A Recommender System for Healthy and Personalized Recipe Recommendations

Florian Pecune *
florian.pecune@glasgow.ac.uk
University of Glasgow

Lucile Callebert *
lucile.callebert@glasgow.ac.uk
University of Glasgow

Stacy Marsella
stacy.marsella@glasgow.ac.uk
University of Glasgow

ABSTRACT

Unhealthy eating behavior is a serious public health issue with massive repercussions on an individual's health. One potential solution to this problem is to help people change their eating behavior by developing systems able to recommend healthy recipes that can influence eating behavior. One challenge for such systems is to deliver healthy recommendations that take into account users' needs and preferences, while also informing users about the healthiness of the recommended recipes. In this paper, we investigate whether introducing a healthy bias in a recipe recommendation algorithm, and displaying a healthy tag on recipe cards would have an influence on people's decision making. To that end, we build three different recipes recommender systems: one that recommends recipes matching users' preferences, another one that only recommends healthy recipes, and a third one that recommends recipes that are both healthy and match users' preferences. We evaluate these three systems through a user study in which we asked participants online to select from a list of recipes the ones they like the most.

KEYWORDS

food recommender system; healthcare; collaborative filtering

1 INTRODUCTION

Unhealthy eating is a major public health burden that may be reduced in part by helping people select healthier dietary choices. However, picking appropriate food to eat implies complex decision making processes [4], including being aware of healthy options and choosing among them [17]. With people growing increasingly familiar with interacting with machines in their everyday life, one solution to overcome this issue and help people to make healthier choices is to develop health-aware food recommender systems [21, 22]. One of the most important challenges for such a system is to deliver accurate and personalized recommendations to their users. Although most of the popular recipes found on Internet are unhealthy as defined by the United Kingdom Food Standard Agency (FSA) [23], significant effort has been put recently into optimizing food recommendation algorithms and try to reconcile users' preferences with healthy recipe recommendation [2, 8, 23]. By analyzing people's eating behavior, authors in [9] found that the fat and calorific content of a recipe were the best rating predictors for people interested in eating healthy. However, this information is not always available, and research has shown how hard it is for people to infer the healthiness of a recipe simply from its picture [5], even when the recipe has been categorized as *healthy* [23]. Based on these findings, it becomes important to build systems that not only recommend healthy and personalized recipes, but that also precisely display how healthy these recipes are.

In this paper, we present an experiment in which we investigate whether people would be likely to select recipes that are healthier than the recipes they usually cook. More specifically, our work focuses on investigating how introducing a healthy bias in the recommendation algorithm and the presence of a *healthy tag* would influence users' likelihood to pick recommended recipes. We first describe the different recommendation algorithms we evaluate in our experiment. Then we describe our experimental design and present our results.

2 RELATED WORK

Food recommender systems traditionally rely on two distinct approaches to deliver personalized recipes recommendations. Systems relying on the content-based approach recommend recipes based on their description and users' preferences. In [6] the authors developed a system that infers people's preferred ingredients based on the recipes they like. The system then recommends new recipes containing the previously inferred ingredients. Rather than relying on recipes ingredients, [13] proposed an Ensemble Topic Modeling based approach that relied on features that were previously extracted from a recipe database to deliver recommendations. Their system performed significantly better than a conventional content-based system. Another approach is described in [25], in which the authors implemented a goal-oriented recipe recommender system providing nutrition information. The system first collects the user's goal (e.g. *I want to prevent a cold*) before finding a nutrient that matches that goal. The system then picks the ingredient containing the most of the nutrient previously selected. Finally, the system recommends a recipe containing that specific ingredient. YumMe, the recommender system developed in [27], rely on dietary information to recommend recipes that would match users' needs. The system automatically extracts dietary information from pictures of recipes to form a user profile. The system then relies on this user profile to deliver subsequent recommendations.

Systems relying on the collaborative filtering approach predict recommendation ratings for a user based on ratings from other users. In [7], the authors developed a system that collects users' preferences by asking them to rate and tag the recipes they usually cook at home. The system then relies on users' preferences to rank recipes and deliver recommendations. Authors found that their improved matrix factorization algorithm outperformed the

*The authors contributed equally to this paper.

