

# Towards Automatic Meal Plan Recommendations for Balanced Nutrition

David Elsweiler<sup>1</sup> Morgan Harvey<sup>2</sup>

<sup>1</sup> University of Regensburg, Germany

<sup>2</sup> Northumbria University, Newcastle, United Kingdom

e-mail: david@elsweiler.co.uk, morgan.harvey@northumbria.ac.uk

## ABSTRACT

Food recommenders have been touted as a useful tool to help people achieve a healthy diet. Here we incorporate nutrition into the recommender problem by examining the feasibility of algorithmically creating daily meal plans for a sample of user profiles (n=100), combined with a diverse set of food preference data (n=64) collected in a natural setting. Our analyses demonstrate it is possible to recommend plans for a large percentage of users which meet the guidelines set out by international health agencies.

## Keywords

Health, Recommender Systems, Planning, Prevention

## 1. INTRODUCTION AND MOTIVATION

Poor dietary habits are a major cause of today's world health problems. Evidence shows that issues such as obesity and diabetes, as well as other lifestyle-related illnesses, can be prevented and sometimes even reversed through good nutrition [3]. People are often very poor at judging the healthiness of their own diet [2] and need support to implement positive changes [5]. Nutritionalists can create long-term plans and help people learn to make better choices, however this solution is neither practical nor economically feasible for everyone. As a result, food recommender systems (RS) have been touted as a potential means to assist people in nourishing themselves more healthily [4, 6].

Food recommenders make sense as part of a strategy for behavioural change as suggesting a change that is less painful, i.e. based on something the user might like, is more likely to be accepted and followed. Recommenders are likely to be effective at predicting which changes will be painful or not but they have a serious drawback when aiming for positive change: they learn user preferences for ingredients and food styles. This leads to users who like fat- and calorie-laden meals being recommended fat- and calorie-laden meals [6] - an outcome not conducive to improving nutritional habits. Here we begin to address a fundamental question: to which

extent is it possible to recommend users recipes they will like, but at the same time fit into a balanced diet?

Our approach is to automatically create daily meal plans by first calculating the nutritional requirements of the user based on their personal *personas* (age, gender, height, etc.). We then take the top recommendations, as estimated by a state-of-the-art recipe recommendation algorithm [6], and attempt to combine these algorithmically such that each plan corresponds to guidelines published by international health agencies.

## 2. RELATED WORK

Early work on recommending recipes include JULIA [7], which used case-based planning to generate a meal plan to satisfy multiple, interacting constraints. More recent efforts try to better understand the user's tastes [4] and how various factors in the rating process influence a user's choice [6]. Although food RS has been mooted as a potential tool to assist users in achieving good nutrition, to our knowledge no published work has explained exactly how to achieve this. We therefore present a first approach to integrating nutrition in the RS by moving beyond recommending individual items - which alone may be considered "unhealthy" - to groups of items, which together may represent a more balanced whole.

Recommending groups of items requires a balance to be struck between the items with the highest individual utility and those which, while they may have lower predicted ratings in isolation, join together to give a better combination. Examples include recommending complete, complementary playlists of music tracks [8] or combining recommended tourist attractions into a plan for a user's stay [9].

Schaller et al. [11] combine event recommendations given defined start and end times and locations, as well as the user's preferences. A schedule is built such that the user can visit as many of their preferred events as possible, whilst adhering to time constraints imposed by the duration of each event and travelling times. In many ways this is analogous to our situation where we are trying to combine a user's preferred meals given a number of nutritional constraints.

Little literature exists dealing with combined meal recommendations. JULIA [7] creates meal plans, although not for health or nutritional reasons. A second exception is an interface allowing users to manually combine recommendations into a food plan [1]. This work is interesting as it provides an example situation where our algorithms could be used. Our goal is slightly different. We wish to produce plans which are both pleasing to the user and, at the same time, meet the user's daily nutritional needs.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

### 3. METHOD

To collect food preferences data we created a food portal website, [anonymised], where users could upload, browse, share and search recipes, calculate the nutritional properties of a recipe and receive recipe recommendations based on their profile. Users can rate recipes from 1 (strongly dislike) to 5 stars (strongly like). Over 3 years 148 users gave 4,549 ratings (median 8 per user) of 957 recipes, consisting of diverse styles of food. The median number of ratings per recipe is 4 and each recipe is classified as being either a main meal, a side dish, a breakfast or a dessert.

