

Research Article

Anticancer Recipe Recommendation Based on Cancer Dietary Knowledge Graph

Jianchen Tang,¹ Bing Huang,² and Mingshan Xie ¹

¹College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China

²Department of Thoracic Oncology, Second Affiliated Hospital of Zunyi Medical University, Zunyi 563000, China

Correspondence should be addressed to Mingshan Xie; hnmingshanxie@163.com

Received 9 July 2023; Revised 14 September 2023; Accepted 28 September 2023; Published 18 October 2023

Academic Editor: Reza Izadpanah

Copyright © 2023 Jianchen Tang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Many recipes contain ingredients with various anticancer effects, which can help users to prevent cancer, as well as provide treatment for cancer patients, effectively slowing the disease. Existing recipe knowledge graph recommendation systems obtain entity feature representations by mining latent connections between recipes and between users and recipes to enhance the performance of the recommendation system. However, it ignores the influence of time on user taste preferences, fails to capture the dependency between them from the user's dietary records, and is unable to more accurately predict the user's future recipes. We use the KGAT to obtain the embedding representation of entities, considering the influence of time on users, and recipe recommendation can be viewed as a long-term sequence prediction, introducing LSTM networks to dynamically adjust users' personal taste preferences. Based on the user's dietary records, we infer the user's preference for the future diet. Combined with the cancer knowledge graph, we provide the user with diet recommendations that are beneficial to disease prevention and rehabilitation. To verify the effectiveness and rationality of PPKG, we compared it with three other recommendation algorithms on the self-created datasets, and the extensive experimental results demonstrate that our algorithm performance performs other algorithms, which confirmed the effectiveness of PPKG in dealing with sequence recommendation.

1. Introduction

Worldwide, cancer is a major public health problem [1]. The fight against cancer is one of the greatest challenges facing humanity. It is estimated that 30–40% of tumors can be through the right lifestyle and diet prevention [2]. Dietary factors are thought to account for approximately 30% of cancers in Western countries [3]. The contribution of diet to the risk of cancer is thought to be low in developing countries, perhaps around 20% [4]. It has been estimated that 30–40% of all cancers can be prevented by lifestyle and dietary measures alone [5]. Promoting physical activity and healthy diet is important in helping to manage non-communicable diseases (including cardiovascular disease, cancer, chronic respiratory disease, diabetes, or mental illness), burden, and reducing mortality are critical [6]. Knowledge graph techniques are applied to cancer by integrating structured and unstructured cancer resources to

build a knowledge graph. Query and reasoning capabilities of knowledge graphs support precision medicine and personalized treatment design. Bohlscheid-Thomas et al. [7] developed the food frequency questionnaire of the German part of the EPIC project, which provided an important basis for obtaining population diet structure data based on questionnaires, enabling nutritional epidemiology to more accurately study the relationship between diet and cancer risk and promote the development of cancer epidemiology. Key et al. [8] conducted a systematic review summarizing the relationship between diet, nutrition, and cancer prevention. The review systematically evaluated the impact of diet on cancer prevention and provided scientific evidence for diet and nutrition policies. Daowd et al. [9] discovered implicit semantic associations from the large-scale biomedical literature using deep learning techniques and constructed a causal knowledge graph for chronic diseases and cancer. This provides support for in-depth understanding of disease

pathogenesis, drug mechanisms of action, and so on, which is of great significance for future knowledge graph construction and personalized medicine. Zhu et al. [10] built an ovarian cancer knowledge graph by integrating data from multiple sources and successfully applied the knowledge graph to complete disease cause analysis and prediction tasks, providing reference for future applications of KG in cancer research and clinical practice. Diet and medicated diet are ideal and effective medical care measures. Diet can be used for the prevention and treatment of cancer; appropriate diet is helpful to prevent and cure the occurrence of cancer and promote the improvement of cancer and play a certain protective role in health. Compared with bitter drugs, a delicious medicinal diet is easier for patients to accept and adhere to long-term consumption, especially suitable for the regulation and treatment of chronic diseases such as tumor.

Knowledge graph has achieved good results in the recommendation system, such as in the fields of movies and music. Knowledge graph can be added as auxiliary information to the recommendation system to enhance the representation of items and users, thus improving the accuracy of recommendation. And the recommendation of recipes has also attracted more and more attention; existing recipe recommendation systems are generally based on content or collaborative filtering algorithms to recommend recipes to users. Freyne and Berkovsky [11] used the content strategy to simply decompose and construct to associate recipes and ingredients to achieve high coverage and reasonable accuracy of recommendation. Yuan and Luo [12] used the k-means clustering algorithm to divide the food set into multiple nonoverlapping subsets and then used the user-based collaborative filtering algorithm to recommend food that the user may like, with an accuracy of more than 70%. These recommendation systems ignore the connection between recipes and between recipes and users, and the recommendation results often fail to achieve the desired result. Different from the method of recommending recipes using content-based or collaborative filtering methods, the relationship information between users, recipes, and food is rarely explored. In recent years, it has become a mainstream research direction to integrate knowledge graph as auxiliary information into the recommendation system to improve the recommendation performance of the system. Care-Graph [13] has used knowledge graph to alleviate the cold start problem, improving the prediction accuracy under cold start, and the accuracy is 5% higher than the baseline model. Min et al. [14] proposed a unified food recommendation framework and determined the main problems affecting food recommendation, including integrating various background and domain knowledge graph, constructing personal models, analyzing unique food features, and expounding the research challenges and future directions in the recipe field.

Huang et al. [15] established a knowledge graph through a web crawler and constructed a diet knowledge graph integrating multidomain information by using the rich semantics of the knowledge graph. Huang et al. [16] integrated data from different sources and formats,

organized extracted knowledge into appropriate representations, and proposed a healthy diet knowledge graph construction model. A better method for knowledge management in intelligent healthy diet can be provided by the research model using machine learning and natural language processing methods. The user's personal preference is considered in the recommendation system, and the user's personal information is used as auxiliary information to make recommendations. RippleNet [17] considered its historical interests as a seed set in KG and then iteratively extended a user's interests along KG links to discover its hierarchical latent interests regarding candidates. Rastogi and Zaki [18] proposed the personal health knowledge graph to assist the recipe recommendation system and improve recommendation accuracy. Researchers enhance the embedding representation of the knowledge graph to improve the recommendation performance. Yuan et al. [19] adopted a translation-based model as the knowledge graph embedding method to learn the embedding representation of entities and then introduced these embeddings into the recommendation module to enrich the expression of items. Ma et al. [20] established a recommendation system based on knowledge graph attention to help learn fine-grained user and recipe embedding by modeling diverse user preferences from user behavior.

Some studies have made recommendations by analyzing potential associations between recipes. Gao et al. [21] proposed a new food recommendation model based on graph convolutional network (FGCN), which deeply explores the relationship between ingredients, ingredient recipes, and user recipes. FGCN adopts the information propagation mechanism and employs multiple embedding propagation layers to model the high-order connectivity of different food relations and enhance the representation. Tian Y. et al. [22] utilized relational information for recipe recommendation and proposed HGAT, a new hierarchical graph attention network for recipe recommendation. The model can through several neural network modules capture user historical behavior, recipe content, and relational information, including type-specific transitions, node-level attention, and relational-level attention. Lei et al. [23] adopted the method combining multimodal and hierarchical ideas and constructed a knowledge graph considering multiple factors as the center. It not only considered the potential demands of users but also excavated the deep relationships between users and recipes, and between users and recipes. The authors proposed a novel multimodal recipe recommendation approach based on multispect node representation and demand-based multirelational graph structure extraction of the knowledge graph.

