Plate and Prejudice: Gender Differences in Online Cooking

Markus Rokicki, Eelco Herder
L3S Research Center
Hannover, Germany
{rokicki,herder}@L3S.de

Tomasz Kuśmierczyk
NTNU
Trondheim, Norway
tomaszku@idi.ntnu.no

Christoph Trattner*
Know-Center
Graz, Austria
ctrattner@know-center.at

ABSTRACT

Historically, there have always been differences in how men and women cook or eat. The reasons for this gender divide have mostly gone in Western culture, but still there is qualitative and anecdotal evidence that men prefer heftier food, that women take care of everyday cooking, and that men cook to impress. In this paper, we show that these differences can also quantitatively be observed in a large dataset of almost 200 thousand members of an online recipe community. Further, we show that, using a set of 88 features, the gender of the cooks can be predicted with fairly good accuracy of 75%, with preference for particular dishes, the use of spices and the use of kitchen utensils being the strongest predictors. Finally, we show the positive impact of our results on online food recipe recommender systems that take gender information into account.

Keywords

online food; cooking; gender differences; classification; food recommender systems

1. INTRODUCTION

There are numerous suggestions in the literature that men and women behave differently and have different preferences on when, how, and why they cook. As discussed in the related work, professional cooking has traditionally been the domain of men and domestic cooking was considered a woman's duty [3]. Several studies and anecdotal evidence indicate that similar differences can still be observed: men tend to cook more for special occasions and their recipes are more ambitious and elaborate; women concentrate more on everyday recipes and pay more attention to health and balanced meals [5]. If such differences also exist in online recipe communities, it might be beneficial to take gender into account for usage analysis or personalization purposes.

In this paper, we look for quantitative evidence on differences between men and women in the domain of cooking, making use of recipes and interaction data from a large online recipe community.

*Work was carried out during the tenure of an ERCIM "Alain Bensoussan" fellowship at NTNU, Norway.

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

We will discuss to what extent popular stereotypes hold and how this translates into recipe uploading and commenting behavior. Our intention is not to speculate about the origin of such differences - there is plenty of sociological and popular literature on that topic - but to show that they do exist and that they are - quantitatively - significant. In addition, we collected a large number of features on the popularity, textual description, composition, and complexity of recipes and used these features to create models that are able to predict a user's gender in a quite reliable manner.

We believe that confounding factors, such as gender, impact the way popular recommendation methods - like collaborative filtering - work and that these factors generate biases that need to be taken into account [19, 2, 4]. Recommending food is believed to be more complex than recommending books or movies, for several reasons. Among others, food preferences are not only guided by taste, but also by dietary needs, seasonality, availability of ingredients, and societal conventions and expectations. As discussed in the related work, there are many indications that there are (strong) differences in societal expectations, dietary choices and cooking preferences between men and women. However, to the best of our knowledge, these differences have not yet been sufficiently quantified.

Contributions. We investigate gender differences with respect to cooking in three different ways. First, we statistically analyze gender differences, by illustrating and confirming or refuting *six prejudices* that are commonly mentioned: Men are better cooks (*H1*), men cook for impressing (*H2*), women cook sweet dishes and men meat dishes (*H3*), women use spices more subtly (*H4*), men use more gadgets (*H5*), and men are more innovative (*H6*). Second, we investigate to what extent features related to these prejudices are useful for *gender classification*. Finally, we investigate how *food recommendation* can be improved by taking gender into account. Using these three different approaches, we aim to provide more insight on the nature as well as the impact of gender differences in the field of cooking and food preferences.

Outline. The structure of this paper is as follows: Section 2 highlights relevant related work in the field. Section 3 introduces our data set and Section 4 features the results of our empirical data analysis. Section 5 and Section 6 present results of our classification and recommendation experiments while Section 7 finally concludes the paper with a summary of our findings and future directions of our work.

2. RELATED WORK

Genderification of Cooking and Eating. In Western culture, there has always been a separation between genders when it comes to cooking [3]: everyday, private domestic cooking was the domain of women, whereas professional cooking and haute cuisine were strictly the domain of men. In earlier days, it was unacceptable to have a female chef in a professional kitchen. This separation is not as strict anymore, but many differences are believed to still hold - not only in the form of prejudices or stereotypes, but also in actual differences in food preference.

According to sociologist Eva Barlösius [3], actual differences in eating preferences between men and women are relatively small and hard to quantify in small-scale studies. In contrast, food itself is often classified as 'male' or 'female'. For example, if a restaurant order comprises roast, baked potatoes and a beer for one person, and salad with grilled chicken and a white wine for the other person, it seems obvious to virtually all that the first person is a man and the second person a woman[1].

Still, it is undeniable that males talk and write differently about food than females. In a non-representative but still telling study[2], panelists concluded that gender 'certainly affects how chefs cook', but they could not articulate how and why exactly. Common prejudices were that women chefs use spices more subtly and that male chefs tend to cook to impress.

Cavazza et al. [5] carried out a study that confirmed that women preferred 'feminine food' to more 'male food'. In short, smaller and more elegantly presented meals were considered more feminine than larger, rough meals; meat was more associated with masculine meals. It has been speculated that these differences in preference might be partially related to preferences in self-presentation - in other words, how to conform (or not to conform) to expectations from society and peers. However, Dibb-Smith and Brindal [7] did not find any significant differences in food choice in different user contexts with different table companions.

Our study complements these works, providing large-scale empirical evidence on gender differences in cooking behavior.

Analysis of Patterns in Cooking and Eating Preferences. An in-depth analysis on how users choose and adapt recipes is given by Teng et al. [25]. Making use of complement and substitution networks, they show which ingredients users add, remove, pair or substitute. This allows them to predict which variation of a recipe will receive the best ratings.

Kusmierczyk et al. and Trattner et al. analyzed data from the German community platform Kochbar.de and found clear seasonal and weekly trends in online food recipe production, both in terms of nutritional value (fat, proteins, carbohydrates, and calories) [15, 26] and in terms of ingredient combinations and experimentation [14]. Similar patterns were observed by Wagner et al. [28] when investigating viewing logs. West et al. [29] found slightly different patterns for the American population. They also found correlations between search preferences and real-world health related issues. Similar observations were made recently by Said & Bellogin [23], De Coudhury et al. [6] and Abbar et al. [1] in the context of Allrecipes.com, Instagram and Twitter.

