

The Analysis and Recommendation in Dining of College Students

Kewen Tang, Zhenyu Liu and Hongwei Jia

School of Software Engineering, Beijing University of Posts and Telecommunications, Beijing, 100876, China

Abstract—Contemporarily, the diet in universities is so well-focused that we build up the system-“Clean-plate Movement”-to make the diet recommendation for the college alumnus. Born out of the conviction of nutriology and resembled user throng, our recommending system acquires the data and logs and then divides the scope into content-based, associated filter and nutrient recommendation, in which results in the fittest and salutary diet.

Keywords- Content-based, associated filter, resembled degree, recommendation, nutrition.

I. INTRODUCTION

College students get used to eat in the dining hall, where the university almost provides the constant recipe for the students. Therefore, our system, “Clean-plate Movement”, counts when dealing with the issue about the eating habit, personal preference and nutrient arrangement. Considering the factor in the nutriology, real-time eating record and resembled users, we assuage the predicament that college student usually waste food and finally help students always have their comparatively optimal choice on eating.

II. METHOD

A. Content-based Recommendation

(1) *Resembled Degree of Dishes*: As for miscellaneous kinds of food, we classify them with taste and materials so that we bring in $\text{sim}(A1, A2)$ to represent the resembled degree of dishes using the following:

$$\text{Sim}(A1, A2) = p_1 \times \frac{nt}{4} + p_2 \times \frac{nm}{x}$$

In the formula, nt symbolizes the number of resemblance of dishes tastes calculated by four Boolean flavors- if it has the attributes, it equals 1; otherwise, it equals 0-defined in the dishes named Sour, Sweet, Hemp, Spicy Taste. nm indicates the number of resemblance of dishes and x indicates the total material number of $A1$. Besides, p_1 and p_2 embody the influence of the material and taste for the user, able to be obtained in the data mining afterwards.

(2) *Recommendation based on Dishes*: Define $C = \{C_1, C_2, \dots, C_n\}$ as user sets, $S = \{S_1, S_2, \dots, S_n\}$ as dishes sets. Assuming that $P(C_1, S_1)$ is fictionally how much user C_1 like the food S_1 and ranged from 0 to 1 and $\text{sim}(S_1, S_2)$ is how much food s_2 is resembled to food s_1 , consequently we get the $P(C_1, C_2)$. Similarly, all the value

of how much user like the food are accessible with the following:

$$P(C_i, S_n) = P(C_i, S_m) \times \text{sim}(S_n, S_m)$$

(3) *Recommendation based on Rating*: Having taken the rating given by users into account, the system will synthesize an integrated value to offer the right recommendation, which is easy to judge the total popularity.

B. Associated Filtering Recommendation

1) *Resembled Degree of Users*: The classification, which is hard to integrate by several countable factors, is defined by the result based on the clustering. Define $S = \{S_1, S_2, \dots, S_n\}$ as the user sets, $C_1 = \{C_1, C_2, \dots, C_n\}$ as the set of daily eating record of user S_n and $C = \{C_1, C_2, \dots, C_m\}$ as the sets of the most frequent food consuming by this user. Equivalently define $C_2 = \{C_1, C_2, \dots, C_i\}$ and $C = \{C_1, C_2, \dots, C_j\}$ as the one of user S_m . So define user S_n and user S_m as resembled users, whose value of resemblance is calculated by $\text{clu}(C_1, C_u)$ the degree of clustering of the users for the food set C which represents the distance between the elements of set C_1 and the one in set C with the method:

$$d = \sqrt{(E_{ci} - E_{cj})^2 + (Z_{ci} - Z_{cj})^2}$$

In the formula, E_{ci} represent the eating number of food C_i and Z_{ci} represents the rating. Based on the degree of clustering, then we gain the value of user similarity:

$$\text{Sim}(S_n, S_m) = 1 - \frac{|\text{clu}(C_1, C) - \text{clu}(C_2, C)|}{\text{clu}(C_1, C)}$$

2) Recommendation based on Resembled Users

Firstly, define the followings:

- a) $C = \{C_1, C_2, \dots, C_n\}$: the set of users
- b) $S = \{S_1, S_2, \dots, S_n\}$: the set of products
- c) $r_{c,s}$: The rating of product S given by the user C – remains unknown and use algorithm to detect. Generally speaking, this rating is calculated by the points that the other users comment on.

