Investigating and Predicting Online Food Recipe Upload Behavior

Christoph Trattner[b,*], Tomasz Kusmierczyk[a], Kjetil Nørvåg[a]

[a] *Norwegian University of Science and Technology, Norway*
[b] *University of Bergen, Norway*

Abstract

Studying online food behavior has recently become an active field of research. While there is a growing body of studies that investigate, for example, how online recipes are consumed, in the form of views or ratings, little effort has been devoted yet to understand how they are created. In order to contribute to this lack of knowledge in the area, we present in this paper the results of a large-scale study of nearly 200k users posting over 400k recipes in the online recipe platform Kochbar.de. The main objective of this study is (i) to reveal to what extent recipe upload patterns can be explained by socio-demographic features and (ii) to what extent they can be predicted. To do so, we investigate the utility of several features such as user history, social connections of the users, temporal aspects as well as geographic embedding of the users. Statistical analysis confirms that recipe uploads can be explained by socio-demographic features. Extensive simulations show that among all features investigated, the social signal, in the form of friendship connections to other users, appears to be the strongest one and henceforth is the best to predict what type of recipe will be uploaded and what ingredients will be used in the future. The research conducted in this work contributes to a better understanding in online food behavior and is relevant for researchers working on online social information systems and engineers interested in predictive modeling and recommender systems.

Keywords: Online food recipes, socio-demographic feature analysis, predictive modeling, recommender systems

1. Introduction

Investigating user behavior online does not only help us in understanding and learning about what people want and need but also what should be changed. In the context of food and in particular nutrition research, a huge body of literature exists that tries to understand how we consume or produce food in our daily lives. These studies are typically performed offline in a survey-based format. Typical methods employed include telephone interviews and surveys (Burt et al., 1995; Alaimo et al., 1994). Though survey-based approaches have the advantage of asking distinctive questions they fail in many ways, such as for instance being inefficient in terms of eliciting data objective (Trattner et al., 2017b). They are also rather time consuming, expensive and typically capture only a small fraction of a population to draw conclusions.

*This is to indicate the corresponding author.
Email address: `christoph.trattner@uib.no` (Christoph Trattner)

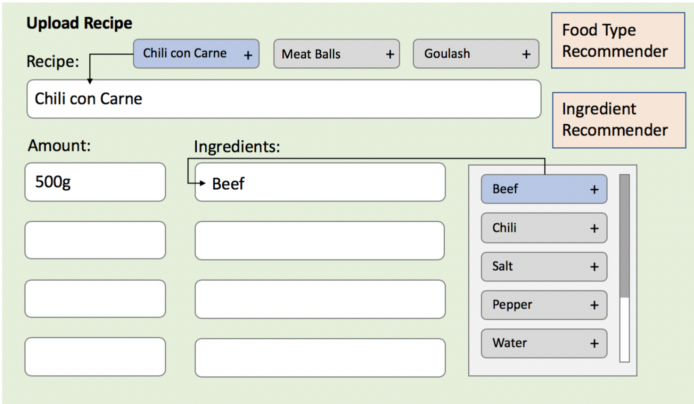


Figure 1: Example of an intelligent user interface that tries to support the user in the recipe upload process by recommending food type and corresponding ingredients to cook this meal.

To overcome this issue recent innovations in research follow a more pragmatic way by mining patterns of what users leave behind on the Web (De Choudhury et al., 2016). This is typically done by investigating online consumption patterns in the form of query logs or implicit and explicit user feedback such as ratings or views from users in online food community forums or social media platforms. The main advantage of such a method is that behavior can be computed without the direct involvement of the user. As such, it allows to learn user behavior or the behavior of a whole population fast, possibly more objective and on a large scale (Trattner et al., 2017b).

Objective. While current research in the online food area is mostly devoted to understand how people consume, i.e., how people search, view or rate online recipes, little research was been devoted yet to understand the producer side. As such, the main objective of this work is to understand and predict online recipe upload behavior. Or more specifically, the types of recipes created by online users and the ingredients to be used. Apart from our own preliminary work in this area (Kusmierczyk et al., 2016), to the best of our knowledge, no other work has been devoted yet to understand and predict online recipe uploading behavior.

Application. We think the study of online recipe upload behavior is useful, since it would not only enable us to understand what are potential food trends as set by a user or a whole community in the future, but would, for example, also help us in the design of novel food recommender algorithms or even intelligent user interfaces, that would support people in the food upload process. Currently, this process is a time-consuming task, and as shown in many other domains such as, e.g., social tagging Kowald et al. (2017), such a system would help the user not only to perform this kind of task more efficiently, but would potentially also increase the user experience or increase the quality of the content being created. Figure 1 highlights such as system, in the context of the online food community platform Kochbar.de. As shown, the system aims at predicting type of recipe (food type) the user is likely going to create as well as the ingredients used to create the recipe during recipe upload time. As such the interface provides a list of recommendations in terms for recipe type and ingredients used. Figure 2 presents a similar idea of an application that takes as input a user profile or a set of user profiles to provide as output a

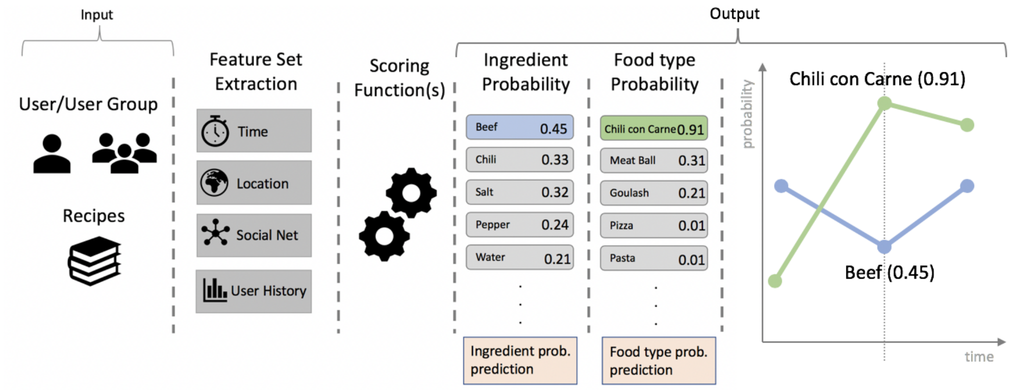


Figure 2: Example of system that tries to predict future recipe uploads given as input a user or user group and a set of recipes uploaded by this user or group.

ranked list of ingredients and food types probably uploaded in the future by the user or the user group. This kind of application may be useful for people interested in potential upcoming food trends as set by a popular online user or a whole community of users.

Research questions. To drive our research, the following high-level research questions were defined:

- RQ1. To what extent do signals and features exist that can explain recipe upload behavior?
- RQ2. To what extent are these features useful to predict online recipe uploads?

To answer these research question, we have taken a two-step approach. In the first step we analyze recipe upload behavior and are in particular interested in understanding to which extent signals and features exist that may explain recipe upload behavior. The second step involves employing these features to predict recipe uploads to reveal their magnitude of importance.

Paper structure. The paper is structured as follows: In Section 2 we highlight relevant related work in the field and highlight what this study adds to the existing literature. Section 3 introduces the dataset we have been using to conduct our research. This section is then followed by Section 4 which features our research methodology and explains amongst other things, data pre-processing, statistical analysis and predictive modeling. Section 5 presents the results coined with our first research question related to explaining online recipe upload behavior and Section 6 presents the results our second research question dealing with the predictability of food types and ingredients. Finally, Section 7 summarizes and discusses the outcomes of our study and Section 8 concludes the paper, with a summary of our main findings and future directions of our work.

2. Background

Studying and understanding online food behavior is a relatively new field of research and only a few previous studies exist. One of the first works in this context were the studies conducted by Ahn et al. (2011). In their work they mined and analyzed three different large-scale online food community platforms from Europe, the US and China to unveil patterns in flavor components that for instance makes Indian food different to the rest of the world.

A second noteworthy study by West et al. (2013) analyzed a large corpus of Microsoft Bing search-logs to discover the extent to which people search and access information in the online food community website Allrecipes.com. West et al. find that people's searches for food on the Web follow weekly and yearly trends and they also find significant differences in terms of search queries between states in the US. Food-related searches are correlated with instances of heart disease in certain regions of the US. Similar findings were reported by Said and Bellogín (2014), who analyzed a crawl of the online food community platform Allrecipes.com showing significant correlations between diabetes in certain regions of the US and the recipes consumed (rated) in these regions. Abbar et al. (2015) found aligning trends in the context of Twitter and Mejova et al. (2015) likewise in Instagram, while De Choudhury et al. (2016) also showed that Instagram data and bad food consumption patterns can be correlated to unhealthy regions in the US. A similar correlation was found in our own research recently in the context of online bookmarking behavior of people in Allrecipes.com (Trattner et al., 2017b).

Wagner et al. (2014) analyzed user preferences in the online food community Kochbar.de by analyzing access log data to the recipes available in this platform. In line with West et al. (2013), they find that the user online food preferences follow temporal trends and vary in certain regions in Germany. Temporal patterns were also discovered in our own study in the context of large-scale analysis of Kochbar.de (Kusmierczyk et al., 2015a,b). There we mined data in the form of ratings, and discovered that recipes follow a power-law distribution when analyzing their interestingness over time with certain recipe categories remaining interesting for longer than others. An interesting follow-up work by Laufer et al. (2015) shows to what extent 31 different language editions on Wikipedia mention food. Among other things they find that neighboring countries "tend to have more similar cultural practices than more distant countries".

Other relevant works are the recent studies of Rokicki et al. (2017) showing how editorial, temporal and social biases affect online food popularity and appreciation. Their recent follow-up work furthermore reveals how certain cues in recipes have an impact on how online recipes are perceived healthily (Rokicki et al., 2018). Also related are the studies of Mejova et al. (2017) who investigate food culture through Instagram posts and our own work (Trattner et al., 2017c) which reveals the extent to which online food preferences and hobbies correlate with food health standards and to what extent popularity of recipes can be predicted from content cues such as recipe title, image, nutrition or ingredients (Trattner et al., 2018).

Also worth mentioning in this context are the studies of Schneider et al. (2013) and our own work (Trattner et al., 2017a; Trattner and Elsweiler, 2017b) that investigated the healthiness of online recipes in food blog post and online food portals, such as Allrecipes.com, or the work of of Yu et al. (2013) and Liu et al. (2014) predicting recipe ratings from, for example, reviews using Machine Learning. Similar interesting work from the recommender research area, are the experiments of Teng et al. (2012), employing a number of features such as cooking style, user geo-location or user history to train a statistical model that is able to predict recipe choices. The most popular work in this strand of research though is work done by Berkovsky and Freyne (2010), Freyne and Berkovsky (2010), who were the first to study online food recipe consumption patterns and preferences, to invent several intelligent systems to recommend recipes to people.

More recent similar studies worth mention here are the works of Trevisiol et al. (2014), Ge et al. (2015), Elsweiler and Harvey (2015), Elsweiler et al. (2017), Rokicki et al. (2015) and the work of Yang et al. (2017) who studied intelligent meal planning and health-aware food (recipe) recommender systems. For an overview of the current state-of-the-art in food recommender systems and future challenges in the area we refer to Tran et al. (2017) and Trattner and Elsweiler (2017a).

