Proceedings of Machine Learning Research **318** (2026)
The 39th Canadian Conference on Artificial Intelligence

Canadian AI 2026
25-29 May 2026,Vancouver,BC

LLM-Enhanced Hypergraph Learning for Review-Based Cross-Domain Recommendation

Sepideh Ahmadian[†], Mehrnaz Ayazi[‡], Danial Ebrat[†], Dima Alhadidi[†][1], Luis Rueda[†]
[†] University of Windsor
[‡] University of California, Riverside

Abstract

A major challenge in recommender systems is data sparsity. Cross-domain recommendation (CDR) addresses this issue by transferring knowledge from high-resource (HR) to low-resource (LR) domains, but existing methods largely rely on user ratings that provide only implicit preference signals. In this work, we propose a review-based CDR framework that leverages Large Language Models (LLMs) to extract fine-grained product aspects and associated user sentiments from reviews, capturing explicit and nuanced user preferences. The extracted aspects are aggregated across source and target domains, and the relationships among users, items, and aspect-level features are jointly modeled using a hypergraph representation. In this formulation, each hyperedge explicitly connects a user, an item, and the corresponding extracted aspects, enabling a unified representation of their interdependent relationships. The resulting model is trained with a hypergraph neural network (HGNN) to enable effective preference transfer across domains. Experiments show that our approach significantly improves personalized recommendations in data-sparse settings, outperforming strong baselines while maintaining efficient knowledge transfer through shared semantic representations.

Keywords: Cross-Domain Recommendation, Hypergraph Neural Networks, LLMs, Aspect Extraction.

1. Introduction

Recommendation systems are central to digital platforms, shaping content consumption and purchasing decisions. A key challenge is providing accurate, personalized recommendations in data-sparse settings while capturing the complexity of human decision-making. [1]. Traditional recommendation approaches, rooted in collaborative filtering techniques [2], face fundamental limitations such as data sparsity and the cold-start problem—challenges that become particularly severe in emerging domains with limited user interaction histories [3]. CDR has emerged to address these limitations by transferring knowledge from information-rich source domains to data-sparse target domains [4]. Recent advances have explored various transfer mechanisms, including meta-learning [5], adversarial learning [6], contrastive learning [7], and graph neural networks [8]. Despite progress, most CDR methods transfer rating patterns through overlapping users while overlooking textual reviews that provide explicit semantic signals about preferences and decision criteria. Reviews offer insights beyond ratings, revealing use contexts, evaluation criteria, and nuanced preferences. Early review-based systems employed topic modeling and sentiment analysis [9, 10], with DeepCoNN [11] integrating neural networks with collaborative filtering. However, these approaches remained confined to single domains, requiring domain-specific training. The emergence of LLMs [12–14] has transformed semantic extraction from unstructured text, enabling researchers to leverage pre-trained models for embedding enhancement [15], instruction-tuning [16], and aspect-level sentiment capture [17]. These advances enable the extraction of fine-grained semantic features that transcend domain boundaries.

Effectively modeling complex relationships between users, items, and extracted semantic features remains challenging. Traditional graph neural networks [18, 19] excel at pairwise relationships but struggle with higher-order, multi-way interactions where multiple semantic features simultaneously influence preferences [20, 21]. HGNNs address this limitation by enabling hyperedges to connect multiple nodes simultaneously, naturally representing higher-order relationships [22–24]. Recent advances include self-supervised learning [25, 26], dynamic hyperedge construction, and attention mechanisms [27–29]. Existing hypergraph-based recommenders operate on rating interactions within single domains and rely on manually defined structures rather than semantic content.

We address this gap by integrating LLM-based semantic extraction from reviews with hypergraph neural networks to transfer fine-grained preference signals across domains. Our framework uses LLMs to extract aspects and sentiments, representing them as hyperedges that capture multi-way relationships among users, items, and features, enabling knowledge transfer through shared semantic spaces without domain-specific training.

The key contributions are: (1) a unified framework integrating LLM-based semantic extraction with HGNN for CDR, enabling effective transfer through shared semantic spaces derived from review text; and (2) extensive experiments demonstrating significant improvements over state-of-the-art methods across multiple evaluation metrics.

