Hybrid Inductive Graph Method for Matrix Completion

Jayun Yong, Sookmyung Women’s University, South Korea
Chulyun Kim, Sookmyung Women’s University, South Korea*

https://orcid.org/0000-0002-6471-5334

ABSTRACT

The recommender system can be viewed as a matrix completion problem, which aims to predict unknown values within a matrix. Solutions to this problem are categorized into two approaches: transductive and inductive reasoning. In transductive reasoning, the model cannot be applied to new cases unseen during training. In contrast, IGMC, the state-of-the-art inductive algorithm, only requires subgraphs for target users and items, without needing any other content information. While the absence of a requirement for content information simplifies the model and enhances transferability to new tasks, incorporating content information could still improve the model’s performance. In this article, the authors introduce Hi-GMC, a hybrid version of the IGMC model that incorporates content information alongside users and items. They present a novel graph model to encapsulate the side information related to users and items and develop a learning method based on graph neural networks. This proposed method achieves state-of-the-art performance on the MovieLens-100K dataset for both warm and cold start scenarios.

KEYWORDS

Deep Learning, Graph Neural Networks, Hybrid Recommendation, Recommendation Systems

INTRODUCTION

In the vast ocean of information, navigating to the desired data has become a formidable challenge, thereby amplifying interest in recommender systems. These systems streamline the process for users, enabling them to discover the information they need more efficiently and swiftly, while also offering companies an avenue to enhance service engagement and foster business advantages (Zamanzadeh Darban & Valipour, 2022).

Recommender systems align predominantly with two main frameworks: content-based methods and collaborative filtering methods. Content-based methods suggest items that mirror the user’s historical preferences. In contrast, collaborative filtering works by leveraging the aggregated preferences of other users to predict what an individual might like. While content-based strategies might limit suggestions to items closely related to those previously preferred by the user, thus narrowing
the breadth of recommendations, collaborative filtering broadens the spectrum of suggestions by reaching beyond similarities in content alone (Almazro et al., 2010).

Among various approaches to implementing collaborative filtering, matrix completion—organizing a matrix with users on one axis and items on the other—has gained prominence (Candès & Recht, 2009). By hypothesizing that the rating matrix is of low rank, numerous leading matrix completion algorithms employ factorization techniques, which have proven to be highly effective. However, matrix factorization faces an inherent limitation: it is transductive (Koren et al., 2009). This means that the latent features learned from users and items in the given dataset cannot be applied to new users or items not seen during training. Consequently, changes in the rating matrix, such as updates or the addition of new data, often necessitate a full retraining to generate new embeddings. To overcome this challenge, various studies have introduced methods to achieve inductive matrix completion (Michalski, 1983). One notable study in this area is the inductive graph-based matrix completion (IGMC) model (Zhang & Chen, 2020). IGMC leverages a graph neural network (GNN) that processes 1-hop subgraphs around user-item pairs from the rating matrix, associating these subgraphs with their respective ratings to enable inductive generalization effectively.

Although the IGMC model successfully introduced inductiveness into matrix completion, it also underscored a critical limitation inherent in collaborative filtering models by strategically avoiding content information to capitalize on its inductive strengths. A critical challenge faced by IGMC is the cold-start problem. This occurs when new users or items are introduced into the system without enough interaction data to form a reliable recommendation. This issue primarily affects collaborative filtering algorithms, which depend on past interactions to make their recommendations. In the absence of such data, a collaborative algorithm is unable to make a certain recommendation. Even with a minimal amount of interaction data, while a collaborative algorithm may still be able to offer recommendations, the reliability and quality of these recommendations are likely to be substandard.

Another key limitation of IGMC lies in its inability to finely integrate and weigh the preferences of neighboring users. Consider the scenario, depicted in Figure 1, where movie B might be recommended to user A. The graphs in Figure 1 illustrate interactions between users and movies with a 1-to-5 rating scale. In the graph on the left, since user A and all other users rate movies highly together except for movie B, the preferences of the other users influence the decision to recommend movie B to user A in collaborative filtering. However, in the graph, ratings for movie B diverge significantly among other users, and thus its recommendation to user A becomes less clear-cut. The addition of content information, like movie genres, can change this situation. The graphs on the right incorporate genre information, suggesting that recommendations for movie B, identified as a romance, could differ on the basis of the genres of movies that other users have rated. Assuming the scenario where ratings for the same genre have more weight than ratings for different genres, movie B might be strongly recommended to user A in the graph at upper right, since the neighbor user’s rating for romance has more weight than that for thriller. Conversely, movie B might not be recommended to user A in the graph at lower right. Therefore, incorporating content information into collaborative filtering not only improves performance but also allows for more tailored recommendations.

In this study, we introduce a cutting-edge hybrid recommendation technique, termed hybrid inductive graph matrix completion (Hi-GMC), which enhances the accuracy of inductive matrix completion and its effectiveness in cold-start scenarios. Unlike IGMC, which foregoes the use of content information to optimize inductive generalization, Hi-GMC leverages additional content-related clues alongside user-item interaction data. On the MovieLens 100K dataset (Miller et al., 2003), Hi-GMC outperformed most contemporary leading models (Code, 2022), establishing a new state of the art in recommendation system performance.

In the subsequent sections of this paper, we first review literature pertinent to our study before detailing the methodologies and mathematical formulations underpinning IGMC. In the method section, we elaborate on the approaches and mathematical frameworks that form the foundation of IGMC, offering comprehensive insights into the mechanisms driving our proposed model. This