International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 11 | November - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Meal Map Pro

Komal[#1] *21BCS4576*
BE CSE
21BCS4576@cuchd.in

Lokendra Kumar[#2]
21BCS4212BE CSE
21BCS4212@cuchd.in

Yuvraj Chaudhary [#3]
21BCS4493BE CSE
21BCS4493@cuchd.in

Ritik Saini [#4]
21BCS7889 BE CSE
21BCS7889@cuchd.in

Harshit[#5]
21BCS4575
BE CSE
21BCS4575@cuchd.in

Puneet Kaur[#6]
E6913
Supervisor
Computer science and engineering
Chandigarh University,

Mohali,Punjab, India

Abstract-- In this paper, we describe the recipe recommendation system in the culinary domain. Due to the widespread use of the internet, the whole world is connected, and different users from different countries share millions of recipes online, all over the world. As a result, users are unaware of all the recipes available on the internet. A recipe contains heterogeneous information's such as ingredients, cooking process, categories, etc. Therefore, we believe that a recipe is an aggregation of these heterogeneous features. The majority of the recipe recommendation systems are based on content or collaborate filtering to predict the new recipe that a user is interested in. Combining with both content and collaborative filtering, we propose an effective and elegant framework for combining recipe recommendation systems. Most recipe recommendation systems use content information for ingredients or cooking process of recipes. To reduce RMSE, we proposed a hybrid approach combining conventional techniques with content and collaborative filtering. This approach adds more heterogeneous information such as cuisine, preparation directions, dietary, etc.

Keywords: Recommendation system, collaborative filtering, hybrid approaches, recipes, content information.

Introduction

In this day and age of the internet, a huge amount of information is deposited on the internet every day. With a huge number of choices available on the internet, information filtering is necessary to get useful information from this raw data. In the culinary field, there are currently 10,000 websites available on the internet. These websites provide different types of information for information filtering purpose such as photos, text, videos etc. However, with the large amount of information deposited on the internet by different users from different countries, information becomes overwhelming. Finding useful recipe which the user may like will be a time-consuming process. Here recipe recommendation system provides the desirable solution to this problem.

Many online stores have recommendation systems, such as Amazon, CD NOW, and Netflix. Most of them use content-based delivery methods and collaborative filtering techniques. Because these websites are so popular on the internet, the products on the website will have higher ratings than less popular recipe websites. Therefore, the role of the recipe recommendation system in the culinary field is to provide some of the challenges. First, there is the issue of sparsity. This means that Recipe Website is not as popular as other websites, so the rating for a particular recipe will be much lower than Recipe Website for popular websites. Second, recipes consist of various features such as ingredients, cooking instructions, cooking methods, and category information. This means that 4,444 large amounts of heterogeneous data are available here. This disparate information is required to understand the recipe. All of this disparate information is needed to understand the

complex opinions about recipe from a user's perspective. Most of the traditional recommendation systems rely heavily on collaborative filtering to find potential connections between different users of the same website using user rating matrices.

However, there are several issues such as cold start, sparsity, popularity bias, and first rater programs. To overcome these issues, we propose a novel hybrid approach that combines both content and collaborative filtering information for personalized usage. Recipe recommendation system. To help users navigate the Internet based on their previous preferences, we propose two her hybrid approaches that use recipe content and a user's rating matrix for recipes. The first hybrid approach is based on KNN collaboration filtering technique with recipe content information. The second hybrid approach is based on a stochastic gradient descent approach that uses as recipe user rating information and recipe content information.

