

The Cholesterol Factor: Balancing Accuracy and Health in Recipe Recommendation Through a Nutrient-Specific Metric*

ALAIN STARKE, Wageningen University & Research, The Netherlands and University of Bergen, Norway
CHRISTOPH TRATTNER, HEDDA BAKKEN, MARTIN JOHANNESSEN, and VEGARD SOLBERG,
University of Bergen, Norway

Whereas many food recommender systems optimize for users' preferences, health is another but often overlooked objective. This paper aims to recommend relevant recipes that avoid nutrients that contribute to high levels of cholesterol, such as saturated fat and sugar. We introduce a novel metric called 'The Cholesterol Factor', based on nutritional guidelines from the Norwegian Directorate of Health, that can balance accuracy and health through linear re-weighting in post-filtering. We tested popular recommender approaches by evaluating a recipe dataset from AllRecipes.com, in which a CF-based SVD method outperformed content-based and hybrid methods. Although we found that increasing the healthiness of a recommended recipe set came at the cost of Precision and Recall metrics, only putting little weight (10-15%) on our Cholesterol Factor can significantly improve the healthiness of a recommendation set with minimal accuracy losses.

Additional Key Words and Phrases: Recipes, Recommender Systems, Health, Offline Evaluation, Nutrients

1 INTRODUCTION

Most food recommender systems to date focus on recommending foods that users like [47]. This includes content-based approaches that are based on historical data, such as by suggesting dairy products to a user if she has previously bought milk. Even though individual ingredients are considered this way [9], the actual nutritional needs of users are often not incorporated [48]. In fact, dietary constraints of users that stem from underlying health conditions, such as hypertension and high levels of cholesterol, have not received much attention to date [37, 47].

Over 90 million adults (20 years or older) in the United States in 2020 have cholesterol levels of 200 mg/dL or higher [51], which is considered unhealthy. Among them, more than 35 million have levels of 240 mg/dL or higher, which puts them at risk for heart disease. Such persons are commonly advised to change their exercise regimen and to attain healthier eating habits. With regard to food recommendation, this would require an approach that incorporates the nutritional content of the recommended internet-sourced recipes [43, 46, 49]. However, an important pitfall is that multiple studies have observed that popular recipes tend to be unhealthy [46], which applies to internet-sourced recipes [29, 35], as well as to popular recipes in other media [30, 39]. This, in turn, can lead to a popularity bias in a recommender system that is at odds with the objective of healthy recipe recommendation (cf. [1]).

This work-in-progress aims to increase the healthiness of recipe recommendations, while maintaining a decent level of accuracy. We focus on mitigating nutrient intake that is associated to high levels of cholesterol, by introducing a 'cholesterol factor', a metric that is based on nutritional guidelines to limit fat, saturated fat, and sugar intake. We apply this in a post-filtering approach to refine an initial set of recipe recommendations, based on competing objectives (cf. [33]), to avoid recipes that contain high levels of nutrients that are associated with high cholesterol. Whereas some multi-objective optimization approaches are tedious to implement and require a lot of computational power [32, 54], we set out a simple post-filtering or post-processing approach that involves linear multiplication (cf. [46]). This is

*Copyright 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). Presented at the MORS workshop held in conjunction with the 15th ACM Conference on Recommender Systems (RecSys), 2021, in Amsterdam, Netherlands.

consistent with previous studies that, among others, seek to mitigate popularity bias or increase recommendation diversity through a post-filtering re-ranking [1–4].

Through offline evaluation, we assess the accuracy of three recipe recommendation approaches: collaborative filtering (CF), content-based (CB), and hybrid. We use our Cholesterol Factor in the most accurate approach (CF; Matrix Factorization SVD) to post-filter the predicted recommendation set on both accuracy and health, aiming to avoid the intake of high levels of fat and sugar. We examine the following research questions:

- **RQ1:** Which recipe recommendation approach has the best performance in terms of different accuracy metrics?
- **RQ2:** To what extent can a nutrient-based post-filtering approach in recipe recommendation balance accuracy and healthiness?

1.1 Contribution

This study presents a nutrient-based post-filtering method to recipe recommendation. Our recommender system is grounded in previous research, testing recommendation methods such as SVD/Matrix factorization that have performed excellently in several independent studies. A food recommender post-filtering approach has also been used by Trattner and Elsweiler [46]. The novelty of our approach is found in our Cholesterol Score, which we have designed based on the guidelines of The Norwegian Directorate of Health, formulated in their ‘Diet Manual’ [15]. Instead of considering the general healthiness of recipes or meals as done in previous studies, such through an aggregate health indicator or calorie counting [12, 43, 46], we assess the presence of multiple nutrients (i.e., fat, saturated fat, and sugar). This way, we specifically target cardiovascular diseases, by re-ranking recipe predictions based on a cholesterol-related metric.

