Implementation of a personalized food recommendation system based on collaborative filtering and knapsack method

Nattaporn Thongsri, Pattaraporn Warintarawej, Santi Chotkaew, Wanida Saetang
Faculty of Science and Industrial Technology, Prince of Songkla University, Surat Thani Campus, Surat Thani, Thailand

Article Info

ABSTRACT

Food recommendation system is one of the most interesting recommendation problems since it provides data for decision-making to users on selection of foods that meets individual preference of each user. Personalized recommender system has been used to recommend foods or menus to respond to requirements and restrictions of each user in a better way. This research study aimed to develop a personalized healthy food recommendation system based on collaborative filtering and knapsack method. Assessment results found that users were satisfied with the personalized healthy food recommendation system based on collaborative filtering and knapsack problem algorithm which included ability of operating system, screen design, and efficiency of operating system. The average satisfaction score overall was 4.20 implying that users had an excellent level of satisfaction.

Keywords:
Collaborative filtering
Food recommendation system
Knapsack method
Personalized

1. INTRODUCTION

According to World Health Organization, 70% of deaths worldwide were from non-communicable diseases (NCDs). The diseases are not caused by germs and cannot be transmissible directly from one person to another but they are from habits or behaviors of humans how they spend their lives, i.e. smoking, alcohol drinking, no exercise, consuming unhealthy food, and air pollution exposure. Such factors lead to cancer, heart disease, and obesity. This is consistent with a study conducted by Department of Disease Control [1] founding that non-communicable diseases are the top leading causes of death of the world and harmful public health threat for developing countries. With regard to literature review, it was found that research studies on food recommendation system for consumers were widely conducted in developed countries. However, non-communicable diseases are harmful public health threat for developing countries. Therefore, research studies including the existence of food recommendation system for consumers should not be ignored or overlooked.

Recommender system is able to provide data for decision-making to users on selection of foods that meet individual preference. The most common filter is collaborative filtering that works by using existing human experience for recommendation [2], [3]. Such principle is different from content-based filtering as recommendation depends on specific characteristics of content [4]. Besides, personalized food recommendation is different, depending on how many calories they should consume each day. The findings from the study showed that knapsack method has been used widely for recommending users the suitable quantity of foods [5]-[7].

Currently, it is found that personalized recommender system is applied to the context of guidelines for food selection increasingly [8]. The system is most likely used to recommend various healthy foods and different menus for specific diseases or even food recipes. The food recommendation system will retrieve users’ favorites from different sources such as users’ historical data, users’ scores or selection of food recipes.
Ge et al. [2] developed a food recommendation system while the recommendation was made on the basis of a dataset of users, the rating users give to an item, and tags. Empirical study results revealed that the use of tags for menu recommendation was accurate for predicting or recommending menus to users at a high level or it can be said that the menus recommended by the system were actually consistent and met users’ requirements. The collaborative filtering concept is to analyze users’ behaviors from selecting or scoring and to search for similar users’ favorites so as to predict which menus users will prefer [9], [10]. After that, similarity among users is measured [11].

However, previous studies were concerned with the food recommendation system for general people, as of today each consumer requires different menus in accordance with their restrictions such as weight and underlying diseases. As a consequence, to achieve the highest efficiency in recommendation according to requirements or restrictions of each user, other techniques must be taken into consideration. In this regard, it is consistent with Gao [12] who conducted a study on recommendation system and found that knapsack technique was suitable for product recommendation under organizational restrictions so as to be a strategy for generating the highest revenue. The knapsack technique is able to respond to organizational requirements and in the meantime it can meet users’ satisfaction. Therefore, this research study applied the collaborative filtering technique for calculating new users’ favorites by using data of the old users as a base. Meanwhile knapsack method was mutually developed for food recommendation in accordance with restrictions of each user. Details of each technique are as follow:

1. Collaborative filtering

Collaborative filtering is calculating similarity among users by using behaviors of persons having similarities with users. Users’ data are determined as a database in working on prediction and recommendation of lists to users accordingly in using the food recommendation system on the basis of collaborative filtering, which can be divided into 2 steps [13], [14] as follows: i) Calculating to figure out similarity of users: This method shall calculate Pearson’s correlation coefficient between the rating given by new users and old users to check similarity based on the correlation (1) [15], [16]:

\[ s(u, v) = \frac{\sum_{i} (I_u \cap I_v)(r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i} (I_u \cap I_v)(r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{i} (I_u \cap I_v)(r_{vi} - \bar{r}_v)^2}} \]  

where: 
- \( s(u, v) \) refers to similarity between new users and old users. 
- \( I_u \cap I_v \) refers to similarity of the rating given by new users and old users. 
- \( r_{ui} \) refers to the rating of new users to the lists. 
- \( \bar{r}_u \) refers to the average rating given by new users. 
- \( r_{vi} \) refers to the rating of old users to the lists. 
- \( \bar{r}_v \) refers to the average rating given by old users.

Calculated values were used to determine statistical correlation between scores of new and old users to figure out prediction value; and ii) Calculating to figure out prediction value: As for prediction or recommendation for new users based on the collaborative filtering between new users and old users, calculation is made to find the nearest value of new users given to the lists that match those of old users. Once the calculation is made, the system is rating user scores for predicting new users [17], [18]. Generally, calculation is based on the average rating of users having several similarities according to (2).

\[ p_{u,i} = \bar{r}_u + \frac{\sum_{u \in N}(s(u,w)(r_{u,i} - \bar{r}_u))}{\sum_{u \in N}[s(u,w)]} \]  

where: 
- \( p_{u,i} \) refers to prediction for new users to the existing lists. 
- \( \bar{r}_u \) refers to the average score of users. 
- \( r_{u,i} \) refers to scores of old users to the lists. 
- \( \bar{r}_v \) refers to the average score of old users. 
- \( s \ (u , u’) \) refers to results from calculations to find similarity between users.

The lists are arranged from what users like the most to what users like the least as to recommend a food menu to new users. In case a lot of menus are available and recommended foods do not match users’ requirements, the next recommended menu will be shown until users’ requirements are met.

2. Knapsack algorithm

Knapsack problem means selecting n items with different weights and values in a knapsack based on the condition that when all values of the selected items are calculated must have the highest total value and the total weight does not exceed the maximum weight that the knapsack can bear [19], [20]. For example, if a thief would like to steal items from a supermarket by selecting items in the shelves in a backpack but the backpack has limited capacity to carry weights. Thus, which items the thief will select in the backpack to gain the highest

Implementation of a personalized food recommendation system ... (Nattaporn Thongsri)
total value of the items they wish to steal as shown in Figure 1 to be called knapsack problem while the total weight must not exceed the knapsack’s capacity; 0-1 knapsack problem when 0-1 mean either take an item completely or leave it completely, substitute taking for 1 and substitute leaving for 0; the fraction of any item cannot be taken. Such problem is called optimization problem. It is selecting the best solution under the available condition. 0-1 Knapsack problem cannot be solved by greedy approach. For example, 3 pieces of items to be selected in a backpack have weights and values as shown in the Table 1, while the backpack’s capacity does not exceed 6 kg [21].

![Figure 1. Knapsack problem](image)

Table 1. Example of product value per kilogram

<table>
<thead>
<tr>
<th>Item</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>10</td>
<td>28</td>
<td>12</td>
</tr>
<tr>
<td>Weight</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Ratio = (Price/Weight)</td>
<td>10</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>

If greedy approach is used, the first item to be selected is A since it has the highest value per kilogram, followed by B, making the total value of A and B is 10+28 = 38 and the total weight of A and B is 5 kilogram which does not exceed the backpack’s capacity. However, it is not the best solution. The items that give the highest value in the backpack are B and C as the total value of B and C is 28+12=40. It can be concluded that the greedy approach is unable to give the best value in this example. Therefore, the method used to solve the problem in this study is dynamic programming approach [22]-[24].

