Market2Dish: Health-aware Food Recommendation

WENJIE WANG, National University of Singapore
LING-YU DUAN, Peking University
HAO JIANG, Shandong University
PEIGUANG JING, Tianjin University
XUEMENG SONG and LIQIANG NIE, Shandong University

With the rising incidence of some diseases, such as obesity and diabetes, the healthy diet is arousing increasing attention. However, most existing food-related research efforts focus on recipe retrieval, user-preference-based food recommendation, cooking assistance, or the nutrition and calorie estimation of dishes, ignoring the personalized health-aware food recommendation. Therefore, in this work, we present a personalized health-aware food recommendation scheme, namely, Market2Dish, mapping the ingredients displayed in the market to the healthy dishes eaten at home. The proposed scheme comprises three components, namely, recipe retrieval, user health profiling, and health-aware food recommendation. In particular, recipe retrieval aims to acquire the ingredients available to the users and then retrieve recipe candidates from a large-scale recipe dataset. User health profiling is to characterize the health conditions of users by capturing the textual health-related information crawled from social networks. Specifically, to solve the issue that the health-related information is extremely sparse, we incorporate a word-class interaction mechanism into the proposed deep model to learn the fine-grained correlations between the textual tweets and pre-defined health concepts. For the health-aware food recommendation, we present a novel category-aware hierarchical memory network–based recommender to learn the health-aware user-recipe interactions for better food recommendation. Moreover, extensive experiments demonstrate the effectiveness of the health-aware food recommendation scheme.

CCS Concepts: • **Information systems** → **Recommender systems**; **Multimedia and multimodal retrieval**; **Collaborative filtering**; • **Applied computing** → **Health care information systems**;

Additional Key Words and Phrases: User health profiling, health-aware food recommendation, recipe retrieval

ACM Reference format:
Wenjie Wang, Ling-Yu Duan, Hao Jiang, Peiguang Jing, Xuemeng Song, and Liqiang Nie. 2021. Market2Dish: Health-aware Food Recommendation. *ACM Trans. Multimedia Comput. Commun. Appl.* 17, 1, Article 33 (April 2021), 19 pages.
https://doi.org/10.1145/3418211

This work is supported by the National Key Research and Development Project of New Generation Artificial Intelligence, No.:2018AAA0102502; the National Natural Science Foundation of China, No.:61772310, and No.:U1936203; the Shandong Provincial Natural Science Foundation, No.:ZR2019JQ23; the Innovation Teams in Colleges and Universities in Jinan, No.:2018GXRC014; the Shandong Provincial Key Research and Development Program, No.:2019JZZY010118.
Authors' addresses: W. Wang, National University of Singapore, Singapore, 117417; email: wenjiewang96@gmail.com; L.-Y. Duan, Peking University, Beijing, China, 100871; email: lingyu@pku.edu.cn; H. Jiang, Shandong University, Qingdao, China, 266237; email: jianghaosdu@mail.sdu.edu.cn; P. Jing, Tianjin University, Tianjin, China, 300072; email: pgjing@tju.edu.cn; X. Song and L. Nie, Shandong University, Qingdao, China, 266237; emails: {sxmustc, nieliqiang}@gmail.com.

1551-6857/2021/04-ART33 $15.00
https://doi.org/10.1145/3418211

1 INTRODUCTION

Food is essential for human beings. Once people's basic requirement for food is satisfied, they focus on the pursuit of a healthier diet. Nowadays, countless people are being plagued by many diseases due to the unhealthy diet. According to the World Health Report 2018,[1] the incidence rate of many diet-related diseases is increasing rapidly over the world, such as diabetes, obesity, and malnutrition. Based on the healthy diet tips,[2] people with different health conditions ought to be endowed with a personalized healthy diet. For example, the diabetics are required to eat more whole-grain cereals and avoid sweet food. However, people are usually caught in a dilemma when deciding what ingredients they should buy from the market according to their physical conditions. Many reasons lead to such dilemma: (1) People's personal knowledge and experience on food are limited, while a healthy and delicious dish is usually complex, involving many ingredients and cooking skills. And (2) many people cannot explicitly and precisely describe their own health conditions, let alone correctly judge what kinds of food are healthy ones.

Despite the great value of a health-aware personalized food recommendation system, there still exist many challenges: (1) The amount of ingredients available in the market is huge. It is non-trivial to learn a mapping function between the ingredients available to users and their expected food. (2) How to obtain the health profile of a given user needs study. After all, the health-related user information we can acquire from users or the Internet is extremely sparse. And (3) jointly considering the existing food knowledge and the health profiles of different people to recommend the personalized healthy ingredients or food is a tough issue. In this work, we try to overcome the aforementioned challenges by proposing a health-aware food recommendation framework, which aims to profile user health, and then recommend the healthy food to users.

Actually, the rapid expansion of Internet has provided much information for us to solve the aforementioned challenges: (1) The popularity of smartphones enables us to capture and record our lives visually and vividly, from which rich ingredient information in the market can be obtained. (2) People are keen to enjoy social networks, such as Twitter[3] and Weibo,[4] and share personal information on these platforms, such as activities, likes, and dislikes. These social media platforms inadvertently expose personal health information, more or less. For example, from their shared tweets, it can be easily found out that one is losing weight while the other might get pregnant. And (3) the explosive growth of online information brings us massive food data as well. Many high-quality and food-related data is available in some recipe-sharing websites (e.g., Yummly[5] and Meishijie[6]). In addition, extensive food-related health knowledge can be acquired in many forms, such as what kind of food will make you fat or lower your blood pressure.

Indeed, several prior methods on food recommendation have paid attention to satisfy users' taste preferences by modeling the historical user-item interactions and predicting the future ones [22]. Nowadays, health-aware food recommendation has become an emerging research topic. For example, Ge et al. [20] leveraged the so-called "calorie balance function" to incorporate calorie counts into the food recommendation method. Elsweiler et al. [15] explored the feasibility of substituting meals recommended to users typically with similar but healthier dishes with the help of the user study. Although great success has been achieved, they failed to build user health profile and recommend them personalized healthy diets based on the available ingredients.

[1] https://www.who.int/gho/publications/world_health_statistics/2018/en/.
[2] https://www.meishij.net/.
[3] https://twitter.com/.
[4] https://www.weibo.com/.
[5] http://www.yummly.com/.
[6] https://www.meishij.net/.

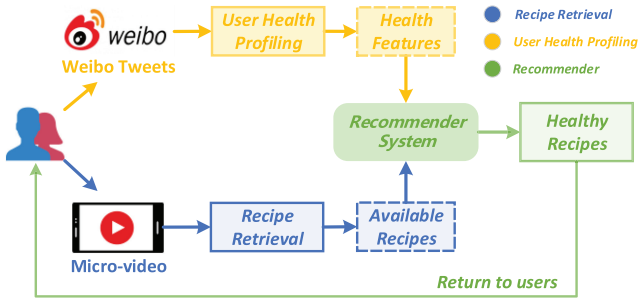

Fig. 1. Illustration of our proposed scheme. Recipe retrieval is applied to retrieving the recipe candidates for users; meanwhile, user health profiling aims to extract users' health features; ultimately, the recommender returns the personalized healthy recipes to users.

To help users decide the personalized healthy diet, we propose a health-aware food recommendation scheme, as illustrated in Figure 1. It comprises three components: recipe retrieval, user health profiling, and recipe recommendation. To be more specific, (1) on recipe retrieval, users have many ways to input their available ingredients from markets, such as selecting from ingredient candidates, voice input, and micro-video input. In this work, we explore the possibility of micro-video input, by which users can record various ingredients in the market by a micro-video conveniently, and then Inception-v3 Net [50] is utilized to capture ingredients via the multi-label image classification. Based on the available ingredients, we can retrieve many recipes for each user from a large-scale recipe dataset. Due to the high accuracy of existing methods on image/video classification, we focus more on the following two components in this work. (2) As for user health profiling, we pre-define some health concepts to represent the common health conditions with the help of domain experts. The health concepts are collected from different perspectives, such as age (e.g., teenagers and old people), occupation (e.g., office workers and students), and diseases (e.g., insomnia, hypertension, obesity, and malnutrition). And then user health profiling aligns users with the pre-defined health concepts by capturing the textual health-related information crawled from users' social accounts (e.g., Weibo). In this way, user health profiling is converted to a text classification problem. However, existing text classification methods usually learn a high-level latent representation for textual content, which is overwhelmed by much noise, since the health-related information collected from social platforms is extremely sparse. To alleviate this issue, we present a word-class interaction-based recurrent convolutional neural network to learn the fine-grained correlations between the words in users' tweets and the health concepts. And (3) the recipe recommendation system aims to jointly consider the retrieved recipe candidates and users' health features to accomplish the personalized food recommendation. Towards this end, we present a category-aware hierarchical memory network to learn the close correlations among users with the same health tags, the relations among the recipes with similar nutritive values, and the interactions between users and recipes. Specifically, the recommender divides the users and the recipes into different categories according to their health tags and nutritive values, respectively. And then it leverages the category-level and recipe-level matching scores to learn the inter-category difference and the intra-category similarity for better food recommendation. Finally, the healthy recipes for users are output by the recommender. To justify our model, we construct two large-scale food-related datasets: user health profiling dataset from Weibo and a health-aware food recommendation dataset. Extensive experiments demonstrate the superiority of our models in two tasks: user health profiling and health-aware food recommendation. Besides, we also develop a demo [27] to testify the effectiveness and efficiency of the whole scheme.

