An Evaluation of Recommendation Algorithms for Online Recipe Portals

Christoph Trattner
University of Bergen
Norway
christoph.trattner@uib.no

David Elsweiler
University of Regensburg
Germany
david.elsweiler@ur.de

ABSTRACT

Better models of food preferences are required to realise the oft touted potential of food recommenders to aid with the obesity crisis. Many of the food recommender evaluations in the literature have been performed with small convenience samples, which limits our confidence in the generalisability of the results. In this work we test a range of collaborative filtering (CF) and content-based (CB) recommenders on a large dataset crawled from the web consisting of naturalistic user interaction data over a 15 year period. The results reveal strengths and limitations of different approaches. While CF approaches consistently outperform CB approaches when testing on the complete dataset, our experiments show that to improve on CF methods require a large number of users (> 637 when sampling randomly). Moreover the results show different facets of recipe content to offer utility. In particular one of the strongest content related features was a measure of health derived from guidelines from the UK Food Safety Agency. This finding underlines the challenges we face as a community to develop recommender algorithms, which improve the healthfulness of the food people choose to eat.

KEYWORDS

Online recipes; recommender systems

ACM Reference Format:
Christoph Trattner and David Elsweiler. 2019. An Evaluation of Recommendation Algorithms for Online Recipe Portals. In *Proceedings of the 4th International Workshop on Health Recommender Systems co-located with 13th ACM Conference on Recommender Systems (HealthRecSys'19) (HealthRecSys '19)* , 5 pages.

1 INTRODUCTION

Food recommenders (e.g. [11, 15]) and studies of online recipes (e.g. [24, 40]) have received increased research attention of late. A key motivation for this is often health, with recommender systems being touted as a means to help people change dietary habits and address costly societal problems, such as diabetes and obesity [7, 11].

Diverse studies have been published, offering insight into the contextual factors influencing recipe preference [28, 40] and the future popularity of recipes [36], as well as providing an understanding of the links between recipe preference and incidence of eating related illness [37]. A further strain of research has attempted to incorporate health in the food recommendation problem by, for example, building nutritional content into the recommendation process [15, 19, 34] or by recommending meal plans, which tailor recommendations to users' nutritional needs over time [6].

Providing healthful food recommendations, using any of the suggested strategies necessitates, however, that we can accurately model and predict the food individual users would actually like to eat. We have yet limited understanding as to which recommender algorithms work best [33] and the studies that have been performed typically focus on one approach in isolation (e.g. recipe ingredients [11] or properties of the associated image [14]). Moreover, past work has tended to employ datasets derived from small scale user studies [11, 19] limiting our confidence in the generalisability of the results. In this work, we test a number of competitive collaborative filtering (CF) and content-based (CB) recommenders on a large scale naturalistic dataset similar to those that have been studied for cultural [24, 40] or epidemiological [37] reasons using data science methods. We formulate the problem as is typically done in recommendation experiments using past feedback from a given user to predict future interactions by that same user [26]. The aim being not only to compare and contrast different models, but also to examine the utility of different facets of content - which are diverse in the case of online recipes - and establish how these influence the recommendation performance. The main findings include that:

- CF methods consistently outperform CB methods over the full dataset.
- CF requires either a small number of highly active users or over six hundred users, selected randomly to achieve competitive performance.
- There is a useful signal in the CB facets, which would be useful in cold-start situations.
- One of the most robust content features is the nutritional healthiness of the recipe as defined by a measure derived from the United Kingdom Food Standards Agency (FSA). This highlights that users are typically consistent in their nutritional preferences over time and emphasizes the challenges faced to change eating habits.

The remainder of the paper is structured as follows: Sections 3 and 4 describe the data basis and experimental setup, respectively. Section 5 continues to report the results of two rounds of experiments, the first of which uses the full dataset and the second employs a bootstrapping approach to test algorithms on sub-samples of the data of various sizes. Section 6 summarises the findings and sets these in context against the literature, which is reviewed in the following section.

