Meta-path Guided Heterogeneous Graph Neural Network For Dish Recommendation System

Jiarui Zhang¹, Yonghua Zhu²

¹School of Computer Engineering and Technology, Shanghai University, Shanghai, 200444, China
²Shanghai Film Academy, Shanghai University, Shanghai, 200072, China
¹email: Astrau@shu.edu.cn, ²email: zyh@shu.edu.cn

Abstract: The traditional recommendation system can hardly utilize the ingredients and flavor characteristics of the dishes, and faces the problems of sparse data and cold start, resulting in inaccurate results of the recommendation system. This paper addresses these problems by constructing a heterogeneous graph network with the interaction data between users and dishes in the food domain and taking the main and auxiliary ingredients as nodes. Also, we capture the higher-order structural information among users, dishes, and main and auxiliary ingredients by using meta-path guidance. In the meantime, we assign weights to the edges associated with user nodes and main and auxiliary ingredients nodes, which can obtain users' preferences for main and auxiliary ingredients by using weighted GCN networks. The experiments conducted on a food domain dataset demonstrate that the meta-path-guided heterogeneous graph dish recommendation algorithm proposed in this paper is improved over the traditional recommendation algorithm.

1. Introduction

In the field of recommendation systems, recommendation of advertisements and commodities is the mainstream research direction, and there are relatively few recommendation studies in the specific field of dishes and foods. On the one hand, it is difficult for the early models to use the characteristics of ingredients and flavors of dishes, and the recommendation system based on collaborative filtering is not effective for the relatively sparse information of users' dietary records; on the other hand, users' dietary records have a high frequency and repetition rate in the category of dishes, which is the type of data distribution that these algorithms are not suitable for.

In recent years, heterogeneous graph networks consisting of multiple types of nodes or links have been used as a powerful modeling approach to fuse complex information and have been successfully applied to many recommendation system tasks. In this paper, we improve and combine various existing models based on the characteristics of users' dietary records, and re-model the heterogeneous graph of user diet records using the main and auxiliary ingredient information of dishes, which alleviates the sparsity of user diet records to some extent; meanwhile, we propose a dish recommendation method based on meta-path-guided heterogeneous graph neural network, which fuses multiple meta-paths of users with dishes, main and auxiliary ingredients. The meta-paths constructed based on the main and auxiliary ingredients classify the categories of dishes, user preferences for ingredients and flavors, and better capture the rich higher-order structural information on the heterogeneous graph; In addition, we add weighted edges to the meta-paths between user and main ingredients, user and auxiliary ingredients,
which improves the accuracy and interpretability of the model. The experimental results show that the model in this paper has improved the AUC and ACC metrics compared with the baseline model.

2. Related work
Many studies related to dish recommendation methods are direct applications of advertising and product recommendation methods in the field of dish recommendation. Mathur et al [11] used graph proximity and graph mining algorithm to perform recommendation. Jin et al [12] used alternating least squares to perform matrix factorization of user-dish items, and then used an association rule matching-based approach to further refine the initial screening recommendation list. Chen et al. [13] fused fuzzy C-mean clustering algorithm and review timeliness factor based on collaborative filtering recommendation. It can be found that in terms of the food domain, the traditional recommendation algorithms cannot make good use of the main and auxiliary ingredient information of the diet data, and the training effect is also poor for the relatively sparse user's dietary data. The heterogeneous graph neural network can solve this problem very well.

