Journal of Information Technology and Digital World (ISSN: 2582-418X)
www.irojournals.com/itdw/

Novel Nutritional Recipe Recommendation

K. Vani[1], K. Latha Maheswari[2]

[1]Department of Computer Science and Engineering, PSG College of Technology, Coimbatore, India.

[2]Department of Electrical and Electronics Engineering, PSG College of Technology, Coimbatore, India.

E-mail: [1]kvi.cse@psgtech.ac.in, [2]klm.eee@psgtech.ac.in

Abstract

Food is essential for living and is the foremost important energy source, making us do all the work. Nowadays, the variability in these food items is increasing. To find out about any new dish or recipe, we mainly depend upon people around us or by trial-and-error method, but neither method tells us about its nutritional content. Since the web has begun to grow, the advent of food recommender systems has made food suggestions easier but these systems work only on the feedback provided by the customer. Hence, here comes a requirement for a nutritional-based recommender system that considers ratings and nutrition and provides the user with an absolute best recommendation so that the users' taste preferences and well-being are given equal priority. This study intends to use graph embedding approaches to develop a food recipe recommender system, which uses the ingredients' nutritional value alongside the recipe's taste and customer feedback. These food recommender systems can impact people's dietary practices, as their suggestions are both healthy and relevant. People can now eat healthily without being compromised on taste.

Keywords: Graph Embeddings, Bipartite Graphs, Healthy Lifestyle, Nutrition, Personalization, Graph based recommendation

1. Introduction

Improper and insufficient diet is one of the multiple reasons for non-wellbeing and malnutrition. A survey conducted by World Health Organization (WHO) reports that around 30% of the world's population suffers from different illnesses, and 60% of deaths in children are related to malnutrition. WHO also reported that lacking and imbalanced nourishment admissions cause about 9% of heart attacks, approximately 11% of chronic heart illness

DOI: https://doi.org/10.36548/jitdw.2023.1.001
Received: 27.12.2022, received in revised form: 27.1.2023, accepted: 13.2.2023, published: 23.2.2023

deaths, and 14% of gastrointestinal cancer issues worldwide. Besides, around 0.25 billion children suffer from Vitamin-A deficiency, 0.2 billion individuals suffer from iron deficiency, and 0.7 billion individuals suffer from iodine insufficiency [10]. According to the forecast by the World Health Organization, the number of overweight grown-ups all over the world has come to a disturbing number, with 2.3 billion by 2015. Altogether, overweight and obese cause numerous chronic health issues. A suitable dietary admission is considered as a critical move to solve this issue. The recommender frameworks are examined as a compelling arrangement to assist users in altering their eating behaviour for more beneficial nourishment choices [9]. Food Recommender Systems (FRS) have a great scope as people are inclined towards a healthy lifestyle. Food Recommender Systems can help users navigate the vast amount of online food/recipe resources and steer them toward healthy options. FRS has the potential to become an effective technology for solving the global obesity and malnutrition epidemic by disseminating dietary recommendations to the general public.

Most existing works on food recommendations suggest food items for individual users by considering only their preferences. However, most of these systems target taste as a priority rather than health, neglecting the importance of nutrients in the food. Some systems even disregard the appropriate choices of users. The existing food recommendation suggests standard recipes to users within the databases, but these do not cater to the needs of users who want to try new recipes similar to the existing ones. Graph theory techniques can be used to recommend better recipes than other existing standard systems in order to express all the data as a graph between users, ingredients and recipes. We create a framework that properly leverages these relationships for finding recipes the user might be interested in. Specifically, we propose an embedding-based ingredient predictor to predict the relevant ingredient. The novel variants of the recipe are recommended by substituting the relevant ingredient in the interested recipe, such that the users' taste preferences and health are given equal priority.

2. Literature Survey

Tran et al. [9] surveyed different types of food recommender systems, namely recommendations considering only user preferences, considering the user's nutritional needs, balancing between user preferences and nutritional needs and food recommender systems to groups. In order to recommend healthy recipe, Chen et al. proposed a new framework called NutRec [6]. This model predicts the required ingredients and their quantities based on the initial user-defined ingredients in order to generate a healthy recipe based on nutritive value.

