Recommending Food: Reasoning on Recipes and Ingredients

Jill Freyne and Shlomo Berkovsky

CSIRO, Tasmanian ICT Center
GPO Box 1538, Hobart, 7001, Australia
firstname.lastname@csiro.au*

Abstract. With the number of people considered to be obese rising across the globe, the role of IT solutions in health management has been receiving increased attention by medical professionals in recent years. This paper focuses on an initial step toward understanding the applicability of recommender techniques in the food and diet domain. By understanding the food preferences and assisting users to plan a healthy and appealing meal, we aim to reduce the effort required of users to change their diet. As an initial feasibility study, we evaluate the performance of collaborative filtering, content-based and hybrid recommender algorithms on a dataset of 43,000 ratings from 512 users. We report on the accuracy and coverage of the algorithms and show that a content-based approach with a simple mechanism that breaks down recipe ratings into ingredient ratings performs best overall.

Key words: Collaborative filtering, content-based, ingredient, recipes

1 Introduction

The World Health Organisation [1] is predicting that the number of obese adults worldwide will reach 2.3 billion by 2015 and the issue is attracting increased attention. Much of this attention is being paid to online diet management systems, which have been replacing traditional pen-and-paper programs. These systems include informative content and services, which persuade users to alter their behaviour. Due to the popularity of diet monitoring facilities, these systems hold a vast amount of user preference information, which could be harnessed to personalize interactive features and to increase engagement with the system and, in turn, the diet program. One such personalized service, ideally suited to informing diet and lifestyle, is a personalized recipe recommender. This recommender

* This research is jointly funded by the Australian Government through the Intelligent Island Program and CSIRO Preventative Health Flagship. The Intelligent Island Program is administered by the Tasmanian Department of Economic Development, Tourism, and the Arts. The authors acknowledge Mealopedia.com and Penguin Group (Australia) for permission to use their data.

could exploit explicit food ratings, food diary entries, and browsing behaviour to inform its recommendations.

The domain of food is varied and complex and presents many challenges to the recommender community. The content or *ingredients* of a meal is only one component, which impacts a user's opinion. Others include *cooking methods*, *ingredient costs* and *availability*, *complexity of cooking*, *preparation time*, *nutritional breakdown*, *ingredient combination effects*, as well as *cultural* and *social factors*. Add to this the sheer number of ingredients, the fact that eating often occurs in groups, and that sequencing is crucial, and the complexity of challenge becomes clear.

Initial efforts in addressing these challenges have resulted in systems, such as Chef [3] and Julia [5], which rely hugely on domain knowledge in their recommendation processes. Conversely, fuzzy logic [9] and active learning and knowledge sources techniques have been applied in [10] to generate recipes from ingredient sets without the need for expensive domain knowledge.

In this work we turn to the traditional recommender technologies to understand their applicability and accuracy in the food domain. We present a preliminary study into the suitability of recommender algorithms for recipe recommendation. The study is based on preferences provided by 512 users on a corpus of recipes. We examine the accuracy of collaborative, and content-based filtering algorithms, and compare them to hybrid recommender strategies, which break down recipes into their ingredients in order to generate more accurate recommendations. We show that solicitation of recipe ratings, which are transferred to ingredient ratings, is an accurate and effective method of capturing ingredient preferences, and that the introduction of simple intelligence can improve the accuracy of recommendations.

2 Recommender Strategies

The aim of this work is to develop recommender algorithms for personalized recipe recommendations. Figure 1 shows the simple recipe to ingredient relationship strategy adopted in this work. We ignore all cooking processes and combination effects and consider all ingredients to be equally weighted within a recipe. Also, we transfer ratings gathered on recipes equally to all its ingredients, and vice versa, from ingredients to their associated recipes. In contrast to previous work [2], here we solely investigate strategies, in which ratings are available on *recipes* and evaluate them on a much larger dataset.

In order to compare our recommender strategies, we implement a baseline algorithm *random*, which assigns a randomly generated prediction score to a recipe. We implemented five personalized recommender strategies. The first is a standard *collaborative filtering* algorithm assigning predictions to recipes based on the weighted ratings of a set of N *neighbours*. Briefly, N neighbours are identified using Pearson's correlation algorithm shown in Equation 1 and predictions for recipes not rated by user u_a are generated using Equation 2.