Rokicki et al. [22] investigated differences in nutritional values between user recipes created by different user groups. They found that recipes from females are, on average, richer in carbohydrates. Further, the amount of carbohydrates decreases with age - as recommended by most nutrition advice centers. Finally, there is the study of Wagner & Aiello [27], who studied gender differences in eating preferences in the context of the online platform Flickr. However, these works do not provide an in-depth analysis on gender differences expressed in recipe publishing behavior.

[1]Example by Barlösius in http://www.zeit.de/2016/06/ernaehrung-kultur-soziologie

[2]http://www.seriouseats.com/2009/06/do-men-cook-differently-than-women-gender-in-the-kitchen-grant-achatz-dana-cowin.html

recipes	405,868	users	199,749
ratings	7,794,868	publishing users	18,212
recipes with at least 10 ratings	240,518	users with at least 10 recipes	4976
ingredients	1485	ratings users	19,444
categories	246		

Table 1: Overview of the dataset.

Online Food Recommendation. In a seminal paper, Pazzani [19] compared the performance of different basic recommender algorithms for creating recommendations from a pool of 58 restaurants to 44 users. Despite the – by current standards – small dataset, they showed that collaborative filtering, content-based methods, and recommendations directly based on demographics – all had their strengths and weaknesses.

Harvey et al. [11] carried out a long-term study to analyze factors that influence people's food choices. Users indicated that reasons for liking or disliking a recipe include particular ingredients or combinations and the preparation time. Reasons for positive ratings include the type of dish and the novelty of the recipe. In addition, there are health-conscious users who also take nutritional information into account – which is only implicitly given in terms of the ingredients used and the quantities. A mobile health-aware food recommender system was recently introduced by Ge et al. [10].

Preferences for particular kinds of dishes and ingredients – as determined by nationality, season, previous experience and other factors – can be captured to a certain extent by collaborative filtering methods, based on food preferences of a neighborhood of similar people. Svensson et al. [24] designed a social navigation system for recipes and found that users liked and acted on aggregated user trails. Users claimed that they were more influenced by user comments than by the reputation of the author or specific ingredients. In a feasibility study on recipe recommendation, Freyne and Berkovsky found that both content-based (e.g., ingredients) and collaborative approaches (taste, context) should be taken into account [8]. In combination, these works motivate and provide a basis for our food recommendation experiments.

3. DATASET

For the purpose of our study, we rely on a large-scale crawl from Kochbar.de[3], a German online food community website to which users can upload and rate cooking recipes, obtained in [14]. The dataset encompasses more than 400 thousand recipes published between 2008 and 2014 (see Table 1). Ingredients are lists of arbitrary strings given as free-form text by users. We resolve word variants, misspellings, etc. in the same way as described in [14].

Almost 200 thousand users provided more than 400 thousand recipes, 2.7 million comments, and 7.7 million ratings. The ratings are on a Likert scale, but – surprisingly – they are overwhelmingly positive (99.1% gave a rating of 5). Hence, ratings were treated as binary feedback in our work, i.e. when there was a rating we counted it as positive. Gender and age information was given by 95 thousand and 57 thousand users, respectively. More than 18 thousand users have also actively contributed recipes to the platform; among them almost 5 thousand have published 10 recipes or more (888 male, 3807 female users).

Data Enrichment. In addition to the information inherent in the data, our analysis of common assumptions in connection to specific types of ingredients (meat and spices), types of dishes (sweet or hearty), and the use of gadgets relies on additional information described in the following paragraphs.

[3]https://www.kochbar.de

To identify red meat ingredients, we manually constructed a list of red meat types and matched them with food items from the USDA nutrition database[4], finding a total of 31 read meat ingredients. We identified spices in our dataset by matching them to a list of spices and herbs obtained from wikipedia[5], yielding 52 spices. By matching these ingredients to USDA food items, we expanded the list to a total of 80 spices.

Likewise, we identified cooking utensils that are mentioned in ingredient lists and preparation instructions. To this end, we performed exact string matching to a list of 350 cooking utensils and 22 categories extracted from the German Wikipedia.[6]

Our approach for identifying sweet dishes was inspired by the ingredient network analysis done by Teng et al. [25]. First, we computed a co-occurrence network of ingredients. Manual inspection of the graph using Gephi[7] revealed patterns similar to those found in their analysis: two dominant clusters around ingredients that are associated with sweet and hearty dishes. To obtain a high precision labeling of sweet dishes (with possibly imperfect coverage), we identified a small set of central ingredients in the sweet cluster and a larger list of ingredients in the hearty cluster. This way, 57 thousand recipes containing sweet ingredients, but none of the hearty ones, were marked as sweet dishes.

4. EMPIRICAL DATA ANALYSIS

In this section, we report the results of our empirical data analysis in accordance to six prejudices and respective hypotheses that are often believed in, when it comes to cooking between men and women .

4.1 Methodology

In our data analysis we compared relevant measurable indicators, such as the number of ingredients, number of downloads or preparation time, of which the means or medians can be compared using statistical tests. Whenever we compare users, we only considered those users who have published at least ten recipes - this ensures a sufficient amount of information on publishing behavior. When comparing recipes published by authors of different genders, we considered only recipes that received at least ten ratings, thus reducing noise in the data. In the statistical tests, the population size may vary due to missing values - mainly user profile data that a user has not provided. In the following paragraphs we introduce measures we rely on in our analysis to capture comment sentiment, ingredient diversity, and innovativeness.

Comment Sentiment. Sentiment of comments was computed using the German version of SentiStrength[8], judging expressed sentiments in terms of a positive sentiment score and a negative sentiment score. Based on this, two measures can be derived to capture *attitude* – the predominant sentiment – and *sentimentality* – the magnitude of sentiments [13].

Ingredient Diversity. Following Hill's work on diversity measures for species diversity in ecology [12], we use two measures to capture different qualities of diversity. First, for each user we measure diversity in terms of the number of different ingredients used (d_0). Second, we employ a measure that weighs rare occurrences less, based on Shannon Entropy ($d_1 = \exp(-\sum p_i \ln p_i)$).