Finally, the dataset for the most similar healthy recipes backed by the pseudo-recipe will be reviewed. In [7], feature identification techniques based on Ensemble Topic Modelling (EnsTM) has been used to perform efficient recipe recommendation and user modelling. In [8], Yamanishia et al. proposed a model to identify alternative-ingredient recommendations leveraging co-occurrence relations between ingredients in the recipe database. The recommendations were based on compatibility between alternative-ingredient and ingredients in a recipe except for exchange-ingredient, and recommendations were based on the similarity between exchange-ingredient and alternative-ingredient. Rehman et al. [10] model recommend an inventory of optimal food items based on users' pathological reports. The model uses Ant Colony Algorithm (ACO) to recommend suitable foods according to the values of pathological reports. In [4], Teng et al. constructed two types of networks, the complement and substitute networks, from the recipe review data. The networks reflect different relationships between ingredients and capture users' knowledge about combining ingredients. The complement network captures which ingredients tend to co-occur frequently using pointwise mutual information. The substitute network, derived from user-generated suggestions for modifications, can be decomposed into clusters of functionally equivalent ingredients and captures users' preference for healthier recipe variants. A nutritional model is developed [5] to recognize the overall nutritive value of a recipe by considering its ingredients. Graph Neural Network (GNN) is used to model the user's rating scores on recipe-user bipartite and ingredient-recipe graphs. These two models are then integrated to recommend more healthy and tasty recipes based on the user preference.

In [8], a list of alternative ingredients is made for all the existing ingredients, but the issue is that if there are no possible alternatives for a particular ingredient, the whole flow of recommendations could be affected. Furthermore, in [10], users' pathological reports are needed for a recommendation, making it unhandy for many users. The user's preference is not considered, thus making the process of recommendation with low personalization [4, 6]. In [5], the system utilizes Graph Neural Network (GNN); the setback of this method is that it is more complex to make recipe recommendations for random users when compared to higher users due to the lack of information availability. [7] Different learning techniques are utilized to develop personalized models but usually lack explainability, which is considered as a potential element of the food recommendation systems.

3. Proposed Method

With this work, we want to propose new recipes that are both delicious and nutritious. A recipe is defined by the ingredients used and is considered new if it is not included in the proposed dataset. While a recipe's tastiness can be roughly estimated by the average ratings supplied by a wide range of users, a recipe's healthfulness can be evaluated by the ratio of good to harmful nutrients (like proteins to sodium). However, we cannot collect the nutrient information and ratings of these recipes as the main focus is dedicated towards providing novel recipe recommendations where these recipes are outside our dataset. As a result, we must employ nutrition and rating models to determine both the healthfulness and tastiness of new recipes by computing the nutritional content and user rating, respectively. Next, using these algorithms, we may offer healthy recipes by assessing new recipes made with various ingredient combinations and aiming for a high healthiness and taste score. As shown in Figure 1, the proposed framework consists of four modules. First, the preprocessed dataset is given as input to the Bipartite graph generation module where user ratings and ingredients recipe bipartite graph is generated. Second, the nutritional model is used to calculate the nutritional value of the ingredients. Third, the rating model determines user preference for the recipe using the rating information. Finally, the novel recipes are recommended by combining the nutritional and user preference through the k nearest neighbour algorithm.

3.1 Data Pre-Processing

The preprocessing of the dataset was done in three stages, firstly the ingredients and their nutritional content are to be cleaned; thus, the extraction of ingredients and removal of the bracketed comments in the ingredients are done as it does not add any value to the data; subsequently, removal of punctuations present in the ingredients is done, and finally, only the nouns are retained in the ingredients as having an adjective or a verb before or after the noun would not add any value

3.2 3.2 Bipartite Graph Construction

From the dataset, two different bipartite graphs are formed, illustrated in Figure 2. One being the user-recipe graph and the other being the ingredient-recipe graph. The presence of rating data helps to form a bipartite graph between the user and recipe, which is constructed in such a way that one set of nodes is the User_id nodes, and the other set is the Recipe_id nodes. The graph constructed is directed as edges must be present only from

User_id to Recipe_id, and weights are added on the edges, which depict the amount of rating given by the user to a recipe.

