# Context-Aware Food Recommendation System

Rui Maia, and Joao C. Ferreira

*Abstract*—**Recommendation systems are commonly used in websites with large datasets, frequently used in e-commerce or multimedia streaming services. These systems effectively help users in the task of finding items of their interest, while also being helpful from the perspective of the service or product provider. However, successful applications to other domains are less common, and the number of personalized food recommendation systems is surprisingly small although this particular domain could benefit significantly from recommendation knowledge. This work proposes a context-aware food recommendation system for well-being care applications, using mobile devices, beacons, medical records and a recommender engine. Users passing near a food place receives food recommendation based on available offers order by appropriate foods for everyone's health at the table in real time. We also use a new robust recipe recommendation method based on matrix factorization and feature engineering, both supported by contextual information and statistical aggregation of information from users and items. The results got from the application of this method to three heterogeneous datasets of recipe's user ratings, showed that gains are achieved regarding recommendation performance independently of the dataset size, the items textual properties or even the rating values distribution.**

*Index Terms*—**Context-Aware, Food, Recommender System, Collaboration**

## I. Introduction

As digital media reaches the domain of gastronomy, e.g., with printed restaurant menus starting to be replaced by digital menus [16], it is likely to see an increased demand for recommendation systems. Moreover, food recommendation systems can also play an important role in persuading users to alter their lifestyle towards healthier nutrition options [6, 26]. It is interesting to notice that the food domain has several unique characteristics that motivate the usage of modern recommendation approaches. The problem deals well with numerous contextual features, given that the consumption of food items is naturally associated to heterogeneous information about ingredients, their chemical composition, nutritional aspects, cooking methods, ingredient combination effects, as well as monetary costs, availability, cultural, social and even environmental factors.

In this work, the user preference on food items is modelled as an aggregation of different features to generate recommendations using context information about user location. Collaborative filtering techniques, e.g., based on matrix factorisation, have been developed for many years, and many variants have also been proposed for specific settings. A popular trend in recent years relates to extending matrix factorisation approaches to incorporate auxiliary information via feature engineering [3]. Factorization Machines (FMs) in particular, originally proposed by Steffen Rendle [19], can combine the high-prediction accuracy of factorisation models with the flexibility of feature engineering, nowadays being commonly employed in the development of context-based recommendation systems [22]. The input data for FMs is described using real-valued features, exactly like in standard supervised machine-learning approaches such as linear regression or discriminative classification.

However, the internal model behind this approach considers factorised interactions between variables, and thus, it shares with other factorisation models the high prediction quality in sparse settings, such as those that are typically found in recommendation systems. It has been shown that FMs can mimic most factorisation models used in recommendation systems just by feature engineering [20], and it has also been shown that this approach outperforms other algorithms proposed for the task of developing context-based recommendation systems [22].

Moreover, to the extent of our knowledge, no methodology has been proposed to deal with very different recipe recommendation contexts. The cultural diversity or region [1] deeply impacts the user's preferences, the recipes or ingredients association, therefore the recommendation process. This work proposes a method involving the usage of factorisation machines and feature engineering over heterogeneous food recommendation problems. The results are reported on experiments using very different features that aggregate information from rating events (e.g., per user/item standard deviation) and specific features of this domain, like ingredients or cuisine type.

## II. Related Work

Good food habits are a critical issue in our daily well-being, and it is fundamental to preventing Diseases, control certain conditions, like diabetes, high blood pressure, among others. Many applications were developed to create healthy eating behaviours, like behaviours have been proposed and studied [27,28], where eating recommendations play an important role [29] and are a hot topic among scientific community with the goal of giving users' advice towards healthy eating behaviour following personalised recommendations [30]. These current food recommendations are mainly based on: 1) surveys, where users' answers questions that are used to create a meal plan, but they do not generate meal recommendations that cater to each person's fine-grained food preferences; and 2) food journaling, still suffer from major limitations, as cool start problem, and it is hard to maintain. To handle this, several models of preference elicitation have been proposed in

Manuscript received March 2018; revised April, 2018.

Rui Maia is PhD Student of Instituto Superior Técnico, Lisbon, Portugal (rui.maia@tecnico.ulisboa.pt)

Joao C. Ferreira is with Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, Lisbon, Portugal and Centro Algoritmi (e-mail: joao.carlos.ferreira@iscte.pt).

recent years, based on decision trees to poll users in a structured fashion [31,32]. Also, there is the possibility of eliciting item ratings directly from the user [33]. Traditional food and recipe RS learn from users' online activity [34], ratings, past recipe choices [35] and browsing history [29,36]. Some of these approaches are: 1) social navigation system that performs recommendation based on previous choices made by the user [37]; 2) the work [29] propose to similarity measurement for a recommendation from crowd card-sorting and make recommendations based on the self-reported meals; and 3) work [38] generate healthy meal plans based on user's ratings towards a set of recipes and the nutritional requirements calculated for the person.

Previous recommenders also seek to incorporate users' food consumption histories recorded by the food logging and journaling systems (e.g., taking food images [39] or writing down ingredients and meta-information [29]). Most of these existing systems for a food recommendation, share a common limitation, the cold-start problem. To overcome this limitation, major commercial applications, such as Zipongo [30] and Shopwell [40] adopt onboarding surveys to more quickly elicit users' coarse-grained food preferences, collecting data related about their nutrient intake, lifestyle, habits, and food preferences for their meals recommendations. Most recent research in recommendation systems has focused on context-aware methods that analyze only the user-item interaction, modelling recommendation as a rating prediction problem, using collaborative information regarding how particular users rate certain items (e.g., on a scale from 1 to 5 stars) to build a model capable of predicting future rating behaviours. In this context, matrix factorization approaches have become very popular, as they usually outperform traditional k-nearest neighbour methods for collaborative filtering [23, 14]. In contrast to the huge literature on standard recommendation systems, there is only little research on context-aware recommendation systems, i.e. systems that take into account additional data about the user (e.g., demographic information, age, profession, gender, etc.), about the item (e.g., attributes like product genres or descriptive keywords), and/or about the situation in which the rating events happen (e.g. the current location, the time, who is nearby, etc.). Recently, FMs have been proposed as an effective approach for contextual recommendation domains [19, 22, 20]. Research works in the area of recommendation systems have focused mostly on applications related to the prediction of movie ratings, using standard datasets such as MovieLens or the one made available in the context of the Netflix challenge [2]. Few previous studies have specifically addressed food recommendation, although this particular domain offers many interesting challenges. For instance, the ingredients of a meal are one of the components which impact a user's opinion and behaviour. Others include cooking methods, monetary costs, and availability, nutritional breakdown, ingredient combination effects, cuisine type, dietary considerations, as well as cultural, social and environmental factors.