Graph neural networks are a new direction in recent years for the domain of recommender systems. Early researches have graph representation learning. For instance, DeepWalk [4] generates sequences from the network in random walks. In order to make the final embedding result express the local structure and global structure of the network, Node2vec[5] combines depth-first search and breadth-first search. On the other hand the study of heterogeneous graphs, Wang[7] et al. proposed the OptRank method to mitigate cold-start methods by using heterogeneous information in social tagging systems. After the concept of meta-path was added to recommender systems[9], Yu [10] et al. used meta-path-based similarity as regularization in a matrix factorization framework. These models are usually utilized in e-commerce recommendation scenarios. For the food domain, the main and auxiliary ingredient information is usually embedded directly as auxiliary information, which cannot learn the user's higher-order preferences (e.g., preferences for ingredients or flavors) well. In this paper, the user's diet record data is reconstructed using a heterogeneous graph approach, integrating the main and auxiliary ingredient nodes and using meta-path guidance for training, which enables the model to obtain richer interaction information between nodes. Also, we add weight information to the graph structure, modify the convolutional graph representation of the GCN network. Besides, we capture the rich higher-order structural information on the heterogeneous graph using the meta-path-guided model. With all that said, we improve the recommendation performance of graph neural networks in the food domain.

3. Method
The proposed GCN network-based heterogeneous graph dish recommendation system is shown in Figure 1 below. we focus on putting the user-dish interaction information into the heterogeneous graph neural network to train, at the same time fusing the different meta-paths to learn the higher-order representation of all aspects of the user and the dish for diet recommendation.

The contributions of this paper are as follows:(1) Since the traditional recommendation system models do not integrate well the main and auxiliary ingredients features of dishes, we model the dish recommendation scenario using heterogeneous graphs, constructs two meta-paths about main and auxiliary ingredients, and builds a heterogeneous graph neural network model to recommend according to the meta-path guidance (2) According to the dish data, there are a large number of condiments such as salt, soy sauce, and sugar (small amount) in the main and auxiliary ingredients that are not very relevant to dish recommendation, and these condiments will affect the subsequent learning of the model, so this paper assigns weights to the edges associated between users and main and auxiliary ingredients in the heterogeneous graph with the help of the concept of Term Frequency - Inverse Document Frequency [8], which improves the model accuracy and interpretability, and improves the graph convolution of GCN layer by the method of [14].
Figure 1. The framework of Meta-path Guided Heterogeneous Graph Neural Network For Dish Recommendation System

3.1. HIN for Dish Recommendation

In this paper, we propose to use heterogeneous graphs to model dish data, first we re-build a heterogeneous graph network, HIN \( G = (V, E) \) which consists of the main and auxiliary ingredients information of dishes as entities of additional nodes, including users \( U = \{u_1, u_2, \ldots\} \), dishes \( D = \{d_1, d_2, \ldots\} \), main ingredient \( M = \{m_1, m_2, \ldots\} \), auxiliary ingredients \( C = \{c_1, c_2, \ldots\} \), two new meta-paths UMU, UCU are constructed through the relationships between users and dishes, dishes and main and auxiliary ingredients.

Due to the huge information of main and auxiliary ingredients of the ingredients themselves, it slows down the sparseness of users’ dietary records on the one hand, and brings some problems on the other hand: the redundancy of main and auxiliary ingredients information is not strongly associated with users. For example, some household condiments such as salt, soy sauce, sugar (small amount), these condiments do not well reflect the flavor or ingredient characteristics of a dish, so in the meta-path UMU and meta-path UCU, we add weights to the neighboring edges of users on these two meta-paths with the help of the concept of term frequency–inverse document frequency (TF-IDF) [8], and finally utilize the normalization operation. Through experiments, our method can have more obvious results in the food dataset.

\[
TF\cdotIDF = TF\cdotIDF^* \\
\text{TF}_{i,j}^* = \frac{n_{i,j}}{\sum_j n_{i,j}} \\
\text{IDF}_{i,j}^* = \log \left( \frac{D_{\text{count}}}{\left| \{ t_i \in [M_j \text{ or } C_j] \} + 1 \right|} \right) \\
W_{i,j} = \frac{\text{TF}_{i,j}^* \cdot \text{IDF}_{i,j}^*}{\text{Count}_{i,j}}
\]

