DOI: 10.54254/2755-2721/5/20230700

The advance of recommendation system with graph neural network

Hui Rao

School of Management, University College London, London, WC1E 6BT, United Kingdom

zczlrao@ucl.ac.uk

Abstract. The field of e-commerce provides customers with many kinds of goods to choose from. It is also a challenge for consumers to choose among tens of thousands of product categories. The emergence of recommender systems solves such problems for consumers to a certain extent because it would recommend products that may fulfil their needs for consumers. Finding user preferences or rankings for a specific item can be aided by recommender systems. It uses specialized software and techniques to offer suggestions for items utilized by various users. To increase customer satisfaction or sell different things for e-commerce platforms, recommender systems advise consumers about specific items they may be interested in or desire to utilize. Graph neural networks generalize effectively to new datasets because they can learn the graph structure of various datasets. To put it another way, this makes Graph Neural Networks (GNNs) the perfect recommender systems for various real-world issues. As a result, the growth and usage of neural graph networks in recommender systems and the limitations of graph neural networks will be covered in detail in this paper.

Keywords: recommendation system, gragh neural network, machine learning.

1. Introduction

Considering how quickly social media and e-commerce have developed, recommender systems are now a crucial tool for many companies. Depending on the business, they can be found in a variety of ways, such as product recommendations on online retail websites, like Amazon, or playlist makers on video and music services, for instance YouTube, Netflix platform. Users rely on recommenders to explore a wide ocean of things that interest them (products, movies, news, or restaurants), helping to reduce the issue of information overload [1]. Therefore, a key component of efficient recommenders is precisely modelling users' preferences and their prior interactions (clicks, views, readings, and purchases).

The recommender systems of most of the data is graph-based by nature. A bipartite graph connecting user and item nodes can be used to describe interaction data in recommender applications, with observed interactions represented by links. Even user behaviour sequence item transitions can be built as graphs. When using structured external data, the advantages of creating recommendations as tasks on the graph become very clear. For example, the social relationship of users and the knowledge graphs of item relation. To model the rich, the recommender systems of heterogeneous data in this manner, the learning of graph gives a unifying viewpoint. The connections between nodes were first described using graph embedding techniques in recommender systems based on graph learning. These can also be separated

DOI: 10.54254/2755-2721/5/20230700

into decomposition, distributed representation, and neural embedding approaches. The remarkable ability of GNNs to comprehend graph-structured data has led to the emergence of numerous GNN-based recommendation models.

However, GNNs are extensively used in recommender systems for many reasons. Still, one of the main ones is that they offer a uniform framework to handle the substantial amount of data in recommender applications. Another reason is that GNNs can naturally and openly encode important collaboration signals (such as topology) to enhance user and item representations, unlike conventional approaches that only implicitly capture collaboration signals (e.g., employ user-item interactions as supervisory signals). Early efforts have shown that interactive things are useful in teaching user representation. These earlier works can enhance user representation learning by utilizing one-hop neighbours given the user-item interaction graph. GNN's advantage is that it offers strong, methodical methods for investigating multi-hop correlations, which are advantageous for recommender systems.

These advantages have contributed to GNN's notoriety in recommender systems over the past few years. Numerous academic studies show that, when applied to open benchmark datasets, GNN-based models outperform traditional methods and yield cutting-edge results. Numerous proposed modifications have been made to several different appraisal system, including session-based recommendation, point of interest (POI) recommendation, community suggestion, interactive suggestion, and package recommendation.

Research on recommender systems is frequently broken down into several tasks based on the type of information utilised to learn user/item representations. By simply utilising user-item interactions, or simultaneously learning user/item representations from paired data, user-item collaborative filtering suggestion tries to capture collaborative signals.

This paper first presents the most common and well-liked collaborative filtering method in recommender systems and then details Collaborative Filtering (CF) and prediction models employing various CF methods. Graph neural networks are frequently utilised to solve recommender systems and have a certain degree of benefit in learning efficient users or goods from interaction and auxiliary information. Finally, while adopting either CF or GNN has helped the recommender system's accuracy to a certain level, many areas might still improve the recommender system.

