Food Recommendation in a Worksite Canteen

Vincenza Carchiolo\textsuperscript{1}\textsuperscript{a}, Marco Grassia\textsuperscript{2}\textsuperscript{b}, Alessandro Longheu\textsuperscript{2}\textsuperscript{c}, Michele Malgeri\textsuperscript{2}\textsuperscript{d} and Giuseppe Mangioni\textsuperscript{2}\textsuperscript{e}

\textsuperscript{1}Dip. Matematica e Informatica, Universit\`a di Catania, Italy
\textsuperscript{2}Dip. Ingegneria Elettrica Elettronica Informatica, University of Catania, Italy

Keywords: Data Analysis, Recommendation System, Machine Learning.

Abstract: Recommendation systems tackle with information overload to assist people in finding their best choice according to their preferences and past behaviour. This occurred in many contexts, including the food sector where culinary inspiration, sales increase or healthy advice motivate the adoption of such a system. In this paper we propose a canteen food recommendation system for workers operating at an innovation hub including more than 20 companies. The system leverages a 30 months data set of past choices, and adopts a content based and a collaborative filtering approach for canteen users, suggesting them with dishes chosen by other similar users. First results for frequent as well as occasional canteen visitors are encouraging to validate the proposed approach.

1 INTRODUCTION

Information overload in decision-making processes exploits recommendation systems (Bobadilla et al., 2013) (Mohamed et al., 2019) both as a tool to help users find products based on their preferences and past choices, and to assist companies in making targeted sales based on customers’ interests; one of the areas where such systems are widely used is the recommendation of meals (Min et al., 2019).

The food sector is important for various reasons (Nations, 2015)(Torreggiani et al., 2018), as improving culinary inspiration, increasing sales, and promoting health improvement. Food recommendation has been studied by many researchers (Jiang et al., 2019), (Merler et al., 2016), (Iwendi et al., 2020); to create an effective system it is crucial to understand how people make food-related choices considering factors as cultural, social, economic and even organic.

In this article we present an application that assists a worker in booking his/her meal at a canteen operating in an innovation hub located in Italy (Carchiolo et al., 2020), (Carchiolo et al., 2021). The canteen menu manager is a native Italian speaker that inserts in the menu local names for dishes, which are usually non self-explanatory; this results in a certain discomfort for some workers when they have to choose. For this reason, a key feature of the app is the engine of a recommendation system that will suggest each worker with the pot he/she will most likely to prefer. We collected data concerning dishes consumed at a worksite during about 2.5 years by more than 200 workers of 22 different companies. Starting from this dataset, we set up a recommendation system aiming to suggest users with dishes chosen by other similar users. We adopted a content based approach and a collaborative filtering approach to manage both frequent and occasional canteen visitors; first results are encouraging to validate our proposal.

In section 2 related works on food recommendation systems are considered, whereas in section 3 the case study is introduced and results from dataset are discussed, and finally in section 4 some conclusions and proposals for future activities are presented.

2 FOOD RECOMMENDATION

Food Recommendation (FR) system are diversified according to their data model, the number of sources data extracted from and users interaction support; they involve disparate factors (e.g. cultural preferences or medical prescriptions), in predicting what

---

\textsuperscript{a} https://orcid.org/0000-0002-1671-840X
\textsuperscript{b} https://orcid.org/0000-0001-5841-6058
\textsuperscript{c} https://orcid.org/0000-0002-9898-8808
\textsuperscript{d} https://orcid.org/0000-0002-9279-3129
\textsuperscript{e} https://orcid.org/0000-0001-6910-0112
people prefer to eat (Elsweiler et al., 2017).

FR system can be used in several scenarios:
- Seasonality of ingredients, products and recipes determined by analyzing historical trends, to help in menu planning or to optimize purchase / inventory decisions for restaurants;
- food Diets, using nutrition data to guide customer preferences according to levels of (or sensitivity to) salt, sugar, fat, etc.;
- increase in efficiency and cost savings in product or recipe formulations by selecting cheaper equivalents of current ingredients;
- customer Trends analysis, gathering which recipes and products users ultimately act or buy, and the time/season/event those recipes or products are associated to;
- recipe generator for food waste reduction. Starting from available food in the fridge or pantry at a given moment and using a recipes dataset, a recurring neural network can generate a complete recipe that includes the instructions, the category and the title (Zhang et al., 2021);
- meal recommendations based on user preferences provided in the form of meal ratings, exploiting the tastes of the individual user or the collaborative approach, which is based on the relationships between users or elements (meals).

2.1 Classification of FR Systems

Based on the existing literature, we classify FR systems into three categories, i.e content-based, collaborative filtering and hybrid approaches.