To sum up, the main contributions of our work are threefold:

- To the best of our knowledge, this is the first work on personalized health-aware food recommendation, mapping the ingredients displayed in the market to the dishes eaten at home. In addition, we release two high-quality food-related datasets and the involved codes[7] to facilitate the research community in this field.
- We present a word-class interaction-based text classification model to profile the users' health conditions via their sparse health-related information posted on the social networks.
- We propose a novel category-aware hierarchical memory network–based food recommendation system to learn the health-aware user-recipe interactions.

2 RELATED WORK

Due to the significance of food to human life and health, extensive research efforts have been dedicated to the food-related study. According to the latest food survey [38], the research in food computing falls into five main tasks, containing perception, recognition, retrieval, recommendation, and monitoring. In particular, food perception [37, 39, 42] studies how people perceive food from its characteristics, while food recognition [2, 29, 40, 43, 58, 59] aims to recognize and detect the categories or ingredients of the meals. Differently, some work also focuses on the recipe modification based on the original recipes [5, 48]. Besides, food retrieval [46] comprises visual image-based retrieval [31], textual recipe-based retrieval [53], and cross-modal recipe-image retrieval [9]. And food recommendation [20, 52] leverages multi-faceted information to recommend healthy and delicious food to users. By contrast, food monitoring [16, 47] is intended to analyze various health-related information, monitor and predict the public health based on the massive data from the social media. For this work, food recommendation, retrieval, and monitoring are three highly relevant research directions.

2.1 Food Recommendation

With the explosive growth of data on the Internet, recommendation systems [10, 11, 34, 55, 57] have been proven effective to alleviate the overloaded information. According to the recent survey [51], the studies in the food recommendation can be divided into five categories, namely, content-based, collaborative filtering-based, context-aware, hybrid-based, and health-aware food recommendation. To be more specific, content-based approaches [18] focus on the recipes, including the ingredients and food images; whereas collaborative filtering-based ones [19] leverage the classic collaborative filtering algorithms [25] to achieve food recommendation, such as Singular Value Decomposition [22] and Matrix Factorization (MF) [19]. For context-aware methods, Reference [32] explored the value of rich context about users in food recommendation, such as gender, hobbies, and culture. Moreover, hybrid-based approaches [19] usually integrate several existing methods for the recipe recommendation. However, all these strategies focus on recommending recipes based on the historical records of users to satisfy their preferences. Recently, incorporating healthiness into the food recommendation has attracted extensive research attention in this community [27]. For example, Ge et al. [20] integrated the calorie consumption into the food recommendation, and Markus et al. [45] estimated the influence of various recipe features on health-aware recipe recommendation. However, existing health-aware approaches cannot profile the user health and make personalized food recommendation at the ingredient level and recipe level due to the lack of large-scale datasets.

[7] https://github.com/WenjieWWJ/FoodRec.

2.2 Food Retrieval

Food-related data are usually presented in multiple modalities, including textual recipes, visual food images, and cooking videos. According to the retrieval types, the existing food retrieval methods comprise three categories: textual recipe-based retrieval [53], visual image-based retrieval [31], and cross-modal recipe-image retrieval [6, 7, 21, 46]. The textual recipe-based retrieval is designed to retrieve recipes by the textual recipe query, whereas the visual image-based retrieval focuses on the understanding of the image query. Lately, many efforts have been dedicated to retrieving the textual recipe based on an image query, namely, cross-modal recipe-image retrieval. For instance, Chen et al. [8] leveraged rich food attributes to retrieve the textual recipe given the food image as a query and achieved promising results. Different from the aforementioned methods, the recipe retrieval in this work is designed to first recognize the ingredients from the micro-videos taken in the market, and then retrieve recipes from the recipe dataset based on the captured ingredients.

2.3 Food Monitoring

According to the recent global digital report issued by *We Are Social*,[8] there are over 4.39B Internet users, including 3.48B social media users in the world today. And it is an inexorable trend that the Internet users are spending more time on social networks. Therefore, large-scale personal data are accumulated on the Internet, providing rich sources for the prediction and analysis of public health. Indeed, many efforts [1, 16] have been dedicated to investigating the public health issues, such as the national obesity and diabetes [1]. However, the existing work is limited in that it only studies some food-related patterns based on the data from social networks, such as the food consumption patterns [35] or the diabetes rates in different regions [14]. In this work, we turned to profile the user health based on the shared personal information in the social media and thereafter generated the personalized health-aware food recommendation.

Indeed, the task of user health profiling has been converted into a multi-label text classification problem by defining many user health tags. Therefore, this work is also closely related to many classic text classification methods, such as Fast Text Classifier (FastText) [28], Convolutional Neural Networks (TextCNN) [30], Recurrent Convolutional Neural Network (RCNN) [33], and Hierarchical Attention Networks (HAN) [61]. Although great success has been achieved in text classification, especially with the rise of the data-driven deep neural models, these representative methods are impotent in this task, because they usually learn a latent vector representation of the text content and then calculate the probability of each class by projecting the latent vector presentation with a fully connected (FC) layer. However, the health-related user information on the social networks is extremely sparse, thus learning high-level user representation from many tweets may be easily overwhelmed by the noise.

3 METHOD

In this article, our proposed health-aware food recommender system aims to profile the user health, and thereafter recommend the personalized healthy food based on the ingredients they could buy from the market. As illustrated in Figure 2, the whole scheme is divided into three components according to the roles, namely, recipe retrieval, user health profiling, and personalized health-aware recipe recommendation. In this section, we will detail them one-by-one.

3.1 Recipe Retrieval

The objective of this component is to acquire ingredients available to users, and then retrieve recipe candidates from large-scale recipe datasets. In the proposed framework, there are many ways to

[8] https://wearesocial.com/.

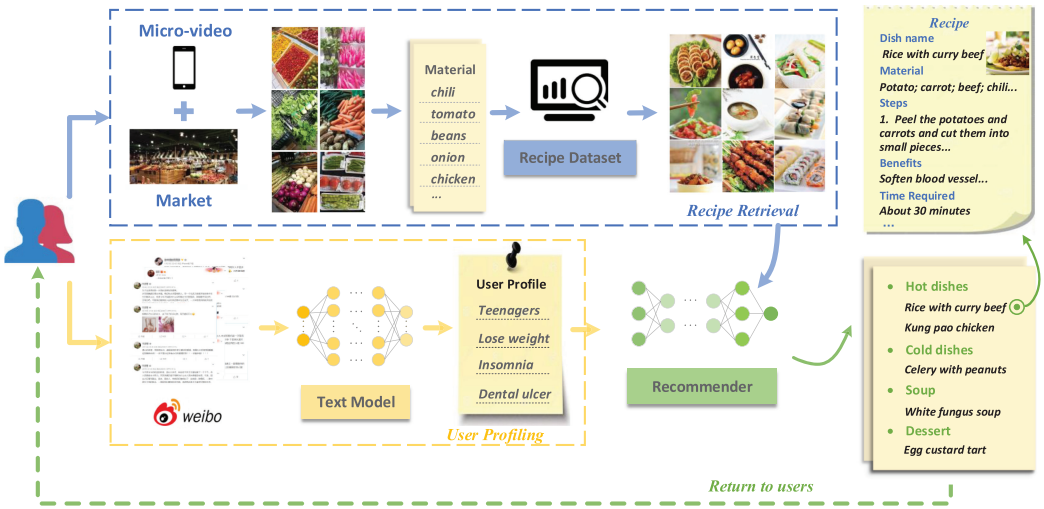

Fig. 2. Schematic illustration of our proposed model, comprising recipe retrieval, user health profiling, and health-aware recommender model. Ultimately, the personalized recipes with the cooking instructions are returned to users.

input ingredients for users, including selecting from the candidates, voice input, and micro-video input. In this work, we explore the possibility of micro-video input due to its higher difficulty. To this end, we sample images (frames) from the micro-videos taken by users in the market, recognize ingredients in the images via the multi-label image classification, and ultimately retrieve available recipes from a large-scale recipe dataset. In particular, we construct a large-scale dataset by human annotation, in which many ingredients are recorded by numerous images captured from the end markets. Based on our analyses and statistics on the ingredient images, there are almost 80 kinds of common ingredients, including various vegetables, seafood, and meat. Notably, an image usually contains multiple ingredients, and thus the ingredient recognition is essentially a task of multi-label image classification. Considering that lots of deep neural networks have achieved promising performance in image classification, we incorporate a pre-trained Inception-v3 Net [50] to tackle this multi-label image classification problem due to its high accuracy. Upon experiments, Inception-v3 Net achieves promising accuracy up to 95.5%, which extremely facilitates the retrieval of available recipes from the recipe dataset. In the light of this, we obtain extensive recipe candidates based on the available ingredients. From the experiments, we find that micro-video input is also a feasible way to input ingredients for users.