In this paper, section II. Describes the existing system, section III. Describes the flow of the system and statistics of dataset, section IV. Describes the proposed hybrid approaches, section V. describes the evaluation result and section VI. Describes the conclusion and future work

EXISTING SYSTEM

At its core, recommendation systems are the result of extensive research in cognitive science [4], approximation theory [9], information retrieval [5], and prediction theory [14], and recommendation systems have emerged as an independent field.
did.
Mid 1990s.
Many His applications using recommendation system to help users find more suitable His products and items from the user's point of view and increase His production and profits from the business point of view exists in the real world.
In the field of cooking, much research has been conducted on recipe recommendation systems.
Yoko et al.
[6] proposed a recipe recommendation system based on the user's schedule, weight, etc.
Farhana et al.
[16] proposed a personalized cancer diet based on the different nutritional needs of cancer patients and provided a table of tips.
Peter Forbes et al.
[7] proposes a recipe recommendation system based on different ingredients and finds his similar ingredients that produce different constitutions.
Tsubasa et al.
[10] proposed a recipe recommendation system that is based on natural language processes (NLP) and uses nutritional information in recipes ().
Dal Inderjeet Kaur et al.
[8] propose a network of ingredients from different cultures.
Correlate various components from different cultures.
Liping Wang et al.
[19] propose the substructure similarity of different cooking directions of recipes and find similar types of recipes.
Jill Frein et al.
[13] propose a recipe recommendation system based on different ingredients of the recipe.
Takuma Maruyama et al.
[17] propose a recipe recommendation system based on object reorganization of recipe materials, and propose different recipes based on the reorganization of materials.
Ahn Young Yeol et al.
[11] proposes a taste network for meal accompaniment and generates alternative ingredients.
Mayumi Ueda et al.
[18] propose a recipe recommendation system based on ingredients and the amount of ingredients in the recipe.
In previous studies, only the composition and nutritional value of ingredients were considered in recipes.
In our proposed work, we added a lot of heterogeneous information about ingredients in the model, such as preparation steps, occasions, cooking, and nutrition.
The conventional model units are used as a benchmark model and compared with the proposed model units.

FLOW OF THE SYSTEM AND STATISTICS OF DATASET

This section describes the general recipe recommendation system flow and extracts information from the dataset.

A. System Flow The general flow of the system is shown in Figure 1. The Web crawler program is required to extract user review information and recipe information from the Web. The information here comes from the website food.com. After the crawling process, a preprocessing step is performed and the useful data is stored in the recipe content database. After crawling, you will get sparse scoring matrices for each user, consisting of users and recipes. We then split the sparse user/recipe matrix into train and test data sets. Next, train recommendations using the training dataset. When you apply the test data set to the Train model, the model evaluates and makes predictions. The dataset was extracted from the website food.com from February 25, 2000 to September 3, 2012. From this data, various recipes and

their content information, as well as user ratings ranging from 1 to 5 for the recipes are obtained. Fig (1)

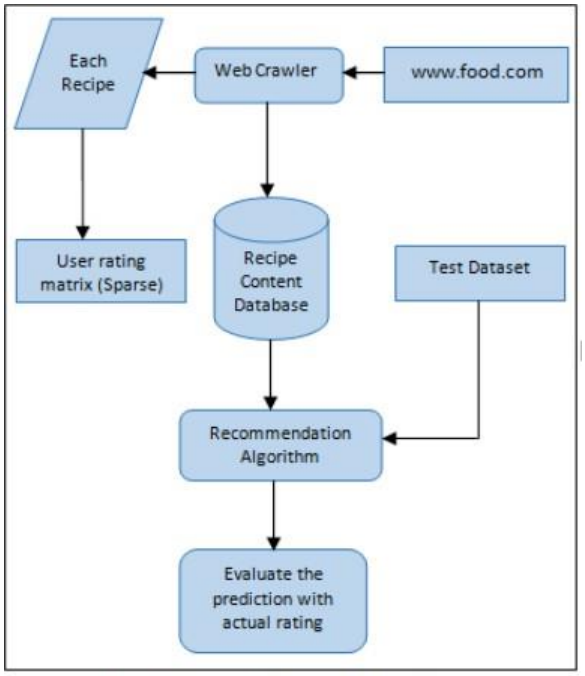


Fig. 1 General flow of the system

In the content information's of the recipes, we get different categorical data like preparation direction, dietary information, different cuisines, occasional information and different courses information. This all information's are used as content information of each recipe. From the dataset, different statistic data and the features information are obtained which is described section III-B.