### A. *Food Recommendation*

Initial efforts in the area of food recommendation have resulted in multiple systems, such as Chef [10] or Julia [12], which hugely relied on domain knowledge and case-based reasoning in their recommendation processes. More recently, Freyne et al. compared different strategies based on the traditional approaches from the recommendation systems domain, which are commonly used on e-commerce websites: user-based collaborative filtering based on the ratings from the k-nearest neighbouring users, a content-based approach that uses information about ingredients, and also hybrid recommendation methods [7]. In their experiments, the authors found that a reasonable accuracy can be achieved through content-based strategies that use a simple break down and construction heuristic, to relate recipes and ingredients. They measure a significant accuracy improvement through the use of content-based techniques over a simple collaborative filtering algorithm, although hybrid methods performed the best overall.

More recently, Freyne et al. also reported on large-scale studies in which they analyzed real user ratings on a set of recipes [8, 9] uncovering several interesting reasoning patterns in the rating data for this domain and suggesting ways to exploit this reasoning in the context of improving food recommendation systems. Their results have shown, for instance, stable user biases towards certain features of recipes (e.g., cuisine type or key ingredient). Harvey et al. identified some important contextual factors which can influence the choice of rating [11], specifically several ingredient factors that have a particularly strong predictive power for ratings. They also found that health factors such as fat and calorie content were less predictive, although were relevant for a subset of the users who give particular attention to nutrition. Still, the approaches proposed by Harvey et al. correspond essentially to extensions of traditional content-based on collaborative filtering approaches.

Moreover, on what regards to large-scale studies over the rating of food items, Yong-Yeol et al. analyzed patterns that can explain the food pairing mechanisms existing in different world regions [1]. The theory states that depending on the culture or world region, some cuisines (or regions) tend to pair ingredients that share many flavor components, while other cuisines tend do avoid component sharing.

Their work underlines the relevance of systematic analysis of user preferences and culinary practices to understand food preparation and user choices better, unraveling the reasons behind the user choices. The authors also observed limitations and structural differences within the recipe preparation information of the datasets, mainly regarding ingredients and instructions. They cite Kinouchi et al. [13] who reports that, on the analyzed cookbooks, the average number of ingredients per recipe does not follow a pattern. Teng et al. have reported on a food recommendation method using ingredient networks together with a discriminative machine learning algorithm, to predict recipe ratings [24]. Specifically, the authors constructed two types of networks to capture the relationships between ingredients, namely: 1) a complement network that captures which ingredients tend to co-occur frequently (e.g., savory, or sweet ingredients); and 2) a substitute network, derived from user-generated suggestions for modifications that capture functionally equivalent ingredients together with user preferences. The method tries to get healthier variants of a recipe.

Structural information from the co-occurrence and substitution networks is used to build features for a discriminative machine method, which corresponds to a pairwise ranking model based on a discriminative classification that, given a pair of recipes, determines which one has the higher average rating. Their experiments indicate that recipe ratings can be well predicted with features derived from combinations of ingredient networks and nutrition information. Some previous studies have also used collaborative filtering approaches based on matrix factorisation for food recommendation [5, 17, 11], often using heuristics to integrate heterogeneous content information about recipes, such as ingredients, dietary facts, cuisine type or occasion, as Forbes and Zhu.

This work proposes to leverage the recently proposed Factorization Machines [19], [22] with statistically inferred features which can be applied multiple heterogeneous rating distribution datasets. The combined use of direct features - as recipes – with derived ones - as standard deviation - applied to factorisation machines, overcome the use of specific context features such as ingredients or dietary type, supporting a sound and theoretically solid approach in all the dataset configurations in the context of food recommendation. Fig. 1 illustrates the proposed method which will be detailed in the Experimental Setup section.

## III. Matrix Factorization using Factorization Machines for the Recipe Recommendation

Besides users and items, in the context-aware recommendation, there is more information about the rating events, such as the location of the user at the time of his rating, his mood, etc... In particular, the food recommendation domain can naturally be approached through context-aware methods, given that the consumption of food items is associated with different types of heterogeneous information.

FMs model all nested interactions up to order d between the p input variables in a case x, using factorised interaction parameters, but do not distinguish between context or general features. Second order factorisation machines are especially appealing because higher-order interactions can be hard to estimate in sparse settings. A factorisation machine model of order d = 2 is defined as follows:

$$\hat{y}(x) = w_0 + \sum_{j=1}^{p} w_j \times x_j + \sum_{j=1}^{p} \sum_{j'=j+1}^{p} x_j \times x_{j'} \times \sum_{f=1}^{k} v_{j,f} \times v_{j',f} \qquad (1)$$

In the formula, k is the dimensionality of the factorization and the model parameters corresponding to $\{w_0, w_1, \ldots w_p, v_{1,1}, \ldots, v_{p,k}\}$ are such that $w_0 \in R$, $w \in R^p$, and $V \in R^{pxk}$.

The first part of the model is similar to standard linear regression, and it contains the unary interactions of each input variable $x_j$ with the target y(x). The second part of the two nested sums contains all pairwise interactions of input variables, that is, $x_j \times x_{j'}$. The important difference to a standard polynomial regression is that the effect of the interaction is not modeled by an independent parameter $w_{j,j}$ but with a factorized parametrization $w_{j,j} \approx \langle v_j, v_{j'} \rangle = \Sigma^{k}_{f=1} v_{j,f}\, v_{j',f}$ which corresponds to the assumption that the effect of pairwise interactions has a low rank. This allows factorization machines to estimate reliable parameters even in highly sparse data (e.g., sparse data regarding user ratings to items, as typically found on recommendation applications), where standard models often fail.