HealthRecSys'20, September 26, 2020, Online, Worldwide

content-based approach proposed by [6]. The extensive comparison performed in [24] confirms that the collaborative filtering approaches performs better than content-based ones. The study also reveals that the FSA score of the recipe was the most important content feature, highlighting that people are usually consistent in their eating habits. Most of these systems focus on delivering personalized recommendations matching a users profile. They do not intend to recommend recipes that not only match their users preferences but are also healthy.

In [8], the authors try to solve this problem by extending their previous recommendation algorithm [7], introducing a health bias based on the balance between the calories that the user needs and the calories of the recipes. Another system reconciling healthiness and personalization is [3], in which the authors propose a method to recommend healthy recipes based on a subset of ingredients given by a user. The system first selects ingredients that are compatible with the given subset, and associate an optimal quantity for each of these ingredients. The system then generates a pseudo-recipe containing the ingredients with the healthiest nutritional value, before picking the existing recipe best matching the pseudo-recipe. DietOS [1] proposes a solution to manage specific health conditions by recommending ingredients matching its users' health profile. The system also presents the nutritional properties for each ingredient as well as their benefits regarding users health conditions'. In [23] the authors weighted the outcome of their Collaborative Filtering algorithm based on the FSA and WHO scores associated with each recipe. The accuracy of such system was lower compared to the best unfiltered collaborative filtering algorithms, but still better than unfiltered algorithms such as MostPopularItem, UserKNN or ItemKNN. Although these systems present interesting approaches to reconcile health with users' preferences, none of them were evaluated by real users.

A subjective evaluation investigating users preferences towards healthy food is proposed in [5]. The authors first paired specific recipes with their healthier version, i.e. similar recipes with healthier substituted ingredients following the method described in [20]. Then, they showed participants the different pairs and asked the latter to pick the one they preferred, and the one they considered to be the healthiest. Results demonstrated that people were less inclined to pick the healthier recipe of the pair, but also how difficult that was for participants to judge the healthiness of a recipe. However, the recipes pairs were the same for all the participants and were not related to their preferences. Therefore, we focus on the following research question:

RQ1: Are people willing to pick recommendations that are both healthy and match their preferences?

RQ2: Does the presence of a healthy tag on the displayed recipes have an influence on people's decision making?

3 MODEL

To investigate our research questions, our first step was to collect a recipe dataset we could use to build our recommender system. We describe our dataset and the recommender system we built in the following section.

3.1 Recipe dataset

We collected our recipes dataset from allrecipes.com with a web crawler in April 2020, limiting ourselves to recipes that had been reviewed by at least 10 users. We collected a total of 13,515 recipes. We chose allrecipes.com as it is one of most popular recipe websites in terms of traffic, with 25 million unique visitors each month, and it provides nutritional information for most of its recipes.

For each recipe, we collected: its title, image link, list of ingredients and quantities, preparation steps, preparation and cooking times, number of servings, nutritional information, ratings data (number of ratings, ratings min and max, average rating) and list of comments (associated with a unique user name and a rating). To reduce data sparsity issues, we selected a subset of recipes/users so that recipes are rated at least 25 times and users who had rated at least 30 recipes. This results in a dense dataset of 1,169 recipes and 1,339 users for a total of 70,945 ratings.

Similar to [5, 23], we used the standards provided by the Food Standard Agency (FSA, UK) and the green, orange and red traffic-light system to evaluate the healthiness of the recipes. The FSA provides standard ranges for low content (green), medium content (orange) or high content (red) of fat, saturates, sugar and sodium. To calculate a health score, we assign to a recipe, for each of the fat, saturates, sugar and sodium elements, one point if the element's quantity is within the low range, 2 for the medium range and 3 for the high range. The health score therefore ranges from 4 (best) to 12 (worst).

