

# What’s On My Plate: Towards Recommending Recipe Variations for Diabetes Patients

Markus Rokicki, Eelco Herder, Elena Demidova

L3S Research Center, Hannover, Germany  
{rokicki,herder,demidova}@L3S.de

**Abstract.** As community-based recipe platforms continue to grow in popularity, recipe recommendation is an active research area. Simultaneously, the analysis of online recipes can provide us with insights on dietary patterns in particular communities. In this paper, we focus on recipe recommendation for a user group that is constrained in terms of choices: diabetes patients need to balance their diet more than average persons and to be aware of the nutritional value of their meals. First, we discuss the type of situations where diabetes-specific food recommendations are desirable. Further, we analyze how people’s age and gender interact with food intake. Based on a large dataset, we explore how variations in ‘canonical meals’ can be exploited for recommending which alternatives better fit the user’s dietary requirements.

## 1 Introduction

Diabetes is a widely spread chronic disease that affects about 10% of the Western population. Patients suffering from diabetes have to take many vital decisions on a daily basis, including: am I allowed to eat this, what is my blood sugar level, how much insulin should I take right now? Particularly those who have just been diagnosed with diabetes experience difficulties facing such decisions.

The GlycoRec project<sup>1</sup> aims to develop a system that provides diabetes patients with personalized support and advices for improving their everyday lives. GlycoRec will provide patients with information and advice regarding their nutrition, physical activities, and the use of medicine. This empowers patients to better communicate their needs with their doctors and advisors, and to better implement advices and stated goals in their everyday lives.

In this paper, we present our first steps towards personalized nutritional advice. Despite the availability of online nutritional information<sup>2</sup>, these databases only provide information on single ingredients and/or standardized products such as ready-made meals. Our goal is to provide diabetes patients with recommendations and feedback in everyday situations, including: (i) How many

---

<sup>1</sup> <https://www.pfh.de/hochschule/forschung/forschungsprojekt-glycorec.html>

<sup>2</sup> The quality and acceptance of databases with nutritional information vary wildly. In Germany, an established resource is <http://www.mri.bund.de/de/service/datenbanken/bundeslebensmittelschlüssel.html>.

carbohydrates can I expect that the Thai curry on the menu contains?, and (ii) Which recipe variation best matches both my dietary restrictions and my taste?

Based on an extensive dataset from the German recipe website Kochbar.de<sup>3</sup>, we show that there are *differences in eating patterns* with respect to gender, age group, and dietary restrictions, such as diabetes. We cluster popular meal names into ‘*canonical meals*’ and analyze to what extent these meals *vary in terms of ingredients and nutritional value*. These insights provide directions for strategies to find the best recipe or for adapting recipes to user needs and preferences.

## 2 Related Work

Diabetes mellitus is a widespread disease that requires constant attention of the patients and their caregivers. Research has shown that effective prevention of diabetes-related complications includes lifestyle changes that include an increased physical activity as well as a diet that is associated with lower blood pressure [6]. Telemedicine and the improved acceptance of smart phones and tablet computers among patients and physicians, contributes to improved patient guidance and self-empowerment [7]. A particular focus of guidance is nutrition.

Online recipe recommenders can play an important role in generating healthy meal plans. Even though the used ingredients are the major reason for liking or disliking a meal, there are health-conscious users who also take nutritional information into account [3]. 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 [2].

An in-depth analysis on how users choose and adapt recipes is given by Teng et al [8]. 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. analyzed data from the German community platform Kochbar and found clear seasonal and weekly trends in online food recipe production, both in terms of nutritional value (fat, proteins, carbohydrates and calories) [5] and in terms of ingredient combinations and experimentation [4]. West et al. [9] analyzed similar patterns for the American population, with slightly different results. They were also able to automatically detect anomalous days - bank holidays and other celebrations - and users who aimed to change their diet.

Making use of these and other insights, the team from IBM Watson created the prototype Chef Watson, which automatically creates recipes that match user preferences, based on existing recipes from the Bon Appetit recipe website [1].

## 3 Dataset and Preprocessing

We use a crawl from Kochbar.de, a German online food community website to which users can upload and rate cooking recipes, provided by Kusmierczyk et

---

<sup>3</sup> <https://www.kochbar.de>

al. [4]. The dataset encompasses more than 400 thousand recipes published between 2008 and 2014. In addition to information on ingredients and preparation, more than 330 thousand recipes also contain nutrition facts. Almost 200 thousand users provided more than 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).

We consider only the 309 thousand recipes that contain valid information on energy (in kJ and kcal), carbohydrates, proteins, and fat. For each recipe, a mostly structured list of ingredients (with quantities) is given. We extracted the ingredients and performed simple normalization by converting the text to lowercase, normalizing whitespaces, removing text in parentheses, and splitting on conjunctions such as “and” and “or”. This process yields more than 300 thousand ingredients, with an average of 10 ingredients per recipe. As only a moderate amount of 2258 ingredients occurs frequently (in at least 100 recipes), this simple preprocessing is sufficient for a first analysis of the dataset.