In a general recommendation problem, the data can be described by a design matrix $X \in R^{nxp}$, where the *ith* row $x_i \in R^p$ of $X$ describes one case (i.e., one rating event) with $p$ real-valued variables, and where $y_i$ is the prediction target of the *ith* case. Fig. 2 represents how, in this work, was

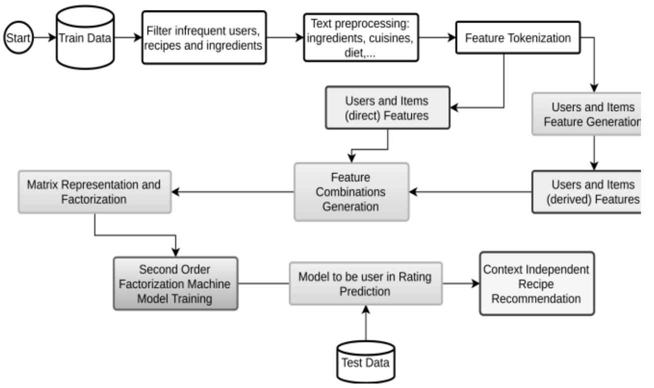


Fig. 1. Representing a recommendation problem with real valued feature

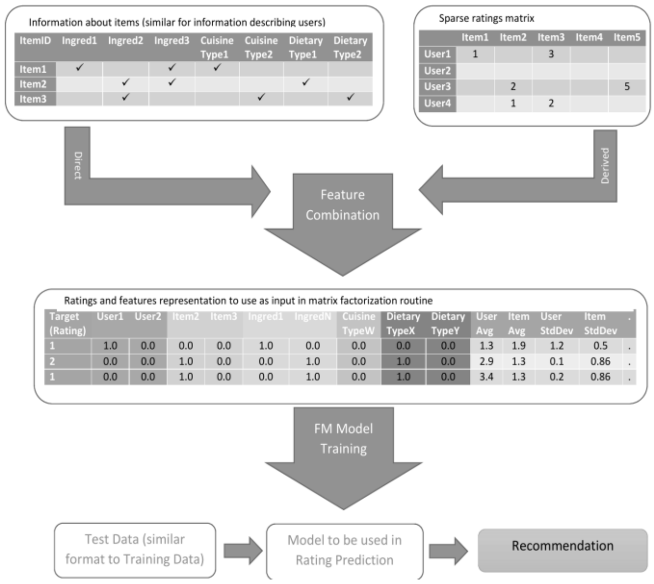


Fig. 2. Representing a recommendation problem with real valued feature vectors

modelled the food recommendation problem through feature engineering, considering multiple context-specific features, but also general statistical features, like the user rating average or standard deviation.

Many algorithms have been proposed for estimating the parameters of FMs, including methods based on Stochastic Gradient Descent (SGD), Alternating Least-Squares (ALS), and Markov Chain Monte Carlo (MCMC) inference. A freely available C++ implementation1 was used in the experiments, which uses MCMC inference as described in a recent publication [21]. By sampling through random draws, a probability distribution is approximated to the target *y* distribution, and the quality of the generated model improves as a function of the number of draws or training iterations. One of the main advantages of MCMC over ALS or SGD is that there is no need to define regularisation parameters or a learning rate. In this study, a feature vector

ISBN: 978-988-14048-1-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

x is chosen to capture important information in the food recommendation domain independently of the dataset characteristics.

## IV. Context-Aware with Beacons

Context-aware is based on user current position that is acquired from mobile device GPS or Bluetooth interaction with a beacon (indoor location approach). This is a new approach for indoor Location device with Bluetooth Low Energy (BLE) technology. This device that broadcasts a Bluetooth signal in a limited and configured range/area. This can be helpful to allow some software applications that can interpret the received signal as Location of the user inside of a building. Since most e the capability to work together giving retailers and other businesses all the information they need to create improved, personalised consumer experiences and we apply this to the creation of context-aware food recommendations. This concept is being applied in every kind of buildings like hospitals, airports, shopping, and museums. Fig. 3, illustrates the working principle, where beacon configured with a Bluetooth transmission radius of 5 meters alerts the user in that range. So, in a proximity of a restaurant food recommendation can be performed based on pre-defined criteria and available user profile (set of predefined preference with possible advice from doctors taking into account medical recording information, not considered in this research work. All information and algorithms run on the cloud server, and the mobile device is used to show the information. Beacon DB is used to correlated beacon with restaurant or food provider. User advice can also be correlated with geographic position.

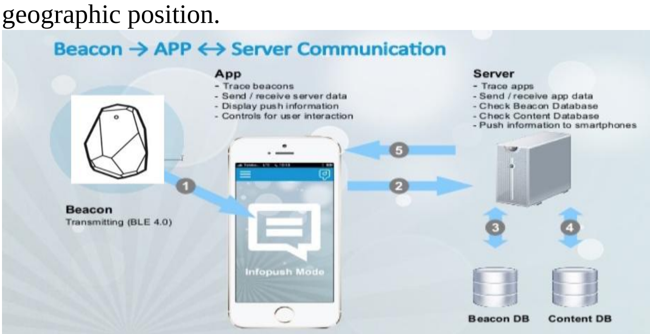

Fig. 3 – Proposed system for context-aware information

## V. Experimental Setup

The experiments reported in this paper relied on three datasets collected from online sources of food recipes to test proposed approach. The developed system was based on this factorisation approach. We use the following datasets: 1) dataset [http://mslab.csie.ntu.edu.tw/˜tim/recipe.zip] previously made available by Lin et al. [17], collected from a large online [http://www.food.com] recipe sharing community;2) dataset [http://students.mimuw.edu.pl /˜tk290810/kochbar_data] made available by Tomasz et al. [15], collected from another large online [http://www.kochbar.de] recipe sharing community; 3) Was constructed by crawling Epicurious website [ttp://www.epicurious.com], another platform where users can share and vote on recipes from other users with basis on their preferences.

In all three cases, the recipes are associated with detailed information on ingredients, preparation directions, cuisine types and dietary categories, all added by users. Notice that recipe is sharing communities, like the ones using the mentioned websites, have no pre-defined ingredient sets, nor predefined categories or standardised units for describing ingredient quantities. Information regarding these variables will, therefore, be noisy and sparse.

The datasets have been filtered to remove recipes rated less than four times, as well as the users with less than 4 rating events. This filtering allows us to maintain most of the dataset information, keeping users and items with relevant presence 2. In the case of the Food dataset, some additional data cleaning was also performed, for instance, typos correction in the names of ingredients and to merge them with similar constituents but with different modifiers (e.g., big red potato and small white potato are both changed to potato). In the case of the Food dataset, the ingredients used no more than three times were filtered out. In what regards the Epicurious dataset, no additional data cleaning or filtering was performed, since the registered ingredients are restricted to the main recipe ingredients, not all the recipe ingredients. It was not possible to do any text preprocessing on the Kochbar dataset since it only has numerical identifiers for the ingredients, dietary groups or cuisine types. By experimenting with different modelling choices for the feature vector *x*, *I* quantify the impact that different types of contextual information can have on the recommendation process. Additionally, the aggregation of user and item rating information was represented as features, specifically considering the user rating average, the standard deviation in user ratings, the item rating average and the item rating standard deviation.