2. Preliminary

By offering individualised, exclusive content and service recommendations, recommender systems address the issue of information overload that users frequently experience. Recent developments in recommender system development have used collaborative, content-based, or hybrid filtering to create a variety of methods. The most popular and advanced filtering method is collaborative filtering. By locating persons with comparable tastes, collaborative filtering makes recommendations for products. It uses the opinions of active users to suggest products to them. Amazon enhances its suggestions by using topic diversification algorithms. The system generates an offline table of related items using an item-to-item matrix, which it then employs as part of a collaborative filtering strategy to solve scaling problems. Based on the user's past purchases, the algorithm then suggests further online products that are comparable.

Technologies based on content relate content resources to user attributes. Predictions are made using user data through content-based filtering approaches, which disregard posts by other users. The system uses a user interface to help users navigate the Internet; it can track users' browsing patterns to predict which pages they might be interested in. It usually generates more meaningful recommendations using different models to find similarities between documents. (To model associations between several documents, vector space models like Term Frequency Inverse Document Frequency (TF/IDF) or probabilistic models like Naive Bayes Classifiers, Decision Trees, or Neural Networks can be used). Users can recommend new items without leaving a rating. In addition, it is capable of short-term adjustments as user preferences change. This technique is used on datasets without including user preferences. Therefore, users gain recommendations without sharing their data to ensure user privacy. However, before they can be recommended to consumers, which is known as limited content analysis,

DOI: 10.54254/2755-2721/5/20230700

they must have thorough and sufficient item descriptions and well-organized user profiles. Descriptive data must therefore be available for CBF to be successful. The over-specialization of content in CBF is a significant issue. Users can only receive suggestions for things that match those listed in their profile as a consequence.

Building a database of user preferences for items is how collaborative filtering technology operates (user-item matrix). Then, it compares user profiles to find others with similar interests and preferences and makes recommendations. The users form a community they call neighbours. The user is given recommendations for items that he has not really reviewed but that locals in his neighbourhood have given optimistic view to. Although the content may not be in the user's profile, CF technology can offer casual recommendations, which means it will still suggest topics that are pertinent to the person. Nevertheless, it may also bring up several issues: chilly starts issues. The recommender does not have enough information about the user or item to generate reliable predictions. This is one of the main problems influencing recommender system performance. Since the new user hasn't rated any items, the system would not know what he could be interested in, so the new person's or item's profile will be blank. The rate of computation normally increases linearly as users and objects increase. When the number of datasets increases, recommendation systems that work well when there are few datasets may not generate adequate recommendations. Application of recommendation algorithms that can effectively scale with the quantity of datasets in the database is therefore crucial. Singular value decomposition (SVD) methods, which can produce accurate and effective recommendations, are based on dimensionality reduction techniques that are used to overcome scaling concerns and speed up suggestion creation.

2.1. Naive Bayes Collaborative Filtering (NBCF)

The core of the research technique is the naive bayes classifier (NBC), a monitored multi-class classification algorithm based on applying the Bayes theorem and the "naive" concept of conditional independence between each pair of variables. It allows us to understand and prove the predictions of the model more simply and to be more competitive with other techniques. For NBC, determine a posterior probability $C = \{c_1, c_2, \dots, c_m\}$ for each potential class given a set of independent variables $X = \{x_1, x_2, \dots, x_n\}$. The user's rating of the item would be the independent variable, and the potential categories would represent each conceivable rating value. The prior probability $P(C)$ and likelihood $P(X|C)$ for each class, which are proportional to the posterior probability, are used by NBC to generate the classification score $P(C|X)$.

$$\mathrm{P(C|X)} \propto \mathrm{P(C)} \prod_{i=1}^{n} P(x_i|C) \tag{1}$$

$$\hat{y} = \ arg \max_{y} P(C|X) \tag{2}$$

2.2. Bayesian Non-negative Matrix Factorization (BNMF)

Over time, user preferences shift. One method of modeling shifts in preference is based on considering the temporal features' evolution. The optimization of opposing parameters is carried out through an evolutionary clustering technique. An effective incremental CF system with a meager computational cost is developed based on the weighted clustering method. By grouping related products together, the suggestion of e-commerce products can be enhanced. The recommendation process is then completed using the created clusters [2]. The network model is formed by associating each user with each item; as a result, the individual's state changes over time. Higher scorers congregate together (find out groups of users with the same taste).