HealthRecSys '19, September 20, 2019, Copenhagen, Denmark

2 RELATED WORK

In this section two bodies of related work are reviewed. The first focuses on the evaluation of food recommender algorithms. The second summarises studies of user interaction with online recipe portals, which provides insight into human food preference and the variables influencing this.

2.1 Food Recommendation

Efforts to design automated systems to recommend meals can be traced to the mid-1980s where case-based planning was employed [18, 21]. More recent efforts have focused on rating prediction, using either aspects of recipe content or ratings data using collaborative filtering approaches. Freyne et al. [11] showed the recommendations could be improved by decomposing recipes into individual ingredients and building user profiles comprising ingredients users liked based on ratings for the recipes containing these ingredients. Harvey et al. extended the approach and improved performance by creating positive and negative profiles for users and reducing the dimensionality of the matrices [19].

Other CB approaches have employed visual signals. Yang and colleagues demonstrated that algorithms designed to extrapolate important visual aspects of food images outperform baseline methods [42, 43]. Elsweiler et al. [8] also show that automatically extracted low-level image features, such as brightness, colourfulness and sharpness can be useful for predicting user food preference.

A second approach has been to exploit ratings data using collaborative filtering (CF) techniques. Freyne and Berkovsky tested a nearest neighbour approach, which offered poorer performance than the content approach described above [11]. Ge et al. [15] tested a matrix factorization solution that fuses ratings information and user supplied tags to achieve significantly better prediction accuracy than content-based and standard matrix factorization baselines. Several studies report that the best results are achieved when CF and CB approaches are combined in hybrid models [11, 14, 19].

A common motivator for food recommendation work has been to promote healthy nutrition. One approach is to rely on rules derived from domain experts to meet daily energy requirements [13] or focus on the nutritional requirements of specific groups such as the elderly care [10] or body-builders [38]. Others have tailored recommendations based on the user's calorific or other nutritional needs [15, 16, 34], existing nutritional habits [31] or combine recommendations to meet requirements [6]. Again, approaches have been published for specific target groups e.g. diabetics [25].

2.2 Studies of Food Behaviour using Online Recipe Portals

While not focusing on recommendation, a large body of recent work sheds light on food preferences by studying interactions with online food portals. Analysing the nutritional content of these portals using metrics derived from the World Health Organisation (WHO) and the United Kingdom Food Standards Agency (FSA) has found recipes to be mainly unhealthy, although healthy recipes can be found [35]. Overall, people tend to interact with the least healthy recipes most often [34]. There is, nevertheless, heterogeneity in the user-base with respect to the nutritional properties of recipes interacted with and a growing body of evidence reports correlations between recipes accessed via search engines, recipes portals and social-media and incidence of diet-related illness [1, 3, 29, 37]. Moreover, clear weekly and seasonal trends can be observed in the way users interact with recipes, both in terms of the contained ingredients and the nutritional value of the recipes (fat, proteins, carbohydrates, and calories) [23, 40]. Other work has reported different interaction patterns for users with different gender [28, 39] and who live in different geographical areas within a country [40, 44]. The number of variables shown to relate to eating habits highlights just how challenging a problem food recommendation is.

Table 1: Basic statistics of the Internet recipes dataset obtained from Allrecipes.com.

Total published recipes	60,983
Recipes containing nutrition information	58,263
Recipes rated	46,713
Ratings	1,032,226
Users providing ratings	125,762

The brief review of literature above has highlighted the increasing popularity of food recsys research and that a key motivator is desire to build systems to promote healthy nutrition. Key takeaways from the review are as follows:

- While several evaluations have CF and CB baselines, no extensive comparison of CF and CB approaches in food recsys domain has been published.
- Moreover, no detailed investigation of different aspects of content that may be useful is available and much of the recipe content (recipe description, cooking steps, cooking time etc.) has not been evaluated.
- Finally, the evaluations performed to date have typically been performed on small artificially generated test collections.

3 MATERIALS

To address the identified gaps in the literature, in this work, we make use of a web crawl of the online platform Allrecipes.com to evaluate diverse CF and CB approaches in the recipe recommendation context.