2 RELATED WORK ON FOOD RECOMMENDER SYSTEMS

The earliest meal-planning systems date back to the 1980s [13, 17], which used case-based reasoning. A more contemporary categorization of food recommender systems differentiates between two types of approaches [28]: one that is optimized towards a user’s preferences and one that considers a user’s nutritional needs. Considering how self-actualization and changes in user preferences come about when interacting with recommender systems [22, 24, 38, 41], only presenting healthy recommendations is an ineffective strategy if these do not align with a user’s preferences – unless users are highly-motivated (cf. [27, 38]). A third type of approach, among others suggested by Tran et al. [45], is to balance user preferences and nutritional needs [7]. The recommender system presented in this work falls into the third category and aims to balance between a user’s preferences and nutritional needs.

Type 1 – User Preferences. This type of food recommender system aims to suggest foods that a user is most likely to enjoy; a common approach in this line of research [47]. For example, Freyne and Berkovsky [9] evaluate the performance of different CB, CF, and hybrid approaches, showing that a content-based approach that deconstructs recipe ratings into ingredient ratings performs best. This means that users are inclined to like recipes that contain similar ingredients (e.g., onion) as recipes they liked in the past. Follow-up studies have improved this approach by accounting for negative evaluations [14], using a hybrid approach of Singular Value Decomposition with user and item biases.

Type 2 – Nutritional Needs. A second type of food recommender systems is optimized towards the nutritional needs of the user [38]. Although the relation between unhealthy food intake and adverse health conditions is well-studied (cf. [10]), how recommender systems can help users to make healthier choices has received less attention [45]. An early example is described by Mankoff et al. [26], who generate food recommendations based on an analysis of the users’ food receipts. The system would suggest foods to buy based on the nutrient a user was lacking. In a more goal-oriented

approach, Ueta et al. [50] propose a recommender system that allows the user to disclose specific health problem that she wants to be addressed, for which the system retrieves the nutrient(s) that co-occur more often with that health problem. From there, it would suggest meals that avoid specific nutrients.

Type 3 – Optimizing Between User Preferences and Nutritional Needs. Although healthy food and ‘tasty food’ are not mutually exclusive categories, there is often an optimization trade-off between nutrient intake and user preferences [47]. Whereas some approaches aim to balance these two factors simultaneously when retrieving recipes [12], most approaches rely on either pre-filtering or post-filtering based on one or more health indicators [5].

Pre-filtering approaches typically involve constraint-based recommender systems, although these are relatively rare in the food domain [47]. Yang et al. [52] describe an interface in which a user could disclose dietary constraints (e.g., halal, vegetarian, or vegan) that led to an initial selection of meals, after which user preferences were elicited to re-rank the initial set. In a similar vein, Toledo et al. [44] employ a multi-criteria decision analysis to filter out foods that do not meet a user’s health requirements, before considering the user’s overall preferences.

A more common approach is to apply post-filtering in food recommender systems [47]. Such recommenders retrieve a relevant set of recipes based on user preferences (e.g., through content-based similarity [9]), after which a feature-based or nutrient-based re-ranking or multiplication would be conducted [42, 43, 46]. Elswailer et al. [8] set out such a re-ranking approach, by retrieving all recipes that score above a certain user preference threshold and re-ranking them on one or more health indicators afterwards. Trattner and Elswailer [46] describe a post-filtering approach that re-weights the predicted score of a user X for a recipe Y, based on an aggregate health indicator (i.e., a recipe’s WHO or FSA score; see also [43]). While some approaches have post-filtered on health by considering a meal’s calorie content [12], this study focuses on nutrient intake, because it is a more accurate predictor of health outcomes [6, 31].

3 METHODOLOGY

We assessed to what extent accuracy and health (in terms of cholesterol-related nutrient intake) could be balanced in recipe recommendation through a nutrient-based post-filtering approach. We first describe what recipe dataset and which recommender approaches were used. Subsequently, we explain the rationale of our cholesterol post-filtering metric and how we performed offline evaluation of our results.

3.1 Recipe Dataset

We employed a dataset that comprised 1,031 unique recipes, which were obtained from the website Allrecipes.com, one of the largest recipe websites. Recipes were annotated with nutrient-specific metadata, including the contents in grams (i.e., carbohydrates, (saturated) fat, fiber, protein, sugar), as well as a recipe’s caloric content. Moreover, it specified ingredients, cooking directions, and the average recipe rating given by users on the website.

In total, we had access to 50,681 ratings given to recipes in our dataset. This only included users that had given at least 20 ratings ($M = 63.02$ ratings, $SD = 54.97$). The provided ratings, which were given on a 5-point scale, were relatively high: 55.95% of the given ratings were 5 out of 5 and 30.9% were 4 out of 5 ($Mean = 4.39$, $SD = 0.82$). We therefore expected that classification metrics (i.e., precision, recall) would reach relatively high values.

3.2 Recommendation Approaches

We evaluated our dataset through three recommender approaches. Each approach was founded in previous research conducted in the food recommender domain, comparing approaches from [14], [9], and a hybrid approach.