[4]http://ndb.nal.usda.gov

[5]https://en.wikipedia.org/wiki/List_of_culinary_herbs_and_spices

[6]https://de.wikipedia.org/wiki/Liste_von_K%C3%BCchenger%C3%A4ten

[7]https://gephi.org/

[8]http://sentistrength.wlv.ac.uk

level	% recipes	% Men	% Women
easy	94.4	19.0	81.0
moderate	5.3	28.6	71.4
difficult	.3	35.3	64.7

Table 2: Ratio of recipes published by male and female users for easy, moderate and difficult recipes.

Innovation. We compute innovativeness of users in a similar way to [14]. Innovation of recipe r captures to what extend it differs from the most similar of previous recipes r':

$$IF(r) = 1 - \max_{r' \prec r} sim(r, r'),$$

where $\prec$ is temporal precedence and $sim \in [0, 1]$ measures similarity between two recipes - in our case the *Jaccard Similarity* over ingredients is employed. In order to study user-level innovation, we compare the mean innovation over recipes of users, computed over all recipes in the dataset.

4.2 H1. Men Are Better Cooks

'Better' is a very broad term and encompasses both objective and subjective measures of goodness, including self-judgement. It is known from the literature that 'professional cooking' historically is the domain of men; everyday, domestic cooking has traditionally been a woman's job. This leads to the expectation that when men cook, this will usually be for special occasions - and consequently be more festive, less everyday. As a result, one would expect that recipes from men are more time-consuming, more complicated, and more appreciated.

As a first step, we investigate the ratio between male and female users among – self-reported – difficulty levels, as shown in Table 2. Most published recipes are labeled as 'easy'. Apparently, Kochbar.de users have a preference for everyday recipes that do not require that much effort. Interestingly, the percentage of male authors is significantly higher for moderate and difficult recipes – $\chi^2(2, N = 268856) = 918.7, p < .001$.

If men are indeed better cooks, we would expect that their recipes are more popular. In order to verify this assumption we compared several popularity indicators. Recipes from men indeed seem to attract more comments than recipes from females ($M = 4.12$ versus $M = 3.30$; $W = 10268000$, $p < .001$, $r = .12$) as well as more views ($M = 825$ versus $M = 771$, $W = 11270000$, $p < .001$, $r = .04$). As expected, the average rating was similar between both genders (4.95 stars), due to the overwhelming amount of five-star ratings.

Interestingly, the sentiment of comments on recipes of both male and female authors shows a different picture. Female recipe authors receive more positive comments in terms of attitude ($M = .257$) compared to male recipe authors ($M = .238$) – $W = 1053998$, $p < .001$, $r = .11$. In addition, we also compare average sentimentality. Surprisingly, male recipe authors receive more sentimental comments ($M = .405$) than female recipe authors ($M = .390$) - $W = 1261516, p < .001, r = .07$. This indicates that male recipes elicit more controversial feedback.

In summary, the indicators show that men tend to publish more difficult recipes (or at least label them as more difficult). Recipes from male authors attract more views and more comments, but the sentiment of these comments is more diverse and on average less positive than comments on recipes from females. This effect may be explained by our second hypothesis: men cook for impressing, which may not always lead to better results.

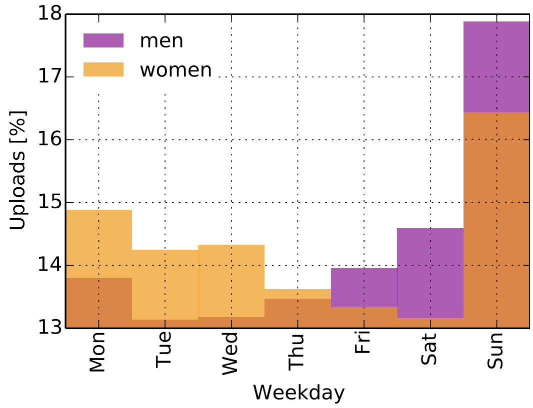

Figure 1: Comparison of women and men (with at least 10 uploads) in terms of activeness over the week. Both genders follow weekly rhythms – men are relatively more active on weekends, whereas women upload more between Mondays and Wednesdays.

4.3 H2. Men Cook for Impressing

Even though simple meals are often the best, elaborate meals are arguably the more impressive ones. Therefore, we compare the number and variety of ingredients used, the preparation time and the length of the recipe descriptions. We also investigate on which days of week both genders cook (=upload recipes in Kochbar.de).

The average number of ingredients that men use per recipe ($M = 10.22$) is slightly but significantly higher than for recipes by females ($M = 9.66$; $W = 9659100$, $p < .001$, $r = .17$). However, with respect to *diversity* of the used ingredients – as measured by the number of different ingredients used in a random sample of recipes (d_0, see Section 4.1) – we observe only moderate differences between men ($M = 57.3$) and women ($M = 54.7$; $W = 1905100$, $p < .001$, $r = .13$). The median preparation time is significantly higher (37.14 minutes versus 30.51 minutes; $W = 7840200$, $p < .001$, $r = .33$), confirming our previous observation on higher self-reported difficulty levels in male recipes.

Differences can also be found in how preparation instructions are written. Men use more words in preparation instructions ($M = 101.9$ versus $M = 86.8$; $W = 2044384$, $p < .001$, $r = .21$) and their instructions contain significantly more sentences ($M = 9.3$ versus $M = 8.8$; $W = 1801611$, $p < .001$, $r = .07$). These are indications that men indeed cook slightly more complex, time-consuming meals than women - rather than everyday meals attributed to women.

As men tend to cook more elaborate meals, they probably also cook more often for special occasions or during the weekend. Therefore, we expect different temporal behavior for males than for females, who tend more to provide everyday recipes. We compared user activity in terms of recipe uploads over the course of the week in Figure 1. The observed differences show that users indeed follow temporal patterns - the χ^2 test for uniformity strongly rejects the hypothesis that differences between days of the week are caused by chance with $\chi^2(6, N = 31805) = 368.83$, $p < .001$ for men and $\chi^2(6, N = 122104) = 649.33$, $p < .001$ for women. What is more, significant differences in patterns can be observed between genders, $\chi^2(6, N = 153909) = 66825.92$, $p < .001$. Men show relatively more active behavior during the weekends, whereas women upload more between Monday and Wednesday, supporting our initial hypothesis that motivations for preparing food of men and women differ.