The following bipartite graph is constructed using the recipe data and the ingredients, which is constructed in such a way that one set of nodes is the ingredient nodes, and the other set is the Recipe_id nodes. This graph is also directed as edges must be present only from ingredients to Recipe_id, and edges depicting the presence of a particular ingredient in the recipe.

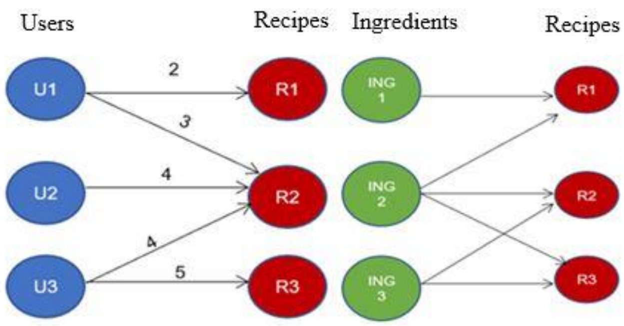

Figure 1. User-Recipe and Ingredient-Recipe Bipartite Graph

3.3 Nutritional Model

The dataset lacked the Nutritional content of individual ingredients, so the Nutritional model was built to solve this problem.

$$R.X = Y \quad (1)$$

So, an equation (1) was formed, where R^{MXI} is a sparse matrix where M and I represent the number of recipes and ingredients, respectively, Y^{MXN} is a two-dimensional matrix where N represents the nutritional content of the recipes. It gives the amount of Nutritional content present (i.e., “Sugar, Fat, Protein, Sodium”) in each recipe. X^{IXN} is a resulting matrix when eq. (1) is solved using linear regression.

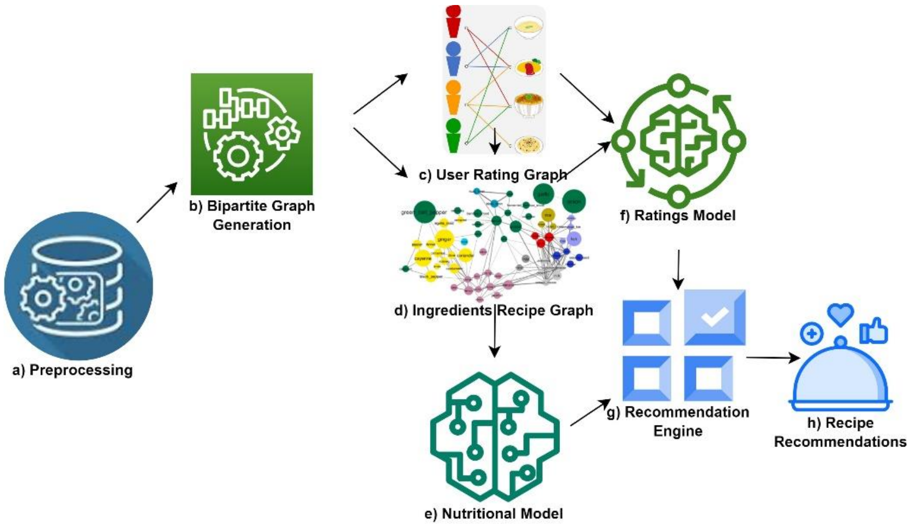

Figure 2. Flow diagram of the proposed model

3.4 Rating Model

The main objective of this research work is to recommend recipes to users in order to satisfy both health and taste preferences. The ratings play a significant role in modelling the users' taste preferences. The model takes User-Recipe and Ingredient-Recipe bipartite graphs as the input and helps predict the recipe's score.