3.2 User Health Profiling

As introduced in Reference [38], the online social networks with billions of users have provided extensive user data for the food-related prediction and monitoring. And as for user profiling, some people have proposed special ways to learn user behavior profiles in social networks [26, 62, 63]. In this work, we design a supervised neural model to profile the user health by the rich user-generated content distributed over the social networks. The experts in the health domain jointly consider various factors, such as gender, age, and common diseases, and then help us propose many health tags to characterize the common identities and health conditions of different people, such as teenager, fitness enthusiast, and the diabetic. Notably, each user has at least one health tag (related to the job or age) and most of them have multiple ones. Therefore, the task of user health profiling will be converted into a multi-label classification problem: Given the user generated content on the

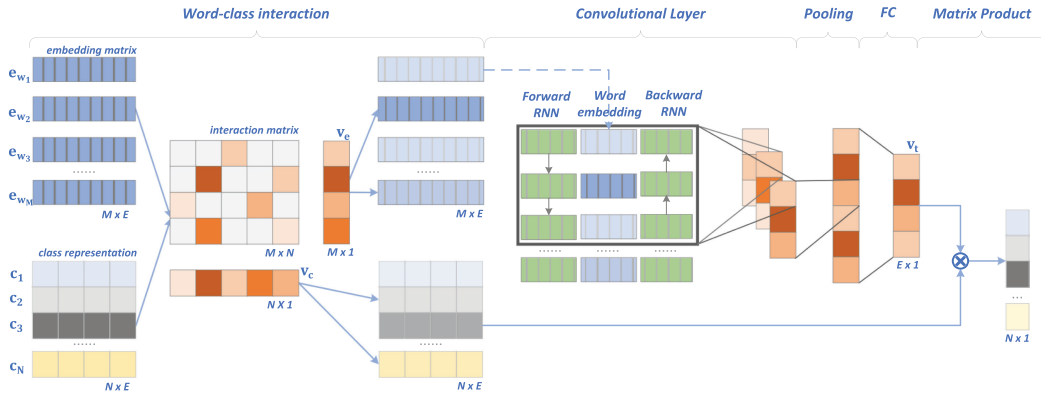

Fig. 3. Schematic illustration of our proposed WIRCNN. WIRCNN incorporates an interaction mechanism to learn the fine-grained correlations between the words and the classes, and leverages Bi-RNN and CNN to distill the textual contents for more distinguishable features.

social networks, our proposed model will classify the user into one or several classes (i.e., different user health tags).

To learn the correlation between the sparse health-related information and the health tags, we propose a Word-class Interaction based Recurrent Convolutional Neural Network, named WIRCNN for short, as shown in Figure 3. Motivated by the phenomenon that only several keywords in the tweets have direct correlation with the health tags, we introduce the word-class interaction [54] to learn the fine-grained matching relations between the keywords and the classes. Thenceforth, WIRCNN leverages Bidirectional Recurrent Neural Networks (Bi-RNN) and Convolutional Neural Network (CNN) to encode the weighted word embeddings, and then outputs the probabilities over the classes via the matrix multiplication.

Formally, given the personal information and many tweets of a user, WIRCNN first concatenates them into a sequence of ordered tokens $\{w_1, w_2, \ldots, w_M\}$ by several special separators (e.g., using _eos_ to represent the end of sentence). Note that in the sequence, the tokens regarding personal information (e.g., age) are ranked before the tweets. We regard the matrix $\mathbf{W_c} \in \mathbb{R}^{N \times E}$ as the class representation, where N and E refer to the number of classes and the feature size, respectively. It is worth noting that the aforementioned FC layer in the existing methods can be interpreted as the class representation $\mathbf{W_c}$, and it essentially estimates the similarity between the latent vector representations of textual contents and the class representations via the matrix multiplication [44]. Thereafter, WIRCNN embeds the given tokens into an embedding matrix $\mathbf{W_e} \in \mathbb{R}^{M \times E}$ and computes the interaction matrix $\mathbf{I} \in \mathbb{R}^{M \times N}$ by the following equations:

$$\begin{cases} \mathbf{W_e} = [\mathbf{e}_{\mathbf{w}_1}, \mathbf{e}_{\mathbf{w}_2}, \ldots, \mathbf{e}_{\mathbf{w}_M}]^T, \\ \mathbf{I} = \mathbf{W_e}\mathbf{W_c^T}, \end{cases} \tag{1}$$

where M is the number of given tokens, and $\mathbf{e}_{\mathbf{w_i}}$ denotes the embedding of the token w_i. Then the max pooling function is, respectively, applied to two different dimensions of $\mathbf{I}$. In the light of this, we can get two vectors $\mathbf{v_e} \in \mathbb{R}^M$ and $\mathbf{v_c} \in \mathbb{R}^N$. Based on these two vectors, the matrices $\mathbf{W_c}$ and $\mathbf{W_e}$ are weighted as follows:

$$\begin{cases} \mathbf{W'_e} = \mathbf{v_e} \odot \mathbf{W_e}, \\ \mathbf{W'_c} = \mathbf{v_c} \odot \mathbf{W_c}, \end{cases} \tag{2}$$

where $\odot$ refers to the element-wise multiply, $\mathbf{W'_e}$ and $\mathbf{W'_c}$ denote the weighted matrices of $\mathbf{W_c}$ and $\mathbf{W_e}$, respectively.

Afterwards, WIRCNN encodes the weighted word embeddings $\mathbf{W}'_{\mathbf{e}}$ via a Bi-RNN step-by-step and concatenates the bidirectional hidden states with the weighted word embedding at each step for the convolutional layer. In the convolutional layer, WIRCNN uses multiple filters with varying window sizes to obtain the text features, and then applies a max-overtime pooling operation [12] over each feature map, which takes the highest feature value as the representation of this feature map. At last, a FC layer is used to project features from the pooling layer into a high-dimensional space in the feature size E. And the probability $\mathbf{p}$ over all classes is calculated as

$$\mathbf{p} = Sigmoid(\mathbf{v}_{\mathbf{t}}^{\mathrm{T}} \mathbf{W}'_{\mathbf{c}}), \tag{3}$$

where $\mathbf{v_t}$ is the text feature from the FC layer, $\mathbf{W}'_{\mathbf{c}}$ refers to the weighted class representation, and the Sigmoid function maps the feature values into the interval $[0, 1]$. Ultimately, a cross-entropy loss is applied to optimizing the whole neural model.

3.3 Health-aware Recipe Recommendation

The recommendation task in this component can be formulated as: Given a user u and the available recipe candidates $\{r_1, r_2, \ldots, r_{N_r}\}$, the recommender ranks these recipes based on the health tags of the user u and the nutritive value of the recipes. For each user, we construct the positive samples $\mathcal{Y}^+ = \{r_1^p, r_2^p, \ldots, r_{N_p}^p\}$ and the negative ones $\mathcal{Y}^- = \{r_1^n, r_2^n, \ldots, r_{N_n}^n\}$ from the recipe candidates according to the food-related health tips. Thereafter, the recommender can be trained by the health-aware user-recipe interactions.

In fact, close correlations not only exist among the users with similar health tags, but also the recipes sharing the same nutritive values. Intuitively, the users under the similar health conditions usually have similar diet habits; in other words, dishes made from the similar ingredients will be suitable for the same class of users. To leverage these correlations to improve the performance of food recommendation explicitly, we propose a category-aware hierarchical memory network to learn the intra-category similarity and the inter-category difference. All users fall into N categories with different health tags, and the recipes are also divided into N_c categories for different health needs, such as low calorie and nutritional supplements. Notably, each user and recipe could belong to multiple categories due to the multiple health tags or various nutritive values.