Furthermore, use C' as the user set which is highly similar to C , there are some functions for detection:

$$r_{c,s} = \frac{1}{N} \sum_{c' \in C'} r_{c',s} \quad (1)$$

$$r_{c,s} = k \sum_{c' \in C'} \text{sim}(c, c') r_{c',s} \quad (2)$$

$$r_{c,s} = k \sum_{c' \in C'} \text{sim}(c, c') (r_{c',s} - \text{ave}(r_c)) + \text{ave}(r_c) \quad (3)$$

Among the formula, we note $\text{Sim}(C, C')$ to show the degree of similarity and K to be a standard factor.

Finally, the result is:

In the formula (1), we just simply average the rating to get result. Clearly convenient but lack of accuracy.

In the formula (2), we refine the result with multiplying the rating by the degree of similarity.

In the formula (3), having brought in $r_{c,s}\text{-ave}(r_c)$ as rectification handling with the characteristics of users, we optimize the result that the recommend points can really reflect users' own eating habits.

C. Recommendation based on Nutriology

When it comes to accommodation recommendation, people always take the nutrition and the need into consideration. So, how to get the optimal result?

1) *By Previous Food Choices*: Monotonous eating style causes unhealthy status, particularly user's emotional repulsion. The system classifies the food selection and analyzes the interconnection among these. Assuming $V = \sum_{i=1}^n V_i$ is regarded as the vitamin while $C = \sum_{i=1}^n C_i$ is referred as the calorie during the eating cycle in the latest n days – The two variables had been set after the food material was put into archives. Depending on how much exercise cost takes into account, the system will automatically judge if our body has reached a balance where the input and output of the calories remain to be almost the same. If not, the system will recommend some low-calorie food to try to keep a good diet. As the experts says, it is best for every people that he or she assimilates the Vitamin C ranging from 200mg to 300 mg and at least more than 60ug, which the system rely on to decide if the recommendation is amplified with the Vitamin C.

2) *By Physiological Factors in One Duration*: Unlike the previous scopes, we concentrate on the height, weight and other physiological factors of user's body. For example, if the weight rises up severely, the system will mention the user to exercise more frequently and give the recipe with low-calorie food. Also, if the user prefers the spicy food in a long term, susceptible to break up the balance in the body and cause many bad chain reactions, the food with less spice, pepper and other condiments will fit more for these users.

III. IMPLEMENTATION AND VERIFICATION

A. Recommendation based on Dishes

According to the formula:

$$P(C_i, S_n) = P(C_i, S_m) \times \text{sim}(S_m, S_n)$$

We do the simulation with the test case: assuming that user rate 0.8 on the food named XiaoChaoTiGuRou, the other rating for other dishes is ready.

The table below shows the similarities between various dishes and dish A (the reference dish) hot pork. The 4th column shows the tastes similarity. The 5th column shows the same number of ingredients, you can see the ingredients of XiaoChaotigurou is the exact same as A. The 6th column

showed that, the highest similarity with the reference dish is **B** whose similarity reached 86.85% (about 10% higher than the similarity of the second).

TABLE I. DISH SIMILARITY

dish	Taste	ingredients	Dishes similarity	Score
B	0010	pork, pepper	86.85%	0.7
C	0010	beef, pepper	63.15%	0.5
D	0011	tofu	52.60%	0.4
E	0000	eggplant, pepper	50.00%	0.4
F	0110	chicken, peanut	26.30%	0.2

B. Recommendation based on Resembled Users

Equivalently, the formula is:

$$\text{Sim}(S_n, S_m) = 1 - \frac{|\text{clu}(C_{1,C}) - \text{clu}(C_{2,C})|}{\text{clu}(C_{1,C})}$$

The result of simulation and test is listed in the following table:

TABLE II. USER SIMILARITY

User	All dish	C	Cluster	similarity
A	1, 2, 3, 4, 5, 6,	2,5	68	-
B	2, 5, 7, 9, 10, 11, 13	2,5	76	0.88
C	2, 5, 3, 8, 12	2,5	53	0.78
D	15, ,16	2,5	0	0

Apparently, this table shows the difference between A and the other three users named B, C, D and get the resembled degree.