3.2.1 SVD (Matrix Factorization). We used a Matrix Factorization model to discover latent factors in our recipe dataset (cf. Koren et al. [23] for mathematical details). It involved the SVD algorithm [11], as defined in SciKit Surprise [18]. This approach was analogous to probabilistic matrix factorization (cf. [36]), but also included additional bias parameters for users and items. A related study by Harvey et al. [14] on recipe recommendation showed that a singular value decomposition algorithm, which could be considered a method analogous to Matrix Factorization SVD, outperformed other non-hybrid recommender approaches.

3.2.2 Content-Based. We also employed a content-based algorithm (CB) that exploited the available item descriptions. Content-based approaches were typically used in food recommender systems that optimized for user preferences [28]. For one, Freyne and Berkovsky [9] used an algorithm that deconstructed the recipe ratings into ingredient ratings. For example, if a recipe was given 4 stars out of 5, this rating would be counted for its ingredients (e.g., a rating of 4 for tomato and cucumber). We employed a similar approach by predicting recipe ratings based on the average ratings given by a user u to its j ingredients ing_j . Analogous to [9], we predicted a user’s rating for a recipe r using the average of the ingredient ratings $score(u, ing_j)$, which was computed as follows:

$$pred(u, r) = \frac{\sum_{j \in r} score(u, ing_j)}{j} \quad (1)$$

3.2.3 Hybrid. Our hybrid approach combined our two other algorithms. In line with [9], we used a collaborative approach to overcome sparsity issues of the content-based approach for recipes with few ratings or ingredients. Whereas Freyne and Berkovsky [9] used a Nearest Neighbor approach, we fitted our SVD recommender to a training set to estimate recipe scores for all user-recipe pairs that did not have a true rating. Due to the used training splits, this expanded the training data by a factor 32. Subsequently, we fit the content-based recommender to the expanded training set as described above, which led to a drastically longer computation time than for the other approaches.

3.3 Designing a Post-Filtering Metric: The Cholesterol Factor

We sought to balance user preferences and nutritional needs through a post-filtering approach. On the one hand, if too much weight would be put on nutritional needs, users would be likely to abandon the recommender system due to a mismatch in taste. On the other hand, if only weight is placed on user preferences (as done in some food recommenders [28]), healthiness might be lost due to the popularity of unhealthy internet-sourced recipes [48].

3.3.1 Nutrient Intake Guidelines. To help users to avoid nutrient intake associated with high levels of cholesterol, we designed a metric to assess a recipe’s healthiness. We followed nutritional guidelines from the Norwegian Directorate of Health [15, p.173], an organization tasked with monitoring research in the field of nutrition. Its guidelines were in line with other nutrition authorities in Europe, such as the Dutch Voedingscentrum¹. Whereas in many European countries (e.g., United Kingdom²) guidelines are along the lines of “Saturated fats should be swapped with unsaturated fats”, the Norwegian advice is formulated more specifically in terms of nutrient intake levels as a percentage of calories per day.

Table 1 provides an overview of nutrient-specific guidelines that the Norwegian Directorate for Health has formulated for people with high cholesterol. The guidelines are formalized as percentages of the total calorie content of a meal, either as a recommended interval or as an upper bound only (the amount for fiber was denoted in grams). Three important nutrients (i.e., sugar, fat, saturated fat) did not have an explicit lower bound [15], which made them particularly useful

¹<https://www.voedingscentrum.nl/nl/service/vraag-en-antwoord/aandoeningen/wat-mag-ik-eten-bij-een-te-hoog-cholesterol-.aspx>

²<https://www.gov.uk/government/news/reducing-saturated-fat-lowers-blood-cholesterol-and-risk-of-cvd>

Table 1. Daily nutrient intake guidelines for people with a high level of cholesterol, as formulated by the Norwegian Directorate for health [15, p.173]. Guidelines used for our Cholesterol Factor are denoted in bold.

Macronutrient	Amount	Nutrient-specific Guidelines
Carbohydrates	55-60%	Sugar: <10%
Protein	10-15%	Fiber: 25-35g
Fat	<30%	Monounsaturated Fat: 10-15%; Omega-3: 1%; Polyunsaturated Fat: 4-9%; Saturated Fat: <10%

to include in a continuous health score, for which positive values would indicate one’s nutrient intake to be below the recommended guidelines, while a negative score would exceed those guidelines.

The dataset was made compatible with these guidelines by converting grams to daily calorie allowance percentages. While a gram of fat amounted to 9 kcal, a gram of protein or carbohydrates was equivalent to 4 kcal [25]. By dividing these by the total amount of kcal per recipe, we obtained the kcal percentage of a meal.

3.3.2 Post-filtering Through a Cholesterol Factor. We proposed a Cholesterol Factor to apply post-filtering on a recipe recommendation set. This entailed a re-weighting of all the predicted scores based on a recipe’s nutrient content, in relation to avoiding high levels of cholesterol. Recipes with relatively low levels of fat, saturated fat, and sugar received higher rating predictions, and vice versa.