Another approach is to enhance the performance of recommendation systems through interpretability. Semantic modeling [24] proposed the Food Explanation Ontology (FEO) for modeling the explanations of food-related recommendations for users, which can provide multiple explanations to accurately represent the explanations of food recommendations while preserving important semantic details. Y. Chen et al. [25] proposed a new food recommendation problem model, which modeled recipe

recommendation as a constrained question answering on a large-scale food knowledge graph and uniformly handled user dietary preferences and personalized needs of health guidelines as additional constraints of the QA system.

Li et al. [26] constructed recipe nutrition and user preference into two knowledge graphs, integrated recipe nutrition into the task of recipe representation and recommendation, used knowledge transfer scheme to realize the transfer of useful semantic information across preferences and health, and fused the important information of the two knowledge graphs, thus achieving the goal of recommending both “delicious” and “healthy” food for users.

All the methods proposed above ignored the influence of time on users’ taste preferences, and users’ daily recipe information cannot be extracted from users’ historical diet records, which may lead to the same recipe being recommended to users several times in a continuous period of time. Our proposed PPKG model not only provides the user with healthy recipes but also considers the user’s eating habits. It meets the personalized demands of users and recommends satisfactory recipes for users.

The main contributions of this paper are as follows:

- (1) The recipe recommendation is integrated with the time factor to fully consider the user’s taste preference.
- (2) We introduced LSTM [27] into the knowledge graph to dynamically predict the user’s preference and represent the recommendation in the way of sequence prediction to obtain the recipe recommendation that meets the user’s demands.
- (3) In order to effectively verify the importance of recommendation in the time factor, we conduct extensive experiments on our self-created datasets to demonstrate the effectiveness of our proposed PPKG model.

In the following, Section 2 describes the problem formalization, Section 3 describes the proposed framework, and Section 4 presents the experimental setup and analyzes the results. Finally, there are conclusions in Section 5.

2. Problem Formalization

The recommendation system requires two types of information, the attribute information of the items and the historical dietary records of the users. We use the public knowledge graph to represent the relationship between various entities. We define entities in sets as $e \in I$; the embedding representation of an entity defined $v_e \in R^d$, where v_e is the embedding of entity e and d is the dimension of the embedding. The knowledge graph is defined as $G = \{(h, r, t) | h \in \mathcal{E}, r \in \mathcal{R}, t \in \mathcal{E}\}$. The triplet denotes that there is a relationship r between the head entity h and the tail entity t (e.g., bitter melon, efficacy, and anticancer), and it denotes that bitter melon has the efficacy of anticancer, and explores potential relationships between entities in the knowledge graph by learning embedding of entities and relationships. We adopt heterogeneous graphs to represent

diverse nodes and relations. To obtain the embedding for the current node, we aggregate features from its neighbors of different relation types. This heterogeneous architecture allows effective feature propagation. We construct a knowledge graph with five different types of nodes and four different types of relationships between them. Then, the items of the knowledge graph are mapped to the recommendation system, and the recommendation result in the recommendation system is the recipe entity in the knowledge graph. We use formula (1) to illustrate the task of the recommendation system:

$$Ui.item = F(G, Ui), \quad (1)$$

where the input of function F is knowledge graph G , which contains various ingredients, recipes, and diseases, and Ui denotes the personal dietary records of the user i . The final output is the recipe recommended to the user.

3. Methodology

3.1. Personal Preference Knowledge Graph (PPKG). Our proposed PPKG model adopts two modules for the recommendation. The first module is the embedding representation modeling of users and items, as shown in Figure 1(a), and the second module is the recommendation of user taste preferences, that is, the model prediction module, as shown in Figure 1(b).

In the first module, we use the KGAT [28] architecture to learn the embedding representation of the items and we enhance the representation of the items to achieve more precise recommendations by representing the features of each node in the knowledge graph with message-passing and update functions. In the recommendation module, it is also necessary to consider the influence of time on users’ taste preferences. After obtaining entity embeddings in the knowledge graph, we incorporate an LSTM network. This captures users’ dietary habits from their historical dietary records. Then, we predict recipes which users may like based on their habit model. Ultimately, the recommended recipes will better match users’ taste preferences.

3.2. User and Item Embedding Modeling Module. In this section, we embed the users and entities in the knowledge graph, as shown in Figure 2. We consider that each recipe in the recipe knowledge graph has a variety of different therapeutic effects, and each effect can treat different symptoms. The recipe can be aggregated into therapeutic effect features, symptom features, and ingredients from different surrounding nodes. The feature representation of the current recipe is obtained. We adopt the TransR [29] algorithm to embed the structured information of entities and their rich relationships in the knowledge graph. We update the current node feature by combining the information of the surrounding nodes, as shown in the following equation:

$$h_i^{(l)} = \rho(\{e_j^{(l+1)} : j \in N(i)\}), \quad (2)$$

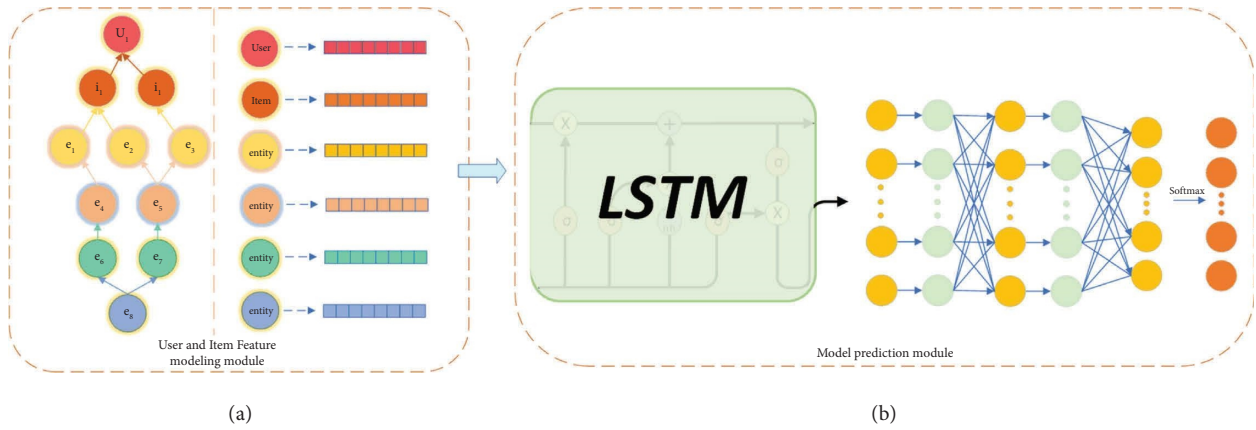


FIGURE 1: Illustration of the proposed PPKG.