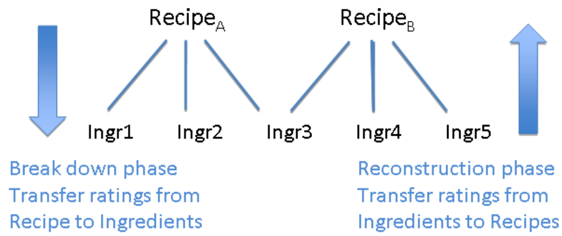

Fig. 1. Recipe - ingredient breakdown and reconstruction

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

$$pred(u_a, r_t) = \frac{\sum_{n \epsilon N} sim(u_a, u_n) rat(u_n, r_t)}{\sum_{n \epsilon N} sim(u_a, u_n)} \tag{2}$$

The second is a *content-based* algorithm, which breaks down each recipe r_i rated by u_a into ingredients $ingr_1, ..., ingr_x$ (see Figure 1) and assigns the ratings provided by u_a to each ingredient according to Equation 3. The strategy then applies a content-based algorithm shown in Equation 4 to predict a score for the target recipe r_t based on the average of all the scores provided by user u_a on ingredients $ingr_1, ..., ingr_j$ making up r_t.

$$score(u_a, ingredient_i) = \frac{\sum_{l \ s.t. \ ingr_i \epsilon r_l} rat(u_a, r_l)}{l} \tag{3}$$

$$pred(u_a, r_t) = \frac{\sum_{j \epsilon r_t} score(u_a, ingr_j)}{j} \tag{4}$$

We also implemented two *hybrid* strategies. Both of these break down each recipe rated by u_a into ingredients and exploit collaborative filtering techniques to reduce the data sparsity of the ingredient matrix by generating predictions for ingredients on which we have no information. The first strategy, $hybrid_{recipe}$, identifies a set of neighbours based on ratings provided on recipes as in Equation 1 and predicts scores for unrated ingredients using Equation 2 (applied to ingredient scores rather than recipe ratings). With the denser ingredient data, the content-based prediction shown in Equation 4 is used to generate a prediction for r_t. The second strategy, $hybrid_{ingr}$, differs from $hybrid_{recipe}$ only in its neighbour selection step. In $hybrid_{ingr}$, user similarity is based on the ingredients scores obtained after the recipe break down rather than on the recipe ratings as in $hybrid_{recipe}$.

3 Evaluation

We gathered a set of 43,893 recipe ratings from 512 users through the Amazon owned online HCI task facilitator Mechanical Turk (www.mturk.com). Online surveys, each containing 36 randomly selected recipes, were posted to the system and users were allowed to answer as many surveys as they choose.

3.1 Set-up

The corpus of recipes used was sourced from the CSIRO Total Wellbeing Diet Books [7, 8] and the online meal planner Mealopedia (www.mealopedia.com). We extracted 404 recipes, which corresponded to 479 unique ingredients. On average, each recipe was made up of 9.52 ingredients (stdev 2.63) and the average number of recipes that each ingredient was found in was 8.03 (stdev 19.86). We gathered opinions of 512 users regarding the available recipes. Users were asked to provide their preferences on how much each recipe appealed to them. Each user provided at minimum 36 recipe ratings and also their demographical information. All ratings were captured on a 5-Likert scale, spanning from "not at all" to "a lot" (6274 recipes rated *not at all*, 6272 – *not really*, 8445 – *neutral*, 10873 – *a little*, and 12029 – *a lot*). In total 43,893 preferences were gathered with an average of 85.73 per user, such that the ratings matrix was 21.22% complete.

We conducted a traditional leave one out off-line analysis, which took each $\{u_i,\ r_t,\ rat(u_i, r_t)\}$ tuple from a user profile and used the algorithms presented in Section 2 to predict the rating $rat(u_i, r_t)$. A set of 20 neighbours were selected only once for each user, based on the entire set of ratings provided. The performance of the recommenders was evaluated using the normalized MAE measure [4] and coverage, i.e. their ability to generate recommendations.

3.2 Results

The *content-based* and both *hybrid* strategies obtained over 99% coverage. For *collaborative filtering*, the coverage was 95.9%, and for *random* recommendations it was obviously 100%. Hence, there is no significant coverage benefit gained by any approach on this dataset.

The lighter bars in Figure 2 show the MAE score obtained for each strategy. As expected, the *random* algorithm performed worst with an MAE of 0.399. The poorest personalized strategy was the *collaborative filtering* algorithm with an MAE of 0.328. Producing MAE scores of 0.309 and 0.330 are the two hybrid strategies, $hybrid_{recipe}$ and $hybrid_{ingr}$, respectively. In comparison to the best performing, *content-based* algorithm, which obtained an MAE of 0.262, both *hybrid* strategies introduce noise in the ingredient scores during the collaborative filtering step. This finding is consistent with that of Melville *et al.*, who also used content-boosted collaborative filtering [6]. The differences in MAE are significant at $p < 0.05$ across all pairings.

A comparison between the *collaborative filtering* algorithm, which treats each recipe as one entity and ignores its ingredients, and the *content-based* algorithm, which considers the ingredients, shows that even the naive break down and reconstruction rules applied here offer significant performance benefits in accuracy.