The recipes in our dataset are rather unhealthy: the health score ranges from 6 to 12, and 75.36% of the recipes have a health score of 8, 9 or 10. Only 2.31% of the recipes are healthy (green category). That is consistent with the observations from [5] showing that most popular recipes tend to be unhealthy (high fat content).

3.2 Recommender systems

To answer our research questions, we built a recommender system that takes into account both the users' preferences and the healthiness of the recipes. Users' preferences are learned via collaborative filtering (CF), a popular approach that relies on user ratings and that reports better results compared to content-based approaches [24]. CF methods allow a recommender system to rank recipes according to a score that represents how likely the recipe is to correspond to the user's preferences.

We used implicit feedback transferring all the ratings to positive feedback from users, indicating a preference of the user for the rated recipes compared to the not-rated ones. The user ratings were then turned into confidence levels on how much the user actually liked the rated recipe. This preference-confidence approach has shown to perform well [10].

We used the Implicit python library and tested three popular CF algorithms: Alternating Least Squares (ALS) [19], Bayesian Personalized Ranking (BPR) [16] and Logistic Matrix Factorization (LMF) [12]. We also compared the performances of those algorithms with a simple Most Popular recommender. We split our dataset into a train, cross-validation and test sets to evaluate the performance in terms of AUC [18] of each algorithm. We ran each experiment 100 times and report the average values in Table 1. The best performing algorithm is ALS, with a performance comparable to the

Algorithm	AUC
ALS	0.694
BPR	0.649
Most popular	0.644
LMF	0.617

Table 1: Performance of the CF algorithms

System	w_p	w_h	Mean	Most Common
Pref	1	0	8.750	9
Healthy	0	1	6.000	6
Hybrid-11	1	1	6.275	6
Hybrid-21	2	1	7.175	6
Hybrid-31	3	1	7.662	8

Table 2: Mean and most common value of FSA health scores of recipes recommended to a user by our recommender systems for different values of w_p and w_h.

one reported in [24] on a similar dataset. Our recommender system therefore relies on this algorithm to output, for each recipe and for each user, a preference score $s(r,u)_p \in [0,1]$.

Each recipe is also assigned a health score $s(r)_h \in [0,1]$ that corresponds to the normed FSA health score of the recipe as calculated in section 3.1 and is independent of users' preferences.

The preference and health scores are then combined to calculate a final score $s(r,u) \in [0,1]$ like in equation 1:

$$s(r,u) = (w_p \times s(r,u)_p + w_h \times (1 - s(r)_h))/(w_p + w_h) \qquad (1)$$

where w_p and w_h are weights to assign to the preference and health scores respectively. $s(r,u)$ is then used to rank the recipes and give a recommendation to the user.

We then implemented three different recommender algorithms by adjusting the weights w_p and w_h.

Preference-based recommender. The preference-based recommender system only takes into account the preferences of the user, ignoring the health scores of the recipes; i.e. $w_h = 0$. In our pilot experiments, the average health scores of the recipes recommended to users with this system was 8.750 while most recipes recommended had a health score of 9 (see Table 1).

Healthy recommender. As opposed to the preference-based recommender, the healthy recommender ignores the preferences of the user; i.e. $w_p = 0$. This system therefore always recommend healthy recipes according the FSA standards (i.e. FSA green category with FSA health score of 6 or below). Our dataset contains 27 healthy recipes, all of them associated with a health score of 6. Therefore our healthy recommender system randomly selects five recipes amongst these 27 ones to recommend to the user.

Hybrid recommender. To fulfil our objective of helping people to gradually shift their eating behaviors towards healthier habits, the system should take into account users' preferences but recommends health*ier* recipes compared to the Preference-based recommender. We tested different values for w_p and w_h and ran the system with 10 users for each condition. Table 2 sums up, for each system, the mean and most common health scores of the recipes recommended to the 10 users. Notice that with the Hybrid-11 and Hybrid-21 systems, the most common health score value of recommended recipes is 6, meaning that the systems mostly recommend healthy recipes. Yet, as explained in section 3.1, healthy recipes represent only 2.31% of our database. This strongly limits the possibilities of personalization based on users' preferences and makes those two systems very similar to the healthy recommender. We therefore decided to use the Hybrid-31 recommender system.

