**1. Introduction to Counterfactual Analysis in Recommender Systems**

**2. Counterfactual Methods in Different Types of Recommender Systems**

Counterfactual reasoning has been employed in recommender systems, serving a variety of purposes including bias reduction and enhancing explainability. By simulating conditions where certain variables are altered, counterfactual reasoning helps illuminate how such changes could affect outcomes, thereby offering insights into the underlying mechanics of recommender systems. This capability is crucial not only for refining system accuracy but also for ensuring the fairness and transparency of the recommendations provided.

In their study, Wei et al. (2023) introduce the KGCR model, an innovative approach designed to counteract bias in graph-based recommender systems by embedding causal inference within knowledge graph structures. This model enhances the accuracy of reflecting true user preferences through the use of Graph Convolutional Networks (GCNs) to refine the embeddings of users, items, and attributes. These enriched embeddings allow for more contextually informed representations. To mitigate biases, especially those originating from prior user interactions with specific attributes, the KGCR model constructs a causal graph. Interventions using do-calculus are applied to 'cut' edges representing biased influences, thereby creating counterfactual scenarios where such biases are excluded. This adjustment enables the recalibration of similarity scores, assessing potential outcomes had the user not engaged with the biasing attributes.

Further advancing the field, Tran et al. (2021) developed the ACCENT framework, aimed at generating actionable and transparent counterfactual explanations within neural recommender systems. This framework emphasizes the influence of user-item interactions on recommendation outputs. By employing extended influence functions, the ACCENT framework assesses item pairs to understand how modifications in these interactions could alter model predictions. It utilizes Fast Influence Analysis (FIA) to efficiently compute the impact of individual data points, significantly reducing the computational demands typical of large neural networks. This process facilitates the identification of the minimal set of user actions that, if altered, could change the recommendation outcomes, implemented across both Neural Collaborative Filtering (NCF) and Relational Collaborative Filtering (RCF) systems.

**3. Techniques and Models for Generating Counterfactuals**

In the domain of graph-based knowledge systems, generating counterfactual scenarios poses challenges due to the complexity and interconnectedness of data elements. Techniques and models for generating counterfactual explanations are evolving to address these challenges, through different methods namely optimization-based and learning-based approaches.

The CLEAR framework represents an optimization-based approach, utilizing a graph variational autoencoder (VAE) to generate counterfactual explanations for graph-level prediction models. By encoding node features and graph structures into a latent space and decoding to construct minimal yet effective modifications in graph data, CLEAR significantly advances the generation of counterfactual explanations. This approach ensures optimization, generalizes across different graphs, and adheres to causal relationships.

On the other hand, the learning-based approach is exemplified by the Counterfactual Explainable Recommendation (CERec) system, which employs reinforcement learning to navigate and optimize paths within a knowledge graph. This system focuses on generating intuitive, attribute-based counterfactual explanations. With the help of an adaptive path sampler and a two-step attention mechanism, CERec efficiently narrows down the vast search space to the most promising counterfactual paths. These paths show how subtle changes to item attributes can lead to significant shifts in recommendation outcomes.

**4. Application of Counterfactuals for Fairness and Bias Mitigation**

Counterfactuals are increasingly utilized in fairness and bias mitigation strategies to address inherent biases within machine learning models, ensuring that predictions do not perpetuate or amplify discriminatory practices based on sensitive or irrelevant attributes. These methods allow systems to simulate 'what-if' scenarios that challenge the status quo, helping to identify and correct biases across various applications, from graph neural networks (GNNs) to recommender systems.

In the realm of graph-structured data, Guo et al. (2023) implement counterfactual reasoning within Graph Neural Networks (GNNs) to promote fairness. Their approach leverages a Graph Variational Autoencoder (GraphVAE), along with algorithms like NIFTY and GEAR, to modify sensitive node attributes and generate realistic counterfactual scenarios. These scenarios, produced through perturbations or through the GraphVAE's encoding and decoding processes, allow the model to disentangle sensitive attributes from other features. The GNNs are trained to minimize discrepancies between the outputs of original nodes and their counterfactual counterparts, ensuring that predictions remain consistent regardless of changes in sensitive attributes.

Similarly, Wei et al. (2021) address popularity bias in recommender systems through their Model-Agnostic Counterfactual Reasoning (MACR) framework. This framework utilizes a causal graph to map out and isolate the direct effects of item popularity on recommendation scores, employing multi-task learning for more precise measurements. Counterfactual scenarios are generated by simulating recommendation scores as if item popularity were absent, allowing the system to adjust actual scores by removing the estimated popularity influence. This method ensures that recommendations reflect genuine user preferences more accurately, mitigating the dominance of popular items.

Both approaches exemplify the versatility and efficacy of counterfactual reasoning in enhancing the fairness of machine learning models, each targeting specific biases within their respective frameworks to improve the overall fairness and reliability of their predictions. By adjusting the systems to respond to hypothetical, bias-free scenarios, these methods help create more equitable and just technological solutions.