Here we have adjusted the meaning of the parameters in the TF-IDF formula so that the adjusted formula can represent the importance of main and auxiliary ingredient node \( i \) for user node \( j \), which means how much the user likes a certain main and auxiliary material. \( n_{i,j} \) indicates the number of dishes...
containing main and auxiliary ingredient $i$ that user $j$ has eaten, and $\sum_{k} h_{k,j}$ indicates all dishes that user $j$ has eaten. $D_{\text{count}}$ indicates the total number of dishes, $\{j: t_j \in \{M_i \text{ or } C_i\} + 1\}$ indicates the total number of dishes containing main and auxiliary ingredient $i + 1$, $\text{Count}_{i,j}$ indicates the total number of main and auxiliary ingredients associated with the user, and $W_{i,j}$ indicates the weight of user node $j$ regarding the upper edge of main and auxiliary ingredient $i$.

3.2. Metapath-guided Heterogeneous Graph Neural Network
Inspired by the basic idea of a GCN that generates object embeddings based on local neighbors [6], we propose a meta-path guided heterogeneous GNN. We use meta-paths to obtain the neighbors of various aspects related to the user's choice of the dish such as ingredients, flavors, and other factors, and the obtained embeddings of users and dishes are the aggregation of their neighbors under different meta-paths.

3.2.1. Graph Convolutional Network for Weighted Graphs (WGCN)
It is well known that GCN [6] is a multilayer neural network that works directly on homogeneous graphs and constitutes an embedding vector of nodes according to their neighborhood properties. Formally, consider a graph $G = (V, E)$, where $V$ and $E$ denote the set of nodes and edges, respectively. Let $X \in \mathbb{R}^{q \times v}$ be a matrix containing nodes with feature $x_v \in \mathbb{R}^q$ (each row $x$ is a feature vector of node $v$). For graph $G$, we introduce its adjacency matrix $A' = A + I$ (with that node itself and degree matrix $D$ where $D_{ii} = \sum_j A_{ij}$). The hierarchical propagation rule is: $H^{(l+1)} = \sigma(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} H^{(l)} W^{(l)})$.

Here $\tilde{A} = D^{-1/2} A D^{-1/2}$ denotes the symmetric normalized adjacency matrix. $W^{(l)}$ is the trainable weight matrix of a neural network layer. $\sigma(\cdot)$ denotes the activation function, such as ReLU. $H^{(0)} \in \mathbb{R}^{q \times v}$ denotes the hidden representation of the node at layer $l$. Initially $H^{(0)} = X$.

However, GCN cannot be used directly with heterogeneous weighted graphs. Therefore, in this paper, we design a meta-path guided heterogeneous graph model by referring to the undirected weighted graph GCN model in [14]. This GCN model (Hereinafter referred to as WGCN) can better handle weighted graphs and learn multi-resolution vertex representations by multi-layer graph convolution, which improves the performance of relevant recommendations.

When $L = L_{rw} = I - D^{-1} A$, the graph
convolution layer can be expressed as:

$$H^{(l+1)} = \sigma(\widetilde{D}^{-1}(A + \lambda I)H^{(l)} \cdot W^{(l)})$$  \hspace{1cm} (5)

Here the degree matrix is $$\widetilde{D} = \lambda + \sum_i A_{ii}$$

The parameter $$\lambda$$ controls the balance between the central vertex and its neighboring vertices. When $$\lambda$$ is larger, the central vertex will involve more convolution operations. If $$\lambda$$ is equal to zero, the center vertex does not contribute to its vertex convolution result. The experiments by [14] show that weighted graphs with weights between 0 and 1 and $$\lambda$$ taking values around 1 work best, and since the main auxiliary-user graph in this paper normalizes the weights, $$\lambda$$ is taken to be 0.95 in the experiments.