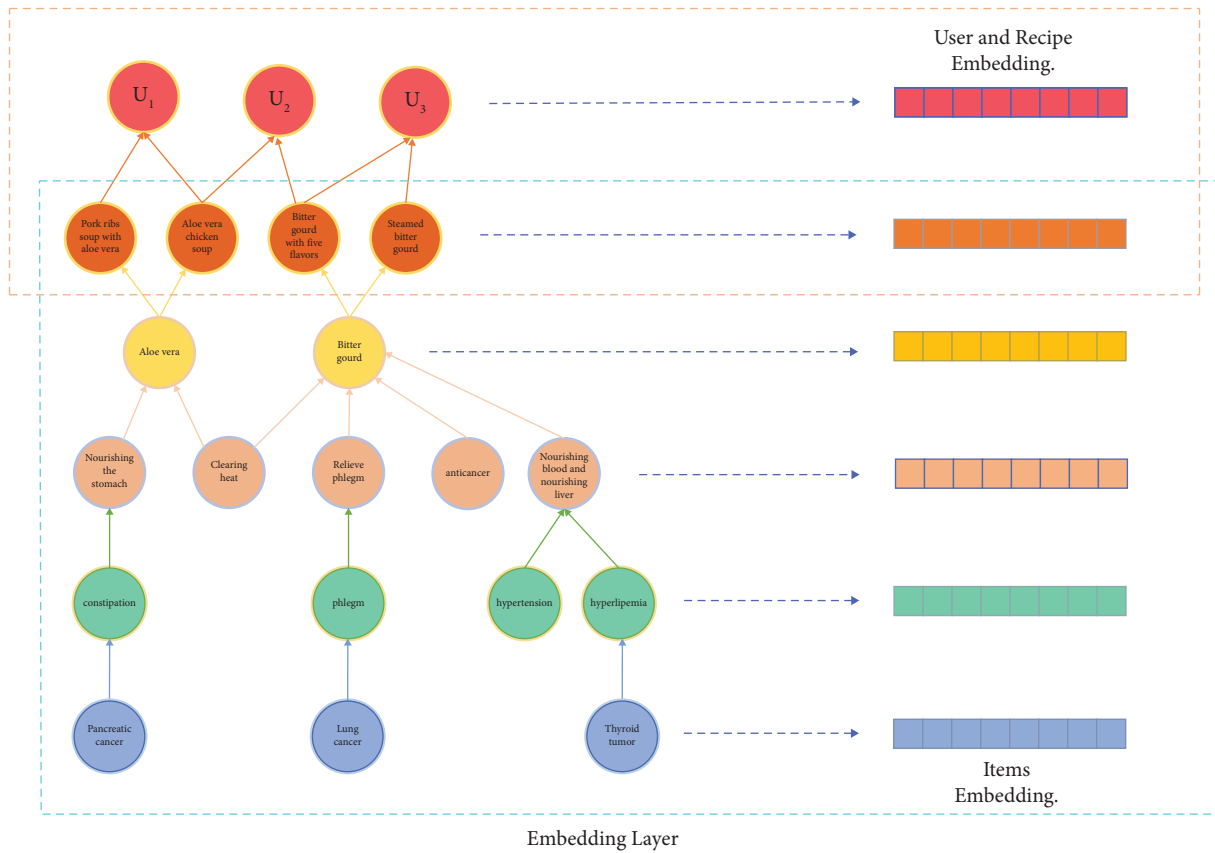


FIGURE 2: Illustration of user and item characteristics modeling module.

where $N(i)$ denotes the set of neighbor nodes connected to the node i , ρ denotes the aggregation function on surrounding nodes, and $h_i^{(l)}$ denotes the hidden feature representation at the l -th layer of the node i .

Then, we fuse the hidden features of l -th layer and the feature representation of $(l+1)$ -th layer of the node i , as shown in the following formula:

$$e_i^{(l)} = \phi(h_i^{(l)}, e_i^{(l+1)}), \quad (3)$$

where $e_i^{(l)}$ denotes the feature representation at the l -th layer of the node i , $e_i^{(l+1)}$ denotes the feature representation at $(l+1)$ -th layer of the node i , and ϕ denotes the aggregation mode of the node i . Here, we can choose three types of aggregators:

GCN aggregator is as follows [30]:

$$\phi_{\text{GCN}} = \text{LeakyReLU}(W(h_i^{(l)} + e_i^{(l+1)})), \quad (4)$$

where LeakyReLU [31] is used as the activation function, W is the weight matrix, $h_i^{(l)}$ is the hidden feature of the node l -th layer, and $e_i^{(l+1)}$ is the feature representation of the node $(l+1)$ -th layer.

GraphSage aggregator is as follows [32]:

$$\phi_{Bi\text{-interaction}} = \text{LeakyReLU}(W_1(h_i^{(l)} + e_i^{(l+1)})) + \text{LeakyReLU}(W_2(h_i^{(l)} \odot e_i^{(l+1)})), \quad (6)$$

where W_1 and W_2 denote the weight parameter matrix and \odot denotes the multiplication of the corresponding position elements of the matrix. We aggregate the features of the surrounding nodes into the hidden layer and then update the embedding with the features of the previous layer of node i , and finally obtain the final representation of the features of the current layer of node i . In this way, recipe nodes are able to get information transmitted from symptoms and efficacy and get richer feature representations. In the triplet of the knowledge graph, the relationship between each entity are defined as following formula:

$$\begin{aligned} \varphi(h, r, t) &= \|h + r - t\|_2^2, \\ \varphi(h, r, t') &= \|h + r - t'\|_2^2, \end{aligned} \quad (7)$$

where (h, r, t) denotes the true triplet, (h, r, t') denotes the triplet sampled from the negative sample, and $\varphi(h, r, t)$ denotes the credibility score of the triplet, which is between 0 and 1, and close to 1 indicates high credibility between the triplet.

Through the interactive relationship between users and items, the relationship between u_i and $item_i$ is formulated as follows:

$$\eta = \begin{cases} \delta(u_i, i), & i \in i_j^+, \\ \delta(u_i, i), & i \in i_j^-, \end{cases} \quad (8)$$

where u_i denotes the user i , i_j^+ denotes the recipe sets that the user likes, i_j^- denotes the recipe sets that the user does not like, δ denotes the function of the inner product of user and item representations, and η denotes the score between the user and the entity, which is between 0 and 1, and close to 1 indicates that u_i likes $item_i$ more.

3.3. Model Prediction Module. In Section 3.2, we model the interaction between users and items and the relationship between items and entities in the knowledge graph. After obtaining the embedding representations for each item, we incorporate temporal factors in the user's history. We represent each recipe in the user's diet history with its entity embedding from the knowledge graph. The temporal element enables modeling the evolution of user preferences. As shown in Figure 3, the historical dietary records are taken as input, and LSTM is used to extract the long-term dependencies between dietary records, and the output is obtained through LSTM, which is formulated as follows:

$$\phi_{\text{GraphSage}} = \text{LeakyReLU}(W(h_i^{(l)} \| e_i^{(l+1)})), \quad (5)$$

where $\|$ denotes the concatenation operation.