Dynamic programming approach is problem-solving to get the best solution. Components of considered items are divided into sub-components and results from the previous calculation are kept in the form of tables for being considered in the next time. It is called a recursive process or calling a function from itself until the final solution is obtained. Solving knapsack problem using the dynamic programming approach determines there are n items where each item I (i=1…n) has a weight wi and an integer value vi. The maximum weight a backpack can carry is W, the problem can be written in the form of optimization problem [25] as follow:

\[
\text{Maximize } \sum_{i=1}^{n} x_i v_i \\
\text{Subject to } \sum_{i=1}^{n} x_i w_i \leq W \\
\text{When } 0 \leq x_i \leq 1 \text{ refers to each item I, the value of } x_i
\]

where n is the number of all items,

\[w = \{w_1, w_2, ..., w_n\}\] refers to a weight of each item i.

\[v = \{v_1, v_2, ..., v_n\}\] refers to the value of each item i.

Thus, the steps to solve problem how to select items to gain the optimum value and the total weight does not exceed the backpack’s capacity are shown as follow:

KnapSack(v, w, nW) when n=3 and W=60, i=1..6 w={1,4,2} v={10,28,12}. The calculation of V[i, w] as shown in the Table 2.
Table 2. Calculation of $V[i, w]$

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i=1, $w_1=1, v_1=10$</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>i=2, $w_2=4, v_2=28$</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>max(28+0, 10)</td>
<td>max(28+10, 10)</td>
<td>max(28+10, 10)</td>
</tr>
<tr>
<td>i=3, $w_3=2, v_3=12$</td>
<td>max(12+0, 10)</td>
<td>max(12+10)</td>
<td>max(12+10, 28)</td>
<td>max(12+10, 38)</td>
<td>max(12+28, 38)</td>
<td></td>
</tr>
</tbody>
</table>

It can be explained by considering the first item $i=1$ having a weight of 1 that can be selected in a backpack having a weight from 1 to 6. The value shown in the table is the total value of items put in the backpack at the weight of $w_i$, $i=2$ at $w_2=4$ and $v_2=28$. The first backpack can carry $w=1$ and it is found that it can carry the 2nd item. Thus, the value $V[2, 1]=V[1, 1]$ is the value of the item selected previously. The backpack $w=2$ and 3 – the weights of the backpacks are less than the items put in the backpacks, the items are not put in the backpacks but the previous item is maintained. The backpack that can carry $w=4$ is found that it can carry the 2nd item. Therefore, the value $V[1, 4]=max\{V[1, 4], v[2]+V[1, 0]\}=max\{28+0, 10\}=28$. The whole table is calculated in the same way in all boxes. The optimum value in the last box is 40 which is obtained from the 3rd item having the value of 12 in combination with the 2nd item having the value of 28.

2. RESEARCH METHOD

The development of a personalized food recommendation system based on collaborative filtering and knapsack problem algorithm in terms of system architecture and design is shown in Figure 2. As illustrated in Figure 2, users fill their personal data comprising sex, weight, height, favorite food rating. Next, the system shall bring personal data variable of users in conjunction with knapsack method to recommend them menus that match their preferences and restrictions such as underlying diseases of each user under the number of calories they need each day. The food recommendation system based on collaborative filtering is divided into 2 steps as follows:

2.1. Calculating to find similarity of users by searching for similarity between old users and new users using

By searching for similarity between old users and new users using, the below formula is used.
\[ s(u, v) = \frac{\sum_{i \in I} u_i - u_i \cdot r_i - r_i \cdot (r_u - r_i)}{\sqrt{\sum_{i \in I} u_i \cdot (r_u - r_i)^2} \sqrt{\sum_{i \in I} v_i \cdot (r_v - r_i)^2}} \]

### 2.2. Calculating to figure out prediction value

The calculation to determine the similarity of users as shown in the Figure 3.