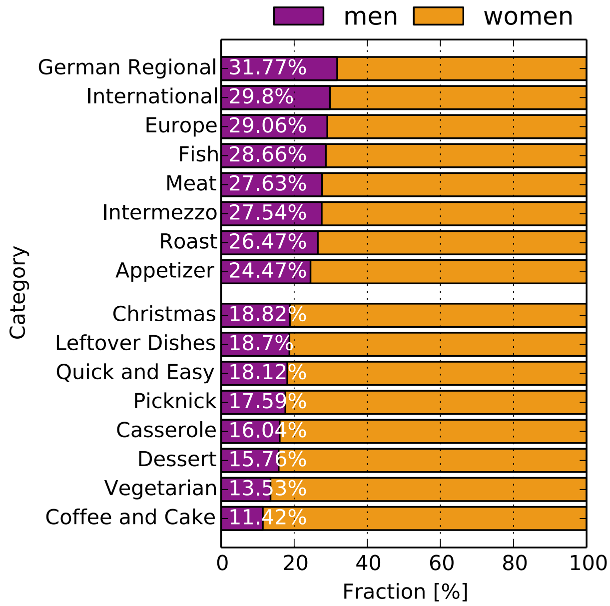

Figure 2: Popular categories with the highest and lowest percentages of recipes provided by men.

4.4 H3. Women Prefer to Cook Sweet Dishes, Men Prefer to Cook Meat Dishes

A common prejudice is that men tend to eat – and probably cook – heftier dishes, preferably with meat and fish [20]. There is indeed a significant difference in the distribution of male and female recipes over the 244 categories in Kochbar.de ($\chi^2(2, N = 2256419) = 22698$, $p < .001$). Figure 2 illustrates these differences by showing the most popular categories (containing more than 10,000 recipes) with the highest and the lowest percentages of recipes provided by men. Males indeed appear to prefer meat-related categories (meat, roast), whereas dessert, coffee and cake are categories that attract mainly women.

To more closely examine this aspect, we analyzed the use of red meat in particular. In order to avoid a bias towards sweet recipes for women, we only considered recipes in the main dish category. Male authors use red meat in 40.8% of their main dish recipes – significantly more than female authors (34.4%; $\chi^2(2, N = 64026) = 141.5$, $p < .001$). Taking a closer look at the stereotypical male ingredient bacon, we would expect even more pronounced differences. This is not the case, though: men use bacon in 10.6% of their main dish recipes, women in 9.3% of their main dishes, $\chi^2(2, N = 64026) = 26.0$, $p < .001$.

We now turn to female preference for sweet dishes. We compare the fraction of sweet dishesusing the labeling introduced in Section 4.1. Among recipes published by female cooks, 16.5% were identified as sweet dishes, significantly more than the fraction of 7.8% for male cooks, $\chi^2(2, N = 226835) = 2068.7$, $p < .001$. Our findings thus confirm both aspects of this hypothesis: men tend to cook meat dishes, women have a preference for sweet dishes.

4.5 H4. Women Use Spices More Subtly

Are there differences in how men and women employ spices? We investigate this by comparing the average number of spices used per recipe, the diversity of spices used by recipe authors of different genders, and which spices are used more by females and which are

		d_0		d_1	
Gender	N	Mean	Sd	Mean	Sd
Men	321	11.97	3.87	8.88	3.03
Women	1171	10.45	3.51	7.78	2.65

Table 3: Descriptive statistics for spice diversity results in main dish recipes.

used more by males. Our analysis is based on recipes from the main dish category from authors who published at least 10 main dishes.

Females use a lower number of different spices per main dish ($M = 2.30$) compared to males ($M = 2.61$), $W = 139129.5$, $p < .001$, $r = .26$. In this light, female use of spices in main dishes appears to be more reserved, but not necessarily more subtle.

Do men, apart from using more spices, also use a greater variety of spices? We investigate this using the diversity measures introduced in Section 4.1. To avoid biases due to differences in the number of published recipes, our computations are based on a random sample of 10 recipes per author. Table 3 shows descriptive statistics of the results on spice diversity for the main dishes. Men indeed use a significantly higher number of different spices (d_0) in their main dish recipes than women, $W = 232101$, $p < .001$, $r = .23$. In addition, male authors also achieve significantly higher d_1 diversity values, $W = 229959$, $p < .001$, $r = .22$.

A greater number and a higher diversity of spices are indications that men use spices in a less subtle manner. However, the choice of spices is of influence as well. Therefore, we further investigated spices that were used in at least 500 main dishes by ordering them according to the fraction of recipes that were published by male cooks. The results are shown in Figure 3. Spices that attract the highest fraction of males include heavy spices that are often used in hefty dishes and stews, such as caraway, bay leaf, rosemary and cloves. Spices with the lowest fraction of recipes from males, the 'female spices' appear to be more everyday (pepper, nutmeg, paprika) and commonly used in salads, soups and other light dishes (mustard, dill, basil). Interestingly, the use of the general word 'spice' as a placeholder seems to be be mainly used by women - which might indicate that women consider spices as something to support the taste of a meal, whereas men use spices more to influence the taste. This observation is in line with the previously observed male preference for more complex, impressive dishes for special occasions.

4.6 H5. Men Use More Gadgets for Cooking

In the introduction of a typical cookbook targeted at men [20] it is stated that '*you should know how to execute kitchen tasks with confidence, aplomb, and – I dare say – showmanship*. Typical male kitchen tasks involve the use of sharp knives and other impressive devices – or gadgets.