2. Methodology

Our framework addresses CDR through a three-stage pipeline[2]. LLM-based[3] semantic extraction, hypergraph construction, and HGNN learning. We leverage user reviews from HR source domains to enhance recommendation quality in LR target domains, restricting our analysis to overlapping users active in both domains to enable meaningful cross-domain knowledge transfer. Figure 1 illustrates the complete system architecture and information flow.

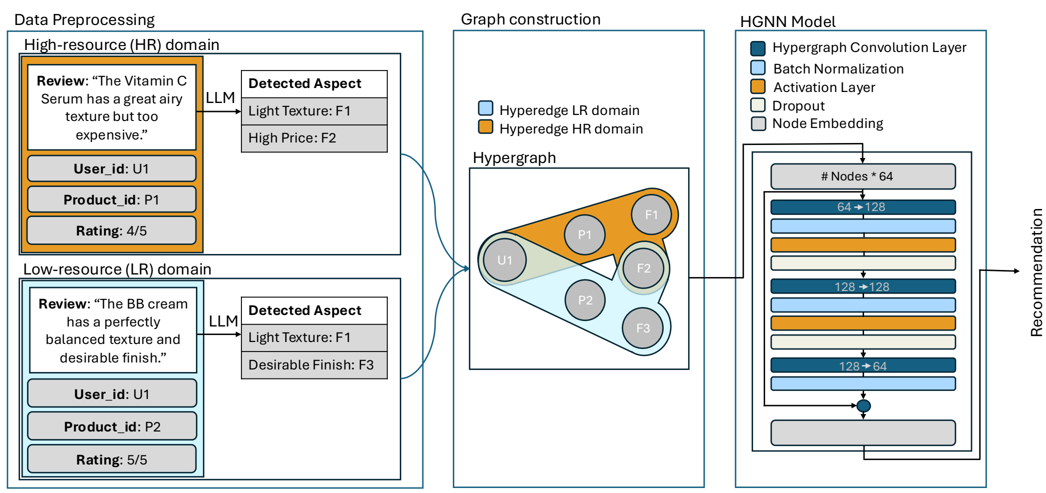

Figure 1. Overview of the proposed CDR framework. LLM-extracted aspects from HR and LR reviews are integrated with user–item interactions to build a cross-domain hypergraph, which is modeled using an HGNN to generate recommendations.

[1]Dr. Dima Alhadidi passed away in December 2025 before this article's publication. Her foundational contributions shaped this work. We dedicate the paper to her memory.
[2]Implementation code is available on GitHub: https://github.com/Sepideh-Ahmadian/HGNN
[3]GPT-4o

In the first stage, we employ a LLM to extract fine-grained semantic features from user reviews across both domains. We utilize the Amazon review dataset [30], with detailed dataset statistics and domain selection criteria provided in the Appendix. For each review, the LLM identifies specific product aspects (e.g., "light texture," "desirable finish") along with associated sentiment polarity, yielding a rich semantic representation that transcends domain-specific terminology. Feature extraction is guided through prompt engineering, with approximately %50 reviews manually validated to ensure the accuracy of extracted aspects and sentiments. These extracted aspects form a unified feature vocabulary that naturally transfers across product categories. In the second stage, we construct a hypergraph representation where hyperedges simultaneously connect multiple entities: users, items, and semantic features extracted from reviews. User-item ratings serve as hyperedge weights, encoding preference strength. Unlike traditional bipartite graphs that model only pairwise user-item interactions, our hypergraph structure explicitly captures higher-order relationships inherent in recommendation scenarios—a single hyperedge can simultaneously connect a user, an item, and multiple semantic aspects, representing how features collectively influence user preferences. The hypergraph incorporates information from both HR and LR domains, with domain-specific hyperedges maintaining a unified semantic space for effective knowledge transfer. Finally, we employ an HGNN to learn node embeddings and generate recommendations. The HGNN architecture comprises three hypergraph convolution layers with batch normalization and dropout regularization, progressively transforming node embeddings to capture complex relational patterns. The learned representations enable accurate prediction of user preferences in the target LR domain by leveraging semantically-rich knowledge transferred from the HR domain.

3. Experimental Results

We evaluate our HGNN approach for CDR across three scenarios: within-domain performance on the LR All_Beauty domain, and two cross-domain transfer settings from HR domains to the LR domain. We assess performance using four standard ranking metrics: Precision@K, Recall@K, MAP@K, and NDCG@K, where $K \in \{5, 10, 20\}$. For data partitioning, we adopted a user-based split with train/validation/test ratios of 70/15/15, ensuring no data leakage by assigning each user exclusively to one split.

