

Research on the Self Evolution Construction and Interpretable Recommendation of Cross Domain Knowledge Graph for E-Commerce Products Based on Multimodal Large Language Model

Li Ma, Yuanli Cui

Department of Information and Business, Jinan Preschool Education College, Jinan 250000, Shandong, China

Corresponding author: Yuanli Cui (e-mail: Cxueshu765@163.com)

ABSTRACT Cross-domain e-commerce recommendation faces challenges from multimodal product heterogeneity, sparse intercategory associations, and opaque recommendation reasoning. To improve accuracy and interpretability, this study proposes a multimodal large-language-model-driven framework for self-evolving cross-domain product knowledge graphs and explainable recommendation. Product images, titles, and attributes from apparel, home-furnishing, and digital-product domains are encoded using a multimodal large model, mapped into a shared latent space, and aligned through contrastive learning for cross-domain entity and semantic association extraction. A self-evolution mechanism uses the large language model as a relation verifier and reasoning engine to validate, buffer, prune, or extend graph edges according to confidence and interaction feedback. A path-aware graph neural network then samples multi-hop cross-domain paths, encodes product sequences through gated recurrent units, and fuses graph representations with original multimodal embeddings. Recommendation explanations are generated from high-weight inference chains under loyalty constraints to ensure factual consistency. Experiments on a large e-commerce dataset show that the proposed method achieves Hits@10 of 0.892 and MRR of 0.537 for relation completion, while CTR, CVR, and Recall@20 reach 12.4%, 6.2%, and 22.1%, respectively. Explanation fidelity and perceived usefulness score 4.23 and 4.15. The framework supports semantic alignment, graph-based reasoning, and interpretable recommendation in multimodal information systems.

INDEX TERMS multimodal large language model, cross-domain knowledge graph, semantic alignment, graph neural network, explainable recommendation

I. INTRODUCTION

WITH the continuous expansion of e-commerce, platform product categories are becoming increasingly refined and diversified, and cross-category consumption behavior among users is becoming more and more common. However, current e-commerce recommendation systems generally face three structural dilemmas: First, product information is distributed across heterogeneous modalities such as images, text, and attributes, creating a natural barrier to cross-modal semantic alignment; second, product association knowledge between different categories is extremely sparse, making it difficult for traditional collaborative filtering methods to capture cross-domain complementary logic such as “hiking boots - portable oxygen cylinders”; third, the generation process of recommendation results lacks

transparency and traceability, leaving users unable to know the rational basis for “why this product is recommended,” severely restricting the trust building and decision-making efficiency of the recommendation system [1,2]. These problems collectively point to a core challenge: how to construct a recommendation mechanism in a dynamic and heterogeneous e-commerce environment that can both mine implicit product associations across category boundaries and provide users with credible explanations. Traditional methods either rely on shallow statistical co-occurrence of a single modality or use static knowledge graphs for relationship representation, neither of which can adequately address the dual requirements of continuous evolution of product relationships and deep coupling of explanation generation. To address the problem of cross-domain product relationship mining,

scholars have conducted extensive research on knowledge graph-enhanced recommendation. For example, Huang et al. [3] proposed a path-enhanced knowledge graph interpretable recommendation method, which improves the interpretability of recommendations by modeling the semantic path between users and items; Caro-Martínez et al. [4] reduced the dependence of interpretable recommendations on prior knowledge from the perspective of graph structure; Lin et al. [5] introduced a self-supervised reinforcement learning mechanism to realize the knowledge-aware recommendation reasoning process; Zhao et al. [6] combined graph neural networks and knowledge graphs to improve the interpretability and representation ability of recommendation systems; Fang et al. [7] and Lyu et al. [8] enhanced the rationality of explanation generation and user interaction ability from the perspectives of counterfactual reasoning and dialogue recommendation, respectively. However, the explanatory texts generated by the above methods are mostly derived from fixed template filling, lacking dynamic restoration and fluent expression of cross-domain reasoning links.