C. Recommendation based on Nutriology

Differed from others, we analyze the case with data-fitting method. Thus we simulate one set of data, receiving the report that the overeating for the meat. So the system will remind this user to eat more vegetables at the specific time to mediate the PH value in the body.

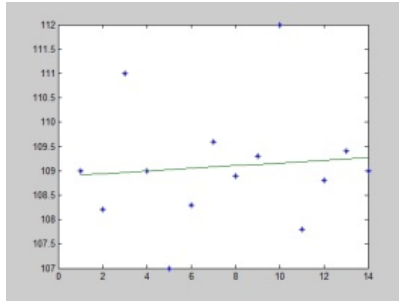


Figure 1. Data-fitting

IV. CONCLUSION

The system enables the users to eat exactly the amount of food they need. In this way, they can avoid the waste of food and keep a balance in nutrition. Through real-time monitoring of the user's weight, interval between meals and special situations such as women's physical periods, the system can automatically adjust its recommendation on food amount, thus ensuring the calories and nutrients that the user takes in remains within a reasonable range. With the assist of our system, it will be easier for the users to develop a more healthy diet.

Obviously, the recommendation model of this system is not yet perfect. In our opinion, if the users' preferred dishes have changed over time, the recommendation system should be able to sense the changes by analyzing the users' behaviors and changes its recommendations accordingly. This mechanism will avoid the problem of the system always recommending the same type of dishes. This function should be developed in the next stage. Besides, we haven't figured out how much and in what way each factor (tastes, ingredients) influences the users' choice of a specific dish. Our next step will be perfecting the existing model by independently analyzing the importance of each factor during the process of a user making his/her choice, so that a more accurate recommendation can be made. The third

problem is that some environmental factors such as the weather and seasons haven't been added the model yet since it's difficult to know how these factors influence users' choice exactly. In the next step we will consider using the technique of correlation analysis to test the statistical significance of the association between these environmental factors and the users' choices. In this way we will be able to distinguish which factors really influence the users' choice and which do not.

When it comes to the recommendation for the dishes, the algorithms or strategies are enormous but lack of specific target. Comparatively, our solution narrows down the scope into the college and the pertinence is very cogent. However, there are still some weaknesses in our system. The recommendation for resembled users is the most evident case that this function acquires tremendous data sizes so that deficiency in data cannot obtain the optimal results. In conclusion, this plan does help students to pick up their own preference on dishes and to some extent mitigate the quandary in food waste in the campus.

REFERENCES

- [1] Deng Ailin, Zhu Yangyong, Shi Bole, A Collaborative Filtering Recommendation Algorithm Based on Item Rating Prediction[J], JOURNAL OF SOFTWARE, 2003, 14 (9) : 162121628.
- [2] Gao Jing, Ying JiKang, A Recommendation System Based on Artificial Immune System[J], COMPUTER TECHNOLOGY AND DEVELOPMENT, 2007, 17 (5) : 1802183.
- [3] Li GuangYuan, Li WenJing, A clustering analysis method based on the similarities of the attributes(in Chinese) [J], COMPUTER AND INFORMATION TECHNOLOGY, 2002,4 (5) :530001. [4]Wang WeiPing, Liu Ying, Recommendation Algorithm Based on Customer Behavior Locus[J], COMPUTER SYSTEMS & APPLICATIONS, 2006, 15 (9) : 35238.
- [4] Wang ZhiMei, Yang Fan, P2P Recommendation Algorithm Based on Hebbian Consistency Learning[J], COMPUTER ENGINEERING AND APPLICATIONS, 2006, 42 (36) : 1102113.
- [5] Xu YiFeng, Xu QingYun, Liu XiaoPing, A Recommender Arithmetic Based on The Order of Time[J], COMPUTER SYSTEMS & APPLICATIONS, 2006, 15 (10) : 22225, 29.