As illustrated in Figure 4, the proposed recommender includes four components, namely, general memory, personal memory, category embedding, and recipe embedding. In particular, recipe embedding and category embedding are utilized to encode the recipes and N_c recipe categories into vectors, respectively. In addition, each user has a personal memory while a health tag corresponds to a general memory. The personal memory and the general one share the same internal structure, including a high-level memory vector and N_c low-level memory vectors. Intuitively, the high-level memory vector records the health-aware preference of users for the recipe categories while the low-level ones remember the preference for each recipe in N_c categories. In particular, the high-level memory corresponds to the category embedding and the low-level one corresponds to the recipe embedding. The general memory is applied to learning the common characteristics among users with the similar health conditions. For instance, the users who are losing weight should eat some low-fat food.

Formally, given a user u and a recipe i, the recommender first obtains the personal memory of the user i, the recipe embedding $\mathbf{v}_{\mathbf{i}}^{\mathbf{r}}$, and the category embedding $\mathbf{v}_{\mathbf{i}}^{\mathbf{c}}$ of the recipe i. Afterwards, we calculate the score $\hat{y}_{u,i}$ for the user u and the recipe r as

$$\hat{y}_{u,i} = \alpha \times Sim(\mathbf{v}_{\mathbf{u}}^{\mathbf{h}}, \mathbf{v}_{\mathbf{i}}^{\mathbf{c}}) + (1 - \alpha) \times Sim(\mathbf{v}_{\mathbf{u}}^{\mathbf{l,i}}, \mathbf{v}_{\mathbf{i}}^{\mathbf{r}})), \tag{4}$$

where $\mathbf{v}_{\mathbf{u}}^{\mathbf{h}}$ refers to the high-level memory vector of the user u, $\mathbf{v}_{\mathbf{u}}^{\mathbf{l,i}}$ denotes the low-level memory vector of the user u regarding the category of the recipe i, and α is a hyper parameter to adjust

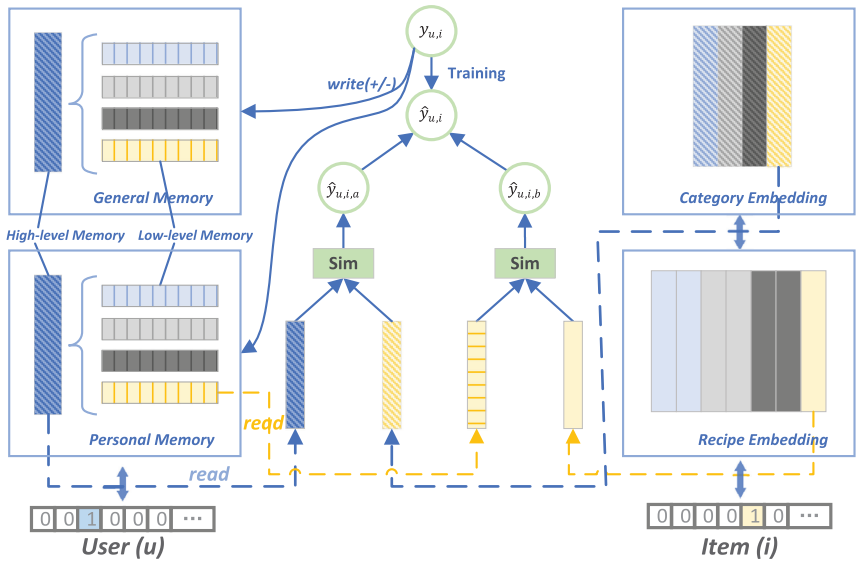

Fig. 4. The proposed recommender, consisting of four components: general memory, personal memory, category embedding, and recipe embedding. Given a user u and a recipe i, the recommender first obtains the personal and general memories of the user u and the recipe and category embeddings of the recipe i. Thereafter, it calculates the matching score hierarchically.

the contribution of the high-level and low-level similarities. As to the $Sim(\mathbf{a}, \mathbf{b})$ function, we have tried several operations to calculate the similarity between $\mathbf{a}$ and $\mathbf{b}$, such as cosine similarity, dot product, and multi-layer perceptron (MLP). Besides, it is worth noting that if the recipe i belongs to multiple categories, the category embedding $\mathbf{v}_\mathbf{i}^\mathbf{c}$ and the low-level personal memory vector $\mathbf{v}_\mathbf{u}^\mathbf{l,i}$ will be computed by taking the mean value of the corresponding vectors. Finally, a binary cross entropy loss is applied to optimizing the recommender, which can be formulated as

$$\ell = -\sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{Y}^+} y_{u,i}\log\hat{y}_{u,i} - \sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{Y}^-}(1-y_{u,i})\log(1-\hat{y}_{u,i}), \tag{5}$$

where $\mathcal{U}$ represents the set of all users, $y_{u,i}$ is the label of the recipe i for the user u. Meanwhile, $y_{u,i}$ is 1 if the recipe i is a positive sample, otherwise, 0.

In addition, we define a write operation to update the personal memory and the general memory dynamically, and then leverage the general memory to update the corresponding personal memories with the same health tags at a certain frequency in the training. This operation will change the preference of the user u for the recipe i explicitly and leverage some common characteristics in the general memory to modify the preference of the user u. Specifically, assuming that the user u has a health tag t corresponding to the general memory g_t, the general memory g_t will be updated through the following equation:

$$\begin{cases} \mathbf{v}_\mathbf{t}^\mathbf{h} \leftarrow \mathbf{v}_\mathbf{t}^\mathbf{h} + z_{u,i}\beta^h\mathbf{v}_\mathbf{i}^\mathbf{c}, \\ \mathbf{v}_\mathbf{t}^\mathbf{l,c} \leftarrow \mathbf{v}_\mathbf{t}^\mathbf{l,c} + z_{u,i}\beta^l\mathbf{v}_\mathbf{i}^\mathbf{r}, \end{cases} \tag{6}$$

where $\mathbf{v}_\mathbf{t}^\mathbf{h}$ denotes the high-level memory vector of g_t, $\mathbf{v}_\mathbf{t}^\mathbf{l,c}$ refers to the low-level one of g_t corresponding to the category c of the recipe i, β^h and β^l are hyper parameters, $\mathbf{v}_\mathbf{i}^\mathbf{c}$ and $\mathbf{v}_\mathbf{i}^\mathbf{r}$ are the category embedding and the recipe embedding of the recipe i, respectively. In addition, $z_{u,i}$ equals to 1 if $i \in \mathcal{Y}^+$, otherwise, -1. Moreover, the personal memory of the user u is also updated by $\mathbf{v}_\mathbf{i}^\mathbf{c}$

and $\mathbf{v}_i^{\mathrm{r}}$, similar to the general memory g_t. Intuitively, the write operation will increase/decrease the similarity between the memories of the user u and the embeddings of the recipe i, which changes the preference of the user u for the recipe i explicitly. Notably, if a user has multiple health tags, all related general memories will be updated in this manner. Subsequently, the personal memory of the user u can be updated by the general memory g_t:

$$\begin{cases} \mathbf{v}_u^{\mathrm{h}} \leftarrow \mathbf{v}_u^{\mathrm{h}} + \lambda^h \mathbf{v}_t^{\mathrm{h}}, \\ \mathbf{v}_u^{\mathrm{l}} \leftarrow \mathbf{v}_u^{\mathrm{l}} + \lambda^l \mathbf{v}_t^{\mathrm{l}}, \end{cases} \tag{7}$$

where λ^h and λ^l are hyper parameters, $\mathbf{v}_u^{\mathrm{h}}$ and $\mathbf{v}_u^{\mathrm{l}}$ denote the high-level and low-level personal memory vectors of the user u, respectively. All low-level personal memories of the user u are updated by the corresponding low-level general memories. In particular, if a user with several health tags has more than one general memory, the personal memory will be updated by the average of these general memories.

