ADScAI
Applied Data Science & Artificial Intelligence Symposium

NeuroFit: Enhancing Fashion Recommendations Through Multi-Modal Graph Neural Networks with Cold-Start User Handling

M. L. Rathnayaka
Department of Software Engieering
Informatics Institute of Technology
Colombo, Sri Lanka
methma.20210501@iit.ac.lk

K. L. P. P. Liyanaarachchi
Informatics Institute of Technology
Colombo, Sri Lanka
prasad.l@iit.ac.lk

Keywords—Cold-start problem, Graph Neural Networks (GNN), Multi-modal recommendation, Personalized recommendations, E-commerce

I. Introduction

E-commerce has transformed consumer shopping habits significantly, enhancing accessibility yet introducing complexity in navigating extensive product catalogs. Personalized fashion recommendations have emerged as a crucial element in improving customer experience and boosting sales. However, traditional recommendation systems (RS) often struggle with cold-start scenarios, where new users or items possess limited interaction histories, and face challenges effectively utilizing multimodal data, including visual, textual, price and demographic information. To overcome these limitations, this research presents a novel multi-modal Graph Neural Network (GNN) recommendation system, NeuroFit, designed explicitly for enhancing personalization and managing cold-start issues.

II. Content

A. Background and Problem Definition

Personalized fashion recommendations rely heavily on accurately interpreting user preferences from multimodal data, such as product images, descriptive text, and user demographics. The cold-start issue remains prominent when systems fail to recommend effectively due to sparse data for new users or items. Current approaches often handle limited modalities separately, inadequately addressing cold-start issues and failing to provide accurate recommendations.

B. Literature Review

Recent advancements in RSs have increasingly focused on multimodal data integration to improve personalization. Multimodal recommender systems (MRS) leverage diverse data such as visual, textual, numeric (price) and demographic information, providing richer representations of products and users. For example, [1] proposed CAMRec, a co-attentionbased multimodal recommendation approach integrating image and text features, achieving superior results compared to single-modality systems but lacking demographic data integration. Similarly, [2] demonstrated that multimodal embedding generated through vision transformers (ViT) and BERT significantly outperform traditional

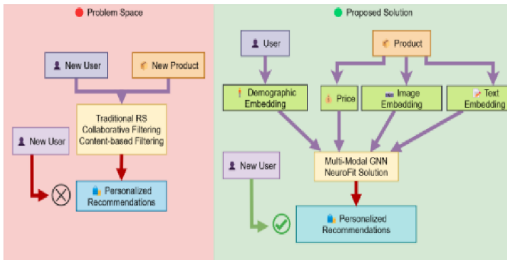

Fig 1 Problem Definition and Research Scope

monomodal approaches. Despite these advances, comprehensive demographic integration remains limited, reducing the system's ability to handle diverse user groups effectively. GNNs particularly Heterogeneous Graph Transformers (HGT), have become powerful tools in RSs due to their ability to model complex user-item interactions as graph structures.

GNNs effectively capture rich relational data, enhancing recommendation accuracy, especially in sparse interaction scenarios [3]. [4] further demonstrated that GNNs significantly improve recommendation quality by modeling higher-order connectivity within user-item graphs, thus capturing subtle patterns missed by traditional approaches. Cold-start challenges remain one of the critical limitations for many RSs, especially within rapidly evolving ecommerce domains. New users or products typically lack sufficient interaction history, significantly limiting the effectiveness of traditional RSs. [5] discussed the potential of leveraging demographic attributes, such as user age, to mitigate cold-start issues, highlighting that incorporating such auxiliary information can substantially enhance initial recommendation quality. Recent surveys [6] have emphasized employing multimodal data and graph-based approaches to effectively address cold-start issues, suggesting substantial improvements over traditional content or collaborative filtering methods. While existing solutions individually tackle multimodal integration, graph-based recommendation, or cold-start handling, there remains a research gap in effectively combining these elements within a unified framework. Specifically, a holistic integration of visual, textual, numeric and demographic features, leveraged through GNN architectures, represent an under-explored yet promising approach. The NeuroFit framework addresses precisely this

https://doi.org/10.31705/ADScAI.2025.52

TABLE I. SUMMARY OF RELATED WORKS

Method	Modalities	Cold-Start Handling	Limitations
CAMRec	Image, Text	No	Ignore demographic information
(ViT+BERT)	Image, Text	Partial	Lacks robust demographic integration
HGT	Graph-based	Partial	Limited multimodal integration
NeuroFit	**Multi-Modal GNN**	**Yes**	**Comprehensive integration**

gap by fusing multimodal embeddings and explicitly targeting cold-start users, thus extending the state-of-the-art capabilities in fashion RSs. Table I shows the summary of related work.