4 EXPERIMENT

To answer to our research questions **RQ1** and **RQ2**, we designed an experiment investigating how our system's recommendation algorithm and the presence of a *healthiness tag* in the recipe card influenced users' recipe selection.

4.1 Experimental Design

For the sake of the experiment, we identified two different independent variables. The first one represents our system's recommendation algorithm (**Reco-Algo**) as a between-subject independent variable with three levels: a preference level (*pref-reco*) in which the system delivers recommendations matching users' preferences, a health level (*healthy-reco*) in which the user only gets the healthiest recommendation, and a hybrid level (*hybrid-reco*) in which the system biases the preference-based recommendations towards slightly healthier options. Those three levels correspond to the three systems described in 3.2. The second between-subject variable (**Tag-Mode**) represents whether the recipe card displayed to the user contains a tag representing how healthy the recipe is and has two levels: a healthy-tag level (*healthy-tag*) in which such a healthy tag is present, and a no-tag level (*no-tag*) in which the recipes do not contain any healthiness tags.

Our experiment has a 3x2 design with **Reco-Algo** and **Tag-Mode** as between subject variables. In each of the six conditions, participants followed the same procedure. After agreeing to participate to our study via a consent form, participants were presented with a short description of the task. Each participant was then randomly assigned to a group according to the different independent variables. The task consisted in two steps. In the *preference elicitation* step, participants were asked to select five recipes that they prefer amongst a list of thirty recipes as represented in fig.1. The five selected recipes were sent as the input to our recommender system which delivered five recommendations in return. The later five recipes corresponding to the output of our recommendation system were then presented to the participants during the *recommendation* step along with 25 randomly selected recipes. The position of the recommended recipes on the grid was randomized. As in the *preference elicitation* step, participants were asked to select the five recipes they preferred. Once their choice was made, participants were asked how satisfied they were with their choice and how easy it was to make this choice. The answers for these two questions were 7-point Likert items (anchors: 0 = very dissatisfied/difficult, 6 = very satisfied/easy). We also asked participants what influenced them the most for their choice using an open-ended question. After

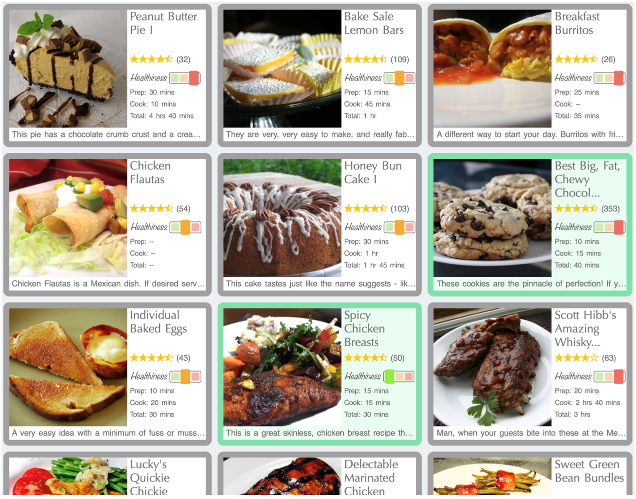

Figure 1: Example of the list of recipes as presented to the participants during the *recommendation step* of the experiment. The cards highlighted in green correspond to the recipes already selected by the participant. In the *preference-elicitation* step, the healthy tag is not displayed.

the end of their task, participants took three surveys: one about what is important to them when looking for a recipe online, another one about their eating habits, and the last one is a demographics questionnaire.

4.2 Measurements

We measured four different constructs in our experiment. (a) We relied on the F1 score to measure the performance of our recommendation algorithm. The F1 score was computed by considering *i)* true positives as the recipes recommended by our system that were selected by the participants, *ii)* false positives as the recipes recommended by our system that were not selected by the participants, *iii)* false negatives as the recipes randomly chosen (i.e. not recommended by our system) that were selected by the participants and *iv)* true negatives as the recipes randomly chosen and that were selected by the participants. (b) To measure the healthiness of the recipes selected by the participants, we calculated the average FSA health score for the five recipes they selected during the recommendation step. Given the nature of our experiment (i.e. selecting items in a list) and based on the results from [26], we also measured (c) whether the participants were satisfied with the five recipes they selected and (d) whether that was easy for them to select the five recipes they liked the most.