B. Statistics for dataset After the crawling process, the data is pre-processed. removes recipes with only 3 or fewer time ratings. Using the string-matching function, similar types of ingredients will be processed into one his ingredients. Therefore, many-to-one relationships are considered here. For example, Cortland apples, dried apples, and green apples are converted to apples. After the preprocessing step, the statistical data shown in Table I, Table II, and Table III are obtained.

TABLE I. STATISTICS OF DATASET

Number of Recipes	10,971
Number of Users	23,807
Total available ratings	3,43,308
Sparsity	0.132%
Average rating/ user	14.42
Average rating/ item	31.29

All statistics for the dataset are displayed data shown are a total of 23,807 users and 10,971 recipes available in the system. There are a total of 3,43,308 reviews. Therefore, the sparsity of the system can be calculated from, or 0.132%.

The other two tables list ingredients, cooking instructions, dishes, courses, and other characteristics.
statistics provided.

PROPOSED HYBRID APPROACHES

The purpose of this study is to find out which algorithm is suitable for personalized recipe recommendation by comparing the evaluation parameter RMSE of various traditional algorithms.

We focused on her two types of data.

These approaches combine each recipe with different characteristics such as ingredients, different cooking styles such as less than 30 minutes, 3 steps or less, less than 60 minutes, low fat, high Different meals such as protein, high carbohydrate, different occasions such as Diwali, birthday, summer, dinner party, different courses such as dessert, main course, salad. One is each feature of the detailed recipes Detailed data and seconds is high-level data rated by users for the recipe.

Here, we proposed two hybrid approaches that use the recipe content and evaluation information described in Sections 4-A and 4-B.

A. Proposed Hybrid Approach-1 This hybrid approach is based on the KNN (k-nearest neighbor) algorithm with use of the content information's of recipes and rating information of users on recipes. So, here one fine grain information of content of recipe is used which is implicit information gathering process and second rating of users on recipe which is explicit information gathering process. So, two matrixes are

TABLE II. STATISTICS OF INGREDIENTS

Total ings. counts in all recipes	56,740
Max. ings. in a recipe	40
Min. ings. in a recipe	2
Average ings. in a recipe	6
Max. appearance of ings. in a recipe	2282
Min. appearance of ings. in a recipe	1
Average appearance of ings. in a recipe	13.05

TABLE III STATISTICS OF FEATURES EXCLUDING INGREDIENTS

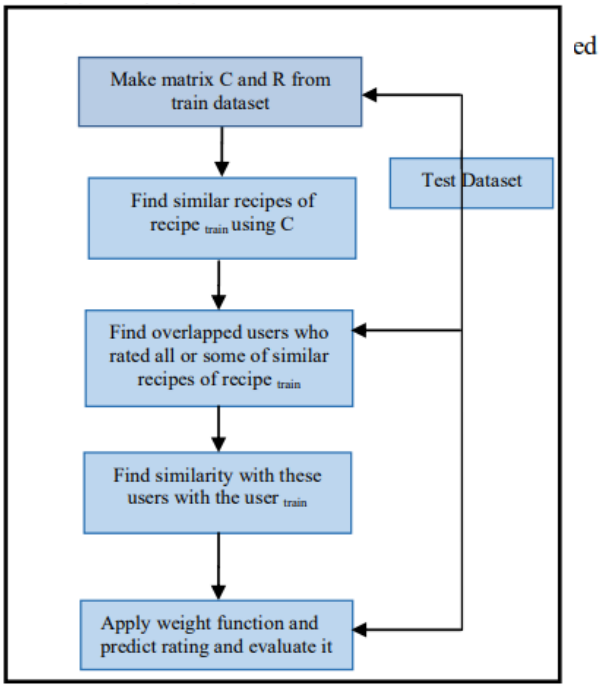
Total feats. counts in all recipes	2,41,259
Max. feats. in a recipe	88
Min. feats. in a recipe	3
Average feats. in a recipe	21.68
Max. appearance of feats. in a recipe	10,909
Min. appearance of feats. in a recipe	1
Average appearance of feats. in a recipe	106.75

generated

C = recipeid × featureid / ingredientid

$$C = \begin{cases} 1, & \forall\, ings, features \in recipeid_{train} \\ 0, & Otherwise \end{cases}$$