3.2.2. Semantic-level attention

Usually, each node in a heterogeneous graph contains multiple types of semantic information, and the node embedding of a particular semantics can only reflect the node in one way. To learn more comprehensive node embeddings, we need to fuse multiple semantics that can be represented by meta-paths. To address the challenges of meta-path selection and semantic fusion in heterogeneous graphs, a semantic-level attention is used to automatically learn the importance of different meta-paths. We use $$Z$$ as the semantic specific node embedding learned from the GCN. For each meta-path $$P = \{\phi_1, \phi_2, \cdots \phi_k\}$$, the importance can be expressed as:

$$w_{\phi} = \frac{1}{|V|} \sum_{i=1}^{V} q^\phi \cdot \tanh(W_i z_i^\phi + b)$$  \hspace{1cm} (6)

$$w_{\phi}$$ is a path-level attention vector. Then, we normalize the importance of all meta-paths by the softmax function.

$$\beta_{\phi} = \text{softmax}(w_{\phi}) = \frac{\exp(w_{\phi})}{\sum_{\phi \in \Phi} \exp(w_{\phi})}$$  \hspace{1cm} (7)

$$\beta_{\phi}$$ denotes the contribution of meta-path $$\phi_i$$ to a specific task $$X$$. Obviously the higher $$\beta_{\phi_i}$$, the more important the meta-path $$\phi_i$$ is. Using the learned weights as coefficients, we can fuse these semantic specific embeddings to obtain the final embedding $$Z$$ as shown in Equation 8 below.

$$Z = \sum_{i=1}^{p} \beta_{\phi_i} \cdot Z_{\phi_i}$$  \hspace{1cm} (8)

We obtain fused user embeddings or dish embeddings by aggregating embeddings based on different meta-paths $$P = \{\phi_1, \phi_2, \cdots \phi_k\}$$ of users or dishes as shown in Equations 9 and 10 below.

$$U = \sum_{i=1}^{p} \beta_{\phi_i} \cdot U_{\phi_i}$$  \hspace{1cm} (9)

$$D = \sum_{i=1}^{p} \beta_{\phi_i} \cdot U_{\phi_i}$$  \hspace{1cm} (10)

3.3 Optimization Objective

We feed these static features to the multilayer perceptron to obtain a representation of the static features $$T_y$$. Then, we stitch together the embeddings of users, dishes, and static features. Finally, we fuse the embeddings into the MLP layer to obtain the prediction scores $$\hat{y}_y$$. The following equation 11.

$$\hat{y}_y = \text{sigmoid}(f(U + D + T_y))$$  \hspace{1cm} (11)

where $$f$$ is the MLP layer with only one output, sigmoid is the sigmoid activation layer, and $$\oplus$$ is the embedded connection operation.

The defined loss function is the log loss function:

$$J = \sum_{i,j \in \mathbb{N}} \left( y_{ij} \log \hat{y}_{ij} + (1 - y_{ij}) \log (1 - \hat{y}_{ij}) \right)$$  \hspace{1cm} (12)
where \( y_g \) is the label of the instance (0 or 1) and \( y^+ \) and \( y^- \) denotes the set of positive and negative instances, respectively.

4. Results & Discussion
In this section, we present the details of the experimental setup and the results. To illustrate the effectiveness of our proposed model, we compare it with some baselines on the recommended tasks. The evaluation metrics are AUC and ACC.

4.1 Data Preprocess
The experiments of this model use the public dataset of Kaggle to verify the effectiveness of the recommendation method in this paper. Food.com Recipes and Interactions[16].

The dataset was processed as follows, considering the limitations of the experimental environment, we randomly selected 1/4 users and their selected recipes for training the data, and since the relationship between users and dishes was initially in rating format, we converted the ratings to binary feedback: ratings with 4-5 stars were converted to positive feedback (denoted as "1"), and other ratings were converted to negative feedback (denoted as "0"). After processing the datasets, we divided each dataset into training/validation/testing sets in a ratio of 8:1:1.

Table 1. Statistics for the FoodRecipe dataset.

| DataSet     | Relations(A-B) | Number of A | Number of B | Numberof A-B | MetaPath |
|-------------|----------------|-------------|-------------|--------------|----------|
| FoodRecipe  | User-Dish      | 11138       | 44566       | 179594       | UDU      |
|             | User-Main Ingredient | 11138       | 8534        | 164844       | UMU      |
|             | User-Condiment | 11138       | 32136       | 518696       | UCU      |
|             | Dish-Main Ingredient | 44566       | 8534        | 93517        | DMD      |
|             | Dish-Condiment | 44566       | 32136       | 317314       | DCD      |

4.2 Baseline methods and their parameter settings
In our comparison experiments, we compare our model with several existing methods used for click-through rate prediction.