4 EXPERIMENTS

4.1 Data Collection

To achieve the personalized health-aware food recommendation based on the available ingredients for users in the market, we constructed two food-related datasets: user health profiling dataset and food recommendation dataset. And all the experimental data have been released to facilitate the research in the food recommendation domain. More details about the datasets and the specific cases can be found in our website.[9]

4.1.1 User Health Profiling Dataset. Nowadays, social networks have become an essential part of our daily life. Extensive users share their activities, feelings, and hobbies on these social network platforms every day, and this user-related information helps us to analyze users' identities, emotions, and even their health conditions. To profile the user health, we crawled extensive users' tweets along with their personal information (e.g., sex, age, and occupation) from Weibo,[10] one of the biggest social media platforms in China. Thereafter, much work on data cleaning and preprocessing was done to improve the quality of the dataset, such as the removal of duplicate and noisy tokens. In addition, we collected the health tags under the guidance of domain experts to cover the common health conditions of the Weibo users from multiple perspectives, such as age and diseases. In particular, the health tags derived from the users' personal information mainly include age-related ones (i.e., school-age child, teen, middle-age users, elderly). And other tags obtained based on tweets are like losing weight, insomnia, and so on. Eventually, we acquired 64,657 Weibo users with 96 health tags (e.g., pregnant women, insomnia, hypertension, and obesity), and then we annotated users with the health tags based on their tweets and personal information from Weibo via a semi-automatic annotation method, integrating the keyword-based filter rules and human inspection. First, we retrieved the health-related tweets from the extensive tweets for each user by some health-related keyword matching rules, and then annotators assigned each user with several appropriate health tags based on the retrieved tweets and the user's personal information. In particular, we defined many health-related keywords such as slimming, exercise, and disease, and selected the tweets containing these keywords for the convenient annotation. Besides, each user has one health tag at least, and usually has multiple ones. On average, each user has more than 40 tweets crawled from Weibo and 3.89 health tags.

[9] https://healthawarerec.wixsite.com/foodrec.
[10] https://www.weibo.com.

4.1.2 Food Recommendation Dataset. Existing recipe websites provide us abundant food-related data for the research in the food domain. We crawled a large-scale Chinese recipe dataset from Meishijie,[11] consisting of over 6.6K recipes, of which each contains rich information, such as the recipe name, ingredients, cooking instructions, dish pictures, benefits, cooking time, and cooking difficulty. In particular, Meishijie provides the benefits of each recipe, such as helps digestion, anti-diarrhea, and treatment of insomnia. To reduce the number of the category, we summarized all the recipes into four categories from the perspective of health (i.e., losing weight, health care, nutritional supplements, and disease recovery). Itis worth noting that a recipe may belong to more than one category due to their multiple nutritional benefits. Besides, we acquired extensive healthy diet tips for people with different health conditions; for example the diabetics should eat more food rich in fiber and vitamins, such as carrots and celery, and meanwhile avoid sausage, lard oil, and other ingredients rich in sugar and cholesterol. Based on the healthy diet tips, we could collect over 20 appropriate and inappropriate ingredients for the users with one kind of health tag. And then by matching the ingredients in recipes, we constructed lots of positive and negative health-aware dish samples from the crawled recipe dataset for each Weibo user. In detail, we collected 28,800 health-aware triples <health tag, suitable recipe, unsuitable recipe>. Each health tag corresponds to at least 300 triples. Then, to simulate the ingredients that users could buy from the market, we assigned many ingredients for each Weibo user randomly to record the available ingredients for him/her. Therefore, the positive and negative samples of the users with multiple health tags were filtered by the available ingredients, and then we acquired extensive user-recipe pairs from the perspective of health. Ultimately, the food recommendation dataset consists of 64,657 training samples, and each of them contains a Weibo user, the health tags, the available ingredients, about 340 positive recipes, and 100 negative recipes.

4.2 Experimental Settings

4.2.1 Hyper Parameters. In WIRCNN, the feature size E, the number of health tags N, and the length of users' Weibo tweets M are 128, 96, and 390, respectively. Besides, the size of hidden states of the Bi-RNN was set as 128. As to the recommender, the numbers of users and recipes are 64,657 and 4,548, respectively. The category of recipes is four. In addition, we chose inner product as the function $Sim(\mathbf{a}, \mathbf{b})$ due to the better experimental performance. The write operation is employed once for each user at each iteration. For the two neural models, we leveraged Adam as the optimizer with the learning rate initialized as 0.001.

4.2.2 Embedding Pre-training. Lots of experiments have proven that parameter initialization plays a great role in many natural language processing and computer vision tasks [36]. In addition, the health tags of users, the recipe images, and the recipe ingredients provide rich context information for the food recommendation. Therefore, we discussed the performance of two kinds of pre-training methods on food recommendation: item2vector [3] and initialization by deep models. In this work, we used them to pre-train the latent embeddings of users and recipes. In particular, item2vector is derived from word2vector [36]. The users and recipes were divided into many groups for the pre-training according to the predefined health tags and recipe categories. As to the deep model initialization, we designed a deep neural model to encode rich context information for the pre-training of the latent embeddings of users and recipes. The latent embeddings of users were calculated from their health tags by a FC layer while the latent embeddings of recipes were acquired by the concatenation of the visual features of recipe images and the textual features of recipe ingredients. In particular, a VGGNet-19 [49] was applied to extracting the visual features

[11] http://meishij.net/.

of recipe images; meanwhile a TextCNN model was incorporated to distill the embeddings of ingredients and output the textual features of ingredients. Notably, the embeddings of ingredients were randomly initialized in this model. Thereafter, we leveraged a MLP model to calculate the matching score between the latent embeddings of users and recipes. By training this deep neural model, we can obtain the latent embedding for each user and recipe. Ultimately, these pre-trained vectors were utilized to initialize the personal memory and the recipe embedding in our proposed recommender. It is worth noting that the high-level and low-level memory vectors of each user were initialized with the same latent embedding. Moreover, the general memory and the category embedding took the mean values of the corresponding personal memories and recipe embeddings, respectively.

4.2.3 Baselines. To evaluate our proposed models, we compared them with several state-of-the-art methods in two tasks, respectively. Regarding the task of user health profiling, we compared the proposed WIRCNN with the following baselines:

- **FastText** [28] averages the word embeddings of textual content, and then predicts the probability of classification.
- **TextCNN** [30] leverages multiple filters to obtain textual features and outputs the probability distribution over all classes by a fully connected softmax layer.
- **RCNN** [33] employs a Bi-RNN to capture the contextual information and a max-pooling layer to acquire the key features.
- **HAN** [61] incorporates the hierarchical attention network to learn the attention weights of textual features at the word level and the sentence level.
- **EXplicit interAction Model (EXAM)** [13] incorporates the interaction mechanism [54] into text classification and achieves promising performance.

As to the health-aware food recommendation, we incorporated several general recommendation frameworks as the baselines:

- **MF** is the most popular collaborative filtering algorithm, conducting recommendation by calculating the inner product between the latent embeddings of the item and the user.
- **Generalized Matrix Factorization (GMF)** [23] generalizes the inner product of MF towards a non-linear neural layer.
- **Neural collaborative filtering (NCF)** [23] employs a MLP model to replace the inner product of MF with the aim of capturing more useful information between the latent embeddings of the user and the item.
- **Neural matrix factorization (NeuMF)** [23] is a fusion model of GMF and NCF (MLP), concatenating the features from GMF and NCF (MLP) to output the probability by a non-linear neural layer.

To learn more representative latent embeddings of users and items, we also computed the latent embeddings of all baselines by deep neural models with the same setting in Section 4.2.2. The difference is that the parameters of deep models in baselines for feature extraction can be optimized.

4.2.4 Evaluation Metrics. In the task of user health profiling, we employed 10-fold cross-validation to estimate the performance of WIRCNN and the baselines. In addition, we adopted *micro-precision* (micro-P), *micro-recall* (micro-R), *micro-F1*, and *macro-F1* to objectively evaluate their performance following the former work [30, 33]. In particular, micro-averaged scores (i.e., micro-P, micro-R, and micro-F1) are calculated by the global confusion matrix, the sum of the confusion matrices of all categories; whereas macro-F1 is computed by first calculating the per-category F1 score and then taking the average of the F1 scores for all categories [60].

Table 1. Performance Comparison between the Baselines and WIRCNN Regarding the Task of User Health Profiling

Methods	Macro-F1	Micro-F1	Micro-P	Micro-R
FastText	0.1014±0.0023	0.6271±0.0036	0.8271±0.0076	0.5049±0.0059
EXAM	0.7322±0.0134	0.8795±0.0067	0.8914±0.0074	0.8680±0.0073
RCNN	0.8783±0.0143	0.9752±0.0009	0.9788±0.0027	0.9717±0.0023
TextCNN	0.8820±0.0120	0.9696±0.0012	0.9630±0.0025	0.9761±0.001
HAN	0.8372±0.0192	0.9612±0.0018	0.9751±0.0025	0.9473±0.0021
Ours	**0.9017±0.0113**	**0.9784±0.0012**	**0.9811±0.0017**	**0.9762±0.0014**

We employ 10-fold cross-validation and report mean±standard deviation. The p-values of the student t-test are much smaller than 0.05, indicating that WIRCNN is significantly superior.