\[ p_{u,i} = \bar{r}_u + \frac{\sum_{u \in \mathbb{N}} S(u, u')(r_{u,i} - \bar{r}_u)}{\sum_{u \in \mathbb{N}} |S(u, u')|} \]

![Figure 3. The calculation to determine the similarity of users](image)

When menu preferences using the collaborative filtering method are calculated, the menus with the highest preference score are selected but they must not exceed the number of calories each user needs each day. This study applied the knapsack algorithm to select the recommended menus by using users’ preference scores given to the menus. \( p(u, i) \) refers to the value of items. The number of calories of each menu represents the weights of items. The number of calories each user needs each day, restrictions, represents the capacity of a backpack. The number of calories each user needs per day is calculated from basal metabolic rate (BMR) based on sex, age, weight, and height.

For male: \( \text{BMR} = 66.5 + (13.75 \times \text{weight in kg}) + (5.003 \times \text{height in cm}) - (6.755 \times \text{age in years}) \)

For female: \( \text{BMR} = 655 + (9.563 \times \text{weight in kg}) + (1.850 \times \text{height in cm}) - (4.676 \times \text{age in years}) \)

The application of knapsack using dynamic programming approach for selecting recommended menus determines there are \( n \) items where each item \( i \) (\( i = 1 \ldots n \)) contains \( c_i \) calories and users’ preference of menus is \( p_i \). The number of calories they need each day = BMR which can be represented by the following symbols:

- \( c \) = \{\( c_1, c_2, \ldots, c_n \)\} represents the number of calories of food menu \( i \)
- \( p \) = \{\( p_1, p_2, \ldots, p_n \)\} represents preference of food menu \( i \)

Selecting which food menus to gain the highest preference value that does not exceed the number of calories they need each day.

Maximize \( \sum_{i=1}^{n} x_i p_i \)

Subject to \( \sum_{i=1}^{n} x_i c_i \leq \text{BMR} \)

When \( 0 \leq x_i \leq 1 \) represents food menu \( i \), the value of \( x_i \). Where \( x_i \) has 2 values; taking an item completely = 1 and leaving an item completely = 0.
KnapSack algorithm for food recommendations update from [26] as follows:

KnapSack(c,p,n,BMR)
for c = 0 to C do
P(0, c) = 0
for i = 1 to n do
P[i, 0] = 0
for c = 0 to BMR do
if c[i] ≤ w and (p[i]+ P[i-1, c-c[i]] > P[i-1, c])
then
P[i, c] = p[i]+ P[i-1, c-c[i]]
keep[i, c] = 1
else
P[i, c] = V[i-1, c]
keep[i, c] = 0
K = BMR
for i = n downto 1
if (keep[i,K] == 1)
output i
K = K- c[i]
return V[n,BMR]

3. RESULTS AND DISCUSSION
In this of collaborative filtering and Knapsack method. The results indicated that some problems in the development of the system were discovered which could be considered as the reference guidelines for further development. First of research, data mining theory was applied to create a personalized food recommendation with the integrated technique all, since collaborative filtering technique requires initial data from previous user preferences as information for potential user recommendation. During initial phase of the system, a cold-start problem was found as the acquisition of user preference information had still very little interactions [7], [27]. Therefore, a certain amount of time is required for data acquisition in order to obtain sufficient data for processing. In this paper, the initial data acquisition problem for use in model learning was resolved by questionnaires. User preferences for food items were randomly collected, which were used as the default information for the system. Later, the system will collect food preferences from users who login to use the system through the web page indefinitely. Another problem found was if there were unavailability of initial data of a user's favorite food list for using to train in the model, the system cannot recommend food items that are close to the user's preferences. Simply put, a solution to this problem requires enough information. Finally, the components of the developed system as a tool for users in making decisions about their menu choices comprise filling out a user profile, user rating towards the menu in the system, rating similarities between new and old users to obtain predictor for the recommended menu suitable for each user according to nutrition information. More importantly, the results of this empirical study implied that for further study, ingredients contained in the system should be used as part of the user choice, especially food allergic users when item recommendation. In addition, any ingredients users are allergic to should deleted from the recommender system. After the method development was finished, the researcher developed the system. The Figure 4 (a)-(e) shows the example of the screen of the developed system (see in appendix).