Bi-interaction aggregator is as follows [28]:

$$v = \text{LSTM}(u_i, G), \quad (9)$$

where v denotes the output of the LSTM, u_i denotes the historical dietary records of the user i , and G denotes the embedded representation of each entity in the recipe knowledge graph.

As shown in Figure 4, we concatenate the output vectors of the LSTM through the connection operation, took them as the input of the fully connected layer, and then extract all the features through the fully connected layer, which is formulated as follows:

$$\begin{aligned} \chi &= \text{concate}(v), \\ \pi_i &= f(\omega_i \cdot \chi + b_i), \end{aligned} \quad (10)$$

where χ denotes the input of the i -th hidden layer, f denotes the RELU [33] activation function, ω_i denotes the weight parameter of the i -th hidden layer, b_i denotes the bias parameter of the i -th hidden layer, and π_i is the output of the i -th hidden layer. Finally, features extracted from the full-connection layer are classified by softmax logistic regression to obtain the user's future dietary prediction result, as shown in the following equation:

$$y = \rho(\pi_n), \quad (11)$$

where ρ denotes softmax logistic regression function, π_n denotes the output of the last layer of the full-connection layer, and y denotes the recipe classification probability by softmax.

3.4. Optimization. We adopt BPR [34] to optimize the model, making the scores between positive and negative samples could be as large as possible so that user u_i and user's preference e_i^+ were higher than e_i^- . We train the embedding representation of users and items as follows:

$$L_1 = \sum_{u_i, e_i^+, e_i^-} -\ln \sigma(\delta(e_i^+, u_i) - \delta(e_i^-, u_i)) + \lambda \| \theta \|_2^2, \quad (12)$$

where u_i represents the user u_i , e_i^+ represents the user's favorite recipes, and e_i^- represents the user's disliked recipes. $\sigma(\cdot)$ is the softplus function, and the L2 regularization parameter makes the model prefer smaller weights, which reduces the complexity of the model and prevents overfitting.

We optimize the triplet representation in the knowledge graph as follows:

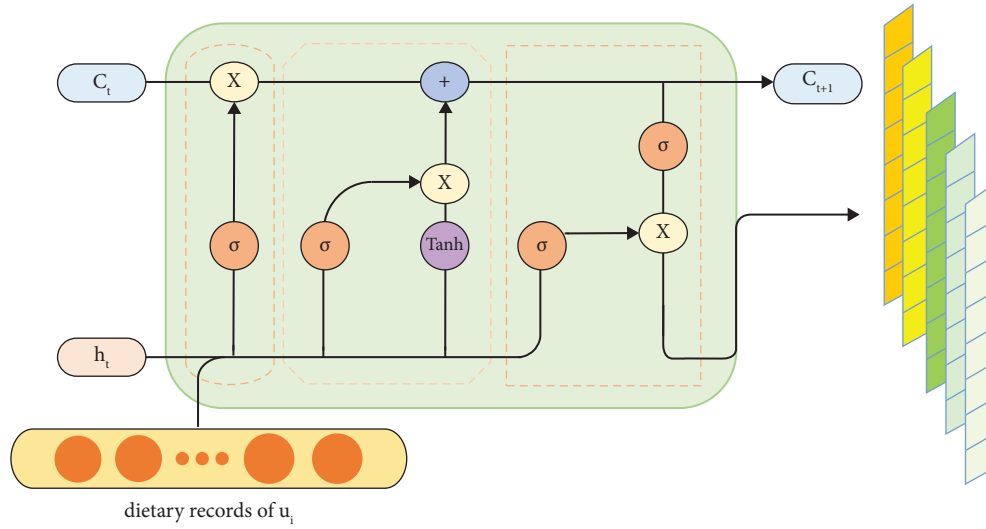


FIGURE 3: Illustration of LSTM layer.

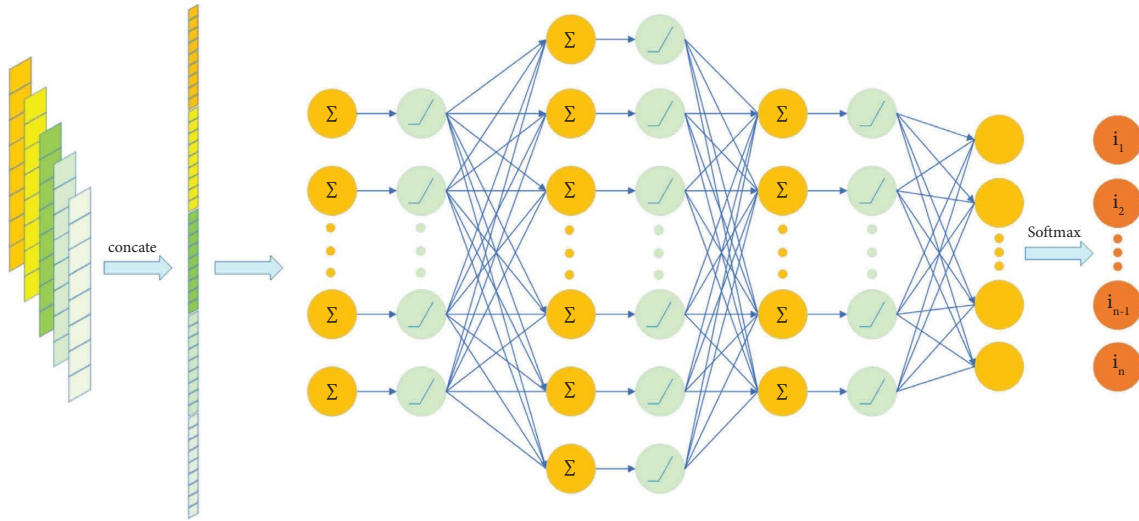


FIGURE 4: Illustration of fully connected layers.

$$L_2 = \sum_{(h,r,t,t') \in G} -\ln \sigma(f(h,r,t) - f(h,r,t')) + \lambda \|\theta\|_2^2, \quad (13)$$

where $(h, r, t, t') \in G$ denotes training sets, G denotes recipe knowledge graph, h denotes head entity, t denotes positive sample tail entity, t' denotes negative sample tail entity, and r denotes the relationship between the head entity h and the tail entity t .

We adopt cross entropy as a loss function to optimize the model, which is formulated as follows:

$$L_3 = -\sum_i y' \log(y_i), \quad (14)$$

where y' denotes the true labels in the training set and y_i denotes the corresponding component in the output vector of the model normalized by softmax. When the classification is more accurate, the component corresponding to y_i will be closer to 1, and thus the L_3 loss will be smaller.

4. Experiments

We conducted extensive experiments on the datasets to answer the following research questions:

RQ1: How does PPKG perform compared to other food recommendation methods?

RQ2: How do different components (namely, model depth and aggregator selection) affect the effectiveness of PPKG?

RQ3: What are the key hyperparameters of PPKG? How does it affect the performance of PPKG?

In the following, we will describe the datasets and experimental settings and then answer the above research questions.

