Temporality in Online Food Recipe Consumption and Production

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

Christoph Trattner
NTNU, Trondheim, Norway
chritrat@idi.ntnu.no

Kjetil Nørvåg
NTNU, Trondheim, Norway
noervaag@idi.ntnu.no

ABSTRACT

In this paper, we present work-in-progress of a recently started research effort that aims at understanding the hidden temporal dynamics in online food communities. In this context, we have mined and analyzed temporal patterns in terms of recipe production and consumption in a large German community platform. As our preliminary results reveal, there are indeed a range of hidden temporal patterns in terms of food preferences and in particular in consumption and production. We believe that this kind of research can be important for future work in personalized Web-based information access and in particular recommender systems.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering*

Keywords

food; communities; user preferences; temporal aspects

1. INTRODUCTION

Food is a fundamental concept in our daily lives and is one of the most important factors that shapes how healthy we are or how good we feel. Although research on the users' food preferences in online communities and social media has gained recently in popularity (e.g., [1, 4, 5]), only little is yet known about the temporal dynamics of the users' online food preferences, and in particular what we like to consume and produce. Current research in this area relies mostly on access-log data and is ignoring explicit preference mechanisms such as food (=recipe) production and rating information (=consumption). We believe that this is important to study, since explicit user feedback is a fundamental concept for personalization and in particular recommender systems that shape nowadays the way we access information on the Web [2, 3]. To fill this gap in the literature, we have started to mine and analyze a set of online food community platforms, to better understand the hidden temporal dynamics of online food preferences and to come up with new models to improve current recommender mechanisms.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742752 .

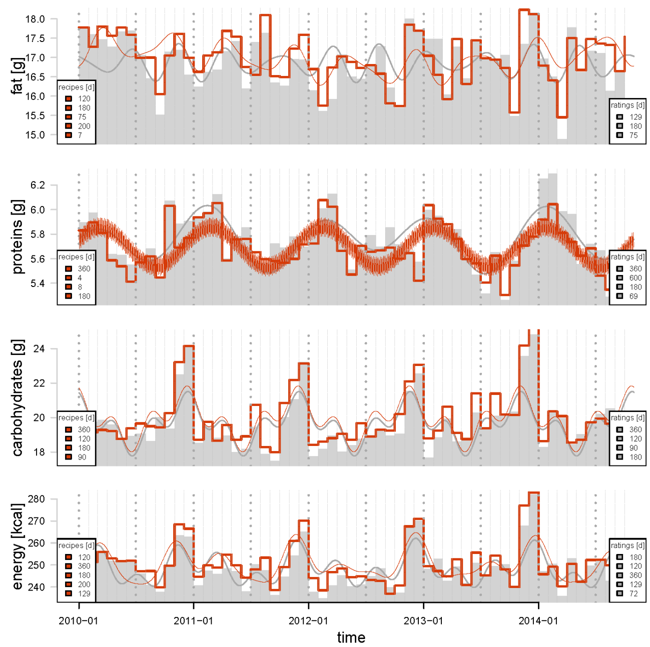


Figure 1: Seasonal trends in online food recipe production (thick contour line) and consumption (background bars) for fat, proteins, carbohydrates and calories per 100g.

2. DATASET

Our study relies on a dataset obtained from the German online food community website kochbar.de[1]; one of the largest of its kind in Europe. The dataset was collected between 2014-11-22 and 2014-12-05, and is available online[2]. Our dataset covers more than 400 thousands recipes published in the years 2010-2014 (170 recipes are published on average per day). Not only information about ingredients and preparation is provided, but also nutrition facts. Recipes are labeled with about 230 categories of 7 classes (here we focus on 4) and were rated by almost 200 thousand users providing 7 million ratings (3300 ratings on average per day). Although people can rate recipes on a likert-scale from 1-5, the community is in general very positive towards the recipes posted by their peers, i.e., we found more than 99% of the recipes are rated with 5 stars. In our analysis, we consider therefore a rating as a Boolean indicator of whether or not a user liked a certain recipe.

[1]http://kochbar.de
[2]Download link available on request.

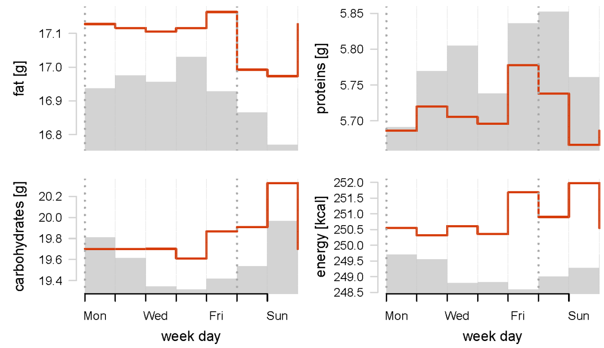
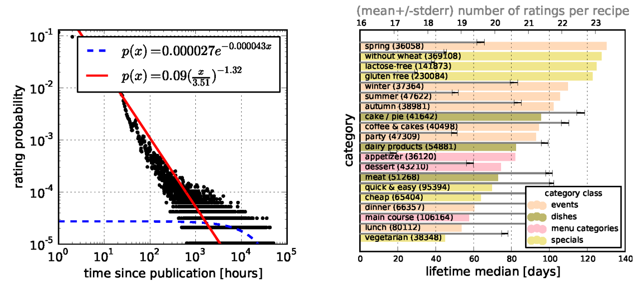