Differences to previous research & contributions. In summary, the work reviewed above reveals that we know little about recipe upload behavior. The current literature mostly investigates how recipes are consumed in the form of ratings or views and how this correlates to health outcomes. Moreover, we know little about which types of features can be exploited to predict recipe uploads. The reviewed literature suggests that online food preferences are biased and that these features can be exploited to predict popularity of online recipes and their nutrition. However, to what extent these features can explain upload behavior or play a role in the predictability of recipe uploads is unknown. This is exactly the gap in the literate the present article is trying to fill by presenting an extensive study of features to explain and predict recipe upload behavior.

The main contribution of this work is the quantitative analysis of one of the largest online food community platforms on the Web comprising more than 200k users and 400k recipes uploaded, to show to what extent socio-demographic features impact recipe upload behavior (see Section 5). The second contribution of this work is an in-depth investigation to show to what extent these features contribute to the prediction of this behavior (see Section 5). A final contribution of this work is the observation that many of the observed patterns can also be linked to offline food behavior (see Summary & Discussions - Section 7).

3. Dataset

Our work relies on a dataset obtained from the German online food community website Kochbar.de[1]. According to Schroeder (2017) Kochbar.de attracts more than 6.6 million unique users every month and is ranked under the top-50 most popular websites in Germany and is the second most popular food website in Europe.

In order to obtain the data, we implemented a Web crawler that by following the link structure of Kochbar.de[2] retrieved all publicly available recipes and related user profiles, in addition to other user-generated contents. The basic statistics of the dataset are presented in Table 1. As shown, the dataset contains more than 400 thousand recipe profiles from the years 2008-2014 and corresponding meta-data, such as categories the recipes were published to, ingredients being used to cook the meal, preparation steps, publication time, recipe title, etc. Furthermore, the dataset comprises nearly 200k user profiles that contain ratings or comments applied to recipes, friendship relations to other users in the community platform of Kochbar.de, and the current home location of the users.

As highlighted in Table 1 the users established more than 195k friendship relations. 18 thousand users out of these were active recipe publishers who have uploaded at least one recipe. Around 19 thousand users also actively rated recipes, resulting in total 7 million user ratings. Interestingly, most of the ratings (99.1%) are 5-star ratings. As such very small divergence in rating behavior does not allow to study preferences sufficiently in detail. We therefore ignore the value of the ratings and just consider all user ratings to be '1' whenever a user applied a rating to a recipe. A value of '0' was chosen, whenever a recipe was found to be not rated by the user. Hence, ratings in Kochbar.de can be seen as bookmarks.

[1] `http://Kochbar.de`

[2] Crawl performed between 2014-11-22 and 2014-12-05 in alignment with the sites Terms of Services.

Table 1: Basic statistics of the Kochbar.de dataset.

Feature	Count
Num. users	199k
Num. ingredients	1,483
Num. recipes	406k
Num. users publishing recipes	18k
Num. food types	2,523
Num. ratings	7,795k
Num. users rating	19k
Num. categories	246

4. Methodology

To achieve our goal, which is to understand recipe uploading behavior and predict it, several steps were necessary. First, the data had to be processed and cleaned. Second, features need to be derived and statistical analysis needs to be performed to explain upload behavior, in our case, type of recipe being created and ingredients used. Third, predictive models need to be derived utilizing these features. Fourth, and finally, a set of experiments need to be conducted in order to evaluate the performance of the models in presence of state-of-the-art baselines.

4.1. Data pre-processing

Ingredient normalization. In our initial crawl of the dataset, ingredients were lists of arbitrary strings provided as free-form text by the users of Kochbar.de. As such, they are noisy and often contain typos or word deviates that do not match each other, although they contain the same meaning. To resolve that disambiguation issue, standard NLP-based pre-processing techniques were employed to normalize all ingredients in the dataset. To do so, we first filtered out stop-words, special characters, verbs as well as amounts and units, which were part of an ingredient string, i.e., '10 gram lovely butter' was transferred into just 'butter'. We also split conjunctions, such as, for example, 'salt and pepper', into 'salt' and 'pepper' (516 in total out of the 334k original ingredients, 311 out of these were related to salt and pepper). After that we matched ingredient names occurring less than 50 times with others starting from the most popular ones. We achieved this by computing a normalized form of Levenshtein's edit distance Yujian and Bo (2007). In this way, for example, the ingredient 'the glass of salted water' was replaced with the two ingredients 'water' and 'salt'.

We also replaced alternatives of two ingredients with the name of the more popular option, for example, 'butter or margarine' was replaced with 'butter'. Finally, we discarded all the ingredients appearing less than 10 times in our data set. The last step involved the manual unification of words (performed by the first and second author or this work), for example, 'penne' and 'spaghetti' both into 'pasta'. The procedure reduced the initial number of over 334 thousand rather noisy ingredients in our data set to 1,483 normalized ingredients which we use to, e.g., compute similarities between recipes/users or to predict these.

Table 2: Top-5 food types and ingredients in the Kochbar.de dataset.

Top-5 ingredients	Count	Top-5 food types	Count
Salt	294,112	Salad	16,313
Pepper	234,063	Muffins	4,731
Sugar	148,308	Dip	4,141
Eggs	127,680	Bread	4,123
Butter	125,787	Pizza	3,531

Food type identification. Furthermore, based on the recipe titles, we mined the types of recipes that we denote further in the paper as 'food types'. We identified those as recipe titles appearing at least 5 times in the whole dataset, i.e., if some title, for example, 'apple pie', appears in 5 or more recipes in the dataset, we assumed that it represents a common food type. Then, employing sub-string inclusion (Alam and Rahman, 2012), we matched all recipe titles to the extracted food types. In this way, for example, the recipe title 'Magic *Apple Pie* by Mrs Schultz' was identified as a special variant of the recipe type '*apple pie*'. Table 2 provides an overview of the top-5 most used ingredients and food types in the Kochbar.de dataset.

Identification of users' geographic origin. The final step involved the calculation of the geographical origin of the users in the Kochbar.de dataset. This is possible because when registering to the platform, the users are asked where they are currently located. However, the granularity and quality of this type of geographic information varies significantly. To resolve this issue, Google's Geocoding API was employed to unify the user 'locations' on three different levels. Specifically, we were able to map the users to a country, region or city.

4.2. Explaining & predicting recipe uploads

Recent literature in the context of online recipes suggests that food preferences in the form of ratings of views are influenced by features such as geographic embedding (Wagner et al., 2014) or time (Kusmierczyk et al., 2015b). To reveal whether we can find similar patterns in the context of recipe uploads and whether they can explain upload behavior, an exploratory data analysis (EDA) was performed. In detail, we report (i) density plots to show distributional trends in the data, (ii) scatter plots to reveal correlations (iii) and statistical tests, for example, in the form of Kolmogorow-Smirnov tests (K-S tests), t-tests or Chi^2 tests, to underpin our finding also statistically.

While the exploratory data analysis, as performed in the first part of this paper, reveals whether certain features are able to explain recipe upload behavior, they do not allow to draw conclusions about which of the features that may be the most useful ones to predict this behavior. To reveal this, we propose a set of scoring functions that model the data based on the investigated features to output the top-n most probable recipe types and ingredients the user will use to create (upload) a new recipe. To evaluate the proposed models, we run a series of offline experiments, as common on recommender systems research, to show the performance of the proposed models. As evaluation metric we employ Normalized Discounted Cumulative Gain (nDCG). Furthermore, we benchmark against well-known recommender algorithms, employing the open-source standard recommender systems library myMediaLite. More details on the exact evaluation protocol can be found in Section 6.

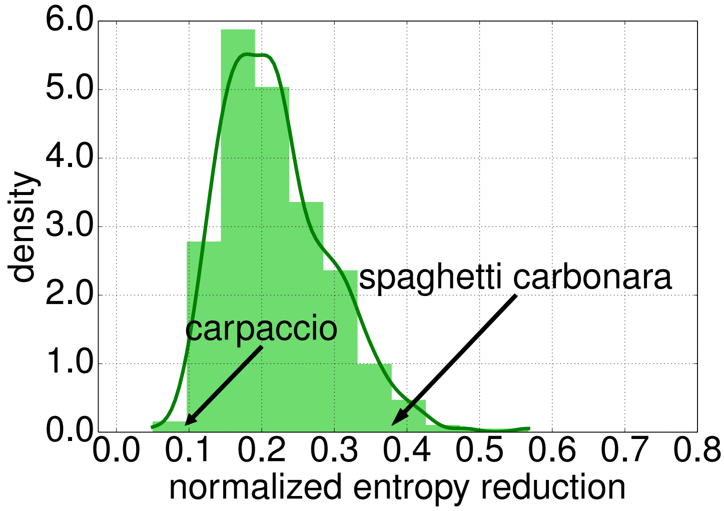


Figure 3: Distribution of reduction in ingredients entropy when food type is known.

5. Explaining Recipe Upload Behavior (RQ1)

This section begins with an investigation to reveal to what extent ingredients can be explained by a given food type (as we aim to predict ingredients from food types later). The second analysis in this section investigates to which extent category labels, as used by the users when uploading recipes, can explain future recipe uploads. Thereafter, an analysis is performed that shall explain the relationship between recipe uploads and preferences in the form of ratings. A further analysis is then coined with the question to what extent users are consistent in their upload behavior. The last three sub-section investigate, social, geographic and temporal biases in the context of recipe uploading behavior.

5.1. Investigating the relation between food types and ingredients

In our study we compare food types and ingredients. The intuition is that some ingredients should be more typical for particular food types than for the others, and therefore the knowledge about what food type was selected should be helpful in ingredients prediction. However, to observe the exact relation between a particular pair of food type and ingredient we need a more specific analysis.

Figure 3 provides a deeper insight into dependencies between food types and ingredients. As a measure for evaluating the discriminative power of food types we use a metric called **n**ormalized **e**ntropy **r**eduction which we define as a ratio between entropy gain when the food type is known and the original entropy value. Hence, $ner_{type}(X) = \frac{H(X)-H_{type}(X)}{H(X)}$. $H(X)$ is the entropy of the ingredients measured over all recipes and $H_{type}(X)$ is the entropy measured only over recipes of particular type. Low values of *ner* indicate that given a certain food type, this food type not allow to explain the ingredient distribution sufficiently. On the other hand, high

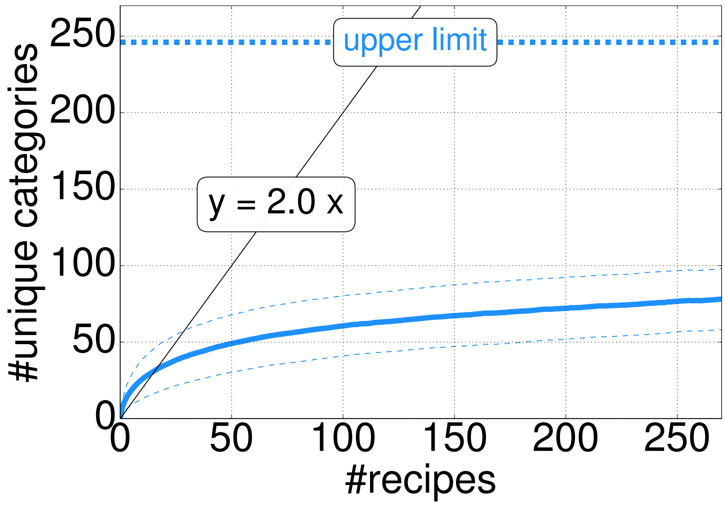


Figure 4: Number (mean and deviation) of unique categories used by users to label recipe uploads.

values mean that by providing a food type the prediction of ingredients is significantly simplified. For example, if for some *type* we observe $ner_{type}(X) = 1$, we know exactly what ingredients to choose. Two illustrative examples are 'spaghetti carbonara' that determines well both the set of ingredients and their frequencies distribution (entropy is reduced by almost 40%), while on the other hand, 'carpaccio' is very general food type (entropy is reduced by only 10%). In general, we find that the average reduction over all recipes is 22%. In other words that means that only 22% of the ingredients in a recipe can be explained by a given food type.