4.1. Dataset. We crawled recipe websites (<https://www.ca39.com/2010/0410/29850.html>) and extracted information from textbooks [35–37] to build knowledge graph datasets of

recipes. We used libraries such as BeautifulSoup in Python to parse web content and extracted recipe names, ingredients, and effects. We used a rule-based method to identify diseases, symptoms, ingredients, and effects in recipes as well as their relationships. We defined rules based on common vocabulary and grammar patterns, such as “beneficial for disease xx,” “alleviate symptom xx,” and “contain ingredient xx.”

We performed data cleaning and preprocessing on the collected cancer knowledge. We conducted name normalization for some entities, such as replacing synonyms or related terms with a uniform standard name, such as replacing “Gastric Carcinoma,” “Gastric Cancer,” “Stomach Cancer,” and “Gastric Carcinoma.” The identified entities and relationships were manually reviewed and corrected by a professional physician to ensure the quality and accuracy of the dataset. And we construct the knowledge graph by transforming illness, recipes, symptoms, ingredients, and efficacy into five types of nodes, respectively. After that, we build four types of edges between these nodes to connect them. We connect each node with the corresponding edge, and the weight of each component is used as the edge weight. We take the triplet with relations between entity nodes as positive samples and then randomly select an entity to replace the tail entity to form the triple as negative samples.

To evaluate the effectiveness of PPKG on recipe recommendations, we used two widely used recommendation evaluation measures: Recall and NDCG.

4.1.1. Baseline Algorithms. To verify the validity of our model, we adopted the following baseline for verification:

BPRMF [34]. A personalized ranking method is based on factorization. It is one of the most advanced methods of recommending implicit feedback data for nonsequential items.

CKE [38]. A typical regularization-based approach leverages TransR-derived semantic embedding to enhance matrix factorization.

CFKG [39]. The model applies TransE on a unified graph including users, items, entities, and relationships and transforms the recommendation task into the credibility prediction of triplet $(u, \text{interact}, i)$, where u denotes user, i denotes item, and interact denotes user interaction with the item.

4.1.2. Parameter Settings. We implemented our model on TensorFlow. We set the embedding dimension of the knowledge graph of the model as 32, the number of hidden neurons of LSTM as 32, and the learning rate as 0.0001. We search the number of LSTM layers in [1–5], the input length of the user’s LSTM dietary records is tuned among [5–9], and the embedding regularization parameters of users and items are fine-tuned in [$1e^{-6}$; $1e^{-5}$; $1e^{-4}$; $1e^{-3}$; $1e^{-2}$; $1e^{-1}$]. Through extensive experiments, we find that the model

achieves the best performance when the number of LSTM layers equals 2, the length of dietary records equals 9, and the regularization parameter is set as $1e^{-3}$.

4.2. Model Comparison (RQ1). To investigate the effectiveness of our proposed approach (PPKG), we studied the performance of PPKG compared to the baseline above. Table 1 shows the performance comparisons. By analyzing the results, we draw the following conclusions.

- (1) The traditional collaborative filtering recommendation method (BPRMF) does not perform well because it does not consider the hidden relationship information or the relationship between each recipe.
- (2) Our proposed PPKG achieves the best performance in both evaluation metrics. This result verifies the merits of our proposed model.
- (3) PPKG is better than other models (CKE and CFKG) in two evaluation indexes, which demonstrate that information propagation mechanism and multilayer architecture of GAT are more effective. The graph attention and multistep propagation in GAT better capture feature representations, enhancing the recommendation algorithm.
- (4) PPKG achieves the best performance in the baseline, which demonstrates that LSTM can effectively enhance recipe recommendations. By incorporating LSTM into the recommendation algorithm, we can improve recipe recommendation performance.

To evaluate the ranking performance of FGCN, we present the performance of Top-N food recommendations in Figure 5, where the ranking positions vary from 5 to 20. Obviously, PPKG consistently outperforms other benchmarks on recall and NDCG metrics. It verified the effectiveness of PPKG on serialization recommended time.

4.3. Study of PPKG (RQ2)

4.3.1. Effect of Model Depth. We investigated whether the model could be improved from multiple layers of LSTMs by varying the number of layers of LSTMs in the PPKG. We vary the number of layers in [1–5]. Table 2 shows the performance of PPKG at different layers, where PPKG-1 denotes that there is one layer of the LSTM network in the model. From Table 2, we can observe that

- (1) Adding LSTM layers can effectively improve the performance of PPKG, and PPKG-2 and PPKG-3 are always better than PPKG-1 in two evaluation metrics, which indicates that adding LSTM layers can effectively extract temporal information from users’ dietary records.
- (2) When layers of PPKG-3 are further stacked, we find that performance degrades quickly, such as PPKG-4 and PPKG-5. The experiment results show that when

TABLE 1: Performance of compared methods.

Methods	Recall@10	NDCG@10
BPRMF	0.04758	0.06210
CKE	0.05176	0.06491
CFKG	0.05416	0.07076
PPKG	0.07213*	0.08902*

the model layer is overmuch, it may introduce noise into modeling, which can easily lead to excessive proposed merger-degraded performance.

To investigate self-representation and neighbor representation, we conduct experiments on three aggregation methods of PPKG (cf. Section 3.2) to study the performance effect of different aggregators on PPKG: PPKG_{Bi}, PPKG_{G_{GCN}}, and PPKG_{G_{Graphsage}}. Table 3 summarizes the performance of the three variations of PPKG. From Table 3, we obtain that PPKG_{G_{Graphsage}} is significantly better than PPKG_{Bi} and PPKG_{G_{GCN}} on both metrics. We analyze that the information transfer framework can aggregate the information of higher-order neighbors and can effectively learn the embedding representation of the node by aggregating the information of the node’s neighbors. It has a better effect in performing information aggregation and propagation feature interaction.

4.3.2. Effect of Aggregators in Ingredient Graph. To investigate the effectiveness of aggregators in knowledge graphs, we conducted experiments on two variants of PPKG. Table 4 summarizes the effects of different aggregators in the knowledge graph, PPKG_{sum} and PPKG_{uni}, on the performance of PPKG.

From Table 4, we observe that PPKG_{sum} always outperforms PPKG_{uni}. This is because the sum operation in PPKG_{sum} connects the individual features. The connections enhance the effectiveness of the PPKG_{sum} model.

4.4. Parameter Sensitivity (RQ3). In this section, we investigate the effect of hyperparameters on the performance of the proposed PPKG. We vary the regularization term λ in ($1e^{-6}$; $1e^{-5}$; $1e^{-4}$; $1e^{-3}$; $1e^{-2}$; $1e^{-1}$). Figure 6 shows the model’s performance when regularization parameters are adjusted. We observe that

- (1) PPKG performed better than CFKG on both measures; PPKG is more effective on recipe recommendations.
- (2) The performance of PPKG and CFKG decreases rapidly when λ is larger than $1e^{-3}$. The results show that proper setting regularization parameter λ is critical for enhancing model performance. This is because improper set λ weights lead to insufficient

regularization or deviation of the objective function. When the weight is too small, it fails to provide adequate regularization, resulting in overfitting, whereas excessively large weights skew the objective function heavily towards the regularization term, failing to properly capture user-recipe relevance and deteriorating the model’s predictive performance.