The platform was crawled between 20^{th} and 24^{th} of July, 2015. We retrieved 60,983 recipes published by 25,037 users between the years 2000 and 2015 through the sitemap that is available in the robots.txt file of the website. In this paper we only make use of the 58,263 recipes where nutrition information was available. The basic statistics of this dataset can be found in Table 1.

In addition to the core recipe components – such as recipe title, ingredient list, number of servings and instructions – we also collected for each recipe the according image, comments provided by users, rating information and nutrition facts[1], such as total energy (kCal), protein (g), carbohydrate (g), sugar (g), salt (g), fat (g) and saturated fat (g) content (measured in 100g per recipe).

[1] Allrecipes.com estimates the nutritional facts for an uploaded recipe by matching the contained ingredients with those in the ESHA research database [9]. The ESHA system is used by popular companies such as MCDonald's and Kellogs.

Allrecipes.com is just one of many online recipe portals. Others popular sites include Food.com, Epicurious.com, Yummly.com and Cooks.com. We chose Allrecipes.com because, at the time of writing, it claims to be the world's largest food-focused social network: the site has a community of over 40 million users from 24 countries who annually visit 3 billion recipes [2]. This claim has been corroborated by services such as eBizMBA, which ranks Allrecipes.com as the most popular recipe website [5]. This means that we not only analyze a large scale dataset, but also the most popular recipe platform on the Web.

4 EXPERIMENTAL SETUP

We ran a series of experiments evaluating the performance of 6 prominent recommender algorithms on the rating data using the LibRec[2] framework. The algorithms tested are: Random item ranking (our baseline), Most Popular item ranking (MostPopular), user- and item-based collaborative filtering (denoted as UserKNN and ItemKNN) [30], Bayesian Personalized Ranking (BPR) [26], Weighted matrix factorization (WRMF) [22] and Latent Dirichlet Allocation (LDA) [17].

For the content-based approaches we induced in total 20 different features, which we used to compute similarities between recipes. Below we briefly summarise these features and their corresponding sets:

- *Title:* For the title feature set, we derived 5 similarity features, based on Levenshein distance, Least Common Sub-Sequence (LCS), Jaro-Winkler distance and bi-gram distance. To obtain a similarity value between two recipes based on these features we calculate $1 - dist(r_i, r_j)$. Furthermore, we employ LDA topic modelling on the recipe titles using Mallet with Gibbs sampling. The number of topics was set to 100 topics. Hence for each recipe we induce a vector of dimension one hundred capturing the topic distribution. To calculate similarities between recipes we employ the cosine similarity metric.
- *Image:* For the image feature set we employed on the one hand side image attractiveness measures such as image brightness, sharpness, contract, colorfulness and entropy as well as deep convolutional neural network (CNN) features from a pre-trained VGG-16 model [32]. For each image we derive one embedding vector of dimension 4096 and calculate cosine similarity between recipes on these vectors. To measure the similarity between two recipes based on the image attractiveness metrics [36] we employ the Manhatten distance, i.e. $1 - |metric(r_i) - metric(r_j)|$.
- *Ingredients:* To calculate similarities between recipes on ingredient level, we inducted four different features. On the one hand side the text itself was used and brought to a TF–IDF representation to calculate cosine similarity between recipes. On the other hand side we also chose to employ LDA again to derive a topic distribution and to calculate cosine similarity between recipes on those vectors. Finally, we employed the normalized ingredient strings, to calculate similarities between recipes using cosine similarity and Jaccard. In the case of cosine we normalized the quantities of each ingredient to 100g of a recipe and used the normalized quantity values as frequency indicator.
- *Directions:* From the directions block we computed two similarity features based again on a LDA topic vector representation of the text as well as on TF–IDF vector representation. Similarities were again computed employing the cosine similarity measure on these vectors.
- *Ratings:* Here we rely on the the number of ratings of a recipe as well the average rating. To compute similarities between recipes on theses indicators we rely again on the inverse Manhatten distance, i.e. $1 - |metric(r_i) - metric(r_j)|$.
- *Health:* In order to measure healthiness of a recipe we rely on the following macro nutrient: 'fat', 'saturated fat', 'sugar' and 'salt' (measured in 100g per recipe). This allows us to measure the healthiness of a recipe according to international standards as introduced in 2007 by The Food Standard Agency (FSA) [12]. There are also other standards that can be applied, such as the ones provided by the World Health Organization (WHO) [41] or the HEI metric as proposed by the CDC [20]. We employ the standards provided by the FSA, as this is currently most robust method to estimate the healthiness of online recipes. The metric was also used in related work [34]. The scale ranges from 4 for very healthy recipes to 12 for very unhealthy recipes. Throughout the paper we refer to this metric as 'FSA score'.