In recent years, the rise of multimodal large language models has provided a new technical paradigm for solving the above problems. For example, Wei et al. [9] and Yu et al. [10] systematically reviewed multimodal and cross-domain recommendation technologies from the perspectives of recommendation system development trends and deep learning modeling, respectively; Zhao et al. [11] proposed a cross-domain recommendation method based on structural alignment, which effectively improved the knowledge transfer capability between different domains. However, existing research focuses on the optimization of a single link and has not yet formed a closed-loop framework from automatic construction and dynamic evolution of cross-domain knowledge to interpretable recommendation generation, especially in the systematic embedding of the reasoning ability of large language models into knowledge graph lifecycle management. This paper is based on the research gap in the above-mentioned cross-domain fields and proposes an integrated method of cross-domain knowledge graph self-evolution and interpretable recommendation with multimodal large language models as the core engine, aiming to achieve adaptive mining of product relationships and transparent expression of recommendation decisions.

The main innovations of this paper are summarized in the following four points: First, we establish a framework of cross-domain product entity alignment and semantic association mining based on a multimodal large language model, which solves the performance bottleneck of traditional methods in heterogeneous modality alignment and cross-category relationship discovery; Second, we design an incremental knowledge graph self-evolution mechanism. Taking the large language model as a dynamic relation inference engine, we endow the knowledge graph with the ability of continuous adaptation and suppression of noises in new consumption scenarios; Third, we propose an interpretable recommendation strategy based on path-aware graph neural

networks and large language model. It can achieve end-to-end generation from graph inference paths to natural language interpretations while maintaining recommendation accuracy; Finally, we finish systematic offline evaluation and online verification on real multi-domain e-commerce datasets. The results demonstrate the effectiveness of our framework from three levels, i.e., recommendation performance, quality of knowledge evolution and credibility of interpretation.

II. MULTIMODAL CROSS-DOMAIN PRODUCT ENTITY ALIGNMENT AND COARSE-GRAINED SEMANTIC ASSOCIATION EXTRACTION

In order to realize uniform representation and initial association of product entities in cross-domain scenarios, including apparel, home furnishing and digital goods, we build a frozen multidimensional large-scale language model encoder (LLaVA-v1.6-34B). Images are resized to 336×336 pixels for the CLIP-ViT-L/14 visual tower, while text is tokenized into sequences of maximum length 512. Firstly, the main image visual information and textual information including title and attribute description are extracted for each domain product. On the visual part, we extract a fixed-dimensional image embedding vector with visual encoding branch of the multidimensional large-scale model for each product. On the text part, the product title and attributes will be concatenated into a natural language sequence and fed into text encoder of the multidimensional large-scale model to obtain semantic representation. For the cross-modal distribution difference, we design a learnable linear projection layer (with input dimension 4096 and output dimension 256) to map the embedding of two modes to the same latent space. We employ a InfoNCE-based contrastive learning objective function with a temperature parameter of 0.07 to narrow the representation distance of semantically similar cross-domain product pairs and push away the irrelevant samples. Further, we adopt a cue learning strategy to construct natural language template containing category context to guide the frozen multidimensional large-scale model to explore coarse-grained semantic association between cross-domain product entities automatically, including the association of co-occurrence, complement or functional substitution, and then obtain initial candidate entity alignment pairs. The whole process does not rely on manually predefined rules or domain ontologies, which greatly improves the automation and coverage of cross-domain association discovery, and finally use the association relationships generated by above process as the framework of subsequent incremental knowledge graph construction and self-evolving reasoning.