In summary, we find that food types and ingredients exhibit some correlations indicating that by knowing what type of food the user is going to create might be helpful in predicting recipe ingredients. However, some food types discriminate much better between ingredients than the others. On average, we find that only 22% of the ingredients in a recipe can be explained if the food type is known.

5.2. Investigating the value of food category labels

When uploading a new recipe, the user is expected to label it with category labels. To reveal to what extent this type of feature may be useful in predicting future recipe uploads we first investigated to which extent the users keep re-using the same category labels for their recipes. Figure 4 provides a better understanding of this pattern. Users introduce new categories at the very beginning and then start reusing them. After some short time, the number of used unique categories saturates. On average recipes are labeled with 2 categories, but the number of used labels ($y = 2.0x$) and the number of used unique categories cross when approximately the 15^th recipe is created. We can expect that before this number (of uploads per user) is reached, category labels for newly created and uploaded recipes may start reoccurring. This also means category labels are only useful to be used to predict future recipe uploads, if at least 15 or more recipes have been created by the user in the past.

To investigate to which extent category labels may explain ingredients and food types we have performed a similar experiment as presented in the previous section. The resulting plots are shown in Figure 5, where the horizontal axis represents the normalized entropy reduction

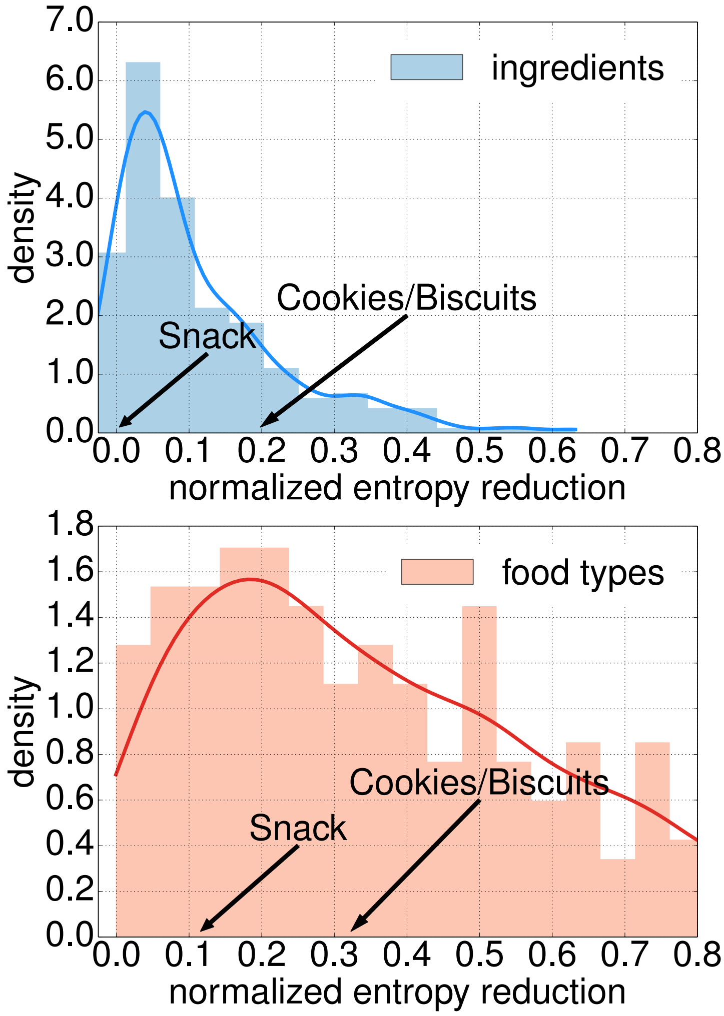


Figure 5: Distributions of reduction in entropy for ingredients and food types when the category of the recipe is known.

(the same measure as in Section 5.1 but here we restrict with categories). Again, the higher the values the more discriminative the category. Some categories are more discriminative than the others, for example, ‘Cookies/Biscuits‘ vs ‘Snack‘. On average we find that the reduction over all recipes for ingredients entropy (first plot) is equal to 11%. For food types (second plot), the entropy distribution is more flat and skewed towards high values. The overall mean here is 40%, which also means that category labels explain food types better than ingredients.

In summary, we find that users tend to reuse categories when labeling recipes and categorical information helps in constraining distributions of both food types and ingredients. On average by knowing how a user is going to label a recipe 11% of the ingredients can be explained though this proxy and there is a 40% probability to explain why a certain type of food may be uploaded.

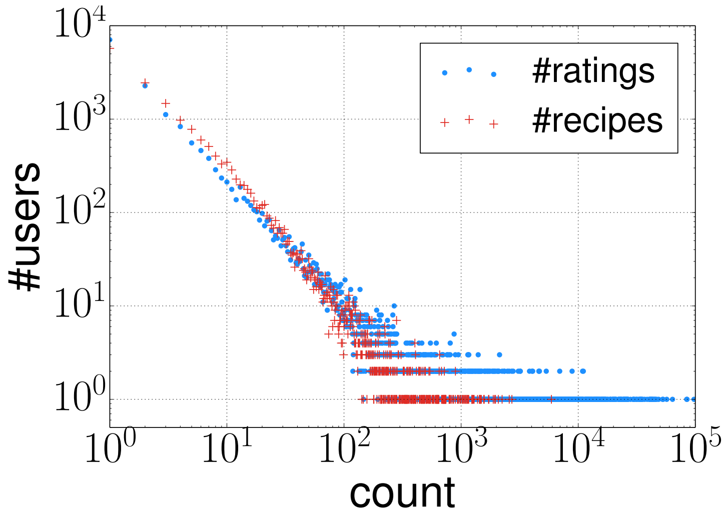


Figure 6: Distribution for the number of published recipes and ratings.

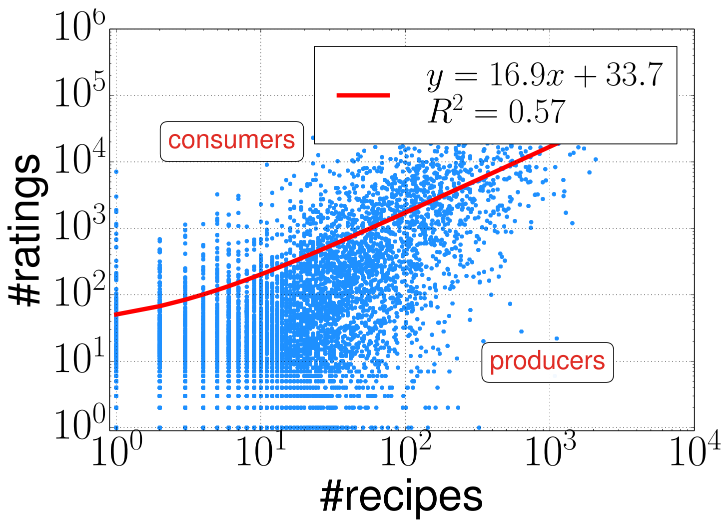


Figure 7: Number of recipe ratings vs uploads.

5.3. Investigating the value of ratings

Current research in the context of online recipes is mostly investigating rating behavior. Apart form our own preliminary previous work in this context (Kusmierczyk et al., 2016), no further investigations have been performed to understand whether recipe uploading behavior, may be different from what people prefer (rate). To further shed light onto this question, we performed a set of experiments. If rating behavior is correlated to upload behavior, we may be able to propose models that exploit the signal present in the ratings to predict future recipe uploads.

In the first experiment we compare rating and upload distributions. Figure 6 presents the results of this experiment. We observe that in both cases the distributions are heavy-tailed and that they follow a power law distribution (in both cases a K-S test can not reject at 0.05 the null hypothesis that the distributions are the same; we report $\alpha = 2.24$ for uploads and $\alpha = 1.39$ for ratings). This means that most users publish/rate only a few recipes and only a few users publish or rate many recipes.

Figure 7 shows the results of the correlation analysis. We observe a strong linear dependency between the two variables ($R^2 = 0.57$ and $p < 0.001$ for the null hypothesis that the slope is zero) meaning that people who are active in publishing recipes are also active in rating them (and vice versa). However, what is also shown in the plot is the nature of this relation: users rate more often than they upload recipes. This is evident from the slope of the linear fit that is almost equal to 17 (the intercept is 34). However, the data points are scattered, and some users deviate from the fit significantly, for an average user the number of ratings is significantly higher than the number of uploads.

In summary, we find that the users on average rate much more than they upload. However, rating behavior and upload behavior is correlated. An R^2 value of 0.57 shows that there is moderate strong correlation between recipe uploads and preferences in terms of ratings.

5.4. Investigating the value of the user's social network

Social networks have been found to be a very useful source of information in many prediction tasks. In particular, in the context of recommender systems, much research has recently shed light on the usefulness of social network relations to, for example, recommending items such as books, movies or music (Guy, 2015; Golbeck, 2006; Ma et al., 2011). However, less attention has yet been paid to what extent social networks can be exploited and can be useful in predicting food behavior online.

In the context of the online food community platform Kochbar.de, social connections are expressed by explicit friendship relations as well as by common interest groups. Similar concepts can also be found in other popular online community platforms such as, for example, Allrecipes.com [3] or Cookpad.com [4] where people can follow each other.

To reveal, to what extent recipe uploads are socially correlated, we performed a simple analysis that compared the user profiles of friends and non-friends to each other. The same was then repeated for users being connected with each other over the same groups.

From social connections we intend to obtain useful information about the users' general preferences towards particular ingredients or food types. Figure 8 shows the results of this experiment.

We compare the users to their friends and members of groups where the users and their friends are both members of (horizontal coordinate) and to other users that they are not connected to (vertical coordinate). To measure the similarities between the users, we employed cosine similarity in the following way: $x(u) = avg_{(u,u')\in\mathcal{S}} sim(u,u')$, $y(u) = avg_{(u,u')\notin\mathcal{S}} sim(u,u')$ where $\mathcal{S}$ stands for chosen social relation between users and $sim(u,u')$ measures similarity of items popularity in users' recipes, i.e., means a cosine similarity between count vectors of respectively ingredients and food types in users' recipes.

For both ingredients and food types we find a significant bias towards recipes uploaded by related users: in all cases a K-S test strongly rejects with $p < 0.001$ the null hypotheses that the distributions along the horizontal and vertical axes are equal.