1. LR [1]: The logistic regression model is the most widely used model during industrial practice before the popularity of deep network models.
2. PNN [2]: PNN incorporates vector product computation between the embedding layer and the deep network to capture higher-order interaction information between features.
3. DeepFM [3]: DeepFM model combines FM model and deep network model together.
4. DeepWalk [4][16]: Random walk is a classical method for embedding homogeneous graph networks, here we use[16] method to train.
5. Metapath2vec [15]: A heterogeneous graph embedding method that uses meta-path-based random walk and embeds heterogeneous graphs using skip-gram.
6. GCN [6]: Graph convolution network a semi-supervised neural network that considers a graph convolution model on homogeneous graphs. Here we transform the heterogeneous graph into a homogeneous graph for training and remove the dish property.
7. GCN+MetapathAtt: A GCN+metapath-guided approach without fusing weight information.
8. WGCN+MetapathAtt: The model in this paper replaces the GCN network with WGCN and bringing the weights on the main and auxiliary ingredients meta-paths into the neural network for training.

The models in this paper use Adam optimizer for parameter optimization, and the LR model uses FTRL for parameter optimization; the PNN and DeepFM models have three neural network layers of
400, 400, 400 dimensions respectively and are initialized using Xavier algorithm, DeepWalk, Metapath2vec and other models we set the window size We set the window size to 5, the walk length to 100, each node walk to 40, and the number of negative samples to 5. For a fair comparison, we set the embedding dimension of all the above algorithms to 64.

| Method           | AUC  | ACC  |
|------------------|------|------|
| LR               | 0.6705 | 0.6325   |
| PNN              | 0.7079 | 0.6621   |
| DeepFM           | 0.7096 | 0.6691   |
| DeepWalk         | 0.7184 | 0.6699   |
| Metapath2vec     | 0.7199 | 0.6735   |
| GCN              | 0.7233 | 0.6752   |
| GCN+MetapathAtt  | 0.7259 | 0.6782   |
| WGCN+MetapathAtt | 0.7388 | 0.6902   |

4.3 Results & Discussion

The experimental results are shown in Table 2 above, and it can be found that:

(1) Compared with other baseline models, the model of this paper, WGCN+MetapathAtt, performs better in both AUC and ACC in food domain data.

(2) The recommendations based on graph neural networks (Metapath2vec, GCN, GCN+MetapathAtt, WGCN+MetapathAtt) are all basically better than traditional deep learning (LR, PNN, DeepFM) recommendations, which can feedback that in the food domain, for the case of sparse user dietary records, traditional deep learning models are not able to fully exploit the interaction information between higher-order nodes, and graph neural networks are significantly better than traditional deep learning recommendations for capturing the higher-order correlations between users and dishes.

(3) For the heterogeneous graph dataset reconstructed in this paper, the model WGCN+MetapathAtt in this paper uses the new meta-path information and the weight information between main-and-auxiliary ingredients and users to improve the AUC by 1.77% compared with the GCN+MetapathAtt model, which indicates that the model in this paper can learn more accurate neighbor relationships. It can also fuse the importance of multiple metapaths, and the weighted graph outperforms the rest of the models in terms of interpretability.