2.3. Biclustering Hybrid Collaborative Filtering (Bi-CF)

By classifying consumers into various clusters according to their purchases, recommend new items that are more likely to purchase. E-commerce businesses can boost income by bringing customers' attention to products they may buy by making recommendations based on comparable interests [3]. According to studies, combining users and items simultaneously on both dimensions indicates this dualism between them. This algorithm matches a portion of the user's preferences using a novel similarity metric [4]. The metrics are shown in formula (3).

$$sim(u,b) = \frac{|I_u \cap I_b|}{I_b} \tag{3}$$

2.4. Gaussian-Gamma Model (GGM)

Collaborative filtering recommender systems may perform less well in distributed settings when the data is scarce. A straightforward recommender that explores two distinct probabilistic modeling methods using the Naive Bayes technique. The findings demonstrate the effectiveness of the Gaussian model for assessing behavior, and after including a Gaussian gamma prior, it continues to perform well even with sparse data [5]. The Gaussian model's expression is shown in formulas (4), (5), and (6).

$$\mathrm{X}|\mathrm{T} \sim \mathrm{N}(\mu, 1/(\lambda\mathrm{T})) \tag{4}$$

$$\mathrm{T}|\alpha,\beta \sim \mathrm{Gamma}(\alpha,\beta) \tag{5}$$

$$(\mathrm{X},\mathrm{T}) \sim \mathrm{NormalGamma}(\mu,\lambda,\alpha,\beta) \tag{6}$$

2.5. Improved Naïve Bayesian Method (INBM)

The Naive Bayes method is the foundation of the INMB algorithm, a collaborative filtering recommendation system. The method is used in cases when the conditional independence requirement is not strictly upheld, in contrast to the original Naive Bayes method [6].

2.6. Non-Negative Matrix Factorization (NMF)

The learning rate changes after each parameter update to cancel out the negative components, preserving only the non-negative components. This is to make sure NMF satisfies the non-negativity condition. According to the convergence premise, initial non-negative parameters are fitted to the training set and remain non-negative after such training [7]. The process of the NMF algorithm is shown in formulas (7) and (8). NMF permits the use of positive elements, which makes the interpretation of experimental results easier.

$$H_{|i,j|}^{n+1} \leftarrow H_{|i,j|}^{n} \frac{((W^n)^T V)_{|i,j|}}{((W^n)^T W^n H^n)_{|i,j|}} \tag{7}$$

$$W_{|i,j|}^{n+1} \leftarrow W_{|i,j|}^{n} \frac{(V(H^{n+1})^T)_{|i,j|}}{(W^n H^{n+1}(H^{n+1})^T)_{|i,j|}} \tag{8}$$

2.7. Directed/undirected graphs

A directed graph is one in which every edge connects every node to every other node. When two nodes are connected, there are two opposite edges in an undirected graph, which is a particular case of a directed graph. Homogeneous and heterogeneous graphs have the same number of nodes and edges, whereas homogeneous graphs have a variety of nodes and edges.

2.8. Hypergraph

A graph that allows any number of vertices to be connected by an edge is known as a hypergraph. The fundamental idea behind GNN is to combine the current centre node representation with the aggregated feature data during propagation. When doing this, graph data are used. From the standpoint of network design, GNNs stack several propagation layers that are made up of aggregation and update procedures.

DOI: 10.54254/2755-2721/5/20230700

3. Recommendation System with Graph Neural Network

The key challenge for recommender systems is learning efficient user or item representations from transactions and auxiliary data. In recent years, GNN technology has proliferated in recommender systems since most of the information in these systems has a graph structure by nature and that GNNs have benefits in learning graph representations.

To iteratively aggregate information from neighbours, graph convolution networks (GCN) simulate the first order eigen decomposition of the graph Laplacian. The core premise behind the GCN network is that it employs a graph convolution operation to consider both node attributes and neighbourhood attributes, allowing the model to improve its regional representation of each node and achieve state-of-the-art performance in tasks involving the data of graph-structure. Based on the GCN concept, the study creates a GCN recommendation model for a traveling DCRS application that generates the data of graph structure to integrate hotels, bars, and other travel facilities to provide customers with complementary services.