III. INCREMENTAL KNOWLEDGE GRAPH SELF-EVOLUTION MECHANISM BASED ON LARGE LANGUAGE MODEL REASONING

After obtaining cross-domain product entity alignment and coarse-grained association extraction, the first knowledge

graph still needs to evolve in two directions: First, new product launches and user behaviors on ecommerce platforms are constantly emerging, and the graph needs to continuously absorb new entities and relationships; Second, since the relationships in coarse-grained associations are multimodal, a certain amount of noise edges will inevitably appear in the downstream graph due to alignment errors, and need further screening and correction. Therefore, in this section, we design an incremental self-evolution mechanism based on multimodal large language model as the reasoning agent. The basic idea is to use the large language model as the verifier of relationship rationality. The LLM is prompted with a structured template: ‘Given Product A (attributes) and Product B (attributes), is the relationship R (type) commercially logical? Answer only with Valid, Questionable, or Invalid.’

Specifically, the evolution event is triggered by two types of events: First, new products are listed or product information changes significantly; Second, the co-purchase / browsing sequence accumulated in user interaction log reaches the preset threshold. After triggering, on one hand, we extract the multimodal description information of the product to be evolved and the existing graph context of candidate neighbor nodes, and construct the structured prompt text containing product attribute information / category characteristics and existing relationship edge information as input of the multimodal large language model; On the other hand, we utilize the multimodal large language model as a verifier, augmented by the low-rank adaptive feedback fine-tuning module to instill e-commerce domain-specific logic. The model outputs a three-value judgment of ‘valid’, ‘questionable’, or ‘invalid’. To ensure rigorous data retention, only edges with a ‘valid’ output and a model-generated confidence score exceeding a threshold $\theta = 0.85$ are integrated into the primary graph, while ‘questionable’ edges are subjected to a secondary validation via the interaction log frequency check, and this three-value judgment is operationalized by filtering out ‘Invalid’ edges and placing ‘Questionable’ ones in a buffer. Quantitative evaluation shows the LLM achieved a 91.4% verification accuracy against three human experts. The edges judged as “questionable” are temporarily not included in the formal graph, but moved to the observation buffer pool; The edges that are obviously judged as “invalid” are removed. After completing relationship verification, we further use the thought chain reasoning ability of large language model to complete cross-domain multi-hop paths. For first-order direct associations that are passed verification, we construct the path expansion hint containing the product entity as the starting point to guide the model to infer the second order implicit association even the third order implicit association under the condition that the parameters are frozen. For example, if the complementary relationship between ‘hiking shoes’ and ‘hiking poles’ is known, a constrained scenario co-occurrence between ‘hiking shoes’ and ‘outdoor insurance’ can be inferred. To prevent ‘semantic explosion’

and maintain recommendation relevance, we apply a decay-based path pruning mechanism that restricts the maximum inference depth to 3 hops and enforces a ‘category-affinity’ constraint, ensuring that the inferred nodes remain within the semantic neighborhood of the user’s core intent. The inference paths generated by the model are returned as sequence of triples. After deduplication and confidence ranking, the paths with value will be added to the candidate set of new relationships returned.

Finally, the graph structure is updated. Due to the large fluctuation of graph states caused by batch updating graph, this mechanism uses sliding window strategy to submit new edges and deleted edges in batch, and records the source identifier (direct extraction, model inference, manual verification) and confidence score of relationship extraction.

To alleviate concept drift and error accumulation in evolution, we design a low-rank adaptive feedback fine-tuning module: Periodically sample recommendation links with good feedback from recent interaction data, and then retrieve corresponding relationship paths in graph in reverse. Take these as positive sample to adjust some adapter parameters of inference module of large language model’s adaptation, so as to improve the discrimination sensitivity to e-commerce domain specific logic of large language model continuously. The main body of model will be frozen. Through this mechanism, the graph achieves automatic growth and self-healing of noise, evidenced by a 15.2% reduction in false-positive edges compared to initial extraction.