5. Conclusions

In this paper, we address the problems that the existing recommendation system models in the food domain cannot adapt well to the sparsity of the diet data and cannot make effective use of the main and auxiliary ingredient information of food. We re-model the existing users' dietary records, and the TF-IDF is used to assign weights to the user-main and auxiliary ingredient element path edges, while the convolutional representation of the GCN layer is improved to adapt the weighted graph, and a semi-supervised heterogeneous graph neural network based on the attention mechanism is proposed. The model in this paper can capture the complex structure and rich semantics behind the heterogeneous graph, and capture the importance of and meta-paths using and semantic-level attention, respectively. Meanwhile, the improved WGCN layer can learn the weight information of heterogeneous graph edges,
which makes the recommendation effect more accurate. The experimental results show that the model in this paper is superior in terms of effectiveness and interpretability of recommendations. In the future, we consider finding the interaction information of higher-order structural features between user-dish-main and auxiliary ingredients in graph neural network to enhance the recommendation effect.

Acknowledgment
This work is supported by the National Key R & D Program topic "Food Flavor Perception and Preference Influence Mechanism and Construction of Chinese Flavor Map" (Project No. 2017YFD0400101-01).

References
[1] Chapelle O, Manavoglu E, and Rosales R. (2014). Simple and Scalable Response Prediction for Display Advertising. ACM Transactions on Intelligent Systems and Technology, 5(4):1-34.
[2] Qu Y, Cai H, Ren K, et al. (2016) Product-Based Neural Networks for User Response Prediction. In: IEEE 16th International Conference on Data Mining. Barcelona. pp. 1149-1154.
[3] Guo H, Tang R, Ye Y, et al. (2017) DeepFM: a factorization-machine based neural network for CTR prediction. In: The Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne. pp.1725-1731.
[4] Perozzi B, Alrfou R, Skiena S, et al. (2014) DeepWalk:online learning of social representations. In: The 20th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.New York. pp. 701-710.
[5] Grover A, Leskovec J. (2016) node2vec: Scalable Feature Learning for Networks. In: The 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. San Francisco. pp. 855–864.
[6] Thomas N, Kipf, Welling M.(2017).Semi-Supervised Classification with Graph Convolutional Networks. In: The 5th International Conference on Learning Representations.
[7] W. Feng and J. Wang. (2012) Incorporating heterogeneous information for personalized tag recommendation in social tagging systems. In:The 18th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.Beijing. pp.1276–1284.
[8] David M, Blei, Andrew Y Ng, and Michael I Jordan.(2003). Latent dirichlet allocation. Journal of Machine Learning Research, 3(6):993–1022.
[9] Sun Y, Han J, Yan X, et al. (2011) PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks. Proceedings of the Vldb Endowment, 4(11):992-1003.
[10] Yu X, Ren X,Gu Q, Sun Y, and Han J.(2013) Collaborative filtering with entity similarity regularization in heterogeneous information networks. In: International Joint Conference on Artificial Intelligence. Beijing.
[11] Mathur A, Juguru SK, Eirinaki M. (2019) A Graph-Based Recommender System for Food Products, In: 2019 First International Conference on Graph Computing. Laguna Hills. pp. 83-87.
[12] QianJ,XianChuan W, Liang P, et al. (2019) Alternate Least Squares And Rule Matching Based Dish Recommendation. Journal of Fuyang Normal University(Natural Science), 36(3):49-54.
[13] XiaoYu C, XiaoJin L, KeCheng L.(2015) Online food ordering recommendation algorithm based on fuzzy clustering and review timeliness. Journal of Nanyang Institute of Technology, 7(2):38-41.
[14] Zhang T, Liu B, Niu D, et al.(2018) Multiresolution Graph Attention Networks for Relevance Matching.In: The 27th ACM International Conference on Informationand Knowledge Management. Toronto. pp. 933-942.
[15] YuXiao D,Chawla, Nitesh V. C,Swami, Ananthram S.(2017). metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In: The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax. pp. 135-144.
[16] Majumder B P, Li S, Ni J, et al. (2019) Generating Personalized Recipes from Historical User Preferences. In: Proceedings of the 9th International Joint Conference on Natural Language Processing. Hong Kong. pp. 5975-5981.
[17] Wang J, Huang P, Zhao H, et al. (2018) Billion-scale Commodity Embedding for E-commerce Recommendation in Alibaba. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London. pp. 839-848