The Cholesterol Factor is composed of two factors: a ‘Cholesterol Weight’ and a ‘Cholesterol Score’. We used the following formula to post-filter our predicted ratings:

$$\text{Post-Filtered Rating} = \text{Predicted Rating} + (\text{Cholesterol Score} \times \text{Cholesterol Weight}) \quad (2)$$

The Cholesterol Weight is a number that can be chosen depending on how much a system designer wants cater to a user’s health objectives. Higher levels of Cholesterol Weight will lead to higher rating predictions for healthy recipes, presumably at the cost of accuracy. The assigned Cholesterol Weight must be considered relative to the weight attributed to the rating predictions (e.g., a weight of 1 balances health and preference ratings). In contrast, the Cholesterol Score is computed per recipe, based on the nutritional content for fat, saturated fat, and sugar. This is formulated as follows:

$$\text{CholesterolScore} = \text{Fat Points} + \text{Saturated Fat Points} + \text{Sugar Points} \quad (3)$$

Table 2 shows how the points for the three nutrient categories are scored on a scale from -5 to 5. If a recipe scores above 0 in a nutrient category, it indicates that the intake for that nutrient complies with healthy eating guidelines (cf. [15, p.173]). The total Cholesterol Score is the sum score of all three categories and, as such, ranges from -15 to 15. This implies that negative sugar scores could be compensated by positive fat scores. Compared to the commonly used FSA score metric (4-12) for nutrient intake [43, 48], our metric had a larger scale resolution and did not consider salt.

3.4 Evaluation

We performed our recommender evaluation using 5-fold cross validation, using the Surprise Sci-kit [18]. To assess our recommender system approaches (i.e., SVD, content-based, hybrid), we used the three performance metrics in Scikit-learn [34]: Precision (P), Recall (R), and Mean Absolute Error (MAE). K was set at 10, evaluating the top-10 retrieved recipes in terms of their relevance. Recommendations were deemed relevant if their rating was at least 4 out of 5. In addition, our recommender approaches were also compared to a Random Item Ranking baseline.

Table 2. Scoring table per nutrient to compute the Cholesterol Factor of a recipe, which was the sum score. A negative score indicates that a recipe was relatively unhealthy. The amounts denote the percentage of the total caloric content. For example, a recipe with 200kcal and 2g of sugar, contains 8kcal of ‘sugar calories’, which is 4% of the recipe, resulting in 1 sugar point. The scores are based on nutritional guidelines from the Norwegian Directorate of Health [15].

Fat		Saturated Fat		Sugar	
Amount	Points	Amount	Points	Amount	Points
<=1	5	<=1	5	<=1	5
<=5	4	<=3	4	<=3	4
<=10	3	<=5	3	<=5	3
<=20	2	<=7	2	<=7	2
<=30	1	<=10	1	<=10	1
>30	0	>10	0	>10	0
>=40	-1	>=13	-1	>=13	-1
>=50	-2	>=16	-2	>=16	-2
>=60	-3	>=19	-3	>=19	-3
>=70	-4	>=22	-4	>=22	-4
>=80	-5	>=25	-5	>=25	-5

Accuracy of classifications. We used the common classification metrics Precision and Recall to assess how representative the predicted Top-N recommendations were [19, p.180]. Precision indicated the proportion of recommended recipes relevant (*Hits*) for user u in the top-10 recommendation set (*RecSet*), while Recall referred to the proportion of relevant recommended recipes (*Hits*) compared to the total set of relevant recommendations (*RelevantSet*):

$$Precision = \frac{Hits_u}{RecSet_u} \quad Recall = \frac{Hits_u}{RelevantSet_u} \quad (4)$$

Accuracy of predictions: We evaluated the capacity of each algorithm to accurately predict to what extent a user will like a certain recipe. We employed the Mean Square Error (MAE) to assess the accuracy of the rating predictions in our full *TestSet* [19, p.179]. Across all users U and recipes i , the mean deviation between the actual scores given by users and their predicted scores was computed as follows:

$$MAE = \sum_{u \in U} \frac{\sum_{i \in TestSet_u} |rec_i(u, i) - r_i|}{|TestSet_u|} \quad (5)$$

Healthiness of recommendations: Finally, we assessed the healthiness of the predicted recommendation set through the Cholesterol Score, as computed in Table 2. Hence, it could take values between -15 and 15. To contextualize our findings, the average ‘Health Score’ of the dataset was found to be -.53. Please note that because of this scale range, the MAE was less appropriate to evaluate accuracy for post-filtered scores (cf. Table 4).

4 RESULTS

We first evaluated our different recommender approaches in terms of classification and accuracy metrics (RQ1). Subsequently, we investigated how recommender accuracy and health could be balanced using a post-filtering approach through our Cholesterol Factor (RQ2).