[3] `http://www.allrecipes.com`
[4] `http://www.cookpad.com`

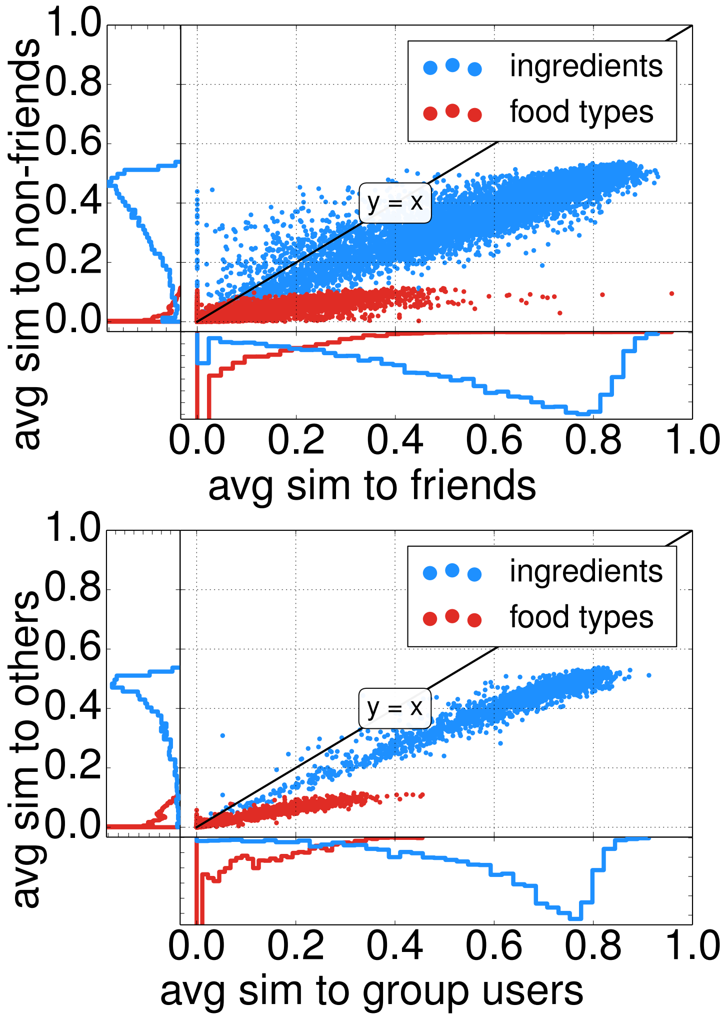


Figure 8: Similarities (measured over ingredients and food types) to recipes uploaded by socially connected users (friendship and group membership).

For example, approximately 95% of the users use the same ingredients ($x > y$) in their uploaded recipes as their friends, while the opposite ($x < y$) is true for only 5% of them. Similarly was obtained for food types, where we observe that 72% of the users being in a friendship relation also upload the same type of food. In the case of common groups these biases are even stronger: 99% of the users in the same group use the same ingredients and 95% upload the same type of food. This means that for the majority of people, being in a friendship relation to a person is a strong indicator of similar food preferences and being a member of a common interest group, is even a stronger indicator for the same food preferences.

In summary, we find that there are strong correlations between social factors such as friendship and group membership and recipes being uploaded by the users. 95% of the users use the same ingredients as their friends and 72% upload the same type of food.

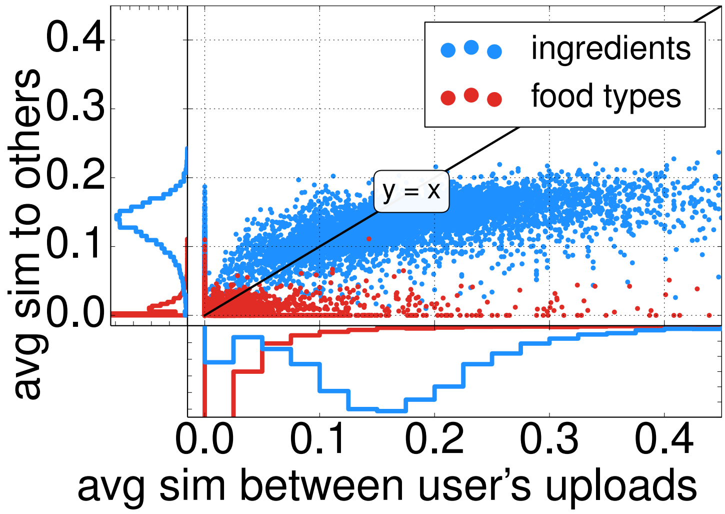


Figure 9: Similarities (measured over ingredients and food types) between the users own recipes and to other users'.

5.5. *Investigating the value of the users' histories*

Historical information is in many cases a very useful source of information to predict future events. In particular, in the context of recommender systems, the users' previous interactions with items, for example, in the form of ratings, are usually exploited to train a model that can inform future interaction behavior Ricci et al. (2011). Henceforth, we were interested in understanding to which extent the users' histories may be able to explain future recipe creation. To do so, we conducted an experiment, where we calculated for each user the mean pairwise cosine similarity between the ingredients used and the type of food the user uploaded (details below). As baseline we chose recipes uploaded by other users, in order to see if the users follow their own preferences rather the ones of their peers.

Figure 9 presents the results of this experiment. Each user u is represented by a single data point (one for ingredients and one for food types) and characterized by two values: mean similarity between recipes the user uploaded (horizontal axis): $x(u) = avg_{r,r' \in U_u} sim(r, r')$ with $r \neq r'$, where U_u is the set of uploads by u, and mean similarity to recipes uploaded by the other users u' (vertical axis): $y(u) = avg_{r \in U_u, r' \in U_{u'}} sim(r, r')$ with $u' \neq u$. Cosine similarity between two recipes $sim(r, r') = \cos(r, r')$ was calculated either over ingredients from r and r' (blue dots) or over food types assigned to them (red dots).

The plot shows strong biases towards the horizontal axis, i.e., similarities are higher between the users' own uploads rather than to other users (in most cases we observe that $x > y$). As anticipated, both a K-S test and t-test strongly reject with $p < 0.001$ the null hypotheses that the distributions along x and y axes are equal, i.e., that users are equal in terms of their styles and preferences towards ingredients or food types. In fact, users tend to have their own set of preferred ingredients and food types. For 52% of users (for ingredients) and 14% of users (for food types), the mean similarity between their uploads is (statistically) significantly higher than to those uploaded by other users, i.e., $x > y$. This means that for 52% and 14% users, personalized recommendations may make more sense. For 28% (for ingredients) and 82% (for food types), we were not able to distinguish between single user's uploads and uploads by the

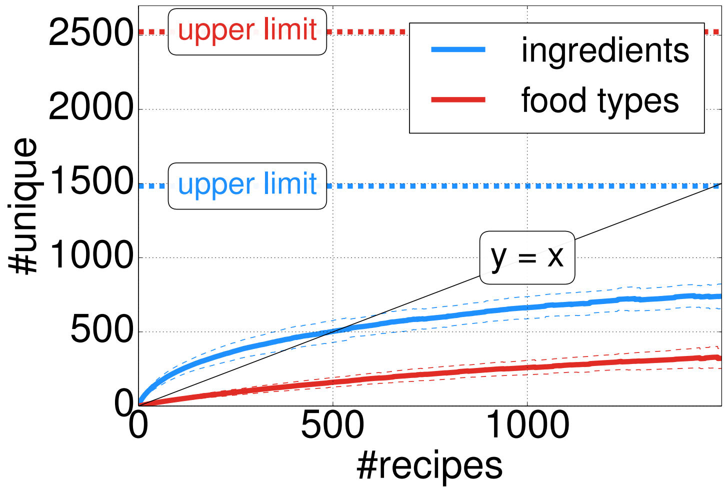


Figure 10: Number (mean and deviation) of unique ingredients used by the users in the recipes and food types created.

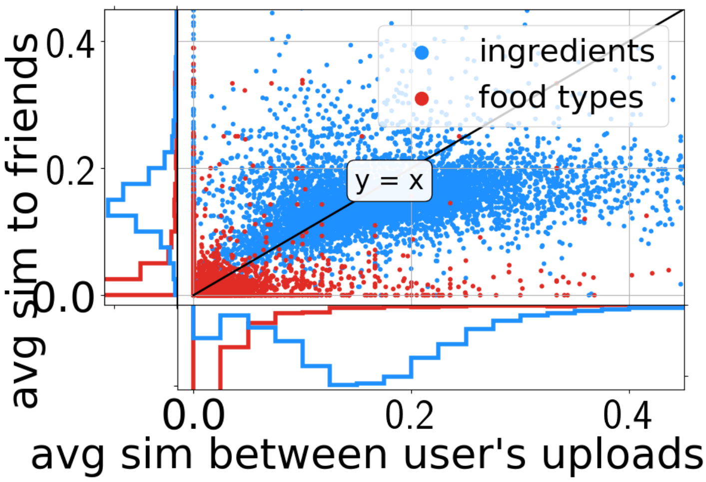


Figure 11: Similarities (measured over ingredients and food types) to recipes uploaded by socially connected users (friendship) vs own uploads.

other users – a t-test was not able to reject the means equality null hypothesis H0: $x(u) = y(u)$ at $\alpha = 0.05$. Finally, we found a group of users (respectively 20% for ingredients and 4% for food types) that behave in the opposite of the expected way, i.e., $x < y$: their uploads resemble more what others produce, and the difference is statistically significant in terms of mean values. For these users it does not make sense to personalize recommendations.

Other trends can be found when investigating Figure 11 where we consider only users which are socially connected to each other (friends). Here we can see that for ingredients we can find similar trends as in the previous investigation though less pronounced. 44% of the users are below the equality line (x=y) and 22% above meaning that the users are more to re-use their own ingredients than those of their friends. Interestingly when it comes to food types the pattern is different. 35% of the food types upload by the users is better explained by their friends rather than their own uploads (7%). Or in other words, people tend to have their set of known and available ingredients but do not want to repeat themselves by uploading the same set of recipes again and again.

Additional insights into the already observed patterns that the users follow their own cooking styles rather than others, is shown in Figure 10. The figure is intended to show the relation between the number of recipes created by the user and the number of ingredients used and different food types created. As highlighted, both the number of food types and ingredients grow with the number of published recipes.

What is also important to note here is that the single user never reaches the upper limits of all known ingredients (1483) and food types (2523) in the whole database.

This means that different users only partially share ingredients and types of food they upload with the others. In consequence, we expect the predictors that rely on the user's prior content to perform better than predictors that rely on recipe content published by other users, as this is, for example, the case when applying a item-based collaborative filtering approach Sarwar et al. (2001).

In summary, we find that the users upload histories are more or less consistent. For 52% of the users (for ingredients) and 14% of the users (for food types), the mean similarity between their uploads is (statistically) significantly higher than to those uploaded by other users. Similar can be found when comparing the users uploads to their friends although the pattern changes when we look at food types. In this case, most of the users' uploads can be explained through their social connections (friends).

5.6. Investigating the impact of the users' geographic embedding

Cultural and geographic differences in terms of cooking are well researched, and there are many studies that show evidence that there are significant differences how people from different regions all over the world prepare or prefer food (Steptoe et al., 1995). Yet, however, little evidence exists in the context of online cooking. To date, only two other study have shed light onto this question (Zhu et al., 2013; Wagner et al., 2014), showing via log file analysis that online recipe preferences vary significantly regionally. To observe whether this pattern can also be seen in the context of Kochbar.de, we investigated the users upload patterns from different countries, regions and cities.

As the results are similar for all three types of geography investigated (country, region and city), in order to improve readability of the results we present only those on country level in this section.

Similar to the methodology presented in Section 5.4, we investigated whether the users geographic embedding correlates with the types of recipes uploaded and the ingredients used in the recipes. Figure 12 (left plot) presents the first set of results in that respect, showing how similar the users are when considering their geographic origin. For both ingredients being used in the recipes, and food types, we observe a strong bias towards fellow countrymen. 97% of the users are on average more similar to people from the same country ($x > y$) in terms of ingredients used. In terms of food types this fraction is similar and amounts to 92%. However, a few users (respectively 3% and 8%) behave in the opposite way ($x < y$). Interestingly, according to a t-test with $\alpha = 0.05$ all the observed differences were found to be statistically significant.