IV. DEEP FEATURE FUSION OF PATH-AWARE GRAPH NEURAL NETWORK AND CROSS-MODAL SEMANTICS

The last two sections have finished the cross-domain product entity alignment and self-evolutionary construction of knowledge graph, which is a graph containing rich cross-category product association structures. However, how to aggregate the discrete graph relationships into continuous vector representations which can be directly utilized by recommendation model, and fully preserve the information of cross-domain semantic paths in representation learning which is critical to the recommendation accuracy. While traditional graph neural networks can well perform neighbor node information aggregation, there are two obvious limitations when applying them in cross-domain scenario: first, multi-hop semantic paths would inevitably experience information decay when performing layer-by-layer convolution, and cross-category association signals would decrease rapidly when hop length enlarges; second, multimodal visual-text semantics would be reduced to single scalar weights during graph aggregation process, and simple deep fusion with graph structure cannot be achieved. Therefore, in this section we design a path-aware graph neural network which utilizes cross-modal semantics as a priori constraint for edge weight modelings and explicitly encode multi-hop association paths between product nodes.

At the model architecture level, we first construct a

heterogeneous product relationship graph based on evolved knowledge graph, where nodes represent product entities from different categories and edge types include coarse-grained association information such as complementarity, substitution and scene co-occurrence, as well as inference relationships completed by self-evolution. Different from traditional graph convolutional networks, for any product association edge, the cosine similarity between the two end products in visual-text joint embedding space would be re-calculated by the cosine similarity in multimodal large-model encoder frozen in Section 2.1, and then normalized as an initial semantic weight of this edge. Therefore, the graph topology not only reflects relationship type between two end nodes, but also contains fine-grained cross-modal content affinity information between two end products. For any target product node, instead of aggregating directly information of its first-order neighbors, multiple cross-domain paths of a maximum length of 3 hops would be sampled (up to 10 paths per node using random walks) from this target node, and each path contains several continuous relationship edges, which represents an potential cross-category inference chain. For any sampled path, gated recurrent unit would encode the feature sequences of each node on this path, and utilize the hidden state at the endpoint of this path as aggregated representation of this path. Subsequently, a path-level self-attention mechanism would be introduced to weightedly fuse different path representations, which allows the model to automatically select the most discriminative cross-domain inference chain for current recommendation task. For example, when user is going to view or purchase portable oxygen cylinders, the model would attend more attention to path hiking boots $\rightarrow$ outdoor gear $\rightarrow$ high-altitude travel supplies.

Regarding cross-modal feature fusion, the graph structure representation extracted from above path aggregation will be interacted with its original multimodal embedding representations of products in a collaborative way. In detail, we will add the path-enhanced node representation to the initial product embedding representation directly output by large multimodal model and then apply layer normalization on them before extracting them into feedforward network for further nonlinear transformation. To enhance the guiding effect of cross-modal semantics for deep propagation more, we will also add one contrastive auxiliary loss term in model training to enforce the structural representation of same product extracted from graph aggregation to be proximal relationship with its original multimodal embedding representation in vector space. Thus, the semantic drift problem happened in deep propagation of graph neural networks is alleviated greatly. Finally, through multi-layer path-aware aggregation and cross-modal fusion, the final embedded representation of each product node will integrate abundant local neighbor information, multi-hop inference paths and multimodal content semantics together. A single-layer gated recurrent unit (hidden size 256) would encode the feature sequences. The GNN component contains 1.2M trainable

parameters and requires 42 minutes for a full training epoch on the 126,820-node graph. The computational complexity scales linearly with the number of sampled paths, ensuring the final embedded representation. And we will directly apply them in following recommendation retrieval and ranking stage, while the representation consisting of product structured representation will also be utilized as the explainable recommendation module input in a visualizable way, and meanwhile, we will also provide vectorized support for generating natural language explanations.

V. CONTEXT-EXPLAINABLE RECOMMENDATION GENERATION BASED ON GRAPH INFERENCE CHAINS

After the aforementioned modules complete cross-domain product representation learning and recommendation ranking, the underlying logic of the recommendation results needs to be presented to the user in natural language to achieve truly explainable recommendations. Unlike traditional methods based on template filling or attention heatmaps, this paper fully utilizes the explicitly stored cross-domain inference paths in the knowledge graph to design an end-to-end generation mechanism from the graph inference chain to fluent explanatory text.