Using rating of users on the recipes, user's rating matrix R is generated:



ed

Fig. 2 Flow of Proposed Hybrid approach-1

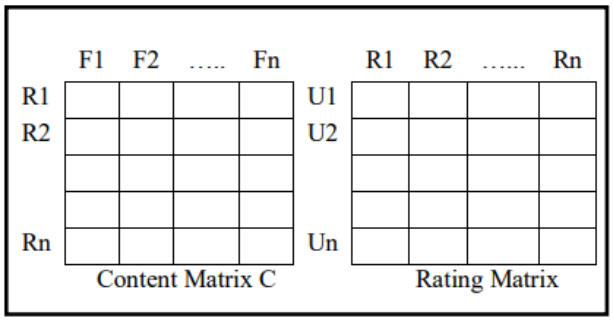


FIG. 3 Structure of content Matrix C and rating Matrix R

For given $recipeid_{train}$, find the similar type of other recipes from the content matrix C.

Similarity between recipes using KNN approach using Content matrix C:

$$\text{sim}(r_a, r_b) = \frac{\sum_{i=1}^{N} (r_{a_i} - \overline{r_a})(r_{b_i} - \overline{r_b})}{\sum_{i=1}^{N} (r_{a_i} - \overline{r_a})^2 \sum_{i=1}^{N} (r_{b_i} - \overline{r_b})^2} \quad (1)$$

Here r_a and r_b indicate the recipe a and recipe b respectively, r_{a_i} indicates the recipe a contains features f (including ingredient I) or not in 0 or 1 binary form. $\overline{r_a}$ indicates the average of recipe a (average of entire raw of recipe a) in content matrix C. Here N indicates the total number of recipes in the system.

Similarity between users using KNN approach using user's rating matrix R:

$$\text{sim}(u_a, u_b) = \frac{\sum_{i=1}^{N} (u_{a_i} - \overline{u_a})(u_{b_i} - \overline{u_b})}{\sum_{i=1}^{N} (u_{a_i} - \overline{u_a})^2 \sum_{i=1}^{N} (u_{b_i} - \overline{u_b})^2} \quad (2)$$

Here u_a and u_b indicate the user a and user b respectively, u_{a_i} indicates the user a contains recipe i or not in the range of 1 to 5. Here, $\overline{u_a}$ indicates the average of user a (average of entire raw of user a) in rating matrix R. Here N indicates the total number of users in the system.

The weight function for finding rating of recipe for given user based on the similarity thresholding value:

$$\text{rat}(u_a, r_i) = \frac{\sum_{n \in Thresholdinglimit} sim(u_a, u_n)\, rat(u_n, r_i)}{\sum_{n \in Thresholdinglimit} sim(u_a, u_n)} \quad (3)$$

Here, $\text{rat}(u_a, r_i)$ indicates the rating prediction for user u_a for recipe id i that is r_i, based on the different thresholding values between 0.1 to 0.9. Here u_n indicates the users who satisfied the particular thresholding values.

B. Proposed Hybrid Approach-2

This hybrid approach is based on the SGD (Stochastic gradient descent) algorithm with the content information as well as rating information of users on the recipes. SGD is the one of the collaborative filtering method and model based method. So, here the model is created first and based on that the prediction is done. Here only one matrix Make matrix A, which contains users and recipe contain information.

$$A = \begin{cases} User\ train\ rating\ table\ (recipeid, userid) = rating \\ 1, \quad \forall\, ings, features \in recipeid_{train} \\ 0, \quad Otherwise \end{cases}$$

Using matrix A, the flow of the algorithm is shown in figure 4. The matrix A defined by both information rating as well as content. The structural view of matrix is shown in figure 5.

For given $recipeid_{train}$ and $userid_{train}$, first the matrix A is created with rating and content information of recipes R. After that SGD (stochastic gradient descent) approach is applied here on rating matrix A is explained here. Using SGD, matrix A is divided into two matrices p and q. p and q is two sub-matrixes, whose multiplicative score try to give original matrix A which try to add some predicted values on the empty cell of matrix A.