Content-based approaches aim to tailor recommendations to the user’s individual tastes. This is achieved building a profile for each user starting from the attributes used to describe the characteristics of the meals. (Freyne et al., 2011) introduce recommendations by extracting from the recipes of the various meals the individual ingredients and the positive evaluations of the users about those ingredients. This means that if an ingredient is present in recipes that a user flagged as liked, other recipes containing the same ingredient will also be suggested. Subsequent work has pursued this approach taking into account both positive and negative ingredient evaluations, rating recipes ingredients (Harvey and Elsweiler, 2017).

Other content-based approaches are specifically suitable to FR systems. For example, since food decisions are often visually guided, images associated with recipes are exploited (Zhang et al., 2020). In (Yang et al., 2017) authors show that basic approaches can be outperformed exploiting (Elsweiler et al., 2017) also show that low-level image functions automatically extracted, such as brightness, color, and sharpness can be useful for predicting a user’s food preference.

Collaborative, filter-based methods for meal recommendation systems have been proposed, in particular in item-based collaborative filtering recommendations are based on how meals resemble each other, whereas in user-based collaborative filtering recommendations are based on the preferences provided by the user and their emerging similarity. In (Freyne and Berkovsky, 2010) authors present an approach using Pearson’s correlation on the classification matrix, but with a worse ranking than the content-based approach described above. Harvey et al. (Harvey and Elsweiler, 2015) demonstrated that Singular Value Decomposition (SVD) outperformed both the content-based and collaborative filtering approaches. A Matrix Factorization (MF) approach for food recommendation systems that merges classification information and user-supplied labels to achieve significantly better prediction accuracy than baselines based on decomposition by content and standard matrices is presented (Ge et al., 2015).

Among Hybrid filtrations (Freyne and Berkovsky, 2010) combined a collaborative user-based filtering method with a content-based method. In their follow-up work involving user groups, the same authors used an hybrid approach to combine three different FR strategies into a single model based on the ratio between the number of rated articles respect to the total number. In (Harvey and Elsweiler, 2017) significant results combining an SVD approach with user and meal bias are achieved.

2.2 FR System Applications

Several commercial solutions for food recommendations are available with different strategies, as (1) simple user ratings or (2) user preferences extracted from their past choices, or (3) similarity between meals based on their ingredients/allergens, (4) similarity between users calculated by defining a profile for each user based on past preferences, or finally (5) exploiting nutritional needs in terms of calories. In the following, some applications are briefly discussed.

Yum-me (Yang et al., 2017) is a FR system based on personalized nutrition; it learns food preferences without relying on the user’s dietary history. The recommender learns users’ food preferences through a simple visual quiz-based interface and then attempts to generate meal recommendations that meet the user’s health goals, food restrictions, and personal appetite for food. It can be used by people who have
food restrictions, such as vegetarian, vegan, etc. and it is based on two steps:

1. Users answer a simple survey to specify their dietary restrictions and nutritional expectations. This is used to filter foods and create an initial set of candidates for recommendations.

2. Users then use an adaptive visual interface to express their food preferences through simple food comparisons. The preferences learned are used to further refine the proposed recommendations.

Another example of a food recommendation system is Caviar (HBS-Digital-Initiative, 2020), a commercial system that can be customized via an user selected optimization function, for example:

- **Recommended for You** algorithm uses a hybrid machine learning algorithm with content-based filtering and similarity among users to suggest restaurants

- **30 minute delivery** algorithm selects restaurants that can fulfill the delivery timeline

- **Appetizers under € 10** algorithm selects popular restaurants that have numerous appetizers below the set price limit.

FooDroid is a recommendation system developed at University of Zurich (Runo and Wattenhofer, 2011), created to provide a unified platform to manage booking of meals by students, having several canteens with different menus available. Users can browse the daily menus offered by the various canteens, select the meals according to their tastes and evaluate them on behalf of the colleagues who will later choose those based on the reviews made. This recommendation system focuses mainly on:

- Detailed evaluation of the menus offered. Indeed, since recommendation accuracy is mainly affected by the (few) users providing the most ratings, some menu choices could not reflect the real taste of an individual. Therefore, a more personalized recommendation system that operates on the preferences of individuals (or small groups) is desirable.

- Distance of the canteen from the customer’s location, as it is assumed that you can only spend a limited time for lunch. Therefore, authors try to obtain a trade off between the quality of the menus and the relative proximity to the customer.

FooDroid aims at obtaining a good combination between the quality of the menu offered by a canteen and users preferences, providing students with a means of evaluating the meals consumed and recommending menus based on these scores in the future.