Its specific process is as follows: Firstly, the high-weight cross-domain association paths triggered in the recommendation stage are mined out based on recommended products to be explained (i.e. the products that users have already interacted with and the products that are recommended by systems). Then, in the path-aware graph neural network computation, several inference paths with higher weights will be automatically selected out by the attention mechanism, and the path with the highest weight is taken as the main explanation framework. Finally, all nodes and relational edges traversed in this path are gradually expanded into a sequence of triples, e.g. “hiking boots—[scenario suitability]—outdoor hiking—[equipment combination] —trekking poles”. This sequence fully records the whole logical process of explanation from the historical product consumption of users across categories to the recommended product.

Subsequently, this sequence of inference chain is further converted into a prompt context that can be understood by the large language model. Its specific template contains three parts: the task description section, namely, explicitly requiring the large language model to generate consumer-oriented explanatory text based on the explained inference path; the path fact section, namely, stating the triple sequence in natural language, e.g. “The hiking boots you viewed are suitable for outdoor hiking scenarios, which are often used with trekking poles”; the style constraint section, namely, requiring the generated explanatory text to be concise and natural, and highlighting the actual usage relationship between products. The whole prompt is fed into the multimodal large language model decoder. To bridge the gap between offline generation and real-time high-concurrency requirements, we implement a two-stage acceleration strategy: first, frequently

triggered inference chains are pre-computed and stored in a low-latency key-value cache; second, for dynamic requests, we employ 4-bit quantization and KV-cache optimization to reduce the latency of the forward inference to sub-100ms levels, allowing the model to generate the final explanation statement word by word without compromising system throughput. To further ensure the explanation facts generated by the model appear in the recommended facts, the loyalty constraint mechanism is introduced. When generating, the decoder is forced to directly output the key entity names (product category, brand, core attributes, etc.) appearing in the inference chain as they are, avoiding the misattribution phenomenon caused by model illusion. Meanwhile, the text generated by the decoder is also validated by rule post-processing module to ensure all product features appearing in explanation can be found in the multimodal description field of input products.

VI. CROSS-DOMAIN RECOMMENDATION ACCURACY AND EXPLANATION QUALITY DUAL-DIMENSIONAL VALIDATION INDEX SYSTEM

A. EXPERIMENTAL SETUP

This paper uses a cross-domain product interaction dataset anonymized and sampled from a large e-commerce platform, covering three heterogeneous categories: apparel, home furnishings, and digital products. The dataset contains 126,820 product nodes, 437 million user interaction records, and product image-text description pairs. The data is divided into training, validation, and test sets in an 8:1:1 ratio based on time stamps. The test set retains 20% of cold-start products to verify the self-evolution capability of the knowledge graph. Baseline methods used for comparison include traditional cross-domain recommendation models CDR and SSCDR, multimodal baselines VBPR and MMGCN, and knowledge graph completion methods TransE and RotatE. All experiments were conducted on a server with four NVIDIA A100-80G GPUs. The multimodal large model backbone uses frozen LLaVA-v1.6-34B as the visual-language encoder, the graph neural network hidden layer dimension is set to 256, and the learning rate is determined to be $1e-4$ through grid search.

The evaluation system consists of two dimensions: objective and subjective indicators. Objective indicators include relation prediction hit rate (Hits@10), mean reciprocal ranking (MRR), and click-through rate (CTR), conversion rate (CVR), and recall rate (Recall@20) for the recommendation task. Subjective indicators are user ratings on a five-point Likert scale of the recommended explanatory text, covering explanatory fidelity and perceived usefulness.