As a proxy for the use of gadgets, we compare gadget mention rates in recipes with at least 10 ratings. Male authors mention any kitchen utensil in 86.7% of recipes, in comparison to 83.5% for recipes published by female authors, $\chi^2(2, N = 226865) = 263.3$, $p < .001$. As these differences may be caused by the type of dishes cooked, we also looked at more constrained subsets. Within the main dish category, kitchen utensils are mentioned more often overall, with a small, but still significant, difference between male and female recipes (90.4% versus 89.7%; $\chi^2(2, N = 64033) = 6.3$, $p < .05$). Also, in the 'female' category desserts, kitchen utensils are mentioned slightly but significantly more often than in male recipes (89.1% versus 86.3%), $\chi^2(2, N = 23831) = 19.5$,

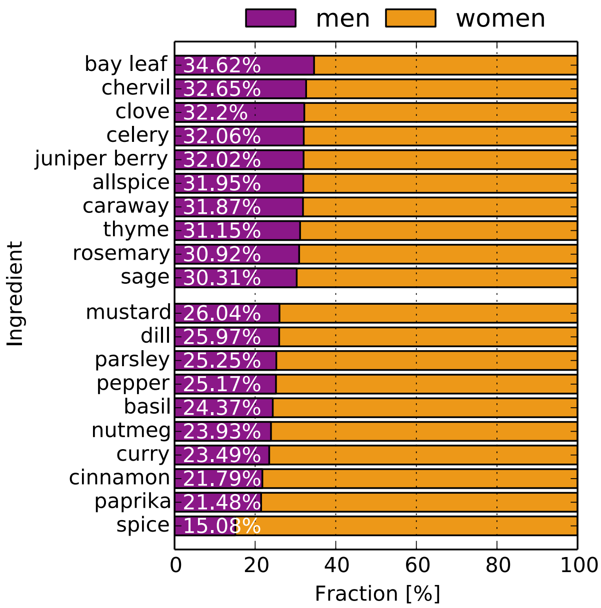

Figure 3: Popular spices used in at least 500 main dish recipes, with the highest percentage of male recipes (at the top) and the highest percentage of female recipes (at the bottom).

$p < .001$. Equally pronounced is the difference between men and women for the dishes containing red meat (89.8% versus 87.6%), $\chi^2(2, N = 48831) = 43.1$, $p < .001$).

We also investigated which gadgets are predominantly used by men and by women. An analysis based on all recipes mainly revealed a bias towards utensils for baking by females (e.g. forms and trays, or measuring devices) and utensils associated with hearty meals by men (e.g. knives), which is in line with our observations regarding H3, but not overly surprising. We therefore restricted our further analysis again to main dish recipes. Table 4 shows the mentioning rates of different gadgets. Utensils mentioned more often in male recipes include knives, utensils for separating, pots, pans and kettles. Although differences between these categories are significant, effect sizes are small. Still small but slightly more pronounced is the difference for forms and trays, which tend to be mentioned more often in female recipes. Surprisingly, we do not find gender differences in mentioning 'electrical devices' (3.62% versus 3.58%; $\chi^2(1, N = 93997) = .04$, $p = .8$).

These differences in gadget use can at least partially be explained preferences for different types of dishes. For instance, women mention casserole dishes more often, evidence for a preference of 'souffles' and other similar dishes – 13.94% versus 9.34%; $\chi^2(1, N = 93997) = 95.1$, $p < .001$. Male cooks, on the other hand, are more likely to employ hatchets (7.59% versus 5.13%; $\chi^2(1, N = 93997) = 188.7$, $p < .001$) and roasters (5.76% versus 3.42%; $\chi^2(1, N = 93997) = 240.0$, $p < .001$) – tools associated with the preparation of roast and other meat dishes. This confirms our observations on male preference for meat dishes.

4.7 H6. Men Are More Innovative

Given the tendencies of men to create more elaborate, time-consuming recipes that attract more polarized comments, it is likely to assume that male recipes are more innovative than recipes from women (for better or for worse). To verify this, we compared the

Category	Male Recipes (N = 21,896)	Female Recipes (N = 72,101)	χ^2
Knives	7.77%	5.21%	201.4*
Separating	10.90%	8.47%	120.4*
Pots, pans, kettles	64.42%	59.92%	143.1*
Hand tools	12.86%	12.01%	11.2*
Warming devices	36.33%	34.85%	16.1*
Containers, bowls	47.51%	46.93%	2.3
Mixing	7.28%	7.98%	11.2*
Forms, trays	9.98%	14.99%	352.6*

Table 4: Mention rates for gadget categories that occur in at least 5000 main dish recipes, ordered such that categories mentioned relatively more often in male or female recipes are at the top and bottom, respectively. Results marked with * are significant at $p < .001$.

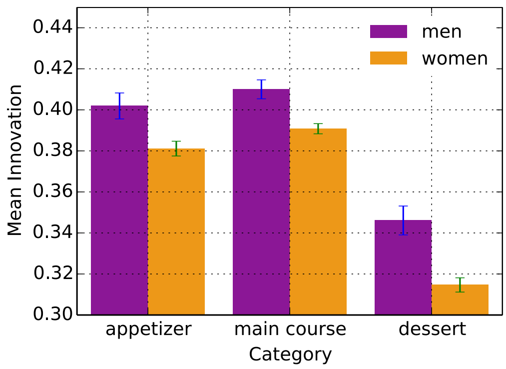

Figure 4: Comparison of innovation in recipes uploaded by women and men (with at least 10 uploads) for three meal categories present in Kochbar.de.

two genders in Kochbar.de in terms of their *innovation factor* (see Section 4.1). To remove the effect of high innovation rates for recipes from the early period (due to the absence of previous content), we evaluated only uploads starting with January 2010. Indeed, we observed a strong, statistically significant difference between genders ($D = .19$, $p < .001$), with men on average being more innovative ($M = .38$) than women ($M = .35$), with a large effect size ($W = 8.99$, $p < .001$, $r = .994$).

Figure 4 illustrates differences in innovation of men and women for three different food categories. The figure shows on the one hand that innovation varies between categories – for example, it seems to be more difficult to innovate in 'desserts' than in appetizers and main dishes. In addition, the innovation gap between genders differs between categories ($W = 3.46$, $p < .001$, $r = .995$ for appetizers, $W = 4.45$, $p < .001$, $r = .996$ for main courses, and $W = 4.18$, $p < .001$, $r = .995$ for desserts). Desserts appear to have the biggest innovation gap between genders.

5. GENDER CLASSIFICATION

In the previous section, we have shown that there are pronounced differences between men and women in terms of cooking preferences. In this section, we present the results of a gender classification experiment that we conducted to find out which features are most discriminative and to what extent these features can be used for identifying male and female cooks based on their cooking style.