Snap-n-eat (Zhang et al., 2015) is a mobile food recognition system based on a deep learning approach. The system can recognize the food and estimate the calories and nutritional content of the food automatically without any user intervention. To identify foods, the system allows the user to simply take a picture of the plate of food. The system detects the salient region, crops the image and subtracts the background accordingly. Basically, the app identifies which segments of the image contain food and then tries to figure out what type of food is present in each segment. In addition, the system determines the portion size which is then used to estimate the calories and nutritional content of the food on the plate. The system is capable of achieving automatic food detection and recognition in real life contexts containing bulky backgrounds. When multiple items of food appear in an image, the system is able to identify them and estimate their portions simultaneously.

### 3. THE CASE STUDY OF WORKSITE CANTEEN

The FR system we propose here advises employees on the choice of meal on a daily basis according to the available menu; a mobile application (not discussed here) allows to perform these and other actions, as news management, meeting rooms booking, workplace lighting control etc.

To set up the FR system, several information from user’s profile are considered together with those related to dishes, as its description, the list of ingredients and/or the contents of the nutritional levels; depending on data available, various machine learning techniques and models for similarity assessment between different data can be taken into account.

In particular, for what concern meals, we must consider that the combinations of dishes on the days are extremely varied and although the dishes are repeated several times during the analysis period, there is not a high degree of repetition in the menus.

In addition, we classify canteen users into Frequent Visitors (FV) or Occasional visitors (OV), where a newcomer is first classified as OV and promoted to a FV after a given number of canteen access; our FR system must address both categories.

#### 3.1 Dataset

The dataset includes data collected during about 2.5 years, from August 2017 to March 2020, concerning dishes consumed at the worksite, an innovation hub.
where workers of 22 different companies actually operate.

The study of the employees food intake we focus here is part of a more extensive project for life quality improvement of the employees at the worksite, increasing their productivity through a better sense of belonging (Carchiolo et al., 2019).

The number of meals consumed was about 50,000 (49,539), chosen by 264 employees. For a detailed view on the number of employees of the companies and the number of meals consumed, we refer the reader to Table 1. The common canteen is accessible by all the employees through a cross-platform proprietary App, which provides the canteen management with statistical information while retaining the order and access history of each employee. Employees are identified with a “employer code” that permits us to know the company where they work, their access to the canteen and the dishes consumed. Records with missing values are pruned away from the data set; this occurred for about 20% of meals. The dishes proposed are organized in 5 categories that mimic the Italian mean of food. The total number of different dishes offered by the canteen is 314.

We note that the multiplicity of proposals in the different categories is not homogeneous and two of them are way larger than others (see Table 2). Each dish is described by:
- the ingredients
- the type of cooking
- the features of dietary restrictions
- the heat intake and nutrients present

The ingredients are obtained on the most common Italian recipes. In the proposed dish, 143 ingredients are used and the number of main ingredients in each dish ranges from 1 to 11, as depicted in Fig 1. In fig. 2 the distribution of ingredients in dishes is reported. As shown, some ingredients are very common while others appear only occasionally in different dishes. The most widely used ingredient is extra virgin olive oil used for the preparation of 83% (in fig. 2 the red circle on the top) of the dishes; other 4 ingredients are very common in the dishes (they are highlighted with the second red circle) and are present in about 30% of the dishes. These ingredients reflect the tradition of Italian cuisine.

We have also identified 5 different types of cooking: fried, boiled, stew, grilled and baked. Each dish can have a combination of these types of cooking (even none of them). In fig. 3 the distribution of types of cooking in the dish is shown. In particular, the impact of the single cooking method and the multiplicity of that in a dish are illustrated.

As a restriction to the diet, those belonging to the “gluten free” and “vegetarian” categories are identified as the characterization of the dishes. In fig. 4 it is shown the distribution of veggie and gluten free dishes.
3.2 Food Recommendation System

As previously discussed, the average number of meals consumed in the canteen is 187, but only for about 80% of them the dataset contains details of the choice. An analysis of the dataset shows that the distribution of accesses is not uniform at all. Fig. 5 shows that many employees access the canteen infrequently (gray bar), while others access a much higher number of times. Fig. 5 shows in blue the variety in the choice of dishes by each person.

As previously discussed, the average number of meals consumed in the canteen is 187, but only for about 80% of them the dataset contains details of the choice. An analysis of the dataset shows that the distribution of accesses is not uniform at all. Fig. 5 shows that many employees access the canteen infrequently (gray bar), while others access a much higher number of times. Fig. 5 shows in blue the variety in the choice of dishes by each person.