For each of the features described above, we derive a scoring function that computes as follows:

$$score(u,i)_{feature} = \frac{\sum\limits_{p \in P_u} sim(i,p)}{|P_u|}, \quad (1)$$

where P_u is the set of items of a user u, i an arbitrary item, and $sim(i,p)$ is any of the above mentioned similarity metrics between item i and p.

For each feature set we calculate scores based on the linear combination of the similarities[3].

As in previous work [26], we operationalise the experiments as a personalized ranking problem (item recommendation). The aim here is to provide a user with a ranked list of items where the ranking has to be inferred from the implicit behavior of the user (e.g. recipes rated in the past). Implicit feedback systems, such as those studied in [26] are challenging as only positive observations are available. The non-observed user-item pairs – e.g. a user has not cooked a recipe yet – are a mixture of real negative feedback (the user is not interested in cooking the recipe) and missing values (the user might want to cook the recipe in the future). We use 5-fold cross validation as protocol for all the experiments and report the recommendation performance results employing AUC as a performance metric [27].

To reduce data sparsity issues, a well-known issue in collaborative filtering-based methods [27], in the first experiments we apply a p-core filter approach [4] using only user profiles with at least 20 rating interactions[4] and recipes that have been rated at least 20 times by the users, resulting in a final dense dataset comprising 1273 users, 1031 items and 50,681 interactions. To study the effects of different levels of users on performance we report a second set

[2]http://www.librec.net/

[3]Parameters were tuned to the optimum using grid search.

[4]We transfer all ratings to positive feedback, i.e. any rating is counted as positive feedback and any none interaction as negative feedback. This makes sense as 95% of all ratings in the Allrecipes.com dataset are 5-star ratings, see also [36].

Table 2: Results of the recommender experiment – collaborative (CF) vs content-based (CB) – in the dense data sample with all users. Best features in each set (CF and CB) are bolded. Top-5 (↑) and Bottom-5 (↓) single content features are also marked.

Method	Algorithm	AUC
CF	BPR	**.7094**
	WRMF	.6881
	UserKNN	.6962
	ItemKNN	.6909
	MostPopular	.6864
	LDA	.6863
CB	Title:Levenstein-Distance	.5468 (↑)
	Title:Bigram-Distance	.5500 (↑)
	Title:LCS-Distance	.5424
	Title:LDA-Text-Cosine	.5353
	Title:Jaro-Winkler-Distance	.5324
	Title:All	.5523
	Image:Cosine-Embeddings	.5322
	Image:Colorfulness-Distance	.5072 (↓)
	Image:Contrast-Distance	.5175
	Image:Sharpness-Distance	.5109
	Image:Entropy-Distance	.5080 (↓)
	Image:Brightness-Distance	.4991 (↓)
	Image:All	.5425
	Ingredients:Cosine-Text	.5547
	Ingredients:Cosine-LDA-Text	.5653 (↑)
	Ingredients:Jaccard	.5502
	Ingredients:Cosine	.5575
	Ingredients:All	.5718
	Directions:Cosine-LDA-Text	.5606 (↑)
	Directions:Cosine-Text	.5210
	Directions:All	.5731
	Ratings:Number-Distance	.4789 (↓)
	Ratings:Average-Distance	.4832 (↓)
	Ratings:All	.5249
	Health:FSA	.5775 (↑)
	CB:All	**.5883**
	Random	.4989

of bootstrapped experiments using smaller dense samples of heavy users (using the same criteria as above), and varying collection sizes using standard random sampling, referred to as 'sparse samples' in the text. These experiments were repeated 100 times each and the average performance reported.

