decision-making enables the system to efficiently and effectively find relevant items, offering explanations for recommendations based on the sequences of relations traversed.

2.6 Recommender System Integration

To integrate the knowledge graph into the framework, I have extended the implementation of CAFÉ's knowledge graph to enable the handling of a subgraph of the original, very large system. Sampling is based on the number of users and is performed using a snowball strategy, sampling connected products and their attributes. A pruning step follows this, which eliminates all entities that are not in a complete relationship with the sampled knowledge graph. For example, entities where only one side of a relationship is sampled, and the reverse does not exist, are removed. Since the CAFÉ recommender system uses predefined embeddings, after pruning, the conversion of IDs from old to new is performed. This revised version improves clarity, flow, and academic tone, and structurally it emphasizes a logical progression from general objectives to specific methodologies and integration steps.

2.7 Framework Design

The counterfactual analysis begins by examining the top 10 recommendations provided by the recommender system. For a given recommended path—which comprises entities and relationships defined by the recommender system's predefined metapaths—this path serves as the input for the counterfactual analysis. The analysis focuses on first-level attributes and entities that are connected to the attributes and related products of the recommended path, ensuring that the selections are highly relevant to the product under consideration.

The recommender system is designed to learn the tastes and behaviors of the user based on their purchased products, which is reflected in the design of the metapaths. After identifying related entities and attributes, a specific metapath associated with each entity is selected for analysis. For example, to explore whether a product would still be recommended if it were associated with a different brand, the metapath might be: user → purchase → product → produced_by → brand → recommended product. The viability of this scenario is assessed by calculating the average score of the steps along the path; if this score exceeds the score of the least recommended product (the 10th product), the scenario is considered plausible.

2.8 Knowledge Graph Entity Stats

Some nodes in the knowledge graph, such as those connected to a "beauty" category, exhibit high connectivity degrees but provide limited specific insight for the analysis due to their general nature. These nodes can create computational overhead. To address this, we calculate the connectivity degree for each entity across all related entity types. Nodes with z-scores exceeding a specified threshold (typically set at 1) are identified for potential exclusion from the analysis to streamline computations.

2.9 Community Identification

To address the computational overhead of some selected entities, for example words that are relatively higher in number connected to products, the study furthers the filtering process. If the number of selected entities exceeds a specified threshold, filtering occurs based on their community membership relative to the recommended product. To identify these communities within the knowledge graph, we employ the Louvain method. This method optimizes modularity, effectively grouping nodes into communities where connections are denser internally than with external nodes. The iterative process starts with each node as its own community and progressively merges them to maximize modularity. This is particularly beneficial for your counterfactual analysis as it helps identify clusters of products with similar attributes or relationships. Understanding these community structures allows you to analyze how alterations in product attributes might impact recommendation patterns, providing deeper insights into the factors that influence product categorization and recommendations.

References

Ai, Q., Azizi, V., Chen, X., & Zhang, Y. (2018). Learning heterogeneous knowledge base embeddings for explainable recommendation. *Algorithms*, *11*(9), 137. https://doi.org/10.3390/a11090137

Balloccu, G., Boratto, L., Fenu, G., & Marras, M. (2022). Post processing recommender systems with knowledge graphs for recency, popularity, and diversity of explanations. *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 646–656. https://doi.org/10.1145/3477495.3532041

Chiappa, S. (2019). Path-specific counterfactual fairness. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(1), 7801–7808. https://doi.org/10.1609/aaai.v33i01.33017801

Geng, S., Fu, Z., Tan, J., Ge, Y., De Melo, G., & Zhang, Y. (2022). Path language modeling over knowledge graphsfor explainable recommendation. *Proceedings of the ACM Web Conference 2022*, 946–955. https://doi.org/10.1145/3485447.3511937

Ghazimatin, A., Balalau, O., Saha Roy, R., & Weikum, G. (2020). PRINCE: Provider-side interpretability with counterfactual explanations in recommender systems. *Proceedings of the 13th International Conference on Web Search and Data Mining*, 196–204. https://doi.org/10.1145/3336191.3371824

Guo, Z., Li, J., Xiao, T., Ma, Y., & Wang, S. (2023). Towards fair graph neural networks via graph counterfactual. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 669–678. https://doi.org/10.1145/3583780.3615092

Liu, Y., Yen, J.-N., Yuan, B., Shi, R., Yan, P., & Lin, C.-J. (2022). Practical counterfactual policy learning for top-k recommendations. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1141–1151. https://doi.org/10.1145/3534678.3539295

Medda, G., Fabbri, F., Marras, M., Boratto, L., & Fenu, G. (2024). GNNUERS: Fairness explanation in GNNs for recommendation via counterfactual reasoning. *ACM Transactions on Intelligent Systems and Technology*, 3655631. https://doi.org/10.1145/3655631

Tan, J., Xu, S., Ge, Y., Li, Y., Chen, X., & Zhang, Y. (2021). Counterfactual explainable recommendation. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 1784–1793. https://doi.org/10.1145/3459637.3482420

Tran, K. H., Ghazimatin, A., & Saha Roy, R. (2021). Counterfactual explanations for neural recommenders. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1627–1631. https://doi.org/10.1145/3404835.3463005

Wang, X., Wang, D., Xu, C., He, X., Cao, Y., & Chua, T.-S. (2019). Explainable reasoning over knowledge graphs for recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(1), 5329–5336. https://doi.org/10.1609/aaai.v33i01.33015329

Wei, T., Feng, F., Chen, J., Wu, Z., Yi, J., & He, X. (2021). Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 1791–1800. https://doi.org/10.1145/3447548.3467289

Wei, Y., Wang, X., Nie, L., Li, S., Wang, D., & Chua, T.-S. (2023). Causal inference for knowledge graph based recommendation. *IEEE Transactions on Knowledge and Data Engineering*, *35*(11), 11153–11164. https://doi.org/10.1109/TKDE.2022.3231352

Xian, Y., Fu, Z., Muthukrishnan, S., De Melo, G., & Zhang, Y. (2019). Reinforcement knowledge graph reasoning for explainable recommendation. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 285–294. https://doi.org/10.1145/3331184.3331203

Xian, Y., Fu, Z., Zhao, H., Ge, Y., Chen, X., Huang, Q., Geng, S., Qin, Z., De Melo, G., Muthukrishnan, S., & Zhang, Y. (2020). CAFE: Coarse-to-fine neural symbolic reasoning for explainable recommendation. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 1645–1654. https://doi.org/10.1145/3340531.3412038

Xu, D., Ruan, C., Korpeoglu, E., Kumar, S., & Achan, K. (2020). Adversarial counterfactual learning and evaluation for recommender system [Publisher: arXiv Version Number: 1]. https://doi.org/10.48550/ARXIV.2012.02295

Yang, M., Dai, Q., Dong, Z., Chen, X., He, X., & Wang, J. (2021). Top-n recommendation with counterfactual user preference simulation [Publisher: [object Object] Version Number: 2]. https://doi.org/10.48550/ARXIV.2109.02444

Appendix

You are encouraged to put in appendices in your final report. In an appendix you can include things such as large tables or background information. Anything that is useful to know for the reader, but prevents the reader to read your main text in a fluent manner. Each appendix should have a number and a self-explanatory title.