Micro-averaged scores give the equal weights to each sample-category pair from the micro perspective, and hence the samples belonging to more categories will occupy larger weights; whereas the macro-averaged ones treat each category equally at the macro level [60].

As to the food recommendation task, we employed *leave-one-out* evaluation method, which is widely adopted in existing recommendation studies [4, 23, 24]. In particular, for each user, we left a positive sample and 50 negative samples as the test data and leveraged the remaining positive and negative samples for training. We ran all models 10 times and randomly chose different test data every time. To judge the proposed recommender, we utilized *Hit Ratio* (HR), *Normalized Discounted Cumulative Gain* (NDCG), and *Area Under the Roc Curve* (AUC) to justify the performance. In the testing period, we selected the top-5 and top-10 recipes from the recommendation list for each user, and then, respectively, calculated the averaged HR@5, NDCG@5, HR@10, and NDCG@10 for all the testing users. Intuitively, AUC measures the probability that the recommender ranks a randomly chosen positive sample higher than a negative sample, and HR@k represents the hit ratio that a positive sample is ranked at the top-k positions, while NDCG assigns higher weights to the hits at the top ranks. For all metrics, higher scores denote better performance.

4.3 Overall Performance

4.3.1 Evaluating the User Health Profiling. Table 1 summarizes the performance of all baselines and WIRCNN with respect to several standard metrics on the task of user health profiling. Following prior work [17, 56, 57], we conducted significance test to evaluate the stability of the proposed method. From Table 1, we can observe the following points: (1) WIRCNN consistently achieves the superior performance than all the baselines, especially under the metrics of Micro-F1 and Macro-F1. This reflects that interaction mechanism, Bi-RNN, and CNN in WIRCNN can capture more health-related user information from the Weibo tweets for the classification, demonstrating the effectiveness of our proposed WIRCNN on the sparse user health profiling dataset. (2) HAN outperforms the other baselines by attentively learning the textual representation at the word and sentence levels. Incorporating the success of the interaction mechanism in WIRCNN, we could conclude that attentively extracting textual features is really helpful in this task. And (3) FastText is the worst one among the baselines, probably because it treats all the words equally in the user's tweets and hence incorporates much noise.

4.3.2 Evaluating the Health-aware Food Recommendation. The performance comparison between the baselines and our proposed recommender is presented in Table 2 and Figure 5. Notably, the latent embeddings of users and recipes are obtained by the deep model in the baselines and our recommender. By contrast, we can have the following findings: (1) the proposed recommender significantly outperforms the baselines. Besides, in Figure 5, the performance of our recommender

Table 2. Performance Comparison between the Baselines and Our Proposed Recommender on the Task of Health-aware Food Recommendation

Methods	HR@5	NDCG@5	HR@10	NDCG@10	AUC
MF	0.8393±0.0102	0.7491±0.0190	0.9095±0.0048	0.7717±0.0168	0.9195±0.0071
GMF	0.8331±0.0121	0.7380±0.0219	0.9059±0.0026	0.7618±0.0196	0.9301±0.0042
NCF(MLP)	0.7837±0.0113	0.6870±0.0137	0.8624±0.0093	0.7123±0.0111	0.9205±0.0047
NeuMF	0.8416±0.0086	0.7560±0.0076	0.9086±0.0051	0.7786±0.0060	0.9300±0.0047
Ours	**0.9008±0.0012**	**0.8046±0.0017**	**0.9548±0.0014**	**0.8208±0.0033**	**0.9570±0.0056**

We randomly change the test data, run all models 10 times, and then report mean ± standard deviation. The proposed recommender significantly outperforms the baselines according to the student t-test (p-value < 0.05).

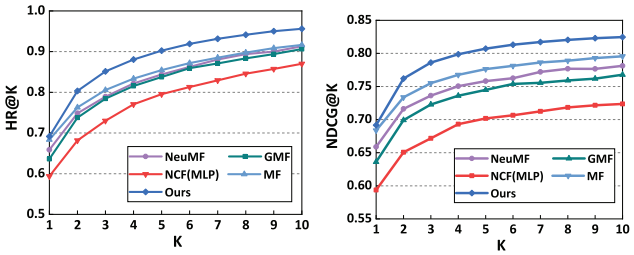

Fig. 5. Performance comparison with respect to Top-K item evaluation where K varies from 1 to 10.

Table 3. Performance of MF and the Proposed Recommender with Different Pre-training Methods

Pre-training Methods		HR@5	NDCG@5	HR@10	NDCG@10
MF	No pre-training	0.7195	0.6071	0.8470	0.6461
	Item2vector	0.8291	0.7382	0.8977	0.7606
	Deep model	**0.8545**	**0.7763**	**0.9165**	**0.7956**
Ours	No pre-training	0.8882	0.7984	0.9497	0.8184
	Item2vector	**0.9111**	**0.8274**	**0.9606**	**0.8436**
	Deep model	0.9025	0.8072	0.9561	0.8247

keeps higher than the baselines regarding HR@k and NDCG@k when k varies from 1 to 10. This verifies that the recommender generates more appropriate recipes for the users by learning the similarity and difference at the category level. And (2) the fusion model NeuMF, combining GMF with NCF (MLP), yields better performance than the single model GMF and NCF (MLP). The observed results are consensus with the work in Reference [23].

4.4 Discussion

4.4.1 Utility of Pre-training. To evaluate the utility of pre-training, we compared the performance of the proposed recommender and the best baseline MF with different pre-training methods. From the results in Table 3, we can summarize the conclusions as follows: (1) The proposed recommender without the pre-training still surpasses the pre-trained MF, illustrating that the category-level correlations are relatively crucial in this task. Besides, the recommender leverages the fixed pre-training embeddings, whereas MF employs the same deep model to extract the features of users and recipes dynamically. Therefore, the experimental results clearly show the superiority of our proposed recommender. (2) The features extracted by the proposed deep model reinforce MF

Table 4. The Results of Ablation Test on WIRCNN and the Recommender

WIRCNN	Macro-F1	Micro-F1	Micro-P	Micro-R
No interaction	0.843	0.977	0.981	0.973
No Bi-RNN	0.649	0.941	0.957	0.927
Ours	**0.881**	**0.981**	**0.983**	**0.978**
Recommender (Deep model)	**HR@5**	**NDCG@5**	**HR@10**	**NDCG@10**
No general memory	0.8946	0.8052	0.9505	0.8235
No category embedding	0.8227	0.7138	0.9080	0.7416
Ours	**0.9025**	**0.8072**	**0.9561**	**0.8247**
Recommender (Item2vector)	**HR@5**	**NDCG@5**	**HR@10**	**NDCG@10**
No general memory	0.8998	0.8231	0.9519	0.8401
No category embedding	0.8767	0.7915	0.9332	0.8100
Ours	**0.9111**	**0.8274**	**0.9606**	**0.8436**

and our recommender with better performance, which reflects that incorporating the rich context of recipes and users is significant. (3) MF is more sensitive to the user and recipe features extracted by the deep model than our proposed recommender. This demonstrates that the incorporation of the category-level information into our recommender partly reduces its dependence on the pre-training. And (4) the Item2vector initialization provides more category-level information about users and recipes for the proposed recommender and makes it achieve the best performance.

4.4.2 Model Ablation. To investigate the effectiveness of multiple components in our proposed WIRCNN and recommender, we conducted the ablation test on two models [41]. Table 4 lists the results of the ablation test. Specifically, for WIRCNN, we removed the interaction mechanism and Bi-RNN, respectively. From Table 4, we can observe that the performance drops significantly when removing the Bi-RNN or interaction mechanism, indicating that Bi-RNN is important for WIRCNN to capture the contextual information, and interaction mechanism significantly promotes WIRCNN regarding the F1 scores. As to the recommender, we evaluated the effectiveness of the general memory and the category embedding by removing them one-by-one. From Table 4, we can find that both the general memory and the category embedding greatly improve the performance of the recommender with two kinds of per-training methods. In addition, the recommender without category embeddings performs similarly to the baselines. Therefore, we can conclude that the similarity and difference at the category level are the foundation of our superior performance.