The experiments start by using data about two categorical variables only, namely information regarding users and food items, as commonly done in collaborative recommendation systems based on matrix factorisation. A set of users $U$ and a set of food items $R$ are assumed, being the feature vectors $x$ Ɛ $R^{|U|+|R|}$ composed of binary indicator variables for the specific user and food item involved in a rating event.

Information about ingredients, cuisine types and dietary groups, was also represented as binary indicator variables, capturing the fact that a given food item is composed of a set of ingredients $|I|$, associated with a set of cuisine types $|C|$ and dietary groups $|G|$. These variables were tested independently with a feature vector $x$ Ɛ $R^{|U|+|R|+|J|}$, $x$ Ɛ $R^{|U|+|R|+|G|}$, $x$ Ɛ $R^{|U|+|R|+|G|}$, and also on different combinations of these variables, with the corresponding changes in the feature vector dimension.

The user and item rating averages and standard deviations were captured using real values as model features. These variables were tested independently with a feature vector $x$ Ɛ $R^{|U|+|R|+1}$ and in different feature vector combinations.

Assuming the intuition that ingredients might have valuable information for the recommendation process, the experiments were separated into two groups: one not including the ingredients as a feature, and the other including them. In summary, the features, categorical or real-valued, were tested with the information regarding users and food items in different combinations. The feature

ISBN: 978-988-14048-1-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

vector could be the most basic $x \in R^{|U|+|R|}$, or a combination of all the available features, being, in this case, $x \in R^{|U|+|R|+|G|+|J+4|}$. Due to a large number of possible combinations, the different characteristics of the datasets, and considering the preprocessing tasks and processing time involved, the efforts were dedicated to a small selection of combinations that could lead to relevant conclusions.

Table I compares the possible contextual feature combinations with the number tested combinations. Table II, presents, a statistical characterisation of the three datasets used in the experiments, showing that they are much sparser than the typical benchmarks used for evaluating recommendation algorithms (e.g., the Netflix dataset has a rating matrix sparsity of 1.18% [2]). The average number of ratings, per users or items, it is also much smaller in the experimented datasets. Such sparsity limits the effectiveness of a conventional collaborative filtering model and justifies the need of adding contextual information into the recommendation process

. TABLE I
NUMBER OF CATEGORICAL OR REAL-VALUED FEATURES USED IN THE TESTS AND THE NUMBER OF POSSIBLE COMBINATIONS

| Categorical/Real Valued features used | Possible combinations | Tests made |
|---|---|---|
| 1 | 7 | 5 |
| 2 | 21 | 4 |
| 3 | 35 | 4 |
| 4 | 35 | 2 |
| 5 | 21 | 1 |
| 6 | 7 | 0 |
| 7 | 1 | 0 |
| Total | 127 | 16 |

Fig. 4 shows the distribution of the number of user and recipe ratings on a logarithmic scale. It is also shown the distribution of values for the rating events on a histogram. Note that the Epicurious dataset has a Likert-scale from 1-4, while datasets Food and Kochbar have a 1-5 scale. These graphs show that there are significant differences between the three datasets on the rating behaviours. The left graph of Fig. 4 shows the differences of the total number of users and user ratings. It underlines that the Kochbar dataset has a greater number of users, which are more active, in comparison with the Epicurious or Food datasets (as can be seen in Table III - Average number of ratings per user).

The right graph from Fig. 4 shows that the users of the Epicurious dataset (collected for this work) have the stronger rating variation, while user ratings in the Food and Kochbar datasets are much more constant. In this last case, i.e., the Kochbar dataset, there is a clear tendency for users in giving the maximum rating value in their reviews. This confirms what Tomasz et al. stated about Kochbar rating behaviours, where more than 99% of the recipes are rated with 5 (the maximum value in the Likert-scale used in Kochbar). Although the used filter for users with less than 4 rating events, the tendency of a constant rating value prevails in the Kochbar dataset. Fig. 5 illustrates the distribution of ingredients, cuisine types and dietary groups, in the three datasets. The Kochbar dataset has a unique category, including the cuisine type and dietary group, and that's why it does not appear in the right graph (which is only for Dietary Groups). Besides the different number of ingredients, cuisine types and dietary groups in the datasets, it seems to exist a homogeneous pattern in their distribution. Taking into account the very significant difference in the size of the datasets, the tests were also run over a random subset of the Food and Kochbar datasets.

TABLE II
STATISTICAL CHARACTERIZATION OF THE DATASETS USED IN THE EXPERIMENTS

| Dataset Origin<br>Made available | Food<br>Lin et al. | Kochbar<br>Tomasz et al. | Epicurious<br>Maia |
|---|---|---|---|
| Number of users | 22598 | 8798 | 8203 |
| Number of food items | 59565 | 344341 | 7997 |
| Number of rating events | 692480 | 7658229 | 73468 |
| Number of ingredients | 4756 | 418593 | 352 |
| Number of cuisine types | 106 | 195 | 79 |
| Number of dietary groups | 28 | na | 21 |
| Rating value (Avg) | 4.72 | 4.99 | 3.38 |
| Ratings per user (Avg) | 30.64 | 870.45 | 8.96 |
| Ratings per item (Avg) | 11.63 | 22.24 | 9.19 |
| Ingredients per item (Avg) | 98.90 | 248.27 | 37.89 |
| Cuisine types per item (Avg) | 12.37 | 22.24 | 6.60 |
| Dietary groups per item (Avg) | 30.7 | na | 5.83 |
| Sparsity on the ratings matrix | 0.051% | 0.253% | 0.112% |

Trimming the datasets to approximately 73 thousand rating events (the size of Epicurious dataset), it was possible to analyze and compare the results of the experiments on different user communities with different rating behaviours.

The factorization models were tested with two different values on $k$ parameter of the Equation 1 ($k = 5$ and $k = 25$) which corresponds to the dimensionally of the factorised 2-way interactions in the factorisation machines. Each training session involved 100 iterations of the MCMC learning algorithm introduced in Section 3.

Regarding validation, the experiments followed a k-fold cross-validation method, using $k = 5$ folds, in which each dataset is first divided into $k$ equally sized, mutually exclusive subsets of randomly selected entries of rating events. Each event is contained in only one of the subsets. This method uses each event $k - 1$ times for training and one time for testing. The final results of each experiment are calculated on the average of the k iterations. To evaluate the performance of each model, this work followed Koren et al. [14] and the common literature, using the root-mean-squared error (RMSE) and mean absolute- error (MAE) metrics to get the average of the results over k-fold experiments.