An additional analysis is provided in Table 3. In order to reduce the impact of noise in our data, we present in this table only countries with at least 500 recipes and with at least 10 users

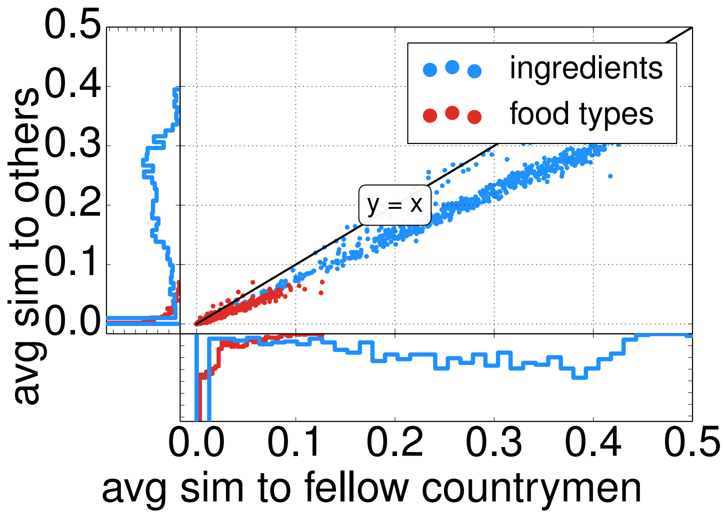


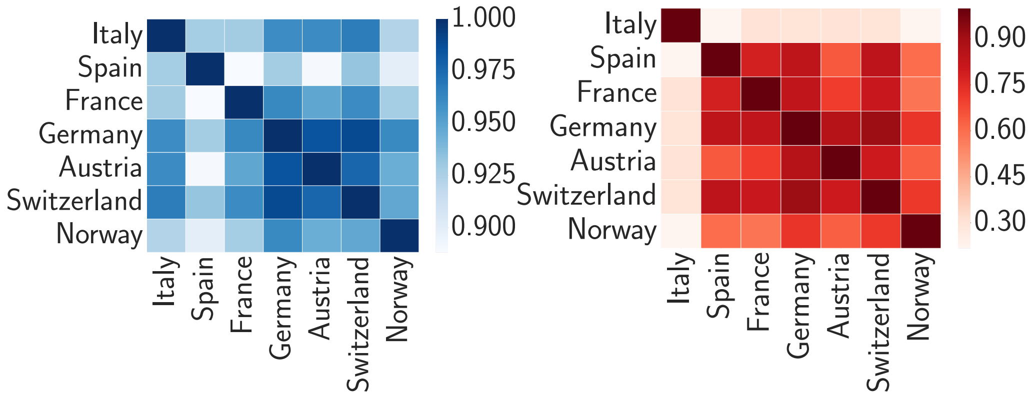
Figure 12: Similarity to recipes uploaded by user from the same country vs others.

Table 3: Fraction of users more consistent with their recipe upload behavior (ingredients and food types) to fellow citizens.

Country	**Ingredients**	**Food types**
Italy	100%	85%
Spain	95%	90%
France	100%	83%
Germany	100%	90%
Austria	54%	54%
Switzerland	90%	51%
Norway	87%	83%

in the particular country. The table shows the fractions of users more similar to the fellow countrymen than to foreigners ($x < y$). In most countries, such users constitute the strong majority, for example, up to 100% for ingredients and up to 90% for food types. In particular, users from Germany, Italy and Spain are remarkably consistent to their fellows. On the other hand, we have Austria and Switzerland that exhibit different trends, i.e., many users from these countries do not replicate local cooking patterns. A possible explanation for this behavior is the strong regional influence from neighboring countries such as Germany and Italy.

Further insight into the dependencies and similarities between countries in respect to ingredients being used and food types created is provided in Figure 13. The figure shows an analysis that measures the cosine similarity between the users in the countries in a heath map. As highlighted, similarities are higher when food types are considered rather than ingredients. What the two plots also show is that countries such as Germany and Switzerland are more similar to other countries. The highest similarities though (ingredient similarity = 99%, food type similarity =



(a) Ingredients

(b) Food types

Figure 13: Cosine similarity between distributions of ingredients and food types for countries with at least 500 recipes from at least 10 users.

91%) can be found between Austria and Germany. This means that the users in these countries create the most similar recipes. The countries with the smallest overlap in terms ingredient use and foods created is Spain and France (ingredient similarity = 88%) and Spain and Italy (food type similarity = 21%).

Other interesting results are obtained when investigating the top-5 most popular (normalized by $1/totalcount$) food types and the most popular ingredients used in the recipes uploaded on country level as shown in Table 4.

As shown, differences (and stereotypes) for the top-5 ingredients in each country can be found, although they overlap sometimes. For example, we find 'butter' in Germany and Switzerland among the top-5 as well as 'cream' in Austria and Switzerland. Also, the ingredient 'eggs' is prominent in Norway and Switzerland, as well as 'sugar' in Austria and Germany.

More clear, differences (and cultural food stereotypes) can be found among the most popular food types. For example, in Italy the top-5 food types are among others 'antipasti', 'limoncello' or 'bruschetta'. Similar can be found for other countries such as Spain, where we find 'paella' among the top-5 food types or Austria, where we find 'strudl' (a popular Austrian dessert) or in the Switzerland, where we find 'rösti' or 'cheese fondue' among the top-5. In Germany we find oxtail, a popular German cabbage salad.

Statistical significance of the observed differences between countries (also regions and cities) were verified with the χ^2 test of independence applied both to ingredients and food types (after limiting to subsets with sufficient counts, i.e., at least 5 per country/region/city). In all cases geographic influence on items popularity was confirmed by strongly rejecting with $p < 0.001$ the null hypotheses H0: countries/regions/cities are indifferent in terms of ingredients/food types popularity.

In summary, we find that users of the same country appear to cluster together. They upload the same types of recipes and use the same ingredients. However, for some countries this pattern is less pronounced than for others. For example, users from Switzerland and Germany are more likely to be similar with other users in other countries. The most similar countries are Austria and Germany (ingredient similarity = 99%, food type similarity = 91%) and dissimilar countries are Spain and France (ingredient similarity = 88%) and Spain and Italy (food type similarity =

Table 4: The top-5 ingredients and food types used in the country of origin of the users.

Country	Top-5 ingredients	Top-5 food types
Italy	peperoncini, pancetta, anchovy, clams, saffron	antipasti, limoncello, bruschetta, carpaccio, polenta
Spain	sald, gambas, ham, sardine, anchovy	fish stew, white bread, vegetable rice, gazpacho, paella
France	cheese, wine, champagne, corn starch, bread	salmon rolls, fennel salad, gorgonzola sauce, garlic bread, curry rice
Germany	sauce, butter, sugar, paprika, oil	oxtail, ragout, soup, stew, spread
Austria	sugar, cream, eggs, meat, marmalade	pancake, strudl, scheiterhaufen, kaiserschmarren, cup cake
Switzerland	cream, butter, flour, broth, eggs	salad sauce, cheese fondue, rösti, sour cream, orange salad
Norway	bouillon, meat, cheese, eggs, syrup	bbq sauce, bread, baked potatoes, cake, soup

21%), which is in line with Laufer et al. (2015) showing similar trends in Wikipedia.

5.7. Investigating the variable of time

Recent research in the context of online food has shown that there are observable temporal patterns how online food is accessed or searched for West et al. (2013); Wagner et al. (2014); Kusmierczyk et al. (2015b). While these analyses provide evidence that there are observable patterns in terms of, for example, how calorie content of the recipes searched for varies over the course of a week, these papers do not provide evidence to what extent the variable of time has an impact on recipe upload behavior.

To start with, we first focused on exploring the popularity of the ingredients used and food types created by manually checking visually for a random selection of 100 ingredients and recipe types. What we observed is that in most of the cases there are strong seasonal and weekly trends. Or in other words, we noticed that some ingredients and recipe types are used in the recipes uploaded more often in some periods of the year than others. Figure 14 presents sample situations where seasonality determines ingredients and food types use. All of the presented ingredients and food types follow strong seasonal trends. However, patterns between different items are shifted: ‘sugar’ is used more often than ‘onion’, however, in January it is ‘onion’ that is always popular. Similarly, ‘apple pie’ becomes more popular than ‘pasta salad’ every year at the beginning December.

In order to validate statistically our observations we applied a χ^2 test for uniformity (with $\alpha = 0.05$) over months and weekdays for all ingredients and food types in our dataset. Unfortunately, for many ingredients and food types, the counts were *not sufficient* (at least 60 (=5*12) for uniformity test over months and at least 35(=5*7) for uniformity test over weekdays). Among the 1287 (87%) ingredients with sufficient statistics, 1164 (90%) were season-dependent. For food types, only 666 (26%) had sufficient counts and 364 (54%) were season-independent.

On the level of weekdays, 1348 (91%) ingredients had sufficient counts, among which 462 (34%) reveal weekday-dependent characteristic. Respectively, among the 1069 (42%) food types with sufficient counts 122 (11%) were identified as dependent on a weekday. The overall results of this analysis (in percents of total number of ingredients/food types) is presented in Figure 15. Overall, we can conclude that there are ingredients and food types available in the dataset which are depended on the variable of time (e.g., the season of the year).

In summary, we find that ingredients and food types are season-dependent: 90% of the ingredients and 54% of the food types in Kochbar.de are affected in that way. However, when weekdays are considered, these temporal patterns are much weaker: only 34% of the ingredients and 11% of the food types are affected.