As shown in Figure 7, model performance first increases then decreases as embedding dimension d grows. When dimension is small ($d=16$), increasing dimension can enhance the representation capability and encode more feature information, thus improving performance. However, when d continues growing ($d=64$), the increased model complexity may lead to overfitting and therefore performance drops. Appropriately increasing d can improve the model’s expressiveness and generalization, but excessively large d causes overfitting.

4.5. Case Study. In this section, we evaluate in detail the model’s recommendations for patients with different cancers, as shown in Figure 8.

Recommendations for Lung Cancer Patients. For lung cancer patients with pain, the model recommends bitter gourd with five flavors. The PPKG model utilizes the features of the user’s lung cancer, matches food ingredients with blood circulation promotion and pain relief effects, and finally generates recipes for lung cancer patients with pain. Through the connections between bitter gourd and efficacy nodes, the model obtains the feature of blood circulation promotion. Bitter gourd is connected with ingredients’ nodes to get its feature. The model calculates the similarity between lung cancer patients and bitter gourd and finally recommends bitter gourd with five flavors to lung cancer patients. Bitter gourd can promote blood circulation and relieve pain, so bitter gourd with five flavors can provide nutrition and alleviate symptoms for these lung cancer patients.

Recommendations for Patients with Gastric Carcinoma. For stomach cancer patients with pain, the model recommends chicken soup with garlic. For stomach cancer patients with pain, the PPKG model identifies recipes containing ingredients with anticancer and blood pressure lowering effects based on the user’s condition. Garlic has anticancer and blood pressure lowering effects. The model finds related recipes through this ingredient. Therefore, chicken soup with garlic is suitable to relieve discomfort for stomach cancer patients with pain.

Through these specific case studies, the applicability of the model recommendation results to different cancer patients can be more intuitively checked and the actual effect of the model can be evaluated.

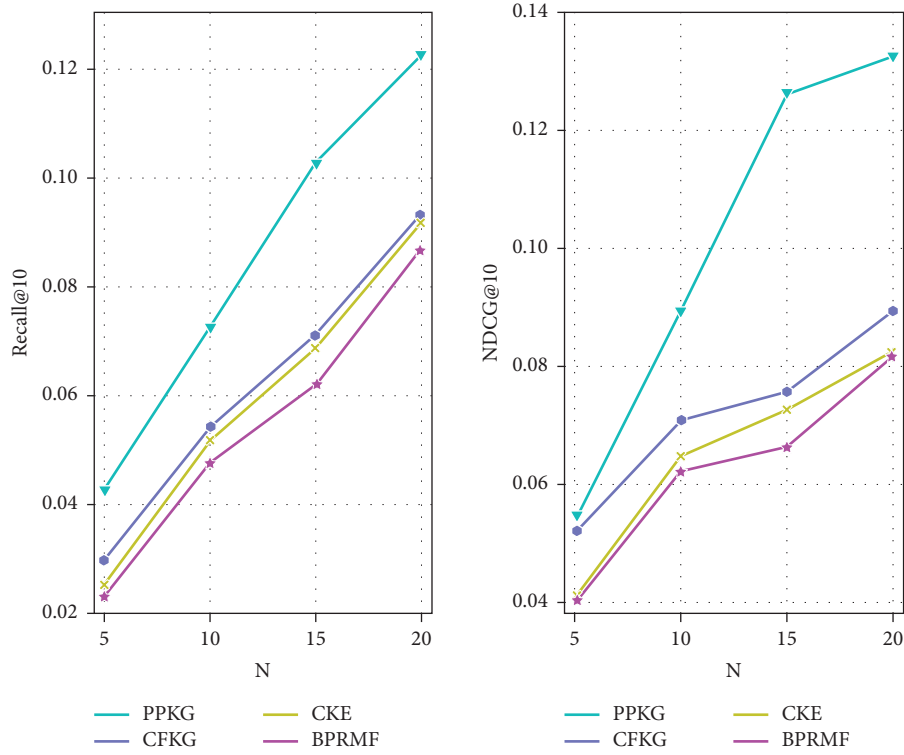
FIGURE 5: Performance of top- N food recommendation.

TABLE 2: Effect of LSTM layer numbers.

	Recall@10	NDCG@10
PPKG-1	0.06290	0.07485
PPKG-2	0.07213*	0.08902*
PPKG-3	0.06808	0.07945
PPKG-4	0.05783	0.07277
PPKG-5	0.05227	0.06314

TABLE 3: Effect of aggregators for ego representations and neighbor representations.

Aggregator	Recall@10	NDCG@10
PPKG _{Bi}	0.04758	0.06062
PPKG _{GCN}	0.06100	0.07630
PPKG _{GraphSage}	0.07213*	0.08902*

TABLE 4: Effect of aggregators in the ingredient graph.

Aggregator	Recall@10	NDCG@10
PPKG _{uni}	0.06123	0.08166
PPKG _{sum}	0.07213*	0.08902*

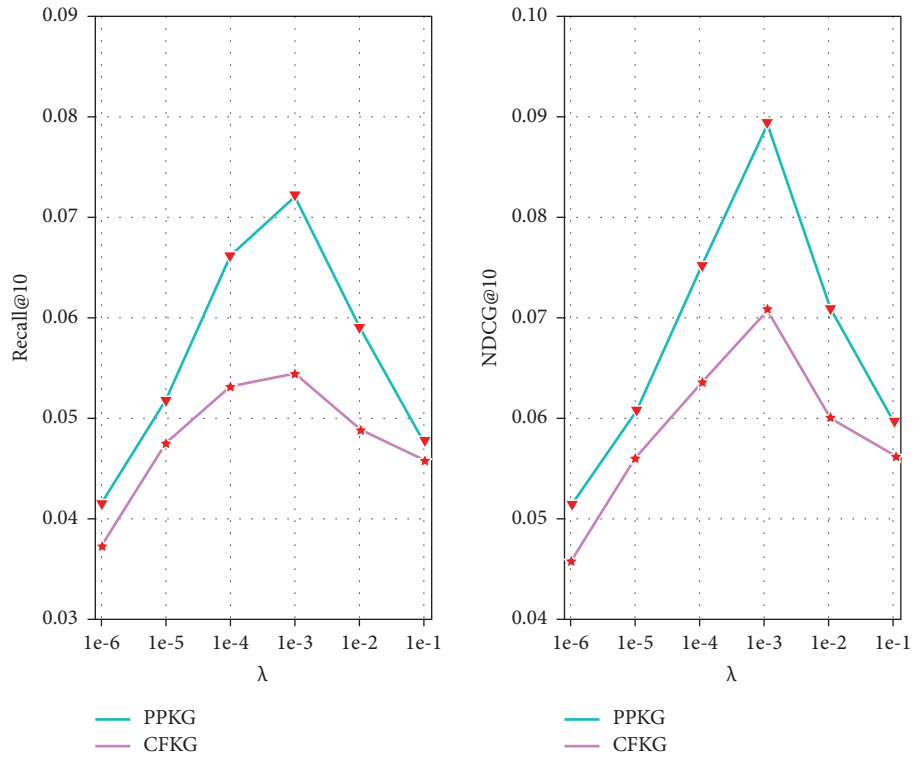


FIGURE 6: Performance of PPKG regarding Recall@10 and NDCG@10 as adjusting the strength of regularization.