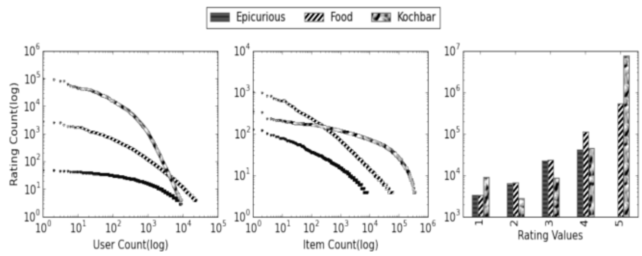

Fig. 4. Users, Items, and Rating Log Distributions

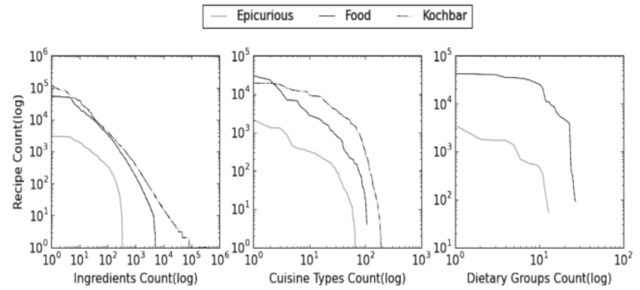

Fig. 5. Ingredients, Cuisine Types, and Dietary Groups Distributions

## VI. Results

Table 3 presents the RMSE and MAE results obtained for each of the modelling choices on the three datasets that were presented in the previous section. It also presents the results when using the user average rating score as a prediction value. These results are used as a baseline against the results obtained with the factorization machines. Table IV presents the results using approximately 73.000 random selected rating events for the Food and Kochbar datasets. Epicurious maintains the original setup of 73.468 rating events. The results on Table III show that by increasing the usage of contextual attributes, gains regarding food recommendation performance can be achieved. The best results, represented in bold font, correspond to RMSE values of 0.5777 for the Food dataset using either k = 5 or k = 25, RMSE of 0.1638 in the Kochbar dataset, and RMSE of 0.7798 using k = 5 in the Epicurious dataset.In all cases the use of recipe context attributes enhances the recommendation results, lowering the RMSE and MAE when comparing to the basic test using only information about users and items, represented in the first line of the first group of tests using Factorization Machines on Table III. The Kochbar dataset is clearly a different case, due to the high prevalence of items with maximum rating. It is clear that the use of statistical features (user and item rating averages and standard deviations) improves the results in all datasets. In the Food and Epicurious datasets the use of the item and user rating averages also outperforms all other features. It was not possible to run all the tests against the Kochbar dataset regarding cuisine types and dietary groups since these properties are merged in the dataset. The use of recipe ingredients shows unclear impact, having different results on recommendation performance over the three datasets, possibly depending on the relevance of the registered ingredients, or due to the existence of more misspelt entries in some datasets. As stated in Section 2, the Epicurious dataset includes only the ingredients considered to be the most relevant by the users, while all the recipe's ingredients are found in Food and Kochbar datasets.

As shown in Fig. 6, it seems clear that the model performs best when considering a large quantity of the users, even if they have only few recipe ratings and larger error dispersion. The model tends to be less precise for users with more rating events, and this can be due to the change of the rating profile of the users along the time. Moreover, this cannot be confirmed as there is no time-stamp associated with the rating events and as most of the events are associated with occasional users (x < 10). Looking at Fig. 7, there is a small reduction of the error distribution in the group of the users who rated more than 300 recipes. The

TABLE III - Results for the complete datasets.

| | Food | | | | Kochbar | | | | Epicurious | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | | MAE | | RMSE | | MAE | | RMSE | | MAE | |
| Average User Rating | 0.5947 | | 0.3741 | | 0.1513 | | 0.0204 | | 0.8245 | | 0.6151 | |
| Linerarly Regression Ensemble | 0.5813 | | - | | - | | - | | - | | - | |
| Content Driven Matrix Factorization | 0.6013 | | - | | - | | - | | - | | - | |
| | *k*=5 | | *k*=25 | | *k*=5 | | *k*=25 | | *k*=5 | | *k*=25 | |
| Factorization Machines | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| Information on users and items | 0.5961 | 0.4638 | 0.5964 | 0.4645 | 0.1818 | 0.0553 | 0.1747 | 0.0590 | 0.8514 | 0.7193 | 0.8500 | 0.7179 |
| and cuisine types | 0.6173 | 0.4551 | 0.6174 | 0.4570 | 0.1820 | 0.0376 | 0.1726 | 0.0420 | 0.8034 | 0.6632 | 0.8025 | 0.6622 |
| and dietary groups | 0.6089 | 0.4556 | 0.6086 | 0.4562 | 0.1713 | 0.0416 | 0.1697 | 0.0456 | 0.7860 | 0.6346 | 0.7862 | 0.6352 |
| and user average | 0.5812 | 0.3754 | 0.5804 | 0.3768 | 0.1659 | 0.0444 | 0.1759 | 0.0488 | 0.8071 | 0.6102 | 0.8068 | 0.6105 |
| and user standard deviation | 0.5823 | 0.3701 | 0.5825 | 0.3724 | 0.1674 | 0.0416 | 0.1760 | 0.0482 | 0.8001 | 0.6027 | 0.7997 | 0.6030 |
| and item average | 0.5966 | 0.3859 | 0.5952 | 0.3863 | **0.1638** | 0.0426 | 0.1759 | 0.0484 | 0.8268 | 0.6360 | 0.8257 | 0.6358 |
| and item standard deviation | 0.5988 | 0.3752 | 0.5978 | 0.3766 | 0.1661 | 0.0437 | 0.1784 | 0.0475 | 0.8179 | 0.6256 | 0.8164 | 0.6253 |
| and user average and user standard deviation | 0.5944 | 0.3662 | 0.5952 | 0.3702 | 0.1702 | 0.0389 | 0.1790 | 0.0470 | 0.8123 | 0.6000 | 0.8115 | 0.6017 |
| and item average and user average | **0.5777** | 0.3727 | **0.5777** | 0.3765 | 0.1652 | 0.0411 | 0.1775 | 0.0491 | **0.7798** | 0.5954 | 0.7800 | 0.5957 |
| and cuisine types and item average and user average | 0.5862 | 0.3868 | 0.5864 | 0.3892 | 0.1784 | 0.0370 | 0.1746 | 0.0442 | 0.7829 | 0.5991 | 0.7828 | 0.6003 |
| and dietary groups and item average and user average | 0.5822 | 0.3821 | 0.5845 | 0.3859 | 0.1685 | 0.0404 | 0.1744 | 0.0485 | 0.7823 | 0.6001 | 0.7812 | 0.6002 |
| Information on users, items and ingredients | 0.6074 | 0.4222 | 0.6100 | 0.4299 | 0.1720 | 0.0526 | 0.2988 | 0.1066 | 0.7892 | 0.6254 | 0.7943 | 0.6319 |
| and cuisine types | 0.6055 | 0.4191 | 0.6076 | 0.4259 | 0.1689 | 0.0456 | 0.1959 | 0.0647 | 0.7925 | 0.6280 | 0.7983 | 0.6353 |
| and cuisine types and dietary groups | 0.6044 | 0.4184 | 0.6063 | 0.4248 | 0.1665 | 0.0431 | 0.1919 | 0.0603 | 0.7951 | 0.6294 | 0.8005 | 0.6370 |
| and item average and user average | 0.5919 | 0.3992 | 0.5934 | 0.3999 | 0.1682 | 0.0476 | 0.6143 | 0.4209 | 0.7830 | 0.6011 | 0.7911 | 0.6095 |
| and cuisine types and item average and user average | 0.5933 | 0.4017 | 0.5951 | 0.4017 | 0.1670 | 0.0393 | 0.3458 | 0.1624 | 0.7839 | 0.6034 | 0.7927 | 0.6115 |
| and cuisine types and dietary groups and item average and user average | 0.5947 | 0.4016 | 0.5949 | 0.4006 | 0.1663 | 0.0423 | 0.1867 | 0.0602 | 0.7826 | 0.6031 | 0.7933 | 0.6128 |