As soon as the system was completely finished, the researcher brought the system to users comprising 90 public health work operators to test and rate satisfaction scores towards the system. The highest score = 5 points meaning extremely satisfied, 1 point = absolutely dissatisfied. The mean score of assessment results of satisfaction towards the operating system, screen design, efficiency of the operating system was 4.20, considered users were satisfied with the system at a high level. The study results and body of knowledge in different aspects used to develop the food recommendation based on collaborative filtering and knapsack problem algorithm enable the food recommendation system to be complete and actually functional. The system can be used to be a part of eliminating the problems related to non-communicable diseases, a harmful threat in developing countries. Besides, the system can give a guideline to relevant persons such as government agencies, Ministry of Public Health including software developers to study and further develop the system to meet current people’s requirements increasingly.

4. CONCLUSION
The study on personalized food recommendation system based on collaborative filtering and knapsack method provides 3 contributions: The first one is the study generated the food recommendation system suitable for individuals to help consumers make their decision on selection of foods that meet their restrictions and needs of nutrients different in each user by presenting data of users such as food favorites and personal data of users as a helper for recommendation that meet nutrition principles more and more. The second one is this study integrated two techniques; collaborative filtering and knapsack problem algorithm, to develop the system.
The last one is this study generates value to the food recommendation system in developing countries like Thailand. The guideline obtained from the study can be disseminated to be a practice guideline for other countries, especially the developing ones to use it widely.

APPENDIX

Figure 4. Personalized food recommendation system screen shots: (a) The main page for login to system, (b) The page for user registration, (c) The page for administrator to manage menu, (d) The page for user to rate the menu, (d) The page for user recommendation food item, (e) the article pages provide knowledge about health

REFERENCES


BIOGRAPHIES OF AUTHORS

Nattaporn Thongsri is currently a lecturer in the Faculty of Science and Industrial Technology, Prince of Songkla University, Surat Thani Campus, Thailand. She received Ph.D. degree major in Management Science and Engineering from the School of Management, Huazhong University of Science and Technology, Wuhan, China. Her teaching and research interests include Management Information System, Technology Acceptance, Educational Informatics and Electronic Commerce. Her research has been published in journals such as Behaviour and Information Technology, Innovations in Education and Teaching International, Journal of Systems and Information Technology and The International Journal of Information and Learning Technology. She can be contacted at email: nattaporn.th@psu.ac.th.

Pattaraporn Wairintaraweew, is currently a lecturer in the Faculty of Science and Industrial Technology, Prince of Songkla University, Surat Thani Campus, Thailand. She received Ph.D. degree major in Computer Science from the Université Montpellier-II, France. Her research interest include data mining, mobile programming and web programming. She can be contacted at email: pattaraporn.w@psu.ac.th.

Implementation of a personalized food recommendation system ... (Nattaporn Thongsri)
Santi Chotkaew is currently a lecturer in the Faculty of Science and Industrial Technology, Prince of Songkla University, Surat Thani Campus, Thailand. He received master degree major in Computer Science, Kasetsart university, Thailand. His research interest include data mining, decision support system. He can be contacted at email: santi.c@psu.ac.th.

Wanida Saetang is currently a lecturer in the Faculty of Science and Industrial Technology, Prince of Songkla University, Surat Thani Campus, Thailand. She received Ph.D. degree major in Information Technology from King Mongkut’s University of Technology North Bangkok, Thailand. Her research interests are Information Management, Information Theory, Innovations Diffusion, and Technology Acceptance Model. She can be contacted at email: wanida.sae@psu.ac.th.