4.1 Evaluation of Recommender Approaches in terms of Accuracy (RQ1)

We evaluated our recommender models through different metrics. Table 3 describes the values for recommender classification (Precision@10, Recall@10)³, accuracy (Mean Square Error (MAE)) and the Cholesterol Score. Although all approaches clearly outperformed the baseline in terms of Precision, Recall and MAE, there was not much between them. Due to the high proportion of 5-star ratings in the dataset, we focused on precision and MAE as the key indicators, also because the domain of recipe recommendation is less concerned with false positives [16]. Therefore, we proceeded to our post-filtering evaluation using the SVD algorithm, also because the hybrid approach was computationally more demanding without showing improvements over SVD.

Table 3. Evaluation of three recommender approaches, as well as baseline (random item ranking).

Method	Precision@10	Recall@10	MAE	Cholesterol Score
SVD	.9687	.9789	.5908	-.1218
Hybrid	.9669	.9771	.6162	-.5615
Content-Based	.9676	.9779	.6117	-.2561
Baseline (RIR)	.6978	.3912	1.728	-.3585

4.2 Balancing Accuracy and Health through Post-Filtering (RQ2)

We moved on to examine to what extent a nutrient-based post-filtering approach could balance accuracy and health in recipe recommendation. We performed multiple evaluations of our Matrix Factorization/SVD algorithm by using different values of the Cholesterol Weight in our post-filtering approach.

Table 4 describes Precision, Recall, MAE_f (for post-filtered scores), and recommendation healthiness (i.e., the Cholesterol Score), each for different values of the Cholesterol Weight. It became evident that even for small weight values up to .1, the Cholesterol Score increased noticeably without sacrificing too much accuracy. In particular, Precision and Recall hardly changed, while a sharp increase in the Cholesterol Score (+5 on a 30-point scale) could be observed. Further increasing the Cholesterol Weight in Table 4 did not increase the Cholesterol score significantly, but did lead to smaller values of Precision and Recall. Moreover, MAE_f increased sharply for weight values above .2, as the Cholesterol Score seemed to further affect the rating predictions. Although this showed that an increase in recommendation healthiness did come at the cost of accuracy for high values of the Cholesterol Weight, the health gains are already achieved for small weight values. Moreover, as argued earlier, Precision might be more important in this case than Recall, which deemed the swap of accuracy for health to be a decent tradeoff when a recommender system designer would like to also focus on healthiness, instead of only accuracy.

Table 4 only reports values of the Cholesterol Weight up 2, as this seemed feasible to apply in a recommender context where accuracy and health are balanced. It was not possible to achieve a healthiness score that was significantly higher than 5, which was arguably due to the healthiness of the recipes available in the dataset. As the average healthiness across the entire dataset was -.53, the findings in Table 4 indicated that our Cholesterol Factor not only improved the healthiness of the predicted recommendations compared to the health score of our baseline SVD approach, but also compared to the mean healthiness of the dataset.

³The values for Recall depended on the k -value, in the sense that shorter lists led to lower levels of Recall. These changes, however, were proportional across the different approaches.

Table 4. Evaluation of an SVD algorithm (as defined in Scikit Surprise [18]) in terms of accuracy and healthiness metrics (i.e., Cholesterol Score), for different values of linear, post-filtering weighting, and $k = 5$. Note: MAE_f denotes the rating prediction accuracy after post-filtering, which can take values beyond the original rating scale of 1 to 5, as the Cholesterol Score scale ranged from -15 to +15. Therefore, MAE_f should only be used for comparisons within the table.

Cholesterol Weight	Precision@5	Recall@5	MAE_f	Cholesterol Score
0	.9695	.8989	.5918	-.1602
.01	.9692	.8975	.5921	.7151
.02	.9695	.8991	.5955	1.9675
.03	.9686	.8986	.5985	2.7395
.05	.9691	.8976	.6121	3.8540
.1	.9686	.8979	.6781	4.7999
.12	.9686	.8977	.7144	4.7235
.13	.9690	.8989	.7377	4.6228
.15	.9675	.8975	.7807	4.7858
.2	.9624	.8933	.9017	4.8817
.25	.9498	.8766	1.0360	5.1436
.3	.9259	.8531	1.1783	4.9227
.5	.8214	.7491	1.7878	4.9614
.7	.7317	.6587	2.4284	5.1686
1	.6513	.5821	3.4098	5.0381
1.5	.6225	.5531	5.0640	4.8586
2	.5549	.4878	6.7267	4.9128

Considering the health-accuracy tradeoff, Table 4 suggests that a Cholesterol Weight of .1 is decent estimate to balance user preferences and recommendation healthiness, particularly if nothing would be known about a user’s health preferences. This is the point where the relative increase in health becomes smaller, while the loss of accuracy and the percentage of correct classifications was steady or even increased. This applied to the Allrecipes dataset, which was representative for U.S.-based and Western Europe recipes.