5 RESULTS

We recruited 118 participants on Amazon Mechanical Turk. We required that participants had at least a 90% HIT acceptance rate on at least 100 HITs. The recipes presented to the users used imperial measurements, therefore, we restricted our evaluation to participants located in the U.S. Participants spent on average 6 minutes and 56 seconds (std=3 minutes and 47 seconds) on the task and were paid USD1.20. Most participants were aged 29 to 47 years old, with 53% female and 48% male. The majority of participants (82%) was employed full-time.

We conducted four different 3x2 factorial ANOVAs (i.e., analysis of variance) with **Reco-Algo** and **Tag-Mode** as between-subject factors. The dependent measures were the F1 score, selected recipes health score, participant's satisfaction and participant's perceived choice easiness.

5.1 F1 score

The factorial ANOVA revealed a significant main effect of **Reco-Algo** ($F(2, 112) = 8.251$; $p < .001$) on the recommender system's accuracy. There was no main effect of **Tag-Mode** ($F(1, 112) = .945$; $p = .33$) on the recommender system's accuracy and the interaction between the two variables was not significant ($F(2; 112) = 0.358$; $p = .7$). For our follow-up analysis, post hoc comparisons after Bonferroni correction indicated that the mean score for both *pref-reco* (M=.235, std=.156) and *hybrid-reco* (M=.200, std=.164) were significantly better than the *healthy-reco* (M=.100, std=.129). This result shows that people are not likely to select healthy recipes if these recipes do not match with their preferences/habits at all.

To better understand our results, we looked at and compared the recipes recommended by our system and the recipes selected by the users. We observed that the recipes selected by participants were much more diverse than those recommended by our system. For example, our system recommended only chicken-based recipes to a participant who eventually selected two recipes containing meat (chicken and pork), one vegetarian recipe and two desserts. As an objective similarity measure, we calculated for each user the cosine similarity $s_{i,j} \in [0, 1]$ of every pair of recommended (resp. selected) recipe titles R_i and R_j, obtaining a 5x5 similarity matrix with 1s on the diagonal (i.e. when $i = j$). We then averaged the values of the similarity matrix, thus obtaining for each user one similarity score $s \in [0, 1]$ for the recommended (resp. selected) recipes. The average similarity value for all users for the recommended recipes is 0.288 (std=0.093) and the average similarity value for all users for the selected recipes is 0.255 (std=0.053). A Student t-Test revealed that the difference in similarity values of the recommended recipes and the selected recipes is significant ($t(117) = 3.969$, $p < .001$).

The low F1 score obtained by all three recommender systems could therefore be explained by the lack of diversity in the recipes recommended, which is coherent with [26] findings.

5.2 Health Score

There was no main effect of **Reco-Algo** ($F(2, 112) = 1.858$; $p = .16$) or **Tag-Mode** ($F(1, 112) = .060$; $p = .81$) on the selected recipes health score. The interaction between the two variables was not significant ($F(2; 112) = 2.362$; $p = .09$).

Although none of these results were significant, the interaction graph in Fig.2 depicts how the presence of a healthy tag on the recipe card had different effects on the health score of the selected recipes depending on the conditions. In the *pref-reco* condition, people selected healthier recipes when the healthy-tag was displayed, unlike in the *healthy-reco* condition, in which people selected recipes that were less healthy when the healthy-tag was displayed. There was almost no impact of the healthy-tag in the *hybrid-reco* condition.