The flow of the algorithm using matrix A is shown in Figure 4. Matrix A is defined by both information rating and content. The structure diagram of the matrix is shown in Figure 5.

Given the given $recipeid_{train}$ and $userid_{train}$, the first matrix is 44 4 Becomes 4 Recipe Created using R's rating and content information A. Next, it will now be explained that the SGD approach (Stochastic Gradient Descent) is applied to the evaluation matrix A.

Using SGD, matrix A is partitioned into two matrices p and q. p and q are two submatrices whose multiplication score attempts to give the original matrix A, and attempts to add the predicted values to the empty cells of matrix A.

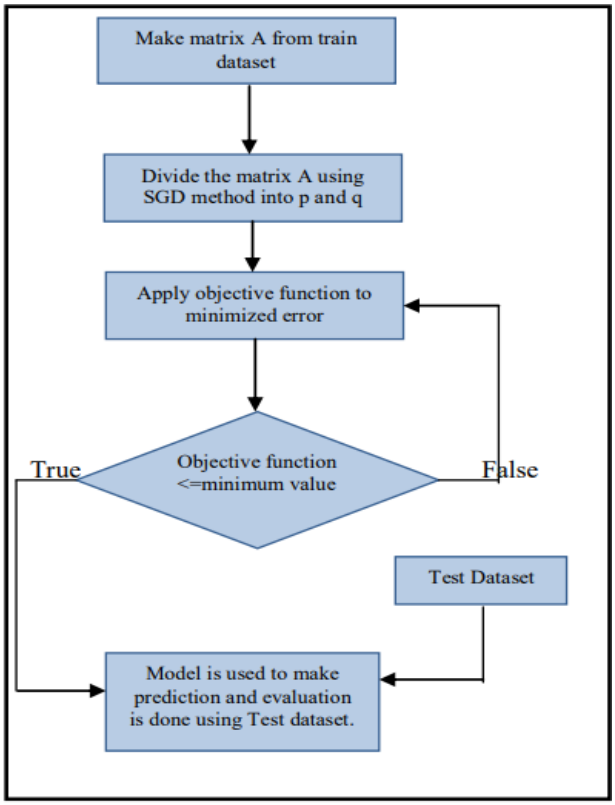


Fig. 4 Flow of Proposed Hybrid approach-2

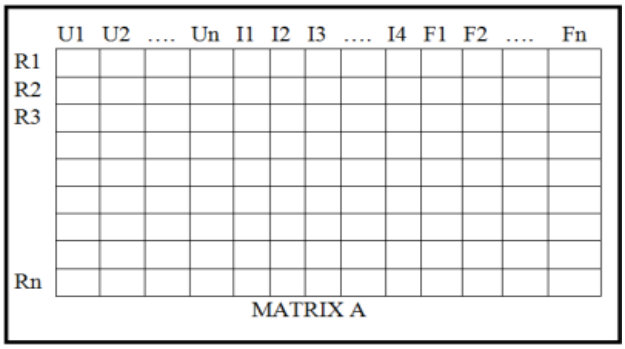


Fig. 5 Structure of matrix A

$$\bar{r}_{ui} = q_i^T p_u \quad (4)$$

The objective (loss) function is defined below is used to minimized the value of two sub-matrices p and q to reduced the error.

Apply the value of the q and p into the objective function:

$$\min_{q^*,p^*} \sum_{(u,i)\in k} (r_{u,i} - q_i^T p_u)^2 + \lambda \left(||q_i||^2 + ||p_u||^2 \right) \quad (5)$$

λ is the regularization parameter, k is the training samples, $r_{u,i}$ is rating of the training samples, p_u is user hidden factors and q_i is recipes hidden factors.