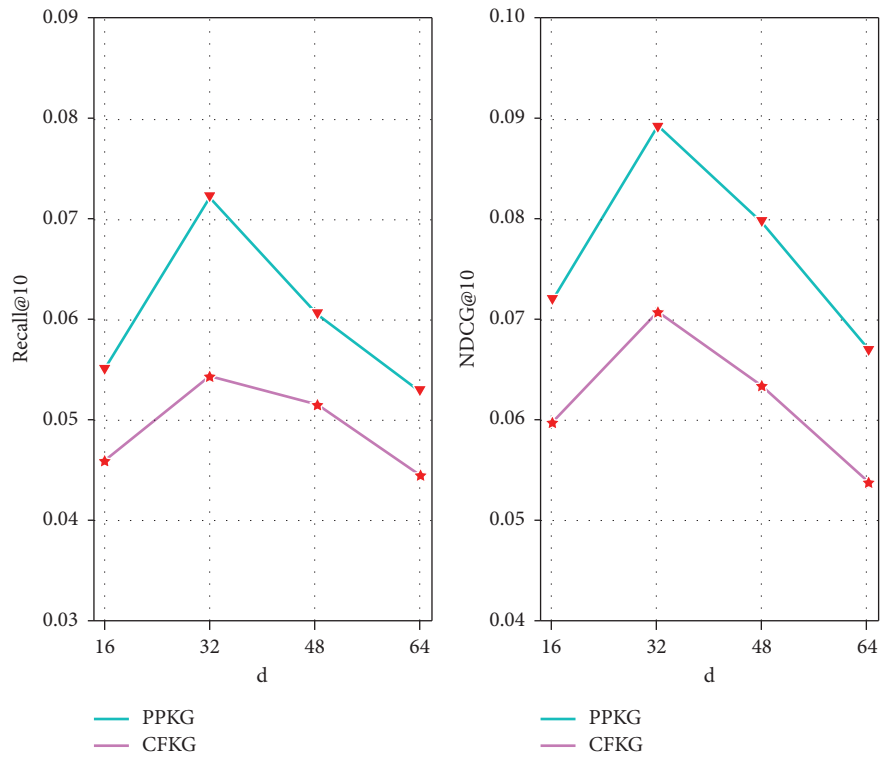


FIGURE 7: Performance of PPKG regarding Recall@10 and NDCG@10 as adjusting the length of dimension.

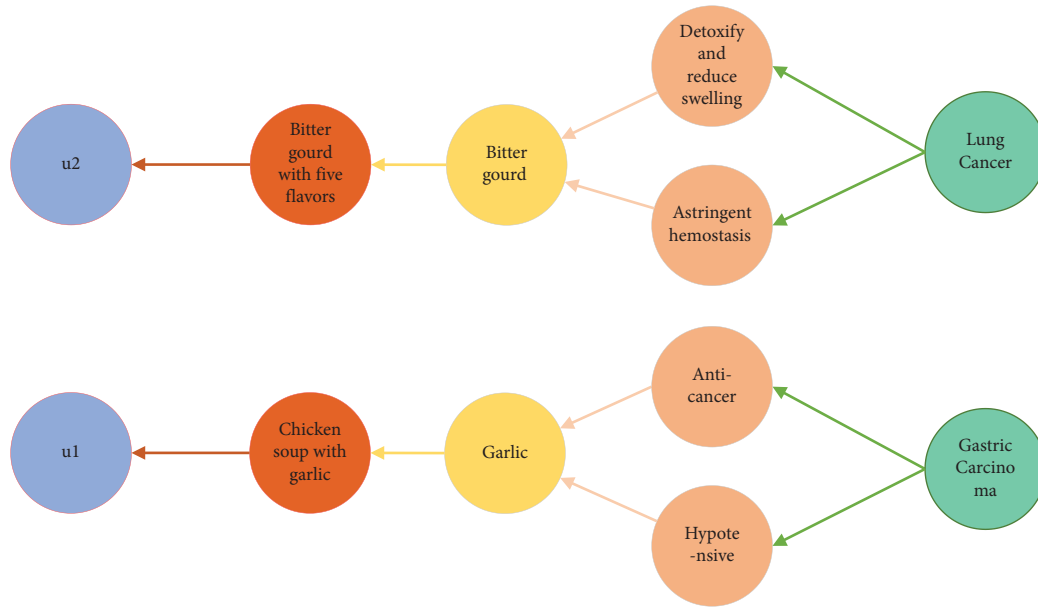


FIGURE 8: Personalized recipe recommendations for patients with cancer.

5. Conclusion

In our works, we use the GAT framework to obtain the entity embedding representation of users, symptoms, recipes, ingredients, and efficacy in the knowledge graph and then take the historical dietary records of users as the input of LSTM to mine the potential connections in the user's dietary records. Recipe recommendation is regarded as a multiclassification problem. The output of LSTM is passed through the fully connected layer and then output through the softmax function, the full-connection layer is used to extract the features from the output of LSTM, and finally the prediction probability of each recipe is obtained by the softmax function. Extensive experiments demonstrate that PPKG outperforms many baseline methods.

Due to the limited knowledge sources, the cancer knowledge graph may still be incomplete and inaccurate. We will discuss the importance of continuously updating the knowledge graph in the future to dynamically refine it and ensure up-to-date and reliable knowledge. Considering the impacts of different cancer stages on diet recommendation, we will specify the highly customized nature of cancer diet in the future and discuss the effects of disease stage factors on the model. In our future work, we plan to leverage users' personal information in two aspects: first, incorporate user data to enhance user-recipe embedding representations. Second, improve recipe recommendation by aggregating more information and increasing diversity. The goal is to provide users with more comprehensive and healthier recipe recommendations tailored to their needs.

Data Availability

The datasets used in this study are available online on the following link anticancer recipe recommendation based on cancer dietary knowledge graph dataset <https://github.com/QuinlandQ/Knowledge-Graph-Recommendation> accessed on 30 May 2023.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Jianchen Tang contributed to development or design of methodology, creation of models, programming and software development, implementation of the computer code and supporting algorithms, and testing of existing code components. Bing Huang contributed to scrub data and maintain research data for initial use and later reuse. Mingshan Xie contributed to ideas and formulation or evolution of overarching research goals and aims.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (grant no. 62266010), in part by the cultivation project of Guizhou University (grant no. [2019]57), in part by the research project of Guizhou University for talent introduction (grant no. [2019]31), in part by the higher education research project of Guizhou University (grant no. GDGJYJ2020014), and in part by the first-class curriculum cultivation project of Guizhou University (grant no. XJG2021023).

References

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 1, pp. 7–34, 2019.
- [2] W. C. Willett, "Diet and cancer," *The Oncologist*, vol. 5, no. 5, pp. 393–404, 2000.
- [3] R. Doll and R. Peto, "The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today," *Journal of the National Cancer Institute*, vol. 66, no. 6, pp. 1192–1308, 1981.