The decomposition of recipes into ingredients implemented in this experiment was simplistic: an ingredient score was computed by averaging the ratings of

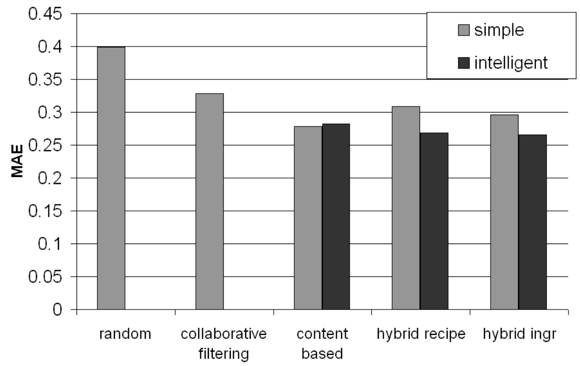

Fig. 2. Normalized MAE score

recipes in which it occurs. This lead to a large number of *mixed* ingredient scores. For example, consider two recipes containing an ingredient i: one is liked and one disliked, but i is not the cause of the dislike. Despite not being the cause of the dislike, i will receive only a neutral score. To address this shortcoming, we implemented a more intelligent strategy, in which only the positive ratings for ingredients that receive mixed ratings are considered.

We see the impact of this assumption in the darker bars in Figure 2. In the *content-based* algorithm, the impact is negligible. However, we see a positive impact of the intelligent break down on the MAE of the two *hybrid* strategies. Now, the intelligent hybrid strategies significantly outperform the *content-based* algorithm with an MAE of 0.269 and 0.265 for the intelligent versions of $hybrid_{recipe}$ and $hybrid_{ingr}$ strategies, respectively. The differences between the previously best performing*content-based* algorithm and both the intelligent versions of $hybrid_{recipe}$ and $hybrid_{ingr}$ strategies are significant at $p < 0.05$, as is the difference between both intelligent strategies. The collaborative filtering step used to generate predictions on the unknown food items has benefited the most from the introduction of this intelligent break down process.

Hence, the best performing algorithm is intelligent $hybrid_{ingr}$, which exploits both content-based and collaborative filtering techniques. Neighbours are determined based on the implied ratings of recipes, which have been transferred down to ingredient scores while reasoning on the presence of mixed ratings. Then, collaborative filtering is used to predict scores for unrated ingredients, and, finally, a recipe prediction is computed by averaging the scores of its ingredients.

4 Conclusions and Future Work

In this work we have investigated the applicability of recommender techniques to generate recipe recommendations. We found that high coverage and reasonable accuracy can be achieved through content-based strategies with a simple break down and construction used to relate recipes and ingredients. We found significant accuracy improvement through use of content-based techniques over a collaborative filtering algorithm. However, the optimal solution was obtained when we bootstrap the recommender process by breaking a recipe down into ingredients, computing ingredient scores, applying the collaborative filtering step to decrease the sparsity of the ingredient scores, and, finally, applying the content-based recipe rating prediction process by examining the scores of individual ingredients.

As noted earlier, there are many factors that influence a user's rating beyond a recipe content. Thus, our future work will focus on extraction of recipe features, such as complexity, time and cooking methods, to examine their impact on user ratings. Furthermore, here we implemented a simplistic idea of what a recipe recommender needs to achieve. We are, however, aware that generating recipe recommendations is a far more complicated task in reality, and we will investigate the issues of group recommendations, where varying social relationships can be at play. In particular, we aim to investigate how family roles and relationships affect compromise and satisfaction with menu plans. Complimentary to this, we need to examine applicability of sequential recommendations. Menu recommendations would not generally be provided in a single shot interaction, but rather users will plan meals over a period of time, such that diversity and satisfaction levels are complex, in particular when groups of users are involved.

References

1. World health organization, chronic disease information sheet. www.who.int/mediacentre/factsheets/fs311/en/index.html, accessed Jan '10.
2. J. Freyne and S. Berkovsky. Intelligent Food Planning: Personalized Recipe Recommendation. In *Proccedings of the International Conference on Intelligent User Interfaces*, pages 321–324, 2010.
3. K. Hammond. CHEF: A Model of Case-Based Planning. In *Proceedings of the National Conference on Artificial Intelligence*, 1986.
4. J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1):5–53, 2004.
5. T. Hinrichs. Strategies for adaptation and recovery in a design problem solver. In *Proceedings of the Workshop on Case-Based Reasoning*, 1989.
6. P. Melville, R. Mooney, and R. Nagarajan. Content-boosted collaborative filtering for improved recommendations. In *Proceedings of the National Conference on Artificial Intelligence*, 2002.
7. M. Noakes and P. Clifton. *The CSIRO Total Wellbeing Diet.* Penguin, 2005.
8. M. Noakes and P. Clifton. *The CSIRO Total Wellbeing Diet Book 2.* Penguin, 2006.
9. J. Sobecki, E. Babiak, and M. Slanina. Application of hybrid recommendation in web-based cooking assistant. In *Proceedings of the Conference on Knowledge-Based Intelligent Information and Engineering Systems*, 2006.
10. Q. Zhang, R. Hu, B. Namee, and S. Delany. Back to the future: Knowledge light case base cookery. Technical report, Dublin Institute of Technology, 2008.