5 RESULTS

The results of the experiments on the full dataset are shown in Table 2. The CF methods clearly outperform the content-based approaches. The best performing CF method (BPR) achieved an AUC score of .7094 and the remaining CF methods demonstrated AUC scores of > .686. This compares to .5883 achieved by the linear combination of content features (= CB:All).

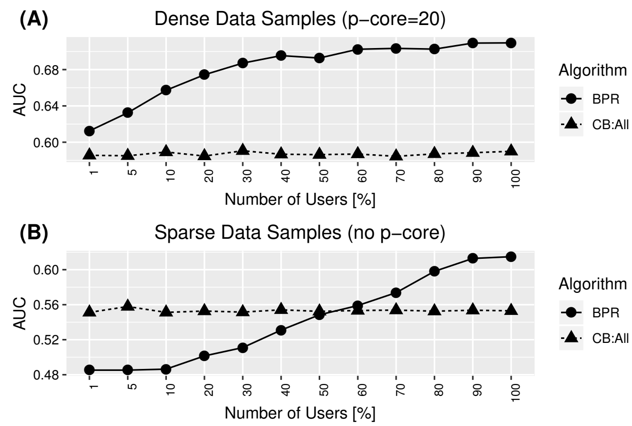

Figure 1: (A) shows the results in the dense data samples (= p-core filtered) where each user has at least 20 item interactions and each item is at least 20-times interacted with, (B) shows the results in the sparse data samples (=no p-core).

Examining the performance of different aspects of content (title, image, ingredients, direction and health) shows that there is a signal in each of these aspects. This is a sign of the consistency, in terms of the properties of recipes, which individual users tend to rate. The fact that the combined model "All" does not achieve a high improvement on these signals individually is perhaps an indication that a linear combination is not the best means to combine these signals. One of the strongest content-based features is the FSA score (AUC=.5775). Again, this hints at consistency in user preference, this time in terms of the healthiness of recipes, which individual users interact with.

To complement these initial results and better understand the relationship between CF and CB methods and the amount of data required to achieve strong recommendation performance with these approaches, we performed the bootstrapping study as described above. The results are presented in Figure 1.

In a first test, see Figure 1 (A), we sampled only from active users, that is, we derived a test size of various sizes where users had rated at least 20 items and the items involved had also achieved at least 20 ratings. Taking this dense sample showed that even a small number of users can attain stable performance. With only 1% of all users (N=13) the CF technique (BPR) is able to outperform the content approach. Nevertheless, when users are selected at random from the dataset and no p-core filter is applied, see Figure 1 (B) – which we argue is a much more realistic setup [4] – many more users are required on average to achieve an equivalent performance. Whereas the CB approaches achieve a consistent performance (AUC=> .54) regardless of the number of users studied, half of the dataset (50%, N=637) is required before the CF methods outperform the CB approach.

6 SUMMARY & CONCLUSION

In this work we have tested competitive recommendation algorithms on a large online recipe dataset. While algorithms of these types have been evaluated before (e.g. [11, 19]), no systematic evaluation has been performed on naturalistic data of this type for only recipes and no results have been published with respect to what signal can be offered by different facets of recipe content.

Our primary finding is that CF outperformed CB in our experiments. This is a different result from the literature - both [11] and [19] report ingredient based CB methods outperforming CF baselines. The small size datasets in these past studies, however, suggests the results to be compatible. It is only after data for several hundred (in our experiment 637) users is available that CF methods start to outperform CB.

With respect to recipe content, the performance of FSA highlights the challenge in changing people's habits. This aligns with past work revealing that the majority of users tend to prefer unhealthy food, a smaller group preferred healthy recipes, but both groups were consistent in their judgments over time [19]. As a community we need to think hard about how these group members can be targeted with recommendations that might alter this situation.

REFERENCES

[1] Sofiane Abbar, Yelena Mejova, and Ingmar Weber. You Tweet What You Eat: Studying Food Consumption Through Twitter. In *Proc. of CHI '15*.