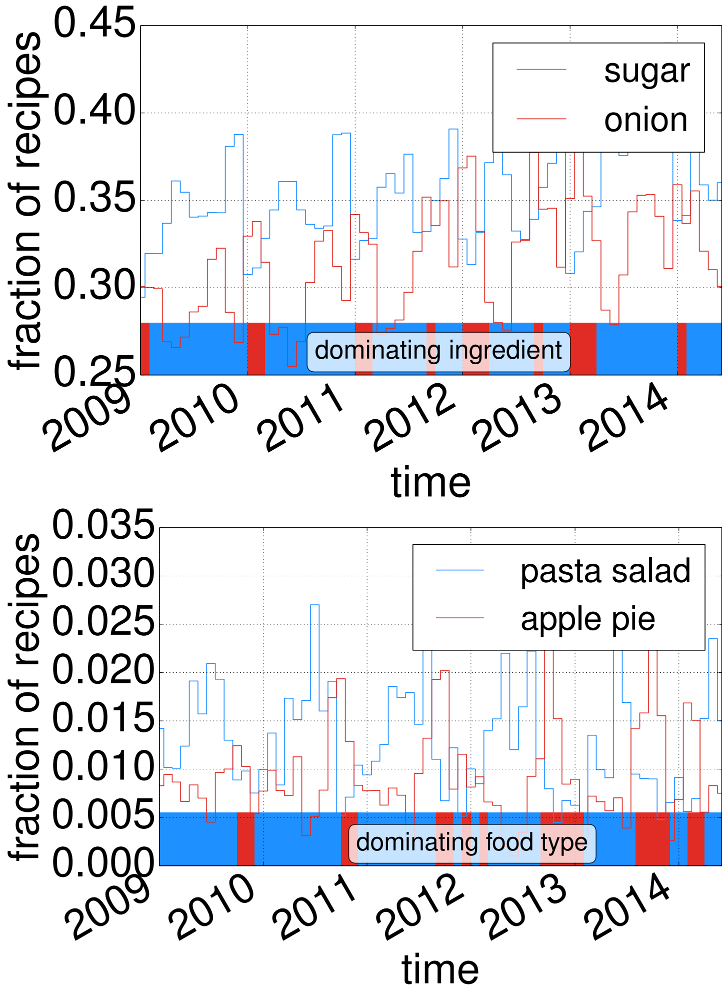


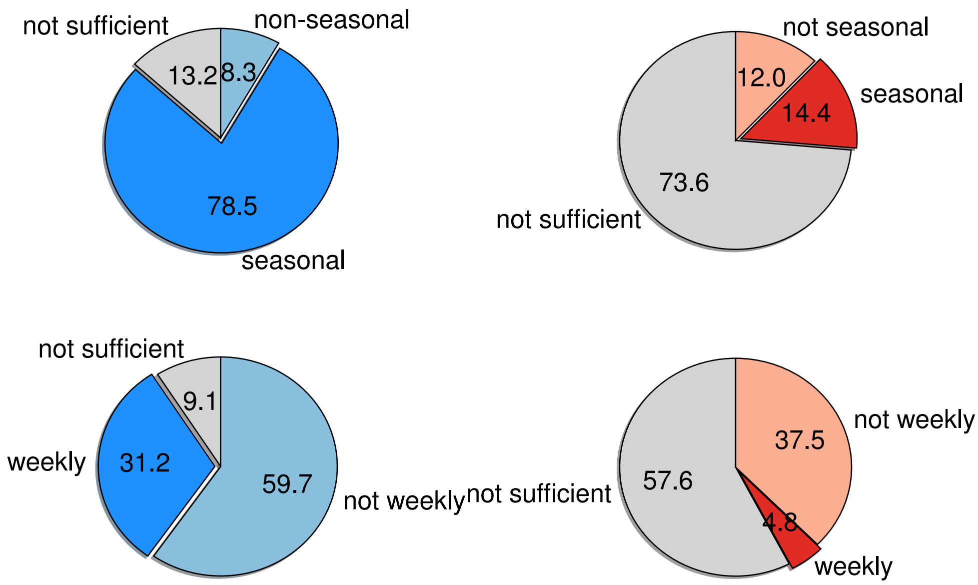
Figure 14: Popularity plots for ingredients (top) and food types (bottom) for two food type and ingredient samples over time.

6. Predicting Recipe Uploads (RQ2)

So far, we have shown that upload behavior can be explained through various features, but we do not know yet to which extent they are useful to predict future recipe upload behavior. In this section, we investigate to what extent the previously studied features can be exploited to predict recipe uploads. Specifically, the two prediction problems, we are trying to tackle in this section, are the following:

- *Given a target user u, what type of recipe r is she going to upload?*
- *Given a target user u, which of the available ingredients i is she going to use?*

The second problem we study in two settings: without recipe type given, and in a context-aware setting when recipe type is already known.



(a) Ingredients (b) Food types

Figure 15: Seasonal and weekly trends in ingredients and food types.

In addition the to set of users $u \in \mathcal{U}$, the set of recipes $r \in \mathcal{R}$, the set of types of recipes $t \in \mathcal{T}$, the ingredients $i \in \mathcal{I}$, we consider factors such as temporality T (seasonal and weekly trends), geographic context GEO (on country, region and city level), friendship F, group information G, and finally categorical information C. The task is then to propose a scoring function $S(u, e)$ (where $e \in \{t, i\}$) that assigns a preference score (predicts ranking) for candidate recipe type t or ingredient i for user u.

The output is then a ranked list of food types or ingredients. This can be used to implement applications as outlined in introduction to this paper (see Section 1), where we have proposed two types of scenarios. One that helps to understand potential food trends on user or group level[5] and the other one is to develop an intelligent user interface that supports the user during the upload with recommendations of recipes types and ingredient recommendations.

6.1. Evaluation protocol

The evaluation protocol we follow in this paper is the one usually used in order to evaluate predictive models and recommender system offline in a time-based manner (Campos et al., 2014). We split the dataset in training and test samples according to the time line, employing the leave-one-out strategy. Hence, the training set contains all the recipes published by a user apart from the last published (this one is put into the test set). In order to determine the quality of our

[5]Please note that for being able to predict food type or ingredient use of a group, a further aggregation step is needed, such as, for example, merging user profiles or user outputs, as for example proposed in Felfernig et al. (2018). In this paper, we focus on the user's individual levels.

predictors we used Normalized Discounted Cumulative Gain. Normalized Discounted Cumulative Gain (nDCG@k) is a ranking-dependent metric that not only measures how many items can be correctly predicted but also takes the position of the items in the recommended list with length k into account[6]. The nDCG metric is based on the Discounted Cumulative Gain (DCG@k) which is given by Järvelin and Kekäläinen (2002):

$$DCG@k = \sum_{p=1}^{k} \frac{2^{B(p)} - 1}{log_2(1 + p)} ,$$

where $B(p)$ is a function that returns 1 if the recommended item at position p in the recommended list is relevant. nDCG@k is calculated as DCG@k divided by the ideal DCG value iDCG@k which is the highest possible DCG value that can be achieved if all the recommended items would be in the correct order. Taken together, it is given by the following equation Järvelin and Kekäläinen (2002):

$$nDCG@k = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{DCG@k}{iDCG@k} ,$$

where we chose $k = 3$ for food types and $k = 10$ for ingredients used.

6.2. Predictors

User history predictors. The first proposed scoring function, based on the findings from Section 5.5, depends on the popularity of the item e (either ingredient i or type t) in historically uploaded recipes and is defined as following:

$$\text{MPU}(u, e) = \sum_{r \in U_u} [e \in r] ,$$

where [*condition*] takes 1 if *condition* is true and 0 otherwise and the expression $e \in r$ means that e (either ingredient, food type or category) is assigned to the recipe r. U_u is the set of recipes uploaded by user u in the past.

Similarly, relying on the findings from Section 5.3 where we showed that upload behavior strongly correlates with rating behavior, we can define:

$$\text{MPU-R}(u, e) = \sum_{r \in R_u} [e \in r] ,$$

where R_u is the set of recipes rated by user u in the past.

For some users the historical data may be very sparse. Hence, it might be better to smooth the scores by incorporating the information from recipes of other users that are somehow related,

[6] In our protocol we leave one item out, however, in contrast to what is usually the case, e.g. movie or music recommender systems, our item is a full recipe that inherently contains a list of ingredients, i.e., we do not evaluate one object (ingredient) but a whole set at the same time. Typically, these lists are of length 10 Trattner et al. (2018), and by using NDCG@10 we evaluate the quality of predicting such an average-length item (recipe). Similarly, each recipe is assigned a set of food types (however, fewer than ingredients). Furthermore, our predicting methods output scores that are then translated into ranks and it matters not only if the top-k ingredients/food types are correct but what is their order. In particular, the beginning of the list with recommended items is far more important than its tail. Precision and similar metrics can not capture that.

for example through common categories (Sections 5.1 and 5.5). The predictor that measures the popularity of item e in categories used by user u is defined as follows:

$$\mathrm{C}(u,e) = \sum_{c \in C} w(u,c) \cdot \left(\sum_{r \in \mathcal{R}} [c \in r \wedge e \in r] \right),$$

where $w(u,c) = \sum_{r \in U_u} [c \in r]$ measures the popularity of category c in user u's recipes. The second part weights the popularity of item e in recipes from the category.

Food types and ingredients are strongly correlated (Section 5.1). When the set $T \subset \mathcal{T}$ of types assigned to the recipe is known, the above scoring function can be adjusted in the following way:

$$\mathrm{C}[T](u,i) = \sum_{t \in T, c \in C} w(t) w(u,c) \cdot \left(\sum_{r \in \mathcal{R}} [c \in r \wedge i \in r \wedge t \in r] \right),$$

where $w(t)$ if the relative weight of a type t. We define $w(t) = \frac{1}{\sum_{r \in \mathcal{R}} [t \in r]}$ to give a higher importance to less popular and more specific types.

Similar to the case of uploads, we define a set of predictors that reflect the popularity in recipes rated by user u: respectively C-R and C-R[T], where U_u is replaced with R_u in the equations.

Social predictors. In Section 5.4 we have shown that social features correlate significantly with types of recipes being upload and ingredients used in the recipes. Hence, we propose to exploit social network relations as follows:

$$\mathrm{F}(u,e) = \sum_{f \in F_u} \sum_{r \in U_f} [e \in r],$$

where F_u is the set of direct friends of user u. Alternatively, we consider the set of groups (denoted G_u) where user u belongs to. In that context, the scoring function is defined over recipes uploaded by the users from the group (but not necessary assigned to the group explicitly):

$$\mathrm{G}(u,e) = \sum_{g \in G_u} \sum_{f \in g} \sum_{r \in U_f} [e \in r]$$

Assuming context (set of types T) to be known, the above scoring function can be adjusted as follows:

$$\mathrm{F}[T](u,i) = \sum_{t \in T} \sum_{f \in F_u} \sum_{r \in U_f} w(t) \cdot [i \in r \wedge t \in r]$$

$$\mathrm{G}[T](u,i) = \sum_{t \in T} \sum_{g \in G_u} \sum_{f \in g} \sum_{r \in U_f} w(t) \cdot [i \in r \wedge t \in r]$$

By replacing U_u with R_u in the above equations, we get the scoring functions over ratings instead of uploads, denoted F-R, G-R, F-R[T], and G-R[T].

Geographic predictors. In Section 5.6, we observed how geographic embedding of the users informs recipe uploads. Location can be represented on various levels of granularity, for example, country, region or city, implying different scoring functions:

$$\mathrm{GEO}[geo](u,e) = \sum_{u' \in origin(u)} \sum_{r \in U_{u'}} [e \in r],$$

where *geo* can by either country, region or city. Accordingly, $origin(u)$ returns either the country, the region or the city of the particular user u. The expression $u' \in origin(u)$ denotes the set of all users from the $origin(u)$. In the scenario with recipe types T known we obtain:

$$\text{GEO}[geo, T](u, i) = \sum_{t \in T} \sum_{u' \in origin(u)} \sum_{r \in U_{u'}} w(t) \cdot [i \in r \wedge t \in r]$$

Furthermore, we can consider the information consumption perspective GEO-R[*geo*] and GEO-R[*geo*, T] by considering ratings $R_{u'}$ instead of uploads $U_{u'}$.

Temporal predictors. In Section 5.7, temporal impact on the online recipe upload behavior was observed on both seasonal and weekly (to a lesser extent) level, implying two variants of the time-dependent scoring function:

$$\text{T}[time](u, e) = \sum_{u' \in \mathcal{U}} \sum_{r' \in U_{u'}} [tm(r') = tm(u) \wedge e \in r'] \ ,$$

where *time* can be either a month or a weekday and $tm(.)$ is a function that returns respectively month or day of week of the recipe rating posting (in case when historical ratings are applied) or recipe upload (in case when historical uploads are used).

Similarly, we can also consider temporal factors for ratings T-R[*time*] and with the context known $\text{T}[time, T](u, i)$.

6.3. Baseline methods

For the first baseline we used a non-personalized scoring function, the so-called *most popular* approach:

$$\text{MP}(u, e) = \sum_{r \in \mathcal{R}} [e \in r]$$

Additionally, we introduced *the most popular based on ratings* score:

$$\text{MP-R}(u, e) = \sum_{u \in \mathcal{U}} \sum_{r \in R_u} [e \in r]$$

Similar to above, we also define context-aware baselines MP[T] and MP-R[T] which measure popularity when the type of food is known.