To investigate meal planning we needed a number of varied user *personas* describing quantities like height, weight, gender, age, nutritional goal (lose/gain/maintain weight) and activity level (from sedentary to highly active). Due to the general variability of the human form, we decided not to randomly generate body measurements. Instead we randomly sampled 50 male and 50 female subjects from a large survey of the US population <sup>1</sup> which reports general demographic information as well as detailed information about weights, heights, BMIs and ages at the time of the survey. Activity levels and goals were assigned to each subject at random assuming unconditional uniform distributions.

The 100 personas have a mean weight of 82.24kg and a mean height of 167.4 cm - an average BMI of 27.44 - meaning 3 are underweight, 27 overweight, 40 obese and the remaining 30 have a healthy weight. Ages vary almost uniformly from 18 to 80 with a median age of 46. This represents the US population well, serving to highlight the obesity present in modern human populations and providing us with a realistic and varied set of data points, allowing us to test the effect individual differences have on our planning approaches.

#### 3.1 Creating Meal plans

For each persona-profile combination we combine meals, which the user has told us he likes and others that we predict he will like based on his profile, such that the combination meets his daily nutritional requirements. This requires several steps: 1) establishing the nutritional requirements for the persona 2) estimating recipe ratings based on each profile 3) combining recipes and establishing if the combination meets the requirements.

We calculate nutritional requirements for a given persona using a version of the Harris-Benedict equation revised by Roza and Shizgal [10] which estimates an individual's basal metabolic rate (BMR) and daily calorie requirements. The estimated BMR value is multiplied by a number between 1.2 and 1.9 corresponding to the individual's activity level giving a recommended daily energy intake to maintain current body weight. We add or remove 500 kilocalories for individuals who wish to gain or lose weight. This is a crude mechanism which would result in the safe gain or loss of 0.45 kgs in 7 days. We assume that 20% of the required energy will come from drinks and between-meal snacks.

Nutritional scientists have established generally-accepted ranges for carbohydrates, fat and protein intake ensuring that a person maintains a balanced diet with sufficient intake of essential nutrients, vitamins, and minerals. 45% to 65% of calories eaten should come from carbohydrates, 20% to 35% from fat and 10% to 35% from protein <sup>2</sup>. As a met-

ric to optimise we target the central value of each of these ranges and set a fixed acceptable error margin of 10%. By accounting for the rate of calories per gram we can calculate how many calories should come from each energy source.

The diversity of user taste profiles can be estimated from the variety of ingredients in rated recipes. A user who has rated recipes with a wide variety of different ingredients likely has a wider range of tastes than another user who sticks to a smaller set. *Simple diversity* is the number of distinct ingredients in the recipes rated by a given user divided by the sum of all ingredients in the same set of recipes. Given a set  $N$  of integer counts, *Simpson diversity* is:  $D(N) = 1 - \frac{\sum_{n \in N} n(n-1)}{\sum_{n \in N} n(\sum_{n \in N} n-1)}$  and measures the probability that two elements randomly selected from a sample will belong to the same class (in this case ingredient). For both metrics 1 is infinite diversity and 0 is no diversity.

Omitting users with very few (< 10) ratings, leaving 64, both diversity metrics show an approximately normal distribution (*Simple diversity* mean = 0.57, sd=0.21; *Simpson diversity* mean = 0.98, sd = 0.004). Therefore most users had a good degree of ingredient diversity with few cases of low diversity. There is a strong linear relationship between ingredient diversity of rated recipes and the ingredient diversity of those recommended by the system ( $r = 0.716$ ), meaning users with very narrow tastes will continue to have narrow tastes if they follow the recommendations.

Most users only have a relatively small number of their rated recipes from the top 50 most popular (<sup>3</sup>/<sub>4</sub> of users drew less than 12% of their recipes from this set). The prevalence of top 50 recipes is negatively correlated with the diversity of the rated recipes ( $r = -0.558$ ), meaning that people who have a preponderance towards popular recipes have a less diverse taste range.