[2] Allrecipes. 2016. Allrecipe.com Press report. available at http://press.allrecipes.com/. Last accessed on 22.03.2019. (2016).

[3] Munmun De Choudhury and Sanket S Sharma. Characterizing Dietary Choices, Nutrition, and Language in Food Deserts via Social Media. In *Proc. of CSCW '16*.

[4] Stephan Doerfel, Robert Jäschke, and Gerd Stumme. 2016. The role of cores in recommender benchmarking for social bookmarking systems. *ACM Transactions on Intelligent Systems and Technology (TIST)* 7, 3 (2016), 40.

[5] Ebizma. 2017. Ebizma rankings for recipe websites. available at http://www.ebizmba.com/articles/recipe-websites. Last accessed on 22.03.2019. (2017).

[6] David Elsweiler and Morgan Harvey. Towards automatic meal plan recommendations for balanced nutrition. In *Proc. of RecSys '15*. 313–316.

[7] David Elsweiler, Morgan Harvey, Bernd Ludwig, and Alan Said. Bringing the "healthy" into Food Recommenders. In *Proc. of DRMS '15*. 33–36.

[8] David Elsweiler, Christoph Trattner, and Morgan Harvey. Exploiting Food Choice Biases for Healthier Recipe Recommendation. In *Proc. of SIGIR '17*. 575–584.

[9] ESHA. 2016. Nutrition Labeling Software. available at http://www.esha.com/. Last accessed on 22.03.2019. (2016).

[10] Vanesa Espín, María V Hurtado, and Manuel Noguera. 2016. Nutrition for Elder Care: a nutritional semantic recommender system for the elderly. *Expert Systems* 33, 2 (2016), 201–210.

[11] Jill Freyne and Shlomo Berkovsky. Intelligent Food Planning: Personalized Recipe Recommendation. In *Proc. of IUI '10*. 321–324.

[12] FSA. 2016. Guide to creating a front of pack (FoP) nutrition label for pre-packed products sold through retail outlet. available at https://www.food.gov.uk/sites/default/files/multimedia/pdfs/pdf-ni/fop-guidance.pdf. Last accessed on 22.03.2019. (2016).

[13] Dhomas Hatta Fudholi, Noppadol Maneerat, and Ruttikorn Varakulsiripunth. 2009. Ontology-based daily menu assistance system. In *Proc. of ECTICON '09*. 694–697.

[14] Xiaoyan Gao, Fuli Feng, Xiangnan He, Heyan Huang, Xinyu Guan, Chong Feng, Zhaoyan Ming, and Tat-Seng Chua. 2018. Visually-aware Collaborative Food Recommendation. *arXiv preprint arXiv:1810.05032* (2018).

[15] Mouzhi Ge, Francesco Ricci, and David Massimo. Health-aware Food Recommender System. In *Proc. of RecSys '15*. 333–334.

[16] Elizabeth Gorbonos, Yang Liu, and Chính T Hoàng. 2018. NutRec: Nutrition Oriented Online Recipe Recommender. In *Proc. of WI '18*. 25–32.

[17] Tom Griffiths. 2002. Gibbs sampling in the generative model of latent dirichlet allocation. (2002).

[18] Kristian J Hammond. 1986. CHEF: A Model of Case-based Planning.. In *AAAI*. 267–271.

[19] Morgan Harvey, Bernd Ludwig, and David Elsweiler. 2013. You are what you eat: Learning user tastes for rating prediction. In *International Symposium on String Processing and Information Retrieval*. Springer, 153–164.

[20] HEI. 2016. Healthy Eating Index. (Oct. 2016). https://www.cnpp.usda.gov/healthyeatingindex

[21] Thomas R Hinrichs. 1989. Strategies for adaptation and recovery in a design problem solver. In *Proc. of Workshop CBR '89*. 115–118.

[22] Yifan Hu, Yehuda Koren, and Chris Volinsky. Collaborative filtering for implicit feedback datasets. In *Proc. of ICDM'08*. 263–272.

