Enhanced Recommendation Systems with Retrieval-Augmented Large Language Model (Abstract Reprint)

Chuyuan Wei[1], Ke Duan[1], Shengda Zhuo[2], Hongchun Wang[1], Shuqiang Huang[2], Jie Liu[3]

[1]Beijing University of Civil Engineering and Architecture
[2]Jinan University
[3]North China University of Technology

Abstract Reprint. This is an abstract reprint of the journal article by Wei, Duan, Zhuo, Wang, Huang, and Liu (2025).

Abstract

Recommender systems have long struggled with challenges such as cold start and data sparsity, which can lead to poor recommendation performance. While previous approaches have attempted to address these issues by incorporating side information, they often introduce noise, lack flexibility for data expansion, and suffer from inconsistent data quality-factors that hinder accurate user preference inference and reduce recommendation performance. With the vast knowledge bases and advanced reasoning capabilities of large language models (LLMs), these models are particularly well-suited to supplement auxiliary information and capture implicit user intent. To address these challenges, we propose a novel framework, ER2ALM, which leverages the capabilities of LLMs enhanced by Retrieval-Augmented Generation (RAG) to improve recommendation outcomes. Our framework specifically addresses the challenges by flexibly and accurately augmenting auxiliary information and capturing users implicit preferences and interests. Additionally, to mitigate the risk of introducing noise, we incorporate a noise reduction strategy to ensure the reliability of the augmented information. Experimental validation on two real-world datasets demonstrates the efficacy of our approach, significantly enhancing both the accuracy and robustness of recommendations compared to state-of-the-art methods. This demonstrates the potential of our framework as a new paradigm for preference mining in recommendation systems.

References

Wei, C.; Duan, K.; Zhuo, S.; Wang, H.; Huang, S.; and Liu, J. 2025. Enhanced Recommendation Systems with Retrieval-Augmented Large Language Model. *Journal of Artificial Intelligence Research*, 82: 1147–1173.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.