Recipe Recommendation System

Wenxin Fang, Xingyu Liu, Victoria Wanjiang Zheng, Yujia Xie, Peicheng Hua, Yifei Wang

March 2017

1 Problem Definition

People always face same problem in their daily life: What to eat and how to eat healthily. What's more, people who love cooking usually don't know how to start when facing too many ingredient pairings. Our Recipe Recommendation System give users nutritious recipe choices which also satisfy their personal taste.

2 Heilmeier Questions

2.1 Project Motivation and Objectives

We propose a system that recommend people recipes appeal to their needs in flavor and nutrition. This recommendation is based on the data clarifying the relation between recipe and ingredients as well as ingredients and compounds. The objective is to help people get the recipes satisfying their needs more easily and time-efficiently.

2.2 Limits of current Actions

Current practice always separate people's need in flavor and nutrition. The current systems usually recommend the recipes which appeal to people's need only in flavor or only in nutrition. The limit is that recipe websites usually recommend recipes without considering nutrition need, and health apps usually calculate calorie but ignore personal flavor.

2.3 Approach

The proposed system will perform better because it takes into consideration not only people's flavor need but also nutrition need. We analyze different flavor of recipe based on ingredients' compounds which goes beyond users' rating and becomes more objective.

2.4 Target Customers

The people who love cooking or care about their healthy or body shape may benefit from our system. We can help them find recipes suitable for them precisely and quickly. Besides, specific people like athletes and models, who need to maintain a good body condition such as appropriate weight or beautiful body shape need this system as well.

2.5 Impact and Evaluation of the Project

Help people with their meal decisions and bring up their health-consciousness. Also, we apply machine learning algorithms in nutritional science field
user studies: user feedback. User number observation
Groundtruth data: used while model building

2.6 Risks and Payoffs

Challenge of data collection and clean. How to balance the time spent on data parsing and data mining. Payoffs: Future opportunities with organic grocery stores and restaurants.

2.7 Cost expenses

The system will be built with open source technologies. Hence, there are no associated apart from the time and effort committed by the six team members over a six week period.

2.8 Project duration

The project can be done in six weeks.

2.9 Quality Assurance Approach

At the end of the third week, a trained model with flavor clusters and nutrition requirement will be complete. The functions can be tested through command lines. At the end of the sixth week, a fully functional prototype with a UI will be implemented. The idea goal would be to open the web page on mobile device and get the recommendation appropriately.

3 Literature Survey

The first challenge in our project is to measure the dietary intake which satisfies different users' daily nutritional goals. One well-known and widely-accepted method for the measurement of dietary intake is to use Estimated Energy Requirements(EER) equation [1] [2] based on the user's weight, height, age, sex and physical activity level. Physical activity level(PAL) [3] is an index related with the intensity and impact of various activities adults do in their daily life. According to the calorie level assessed, we could get daily macronutrients goals for specific users [4]. For those with high Body Mass Index(BMI), a reasonable deduction in calorie level in the first step will be made so as to help them lose weight [5]. By getting these statistics, we are allowed to recommend recipes which suits individual nutrition requirement.

After satisfying nutritional need, another challenge in our project is to obtain an ordered list of recipe recommendations. Our first thought is constraint optimization algorithm, which requires to set up a utility function that measures the usefulness of items to users [6]. However, considering the magnitude of databases and the cost of coordinating databases, we believe it is a better idea to split the problem into selecting feasible set and building a recommender system [7]. First, we let users to choose what flavor they would like today, and shrink the dataset of recipes based on the compounds in the ingredients [8] [9]. Then, the dataset can be further reduced by adding nutritional constraints. Finally, we order the feasible set by hybrid recommending [10]. Like Cotter & Smyth (2000)[11], we plan to merge the content-based information in step one and results of collaborative filtering methods [7][12][13] to produce a final list. As for the cold start problem [14] that is usually faced by recommender system, since content-based approaches [15][16] do not rely on ratings from other users, they can be used to produce recommendations for all items, provided attributes of the items are available. In fact, the content-based predictions of similar users can also be used to further improve predictions for the active user [17].

For our algorithm, recipe data will be collected and studied. For each recipe, we will parse out the information about the ingredients and compounds it contains. Then we will use flavour network [18] to project recipes with ingredients into our flavour space. In order to classify the recipes, different clusters among recipes will be created to select recipes from the cluster of flavor more close to user's preference. There are various ways to perform the clustering [19]. The traditional clustering method (K-means) performs better [20] than neural networks using Kohonen learning using simulated data with known cluster solutions [21]. A popular Lloyd's k-means clustering algorithm is also easy to implement and very efficient, requiring a kd-tree as the only major data structure [22]. In addition, one may consider employing hierarchical clustering, since they have the similar performance but potentially it might be easier to measure the dissimilarity among clusters.