5 DISCUSSION

Food recommender systems face a distinct multi-objective optimization problem [28], which particularly applies to internet-sourced recipes. The challenge at hand is how to optimize between a user’s preferences and the healthiness of food presented. This can be particularly challenging due to the unhealthiness of many popular recipes [48], which also applied to the Allrecipes dataset used in the current study, as well as to users who have liked unhealthy recipes in the past and have changed their preferences [24, 41].

Our main contribution is the development of a metric to balance accuracy and health in a recommender approach; in this study through post-filtering. We have aimed to ‘serve’ recipes to users that are healthier, while maintaining relevance. In doing so, we have taken a nutrient-based approach to optimize the healthiness of recipe recommendations using the Cholesterol Factor, which is based on nutritional guidelines from the Norwegian Directorate for Health. Although such exact guidelines vary between countries, they are rather representative for European countries. Based on our metric, we find many recipes in our dataset that are rather unhealthy, which are best avoided in recommendation sets if users wish to meet nutritional guidelines. This is line with other work that uses an Allrecipes.com dataset [46].

In terms of recommendation approaches, we have followed the work of Freyne and Berkovsky [9]. In doing so, we have observed that an Matrix Factorization SVD algorithm is able to provide decent predictions for our recommendations, compared to hybrid and content-based approaches. This finding is consistent with studies that show the merits of a Singular Value Decomposition approach [14], but different from those that used a content-based approach [9]. Also considering the Cholesterol Score of the retrieved recommendation set prior to nutrient-based filtering (cf. Table 3), we have found MF-SVD to be the best option to examine further in our post-filtering approach.

The good performance of the recommender algorithms in terms of Precision, Recall, and MAE served as a solid foundation for the post-filtering process. We have been able to increase the healthiness of the predicted recommendation set, while maintaining acceptable prediction accuracy. Whereas the healthiness (i.e., the Cholesterol Score) for the top-10 predicted recipes for users started out at -0.38 for the SVD model, we have been able to increase this through post-filtering. Using a weight of 0.12, we have already been able to increase the healthiness score to 4.7. This is still not an extremely healthy score on a scale [-15;+15], but a rather significant increase at a small accuracy cost; and most scores above 0 fall within the recommended guidelines by the Norwegian Directorate of Health. We encourage other researchers to also employ a nutrient-based post-filtering approach to food recommendation, and to expand our work.

The extent to which we can generalize our results is somewhat limited, for we have only used a dataset from a single website. Although the website Allrecipes.com is representative for internet-sourced recipes in the United States and, arguably, some European countries, recent studies suggest that there may be differences in how people perceive and interact with recipes online, depending on their cultural background [21, 53]. Whereas Norwegian nutritional guidelines indicate that our US-based recipe dataset is rather unhealthy on average, US food guidelines tend to be more lenient in terms of some nutritional guidelines [40]. Moreover, our current approach has been specific to the platform and did not consider any user characteristics, which is a problem in more food recommender studies [20]. We are seeking to perform follow-up studies that consider such cultural differences in food, by using different research populations (e.g., in crowdsourcing studies) and datasets from different countries. In a similar vein, we also wish to investigate how our metric performs compared to other summary indicators. For one, the FSA and WHO scores have been used in a similar fashion [46], by estimating the healthiness of recipes on specific nutrient, albeit with a smaller scale size.

Future studies could also consider to further develop the hybrid algorithm used in the current study. The approach in the current study is based on the work of Freyne and Berkovsky [9], which turned out to be computationally demanding. In contrast, we have used a post-filtering approach with linear weights to optimize for health, which is computationally more efficient than incorporating health in the main user model [32]. Although it did not outperform SVD on the metrics in the current study, we aim to further explore its merits in combination with either a pre- or post-filtering approach. In doing so, we aim to investigate whether this leads to better outcomes in terms of accuracy and health, despite its computational demands.

ACKNOWLEDGEMENTS

This work was supported by the Research Council of Norway with funding to MediaFutures: Research Centre for Responsible Media Technology and Innovation, through the Centres for Research-based Innovation scheme, project number 309339.

REFERENCES

- [1] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2019. Managing popularity bias in recommender systems with personalized re-ranking. In *The thirty-second international flairs conference*. 413–418.