TABLE IV - Results on approximately 73 thousand rating events.

| | Food | | | | Kochbar | | | | Epicurious | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | | MAE | | RMSE | | MAE | | RMSE | | MAE | |
| Average User Rating | 0.5947 | | 0.3741 | | 0.1369 | | 0.0137 | | 0.8245 | | 0.6151 | |
| Linerarly Regression Ensemble | 0.5813 | | - | | - | | - | | - | | - | |
| Content Driven Matrix Factorization | 0.6013 | | - | | - | | - | | - | | - | |
| | *k*=5 | | *k*=25 | | *k*=5 | | *k*=25 | | *k*=5 | | *k*=25 | |
| Factorization Machines | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| Information on users and items | 0.8422 | 0.7694 | 0.8557 | 0.7824 | 0.5997 | 0.5861 | 0.6074 | 0.5944 | 0.8514 | 0.7193 | 0.8500 | 0.7179 |
| and cuisine types | 0.6686 | 0.5339 | 0.6659 | 0.5327 | 0.1321 | 0.0256 | **0.1293** | 0.0285 | 0.8034 | 0.6632 | 0.8025 | 0.6622 |
| and dietary groups | 0.6732 | 0.5555 | 0.6733 | 0.5564 | 0.1622 | 0.0849 | 0.1615 | 0.0884 | 0.7860 | 0.6346 | 0.7862 | 0.6352 |
| and user average | 0.5828 | 0.3637 | 0.5825 | 0.3661 | 0.1351 | 0.0270 | 0.1344 | 0.0327 | 0.8071 | 0.6102 | 0.8068 | 0.6105 |
| and user standard deviation | 0.5871 | 0.3637 | 0.5865 | 0.3662 | 0.1356 | 0.0245 | 0.1358 | 0.0304 | 0.8001 | 0.6027 | 0.7997 | 0.6030 |
| and item average | 0.6084 | 0.3943 | 0.6080 | 0.3956 | 0.1398 | 0.0280 | 0.1396 | 0.0353 | 0.8268 | 0.6360 | 0.8257 | 0.6358 |
| and item standard deviation | 0.6071 | 0.3797 | 0.6065 | 0.3811 | 0.1447 | 0.0230 | 0.1454 | 0.0277 | 0.8179 | 0.6256 | 0.8164 | 0.6253 |
| and user average and user standard deviation | 0.5995 | 0.3587 | 0.5996 | 0.3664 | 0.1403 | 0.0231 | 0.1420 | 0.0332 | 0.8123 | 0.6000 | 0.8115 | 0.6017 |
| and item average and user average | 0.5793 | 0.3624 | **0.5759** | 0.3725 | 0.1355 | 0.0267 | 0.1384 | 0.0393 | **0.7798** | 0.5954 | 0.7800 | 0.5957 |
| and cuisine types and item average and user average | 0.5835 | 0.3902 | 0.5860 | 0.3955 | 0.1341 | 0.0255 | 0.1385 | 0.0397 | 0.7829 | 0.5991 | 0.7828 | 0.6003 |
| and dietary groups and item average and user average | 0.5795 | 0.3835 | 0.5826 | 0.3888 | 0.1444 | 0.0363 | 0.1460 | 0.0486 | 0.7823 | 0.6001 | 0.7812 | 0.6002 |
| Information on users, items and ingredients | 0.6204 | 0.4537 | 0.6204 | 0.4602 | 0.2228 | 0.1311 | 0.2393 | 0.1402 | 0.7892 | 0.6254 | 0.7943 | 0.6319 |
| and cuisine types | 0.6158 | 0.4472 | 0.6140 | 0.4527 | 0.1540 | 0.0486 | 0.2399 | 0.0949 | 0.7925 | 0.6280 | 0.7983 | 0.6353 |
| and cuisine types and dietary groups | 0.6154 | 0.4469 | 0.6135 | 0.4518 | 0.1555 | 0.0480 | 0.2814 | 0.1053 | 0.7951 | 0.6294 | 0.8005 | 0.6370 |
| and item average and user average | 0.5891 | 0.4070 | 0.5967 | 0.4131 | 0.1713 | 0.0688 | 0.2115 | 0.0960 | 0.7830 | 0.6011 | 0.7911 | 0.6095 |
| and cuisine types and item average and user average | 0.5894 | 0.4080 | 0.5966 | 0.4161 | 0.1538 | 0.0451 | 0.2461 | 0.1027 | 0.7839 | 0.6034 | 0.7927 | 0.6115 |
| and cuisine types and dietary groups and item average and user average | 0.5896 | 0.4078 | 0.5970 | 0.4154 | 0.1550 | 0.0458 | 0.3600 | 0.1860 | 0.7826 | 0.6031 | 0.7933 | 0.6128 |

ISBN: 978-988-14048-1-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

fact that this group of users is present in a large number of rating events might impact the overall dataset results. Regarding recipes, the model seems to be independent of the recipe rating frequency. Moreover, it is possible to observe that most of the rating events are associated with recipes rated less than 30 times which means that there is an even distribution of the dataset. As previously found for Epicurious, in the Food dataset, the model seems to be independent of information on dietary groups, cuisine types, and ingredients.