However, our literature survey is far from being satisfactory. More work needs to be done on the clustering algorithms and recommender systems. We may also consider the requirement of users with special need such as pregnancy, lactation and living with diabetes in order to extend the range of recipe recommendation system application.

4 Plan of Activity

There is total 6 weeks for us to complete our project. Effort distribution is shown as above in Table 1.

	Week 1	Week 2-3	Week 4-5	Week 6
Wenxin	clean data	try different clustering algorithm	UI design	Improve UI
Xingyu	Build web app framework	try different clustering algorithm	connect algorithm code with UI	deployment and test backend
Wanjiang	clean data	transfer data and combine	add some recommendation algorithm	Iterate over UI interface
Yujia	extra recipe data from API	transfer data and combine	add some recommendation algorithm	Improve UI
Peicheng	extra recipe data from API	code to satisfy nutrition requirement	UI design	test label of clustering algorithm
Yifei	database relation design	code to satisfy nutrition requirement	UI design	test output of recommendation algorithm

References

[1] I. of Medicine, *"ENERGY." Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids.* Washington (DC):National Academies Press, 2002.

[2] B. E. Millen, S. Abrams, L. Adams-Campbell, C. A. Anderson, J. T. Brenna, W. W. Campbell, S. Clinton, F. Hu, M. Nelson, M. L. Neuhouser, *et al.*, "The 2015 dietary guidelines advisory committee scientific report: development and major conclusions," *Advances in Nutrition: An International Review Journal*, vol. 7, no. 3, pp. 438–444, 2016.

[3] I. of Medicine, *Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein and amino acids.* Washington (DC):National Academies Press, 2002.

[4] I. of Medicine., *Dietary reference intakes: the essential guide to nutrient requirements.* Washington (DC):National Academies Press, 2006.

[5] M. D. Jensen, D. H. Ryan, C. M. Apovian, J. D. Ard, A. G. Comuzzie, K. A. Donato, F. B. Hu, V. S. Hubbard, J. M. Jakicic, R. F. Kushner, *et al.*, "2013 aha/acc/tos guideline for the management of overweight and obesity in adults," *Circulation*, vol. 129, no. 25 suppl 2, pp. S102–S138, 2014.

[6] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE transactions on knowledge and data engineering*, vol. 17, no. 6, pp. 734–749, 2005.

[7] G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering," *IEEE Internet computing*, vol. 7, no. 1, pp. 76–80, 2003.

[8] Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, and A.-L. Barabási, "Flavor network and the principles of food pairing," *Scientific reports*, vol. 1, 2011.

[9] C.-Y. Teng, Y.-R. Lin, and L. A. Adamic, "Recipe recommendation using ingredient networks," in *Proceedings of the 4th Annual ACM Web Science Conference*, pp. 298–307, ACM, 2012.

[10] P. Melville and V. Sindhwani, "Recommender systems," in *Encyclopedia of machine learning*, pp. 829–838, Springer, 2011.

[11] P. Cotter and B. Smyth, "Ptv: Intelligent personalised tv guides," in *AAAI/IAAI*, pp. 957–964, 2000.

[12] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, vol. 35, no. 12, pp. 61–70, 1992.

[13] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "Grouplens: an open architecture for collaborative filtering of netnews," in *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pp. 175–186, ACM, 1994.

[14] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and metrics for cold-start recommendations," in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 253–260, ACM, 2002.

[15] R. J. Mooney and L. Roy, "Content-based book recommending using learning for text categorization," in *Proceedings of the fifth ACM conference on Digital libraries*, pp. 195–204, ACM, 2000.

[16] M. Pazzani and D. Billsus, "Learning and revising user profiles: The identification of interesting web sites," *Machine learning*, vol. 27, no. 3, pp. 313–331, 1997.

[17] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-boosted collaborative filtering for improved recommendations," in *Aaai/iaai*, pp. 187–192, 2002.

[18] Y.-Y. Ahn and S. Ahnert, "The flavor network," *Leonardo*, vol. 46, no. 3, pp. 272–273, 2013.

[19] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.

[20] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern recognition letters*, vol. 31, no. 8, pp. 651–666, 2010.

[21] P. S. Balakrishnan, M. C. Cooper, V. S. Jacob, and P. A. Lewis, "A study of the classification capabilities of neural networks using unsupervised learning: A comparison withk-means clustering," *Psychometrika*, vol. 59, no. 4, pp. 509–525, 1994.

[22] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 881–892, 2002.