[23] Tomasz Kusmierczyk, Christoph Trattner, and Kjetil Nørvåg. Temporality in online food recipe consumption and production. In *Proc. of WWW '15*.

[24] Tomasz Kusmierczyk, Christoph Trattner, and Kjetil Nørvåg. 2016. Understanding and Predicting Online Food Recipe Production Patterns. In *Proc. of HT '16*. 243–248.

[25] Maiyaporn Phanich, Phathrajarin Pholkul, and Suphakant Phimoltares. 2010. Food recommendation system using clustering analysis for diabetic patients. In *Proc. of ISA '10*. 1–8.

[26] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. BPR: Bayesian personalized ranking from implicit feedback. In *Proc. of UIAI'09*. 452–461.

[27] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2011. *Introduction to recommender systems handbook*. Springer.

[28] Markus Rokicki, Eelco Herder, Tomasz Kusmierczyk, and Christoph Trattner. Plate and Prejudice: Gender Differences in Online Cooking. In *Proc. of UMAP '16*. 207–215.

[29] Alan Said and Alejandro Bellogín. You are What You Eat! Tracking Health Through Recipe Interactions. In *Proc. of RSWeb '14*.

[30] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proc. of WWW '01*. 285–295.

[31] Hanna Schäfer and Martijn C Willemsen. 2019. Rasch-based tailored goals for nutrition assistance systems. In *Proc. of IUI '19*. 18–29.

[32] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[33] Christoph Trattner and David Elsweiler. 2017. Food recommender systems: important contributions, challenges and future research directions. *arXiv preprint arXiv:1711.02760* (2017).

[34] Christoph Trattner and David Elsweiler. 2017. Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proc. of WWW '17*. 489–498.

[35] Christoph Trattner, David Elsweiler, and Simon Howard. 2017. estimating the healthiness of internet recipes: a cross-sectional study. *Frontiers in public health* 5 (2017), 16.

[36] Christoph Trattner, Dominik Moesslang, and David Elsweiler. 2018. On the predictability of the popularity of online recipes. *EPJ Data Science* 7, 1 (2018), 20.

[37] Christoph Trattner, Denis Parra, and David Elsweiler. 2017. Monitoring obesity prevalence in the United States through bookmarking activities in online food portals. *PloS one* 12, 6 (2017), e0179144.

[38] Piyaporn Tumnark, Filipe Almeida da Conceição, João Paulo Vilas-Boas, Leandro Oliveira, Paulo Cardoso, Jorge Cabral, and Nonchai Santibutr. 2013. Ontology-based personalized dietary recommendation for weightlifting. In *Proc. of Int. WS on Computer Science in Sports*. 44–49.

[39] Claudia Wagner and Luca Maria Aiello. 2015. Men eat on Mars, Women on Venus?: An Empirical Study of Food-Images.. In *Proc. of WebSci '15*. 63–1.

[40] Claudia Wagner, Philipp Singer, and Markus Strohmaier. 2014. The nature and evolution of online food preferences. *EPJ Data Science* 3, 1 (2014), 1–22.

[41] Joint Who and FAO Expert Consultation. 2003. Diet, nutrition and the prevention of chronic diseases. *World Health Organ Tech Rep Ser* 916, i-viii (2003).

[42] Longqi Yang, Yin Cui, Fan Zhang, John P Pollak, Serge Belongie, and Deborah Estrin. 2015. Plateclick: Bootstrapping food preferences through an adaptive visual interface. In *Proc. of CIKM '15*. 183–192.

[43] Longqi Yang, Cheng-Kang Hsieh, Hongjian Yang, John P Pollak, Nicola Dell, Serge Belongie, Curtis Cole, and Deborah Estrin. 2017. Yum-Me: A Personalized Nutrient-Based Meal Recommender System. *ACM Transactions on Information Systems (TOIS)* 36, 1 (2017), 7.

[44] Yu-Xiao Zhu, Junming Huang, Zi-Ke Zhang, Qian-Ming Zhang, Tao Zhou, and Yong-Yeol Ahn. 2013. Geography and similarity of regional cuisines in China. *PloS one* 8, 11 (2013), e79161.