Machine Learning concepts are inefficient to apply directly to graphs. Thus, embeddings are used to apply Machine Learning to graphs. Embeddings help to capture the graph topology, node to node relation by encoding each node and giving a vector representation based on the node similarity and other structural information present in the graph. There are multiple methods for generating embeddings, which is essential for accomplishing the goal of the rating model. We used Node2Vec [11] method in our proposed model because of its superiority over other the state of art methods. Node2vec [11] features a walk bias variable α, parameterized by P and Q, where P represents the probability of a random walk being returned to the previous node prioritizing the breadth-first search (BFS) procedure. Q represents the probability of a random walk discovering an undiscovered part of the graph prioritizing the depth-first search (DFS) procedure. Thus, P controls the discovery of view around the node, whereas Q controls the discovery of a larger neighbourhood. Therefore, the decision of where to walk next is influenced by the probabilities 1/p or 1/q.

3.5 Novel Recipe Recommendation

The embeddings generated from the User-Recipe graph are used as input to K-Nearest Neighbor (KNN) model. KNN model gives the most similar users to the target user employing cosine similarity as a metric using equation (2),

$$\text{Cosine}_{\text{Similarity}}(A, B) = (A.B)/(||A||.||B||) \tag{2}$$

Thus, the recipes associated with similar users are recommended to the target user. For generating novel recipes, the ingredient embeddings are passed to the similarity algorithms resulting in the most similar ingredients for the target ingredient.

$$HR = C1 * \text{Sugar}_R + C2 * \text{Protein}_R + C3 * \text{Fat}_R + C4 * \text{Sodium}_R \tag{3}$$

Considering the assumption that the first ingredient of any recipe is prominent, the Novel recipe recommendation module replaces the recipe's first ingredient with its most similar ingredient and calculates the new health score of such produced recipe using Equation (3). In Equation (3), individual weights are assigned to every nutrient associated with the recipe "R", and HR is set with a constraint value to ensure the healthy recommendation of recipes to a target user. All the constrained and weights are assigned in accordance with the dataset.

4. Evaluation and Results

4.1 Dataset

The dataset considered includes 50,000 recipes retrieved from Kaggle [12]. Recipe data gives information regarding recipe_id, recipe_name, image URL of the recipe, ingredients, cooking_directions and nutritional content of the whole recipe. The rating data gives information such as user_id, recipe_id, the rating given by the user to the recipe, and the date on which the rating was given.

4.2 Evaluation Metrics

4.2.1 Average Precision (AP)

This metric gives us a score based on the number of relevant items in the predicted list. It can be calculated as shown in Equation (4). Here P(k) represents Precision at k calculated by considering the subset till k, rel(k) returns a binary value based on the relevance of the item k, and m represents the total item space.

$$\text{Average Precision (AP)} = 1/m\ (\sum P(k).rel(k)) \quad (4)$$

4.2.2 Mean absolute error (MAE)

This metric gives us the error count between lists containing the predicted and the actual values. It can be calculated as shown in Equation (5). Here yi represents ith prediction, xi represents ith actual value, and n represents the total item space. Here, this metric is applied between health scores of both predicted and actual values. The value obtained will indicate the Mean absolute error of health scores in the recommended ones.

$$\text{MAE} = 1/n\ (\sum |yi-xi|) \quad (5)$$

4.2.3 Root mean squared error (RMSE)

This metric gives us the error count between the lists containing the predicted and the actual values. It can be calculated as shown in Equation (6). Here yi represents ith prediction, xi represents ith actual value, and n represents the total item space. Here, this metric is applied between health scores of both predicted and actual values. The value obtained will indicate the Mean absolute error of health scores in the recommended ones.

$$\text{RMSE} = \sqrt{}\ (\sum (yi-xi)2/n) \quad (6)$$

4.3 Results

The system’s main goal is to maximize the intake of Protein and sodium (stabilized amount) and to minimize the intake of sugar and fat content present in the recommendation of recipes; weights of the nutrients are chosen to suit this requirement. For the recommendation to work, the user’s rating data must be present in the dataset, i.e., the system must be aware of the user’s preferences. Though the system’s main goal is to maximize the intake amount of Protein and sodium, the system’s working can be better understood in two instances.