	Reco-Algo			Tag-Mode	
Variable	**pref-reco**	**hybrid-reco**	**healthy-reco**	**healthy-tag**	**no-tag**
F1 score	$0.235(\pm0.156)^{***}$	$0.200(\pm0.164)^{*}$	$0.100(\pm0.129)^{***/*}$	$0.166(\pm0.172)$	$0.193(\pm0.147)$
Health score	$9.185(\pm0.631)$	$8.965(\pm0.653)$	$8.889(\pm0.8427)$	$9.000(\pm0.735)$	$9.030(\pm0.706)$
Satisfaction	$5.000(\pm0.961)$	$4.950(\pm0.782)$	$4.868(\pm1.234)$	$5.069(\pm0.814)$	$4.817(\pm1.142)$
Choice easiness	$4.175(\pm1.412)$	$4.375(\pm1.314)$	$4.421(\pm1.222)$	$4.362(\pm1.347)$	$4.283(\pm1.290)$

Table 3: Summary of all means and standard errors (in parentheses) for the four dependent variables across conditions. The differences between the means are marked according to their level of significance (* for p < .05 and * for p < .001)**

In our post-study questionnaire, nine participants of the *healthy-tag* condition mentioned that they were mostly influenced by the healthy tag while choosing a recipe. This correlates with both *i)* a significantly higher F1 score (t(12.3)=−2.7, p<0.05) for participants who mentioned they were influenced by the healthy tag (M=.29, std=.15) compared to the other participants (M=.14, std=.17) and *ii)* a significantly lower health score of the selected recipes (t(11.4)=2.8, p<0.05) for participants who mentioned they were influenced by the healthy tag (M=8.42, std=.67) compared to the other participants (M=9.11, std=.70). Both the F1 score and the health score were significantly different between the two Tag-Mode groups in the recommendation phase but not in the preference elicitation phase, confirming that people are poor judges of the healthiness of a recipe [5].

5.3 Perceived satisfaction and choice easiness

There was no main effect of **Reco-Algo** (F(2, 112) = 0.171; p = .84) or **Tag-Mode** (F(1, 112) = 1.850; p = .18) on participants' satisfaction. The interaction between the two variables was not significant (F(2; 112) = 0.186; p = .83). Overall, participants were more satisfied with their choices when the recommended recipes matched their preferences. The presence of a healthy tag on the recipe card increased satisfaction regardless of the recommendation algorithm.

Regarding the perceived ease of use, there was no main effect of **Reco-Algo** (F(2, 112) = 0.391; p = .68) or **Tag-Mode** (F(1, 112) = .110; p = .74) on participants' perceived choice easiness. The interaction between the two variables was not significant (F(2; 112) = 1.697; p = .19). Although the presence of a healthy tag lowered the perceived difficulty in both the *healthy-reco* and the *pref-reco* conditions, such tag made the selection more difficult for people who were recommended recipes in the *hybrid-reco* conditions.

5.4 Discussion

To answer to our research question **RQ1**, our results show that people are slightly less inclined to select recommendations coming from our hybrid recommender compared to the preference-based one. However, although the difference is minimal when no tags are displayed on the recipes, the presence of healthy tags accentuates the difference. The negative impact of the healthy tags on the F1 score of our hybrid algorithm can be linked to the choice easiness. Unlike the healthy and preference-based conditions, the healthy tags made it more difficult for participants to select five recipes in the hybrid condition. One potential explanation is that all recipes in the hybrid condition had very similar health scores (in orange), whereas the two other conditions introduced recipes tagged in green (for the healthy-algo) or in red (for the pref-algo). People who explicitly cared about the health tag were more likely to choose recipes recommended by our system in the *hybrid-reco* and *health-reco* conditions. That confirms the results found in [9] and highlights that need to accurately infer people's eating goals to adapt the recommendation algorithm accordingly. Both hybrid and preference-based recommender systems had a significantly better F1 scores than our healthy recommender system, which shows that people are not likely to select healthy recipes if these recipes do not match with their preferences/habits at all.