5.1 Methodology

Feature Engineering. We selected 88 features that are related to the hypotheses that we investigated in the previous section. Below we briefly summarize these features, with the feature sets corresponding to the hypotheses.

- **Men are better cooks** (*H1*). We derived 11 features that capture the distribution of published recipes across difficulty levels, number of views, ratings, and comments, as well as sentiment and sentimentality in comments on recipes of the user.
- **Men cook for impressing** (*H2*). This hypothesis is covered by 16 additional features, including preparation time, ingredients per recipe, ingredient diversity, average word length, average number of words, characters, and sentences in instructions. Further features capture the temporal behavior in terms of the distribution of uploads across days of the week and regularity of publishing behavior [9].
- **Women cook sweet dishes and men meat dishes** (*H3*). These differences in preferences and cooking practice are captured by 19 features, including the fractions of recipes containing red meat, bacon, and recipes labeled as sweet dishes, as well as the distribution across the categories preferred by male and female cooks (listed in Figure 2).
- **Women use spices more subtly** (*H4*). We derived 3 features based on the use of spices, quantifying the diversity of spices used in a sample of recipes and the average number of spices used per dish.
- **Men use more gadgets** (*H5*). We modeled the use of gadgets with 18 features that capture the frequency of mentioning any gadget, as well as gadgets in the 17 most frequently mentioned categories.
- **Men are more innovative** (*H6*). 19 Features capture innovation, measuring overall innovation, as well as innovation in main dishes and recipe categories preferred by male and female cooks.

Dataset Preparation. Our classification experiments were evaluated on a balanced dataset: we under-sampled women, so their number was equal to the men. Further, to ensure sufficient evidence, we focused our analysis on active users who published at least 10 recipes. Overall, the preprocessing resulted in 888 users in the class 'men' and 888 users in the class 'women'.

Feature Selection. Discriminative power of features was compared with the help of Information Gain (IG) and decrease in Random Forest (RF) accuracy [17]. Information Gain weights features according to their correlation with class attribute (gender) based on entropy. Mean decrease in accuracy of Random Forests measures classification performance in comparison to using a randomly chosen feature.

Classification. The classification experiment was conducted with the help of the Weka[10] machine learning suite. Classifiers employed for the experiment were Random Forests (RF), Logistic Regression (LR), and AdaBoost (AB) with standard parameter settings. The evaluation protocol we employed was 10-fold cross-validation.

[9]The three features capturing difficulty levels are also redundantly included in this category.
[10]http://www.cs.waikato.ac.nz/ml/weka/

feature name	IG	rank	RF	rank	H
sweet recipes	.058	1	9.325	4	H3
‘forms’ gadgets	.045	2	16.122	1	H5
spices per recipe	.043	3	11.019	3	H4
‘pots & pans’ gadgets	.039	4	3.358	18	H5
red meat recipes	.034	5	7.335	6	H3
‘coffee & cake’ recipes	.027	6	13.567	2	H3
bacon recipes	.025	7	2.260	30	H3
distinct spices count	.023	8	1.376	49	H4
preparation time	.021	9	4.181	12	H2
international category	.018	10	1.521	43	H3
spices diversity	.018	11	1.468	47	H4
‘knives‘ gadgets	.016	12	2.112	35	H5
innovation in ‘coffee & cake’	.016	13	8.146	5	H6
regional category	.016	14	−1.086	85	H3
‘pounding’ gadgets	.016	15	.696	61	H5
average words in instructions	.015	16	2.743	23	H2
average char count in instructions	.015	17	2.695	25	H2
Europe category	.015	18	.517	66	H3
innovation	.014	19	5.985	7	H6
dessert category	.012	20	.609	64	H3

Table 5: Top-20 features (out of 88) with gender as a target feature, according to Information Gain (IG) and mean decrease in accuracy of Random Forests (RF). Most of the best features stem from hypotheses H3 (preferences for sweet/meat dishes) and H5 (use of gadgets).

5.2 Results

Feature Selection. The 20 best features along with the evaluation measures (obtained with the FSelector package in R[11]) are presented in Table 5. We note that both measures (IG and RF) correlate to a high extent for the top-ranked features.

Figure 5 shows all 88 features, ranked according InfoGain - this measure is classifier-independent and we already noted that results between the two feature selection methods do not vary to a great extent. We observe a large diversity in feature quality and also differences in the quality of features from the six hypotheses that we investigated. A few top features are related to H3 (men prefer meat whereas women sweet dishes), with the highest InfoGain for ‘sweet recipes’ (the fraction of recipes marked as ‘sweet’). Two more of the best features are related to H5 (men use more gadgets): the use of cooking forms (#2) and pots (#4). The feature ‘spices per recipe’ (H4) ranks at #3. Finally, one feature related to H2 (men cook to impress) is also in the top-10 features: preparation time.

Among the middle-quality features with InfoGain between .001 and .02 we find several other features related to H3, H4, and H5, which delivered the best-performing features. The middle field also contains some features regarding innovation (H6) and cooking for impressing (H2), but popularity and difficulty features related to H1 (men are better cooks) performed quite poorly.

Finally, there is a long tail of a very weak features (with almost zero InfoGain) from almost all hypotheses. This includes almost all features related to hypothesis H1 (men are better cooks than women), which - and this may be of relief to some women - confirms the weakness of such a claim.

[11]https://cran.r-project.org/web/packages/FSelector/FSelector.pdf

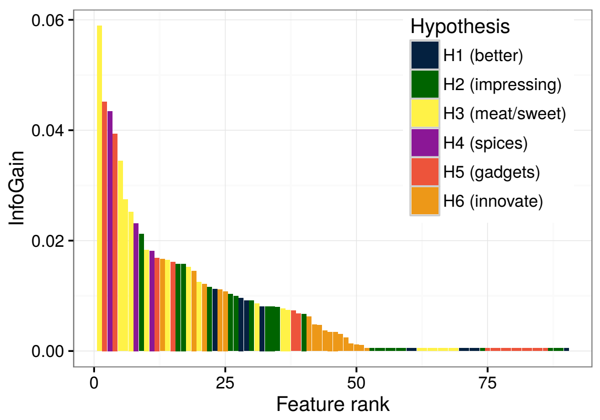

Figure 5: Ranking of the quality of features according to Information Gain. The feature classes are color-coded. The most useful features are related to hypotheses H3, H5 and H4.

