Nutrition Guided Recipe Search via Pre-trained Recipe Embeddings

Diya Li
Rensselaer Polytechnic Institute
lid18@rpi.edu

Mohammed J. Zaki
Rensselaer Polytechnic Institute
zaki@cs.rpi.edu

Ching-Hua Chen
IBM Research
chinghua@us.ibm.com

***Abstract*—The use of machine learning to recommend foods that are both healthy and tasty is an open problem. Fundamentally, it is challenging to balance health goals with preferences in taste, while offering users a large diversity of options. Representing recipes via embedding vectors trained on large-scale food datasets can capture the implicit semantics of a recipe. We utilize pre-trained embeddings to perform recipe search and compare our search results with a keyword based search. We compare the health score, nutritional content and recipe titles returned using both search approaches. Our exploratory experiments show that recipe search via embeddings can return more diverse recipe titles in contrast to keyword based search.**

***Index Terms*—food computing, recipe representation, recipe search, food recommendation**

I. Introduction

Eating well is fundamental to good health and well-being. Food computing mainly utilizes methods from computer science to address food-related issues in medicine, biology, and gastronomy [1]. The fast development of online networks (e.g., social networks and recipe sharing websites) has led to large-scale food datasets with rich knowledge like recipes, food images and food logs [2]. Among existing large-scale recipe datasets [3]–[7], the Recipe1M dataset [7] is the largest one which contains over 1 million recipes consisting of ingredients, cooking instructions, and food images. Recipe1M also provides nutritional information for its recipes by quantifying the ingredients through their measurable units and numerical quantities, and then mapping their nutritional content to the USDA nutrient database (www.ars.usda.gov/ba/bhnrc/ndl) in terms of sugar, salt, saturates, and fat. Next, using this mapped content, a nutritional quality score established by the British Food Standards Agency (FSA; food.gov.uk) is computed. However, we observed that there is very scarce nutritional information in Recipe1M, spanning only about 5% of the recipes. The Recipe1M dataset is collected from over two dozen popular cooking websites and over half of its recipes are scraped from the *food.com* website. To mitigate the sparsity and also inaccuracy of nutrient information in Recipe1M, which contains derived nutrition values, we directly extract the nutrition content from the *food.com* website since it contains high quality nutrition information, thereby generating a large dataset of around half a million recipes with adequate nutritional content for food recommendation.

Popular food recommender systems are typically based on collaborative filtering techniques that utilize past behaviors of 'similar' users [8]. One limitation of this approach is that it reinforces past behaviors, which may be unhealthy. For a user addicted to calorie-rich foods, the recommender system will return suggestions of high-calorie foods. On the other hand, content-based recommender systems that use nutritional content of foods to suggest healthier options may err on the side of being unattractive to users' palates. These systems also have the drawback of requiring explicit representation of knowledge related to health and nutrition. Representing recipes via embedding vectors can be seen as a compromise, because these embeddings can be trained on large datasets to capture the implicit semantics of a recipe [7], [9]. There are a lot of studies to learn cross-modal embeddings for cooking recipes and food images [10]–[13]. Although there has been recent interest in creating embeddings of recipes, meaningful applications remain elusive, such as guiding user behavior, and improving health and understanding culinary culture [1]–[3]. Recently, [14] proposed a joint approach for learning recipe embeddings that uses textual information of both the ingredients and cooking directions from Recipe1M [7].

In this work, we explore the potential uses and limitations of pre-trained recipe embeddings for retrieving a set of recipes that could offer users more meaningful choices with respect to both 'healthiness' and 'taste'. Related work on recipe search mainly focuses on semantic-based features and behavior-oriented requirements [17], [18]. These works have a different task setting from ours, and they do not explore the use of embeddings. In contrast, to better understand the nature of recipe search, we compare the properties of recipes retrieved using the pre-trained embeddings, with results retrieved from contextualized sentence embeddings, and those obtained through keyword-based online search. Our quantitative evaluation, along with human evaluation, reveals that recipe search via embeddings can return more diverse results in contrast to keyword based search.

Our contributions are twofold. First, we examine the properties of recipe search via learned embeddings and show that it can return more diverse results comparing to key-word based search. Second, we expand the largest recipe dataset with high quality nutrition information, which we hope will be a good resource for future nutrition-related studies.