3.2.1 Case 1: Frequent Visitors

To solve our problem either user–based or item–based collaborative filtering can be used. In the latter case, the selection of the item (namely, the dish) is made adopting a similarity based filtering among dishes. The goodness of the results of a recommendation system is affected by the model used for coding the de-
description of the dishes and therefore to the algorithm used to evaluate the similarity; several approaches can be used:

- encoding the string of dish description $D$, by applying the one-hot encoder, to generate the array $A_D$ describing the plate
- using a string $SI$ obtained by concatenating the list of ingredients included in the dish and applying the one-hot encoder to generate the array $A_{SI}$ describing the plate
- using the product between array $A_{SI}$ with the matrix containing the amount of each ingredient in the plate
- adding the description of the dish with the amount of nutrients.

Two dishes will be considered similar according to the first semantic model, which is the simplest to implement. The differences arising from the choice of the model are not analyzed in this paper.

In the case of a user–based collaborative filtering, a common solution is the use of some ranking. Our dataset does not include a ranking among elements, but it is built using the history about the past choices of dishes. In particular, a table is extracted from the dataset in which each row contains the information about the menu offered on a given day, the user who made the reservation and his choices. To solve our problem we made a test with both the classical approaches of collaborative filtering: memory–based and a model–based. For each model, we perform two experiments, one on data about FV’s and another on the complete data set (FV+OV).

In the former case, we used a KNN (K-Nearest Neighbours) with means model. This algorithm uses the similarities between users and/or items as weights to predict a rating for them. This similarity is computed by using the Pearson correlation or cosine similarity function.

The second approach used is based on the SVS Model Based Collaborative Filtering. The Singular-Value Decomposition, or SVD for short, is a matrix decomposition method for reducing a matrix to its constituent parts in order to make certain subsequent matrix calculations simpler. It provides another way to factorize a matrix, into singular vectors and singular values.

To find the rating $R$ that a user $U$ would give to an item $I$, the approach includes (1) finding users similar to $U$ who have rated the item $I$, and (2) calculating the rating $R$ based on the ratings of users found in the previous step; when using KNN with means to remove the bias we take into account the mean ratings of each user.

To evaluate the results we calculate the RMSE (Root Mean Squared Error), the most common metrics used to measure accuracy for continuous variables, and the Cross Validation (CV) of RSME with KNN and SVD model; for each one we made a test on the two above mentioned datasets: FV and (FV + OV). Table 3 summarizes the RSME and CV obtained.

### 3.2.2 Case 2: Occasional Visitors

In defining the recommendation systems, one of the problems described is relating to cold starts, in particular when a new user is introduced into the data set. Models defined above do not make a correct prediction on a user whose past choices are unknown, there-
fore we adopt an alternative solution using a popularity based recommender system that ranks products based on their popularity (i.e. the rating count) at a given moment. If a product is highly rated then it is most likely to be ranked higher and hence will be recommended. As it is based on the products popularity, the same set of products will be recommended for all the users (no personalizing).

Data used to train the model contains User ID and a Meal ID column. Each row represents an observed interaction between the user and the element. Pairs (user, meal) are stored with the model so that they can later be excluded from recommendations if desired. In addition, a Target column is added which represents the Ranking of the choice. This ranking is calculated as a function of the number of times the person consumed a meal with respect to the total number of days that the meal was available and therefore normalized in a range space.

Based on these assumptions, the initial data necessary to create the model are different as the booking history with the daily menu proposed and the choice made is no longer required, rather it is only necessary to obtain the sequence of choices made by each employee over time.

Results of such a model are at a very early stage and the further development step envisaged is the replacement of the simple model based on popularity with one that takes into account the group of people with whom the new employee has already had contacts because these with high probabilities will influence the choices.

Since this is a popularity based recommender model, we are getting similar result for all users. i.e. the model is recommending same products for all the users. Then we will choose meals recommended by most popular persons (fig 8 shows the results).

4 CONCLUSIONS AND FUTURE WORK

In this paper we described a food recommendation system for users of a workplace canteen. Exploiting a dataset of past choices and using a content based and a collaborative filtering approach for canteen users, the proposed system suggests users dishes chosen by other similar users. First results for frequent as well as occasional canteen visitors are encouraging though further steps are need to validate the proposed approach, in particular:

- to leverage dishes complete features, as ingredients and macro nutrients
- to combine time series and machine learning for prediction purposes, to help both canteen manager for a better menu planning as well as users for a better choice
- to test the proposed recommendation system in order to validate it through users feedback

As future extensions, it is planned to use an iot based approach (Loria et al., 2017) to collect other data, such as movements and contacts between different employees, useful for improving user profiling.
REFERENCES