- [2] Himan Abdollahpouri, Masoud Mansoury, Robin Burke, and Bamshad Mobasher. 2020. Addressing the Multistakeholder Impact of Popularity Bias in Recommendation Through Calibration. *arXiv preprint arXiv:2007.12230* (2020).
- [3] Gediminas Adomavicius and YoungOk Kwon. 2011. Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering* 24, 5 (2011), 896–911.
- [4] Arda Antikacioglu and R Ravi. 2017. Post processing recommender systems for diversity. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 707–716.
- [5] Devis Bianchini, Valeria De Antonellis, Nicola De Franceschi, and Michele Melchiori. 2017. PREFer: A prescription-based food recommender system. *Computer Standards & Interfaces* 54 (2017), 64–75.
- [6] Ruth E Brown, Karissa L Canning, Michael Fung, Dishay Jiandani, Michael C Riddell, Alison K Macpherson, and Jennifer L Kuk. 2016. Calorie estimation in adults differing in body weight class and weight loss status. *Medicine and science in sports and exercise* 48, 3 (2016), 521.
- [7] Jefferson Caldeira, Ricardo S Oliveira, Leandro Marinho, and Christoph Trattner. 2018. Healthy menus recommendation: optimizing the use of the pantry. In *Proceedings of the 3rd International Workshop on Health Recommender Systems Co-Located with ACM RecSys*. CEUR, Aachen, DE, 6.
- [8] David Elswiler, Morgan Harvey, Bernd Ludwig, and Alan Said. 2015. Bringing the “healthy” into Food Recommenders. In *DMRS*. CEUR, Aachen, DE, 33–36.
- [9] Jill Freyne and Shlomo Berkovsky. 2010. Recommending food: Reasoning on recipes and ingredients. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 381–386.
- [10] G Frost, AA Leeds, CJ Dore, S Madeiros, S Brading, and A Dornhorst. 1999. Glycaemic index as a determinant of serum HDL-cholesterol concentration. *The Lancet* 353, 9158 (1999), 1045–1048.
- [11] Simon Funk. 2006. Netflix update: Try this at home.
- [12] Mouzhi Ge, Francesco Ricci, and David Massimo. 2015. Health-aware food recommender system. In *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, New York, NY, USA, 333–334.
- [13] Kristian J Hammond. 1986. CHEF: A Model of Case-based Planning. In *AAAI*. 267–271.
- [14] Morgan Harvey, Bernd Ludwig, and David Elswiler. 2013. You are what you eat: Learning user tastes for rating prediction. In *International symposium on string processing and information retrieval*. Springer, 153–164.
- [15] Helsedirektoratet. 2016. Kosthåndboken. <https://www.helsedirektoratet.no/veiledere/kosthandboken>
- [16] Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)* 22, 1 (2004), 5–53.
- [17] Thomas R Hinrichs. 1989. Strategies for adaptation and recovery in a design problem solver. In *Proceedings of the Workshop on Case-Based Reasoning*. 343–348.
- [18] Nicolas Hug. 2020. Surprise: A Python library for recommender systems. *Journal of Open Source Software* 5, 52 (2020), 2174. <https://doi.org/10.21105/joss.02174>
- [19] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. 2010. *Recommender systems: an introduction*. Cambridge University Press.
- [20] Mansura A Khan, Barry Smyth, and David Coyle. 2021. Addressing the complexity of personalized, context-aware and health-aware food recommendations: an ensemble topic modelling based approach. *Journal of Intelligent Information Systems* (2021), 1–41.
- [21] Kyung-Joong Kim and Chang-Ho Chung. 2016. Tell me what you eat, and i will tell you where you come from: A data science approach for global recipe data on the web. *IEEE Access* 4 (2016), 8199–8211.
- [22] Bart P Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. 2016. Recommender systems for self-actualization. In *Proceedings of the 10th acm conference on recommender systems*. ACM, New York, NY, USA, 11–14.
- [23] Yehuda Koren. 2009. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 447–456.
- [24] Yu Liang. 2019. Recommender system for developing new preferences and goals. In *Proceedings of the 13th ACM Conference on Recommender Systems*. ACM, New York, NY, USA, 611–615.
- [25] National Agricultural Library. [n.d.]. How many calories are in one gram of fat, carbohydrate, or protein? <https://www.nal.usda.gov/fnic/how-many-calories-are-one-gram-fat-carbohydrate-or-protein>
- [26] Jennifer Mankoff, Gary Hsieh, Ho Chak Hung, Sharon Lee, and Elizabeth Nitao. 2002. Using low-cost sensing to support nutritional awareness. In *International conference on ubiquitous computing*. Springer, 371–378.
- [27] Jutta Mata, Marlene N Silva, Paulo N Vieira, Eliana V Carraça, Ana M Andrade, Sílvia R Coutinho, Luis B Sardinha, and Pedro J Teixeira. 2011. Motivational “spill-over” during weight control: Increased self-determination and exercise intrinsic motivation predict eating self-regulation. 1 (2011), 49–59.
- [28] Stefanie Mika. 2011. Challenges for nutrition recommender systems. In *Proceedings of the 2nd Workshop on Context Aware Intel. Assistance, Berlin, Germany*. Citeseer, 25–33.
- [29] Cataldo Musto, Alain D Starke, Christoph Trattner, Amon Rapp, and Giovanni Semeraro. 2021. Exploring the Effects of Natural Language Justifications in Food Recommender Systems. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, New York, NY, USA, 147–157.