The Graph Attention Network (GAT) updates each node's vector while distinguishing the contributions of its neighbours by using an attention mechanism, assuming that the influence of neighbours is neither identical nor predefined by the graph topology [8]. Different models have been suggested to successfully learn their patterns for better recommendation results due to the particularity of diverse data kinds in the suggestion, which is a significant issue for model construction. All these tasks can be resolved utilising a single GAT framework when looking at the data in suggestions from a graph perspective. GAT can learn both node representation and graph representation. Also, incorporating additional information is more convenient and flexible than non-graphical perspectives. For example, social networks are integrated into user-item dichotomies as a unified graph. During iterative propagation, both social influence and collaboration signals are captured. Furthermore, to improve user/item representations through the propagation process, GAT explicitly encodes crucial collaboration signals of user-item interactions. It's not a novel concept to use collaborative signals to enhance learning representation. GAT is easier to use and more adaptable than non-graphical models. GAT is more adaptable and practical for modelling multi-hop networks from user-item interactions than non-graphical models. It has been demonstrated that CF signals acquired in high-hop neighbours are useful for referrals.

Since most of the information in the recommender system is graph-structured, graph neural network technology has a certain degree of advantage in learning with graph-structure. In addition, the recommendation models of GCN and GAT are also explained above section. The principles and benefits of various types of neural networks will be thoroughly presented in the paragraphs that follow.

3.1. Graph Attention network-based RS (GATRS)

To discriminately learn the various levels of relevance and effect of other users (things) and target users (items) on a given graph, focus mechanisms are included into GNNs via graph attention networks (GAT). Based on GAT, GATRS is a tool for specifically learning inter user or inter item relationships. In this instance, emphasising more crucial individuals or goods is advantageous for recommendations since it highlights the influence of a particular user or item, which is more in accordance with the actual scenario. Due of its exceptional discriminative ability, GAT is frequently utilised in a variety of graph types, such as social graphs, item discussion graphs, and knowledge graphs.

3.2. Gated Graph Neural Network based RS (GGNNRS)

To develop optimum node representations by iteratively absorbing the effect of other nodes in the graph, Gated Graph Neural Networks (GGNNs) introduce Gated Recurrent Units (GRUs) into GNNs. This fully captures the interactions between nodes. GGNNRS builds on GGNN and learns user or item embeddings for recommendation by carefully considering intricate interactions between users or items. Because they can capture complicated links between nodes and have been attaining good recommendation performance, in several sectors, GGNNs are used to simulate intricate connections between various product categories or to describe complicated transitions between items in sequence graphs for sequential suggestion.

3.3. Graph Convolutional Network based RS (GCNRS)

By utilising graph structure and node feature information, Graph Convolutional Networks (GCNs) can often learn how to repeatedly collect feature information from local map neighbours [9]. In general, GCNs may efficiently aggregate data from neighbours in the network to build information embeddings for people and items by utilising convolution and pooling processes. GCNRS is based on GCN and uses the intricate relationships between persons or things and their content information for recommendations while learning how to incorporate them in graphs.

Due to their excellent extraction of features and learning capabilities, specifically their benefits in combining the structure of graph and the information about node content, GCNs have been widely used in RS for various graph creation GCNRS. This has been found to be extremely successful. For instance, the problem of data sparsity is solved by collaborative filtering, the social graphs of impact dissemination in social recommendation, mining the invisible connection data between user and item on interaction graphs, and gathering these data connections by excavating their associations correlation attributes on the graph of knowledge.

4. Prospects and challenges

User Portrait Concept. A user portrait is a digital depiction of actual users and a model about target users derived from a collection of real data. It was first proposed by Alan Cooper, who is known as the pioneer of interface design. User roles are another name for user portraits. To offer customers high-quality, individualised, and informed knowledge resource services, user data must be collected and mined to identify the feature tags that most accurately define the user [10]. User photos are being employed more and more frequently in a variety of industries lately, particularly e-commerce. Since user portraits can accomplish the following objectives. First, targeted marketing involves researching potential customers to send emails and text messages. Second, user analysis directs product improvement and even feature personalization. Third, offer customized searches and recommendations based on customer requirements. Fourth, in some way, videos and user descriptions are mined along with user tag data and videos to draw connections between user portraiture and company decisions, ranking statistics, geographical analysis, industry trends, and competitor product analysis. This study suggests using movie reviews to compare films based on a user's characteristics and actions. To arrive at the suggestion results, it then compares the user's interest traits with the browsed movies and movie datasets. This strategy aims to solve the personalised suggestions' scant data issue.