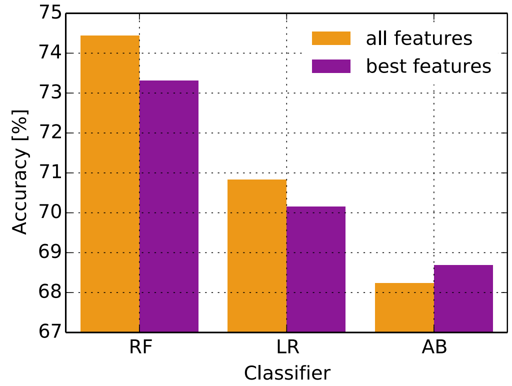

Figure 6: Classification accuracy for Random Forests (RF), Logistic Regression (LR), and AdaBoost (AB), using a complete set of features and the top 20 features (according to Information Gain).

Classification Accuracy. Figure 6 compares the results of our classification experiment in two settings: with all features and with only the top 20 best (according to InfoGain). When both classes (genders) are balanced, a random classifier would achieve exactly 50% of accuracy. As shown, all three classifiers improve significantly over this baseline: the best classifier, RandomForests, reaches an accuracy level of almost 75% when all features are considered. Similar results are observed with the Logistic Regression classifier, which reaches almost 71% accuracy. The worst among the three considered classifiers is AdaBoost reaching almost 69% of accuracy when only the 20 best features are used. Although the differences between methods are meaningful, all of them perform satisfactorily, which demonstrates that our selection of features - particularly the best-performing features related to H3 (meat/sweet), H4 (spices) and H5 (gadgets) - effectively captures differences between men and women.

6. GENDER-AWARE RECOMMENDATION

In the previous sections, we found that there are various effects between the user’s gender and several preferences and tendencies concerning recipes and cooking. Particularly features concerning preferences for meat or sweet dishes as well as the use of spices and gadgets were strong enough to classify users as either male or female. In this section, we will show the impact of 'gendering' on the success of recipe recommendation. We adapt two common recommendation strategies by restricting them to users of the same gender, and to recipes from authors of the same gender.

6.1 Methodology

We first selected users who have rated at least 20 items and items (recipes) that have been rated at least 50 times. This procedure was performed to ensure that enough data is present in the user profile to learn from and to remove the bias of unpopular items in our dataset (see e.g., [18]). After that, we employed the MyMediaLite[12] recommender framework to run two different recommender strategies that are often employed in real-world systems [9]. The first recommender strategy we chose is known as the *MostPopular* approach, which recommends the most popular items (in terms of obtained ratings) to the user. The second strategy we chose was user-based *collaborative filtering* with k-Nearest Neighborhood (KNN) search. To show the effect of gender on the two baseline methods MostPopular (MP) and Collaborative Filtering (CF), we adapted the methods in the following forms.

The first restriction of MP concerns the user group of which the most popular items were drawn: $MP(g)$ recommends the items that are most popular among users of the same gender. As a second restriction, we filter the items based on the gender of the recipe author: $MP(g|g)$ recommends the items that are most popular among users of the same gender, authored by users of the same gender. We restrict CF in a similar manner: $CF(g)$ recommends items that are most popular among the KNN of users with the same gender. $CF(g|g)$ further restricts the recommendation strategy to items most popular among the KNN of users with the same gender, authored by users of the same gender.

To evaluate the proposed methods, an offline experiment was conducted employing MyMediaLite's 5-fold cross validation evaluation protocol and Mean Average Precision (MAP) – over the whole item list – as accuracy metric. The parameter k for the user-based CF method was set to 80, which delivered the best results within the three presented methods (cross-validated on hold-out data).

6.2 Results

As highlighted in Figure 7, all gender-aware methods improve over the baselines MP ($MAP = .011$) and CF ($MAP = .059$). While the results for $MP(g)$ ($MAP = .015$) and $CF(g)$ ($MAP = .062$) are not so pronounced, more significant differences can be observed when additionally filtering the items (recipes) based on the gender of the author, with $MAP = .026$ for the most popular approach $MP(g|g)$ and $MAP = .086$ for the collaborative filtering method $CF(g|g)$. While the proposed methods are rather basic, the obtained results are insightful, showing the value of "gender filtering" in the context of online food recommender systems. Better results might be obtainable by employing better recommender methods, such as Factorization Machines [21], which we see as possible extensions of this work and out of the scope of this paper.

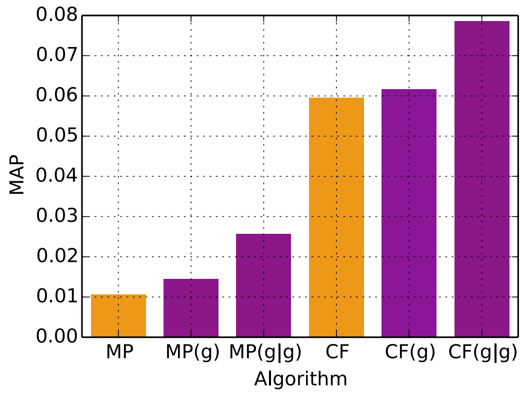

Figure 7: Mean Average Precision (MAP) for the context-blind recommender MostPopular (MP) and User-based Collaborative Filtering (CF) compared to the context-aware methods $MP(g)$, $MP(g|g)$, $CF(g)$ and $CF(g|g)$. As shown, MAP is improving over the baseline methods (highlighted in orange) when user gender (g) and additionally item gender ($g|g$) are considered.

7. CONCLUSIONS

In this paper, we investigated the extent to which differences between men and women in terms of online cooking behavior do exist. While there are many beliefs hinting at significant differences between the two genders in the context of food preparation, no profound, quantitative study has been yet conducted to confirm these prejudices. To contribute to knowledge in this area, we conducted a large-scale empirical study in which we mined and analyzed online traces of almost 200 thousand users and their recipes in one of the largest online cooking platforms available on the Web.