C. *Materials and Methods*

The NeuroFit approach integrates a multi-modal GNN architecture. Primary modalities—visual (product images), textual (product descriptions), numarics (price) and demographic data (user age)—are processed individually and integrated through embedding fusion. Visual embeddings are extracted using pre-trained ResNet architectures, textual embeddings are obtained via DistilBERT transformer models, prices are normalized, and demographic data is encoded through embedding layers. Each modality embedding is normalized and concatenated into a unified multimodal vector representation to construct the heterogeneous user-item interaction graph.

GNN leveraging graph convolutional layers captures complex user-item interactions, generating personalized recommendations. Model performance was evaluated using standard metrics: Recall@10, NDCG@10, and MAP@12 as shown in fig 2.

D. *Results and Discussion*

The NeuroFit system significantly outperformed traditional baseline models, particularly in handling cold-start situations. The results, summarized in Table II, clearly demonstrate improved recommendation accuracy and quality. The considerable improvements highlight the effectiveness of integrating multimodal embeddings and explicitly addressing cold-start challenges. The embedding

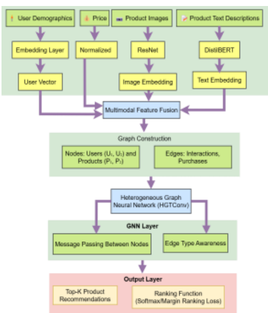

Fig. 2 Proposed Multi-Modal GNN Architecture

TABLE II. PERFORMANCE COMPARISON

Model	Recall@10	NDCG@10	MAP@12
Baseline (No cold start handling)	0.1783	0.1006	0.2076
Proposed Multi-Modal GNN (with cold start handling)	0.4915	0.2743	0.5502

fusion approach allowed richer semantic interactions, directly improving recommendation relevance.

Although NeuroFit has demonstrated good scalability for typical e-commerce scenarios, future research is recommended optimizing performance further when scaling larger, denser user-item graphs.

E. *Conclusion*

This study introduced NeuroFit, a multi-modal GNN recommendation system capable of effectively handling cold-start issues through integrated multimodal data analysis. By combining user demographics, textual descriptions, prices and visual content, NeuroFit significantly enhanced recommendation performance.

Future directions include exploring explicit user feedback post-purchase to refine recommendations dynamically and evaluating NeuroFit's applicability beyond fashion ecommerce, thereby enhancing its robustness and versatility across different domains.

ACKNOWLEDGMENT

First and foremost, I would like to express my sincere gratitude to my supervisor, Mr. Pushpika Liyanaarachchi, for his continuous guidance, encouragement, and valuable feedback throughout this project. I also wish to thank the University of Westminster for providing the resources and academic support necessary for the successful completion of this research. Special thanks to my peers, friends, and family for their constant support and motivation during this journey. Finally, I am grateful to all individuals who provided feedback and insights that helped refine and evaluate this work.

REFERENCES

[1] E. Jeong, X. Li, A. (Eunyoung) Kwon, S. Park, Q. Li, and J. Kim, "A Multimodal Recommender System Using Deep Learning Techniques Combining Review Texts and Images," *Appl. Sci.*, vol. 14, no. 20, p. 9206, Oct. 2024, doi: 10.3390/app14209206.

[2] B. Huang, Q. Lu, S. Huang, X. Wang, and H. Yang, "Multi-modal clothing recommendation model based on large model and VAE enhancement," 2024, *arXiv*. doi: 10.48550/ARXIV.2410.02219.

[3] Z. Hu, Y. Dong, K. Wang, and Y. Sun, "Heterogeneous Graph Transformer," in *Proceedings of The Web Conference 2020*, Taipei Taiwan: ACM, Apr. 2020, pp. 2704–2710. doi: 10.1145/3366423.3380027.

[4] K. Li, L. Xu, C. Zhu, and K. Zhang, "A Multimodal Graph Recommendation Method Based on Cross-Attention Fusion," *Mathematics*, vol. 12, no. 15, p. 2353, July 2024, doi: 10.3390/math12152353.

[5] M. Elahi and L. Qi, "Fashion Recommender Systems in Cold Start," in *Fashion Recommender Systems*, N. Dokoohaki, Ed., in Lecture Notes in Social Networks. , Cham: Springer International Publishing, 2020, pp. 3–21. doi: 10.1007/978-3-030-55218-3_1.

[6] W. Zhang *et al.*, "Cold-Start Recommendation towards the Era of Large Language Models (LLMs): A Comprehensive Survey and Roadmap," 2025, *arXiv*. doi: 10.48550/ARXIV.2501.01945.