Apart from these naive methods, we also compare to state-of-the-art methods from the literature, namely BPR-MF, ItemKNN, and UserKNN (only applicable in context-blind case) in two variants. The first variant relies on uploads and the second (suffix R, e.g, BPR-MF-R) on ratings. We used the implementations from the *MyMediaLite* library with default parameter settings (Gantner et al., 2011).

6.4. Results

Food type prediction. Figure 16 summarizes the results of our food types prediction experiment. The plot shows the results over all users (means incl. standard errors).

As presented, food type prediction problem is in general a hard task. None of the predictors obtain results over 0.075 nDCG@3. However, while the baselines achieve relatively high scores, a few approaches overpower them. One predictor, namely F-R (popularity in ratings by user

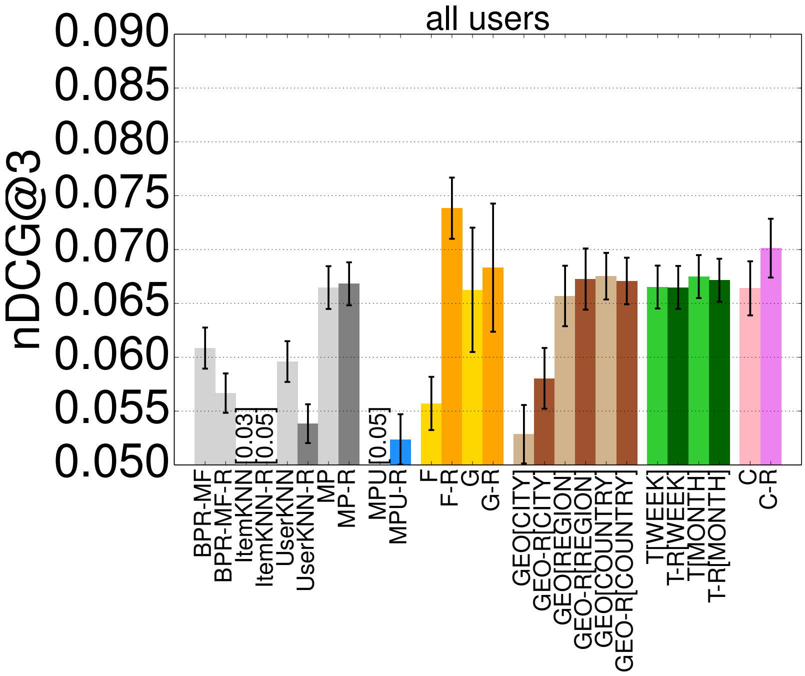


Figure 16: Food types prediction quality (means and standard errors) for all users. Colors indicate different groups of methods.

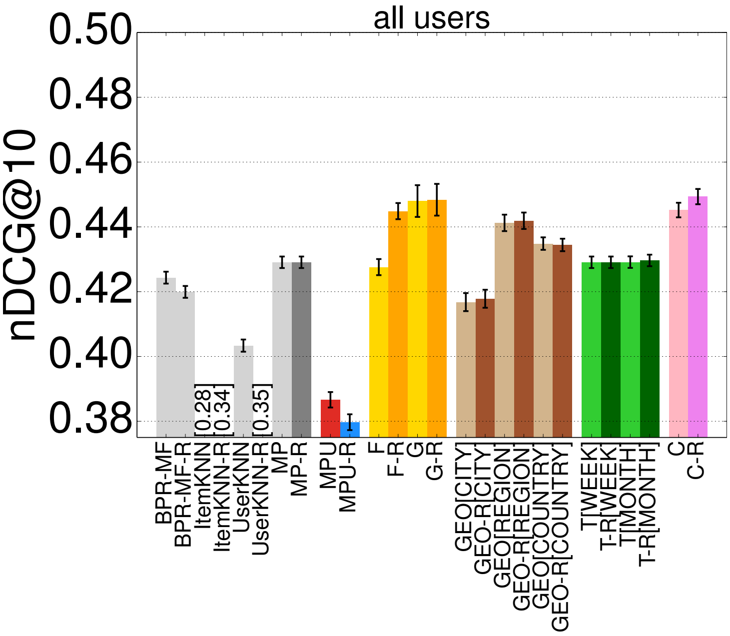
friends), performs significantly better than the others. The second best results are obtained via C-R (popularity in often rated food categories). Time-based and geographic predictor (in particular on city level) perform no better than the baselines. The worst among our approaches are MPU and MPU-R that rely entirely on users' previous uploads and ratings, suggesting that one of the key factors influencing results quality is the data sparsity, for example, just by regularizing with categorical information we can improve from one of the worst approaches (MPU-R) to almost the best one (C-R).

Ingredient prediction. Figure 17 summarizes the results of the ingredients prediction experiment in two cases: (a) recipe type is unknown and in (b) recipe type is given. In general, we find that the ingredient prediction task seems to be less complex which is evident in the much higher prediction performance compared to the food type prediction experiment. When the food type is known the prediction results are the best ($nDCG@10 \sim 0.68$).

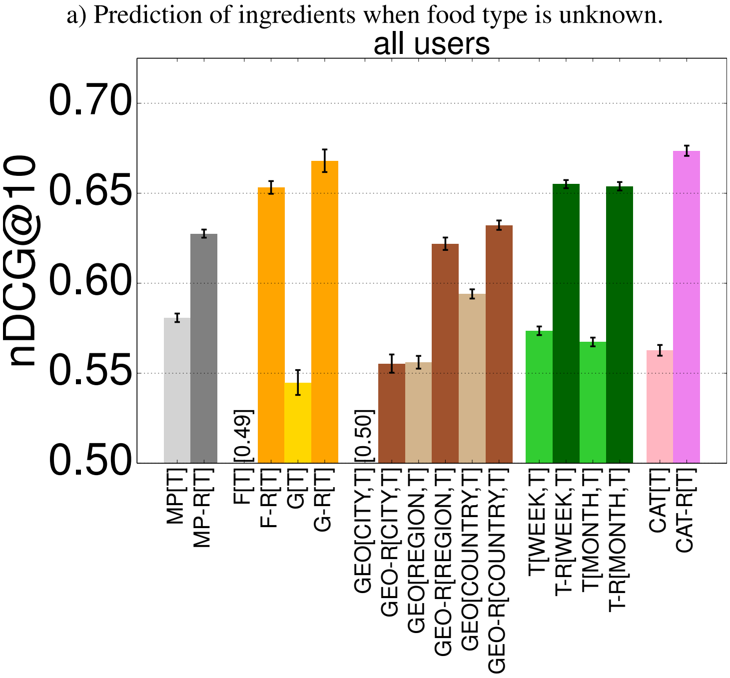
We also note that the best prediction quality is again obtained for scoring functions based on social features such as friendship (F-R[T]), group connectivity (G-R[T]), and for the methods that use categories and historical ratings (C-R[T]). It is interesting to note that complex recommender methods such as BPR-MF perform poorer than these simple scoring functions.

Interesting is also the observation that the rating-based approaches significantly outperform the upload-based approaches when the food type is known. One potential explanation for this behavior may be that people significantly rate more than they upload and as such it may be easier to extract stable patterns that explain the data well.

Summarizing our observations we can conclude, that in general the prediction of ingredients



a) Prediction of ingredients when food type is unknown.



b) Prediction of ingredients when food type is known.

Figure 17: Ingredient prediction quality (means and standard errors) for all users in the dataset.

is a much easier task than the prediction of food types. Furthermore, we find that, by exploiting social factors or categorical factors along with consumption evidence (ratings) we are able to outperform state-of-the-art baseline methods significantly.

7. Summary & Discussion

In summary, our experiments show interesting patterns to explain and predict online recipe upload behavior. On high level, the exploratory analysis (see RQ1 - Section 5) shows that there are a variety of signals and features that can be leveraged to explain upload behavior, such as the users' previous uploads, social network, temporal context as well as geographic embedding of the users. It is also interesting to note that these findings are often in line with the literature conducted in the real world, adding evidence that the online context bears a great potential to study real-world food behavior. For example, in line with Kearney (2010), showing global food patterns from survey data, our study also shows that users in the online context of the same country appear to cluster together. They upload the same types of recipes and use the same ingredients. However, for some countries this pattern is less pronounced than for others. For example, users from Austria and Switzerland are more likely to be similar with other users in other countries. This analysis is also in line with work recently conducted by Laufer et al. (2015) in the context of Wikipedia.

Also in line with Pachucki et al. (2011); Fletcher et al. (2011), showing to what extent social network influence food choice, our study shows that there are strong correlations between social factors such as friendship and group membership and recipes being uploaded by the users. 95% of the users use the same ingredients as their friends and 72% upload the same type of food. Also, in line with Stelmach-Mardas et al. (2016), providing a systematic review and meta-analysis of seasonality of food groups and total energy intake, we find that ingredients and food types are also season-dependent in the online context: 90% of the ingredients and 54% of the food types in Kochbar.de are affected in that way. However, when weekdays are considered these temporal patterns are less pronounced: only 34% of the ingredients and 11% of the food types are affected in that way.

Other interesting findings include that food types and ingredients correlate only to some extent. On average we find that only 22% of the ingredients in a recipe can be explained if the food type is known. Furthermore, we find that users tend to reuse categories when labeling recipes and category labels help in constraining distributions of both food types and ingredients. On average by knowing how a user is going label a recipe 11% of the ingredients can be explained though this proxy and there is a 40% probability to explain why type of food may be uploaded. This is an interesting behavior and in line with recent research on tagging of online resources, such as in Twitter or social bookmarking sites (Kowald et al., 2017; Trattner et al., 2016). Also we find that users on average rate much more than they upload. However, rating behavior and upload behavior is correlated. An R^2 value of 0.57 shows that there is moderate strong correlation between recipe uploads and preferences in terms of ratings. In addition to this we find that the users upload histories are more or less consistent. For 52% of users (for ingredients) and 14% of users (for food types), the mean similarity between their uploads is (statistically) significantly higher than to those uploaded by other users and friends.

Finally, from the prediction experiments (see RQ2 - Section 6) we can conclude that in general the prediction of ingredients is a much easier task than the prediction of food types. For food types we find values up to 0.075 nDCG@3 and remarkable values up to 0.45 nDCG@10 for the ingredient prediction task when the food type is unknown. Furthermore, we find that, by exploiting social factors or categorical factors along with consumption evidence (ratings) we are able to outperform current state-of-the-art methods significantly. The best overall predictor is the one exploiting social factors (= friendship connections). This finding is interesting as it adds also more evidence to the usefulness of exploiting social signals to the recommender literature investigating the potential of social connections to compute recommendations (Guy, 2015). Also of relevance is this finding for research conducted in online biases (Baeza-Yates, 2018) and especially in the online food context, where it was recently shown that popularity of recipes in terms of number of ratings or views is significantly influenced by social online connections (Rokicki et al., 2018).

Limitations, unanswered questions and future research. One important aspect to keep in mind when interpreting our results is that they relate only to one site, albeit one of the largest food portals on the Web, Kochbar.de. The site is primarily used by users from Europe and repeating the analyses with data from sites hosted in other countries may result in different outcomes. We plan to source other datasets and repeat our analyses.