Figure 2: Weekly trends in online food recipe production (contour lines) and consumption (bars) for fat, proteins, carbohydrates and calories per 100g.

3. RESULTS

Figure 1 compares food production (thick contour line) and consumption (background bars) characteristics in terms of the average amount of fat, proteins, carbohydrates and calories. For each factor the most important periods were extracted using spectral density estimation. Corresponding wave plots are plotted with thin continuous lines. Seasonal trends can be easily observed (for all nutrients Kolmogorov-Smirnov (KS) test rejects with $p < 0.001$ the hypothesis that distributions over months is uniform). Half year periods were found in all cases. Similarly to the work of West et al. [5] (focusing on analysis of food consumption temporal trends), one year periods were identified in all plots apart from fat. For carbohydrates and calories significant bursts are observed in November and December. For proteins, bursts are more flat and shifted towards late winter and early spring. What is also interesting to see, is that very short periods (about one week) are dominating there and users are more interested in higher values of protein amount per 100g than what is provided in recipes. In other cases, the gap between what is produced and consumed is the opposite, i.e., the amount of fat, carbohydrates, and calories are higher in production than in consumption. Even though significant differences in means are observed only for calories ($t = -12.67; p < 0.001$) and fat ($t = -2.38; p < 0.05$), distributions of nutrients in production and consumption differ in all cases (KS test: $p < 0.001$). Finally, the plot for fat is the most flat and least seasonal dependent one, indicating that fat food is on the users' preference list in the German speaking world through the year. Similar results but on a weekly basis are presented in Figure 2. Once again KS test rejects with $p < 0.001$ uniformity hypothesis for weekdays. Patterns of different factors in weekly consumption confirms previous results from West et al. [5] only partially. For example, we identify the lowest interest in fatty food in the German speaking world on Sundays whereas in the US people avoid fatty food on Fridays. One possible explanation for this finding is the cultural discrepancy between Europe and USA. However, our analysis is consistent with findings from [4], who also show that weekend food preferences are slightly different from working days.

Apart from the previous analysis which confirmed our hypotheses that online food consumption and production patterns are highly sensitive in time, we were also interested in studying the users' interest shifts over time. Figure 3a presents the distribution of rating probability measured after recipe publication time. As presented, users' interests for newly published recipes follow a power law function, meaning that people rather prefer newly published recipes than those which have been published earlier in time. Since



(a) Fitting exponential (dashed line) and power law (continuous line) models to recipes interest (measured with number of ratings) decay.

(b) Lifetime (thick colored bars) and number of ratings per recipe (thin one-color bars) in 20 the most popular categories (number of recipes assigned provided in brackets).

Figure 3: Recipes lifetimes modeling and distribution among the most popular categories.

this finding was to some extent surprising, we further studied the extent to which this pattern holds true for different types of recipe categories. Figure 3b presents median lifetimes (what is the value for which the fitted cumulative distribution function is equal to 0.9) in the top 20 most used categories from 4 classes (medians instead of means are used to avoid long-tailed recipes influence). We observe that there are indeed different temporal preference patterns in recipe categories showing that *spring* recipes are more persistent in time than recipes for *vegetarians*. What is also provided in this plot, is information about the number of ratings per recipe and number of recipes assigned to those categories. We computed the Spearman rank-order correlation coefficient over 71 categories having assigned more than 1000 recipes. For the comparison of lifetimes and ratings we observe a statistically significant negative correlation with $\rho = -0.62$; $p < 0.001$. Our hypothesis is that more persistent categories earn less attention than these changing faster. On the other hand, no statistically significant correlation was observed between lifetimes and number of recipes published in categories. This implies that lifetime is not driven by production of new recipes but is their internal property.

4. CONCLUSIONS

In this paper we have presented work-in-progress of a recently started research project that tries to understand the temporal hidden dynamics of the users' online food preferences. Based on a large-scale dataset collected from a German food community website, we have shown several hidden temporal patterns in terms of food consumption and production, which we argue play an important role in the context of personalized and time-aware Web-based information access and recommender systems [2].

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

[3] M. Trevisiol, L. Chiarandini, and R. Baeza-Yates. Buon appetito-recommending personalized menus. In *Proc. of HT'14*, 2014.

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

[5] 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*, 2013.