B. VALIDATION EXPERIMENT ON THE ACCURACY OF CROSS-DOMAIN PRODUCT KNOWLEDGE GRAPH SELF-EVOLUTIONARY RELATIONSHIP COMPLETION

Conduct an empirical evaluation to demonstrate the correctness of the proposed cross-domain knowledge graph self-evolutionary mechanism in relation completion task. We

choose Rule-based method, TransE and RotatE as the baselines. We use product data from apparel, home furnishing and digital product domain to predict the new cross-domain relationship. Hits@10 and the average reciprocal ranking are adopted as evaluation metrics, which are commonly used in knowledge graph completion field to assess the dynamically evolving relation links.

As shown in Figure 1: Accuracy Evaluation of Self-Evolving Relation Completion in Cross-Domain Product Knowledge Graph, the proposed cross-domain relation self-evolutionary mechanism based on a multimodal large language model demonstrates a significant advantage in completion accuracy. Specifically, our proposed method achieves an accuracy of 0.892 on the Hits@10 metric, representing improvements of 24.2% and 20.1% respectively compared to traditional embedding models TransE (0.718) and RotatE (0.743), and a 36.8% improvement over rule-based methods (0.652). It also achieves an MRR of 0.537, a 30.3% improvement over RotatE. These data demonstrate that introducing a large language model as a dynamic inference engine effectively captures implicit semantic relationships between cross-category products, significantly enhancing the robustness and coverage accuracy of relation prediction in incremental evolution of the knowledge graph.

C. MULTIMODAL CROSS-DOMAIN EXPLAINABLE RECOMMENDATION PERFORMANCE COMPARISON EXPERIMENT

In this experiment, we test the recommendation performance of our framework in multimodal cross-domain recommendation. We choose CDR, SSCDR, VBPR and MMGCN as the comparison methods, and conduct the offline validation on the apparel, home furnishings and digital products joint dataset. The evaluation metrics are click through rate (CTR) or conversion rate (CVR), and recall@20, which represent user interest accuracy of the model and recommendation list coverage quality in cross domain case, respectively. Specific data is shown in Figure 2: Evaluation of Multimodal Cross-Domain Explainable Recommendation Performance. The comparative experimental results of recommendation performance further validate the performance advantages of the proposed framework in multimodal cross-domain scenarios. In terms of CTR, our method achieved 0.124, an improvement of 8.8% compared to the best baseline MMGCN (0.114), and a 51.2% improvement compared to the traditional cross-domain model CDR (0.082). Regarding conversion rate, our method achieved 0.062, 19.2% higher than MMGCN. Recall@20 reached 0.221, a 12.2% improvement compared to MMGCN. The consistent growth of these three core metrics indicates that introducing a multimodal large language model as the semantic alignment and relational reasoning hub can effectively bridge the representation gap between cross-category products, significantly enhancing the recommendation system's ability to capture user cross-domain interests and its conversion efficiency.

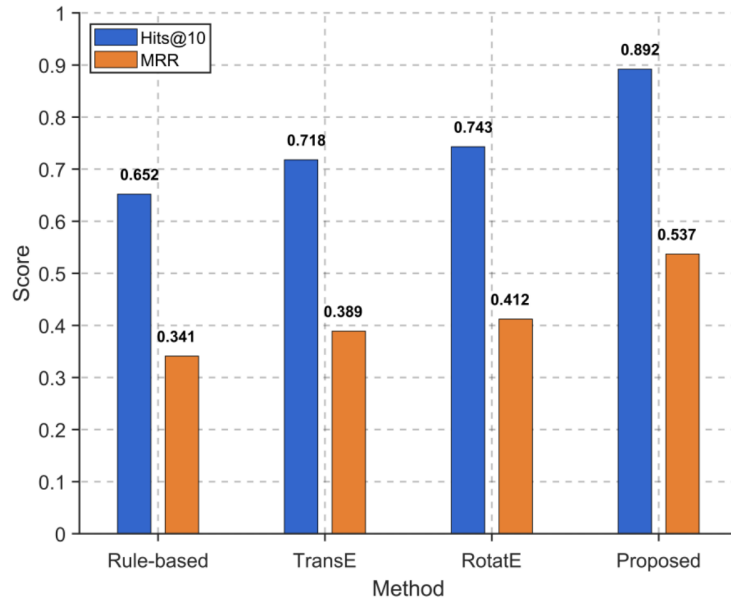


FIGURE 1. Accuracy Evaluation of Self-Evolving Relation Completion in Cross-Domain Product Knowledge Graph