4.4.3 Parameter Sensitivity. To estimate the proposed models' sensitivity to hyper-parameters, we conducted many contrast experiments to measure the performance of WIRCNN and the recommender under different hyper-parameter settings. Note that only one hyper-parameter was changed at a time. The results of WIRCNN and the recommender are reported in Figure 6 and Figure 7, respectively. Regarding WIRCNN, we evaluated it with different sizes of word embeddings and depths of convolutional layers. From Figure 6, we can have the following observations: (1) at the beginning, the performance of WIRCNN is improving with the increase of the word embedding size, but it remains stable when the embedding size exceeds 50. This indicates that WIRCNN is not very sensitive to the change of word embedding size; (2) the performance of WIRCNN drops slightly when the depth of convolutional layer increases, which proves that increasing the complexity of the neural networks blindly does not work well in this task. As to the health-aware recommender, we testified the effect of the similarity functions and the user/recipe embedding size. From the results shown in Figure 7, we find that: (1) inner product is the best

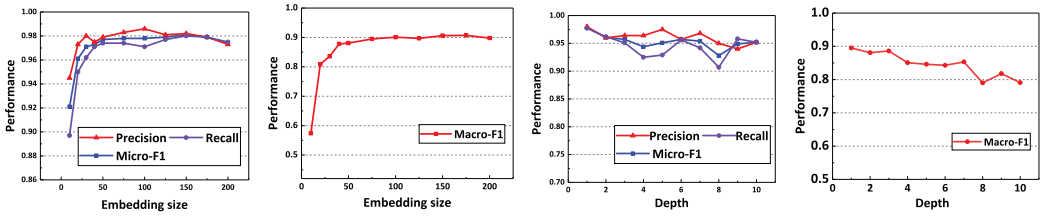

Fig. 6. Performance comparison on WIRCNN regarding word embedding size and convolutional depth, respectively.

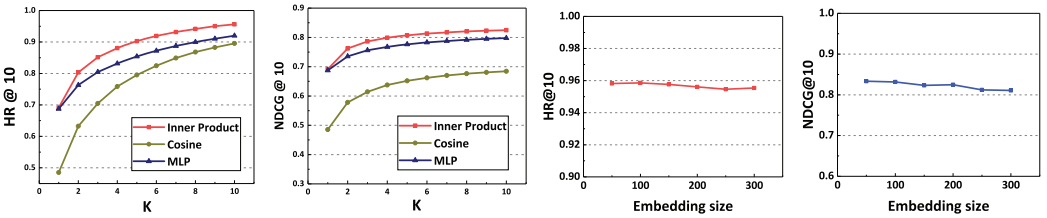

Fig. 7. Evaluation of the proposed recommender with respect to the similarity function and the embedding size of users/recipes.

function to calculate the similarity score in this task and cosine performs the worst; (2) the performance maintains the same level with the increase of embedding size. The proposed recommender is relatively insensitive to the embedding size of users and recipes.

5 CONCLUSION AND FUTURE WORK

In this work, we present a personalized health-aware food recommendation scheme, consisting of three main components, namely, recipe retrieval, user health profiling, and health-aware food recommendation. To justify our proposed deep models, we constructed two high-quality datasets. And experimental results demonstrate the effectiveness of our health-aware food recommendation scheme and the superior performance of the proposed models. Moreover, from the experimental results, we could draw the following conclusions: (1) attentive feature extraction is crucial to the user health profiling based on sparse data. And (2) the category-level information significantly matters in the food recommendation.

This work is only a small step to build the personalized health-aware food recommendation system. In the future, we will continue to perfect the proposed scheme from the following directions: (1) building more mature systems to profile the user health from multiple aspects. (2) Health-aware ingredient recommendation is also a promising research direction to protect the user health. And (3) incorporating more healthy diet knowledge in more effective ways to make recommendation. The embedding and incorporation of knowledge are emerging research topics, whereas the existing work about the food-related knowledge is limited.

REFERENCES

[1] Sofiane Abbar, Yelena Mejova, and Ingmar Weber. 2015. You tweet what you eat: Studying food consumption through Twitter. In *Proceedings of the Conference on Human Factors in Computing Systems*. ACM, 3197–3206.

[2] Yongsheng An, Yu Cao, Jingjing Chen, Chong-Wah Ngo, Jia Jia, Huanbo Luan, and Tat-Seng Chua. 2017. PIC2DISH: A customized cooking assistant system. In *Proceedings of the International Conference on Multimedia*. ACM, 1269–1273.

[3] Oren Barkan and Noam Koenigstein. 2016. Item2Vec: Neural item embedding for collaborative filtering. In *Proceedings of the International Workshop on Machine Learning for Signal Processing*. IEEE, 1–6.

[4] Immanuel Bayer, Xiangnan He, Bhargav Kanagal, and Steffen Rendle. 2017. A generic coordinate descent framework for learning from implicit feedback. In *Proceedings of the International Conference on World Wide Web*. ACM, 1341–1350.

[5] Alexandre Blansché, Julien Cojan, Valmi Dufour-Lussier, Jean Lieber, Pascal Molli, Emmanuel Nauer, Hala Skaf-Molli, and Yannick Toussaint. 2010. Taaable 3: Adaptation of ingredient quantities and of textual preparations. In *Proceedings of the International Conference on Case-based Reasoning Workshop*. 189–198.

[6] Micael Carvalho, Rémi Cadène, David Picard, Laure Soulier, Nicolas Thome, and Matthieu Cord. 2018. Cross-modal retrieval in the cooking context: Learning semantic text-image embeddings. In *Proceedings of the International SIGIR Conference on Research & Development in Information Retrieval*. ACM, 35–44.

[7] Jingjing Chen, Lei Pang, and Chong-Wah Ngo. 2017. Cross-modal recipe retrieval: How to cook this dish? In *Proceedings of the International Conference on Multimedia Modeling*. Springer, 588–600.

[8] Jing-jing Chen, Chong-Wah Ngo, and Tat-Seng Chua. 2017. Cross-modal recipe retrieval with rich food attributes. In *Proceedings of the International Conference on Multimedia*. ACM, 1771–1779.

[9] Jing-Jing Chen, Chong-Wah Ngo, Fu-Li Feng, and Tat-Seng Chua. 2018. Deep understanding of cooking procedure for cross-modal recipe retrieval. In *Proceedings of the International Conference on Multimedia*. ACM, 1020–1028.

[10] Zhiyong Cheng and Jialie Shen. 2014. Just-for-Me: An adaptive personalization system for location-aware social music recommendation. In *Proceedings of the International Conference on Multimedia Retrieval*. ACM, 185–192.

[11] Zhiyong Cheng and Jialie Shen. 2016. On effective location-aware music recommendation. *ACM Trans. Inf. Syst.* 34, 2 (2016).

[12] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* 12, Aug. (2011), 2493–2537.

[13] Du Cunxiao, Chen Zhaozheng, Feng Fuli, Zhu Lei, Gan Tian, and Nie Liqiang. 2019. Explicit interaction model towards text classification. In *Proceedings of the International AAAI Conference on Artificial Intelligence*. AAAI Press, 6359–6366.

[14] Munmun De Choudhury, Sanket Sharma, and Emre Kiciman. 2016. Characterizing dietary choices, nutrition, and language in food deserts via social media. In *Proceedings of the Conference on Computer-supported Cooperative Work & Social Computing*. ACM, 1157–1170.

[15] David Elsweiler, Christoph Trattner, and Morgan Harvey. 2017. Exploiting food choice biases for healthier recipe recommendation. In *Proceedings of the International SIGIR Conference on Research and Development in Information Retrieval*. ACM, 575–584.

[16] Aleksandr Farseev and Tat-Seng Chua. 2017. Tweet can be fit: Integrating data from wearable sensors and multiple social networks for wellness profile learning. *Trans. Inf. Syst.* 35, 4 (2017), 42.

[17] Fuli Feng, Xiangnan He, Yiqun Liu, Liqiang Nie, and Tat-Seng Chua. 2018. Learning on partial-order hypergraphs. In *Proceedings of the World Wide Web Conference*. ACM, 1523–1532.

[18] Jill Freyne and Shlomo Berkovsky. 2010. Intelligent food planning: Personalized recipe recommendation. In *Proceedings of the International Conference on Intelligent User Interfaces*. ACM, 321–324.

[19] Mouzhi Ge, Mehdi Elahi, Ignacio Fernaández-Tobías, Francesco Ricci, and David Massimo. 2015. Using tags and latent factors in a food recommender system. In *Proceedings of the International Conference on Digital Health*. ACM, 105–112.

[20] Mouzhi Ge, Francesco Ricci, and David Massimo. 2015. Health-aware food recommender system. In *Proceedings of the Conference on Recommender Systems*. ACM, 333–334.

[21] Jun Harashima, Yuichiro Someya, and Yohei Kikuta. 2017. Cookpad image dataset: An image collection as infrastructure for food research. In *Proceedings of the International SIGIR Conference on Research and Development in Information Retrieval*. ACM, 1229–1232.