This objective function is checked every time with predefined minimum value and model is created and evaluation is done. To find the p and q value is iterative process which is used the objective function. So the value of p and q is based on the previous value of p and q which is defined as:

$$q_i' \leftarrow q_i + \gamma . (e_{ui} . p_u - \lambda . q_i) \quad (6)$$
$$p_u' \leftarrow p_u + \gamma . (e_{ui} . q_i - \lambda . p_u) \quad (7)$$

Here γ is the step size, e_{ui} is the error, q_i' is new value based on the q_i and p_u' is new value based on the p_u.

EVALUATION

There are various approaches to RS that were already available before in 1990. Therefore, we adopted three traditional models as benchmarks: User User-based collaborative filtering, item-to-item-based collaborative filtering, and rating-based SGD (Stochastic Gradient). descent). We compared our proposed hybrid approach, which includes both rating and content information, with his three traditional models in the culinary domain.

A. Setup There are approximately 10,000 recipe websites available to extract information for each recipe, including ratings provided by users.

For this purpose, I wrote a crawler program that extracts JSON data from the web and extracted the necessary information from it.

The process takes a long time to crawl to extract such a large amount of data.

This web content extraction took approximately 10-15 days.

From this information, we obtained user ratings for the recipes and information about the content of each recipe, such as the differences in ingredients and functions of the recipes.

As a preprocessing after cleaning the data, we discovered co-occurrences of features/components to determine their importance in the system.

After that, we adopted the top 90 ingredients and the top 90 recipe features as content information.

Therefore, we finally obtained content information and recipe evaluation information in the system and used this in the proposed model.

B. Results To determine the performance of the model, RMSE (Root Mean Squared Error) is used here instead of MAE (Mean Absolute Error). RMSE is very common and is ideal as a general-purpose error metric for numerical prediction. Compared to a similar mean absolute error, RMSE amplifies large errors and imposes severe penalties. Here, we adopted, three conventional models as benchmarks. A comparison of is shown in the figure.6 and RMSE of all five models.

From the figure: 6, the first model with collaborative inter-user filtering gives 0.7703 RMSE. This model only uses user rating information about recipes and finds rating predictions based on user similarity. The second model, between-item-based collaborative filtering T reduces the RMSE to 0.6848 compared to between-user-based collaborative filtering.

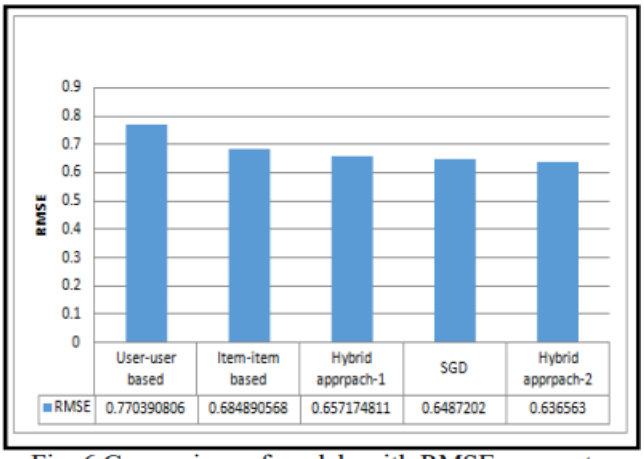


Fig. 6 Comparison of models with RMSE parameter

This model also only uses user rating information based on item similarity. Our proposed hybrid model 1, used content information with rating information and reduced the RMSE to 0.6571.

Here model works based on user similarity and content similarity. Similar to, the SGD model is one of the more powerful dual models because it first builds a precomputed model and then finds rating predictions. In the fourth model, SGD only uses users' rating information about recipes to reduce RMSE to 0.6487.

In the fifth model, the proposed hybrid approach 2 uses the recipe's content information and the recipe's user rating information to reduce the RMSE to 0.6365.

CONCLUSION AND FUTURE WORK

Recommendation systems have been the most advanced field in the past decade, with many content-based collaborative and hybrid approaches being proposed for various industrial growth purposes. In our work, we choose the culinary domain here, but content and collaborative filtering each have their advantages and disadvantages.

To overcome each other's problems, a hybrid approach is proposed here to try to reduce the RMSE. The SGD-based hybrid approach provides better RMSE than other approaches.