**5. User-centric Approaches to Counterfactual Explanations**

**6. Challenges and Future Directions**

**7. Summary and Implications for Future Research**

Ai, Q., Azizi, V., Chen, X., & Zhang, Y. (2018a). Learning Heterogeneous Knowledge Base Embeddings for Explainable Recommendation. *Algorithms*, *11*(9), 137. <https://doi.org/10.3390/a11090137>

Ai, Q., Azizi, V., Chen, X., & Zhang, Y. (2018b). Learning Heterogeneous Knowledge Base Embeddings for Explainable Recommendation. *Algorithms*, *11*(9), 137. <https://doi.org/10.3390/a11090137>

Chen, F., Yin, G., Dong, Y., Li, G., & Zhang, W. (2023). KHGCN: Knowledge-Enhanced Recommendation with Hierarchical Graph Capsule Network. *Entropy*, *25*(4), 697. <https://doi.org/10.3390/e25040697>

Chen, H., Li, Y., Sun, X., Xu, G., & Yin, H. (2021). Temporal Meta-path Guided Explainable Recommendation. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 1056–1064. <https://doi.org/10.1145/3437963.3441762>

Chen, Z., Silvestri, F., Wang, J., Zhang, Y., & Tolomei, G. (2023). The Dark Side of Explanations: Poisoning Recommender Systems with Counterfactual Examples. *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2426–2430. <https://doi.org/10.1145/3539618.3592070>

Chiappa, S. (2019). Path-Specific Counterfactual Fairness. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 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, Q., Zhuang, F., Qin, C., Zhu, H., Xie, X., Xiong, H., & He, Q. (2020). A Survey on Knowledge Graph-Based Recommender Systems. *IEEE Transactions on Knowledge and Data Engineering*, *34*(8), 3549–3568. <https://doi.org/10.1109/TKDE.2020.3028705>

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>

Jaimini, U., & Sheth, A. (2022). CausalKG: Causal Knowledge Graph Explainability Using Interventional and Counterfactual Reasoning. *IEEE Internet Computing*, *26*(1), 43–50. <https://doi.org/10.1109/MIC.2021.3133551>

Li, D., Qu, H., & Wang, J. (2023). A Survey on Knowledge Graph-Based Recommender Systems. *2023 China Automation Congress (CAC)*, 2925–2930. <https://doi.org/10.1109/CAC59555.2023.10450693>

Liu, J., & Duan, L. (2021). A Survey on Knowledge Graph-Based Recommender Systems. *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2450–2453. <https://doi.org/10.1109/IAEAC50856.2021.9390863>

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>

Ma, J., Guo, R., Mishra, S., Zhang, A., & Li, J. (2022). *CLEAR: Generative Counterfactual Explanations on Graphs*. <https://doi.org/10.48550/ARXIV.2210.08443>

Ma, T., Lv, S., Huang, L., & Hu, S. (2021). HiAM: A Hierarchical Attention based Model for knowledge graph multi-hop reasoning. *Neural Networks*, *143*, 261–270. <https://doi.org/10.1016/j.neunet.2021.06.008>

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>

Saito, Y., & Joachims, T. (2021). Counterfactual Learning and Evaluation for Recommender Systems: Foundations, Implementations, and Recent Advances. *Fifteenth ACM Conference on Recommender Systems*, 828–830. <https://doi.org/10.1145/3460231.3473320>

Shang, R., Feng, K. J. K., & Shah, C. (2022). Why Am I Not Seeing It? Understanding Users’ Needs for Counterfactual Explanations in Everyday Recommendations. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 1330–1340. <https://doi.org/10.1145/3531146.3533189>

Stepin, I., Alonso, J. M., Catala, A., & Pereira-Farina, M. (2021). A Survey of Contrastive and Counterfactual Explanation Generation Methods for Explainable Artificial Intelligence. *IEEE Access*, *9*, 11974–12001. <https://doi.org/10.1109/ACCESS.2021.3051315>

Tan, J., Geng, S., Fu, Z., Ge, Y., Xu, S., Li, Y., & Zhang, Y. (2022). Learning and Evaluating Graph Neural Network Explanations based on Counterfactual and Factual Reasoning. *Proceedings of the ACM Web Conference 2022*, 1018–1027. <https://doi.org/10.1145/3485447.3511948>

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>

Tian, L., Zhou, X., Wu, Y.-P., Zhou, W.-T., Zhang, J.-H., & Zhang, T.-S. (2022). Knowledge graph and knowledge reasoning: A systematic review. *Journal of Electronic Science and Technology*, *20*(2), 100159. <https://doi.org/10.1016/j.jnlest.2022.100159>

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., He, X., Feng, F., Nie, L., & Chua, T.-S. (2018). TEM: Tree-enhanced Embedding Model for Explainable Recommendation. *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW ’18*, 1543–1552. <https://doi.org/10.1145/3178876.3186066>

Wang, X., Li, Q., Yu, D., Li, Q., & Xu, G. (2024). Reinforced Path Reasoning for Counterfactual Explainable Recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 1–17. <https://doi.org/10.1109/TKDE.2024.3354077>

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*(01), 5329–5336. <https://doi.org/10.1609/aaai.v33i01.33015329>

Wang, Z., Li, L., Zeng, D., & Wu, X. (2021). Incorporating prior knowledge from counterfactuals into knowledge graph reasoning. *Knowledge-Based Systems*, *223*, 107035. <https://doi.org/10.1016/j.knosys.2021.107035>

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*. <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*. <https://doi.org/10.48550/ARXIV.2109.02444>

Zellinger, L., Stephan, A., & Roth, B. (2024). *Counterfactual Reasoning with Knowledge Graph Embeddings*. <https://doi.org/10.48550/ARXIV.2403.06936>

Zhang, Y., & Chen, X. (2020). Explainable Recommendation: A Survey and New Perspectives. *Foundations and Trends® in Information Retrieval*, *14*(1), 1–101. <https://doi.org/10.1561/1500000066>