Table 1. Most Similar ingredients

Ingredient	**chicken broth**	**bacon**	**turkey**	**mustard**	**quarts water**
Top 4 similar Ingredients	dish pie crusts	head cabbage cut eighths	bars caramel peanut nougat candy	envelope reduced sodium taco mix	lobster mushrooms cubes

	coarse bulgur	creamcheese chives	kielbasa sausage	strawberry syrup	grain rice mix herbs
	beef sausage hillshire farm	salt spice islands beau monde	coarse grain mustard	quarts buttermilk	bottle pimento
	mango	ginger crust	dole carrots	cooking oil	acai berry pulp frozen

Using the default version of the recommender system, the recommendation provided to the user depends upon the attribute weights as follows, Sugar (c1), Fat (c2), Protein(c3), Sodium (c4) and with the values [0.2,0.2,0.34,0.00026] respectively. Finally, a score is calculated using Equation (6), and a minimum value of 10, is fixed to HR (Score of the recipe) to eliminate the recommendation of unhealthy recipes.

recipe_id	recipe_name	ingredients	sugars	fat	protein	sodium	health_score
r23314	Chicken Chimichangas	chicken broth,longgrain	6.050989	83.6897	79.78748	2321.729	45.67953054
r174720	Feta and Bacon Stuffed	bacon,feta cheese,creai	10.12954	84.16102	64.73801	1960.614	41.37879504
r23157	Turkey in a Bag	turkey,salt pepper,flour	1.327612	36.47689	93.23131	309.4594	39.34000524
r8534	Bou's Chicken	mustard,sugar,soy sauc	26.55713	52.09826	66.06461	1986.014	38.70940904
r91456	Good Ole' Southern Frc	quarts water,lemon,me	8.318274	63.31022	66.64442	2644.106	37.67226916

Figure 3. Recommendation of recipes within the database

From Figure 3 it has been observed that for the user ‘u2437545’, under the default setting of the recommender system, the five most recommended recipe is Chicken Chimichangas, Feta and Bacon Stuffed Chicken with Onion Mashed Potatoes, Turkey in a Bag, Bou’s chicken, Good Ole’s Southers Frogmore Stew. In the previous recommendation, the recipe Chicken Chimichangas (r23314) has its first ingredient as chicken broth. Furthermore, using the rating model, the first ingredient’s most similar ingredients are found, and it is used to replace the existing one. The most similar ingredients to chicken broth, bacon, turkey, mustard, and quarts of water are listed in Table 1. So, in the novel recipe, the topmost similar ingredient, i.e., dish pie crusts, is replaced in place of chicken broth. The novel recipe recommended to the user ‘u2437545’ using the proposed model is shown in Figure 4.

recipe_id	recipe_name	ingridients	sugars	fat	protein	sodium	scores_new
nr23314	novel_Chicken Chimichangas	dish pie crusts,longgrain r	6.279759	84.06616	8.309857	6.279759	20.896167
nr174720	novel_Feta and Bacon Stuffed	head cabbage cut eighths	12.34337	79.59479	9.722426	0	21.693258
nr23157	novel_Turkey in a Bag	bars caramelpeanut noug	4.298805	45.85963	5.387329	42.29588	11.874376
nr8534	novel_Bou's Chicken	envelope reducedsodium	26.55849	51.95471	25.79677	0	24.47354
nr91456	novel_Good Ole' Southern Fr	lobster mushrooms cube:	8.318274	56.95509	0.857218	0	13.346126

Figure 4. Generated Novel Recipes

Table 2. Comparison of results for different algorithms

Algorithm	MAE	RMSE	AP
Proposed			
Top 5	0.663	0.883	0.23
Top 10	0.652	0.864	0.38
Top 20	0.653	0.866	0.46
Top 50	1.519	1.803	0.50
KNN CF	0.614	0.885	0.21
SVD CF	0.617	0.819	0.20

We have used two collaborative filtering algorithms as baselines: a model based Singular Value Decomposition (SVD) and a memory-based K-Nearest Neighbours (KNN). The evaluation results are then determined by comparing the health scores of recommended recipes with those of previously rated recipes by users. It can be seen that both approaches perform reasonably similarly on MAE, while SVD CF marginally beats KNN CF due to a lower error on RMSE (0.819 to 0.885). As seen in Table 2, it is observed that the accuracy is high when the number of recommendations is less and low when more recipes are recommended. This is because the number of recommendations for the user increases, the similarity with the recipes decreases, as users connected to the target user is less likely to similar.