Here, we will break down recipe features and ingredients and use them as content information. A 1% reduction in RMSE also brings high effectiveness to RS. We used about 10,000 recipes and 20 3,000 users.

As future work, the large dataset can be used and more features can be added, such as flavor of ingredients by chemical properties for deep filtering purposes, thus allowing for different flavors and alternative ingredients.

REFERENCES

1. http://www.alexa.com/topsites/category/Top/Home
2. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, (8), 30-37.
3. Wang, L., Li, Q., Li, N., Dong, G., & Yang, Y. (2008, April). Substructure similarity measurement in Chinese recipes In Proceedings of the 17th international conference on World Wide Web (pp. 979-988). ACM.
4. E. Rich, "User Modeling via Stereotypes," Cognitive Science, vol. 3, no. 4, pp. 329-354, 1979.
5. G. Salton, Automatic Text Processing. Addison-Wesley, 1989.
6. Mino, Y., & Kobayashi, I. (2009, November). Recipe recommendation for a diet considering a user's schedule and the balance of nourishment. In Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on (Vol. 3, pp. 383-387). IEEE.
7. Forbes, P., & Zhu, M. (2011, October). Content-boosted matrix factorization for recommender systems: experiments with recipe recommendation. In Proceedings of the fifth ACM conference on Recommender systems (pp. 261-264). ACM.
8. Kular, Dal Inderjeet Kaur, Ronaldo Menezes, and Eraldo Ribeiro. "Using network analysis to understand the relation between cuisine and culture. "Network Science Workshop (NSW), 2011 IEEE. IEEE, 2011.
9. M.J.D. Powell, Approximation Theory and Methods. Cambridge Univ. Press, 1981.
10. Ueta, Tsuguya, Masashi Iwanami, and Takayuki Ito. "Implementation of a goal-oriented recipe recommendation system providing nutrition information. "Technologies and Applications of Artificial Intelligence (TAAI), 2011 International Conference on. IEEE, 2011.
11. Ahn, Yong-Yeol, et al. "Flavor network and the principles of food pairing. "Scientific reports 1 (2011).
12. Zhang, R., Liu, Q. D., Gui, C., Wei, J. X., & Ma, H. (2014, November). Collaborative Filtering for Recommender Systems. In Advanced Cloud and Big Data (CBD), 2014 Second International Conference on (pp. 301-308). IEEE.
13. Freyne, Jill, and Shlomo Belkovsky. "Intelligent food planning: personalized recipe recommendation." Proceedings of the 15th international conference on Intelligent user interfaces. ACM, 2010.
14. J.S. Armstrong, Principles of Forecasting—A Handbook for Researchers and Partitioners. Kluwer Academic, 2001.
15. Linden, Greg, Brent Smith, and Jeremy York. "Amazon.

com recommendations: Item-to-item collaborative filtering." Internet Computing, IEEE 7.1 (2003): 76-80.

16. Saladin, F. A., Zakaria, N., & Husain, W. (2010, November). User requirement analysis for personalized cancer dietary planning and menu construction. In Biomedical Engineering and Sciences (IECBES), 2010 IEEEEMBS Conference on (pp. 410-416). IEEE.

17. Maruyama, T., Kawano, Y., & Yanai, K. (2012, November). Realtime mobile recipe recommendation system using food ingredient recognition. In Proceedings of the 2nd ACM international workshop on Interactive multimedia on mobile and portable devices (pp. 27- 34). ACM.

18. Ueda, M., Asanuma, S., Miyawaki, Y., & Nakajima, S. (2014). Recipe recommendation method by considering the user's preference and ingredient quantity of target recipe. In Proceedings of the International Multiconference of Engineers and Computer Scientists (Vol. 1).

19. Wang, Liping, et al. "Substructure similarity measurement in Chinese recipes." Proceedings of the 17th international conference on World Wide Web. ACM, 2008.

20. De Campos, Luis M., et al. "Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks. "International Journal of Approximate Reasoning 51.7 (2010): 785-799.