[22] Morgan Harvey, Bernd Ludwig, and David Elsweiler. 2013. You are what you eat: Learning user tastes for rating prediction. In *Proceedings of the International Symposium on String Processing and Information Retrieval*. Springer, 153–164.

[23] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the International Conference on World Wide Web*. ACM, 173–182.

[24] Xiangnan He, Hanwang Zhang, Min-Yen Kan, and Tat-Seng Chua. 2016. Fast matrix factorization for online recommendation with implicit feedback. In *Proceedings of the International SIGIR Conference on Research and Development in Information Retrieval*. ACM, 549–558.

[25] Jonathan L. Herlocker, Joseph A. Konstan, Al Borchers, and John Riedl. 1999. An algorithmic framework for performing collaborative filtering. In *Proceedings of the International SIGIR Conference on Research and Development in Information Retrieval*. ACM, 230–237.

[26] Jose Antonio Iglesias, Plamen Angelov, Agapito Ledezma, and Araceli Sanchis. 2011. Creating evolving user behavior profiles automatically. *Trans. Knowl. Data Eng.* 24, 5 (2011), 854–867.

[27] Hao Jiang, Wenjie Wang, Meng Liu, Liqiang Nie, Ling-Yu Duan, and Changsheng Xu. 2019. Market2Dish: A health-aware food recommendation system. In *Proceedings of the International Conference on Multimedia*. ACM, 2188–2190.

[28] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*. ACL, 427–431.

[29] Hokuto Kagaya, Kiyoharu Aizawa, and Makoto Ogawa. 2014. Food detection and recognition using convolutional neural network. In *Proceedings of the International Conference on Multimedia*. ACM, 1085–1088.

[30] Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, 1746–1751.

[31] Keigo Kitamura, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2008. Food log by analyzing food images. In *Proceedings of the International Conference on Multimedia*. ACM, 999–1000.

[32] Tomasz Kusmierczyk, Christoph Trattner, and Kjetil Nørvåg. 2015. Temporality in online food recipe consumption and production. In *Proceedings of the International Conference on World Wide Web*. ACM, 55–56.

[33] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent convolutional neural networks for text classification. In *Proceedings of the International AAAI Conference on Artificial Intelligence*. AAAI Press, 2267–2273.

[34] Yongqi Li, Meng Liu, Jianhua Yin, Chaoran Cui, Xin-Shun Xu, and Liqiang Nie. 2019. Routing micro-videos via a temporal graph-guided recommendation system. In *Proceedings of the International Conference on Multimedia*. ACM, 1464–1472.

[35] Yelena Mejova, Hamed Haddadi, Anastasios Noulas, and Ingmar Weber. 2015. # foodporn: Obesity patterns in culinary interactions. In *Proceedings of the International Conference on Digital Health*. ACM, 51–58.

[36] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the International Conference on Advances in Neural Information Processing Systems*. The MIT Press, 3111–3119.

[37] Weiqing Min, Bing-Kun Bao, Shuhuan Mei, Yaohui Zhu, Yong Rui, and Shuqiang Jiang. 2017. You are what you eat: Exploring rich recipe information for cross-region food analysis. *Transactions on Multimedia* 20, 4 (2017), 950–964.

[38] Weiqing Min, Shuqiang Jiang, Linhu Liu, Yong Rui, and Ramesh Jain. 2018. A survey on food computing. *arXiv preprint arXiv:1808.07202* (2018).

[39] Weiqing Min, Shuqiang Jiang, Shuhui Wang, Jitao Sang, and Shuhuan Mei. 2017. A delicious recipe analysis framework for exploring multi-modal recipes with various attributes. In *Proceedings of the International Conference on Multimedia*. ACM, 402–410.

[40] Weiqing Min, Linhu Liu, Zhengdong Luo, and Shuqiang Jiang. 2019. Ingredient-guided cascaded multi-attention network for food recognition. In *Proceedings of the International Conference on Multimedia*. 1331–1339.

[41] Liqiang Nie, Wenjie Wang, Richang Hong, Meng Wang, and Qi Tian. 2019. Multimodal dialog system: Generating responses via adaptive decoders. In *Proceedings of the International Conference on Multimedia*. ACM, 1098–1106.

[42] Ferda Ofli, Yusuf Aytar, Ingmar Weber, Raggi Al Hammouri, and Antonio Torralba. 2017. The food perception gap on Instagram and its relation to health. In *Proceedings of the International Conference on World Wide Web*. ACM, 509–518.

[43] Parisa Pouladzadeh and Shervin Shirmohammadi. 2017. Mobile multi-food recognition using deep learning. *Trans. Multimedia Comput. Commun. Applic.* 13, 3s (Aug. 2017), 36:1–36:21.

[44] Ofir Press and Lior Wolf. 2017. Using the output embedding to improve language models. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*. ACL, 157–163.

[45] Markus Rokicki, Christoph Trattner, and Eelco Herder. 2018. The impact of recipe features, social cues and demographics on estimating the healthiness of online recipes. In *Proceedings of the International Conference on Web and Social Media*. AAAI Press, 310–319.

[46] Amaia Salvador, Nicholas Hynes, Yusuf Aytar, Javier Marin, Ferda Ofli, Ingmar Weber, and Antonio Torralba. 2017. Learning cross-modal embeddings for cooking recipes and food images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*. IEEE, 3020–3028.

[47] Satoshi Sanjo and Marie Katsurai. 2017. Recipe popularity prediction with deep visual-semantic fusion. In *Proceedings of the Conference on Information and Knowledge Management*. ACM, 2279–2282.

[48] Yuka Shidochi, Tomokazu Takahashi, Ichiro Ide, and Hiroshi Murase. 2009. Finding replaceable materials in cooking recipe texts considering characteristic cooking actions. In *Proceedings of the Multimedia Workshop on Multimedia for Cooking and Eating Activities*. ACM, 9–14.

[49] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[50] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*. IEEE, 2818–2826.

[51] Christoph Trattner and David Elsweiler. 2017. Food recommender systems: Important contributions, challenges and future research directions. *arXiv preprint arXiv:1711.02760* (2017).

[52] Christoph Trattner and David Elsweiler. 2017. Investigating the healthiness of internet-sourced recipes: Implications for meal planning and recommender systems. In *Proceedings of the International Conference on World Wide Web*. ACM, 489–498.
[53] Liping Wang, Qing Li, Na Li, Guozhu Dong, and Yu Yang. 2008. Substructure similarity measurement in Chinese recipes. In *Proceedings of the International Conference on World Wide Web*. ACM, 979–988.
[54] Shuohang Wang and Jing Jiang. 2016. Learning natural language inference with LSTM. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*. ACL, 1442–1451.
[55] Wenjie Wang, Fuli Feng, Xiangnan He, Liqiang Nie, and Tat-Seng Chua. 2020. Denoising implicit feedback for recommendation. *arXiv preprint arXiv:2006.04153* (2020).
[56] Wenjie Wang, Minlie Huang, Xin-Shun Xu, Fumin Shen, and Liqiang Nie. 2018. Chat more: Deepening and widening the chatting topic via a deep model. In *Proceedings of the International SIGIR Conference on Research & Development in Information Retrieval*. ACM, 255–264.
[57] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural graph collaborative filtering. In *Proceedings of the International SIGIR Conference on Research & Development in Information Retrieval*. ACM, 165–174.
[58] Hui Wu, Michele Merler, Rosario Uceda-Sosa, and John R. Smith. 2016. Learning to make better mistakes: Semantics-aware visual food recognition. In *Proceedings of the International Conference on Multimedia*. ACM, 172–176.
[59] Shulin Yang, Mei Chen, Dean Pomerleau, and Rahul Sukthankar. 2010. Food recognition using statistics of pairwise local features. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*. IEEE, 2249–2256.
[60] Yiming Yang. 1999. An evaluation of statistical approaches to text categorization. *Inf. Retr.* 1, 1–2 (1999), 69–90.
[61] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*. ACL, 1480–1489.
[62] Xun Yi, Elisa Bertino, Fang-Yu Rao, Kwok-Yan Lam, Surya Nepal, and Athman Bouguettaya. 2019. Privacy-preserving user profile matching in social networks. *Trans. Knowl. Data Eng.* 32, 8 (2019), 1572–1585.
[63] Zhou Zhao, Hanqing Lu, Deng Cai, Xiaofei He, and Yueting Zhuang. 2016. User preference learning for online social recommendation. *Trans. Knowl. Data Eng.* 28, 9 (2016), 2522–2534.

Received September 2019; revised June 2020; accepted July 2020