4) We can create plans for a given user (persona-profile combination) by first taking the top  $x$  recommendations from the RS for the taste profile. This set of recipes is then split into two separate sets, one for breakfasts and one for main meals. A full search is performed to find **every combination** of these recipes in the sequence [breakfast, main meal, main meal] which meets the target nutritional requirements defined above. Combinations with the same meals cannot be repeated, e.g. [R1, R2, R3] and [R1, R3, R2] are treated as only one plan.

This elementary approach is sensible as 1) we are analysing the feasibility of creating plans and this provides all combinations meeting the set criteria and 2) we are dealing with small sets of recommendations so efficiency is not an issue.

### 4. RESULTS

For the following analyses (except in Section 4.3) the number of top recommendations used for planning,  $x$ , is set to 100. Combining taste profiles for users who have rated 10 or more recipes ( $n=64$ ) with the 100 sampled personas yields 6400 simulated users. The method outlined above was able to generate plans for 4025/6400 cases (63%) and at least 1 plan was generated for 58 out of the 64 (91%) user profiles and for all 100 personas. On a per-persona basis, the total number of plans generated and the number of profiles for which any plans can be generated are related. As the number of "easy" profiles increases, the number of plans that can be generated from those profiles increases polynomially, thus increasing the number of recipes users rate will quickly

<sup>1</sup>NHANES - <http://www.cdc.gov/nchs/nhanes.htm>

<sup>2</sup><https://www.nrv.gov.au/chronic-disease/summary>

improve coverage of the planning algorithm.

**For the majority of people to eat in a way conforming to nutritional guidelines, we do not have to change what they eat, rather just when they eat what they like and how they combine meals.** That said, there was still a number of user combinations for whom plans could not be generated. To understand why we defined features describing each profile and used these to compare the difference between the profiles for which we can generate plans and those we can't. Comparing these groups statistically showed that profiles are "difficult" if:

- They tend to only rate highly calorific and fatty recipes
- They rate very few breakfasts, which is a potential bottleneck in our algorithm
- They rate recipes with a lower diversity of ingredients
- The number of recipes they have rated is low

These insights point to a general theme - difficult profiles exhibit less diversity than the easier ones, meaning the set of recipes used by the planning algorithm is itself less diverse.

To understand how good the generated plans are we need to define a quality metric. We use a normalised linear combination of the expected rating from the RS (i.e. whether the user will like the recipe) and the inverse of the error compared to the user's ideal nutritional profile (i.e. distance from what the user should be eating). For a plan  $i$  this is:

$$score(i) = \lambda \cdot \left(1 - \frac{\epsilon_i}{\forall i Max(\epsilon_i)}\right) + (1 - \lambda) \cdot \frac{\hat{\mu}_i}{\forall i Max(\hat{\mu}_i)}$$

where  $\epsilon_i$  is the error for plan  $i$  and  $\hat{\mu}_i$  is its expected rating.  $\lambda$  is set to 0.5 to apply equal weighting to both nutrition and taste. As each plan is actually composed of three recipes (a breakfast, lunch and dinner), the error and expected ratings in the above equation are the average of these values over the three items. The error distribution is normal with mean 0.048 (4.8% of the total) and  $\sigma^2$  0.015. The distribution of ratings is approximately normal with peaks at 4 and 5 (due to the presence of the actual ratings).

To understand what features of profiles and personas have the greatest impact on the plans generated we identified the top ("easy") and bottom ("hard") quartiles of each based on their mean scores. Note that for all of the comparisons in the following section we first assessed the normality of each distribution using the Shapiro-Wilk test and then performed the appropriate statistical test to compare the two distributions: Student's t-test for normal data (parametric) and the Wilcoxon signed-rank test (non-parametric) otherwise.