## 4 Differences in Food Intake

As a first step, we aim to identify differences in user recipes created by different user groups. Among the 200 thousand users, 95 thousand provided information regarding gender (25 thousand male, 70 thousand female users) and 57 thousand provided information regarding their age (mean 42.2, median 42). In addition, we are interested in diabetes patients. However, apparently most patients did not disclose the fact that they suffer from diabetes: from the profile information we are able to identify only 65 users who suffer from diabetes or have close relatives with diabetes - a number too small for an effective analysis. Similarly, only about 3.000 of the recipes have been labeled as ‘diabetes friendly’. By contrast, about 220.000 recipes are marked as gluten free, 137.000 as lactose free and (only) 37.000 recipes as vegetarian.

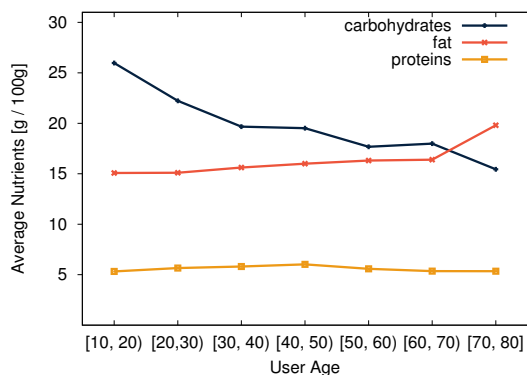


Fig. 1: Average nutritional facts of recipes for different user ages as given by the recipe authors.

Figure 1 shows the average nutritional values of recipes created by different age groups. The levels of fat and protein are quite stable between age groups, but the amount of carbohydrates consistently decreased with age ( $F = 152$ ,  $p < 0.01$ ). Reduction of carbohydrates is a recommendation provided by most nutrition advice centers.

Figure 2 shows the average nutritional values of the recipes provided by different user groups. The most noticeable difference can be observed for carbohydrates: recipes provided by female users are, on average, richer in carbohydrates than recipes provided by male users ( $t=52.108$ ,  $p < 0.01$ ). As most users in Kochbar are female, this is reflected in the averages for all users. There are two explanations for this effect: first, baking recipes are typically written by women; Second, male users are on average older (50.9 years) than the female users (43.8 years) in our dataset ( $t=5.629$ ,  $p < 0.01$ ) - as mentioned earlier, the average intake of carbohydrates decreases with age (Figure 1).

The two rightmost bars in Figure 2 correspond to the recipes provided by self-reported diabetes patients and recipes that are marked as diabetes friendly. Recipes of the self-reported diabetes patients are clearly lower in fat ( $t=2.991$ ,  $p < 0.01$ ) and contain more protein than recipes from other users ( $t = 5.629$ ,  $p < 0.01$ ) - which is in line with recommendations from diabetes information centers. By contrast, recipes marked as ‘diabetes friendly’ seem not to differ from regular recipes.

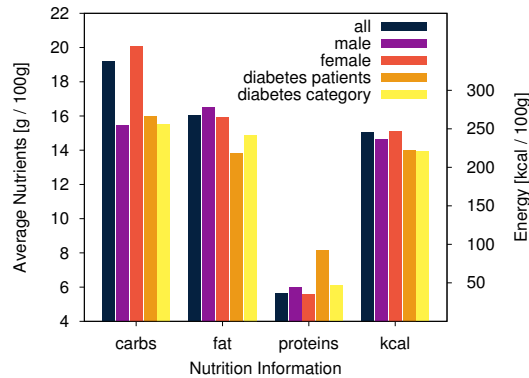


Fig. 2: Average nutritional facts of recipes for different user groups and for the recipes assigned to the category “diabetes”.

## 5 Canonical Meals and their Variations

As discussed by [2], most users do not search for recipes based on their nutritional values, but rather look for recipes with certain ingredients or for a particular dish. As is to be expected, many variations of popular recipes can be found at Kochbar.de. In order to find out differences in nutritional value within a

particular type of dish, we selected the 200 most frequently used recipe titles as ‘canonical meals’, to which we assigned all recipes of which the title contained the title of the canonical meal. The ‘top meals’ are shown in Table 1. From the selection one can clearly see that the user base of Kochbar.de is German.

| Canonical Meal                | Recipes | Ratings | Comments |
|-------------------------------|---------|---------|----------|
| Kartoffelsalat (potato salad) | 1,863   | 41,226  | 14,117   |
| Pizza                         | 1,812   | 38,250  | 13,210   |
| Käsekuchen (cheese cake)      | 1,681   | 36,807  | 12,085   |
| Apfelkuchen (apple pie)       | 1,450   | 34,935  | 11,982   |
| Gulasch (gulasch)             | 1,187   | 28,668  | 10,162   |
| Nudelsalat (pasta salad)      | 1,706   | 30,196  | 9,782    |
| Eierlikör (egg liquor)        | 1,085   | 27,482  | 9,360    |
| Pfannkuchen (pancake)         | 1,221   | 24,492  | 8,550    |
| Lasagne (lasagna)             | 1,187   | 22,570  | 8,034    |
| Tiramisu                      | 1,358   | 22,931  | 7,773    |

Table 1: Canonical meals of which the recipes received most comments.