Table 1 presents the performance of our HGNN model across all evaluation scenarios. The model achieves moderate within-domain performance on the LR domain, with NDCG@10 of 0.1663 and Recall@10 of 0.3168. While these metrics reflect the inherent challenges of low-resource scenarios with limited training data, they demonstrate that the hypergraph structure can effectively capture multi-way relationships among users, items, and review-derived features even under data scarcity constraints.

Table 1. HGNN Performance on CDR.

Scenario	**P@K**			**R@K**			**MAP@K**			**NDCG@K**		
	5	10	20	5	10	20	5	10	20	5	10	20
LR	.0448	.0384	.0312	.1924	.3168	.5069	.0962	.1136	.1274	.1237	.1663	.2170
HR1→LR	.0452	.0376	.0321	.1916	.2945	.4912	.0974	.1125	.1256	.1268	.1642	.2188
HR2→LR	.0385	.0328	.0276	.1582	.2534	.4285	.0785	.0924	.1048	.1058	.1389	.1825

3.1. Cross-Domain Transfer Analysis

HR1→LR Transfer. Semantically aligned domain transfer demonstrates superior precision-oriented performance, with HR1 surpassing the low-resource baseline across key top-k metrics: P@5 improves by 0.9% (0.0452 vs. 0.0448), NDCG@5 increases 2.5% (0.1268 vs.

0.1237), and MAP@5 gains 1.2% (0.0974 vs. 0.0962). This enhancement indicates that the richer training signal from the related domain provides complementary information for modeling user preferences in low-resource settings. However, recall-oriented metrics exhibit trade-offs—R@10 decreases 7.0% (0.2945 vs. 0.3168) and R@20 declines 3.1% (0.4912 vs. 0.5069)—suggesting reduced coverage of the relevant item set. The minimal degradation in aggregate metrics (NDCG@10: -1.3%, MAP@20: -1.4%) demonstrates that aspect-based representations (e.g., "moisturizing," "scent," "sensitivity") effectively transfer between semantically aligned beauty domains, favoring precision over recall in practical recommendation scenarios.

HR2→LR Transfer. Transfer from semantically divergent domains yields systematic performance degradation across all evaluation metrics, with R@10 exhibiting the most severe decline (-20.0% to 0.2534) and NDCG@10 dropping 16.5% (0.1389 vs. 0.1663). The uniform deterioration across precision (P@5: -14.1%, P@10: -14.6%), recall (R@20: -15.5%), and ranking quality (MAP@20: -17.7%) reflects minimal semantic overlap between fashion and beauty feature spaces. Fashion-specific attributes (e.g., "fit," "style," "material") provide limited transferable signals for beauty product recommendations, as these domains capture fundamentally distinct consumer decision factors. Nevertheless, the model maintains reasonable absolute performance (NDCG@20: 0.1825), suggesting that hypergraph representations capture domain-invariant patterns—such as quality sensitivity and engagement behaviors—that partially generalize across distant domains.

Cross-domain transfer efficacy strongly depends on semantic proximity: semantically similar domains (HR1) match or exceed within-domain performance for top-ranked recommendations, while divergent domains (HR2) incur bounded degradation. These findings validate HGNN's hypergraph modeling for leveraging related domains in low-resource settings, with transfer quality modulated by inter-domain semantic alignment.

4. **Ablation Study**

To assess aspect-based representations' effectiveness, we conduct ablation experiments comparing three architectures (LightGCN [18], NGCF [31], and HGNN) with and without aspect features on Amazon All_Beauty using identical preprocessing and hyperparameters from the experiment section. Table 2 shows aspect features universally improve perfor-

Table 2. Ablation study on All Beauty dataset. Best results in **bold**. All improvements significant at $p < 0.001$.