Our results show that there are indeed significant differences between men and women in terms of cooking, arguably even larger than anticipated. For instance, we statistically confirmed that men tend to prepare dishes with more ingredients and a longer preparation time. Women are less inclined to prepare meat dishes and they use spices more subtly than man do. Recipes from males receive more attention and feedback, but females receive more positive and less polarized feedback.

To further reveal the magnitude of the differences and the importance of certain features, we also conduced a classification experiment. Here, we showed in detail the classification power of 88 features – derived from our empirical analysis – to predict the user's gender. Our results reveal that the best features to distinguish between men and women in terms of cooking are food type (sweet dishes for women and meat dishes for men, *H3*), the use of spices (*H4*) and the use of gadgets (*H5*).

Finally, a simple recommender systems experiment shows the usefulness of employing gender as context in the online food recipe recommendation task. Particularly restricting recommendations to recipes from authors of the same gender had a significant impact.

Future Work. One natural extension of this work would be to apply our framework to other online food community platforms to examine cultural differences [16]. Furthermore, one could enhance our classifier framework with a more diverse set of features, such as rating behavior or nutrition values. In addition, it would be interesting to study other user characteristics, such as age or geographic origin of the users. Finally, more extensive experiments in a recommender scenario employing, for example, matrix factorization or learning to rank techniques would provide additional insights on the usefulness of the features studied in this paper.

8. ACKNOWLEDGMENTS

The authors would like to thank Eva Barlösius for fruitful discussions during the study phase.

[12] http://www.mymedialite.net/

9. REFERENCES

[1] S. Abbar, Y. Mejova, and I. Weber. You tweet what you eat: Studying food consumption through twitter. In *Proc. of CHI'15*, 2015.

[2] G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. In *Recommender systems handbook*, pages 217–253. Springer, 2011.

[3] E. Barlösius. *Soziologie des Essens*. Juventa Verlag, 1999.

[4] J. Beel, S. Langer, A. Nürnberger, and M. Genzmehr. The impact of demographics (age and gender) and other user-characteristics on evaluating recommender systems. In *Research and Advanced Technology for Digital Libraries*, pages 396–400. Springer, 2013.

[5] N. Cavazza, M. Guidetti, and F. Butera. Ingredients of gender-based stereotypes about food. indirect influence of food type, portion size and presentation on gendered intentions to eat. *Appetite*, 91:266–272, 2015.

[6] M. De Choudhury and S. S. Sharma. Characterizing dietary choices, nutrition, and language in food deserts via social media. In *Proc. of CSCW '16*, 2016.

[7] A. Dibb-Smith and E. Brindal. Table for two: The effects of familiarity, sex and gender on food choice in imaginary dining scenarios. *Appetite*, 95:492–499, 2015.

[8] J. Freyne and S. Berkovsky. Recommending food: Reasoning on recipes and ingredients. In *Proc. of UMAP'10*, pages 381–386, 2010.

[9] Z. Gantner, S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme. Mymedialite: A free recommender system library. In *Proc. of RecSys'11*, pages 305–308, 2011.

[10] M. Ge, F. Ricci, and D. Massimo. Health-aware food recommender system. In *Proc. of RecSys '15*, pages 333–334, 2015.

[11] M. Harvey, B. Ludwig, and D. Elsweiler. Learning user tastes: a first step to generating healthy meal plans? In *Proc. of LIFESTYLE'12*, page 18, 2012.

[12] M. O. Hill. Diversity and evenness: a unifying notation and its consequences. *Ecology*, 54(2):427–432, 1973.

[13] O. Kucuktunc, B. B. Cambazoglu, I. Weber, and H. Ferhatosmanoglu. A large-scale sentiment analysis for yahoo! answers. In *Proc. of WSDM'12*, pages 633–642. ACM, 2012.

[14] T. Kusmierczyk, C. Trattner, and K. Nørvåg. Temporal patterns in online food innovation. In *Proc. of WWW'15 Companion*, 2015.

[15] T. Kusmierczyk, C. Trattner, and K. Nørvåg. Temporality in online food recipe consumption and production. In *Proc. of WWW'15*, 2015.

[16] P. Laufer, C. Wagner, F. Flöck, and M. Strohmaier. Mining cross-cultural relations from wikipedia-a study of 31 european food cultures. *arXiv preprint arXiv:1411.4484*, 2014.

[17] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts. Understanding variable importances in forests of randomized trees. In *Advances in Neural Information Processing Systems 26*, pages 431–439. Curran Associates, Inc., 2013.

[18] D. Parra-Santander and P. Brusilovsky. Improving collaborative filtering in social tagging systems for the recommendation of scientific articles. In *Proc. of WI'10*, pages 136–142, 2010.

[19] M. J. Pazzani. A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review*, 13(5-6):393–408, 1999.

[20] S. Raichlen. *Man Made Meals*. Workman Publishing Company, 2014.

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

[22] M. Rokicki, E. Herder, and E. Demidova. What's on my plate: Towards recommending recipe variations for diabetes patients. *Proc. of UMAP'15 LBRS*, 2015.

[23] A. Said and A. Bellogín. You are what you eat! tracking health through recipe interactions. In *Proc. of RSWeb'14*, 2014.

[24] M. Svensson, K. Höök, and R. Cöster. Designing and evaluating kalas: A social navigation system for food recipes. *ACM TOCHI*, 12(3):374–400, 2005.

[25] C. Teng, Y. Lin, and L. A. Adamic. Recipe recommendation using ingredient networks. *CoRR*, abs/1111.3919, 2011.

[26] C. Trattner, T. Kusmierczyk, and K. Nørvåg. FOODWEB - studying food consumption and production patterns on the web. *ERCIM News*, 2016(104), 2016.

[27] C. Wagner and L. M. Aiello. Men eat on mars, women on venus? an empirical study of food-images. In *Proc. of WebSci'15 Posters*, 2015.

[28] C. Wagner, P. Singer, and M. Strohmaier. The nature and evolution of online food preferences. *EPJ Data Science*, 3(1):1–22, 2014.

[29] R. West, R. W. White, and E. Horvitz. From cookies to cooks: Insights on dietary patterns via analysis of web usage logs. In *Proc. of WWW'13*, pages 1399–1410, 2013.