- [4] P. Greenwald, C. K. Clifford, and J. A. Milner, "Diet and cancer prevention," *European Journal of Cancer*, vol. 37, no. 8, pp. 948–965, 2001.
- [5] M. S. Donaldson, "Nutrition and cancer: a review of the evidence for an anti-cancer diet," *Nutrition Journal*, vol. 3, no. 1, pp. 1–21, 2004.
- [6] G. McKeon, E. Papadopoulous, J. Firth et al., "Social media interventions targeting exercise and diet behaviours in people with noncommunicable diseases (NCDs): a systematic review," *Internet interventions*, vol. 27, Article ID 100497, 2022.
- [7] S. Bohlscheid-Thomas, I. Hoting, H. Boeing, and J. Wahrendorf, "Reproducibility and relative validity of food group intake in a food frequency questionnaire developed for the German part of the EPIC project. European Prospective Investigation into Cancer and Nutrition," *International Journal of Epidemiology*, vol. 26, no. 90001, pp. S59–S70, 1997.
- [8] T. J. Key, A. Schatzkin, W. C. Willett, N. E. Allen, E. A. Spencer, and R. C. Travis, "Diet, nutrition and the prevention of cancer," *Public Health Nutrition*, vol. 7, no. 1a, pp. 187–200, 2004.
- [9] A. Daowd, M. Barrett, S. Abidi, and S. S. R. Abidi, "A framework to build a causal knowledge graph for chronic diseases and cancers by discovering semantic associations from biomedical literature," in *Proceedings of the 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)*, pp. 13–22, IEEE, Victoria, Canada, August 2021.
- [10] C. Zhu, Z. Yang, X. Xia, N. Li, F. Zhong, and L. Liu, "Multimodal reasoning based on knowledge graph embedding for specific diseases," *Bioinformatics*, vol. 38, no. 8, pp. 2235–2245, 2022.
- [11] J. Freyne and S. Berkovsky, "Recommending Food: Reasoning on Recipes and ingredients," *International Conference on User Modeling, Adaptation, and Personalization*, Springer, Berlin, Heidelbergpp. 381–386, 2010.
- [12] Z. Yuan and F. Luo, "Personalized diet recommendation based on K-means and collaborative filtering algorithm," *Journal of Physics: Conference Series*, vol. 1213, no. 3, Article ID 032013, 2019.
- [13] S. Tangruamsub, K. Kappaganthu, J. O'Donovan, and A. Madan, "CareGraph: a graph-based recommender system for diabetes self-care," 2021, <https://openreview.net/forum?id=rX3rZYP8zZF>.
- [14] W. Min, S. Jiang, and R. Jain, "Food recommendation: framework, existing solutions, and challenges," *IEEE Transactions on Multimedia*, vol. 22, no. 10, pp. 2659–2671, 2020.
- [15] B. Huang, X. Shi, R. Wang, C. Wang, and Y. Han, "A novel recipes recommendation system based on knowledge-graph," in *Proceedings of the 2022 7th international conference on intelligent computing and signal processing (ICSP)*, pp. 1408–1412, IEEE, Xi'an, China, April 2022.
- [16] L. Huang, C. Yu, Y. Chi, X. Qi, and H. Xu, "Towards Smart Healthcare Management Based on Knowledge Graph technology," in *Proceedings of the 2019 8th International Conference on Software and Computer Applications*, pp. 330–337, Penang, Malaysia, February 2019.
- [17] H. Wang, F. Zhang, J. Wang et al., "Ripplenet: propagating user preferences on the knowledge graph for recommender systems," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 417–426, Torino, Italy, October 2018.
- [18] N. Rastogi and M. J. Zaki, "Personal health knowledge graphs for patients," 2020, <https://arxiv.org/abs/2004.00071>.
- [19] Y. Yuan, Y. Tang, L. Du, and X. Li, "Entity2item: leveraging knowledge graph embedding for item recommendation," in *Proceedings of the 2021 international joint conference on neural networks (IJCNN)*, pp. 1–7, IEEE, Shenzhen, China, July 2021.
- [20] X. Ma, Z. Gao, Q. Hu, and M. Abdelhady, "Contrastive knowledge graph attention network for request-based recipe recommendation," in *Proceedings of the ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3278–3282, IEEE, Singapore, May 2022.
- [21] X. Gao, F. Feng, H. Huang, X. L. Mao, T. Lan, and Z. Chi, "Food recommendation with graph convolutional network," *Information Sciences*, vol. 584, pp. 170–183, 2022.
- [22] Y. Tian, C. Zhang, R. Metoyer, and N. V. Chawla, "Recipe recommendation with hierarchical graph attention network," *Frontiers in big Data*, vol. 4, Article ID 778417, 2021.
- [23] Z. Lei, A. Ul Haq, A. Zeb, M. Suzauddola, and D. Zhang, "Is the suggested food your desired?: multi-modal recipe recommendation with demand-based knowledge graph," *Expert Systems with Applications*, vol. 186, Article ID 115708, 2021.
- [24] I. Padhiar, O. Seneviratne, S. Chari, D. Gruen, and D. L. McGuinness, "Semantic modeling for food recommendation explanations," in *Proceedings of the 2021 IEEE 37th international conference on data engineering workshops (ICDEW)*, pp. 13–19, IEEE, Chania, Greece, April 2021.
- [25] Y. Chen, A. Subburathinam, C. H. Chen, and M. J. Zaki, "Personalized food recommendation as constrained question answering over a large-scale food knowledge graph," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pp. 544–552, Jerusalem, Israel, August 2021.
- [26] D. Li, M. J. Zaki, and C. Chen, "Health-guided recipe recommendation over knowledge graphs," *Journal of Web Semantics*, vol. 75, Article ID 100743, 2023.
- [27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [28] X. Wang, X. He, Y. Cao, M. Liu, and T. S. Chua, "Kgat: knowledge graph attention network for recommendation," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 950–958, Anchorage, AK, USA, July 2019.
- [29] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, 2015.
- [30] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, <https://arxiv.org/abs/1609.02907>.
- [31] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," *Proc. icml*, vol. 30, no. 1, p. 3, 2013.
- [32] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [33] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 315–323, Singapore, August 2011.
- [34] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: bayesian personalized ranking from implicit feedback," 2012, <https://arxiv.org/abs/1205.2618>.

- [35] M. Chang, *KangAiYaoShan*, Hunan Science and Technology Press, Beijing, China, 1996.
- [36] L. Xu and J. Lu, *100 Kinds of Anti-cancer Medicinal Diet*, People's Medical Publishing House, Beijing, China, 2014.
- [37] M. Chang, *Practical Anti-cancer Medicinal Diet*, China Medical Science and Technology Press, Beijing, China, 2014.
- [38] F. Zhang, N. J. Yuan, D. Lian, X. Xie, and W. Y. Ma, "Collaborative Knowledge Base Embedding for Recommender systems," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 353–362, San Francisco, CA, USA, August 2016.
- [39] Y. Zhang, Q. Ai, X. Chen, and P. Wang, "Learning over knowledge-base embeddings for recommendation," 2018, <https://arxiv.org/abs/1803.06540>.