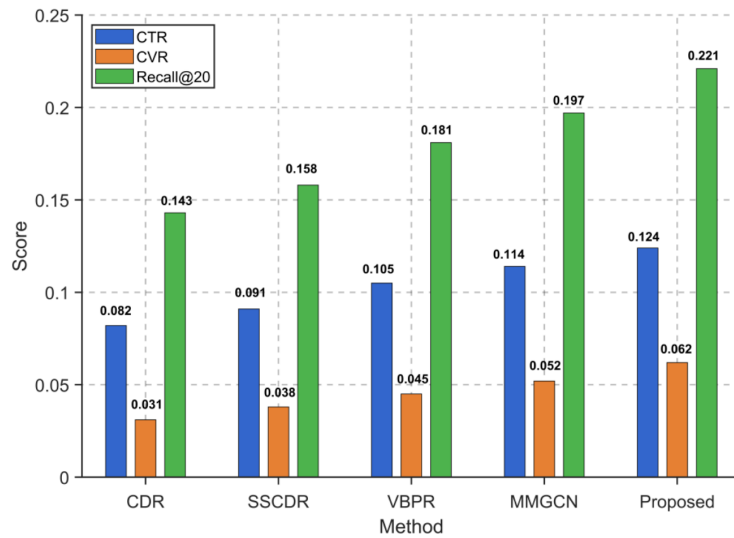


FIGURE 2. Evaluation of Multimodal Cross-Domain Explainable Recommendation Performance

D. EXPERIMENT ON EVALUATION OF RECOMMENDATION EXPLANATION CREDIBILITY AND PERSUASIVENESS BASED ON USER SUBJECTIVE RATINGS

We conduct an experiment to validate the quality of recommendation explanations from the perspective of user subjective perception. We recruit users with real online shopping experience to score the algorithm-generated explanation text with 5-levels Likert scale on two indicators of explanation fidelity and perceived usefulness. We also conduct an ablation test to remove the graph inference chain module and the large language model generation component to check the amount of contribution of each module to the

explanation quality.

TABLE 1. Evaluation of the Credibility and Persuasiveness of Recommendation Explanations Based on User Subjective Ratings

Method Configuration	Explanation Fidelity	Perceived Usefulness
No Explanation (Recommendation List Only)	—	2.14 ± 0.31
Rule-based Explanation with Attribute Templates	3.02 ± 0.28	2.87 ± 0.35
Proposed w/o Graph Reasoning Path	3.41 ± 0.22	3.26 ± 0.29
Proposed w/o LLM Generation	3.18 ± 0.25	2.95 ± 0.32
Proposed Full Framework	4.23 ± 0.18	4.15 ± 0.21

In Table 1, our complete framework achieved scores of 4.23 and 4.15 in explanation fidelity and perceived usefulness, respectively, significantly outperforming all ablation variants. After removing the graph reasoning chain, the scores dropped to 3.41 and 3.26 respectively, confirming the crucial role of cross-domain knowledge paths in supporting the explanatory logic. Removing the large model generation component further reduced the scores to 3.18 and 2.95, indicating that natural language generation capabilities significantly contribute to improving the fluency and persuasiveness of explanations. These data validate the effectiveness of the collaborative mechanism between graph reasoning and multimodal large models in explainable recommendations.