## 4.1 Personas

	mean		significance	
	easy	diff.	test	p value
age	47.8	43.9	t	0.43
weight	83.2	99.4	t	0.013*
height	166	173	t	0.011*
energy	1826	2536	wilcox	$\ll$ 0.001*
nutrition goal	0.68	1.32	wilcox	$\ll$ 0.037*

**Table 1: Summary statistics for personas.**

Table 1 shows the differences between the easy and difficult groups of personas (the physical attributes of the sam-

pled people as well as their nutritional goals and activity levels). A person's physical dimensions play a crucial role in the difficulty of generating plans; in general the larger someone is, the more nutrition they need to consume. This is difficult for the planning algorithm as it is constrained by the need to fulfil all nutritional requirements from just 3 meals and does not have the ability to adjust the portion sizes. Similarly, the person's goal also has an effect on food intake and, in turn, the difficulty of planning. Of the 25 difficult personas, 16 wish to gain weight whereas only 4 personas from the easy group had this goal.

Linear models shows that all of the features in table 1 are able to significantly explain the variance when predicting the score. Gender is also a significant feature with male personas more likely to belong to the difficult group.

## 4.2 Profiles

	median		significance	
	easy	diff.	test	p value
# rated recipes	71	14	wilcox	$\ll$ 0.001*
# rated ingr.	378	56	wilcox	$\ll$ 0.001*
ingr. diversity	0.711	0.393	wilcox	$\ll$ 0.001*
# in top 50	7	2	wilcox	0.003*

**Table 2: Summary statistics for profiles.**

Table 2 displays statistics for the easy and difficult groups of profiles which describe the user's tastes based on the recipes they rated. Recall that recipes returned by the RS to be used for planning also include those already rated 4 or 5 by the user (indicating strong approval of the recipe). The number of rated recipes has an influence on the difficulty of planning, primarily because it increases the number of recipes the planner has at its disposal and also slightly increases the mean rating score and, therefore, the outcomes of the combined score. This is because the mean *predicted* rating (over all users) is 3.38 whereas the mean of all of the 4 and 5 star ratings is 4.43.

## 4.3 Varying the parameter $x$

$x$	# plans	# profiles	plans/comb	score
25	204,523	41	124	0.703
50	273,071	54	97	0.694
75	375,371	56	105	0.681
100	539,818	58	134	0.667

**Table 3: Change in performance when varying  $x$ .**

To make plans, the algorithm takes the recipes each user had liked previously as well as the top  $x$  suggested recipes from the recommender system. In the above analyses we set  $x$  to 100, however it is interesting to consider what effect varying this parameter has on the number of plans that can be generated as well as the "quality" of those plans. As  $x$  increases, the number of plans that can be created increases, albeit not linearly. As  $x$  increases so does the number of profiles for which plans can be generated with a particularly large jump between 25 and 50. The average number of plans generated per user profile-persona combination does not monotonically increase or decrease as either extreme of

the parameters values have the highest outputs for this metric. In the case of  $x = 25$  this is likely because there are relatively few profiles for which it can generate plans at all that the ones it can generate for are quite easy. In terms of average plan score, this decreases almost linearly with the value of  $x$ . This is because as  $x$  is increased the average score of the ratings decreases, causing a corresponding decrease in the scores.

## 5. DISCUSSION AND LIMITATIONS

We have shown that even with a simple algorithm and a recipe collection of modest size, it is possible to create balanced meal plans for many users by combining recipes that they like or which a recommender predicts they will like. For other users the task is more challenging. The results of our analyses suggests ways that we can make changes to our approach in order to increase the number of situations where plan generation is possible.

A simple change that we can make is to encourage users to rate more recipes to increase the pool of potential combinations for the system to work with. Our analyses show that the users who rated the most recipes tended to have more possible plans and even small increases helped. In particular if users can rate recipes with diverse ingredients this can be beneficial as diverse taste profiles were also correlated with the number of the possible plans. It may be helpful to employ a recommender algorithm specifically designed to recommend diverse recommendations as increasing diversity is a popular research topic in RS e.g. [12]. However, this may come at a cost to ratings.

Another specific bottleneck in our approach was the lack of breakfasts. It may be of benefit therefore to use separate lists of top  $x$  breakfasts and top  $y$  main meals as input to the algorithm, although, again, this may come at the cost of lower ratings and hurt quality scores overall.