Recommender systems still have a few flaws that need to be fixed in the future—first, the ambiguity surrounding user preferences. For instance, it would be difficult to effectively recommend content to a new user based on their user history. Second, the recommender system still has a few minor issues in large-scale data graphs. Because there could be more than a billion nodes and edges in the industrial realm. The recommender system still finds processing such a large amount of data challenges. Third, the connections between the things or occasions that consumers are interested in change throughout time. Therefore, one area that might be further researched in the future is implementing and modelling the recommender system in dynamic scenarios. Finally, collaborative filtering algorithms primarily rely on complete user information to produce generally accurate recommendation outcomes. It isn't easy to discern some movies with identical names in the realm of cinema. Therefore, to further increase the accuracy of recommendation results, it is also essential to distinguish between item similarities.

5. Conclusion

By offering customers individualized, exclusive content and service recommendations, recommender systems address the issue of information overload that users frequently experience. They can also help platforms enhance revenue and customer satisfaction. Collaborative filtering is one of recommender systems' most extensively adopted and well-liked techniques. One of the collaborative filtering

DOI: 10.54254/2755-2721/5/20230700

approaches is the hypergraph, a generalization of a graph where one edge connects any number of vertices. In addition, the fundamental principle of a GNN with a graph structure is to propagate feature information that has been iteratively gathered from neighbours and combined with the current centre node representation. Therefore, GNN provides clear advantages and benefits in graph representation. Many parts of recommender systems are still limited and need to be investigated and improved in the future, even though their use is expanding, and more innovative techniques are being developed.

References

[1] Gong, Wu Ye, H., & Tan, H. S. (2009). Combining Memory-Based and Model-Based Collaborative Filtering in Recommender System. 2009 Pacific-Asia Conference on Circuits, Communications and Systems, 690–693.

[2] Bobadilla, Bojorque, R., Hernando Esteban, A., & Hurtado, R. (2018). Recommender Systems Clustering Using Bayesian Non Negative Matrix Factorization. IEEE Access, 6, 3549–3564.

[3] RuLong Zhu, & SongJie Gong. (2009). Analyzing of collaborative filtering using clustering technology. 2009 ISECS International Colloquium on Computing, Communication, Control, and Management, 4, 57–59.

[4] Symeonidis, Nanopoulos, A., Papadopoulos, A. N., & Manolopoulos, Y. (2007). Nearest-biclusters collaborative filtering based on constant and coherent values. Information Retrieval (Boston), 11(1), 51–75.

[5] Barnard, T., & Prügel-Bennett, A. (2011). Experiments in Bayesian recommendation. In Advances in Intelligent Web Mastering–3 (pp. 39-48). Springer, Berlin, Heidelberg.

[6] Wang, & Tan, Y. (2011). A New Collaborative Filtering Recommendation Approach Based on Naive Bayesian Method. Advances in Swarm Intelligence, 6729(2), 218–227.

[7] Luo, Zhou, M., Xia, Y., & Zhu, Q. (2014). An Efficient Non-Negative Matrix-Factorization Based Approach to Collaborative Filtering for Recommender Systems. IEEE Transactions on Industrial Informatics, 10(2), 1273–1284.

[8] Wu, Sun, F., Zhang, W., Xie, X., & Cui, B. (2022). Graph Neural Networks in Recommender Systems: A Survey. ACM Computing Surveys.

[9] Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2019). Graph convolutional networks: a comprehensive review. Computational Social Networks, 6(1), 1-23.

[10] Yao, W., Hou, Q., Wang, J., Lin, H., Li, X., & Wang, X. (2019, July). A personalized recommendation system based on user portrait. In Proceedings of the 2019 international conference on artificial intelligence and computer science (pp. 341-347).