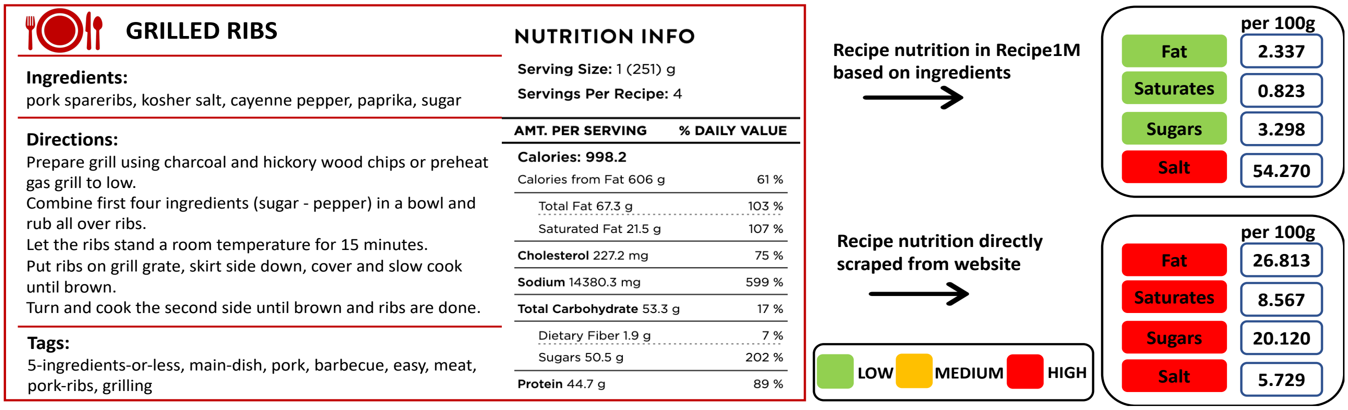

Fig. 1. Example of a recipe comprising ingredients and directions, tags and nutrition information scraped from food.com with FSA score presented in Recipe1M (upper right corner) and new FSA scores directly calculated using recipe nutrition (lower right corner) from food.com. Note that the recipe lists nutrients per 251g (serving size), which we convert to per 100g.

TABLE I
EXAMPLES OF HEALTHY FOOD RECOMMENDATION WITH COLOR-CODED FSA HEALTH QUALITY RATINGS.

Query Recipe	k-NN Results (k=3)	Nutrition Information	Healthy Choice
Grandma's Chocolate Mint Cookies fat saturates sugars salt	Cabernet Sauvignon Chocolate Chip Cookies	fat saturates sugars salt	
	Chocolate Mexican Wedding Cookies	fat saturates sugars salt	
	Spicy Chocolate Cookies	fat saturates sugars salt	✓
Easy Roast Beef fat saturates sugars salt	Bootlegger's Beef	fat saturates sugars salt	✓
	Holiday-Spiced Roast Beef	fat saturates sugars salt	
	Home Made Beef Hash	fat saturates sugars salt	

II. DATASET

Whereas Recipe1M is the largest recipe dataset, and thus ideally suited for learning recipe embeddings from the ingredients and preparation directions contained in the data, only 5% of the recipes in the Recipe1M dataset include nutritional information. For better recommendation based on nutrients, we therefore extract nutrition information directly from the *food.com* website. Fortunately, there is a large overlap between Recipe1M and food.com, namely there are 507,834 recipes that are common to both. The advantage of the recipe information on *food.com* is that it includes nutritional information for each recipe. Using the nutritional content available on *food.com*, we apply the nutritional quality rating established by FSA to each recipe. For each recipe, an FSA color-coded rating (red–bad, amber–caution or green–good) is computed independently for four macronutrients: fat, saturated fat (or saturates), sugar and salt. We also obtain pre-trained embedding vectors for each of the recipes using the recipe representation learning approach [14] as it integrates multiple contents of recipes. Every recipe is represented as a 600-dimensional vector, but the embeddings do not take into account the nutritional content of the recipes, rather only the ingredients (e.g., butter, chicken, onion, etc) and the cooking instructions. Thus, it's fair to study the nutrition-guided recipe search problem as the nutritional content is initially excluded from recipe embeddings. Figure 1 illustrates an example of such content. In the example we can clearly see that the fat, saturates and sugars entries in Recipe1M nutrition panel appear as green (or healthy/good), whereas in fact the dish is unhealthy (red) for all categories. This example shows that the derived nutritional data in Recipe1M is not always reliable, whereas the information directly scraped from food.com has higher quality.

III. EMBEDDINGS-BASED RETRIEVAL

With the additional recipe nutrition information, we obtain the healthiness distribution within the pre-trained recipe embeddings and are able to do food recommendations regrading food diversity and healthiness. For a given recipe query, we retrieve the top-k nearest neighbors based on a distance measure between the query recipe and the other 507,833 recipes in our dataset. The cosine similarity between two recipe embeddings is used to measure the distance. Since the search space is large, to speed up the retrieval process, we alternatively search for the top-k approximate nearest neighbors (ANN), using Faiss [15].

Table I shows two examples of the search results; the corresponding FSA health quality ratings for each of the macronutrients are also displayed. In the first example, for a person who loves *Grandma's Chocolate Mint Cookies*, the system presents several alternatives that exhibit variation in FSA health ratings, and includes recipe titles that appear to be semantically similar, without being potentially redundant. Although our immediate interest is in understanding the nature of the retrieved results, a recommender system might consider recommending *Spicy Chocolate Cookies* as, to some degree, it is 'similar' to the user's favorite dessert and healthier according to fat, saturates, sugar, and salt nutrients. This example shows the effectiveness of our food recommendation approach based on food similarity and the corresponding nutrition information.