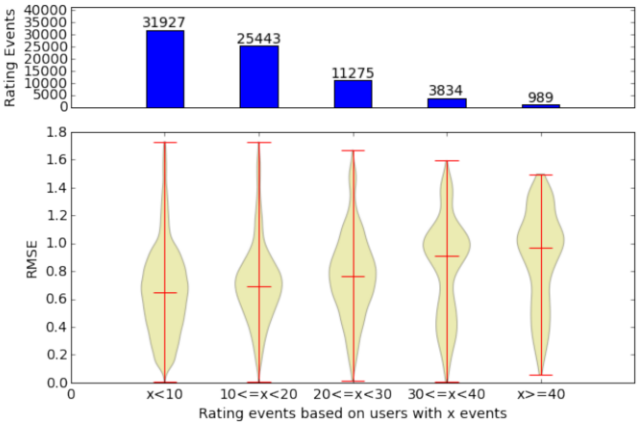


Fig. 6. Epicurious dataset users grouped by number of rating events

It is also possible to observe that there are a few dietary groups and cuisine types with a strong relationship to the rating events, which do not seem to impact on the group results. It is relevant to remember that all the three datasets were extracted from Web sites, and that Food is the largest of the analysed datasets, with almost 8 million rating events and an extremely high average rating, with more than 95% of the events the maximum value in scale, 5.

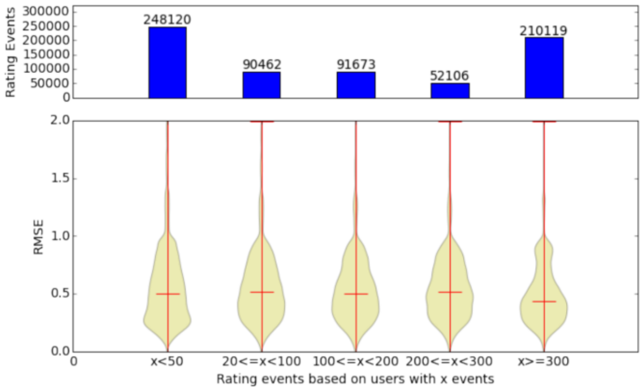


Fig. 7. Food dataset users grouped by number of rating events

In Fig. 8 it is possible to observe that a large number of users have a high rating frequency (with more than 7 thousand rating events each). The model seems to obtain larger error when considering users with few rating events (less than 100). Regarding dietary groups and cuisine types, which in this dataset are merged on a single set, it is clear that most of the rating events are associated with a few numbers of dietary groups. The set of dietary groups and cuisine types having more than 100 thousand associated events have a larger presence in all the observed rating events. Regarding Epicurious, the model is independent of information about cuisine types, dietary groups, ingredients or even recipes. Moreover, the small difference between the results obtained from complete Kochbar and Food datasets, and trimmed ones (see Fig. 5), show that there is a weak association between the dimension of the dataset (i.e., the number of rating events) and the system efficiency.

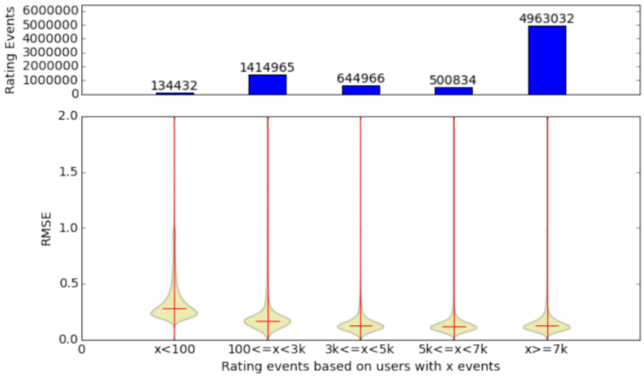


Fig. 8. Kochbar dataset users grouped by number of rating events

## VII. Conclusions and Future Work

The problem of making personalised recommendations in the food domain has several unique characteristics that motivate the usage of modern approaches that usefully handle a large number of contextual features. Given that the consumption of food items is naturally associated with heterogeneous information about ingredients, their chemical composition, nutritional aspects, cooking method. It is possible to associate as well as monetary costs, availability and cultural or social, even environmental factors, the task of finding the correct description of each recipe is crucial for an effective and robust recommendation system methodology. This work proposes a new methodology that explores the most basic recipe features and statistically derived features from leveraging an efficient matrix factorization based robust recommendation system which can be applied to very different food and recipe datasets. We report on experiments that modelled the user preferences on food items as an aggregation of features, latter leveraging factorisation machines to capture latent low-rank interactions between the features for making accurate recommendations. We assess which are the most relevant and independent dataset features that can be used as accuracy improving factors in the food domain context.

Experiments with three extremely different datasets collected from online recipe sharing communities attest the effectiveness of this approach showing that it is possible to achieve recommendation performance gains by considering multiple contextual attributes. Moreover, the use of item and user average rating and standard deviation as model features outperformed almost all other feature combinations, independently of the very different characteristics of the considered datasets.This study concluded that all food datasets are strongly affected by noisy information introduced by users, which limits the relevance of the dataset size. Due to this negative impact, we proved that a small part (near 70.000 entries) of a large dataset could be used with virtually the highest possible efficiency while lowering computational time and resources. The integration of features derived from textual information associated to recipe descriptions can be interesting as it might be less error-prone. It might be possible, for instance, to consider mining text from user reviews to extract relevant food terminology, or use sentiment analysis methods to infer food preferences [25], including this information in factorisation machines enhancements [4, 18].

As future work, the introduction of other feature types can also experiment, namely some that are not directly available in the datasets that were used in the experiments here described (e.g., features computed from the weight of the ingredients in each recipe, or environmental features such as temperature at the time, external medical information can be added). In large datasets is made clear the need of algorithms that can identify and despise the erroneous or noisy contextual information, while keeping the dataset coherent.

## References


[1] Y.Y. Ahn, S. E. Ahnert, J. P. Bagrow, and A.L. Barabasi. Flavour network and the principles of food pairing. Scientific reports,1, 2011.

[2] R. M. Bell and Y. Koren. Lessons from the Netflix prize challenge. SIGKDD Explorations Newsletter, 9(2), 2007.

[3] T. Chen, W. Zhang, Q. Lu, K. Chen, Z. Zheng, and Y. Yu. SVDFeature: A toolkit for feature-based collaborative filtering. Journal of Machine Learning Research, 13(1), 2012.

[4] [4] C. Cheng, F. Xia, T. Zhang, I. King, and M. R. Lyu. Gradient boosting factorization machines. In Proceedings of the ACM Conference on Recommender Systems, 2014.