There is no significant evidence related to the impact of healthy tags on participants' decision making that would help us answer **RQ2**. Indeed, only nine people out of 60 explicitly stated they were influenced by the healthy tag in our post-evaluation questionnaire. However, the results described in section 5.2 suggest that although people are more likely to pick healthier recipes compared to what they would usually pick when informed about recipes healthiness, they are less likely to pick recipes tagged as very healthy. In other words, people will avoid recipes tagged as unhealthy (in red) as well as recipes tagged as healthy (in green). The first part can partially be explained by the fact that people usually associate a feeling of guilt with unhealthy food consumption [11]. Thus, people are less inclined to pick unhealthy recipes if they are explicitly informed about their unhealthiness. The second part can be explained by the "healthy = less tasty" effect which describe how people tend to associate healthy food with low tastiness [15]. Hence, we assume that participants in our experiment were less inclined to pick recipes explicitly tagged as healthy because they thought such recipes would be tasteless. Overall, people were more satisfied with their choices when informed about the recipes' healthiness.

6 CONCLUSION

In this paper, we investigated whether introducing a healthy bias in a recipe recommendation algorithm, and displaying a healthy tag on recipe cards would have an influence on people's decision making. Our results show that a the performance of a recommender system able to combine healthiness with personalization depends on its users eating goals. People already interested in eating healthy are more likely to select recipes coming from such a recommendation system. For the others, our results also suggest that adding a simple yet accurate tag depicting how healthy recipes are might help them to select healthier recipes compared to what they would usually select. Explicitly informing people how unhealthy some recipes

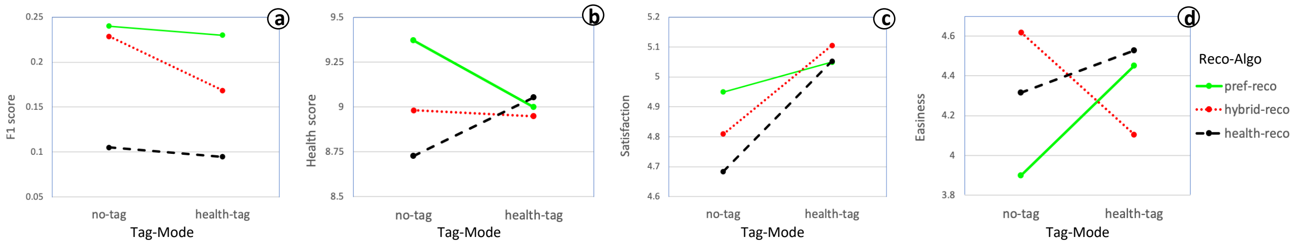

Figure 2: Interaction graphs between Reco-Algo and Tag-Mode regarding the average ⓐ F1 score, ⓑ health score ⓒ satisfaction and ⓓ choice easiness.

are might help them to consciously change their eating habits and prevent them to pick unhealthy recipes.

One potential extension of this work would be to combine our CF approach with a knowledge-based approach to have more control over the diversity of the recommended recipes. As explained in section 5.1, a diverse set of recommendations can positively impact users' experience [26]. The results of a CF-based algorithm could for instance be post-filtered to force the presence of different categories of recipes (e.g. main, vegetarian, dessert) and/or different ingredients (e.g. chicken, pork) in the list of recommended recipes.

The integration of a knowledge-based approach could be done by building a conversational recommender system asking specific questions about users requirements. Appropriate conversational skills can also improve users' experience as well as people's perception of recommended items [14]. Furthermore, such a conversational approach could also help us to know whether users are initially interested in eating healthy so that the system could adapt its recommendations consequently.

REFERENCES

[1] Giuseppe Agapito, Mariadelina Simeoni, Barbara Calabrese, Ilaria Caré, Theodora Lamprinoudi, Pietro H Guzzi, Arturo Pujia, Giorgio Fuiano, and Mario Cannataro. 2018. DIETOS: A dietary recommender system for chronic diseases monitoring and management. *Computer methods and programs in biomedicine* 153 (2018), 93–104.

[2] 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.

[3] Meng Chen, Xiaoyi Jia, Elizabeth Gorbonos, Chnh T Hong, Xiaohui Yu, and Yang Liu. 2019. Eating healthier: Exploring nutrition information for healthier recipe recommendation. *Information Processing & Management* (2019), 102051.