To find out to what extent canonical meals vary in nutritional value, we selected three different, representative meals and calculated the means and standard deviations - see Table 2. We also analyzed which ingredients are associated with recipes that are high and low in carbohydrates, fat, proteins, and energy - by calculating the average levels for all meals and sorting them accordingly.

Potato salad is low in protein, but the standard deviations are relatively high. Ingredients associated with high protein are meat and fish, low-protein recipes contain vegetables instead, such as pickles, radish, olives and asparagus. The same pattern can be found for lasagna. Low-fat cheese cake is associated with low-fat milk products and high-fat with chocolate and cream cheese.

| Meal         | Carbohydrates |         | Fat   |         | Proteins |         | Kcal   |         |
|--------------|---------------|---------|-------|---------|----------|---------|--------|---------|
|              | Mean          | Std dev | Mean  | Std dev | Mean     | Std dev | Mean   | Std dev |
| Potato salad | 10.26         | 8.025   | 13.09 | 17.64   | 3.27     | 3.60    | 171.25 | 151.17  |
| Cheese cake  | 29.93         | 15.68   | 11.97 | 9.29    | 7.32     | 2.97    | 256.29 | 95.96   |
| Lasagna      | 8.44          | 10.81   | 13.23 | 11.00   | 8.06     | 6.18    | 185.43 | 137.25  |

Table 2: Nutritional facts for 3 popular canonical meals.

These findings confirm our expectation that it is possible to estimate the nutritional value of certain dishes from the recipes associated with these dishes and that one can identify ingredients associated with high or low levels of food value. This provides a good base for developing food recommender systems that suggest ‘better recipes’ and alternative ingredients for particular recipes.

## 6 Discussion and Outlook

Existing food databases for diabetes patients are incomplete and inconsistent. Moreover, they only contain ingredients and standardized products. Diabetes patients typically learn over time what they can eat and what not, but this is often not sufficient for many common situations. In restaurants, served meals are often ‘black boxes’ in terms of nutritional value, which causes great uncertainty among diabetes patients, especially when trying out new meals while on vacation in foreign countries.

To provide patients with advice and information on the nutritional value of a meal - and to recommend them alternative meals or ingredients - we aim to exploit recipe sites such as Kochbar.de, which contain user-provided recipes. This paper provides some first insights and confirms the feasibility of the approach.

A particular challenge for the food recommender will be to provide detailed feedback on the precision of its estimations and the resulting recommendations. Particularly for ‘canonical meals’ with many variations, estimations may need to be refined with additional user input and feedback.

## Acknowledgment

The GlycoRec project is funded by the Federal Ministry of Education and Research (BMBF) under the funding scheme Adaptive, Learning Systems (Adaptive, lernende Systeme).

## References

1. FIRTH, N. Cooking by numbers. *New Scientist* 225, 3003 (2015), 19–20.
2. FREYNE, J., AND BERKOVSKY, S. Recommending food: Reasoning on recipes and ingredients. In *User Modeling, Adaptation, and Personalization*. 2010, pp. 381–386.
3. HARVEY, M., LUDWIG, B., AND ELSWEILER, D. Learning user tastes: a first step to generating healthy meal plans? In *First International Workshop on Recommendation Technologies for Lifestyle Change (LIFESTYLE 2012)* (2012), p. 18.
4. KUSMIERCZYK, T., TRATTNER, C., AND NØRVÅG, K. Temporal patterns in online food innovation. In *5th Temporal Web Analytics Workshop (TempWeb) at WWW 2015*. (2015).
5. KUSMIERCZYK, T., TRATTNER, C., AND NØRVÅG, K. Temporality in online food recipe consumption and production. In *Proc. of WWW* (2015), vol. 15.
6. LEHMANN, R., AND SPINAS, G. Screening, diagnostik und management von diabetes mellitus und diabetischen folgeerkrankungen. *Therapeutische Umschau* 57, 1 (2000), 12–21.
7. SCHILDT, J., AND MERTENS, H. Chronic care management of diabetes mellitus—telemedicine as option in a changing supply situation with general practitioners (gp). *Diabetes aktuell für die Hausarztpraxis* 10, 06 (2012), 262–268.
8. TENG, C., LIN, Y., AND ADAMIC, L. A. Recipe recommendation using ingredient networks. *CoRR abs/1111.3919* (2011).
9. WEST, R., WHITE, R. W., AND HORVITZ, E. From cookies to cooks: Insights on dietary patterns via analysis of web usage logs. In *Proc. 22nd Conf. World Wide Web* (2013), pp. 1399–1410.