VII. CONCLUSION

To address the sparse knowledge association and weak explainability in cross-domain e-commerce recommendations, this paper presents a knowledge graph self-evolution and explainable recommendation framework guided by a multimodal large language model as its cognitive center. Experiments validate that the framework can improve cross-modal product semantics alignment, realize dynamic incremental completion of cross-category relationships, and produce logically supported natural language explanations without compromising recommendation accuracy. The proposed framework is still limited in terms of real-time streaming update efficiency and robustness in extreme cold-start scenarios. Our future work will explore the introduction of lightweight model adaptation and online learning mechanisms to facilitate the efficient deployment of the framework in real-world production environments.

AUTHOR CONTRIBUTIONS

Conceptualization – Li Ma and Yuanli Cui; methodology – Li Ma and Yuanli Cui; investigation – Li Ma and Yuanli Cui; writing-original draft preparation – Li Ma and Yuanli Cui. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This work was supported by the 2025 Special Project on Traditional Culture and Economic-Social Development in Shandong Province is titled ‘A Study on the Impact of Differentiated Live Streaming Content on the Transmission of Shandong Cuisine Culture Among Young Audiences’. (Project Number: 25C520024282)

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] C. Kumar, C. R. Chowdary, and A. K. Meena, “Recent trends in recommender systems: a survey,” *International Journal of Multimedia Information Retrieval*, vol. 13, no. 4, pp. 1-26, 2024, doi: 10.1007/s13735-024-00349-1.
- [2] S. Gheewala, S. Xu, and S. Yeom, “In-depth survey: deep learning in recommender systems-exploring prediction and ranking models, datasets, feature analysis, and emerging trends,” *Neural Computing and Applications*, vol. 37, no. 17, pp. 10875-10947, 2025, doi: 10.1007/s00521-024-10866-z.
- [3] Y. Huang, F. Zhao, X. Gui, et al., “Path-enhanced explainable recommendation with knowledge graphs,” *World Wide Web*, vol. 24, no. 5, pp. 1769-1789, 2021, doi: 10.1007/s11280-021-00912-4.
- [4] M. Caro-Martínez, G. Jiménez-Díaz, and J. A. Recio-García, “A graph-based approach for minimising the knowledge requirement of explainable recommender systems,” *Knowledge and Information Systems*, vol. 65, no. 10, pp. 4379-4409, 2023, doi: 10.1007/s10115-023-01903-9.
- [5] Y. Lin, W. Zhang, F. Lin, et al., “Knowledge-aware reasoning with self-supervised reinforcement learning for explainable recommendation in MOOCs,” *Neural Computing and Applications*, vol. 36, no. 8, pp. 4115-4132, 2024, doi: 10.1007/s00521-023-09257-7.
- [6] Y. Zhao, J. Ju, J. Gong, et al., “Cross-domain recommendation via adaptive bi-directional transfer graph neural networks,” *Knowledge and Information Systems*, vol. 67, no. 1, pp. 579-602, 2025, doi: 10.1007/s10115-024-02246-9.
- [7] H. Fang, J. Liang, and L. Sha, “Enhanced multimodal recommendation systems through reviews integration,” *Knowledge and Information Systems*, vol. 67, no. 4, pp. 3459-3486, 2025, doi: 10.1007/s10115-024-02309-x.
- [8] Z. Lyu, Y. Wu, J. Lai, et al., “Knowledge Enhanced Graph Neural Networks for Explainable Recommendation,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 5, pp. 4954-4968, 2023, doi: 10.1109/TKDE.2022.3142260.
- [9] Y. Wei, X. Qu, X. Wang, et al., “Rule-Guided Counterfactual Explainable Recommendation,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 5, pp. 2179-2190, 2024, doi: 10.1109/TKDE.2023.3322227.
- [10] D. Yu, Q. Li, X. Wang, et al., “Counterfactual Explainable Conversational Recommendation,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 6, pp. 2388-2400, 2024, doi: 10.1109/TKDE.2023.3322403.
- [11] C. Zhao, H. Zhao, X. Li, et al., “Cross-Domain Recommendation via Progressive Structural Alignment,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 6, pp. 2401-2415, 2024, doi: 10.1109/TKDE.2023.3324912.