- [30] Yandisa Ngqangashe, Charlotte De Backer, Christophe Matthys, and Nina Hermans. 2018. Investigating the nutrient content of food prepared in popular children’s TV cooking shows. *British Food Journal* 120 (2018), 2102–2115. Issue 9.
- [31] Ruairi O’Driscoll, Jake Turicchi, Kristine Beaulieu, Sarah Scott, Jamie Matu, Kevin Deighton, Graham Finlayson, and James Stubbs. 2020. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. *British Journal of Sports Medicine* 54, 6 (2020), 332–340.
- [32] Umberto Panniello and Michele Gorgoglione. 2011. Context-Aware Recommender Systems: A Comparison Of Three Approaches.. In *DART@ AI* IA. CEUR, Aachen, DE*, 12.
- [33] Umberto Panniello, Alexander Tuzhilin, Michele Gorgoglione, Cosimo Palmisano, and Anto Pedone. 2009. Experimental comparison of pre-vs. post-filtering approaches in context-aware recommender systems. In *Proceedings of the third ACM conference on Recommender systems*. 265–268.
- [34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [35] Markus Rokicki, Eelco Herder, and Christoph Trattner. 2017. How editorial, temporal and social biases affect online food popularity and appreciation. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 11. AAAI, 7. Issue 1.
- [36] Ruslan Salakhutdinov and Andriy Mnih. 2008. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In *Proceedings of the 25th international conference on Machine learning*. 880–887.
- [37] Hanna Schäfer, Santiago Hors-Fraile, Raghav Pavan Karumur, André Calero Valdez, Alan Said, Helma Torkamaan, Tom Ulmer, and Christoph Trattner. 2017. Towards health (aware) recommender systems. In *Proceedings of the 2017 international conference on digital health*. 157–161.
- [38] Hanna Schäfer and Martijn C Willemsen. 2019. Rasch-based tailored goals for nutrition assistance systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 18–29.
- [39] Elizabeth P Schneider, Emily E McGovern, Colleen L Lynch, and Lisa S Brown. 2013. Do food blogs serve as a source of nutritionally balanced recipes? An analysis of 6 popular food blogs. *Journal of nutrition education and behavior* 45, 6 (2013), 696–700.
- [40] Marco Springmann, Luke Spajic, Michael A Clark, Joseph Poore, Anna Herforth, Patrick Webb, Mike Rayner, and Peter Scarborough. 2020. The healthiness and sustainability of national and global food based dietary guidelines: modelling study. *bmj* 370 (2020).
- [41] Alain Starke. 2019. RecSys Challenges in achieving sustainable eating habits.. In *HealthRecSys@ RecSys*. 29–30.
- [42] Alain D Starke, Elias Kløverød Kløverød Brynstad, Sveinung Hauge, and Louise Sandal Løkeland. 2021. Nudging Healthy Choices in Food Search Through List Re-Ranking. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. 293–298.
- [43] Alain D Starke, Martijn C Willemsen, and Christoph Trattner. 2021. Nudging Healthy Choices in Food Search Through Visual Attractiveness. *Frontiers in Artificial Intelligence* 4 (2021), 20.
- [44] Raciél Yera Toledo, Ahmad A Alzahrani, and Luis Martinez. 2019. A food recommender system considering nutritional information and user preferences. *IEEE Access* 7 (2019), 96695–96711.
- [45] Thi Ngoc Trang Tran, Müslüm Atas, Alexander Felfernig, and Martin Stettinger. 2018. An overview of recommender systems in the healthy food domain. *Journal of Intelligent Information Systems* 50, 3 (2018), 501–526.
- [46] Christoph Trattner and David Elsweiler. 2017. Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proceedings of the 26th international conference on world wide web*. 489–498.
- [47] Christoph Trattner and David Elsweiler. 2019. Food Recommendations. In *Collaborative recommendations: Algorithms, practical challenges and applications*. World Scientific, 653–685.
- [48] 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.
- [49] Christoph Trattner, Dominik Moesslang, and David Elsweiler. 2018. On the predictability of the popularity of online recipes. *EPJ Data Science* 7, 1 (2018), 1–39.
- [50] Tsuguya Ueta, Masashi Iwakami, and Takayuki Ito. 2011. A recipe recommendation system based on automatic nutrition information extraction. In *International Conference on Knowledge Science, Engineering and Management*. Springer, 79–90.
- [51] Salim S Virani, Alvaro Alonso, Emelia J Benjamin, Marcio S Bittencourt, Clifton W Callaway, April P Carson, Alanna M Chamberlain, Alexander R Chang, Susan Cheng, Francesca N Delling, et al. 2020. Heart disease and stroke statistics—2020 update: a report from the American Heart Association. *Circulation* 141, 9 (2020), e139–e596.
- [52] 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), 1–31.
- [53] Qing Zhang, Christoph Trattner, Bernd Ludwig, and David Elsweiler. 2019. Understanding Cross-Cultural Visual Food Tastes with Online Recipe Platforms. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 671–674.
- [54] Yong Zheng. 2018. Context-aware mobile recommendation by a novel post-filtering approach. In *The Thirty-First International Flairs Conference*. AAAI, 4.