We have also shown that it is often difficult to generate high quality plans conforming to nutritional guidelines when the user requires large numbers of calories per day. For example when the user wishes to gain weight or they are very tall. It also seems more challenging to create plans for males than for females. These results suggest that for difficult cases what is needed are planning methods with more flexibility, such as the ability to alter portion sizes, the number of meals per day or to consider side dishes. Greater flexibility will require more complicated algorithms to deal with a larger search space and efficiency issues. The methods and metrics we provide here, however, offer a framework and a baseline to study such algorithms in the future.

## 6. CONCLUSIONS AND FUTURE WORK

Previous work has proposed food RS as a valuable tool for assisting people achieve a healthy diet. In this paper we have provided a first attempt at incorporating health and nutrition into the food recommendation problem. We have shown that it is possible to combine recommended recipes into balanced meal plans according to guidelines from nutritional agencies. By analysing situations where it was difficult or impossible to generate plans we have gained clues as to how to design better planning algorithms in the future.

There are several other aspects of this work that open up future research questions. We wish to build on our contributions here by experimenting with more complicated plan-

ning algorithms using the methodology and metrics we have developed here to understand the costs and benefits of different approaches. We are, moreover, looking for nutritional experts to collaborate with to evaluate and refine the approach from this perspective. Such an expert would also be able to evaluate how truly healthy the meal plans are, as just because they fit the various nutritional constraints does not necessarily mean they are balanced. The number of recommendations used has an impact on both the number of plans that can be generated as well as the quality of those plans. In future work it would be useful to investigate in more detail which value or range of values are optimal and how this is influenced by factors such as the total number of ratings available to the system and the quality of the recommender system used. Finally, we want to better understand how to combine recipes and what makes a good plan from the user's perspective. To achieve this we plan to perform a further user study where complete daily plans are rated and not just recipes, in addition to analysing how the planner, already embedded in our food portal, is actually used.

## 7. REFERENCES

- [1] S. Berkovsky, J. Freyne, and G. Smith. Personalized network updates: increasing social interactions and contributions in social networks. In *UMAP*, pages 1–13. Springer, 2012.
- [2] E. Brunner, D. Stallone, M. Juneja, S. Bingham, and M. Marmot. Dietary assessment in whitehall ii: comparison of 7 d diet diary and food-frequency questionnaire and validity against biomarkers. *British J. of Nutrition*, 86(3):405–14, 2001.
- [3] D. Ornish et al. Can lifestyle changes reverse coronary heart disease?: The lifestyle heart trial. *The Lancet*, 336(8708):129 – 133, 1990.
- [4] J. Freyne and S. Berkovsky. Intelligent food planning: personalized recipe recommendation. In *IUI*, pages 321–324, 2010.
- [5] J. F. Guthrie, B. M. Derby, and A. S. Levy. *America's Eating Habits: Changes and Consequences Agriculture Information Bulletin No. (AIB750)*, pages 243–280. US Dept. for Agriculture, 1999.
- [6] M. Harvey, B. Ludwig, and D. Elswailer. You are what you eat: Learning user tastes for rating prediction. In *SPIRE*, pages 153–164, 2013.
- [7] T. Hinrichs. Strategies for adaptation and recovery in a design problem solver. In *Workshop on Case-Based Reasoning*, 1989.
- [8] J.C. et al. Platt. Learning a gaussian process prior for automatically generating music playlists. In *NIPS*, pages 1425–1432, 2001.
- [9] F. Ricci, B. Arslan, N. Mirzadeh, and A. Venturini. Itr: a case-based travel advisory system. In *Advances in Case-Based Reasoning*, pages 613–627. Springer, 2002.
- [10] A.M. Roza and H.M. Shizgal. The harris benedict equation reevaluated: resting energy requirements and the body cell mass. *Am Jour. of clinical nutrition*, 40(1):168–182, 1984.
- [11] R. Schaller, M. Harvey, and D. Elswailer. Entertainment on the go: Finding things to do and see while visiting distributed events. In *IiX*, pages 90–99. ACM, 2012.
- [12] S. Vargas and P. Castells. Rank and relevance in novelty and diversity metrics for recommender systems. In *ACM RecSys*, pages 109–116, 2011.