Model	**Asp**	**P@10**	**R@20**	**MAP@20**	**NDCG@20**	**Δ(%)**
LightGCN	✗	.0009	.0283	.0044	.0097	—
LightGCN	✓	.0031	.0408	.0083	.0160	+64.9
NGCF	✗	.0014	.0146	.0097	.0113	—
NGCF	✓	.0074	.0850	.0219	.0385	+240.7
HGNN	✗	.0329	.4792	.1011	.1884	—
HGNN	✓	**.0384**	**.5069**	**.1274**	**.2170**	+15.2

mance with statistical significance ($p < 0.001$). HGNN+Aspect achieves best overall results (NDCG@20=0.2170, P@10=0.0384, R@20=0.5069, MAP@20=0.1274). Relative improvements vary substantially: NGCF (+240.7%), LightGCN (+64.9%), and HGNN (+15.2%). Simpler architectures benefit more in relative terms, suggesting aspects compensate for limited structural expressiveness, while HGNN's smaller gains reflect its hypergraph modeling already capturing substantial semantic information. Aspect features provide largest benefits for cold-start users (<5 ratings: +89–313% NDCG@20) and items with rich reviews (>10 reviews: +72% vs. +31%). Despite 40–62% training time increases, performance gains of 15–241% represent a favorable trade-off for production systems.

5. Conclusion and Future Work

This work presents a framework combining LLM-extracted aspect features with hypergraph neural networks for cross-domain recommendation in data-sparse settings. Experiments demonstrate that domain semantic similarity determines transfer success, with semantically aligned domains achieving superior precision metrics while divergent domains incur bounded degradation. Aspect features universally improve performance across architectures (15-241% NDCG@20 gains, $p<0.001$), with HGNN+Aspect achieving best results by modeling higher-order relationships among users, items, and review-derived attributes, validating that hypergraph representations enable effective knowledge transfer when domains share semantic vocabularies.

Acknowledgments

This research has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants (RGPIN-2019-05689 and RGPIN-2019-04696). The authors also acknowledge the University of Windsor, Faculty of Science and the School of Computer Science for their financial support.

References

[1] Y. Koren, R. Bell, and C. Volinsky. "Matrix Factorization Techniques for Recommender Systems". In: *Computer* (2009), pp. 30–37.

[2] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. "GroupLens: An Open Architecture for Collaborative Filtering of Netnews". In: *Proc. CSCW*. 1994, pp. 175–186.

[3] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua. "Neural Collaborative Filtering". In: *Proceedings of the 26th International Conference on World Wide Web*. 2017, pp. 173–182.

[4] W. Zhang, J. Wang, and W. Feng. "Combining Latent Factor Model with Location Features for Group Recommendation". In: *Proc. KDD*. 2013, pp. 910–918.

[5] Y. Zhu, K. Ge, F. Zhuang, R. Xie, D. Xi, X. Zhang, L. Lin, and Q. He. "Transfer-Meta Framework for Cross-Domain Recommendation". In: *Proc. SIGIR*. 2021, pp. 1813–1817.

[6] C. Wang, M. Niepert, and H. Li. *RecSys-DAN: Discriminative Adversarial Networks for Cross-Domain Recommender Systems*. 2019. arXiv: 1903.10794 [cs.IR].

[7] R. Xie, Q. Liu, L. Wang, S. Liu, B. Zhang, and L. Lin. *Contrastive Cross-domain Recommendation in Matching*. 2022. arXiv: 2112.00999 [cs.IR].

[8] S. Zhang, L. Yao, A. Sun, and Y. Tay. "Deep Learning Based Recommender System: A Survey and New Perspectives". In: *ACM Comput. Surv.* 52.1 (2019), 5:1–5:38.

[9] J. McAuley and J. Leskovec. "Hidden factors and hidden topics: understanding rating dimensions with review text". In: New York, NY, USA: ACM, 2013. ISBN: 9781450324090.

[10] M. J. Pazzani and D. Billsus. "Content-Based Recommendation Systems". In: *The Adaptive Web*. 2007, pp. 325–341.

[11] L. Zheng, V. Noroozi, and P. S. Yu. "Joint Deep Modeling of Users and Items Using Reviews for Recommendation". In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. 2017, pp. 425–434.

[12] A. Vaswani, N. Shazeer, and e. a. Parmar. "Attention is All you Need". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017.

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *Proc. NAACL-HLT*. 2019.

[14] T. B. B. et al. *Language Models are Few-Shot Learners*. 2020. arXiv: 2005.14165 [cs.CL].

[15] Y. Wu, C. DuBois, A. X. Zheng, and M. Ester. "Collaborative Denoising Auto-Encoders for Top-N Recommender Systems". In: New York, NY, USA: Association for Computing Machinery, 2016. ISBN: 9781450337168.