5. Conclusion

This article has proposed a novel model to provide a healthy recipe recommendation. The model first predicted the top similar healthy recipes and followed by that a novel recipe is created by considering the nutritional values. The novel recipe is recommended by substituting a similar ingredient instead of an original ingredient of the similar top recipes. As evident from the presented results, the proposed model can improve the average healthiness of the recommended recipes (for a top-10 recipe).

Though the present application seems feasible and user-friendly, recommendations can only be made to a known user. An auto-rating prediction functionality could be included based on the ingredients present. Personalized recommendations can be made by tracking the dietary practices of the user, thereby enabling the user to include/exclude ingredients in the recommendation. Based on the user choices, including regional preferences along with seasonal preferences would increase the level of personalization in the recommendation.

References

[1] Adomavicius, Gediminas, and Alexander Tuzhilin. "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." IEEE transactions on knowledge and data engineering 17, no. 6 (2005): 734-749.

[2] Wu, Shiwen, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. "Graph neural networks in recommender systems: a survey." ACM Computing Surveys 55, no. 5 (2022): 1-37.

[3] Fan, Wenqi, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. "Graph neural networks for social recommendation." In The world wide web conference, pp. 417-426. 2019.

[4] Teng, Chun-Yuen, Yu-Ru Lin, and Lada A. Adamic. "Recipe recommendation using ingredient networks." In Proceedings of the 4th annual ACM web science conference, pp. 298-307. 2012.

[5] Tang, Yew Siang, Anita Hanzhi Zheng, and Nicholas Lai. "Healthy Recipe Recommendation using Nutrition and Ratings Models." (2019).

[6] Chen, Meng, Xiaoyi Jia, Elizabeth Gorbonos, Chinh T. Hoang, Xiaohui Yu, and Yang Liu. "Eating healthier: Exploring nutrition information for healthier recipe recommendation." Information Processing & Management 57, no. 6 (2020): 102051.

[7] Khan, Mansura A., Ellen Rushe, Barry Smyth, and David Coyle. "Personalized, health-aware recipe recommendation: an ensemble topic modeling based approach." arXiv preprint arXiv:1908.00148 (2019).

[8] Yamanishi, Ryosuke, Naoki Shino, Yoko Nishihara, Junichi Fukumoto, and Aya Kaizaki. "Alternative-ingredient recommendation based on co-occurrence relation on recipe database." Procedia Computer Science 60 (2015): 986-993.

[9] Trang Tran, Thi Ngoc, Müslüm Atas, Alexander Felfernig, and Martin Stettinger. "An overview of recommender systems in the healthy food domain." Journal of Intelligent Information Systems 50 (2018): 501-526.

[10] Rehman, Faisal, Osman Khalid, Kashif Bilal, and Sajjad A. Madani. "Diet-right: A smart food recommendation system." KSII Transactions on Internet and Information Systems (TIIS) 11, no. 6 (2017): 2910-2925.

[11] Grover, Aditya, and Jure Leskovec. "node2vec: Scalable feature learning for networks." In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 855-864. 2016.

[12] https://www.kaggle.com/elisaxxygao/foodrecsysv1

[13] https://snap.stanford.edu/node2vec/

Author's biography

K.Vani received master degree in Computer Science and Engineering from Anna University in 2014. She is an Assistant Professor in the Department of Computer Science and Engineering at PSG college of Technology. Her research focus includes social network analytics, personalization and privacy mechanisms. She is an associate member of the IE.

K. Latha Maheswari received her Master’s and Bachelor’s degrees in Electrical and Electronics Engineering from Anna University, Chennai, in 2012 and 2010, respectively. Currently, she is working as an Assistant Professor in the Department of Electrical and Electronics Engineering at PSG College of Technology and pursuing her PhD degree at Anna University. In addition, she authored or co-authored various conference papers and book chapters. Her research interests include smart grid protection and control, distribution systems fault diagnosis and service restoration strategies, renewable integration, theory and applications of soft-computing techniques, and microgrid control.