[5] P. Forbes and M. Zhu. Content-boosted matrix factorization for recommender systems: Experiments with recipe recommendation. In Proc. of the ACM Conference on Recommender Systems, 2011.

[6] J. Freyne and S. Berkovsky. Intelligent food planning: Personalized recipe recommendation. In Proceedings of the International Conference on Intelligent User Interfaces, 2010.

[7] J. Freyne and S. Berkovsky. Recommending food: Reasoning on recipes and ingredients. In Proc. of the International Conference on User Modeling, Adaptation, and Personalization, 2010.

[8] J. Freyne, S. Berkovsky, and G. Smith. Recipe recommendation: Accuracy and reasoning. In Proceedings of the International Conference on User Modeling, Adaption, and Personalization, 2011.

[9] J. Freyne, S. Berkovsky, and G. Smith. Rating bias and preference acquisition. ACM Transactions on Interactive Intelligent Systems, 3(3), 2013.

[10] K. J. Hammondn. CHEF: A model of case-based planning. In Proc. of the National Conference on Artificial Intelligence, 1986.

[11] M. Harvey, B. Ludwig, and D. Elsweiler. You are what you eat: Learning user tastes for rating prediction. In Proceedings of the International Symposium on String Processing and Information Retrieval, 2013.

[12] T. R. Hinrichs. Strategies for adaptation and recovery in a design problem solver. In Proceedings of Workshop on Case-Based Reasoning, 1989.

[13] O. Kinouchi, R. W. Diez-Garcia, A. J. Holanda, P. Zambianchi, and A. C. Roque. The non-equilibrium nature of culinary evolution. New Journal of Physics, 10(7):073020, 2008.

[14] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. Computer, 42(8), 2009.

[15] T. Kusmierczyk, C. Trattner, and K. Nørv°ag. Temporality in online food recipe consumption and production. In Proceedings of the International Conference on World Wide Web, 2015.

[16] P. Lessel, M. B¨ohmer, A. Kr¨oner, and A. Kr¨uger. User requirements and design guidelines for digital restaurant menus. In Proceedings of the Nordic Conference on Human-Computer Interaction: Making Sense Through Design, 2012.

[17] C.-J. Lin, T.-T. Kuo, and S.-D. Lin. A content-based matrix factorization model for recipe recommendation. In Proc. of the Pacific-Asia Conf. on Knowledge Discovery and Data Mining, 2014.

[18] B. Loni, A. Said, M. Larson, and A. Hanjalic. 'free lunch' enhancement for collaborative filtering with factorization machines. In Proc. of the ACM Conference on Recommender Systems, 2014.

[19] S. Rendle. Factorization machines. In Proceedings of the IEEE International Conference on Data Mining, 2010.

[20] S. Rendle. Factorization machines with libFM. ACM Transactions on Intelligent System Technology, 3(3), 2012.

[21] S. Rendle. Learning recommender systems with adaptive regularization. In Proceedings of the ACM International Conference on Web Search and DataMining, 2012.

[22] S. Rendle, Z. Gantner, C. Freudenthaler, and L. Schmidt-Thieme. Fast context-aware recommendations with factorization machines. In Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval, 2011.

[23] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In Proceedings of the International Conference on World Wide Web, 2001.

[24] C.-Y. Teng, Y.-R. Lin, and L. A. Adamic. Recipe recommendation using ingredient networks. In Proceedings of the Annual ACM Web Science Conference,2012.

[25] M. Trevisiol, L. Chiarandini, and R. Baeza-Yates. Buon appetito: Recommending personalized menus. In Proceedings of the ACM Conference on Hypertext and Social Media, 2014.

[26] Y. van Pinxteren, G. Geleijnse, and P. Kamsteeg. Deriving a recipe similarity measure for recommending healthful meals. In Proceedings of the International Conference on Intelligent User Interfaces, 2011.

[27] Kerry Shih-Ping Chang, Catalina M. Danis, and Robert G. Farrell. 2014. Lunch Line: Using public displays and mobile devices to encourage healthy eating in an organization. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 823–834.

[28] Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, et al. 2008. Activity sensing in the wild: aA field trial of ubifit garden. In Proceedings of the Special Interest Group on Computer-Human Interaction Conference on Human Factors in Computing Systems. ACM, 1797–1806.

[29] Youri van Pinxteren, Gijs Geleijnse, and Paul Kamsteeg. 2011. Deriving a recipe similarity measure for recommending healthful meals. In Proceedings of the 16th ACM International Conference on Intelligent User Interfaces. 105–114.

[30] zipongo. 2015. Personalizing Food Recommendations with Data Science.

[31] Mahashweta Das, Gianmarco De Francisci Morales, Aristides Gionis, and Ingmar Weber. 2013. Learning to question: leveraging user preferences for shopping advice. In Proceedings of the Conference Knowledge Discovery in Databases (KDD'13).

[32] Mingxuan Sun, Fuxin Li, Joonseok Lee, Ke Zhou, Guy Lebanon, and Hongyuan Zha. 2013. Learning multiple-question decision trees for cold-start recommendation. In Proceedings of the ACM International Conferene on Web Search and Data Mining (WSDM'13)

[33] Shuo Chang, F. Maxwell Harper, and Loren Terveen. 2015. Using groups of items for preference elicitation in recommender systems. In Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW'15).

[34] David Elsweiler and Morgan Harvey. 2015. Towards automatic meal plan recommendations for balanced nutrition. In Proceedings of the 9th ACM Conference on Recommender Systems. ACM, 313– 316

[35] Gijs Geleijnse, Peggy Nachtigall, Pim van Kaam, and Lucienne Wijgergangs. 2011. A personalized ¨ recipe advice system to promote healthful choices. In Proceedings of the 16th International Conference on Intelligent User Interfaces. ACM, 437–438.

[36] Mayumi Ueda, Syungo Asanuma, Yusuke Miyawaki, and Shinsuke Nakajima. 2014. Recipe recommendation method by considering the users preference and ingredient quantity of target recipe. In Proceedings of the International MultiConference of Engineers and Computer Scientists, Vol. 1

[37] Martin Svensson, Kristina Ho¨ok, and Rickard C ¨ oster. 2005. Designing and evaluating kalas: A social ¨ navigation system for food recipes. ACM Trans. Comput.-Hum. Interact. 12, 3 (2005), 374–400

[38] David Elsweiler and Morgan Harvey. 2015. Towards automatic meal plan recommendations for balanced nutrition. In Proceedings of the 9th ACM Conference on Recommender Systems. ACM, 313– 316.

[39] Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. 2015. Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, 3207–3216.

[40] Shopwell. 2016. Retrieved from http://www.shopwell.com