Another important aspect that needs to be considered in future work are the different ways to model features in a recommender scenario (system). While the scoring functions as proposed in this study are easy to interpret and to compute, it may be also worthwhile to employ more sophisticated modeling approaches in the form of Factorization machines (Rendle, 2012) or learning to rank approaches (Macedo et al., 2015), to further reveal interactions or combinations of features which was totally neglected in this work. The primary goal was to investigate upload behavior and to reveal interesting patterns and potentially exploit these to predict future behavior. As such the goal was not to find the best method to predict recipe uploads but rather to explain it. Hence, a potential open research question is what is the best model to predict recipe uploads and what is the best feature combination.

Finally, it would be interesting to implement one of the systems, as highlighted in the introduction (see Section 1) based on the findings reported in that study and test it with real users. So far our experiments are based on simulations and statistical analysis and we have by no means an understanding, whether the predictions produced by our system would also be acceptable for the user. We hope to do this in our future work.

8. Conclusions

In this paper we have analyzed upload behavior of Kochbar.de, Germany's second largest online food community platform, comprising more than 400k recipes uploaded by nearly 200k users and revealed to what extent upload behavior may be explained and predicted. Our data analysis and prediction experiments show that social connections in the form of online friendship relations bear a great potential to predict what types of recipes user are going to create in the future and which ingredients are going to be used. This signal seems to be more important than geographic embedding of the user or the temporal context (= the time when the recipes uploaded). Exploiting the user's history is only beneficial if category labels are considered. The research conducted in this work contributes to a better understanding in online food behavior and is relevant for researchers working on online social information systems and engineers interested in predictive modeling and recommender systems.

References

Abbar, S., Mejova, Y., and Weber, I. (2015). You tweet what you eat: Studying food consumption through Twitter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3197–3206.

Ahn, Y.-Y., Ahnert, S. E., Bagrow, J. P., and Barabási, A.-L. (2011). Flavor network and the principles of food pairing. *Scientific Reports*, 1(196).

Alaimo, K., McDowell, M. A., Briefel, R., Bischof, A., Caughman, C., Loria, C., and Johnson, C. (1994). Dietary intake of vitamins, minerals, and fiber of persons ages 2 months and over in the United States: Third national health and nutrition examination survey, phase 1, 1988-91. *Advance Data*, 1994(258):1–28.

Alam, M. R. and Rahman, M. S. (2012). The substring inclusion constraint longest common subsequence problem can be solved in quadratic time. *Journal of Discrete Algorithms*, 17:67–73.

Baeza-Yates, R. (2018). Bias on the Web. *Communications of the ACM*, 61(6):54–61.

Berkovsky, S. and Freyne, J. (2010). Group-based recipe recommendations: analysis of data aggregation strategies. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, pages 111–118.

Burt, V. L., Whelton, P., Roccella, E. J., Brown, C., Cutler, J. A., Higgins, M., Horan, M. J., and Labarthe, D. (1995). Prevalence of hypertension in the us adult population: results from the third national health and nutrition examination survey, 1988-1991. *Hypertension*, 25(3):305–313.

Campos, P. G., Díez, F., and Cantador, I. (2014). Time-aware recommender systems: A comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction*, 24(1-2):67–119.

De Choudhury, M., Sharma, S., and Kiciman, E. (2016). Characterizing dietary choices, nutrition, and language in food deserts via social media. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 1157–1170.

Elsweiler, D. and Harvey, M. (2015). Towards automatic meal plan recommendations for balanced nutrition. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 313–316.

Elsweiler, D., Trattner, C., and Harvey, M. (2017). Exploiting food choice biases for healthier recipe recommendation. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 575–584.

Felfernig, A., Boratto, L., Stettinger, M., and Tkalčič, M. (2018). *Group Recommender Systems: An Introduction*. Springer.

Fletcher, A., Bonell, C., and Sorhaindo, A. (2011). You are what your friends eat: systematic review of social network analyses of young people's eating behaviours and bodyweight. *Journal of Epidemiology & Community Health*, pages jech–2010.

Freyne, J. and Berkovsky, S. (2010). Intelligent food planning: personalized recipe recommendation. In *Proceedings of the 15th International Conference on Intelligent User Unterfaces*, pages 321–324.

Gantner, Z., Rendle, S., Freudenthaler, C., and Schmidt-Thieme, L. (2011). MyMediaLite: a free recommender system library. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, pages 305–308.

Ge, M., Ricci, F., and Massimo, D. (2015). Health-aware food recommender system. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 333–334.

Golbeck, J. (2006). Generating predictive movie recommendations from trust in social networks. In *International Conference on Trust Management*, pages 93–104.

Guy, I. (2015). Social recommender systems. In *Recommender Systems Handbook*, pages 511–543. Springer.

Järvelin, K. and Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446.

Kearney, J. (2010). Food consumption trends and drivers. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1554):2793–2807.

Kowald, D., Pujari, S. C., and Lex, E. (2017). Temporal effects on hashtag reuse in twitter: A cognitive-inspired hashtag recommendation approach. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1401–1410.

Kusmierczyk, T., Trattner, C., and Nørvåg, K. (2015a). Temporal patterns in online food innovation. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1345–1350.

Kusmierczyk, T., Trattner, C., and Nørvåg, K. (2015b). Temporality in online food recipe consumption and production. In *Proceedings of the 24th International Conference on World Wide Web*, pages 55–56.

Kusmierczyk, T., Trattner, C., and Nørvåg, K. (2016). Understanding and predicting online food recipe production patterns. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, pages 243–248.

Laufer, P., Wagner, C., Flöck, F., and Strohmaier, M. (2015). Mining cross-cultural relations from Wikipedia: a study of 31 European food cultures. In *Proceedings of the ACM Web Science Conference*, page 3.

Liu, C., Guo, C., Dakota, D., Rajagopalan, S., Li, W., Kübler, S., and Yu, N. (2014). "My curiosity was satisfied, but not in a good way": Predicting user ratings for online recipes. In *Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP)*, pages 12–21.

Ma, H., Zhou, D., Liu, C., Lyu, M. R., and King, I. (2011). Recommender systems with social regularization. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, pages 287–296.

Macedo, A. Q., Marinho, L. B., and Santos, R. L. (2015). Context-aware event recommendation in event-based social networks. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 123–130.

Mejova, Y., Benkhedda, Y., and Khairani, K. (2017). #halal culture on instagram. *Frontiers in Digital Humanities*, 4:21.

Mejova, Y., Haddadi, H., Noulas, A., and Weber, I. (2015). # foodporn: Obesity patterns in culinary interactions. In *Proceedings of the 5th International Conference on Digital Health 2015*, pages 51–58.

Pachucki, M. A., Jacques, P. F., and Christakis, N. A. (2011). Social network concordance in food choice among spouses, friends, and siblings. *American Journal of Public Health*, 101(11):2170–2177.

Rendle, S. (2012). Factorization machines with libFM. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(3):57.

Ricci, F., Rokach, L., and Shapira, B. (2011). *Introduction to recommender systems handbook*. Springer.

Rokicki, M., Herder, E., and Demidova, E. (2015). What's on my plate: Towards recommending recipe variations for diabetes patients. In *Posters, Demos, Late-breaking Results and Workshop Proceedings of UMAP'15*, pages 1–6.

Rokicki, M., Herder, E., and Trattner, C. (2017). How editorial, temporal and social biases affect online food popularity and appreciation. In *Proceedings of the International Conference on Weblogs and Social Media*, pages 192–200.

Rokicki, M., Trattner, C., and Herder, E. (2018). The impact of recipe features, social cues and demographics on estimating the healthiness of online recipes. In *Proceedings of the International Conference on Weblogs and Social Media*, pages 310–319.

Said, A. and Bellogín, A. (2014). You are what you eat! tracking health through recipe interactions. In *RSWeb@ RecSys*.

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, pages 285–295.

Schneider, E. P., McGovern, E. E., Lynch, C. L., and Brown, L. S. (2013). Do food blogs serve as a source of nutritionally balanced recipes? an analysis of 6 popular food blogs. *Journal of nutrition education and behavior*, 45(6):696–700.

Schroeder, J. (2017). Agof: Chefkoch erstmals mit mehr als 20 mio. nutzern, rekorde auch für kochbar, essen, trinken und lecker.

Stelmach-Mardas, M., Kleiser, C., Uzhova, I., Peñalvo, J., La Torre, G., Palys, W., Lojko, D., Nimptsch, K., Suwalska, A., Linseisen, J., et al. (2016). Seasonality of food groups and total energy intake: a systematic review and meta-analysis. *European journal of clinical nutrition*, 70(6):700.

Steptoe, A., Pollard, T. M., and Wardle, J. (1995). Development of a measure of the motives underlying the selection of food: The food choice questionnaire. *Appetite*, 25(3):267–284.

Teng, C.-Y., Lin, Y.-R., and Adamic, L. A. (2012). Recipe recommendation using ingredient networks. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 298–307.

Tran, T. N. T., Atas, M., Felfernig, A., and Stettinger, M. (2017). An overview of recommender systems in the healthy food domain. *Journal of Intelligent Information Systems*, 50:501–526.

Trattner, C. and Elsweiler, D. (2017a). Food recommender systems: Important contributions, challenges and future research directions. *arXiv preprint arXiv:1711.02760*.

Trattner, C. and Elsweiler, D. (2017b). Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proceedings of the 26th International Conference on World Wide Web*, pages 489–498.

Trattner, C., Elsweiler, D., and Howard, S. (2017a). Estimating the healthiness of internet recipes: A cross sectional study. *Frontiers in Public Health*, 5:16.

Trattner, C., Kowald, D., Seitlinger, P., Ley, T., Kopeinik, S., et al. (2016). Modeling activation processes in human memory to predict the use of tags in social bookmarking systems. *J. Web Science*, 2(1):1–16.

Trattner, C., Moesslang, D., and Elsweiler, D. (2018). On the predictability of the popularity of online recipes. *EPJ Data Science*, 7(1):20.

Trattner, C., Parra, D., and Elsweiler, D. (2017b). Monitoring obesity prevalence in the united states through bookmarking activities in online food portals. *PloS one*, 12(6):e0179144.

Trattner, C., Rokicki, M., and Herder, E. (2017c). On the relations between cooking interests, hobbies and nutritional values of online recipes: Implications for health-aware recipe recommender systems. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, pages 59–64.

Trevisiol, M., Chiarandini, L., and Baeza-Yates, R. (2014). Buon appetito: recommending personalized menus. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, pages 327–329.

Wagner, C., Singer, P., and Strohmaier, M. (2014). The nature and evolution of online food preferences. *EPJ Data Science*, 3(1):1.

West, R., White, R. W., and Horvitz, E. (2013). From cookies to cooks: Insights on dietary patterns via analysis of web usage logs. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1399–1410.

Yang, L., Hsieh, C.-K., Yang, H., Pollak, J. P., Dell, N., Belongie, S., Cole, C., and Estrin, D. (2017). Yum-me: a personalized nutrient-based meal recommender system. *ACM Transactions on Information Systems (TOIS)*, 36(1):7.

Yu, N., Zhekova, D., Liu, C., and Kübler, S. (2013). Do good recipes need butter? predicting user ratings of online recipes. In *Proceedings of the IJCAI Workshop on Cooking with Computers*, pages 1–6.

Yujian, L. and Bo, L. (2007). A normalized Levenshtein distance metric. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 29(6):1091–1095.

Zhu, Y.-X., Huang, J., Zhang, Z.-K., Zhang, Q.-M., Zhou, T., and Ahn, Y.-Y. (2013). Geography and similarity of regional cuisines in China. *PloS one*, 8(11):e79161.