[4] Sally Jo Cunningham and David Bainbridge. 2013. An analysis of cooking queries: Implications for supporting leisure cooking. (2013).

[5] David Elsweiler, Christoph Trattner, and Morgan Harvey. 2017. Exploiting food choice biases for healthier recipe recommendation. In *Proceedings of the 40th international acm sigir conference on research and development in information retrieval*. 575–584.

[6] Jill Freyne and Shlomo Berkovsky. 2010. Intelligent food planning: personalized recipe recommendation. In *Proceedings of the 15th international conference on Intelligent user interfaces*. ACM, 321–324.

[7] Mouzhi Ge, Mehdi Elahi, Ignacio Fernaández-Tobías, Francesco Ricci, and David Massimo. 2015. Using tags and latent factors in a food recommender system. In *Proceedings of the 5th International Conference on Digital Health 2015*. ACM, 105–112.

[8] Mouzhi Ge, Francesco Ricci, and David Massimo. 2015. Health-aware food recommender system. In *Proceedings of the 9th ACM Conference on Recommender Systems*. 333–334.

[9] 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.

[10] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE International Conference on Data Mining*. Ieee, 263–272.

[11] JungYun Hur and SooCheong Shawn Jang. 2015. Anticipated guilt and pleasure in a healthy food consumption context. *International Journal of Hospitality Management* 48 (2015), 113–123.

[12] Christopher C Johnson. 2014. Logistic matrix factorization for implicit feedback data. (2014), 78 pages.

[13] Mansura A Khan, Ellen Rushe, Barry Smyth, and David Coyle. 2019. Personalized, Health-Aware Recipe Recommendation: An Ensemble Topic Modeling Based Approach. *arXiv preprint arXiv:1908.00148* (2019).

[14] Florian Pecune, Shruti Murali, Vivian Tsai, Yoichi Matsuyama, and Justine Cassell. 2019. A model of social explanations for a conversational movie recommendation system. In *Proceedings of the 7th International Conference on Human-Agent Interaction*. 135–143.

[15] Rajagopal Raghunathan, Rebecca Walker Naylor, and Wayne D Hoyer. 2006. The unhealthy= tasty intuition and its effects on taste inferences, enjoyment, and choice of food products. *Journal of Marketing* 70, 4 (2006), 170–184.

[16] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2012. BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618* (2012).

[17] Benjamin Scheibehenne, Rainer Greifeneder, and Peter M Todd. 2010. Can there ever be too many options? A meta-analytic review of choice overload. *Journal of consumer research* 37, 3 (2010), 409–425.

[18] Gunnar Schröder, Maik Thiele, and Wolfgang Lehner. 2011. Setting goals and choosing metrics for recommender system evaluations. In *UCERSTI2 workshop at the 5th ACM conference on recommender systems, Chicago, USA*, Vol. 23. 53.

[19] Gábor Takács and Domonkos Tikk. 2012. Alternating least squares for personalized ranking. In *Proceedings of the sixth ACM conference on Recommender systems*. 83–90.

[20] Chun-Yuen Teng, Yu-Ru Lin, and Lada A Adamic. 2012. Recipe recommendation using ingredient networks. In *Proceedings of the 4th Annual ACM Web Science Conference*. 298–307.

[21] 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.

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

[23] 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.

[24] Christoph Trattner and David Elsweiler. 2019. An Evaluation of Recommendation Algorithms for Online Recipe Portals.. In *HealthRecSys@ RecSys*. 24–28.

[25] Tsuguya Ueta, Masashi Iwakami, and Takayuki Ito. 2011. Implementation of a goal-oriented recipe recommendation system providing nutrition information. In *2011 International Conference on Technologies and Applications of Artificial Intelligence*. IEEE, 183–188.

[26] Martijn C Willemsen, Mark P Graus, and Bart P Knijnenburg. 2016. Understanding the role of latent feature diversification on choice difficulty and satisfaction. *User Modeling and User-Adapted Interaction* 26, 4 (2016), 347–389.

[27] 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.