OUKA
The University of Osaka
Institutional Knowledge Archive

Title	Research on Improving GNN-based Recommender System Involving Multi-modal Information
Author(s)	Jiao, Shengzhe
Citation	大阪大学, 2026, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/105505
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

Form 3

Abstract of Thesis

Name (Shengzhe Jiao)	
Title	Research on Improving GNN-based Recommender System Involving Multi-modal Information (マルチモーダル情報を用いたGNNベース推薦システムの性能向上に関する研究)

Abstract of Thesis

With the rapid growth of online platforms and the large number of products and services they provide, users are increasingly overwhelmed by information and face difficulties in identifying items that match their preferences. Recommender systems (RSs) have been developed to address this problem by modeling user-item interactions, such as views, clicks, and purchases, and generating personalized recommendations. By predicting user preferences from historical behaviors, RSs improve user experience and provide substantial commercial value to service providers. As a result, RSs have become a fundamental component of modern online ecosystems, including e-commerce platforms, streaming services, and social networks.

Recent advances in deep learning have significantly enhanced the ability of RSs to capture complex and nonlinear relationships between users and items. Deep models can automatically learn high-dimensional representations that reflect both explicit and implicit factors of user preferences, leading to improved recommendation accuracy. However, their effectiveness is often limited by the sparsity of interaction data: in real-world systems, users typically interact with only a small subset of items, and many items receive limited feedback. This sparsity poses a major challenge for learning robust representations and reduces the practical performance of deep learning-based RSs.

To alleviate data sparsity, multi-modal recommender systems have been proposed to incorporate additional side information, such as images, textual descriptions, and categorical attributes. These heterogeneous modalities provide complementary perspectives on user preferences and item characteristics, enabling richer representations. Despite their potential, existing multi-modal approaches still face several fundamental challenges that limit their effectiveness, robustness, and explainability.

First, many methods integrate different modalities in a shallow manner, often processing each modality independently and combining them through simple fusion strategies. Such designs fail to capture the underlying relationships between modalities and overlook modality-specific relations, such as visual similarity or textual correlation among items, leading to incomplete item representations. Second, there is often a misalignment between user behaviors and side information: interaction-based embeddings and content-based features are learned separately, which may result in representations that are semantically inconsistent with actual user preferences. Third, existing RSs generally lack explainability. While achieving high accuracy, they operate as black boxes and provide limited insight into why particular items are recommended, which reduces user trust and restricts their applicability in sensitive domains.

In this thesis, we investigate multi-modal graph-based learning approaches to address these challenges and improve both the accuracy and explainability of recommender systems. Specifically, we develop three models that target different aspects of the recommendation process.

In Chapter 2, we propose a framework that deeply integrates multi-modal information by constructing modality-specific relational graphs that jointly model user-item interactions and item-item relations derived from side information. By enabling information propagation across modalities, the model captures richer dependencies between users and items and improves robustness under sparse interaction settings.

In Chapter 3, we address the misalignment between behavioral and semantic representations by introducing a joint learning strategy that explicitly aligns interaction-based embeddings with similarity structures derived from side information. Through multi-task optimization, the model learns representations that are consistent

with both user behaviors and item semantics, leading to improved generalization when interaction data are limited.

In Chapter 4, we focus on enhancing explainability by enriching item semantics with large language model (LLM)-generated entities and introducing a personalized reasoning mechanism. By augmenting side information and estimating the causal influence of semantic entities on user preferences, the proposed approach improves both recommendation accuracy and transparency.

Experimental results on multiple real-world datasets demonstrate that the proposed methods consistently outperform state-of-the-art baselines, confirming their effectiveness in modeling sparse interactions, heterogeneous modalities, and explainable reasoning.

Finally, Chapter 5 summarizes the overall contributions of this thesis, discusses the limitations of the three proposed models, and outlines several promising future research directions.

様式７

論文審査の結果の要旨及び担当者

氏　名（　Shengzhe Jiao　）			
論文審査担当者		（職）	氏　名
	主 査	教授	原　隆浩
	副 査	准教授	佐々木　勇和
	副 査	准教授	大倉　史生
	副 査	教授	五十部　孝典
	副 査	教授	鬼塚　真
	副 査	教授	中野　珠実
	副 査	教授	村田　忠彦
	副 査	教授	伊達　進

論文審査の結果の要旨

オンラインプラットフォームの急速な成長と、それらが提供する製品やサービスの膨大な数により、自分の嗜好に合ったアイテムを見つけるのが困難になっている。この問題に対処するために、推薦システム（RS）が盛んに開発されている。推薦システムは、閲覧、クリック、購入といったユーザーとアイテムのインタラクションをモデル化し、パーソナライズされたレコメンデーションを生成する。RSは、現代のオンラインエコシステムの基本的な構成要素となっている。

近年の深層学習技術の進歩により、RSはユーザーとアイテム間の複雑かつ非線形な関係を捉える能力が大幅に向上した。しかし、その有効性は、インタラクションデータのスパース性によって制限される。データのスパース性を軽減するために、画像、テキストによる説明、カテゴリ属性などの追加のサイド情報を組み込むマルチモーダル推薦システムが提案されている。しかし、既存手法の多くは、各モダリティを個別に処理し、単純な融合戦略によってそれらを組み合わせている。このような設計では、モダリティ間の根本的な関係性を捉えることができず、アイテム間の視覚的な類似性やテキストの相関関係といったモダリティ固有の関係を見落とし、不完全なアイテム表現につながる。さらに、インタラクションベースの埋め込みとコンテンツベースの特徴は別々に学習されるため、実際のユーザーの嗜好とは意味的に矛盾する表現が生成される可能性がある。また、既存のRSは一般的に説明可能性に欠けている。

本論文では、これらの課題に対処し、レコメンデーションシステムの精度と説明可能性の両方を向上させるために、マルチモーダルグラフベース学習アプローチに取り組んだ。具体的には、推薦プロセスの異なる側面をターゲットとする以下の3つのモデルを開発した。

1. ユーザーとアイテムのインタラクションとサイド情報から得られるアイテム間の関係を共同でモデル化する、モダリティ固有の関係グラフを構築することにより、マルチモーダル情報を統合するフレームワークを提案した。モダリティ間での情報伝播を可能にすることで、このモデルはユーザーとアイテム間のより豊富な依存関係を捉え、疎なインタラクション環境下における堅牢性を向上させた。
2. インタラクションベースの埋め込みをサイド情報から得られる類似性構造と明示的に整合させる共同学習戦略を導入することで、行動表現と意味表現の不整合に対処した。マルチタスク最適化を通じて、モデルはユーザー行動とアイテムの意味の両方に一致する表現を学習し、インタラクションデータが限られている場合でも汎化性能を向上させた。
3. 大規模言語モデル（LLM）生成エンティティを用いてアイテムの意味を拡充し、パーソナライズされた推論メカニズムを導入することで、説明可能性の向上に焦点を当てた。サイド情報を拡張し、意味エンティティがユーザーの嗜好に及ぼす因果的影響を推定することについて説明し、いくつかの有望な将来の研究の方向性を概説した。

以上のように、本論文は既存の推薦システムの課題を解決する先駆的な研究として、情報科学に寄与するところが大きい。よって本論文は博士（情報科学）の学位論文として価値のあるものと認める。