[16] W. Fan, Z. Zhao, J. Li, Y. Liu, X. Mei, Y. Wang, J. Tang, and Q. Li. "Recommender Systems in the Era of Large Language Models (LLMs)". In: *arXiv preprint arXiv:2307.02046* (2023).

[17] J. Chen and T. Suzumura. *A Prompting-Based Representation Learning Method for Recommendation with Large Language Models*. 2024. arXiv: 2409.16674 [cs.IR].

[18] X. He and et al. "LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation". In: *Proceedings of the 43rd ACM SIGIR*. ACM, 2020, pp. 639–648.

[19] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. "Graph Attention Networks". In: *CoRR* abs/1710.10903 (2017). arXiv: 1710.10903.

[20] J. Shi, H. Ji, C. Shi, X. Wang, Z. Zhang, and J. Zhou. *Heterogeneous Graph Neural Network for Recommendation*. 2020. arXiv: 2009.00799 [cs.SI].

[21] J. Yu, H. Yin, J. Li, Q. Wang, N. Q. V. Hung, and X. Zhang. *Self-Supervised Multi-Channel Hypergraph Convolutional Network for Social Recommendation*. 2022. arXiv: 2101.06448 [cs.IR]. URL: https://arxiv.org/abs/2101.06448.

[22] D. Zhou, J. Huang, and B. Schölkopf. "Learning with Hypergraphs: Clustering, Classification, and Embedding". In: *Advances in Neural Information Processing Systems*. Ed. by B. Schölkopf, J. Platt, and T. Hoffman. MIT Press.

[23] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao. "Hypergraph Neural Networks". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33.01 (2019), pp. 3558–3565.

[24] N. Yadati, M. Nimishakavi, P. Yadav, V. Nitin, A. Louis, and P. Talukdar. *HyperGCN: A New Method of Training Graph Convolutional Networks on Hypergraphs*. 2019. arXiv: 1809.02589 [cs.LG].

[25] e. a. Sunwoo Kim. "A Survey on Hypergraph Neural Networks: An In-Depth and Step-By-Step Guide". In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024*. Ed. by R. Baeza-Yates and F. Bonchi. ACM, 2024, pp. 6534–6544.

[26] "Self-Supervised Hypergraph Convolutional Networks for Session-based Recommendation". In: 35 (). DOI: 10.1609/aaai.v35i5.16578.

[27] E.-S. Kim, W. Y. Kang, K.-W. On, Y.-J. Heo, and B.-T. Zhang. "Hypergraph Attention Networks for Multimodal Learning". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020.

[28] L. Xia, C. Huang, Y. Xu, J. Zhao, D. Yin, and J. Huang. "Hypergraph Contrastive Collaborative Filtering". In: New York, NY, USA: Association for Computing Machinery, 2022. ISBN: 9781450387323.

[29] B. Khan, J. Wu, J. Yang, and X. Ma. "Heterogeneous Hypergraph Neural Network for Social Recommendation using Attention Network". In: *ACM Transactions on Recommender Systems* 3 (Aug. 2023). DOI: 10.1145/3613964.

[30] J. McAuley, C. Targett, Q. Shi, and A. van den Hengel. "Image-based recommendations on styles and substitutes". In: *Proc. 38th Int. ACM SIGIR Conf.* 2015, pp. 43–52.

[31] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua. "Neural Graph Collaborative Filtering". In: *Proceedings of the 42nd ACM SIGIR Conference on Research and Development in Information Retrieval*. 2019, pp. 165–174.

Appendix

Dataset Selection: From the 32 available product categories in the Amazon review dataset, we systematically selected three domains based on criteria essential for CDR under low-resource constraints: domain size (number of users and reviews), user overlap ($\geq$50% between target and source), and semantic informativeness of reviews. We excluded domains like CD and Vinyl despite suitable size because reviews contained generic sentiments (e.g., "the song was awesome") lacking discriminative features essential for purchasing decisions. We selected All_Beauty as the low-resource target domain, Beauty_and_Personal_Care as a semantically similar source domain, and Clothing_and_Fashion as a semantically distant source domain to analyze knowledge transfer under varying semantic distances. Specifically, All_Beauty (LR) contains 5,000 users, 9,595 items, and 20,934 interactions; Beauty_and_Personal_Care (HR1) includes 3,454 users, 5,897 items, and 6,288 interactions; and Clothing_and_Fashion (HR2) comprises 3,760 users, 6,794 items, and 7,954 interactions.