TABLE II
EXAMPLE OF RECIPE SEARCH COMPARISON.

Query Recipe	k-NN Results From Recipe Embeddings (k=5)	Nutrition Information
Roasted Brussels Sprouts with Pine Nuts	Roasted Brussels Sprouts with Pear and Pistachio	fat saturates sugars salt
	Roasted Green Beans with Lemon Pine Nuts Parmigiano	fat saturates sugars salt
	Herbed Green Bean Casserole	fat saturates sugars salt
	Roasted Brussels Sprouts with Kielbasa	fat saturates sugars salt
	Make the Bo-Beau Brussels Sprouts Home Edition!	fat saturates sugars salt
	Retrieved Results From food.com	**Nutrition Information**
	Roasted Brussels Sprouts!	fat saturates sugars salt
	Roasted Brussels Sprouts	fat saturates sugars salt
	Roasted Brussels Sprouts and Red Onions	fat saturates sugars salt
	Roasted Brussels Sprouts with Browned Garlic	fat saturates sugars salt
	Roasted Brussels Sprouts and Garlic	fat saturates sugars salt

TABLE III
RESULTS OF PROPERTY ANALYSIS. THE HIGHER EVALUATION SCORES ARE PRESENTED IN BOLD.
QUANT. INDICATES THE QUANTITATIVE MEASUREMENTS.

Type	Healthiness		Nutri Diversity		Title Diversity	
	quant.	human	quant.	human	quant.	human
Key-word Based Search						
main	2.46	2.83	0.33	2.43	12.8	3.04
veg.	2.55	3.11	0.32	2.00	13.0	2.86
soup	2.62	2.71	0.34	2.45	10.8	2.07
dessert	1.94	0.34	0.66	3.87	8.9	1.78
bev.	2.50	1.93	0.46	2.36	8.5	1.83
total	2.41	2.19	**0.42**	**2.62**	10.8	2.32
Pre-trained Recipe Embedding						
main	2.37	2.42	0.34	2.53	13.3	3.23
veg.	2.61	3.12	0.27	2.10	13.3	2.87
soup	2.66	2.83	0.25	1.89	12.0	2.67
dessert	1.96	0.53	0.66	3.42	12.2	2.56
bev.	2.53	2.17	0.39	1.79	11.1	2.13
total	**2.42**	2.22	0.38	2.35	**12.38**	**2.70**
Sentence Embedding						
main	2.44	2.74	0.36	2.43	12.9	3.00
veg.	2.47	3.09	0.32	2.05	13.3	2.84
soup	2.59	2.71	0.27	2.17	10.1	2.50
dessert	1.93	0.47	0.61	3.19	10.2	2.00
bev.	2.60	2.21	0.51	2.40	8.7	1.94
total	2.41	**2.24**	0.41	2.45	11.04	2.46

IV. PROPERTY ANALYSIS OF RECIPE SEARCH

Our main concern is regarding the nutrition and diversity of recipe search, since the system should offer users more choices in regard to 'healthiness' and 'taste'. In this section, we aim to compare the properties of recipes retrieved from different measures. Besides using the learned recipe embeddings [14] which integrate the ingredient and cooking instructions, we also create sentence embeddings directly from recipe titles. We observe that the recipe titles, which are typically short phrases, can be represented as sentence embeddings from pre-training deep language models. Since the popular pre-trained language model BERT [16] has shown competitive power for language understanding, we adopt BERT to get the contextualized embeddings of words in recipe titles and then obtain the sentence embeddings by average pooling. Finally, we compare the search results retrieved from pre-trained embeddings, sentence embeddings, and those obtained directly through the keyword-based search engine on the *food.com* website.

The properties that we compare them on include **healthiness**, **nutritional diversity**, and **title diversity**. We randomly select 50 recipes from our dataset, which are evenly chosen from 5 different recipe categories (i.e., main-dish, vegetable, soup, dessert, and beverage). Each of the sampled recipes serves as a query recipe, and we retrieve the top-k (i.e., k=5) results based on the pre-trained recipe embeddings from [14], sentence embeddings from BERT, as well as the first k results returned by food.com's keyword-based search engine.

In order to generate meaningful results using food.com's keyword-based search engine, we use appropriate subsets of the words in the recipe title to generate results instead of entering the complete recipe name since the latter produces only one exact match or redundant recipes sharing the same name. For instance, for the query recipe *Roasted Brussel Sprouts with Pine Nuts*, we search[1] for 'roasted Brussels sprouts'. Table II shows that for the query *Roasted Brussels Sprouts with Pine Nuts*, results based on our recipe embeddings are more diverse both in title words and nutritional content.

For the comparison across all 50 queries, we use the average FSA ratings to measure the healthiness of the retrieved recipes. FSA color rating is quantified to numeric values (red-1, amber-2, green-3). We use the standard deviation of FSA ratings to measure the nutritional diversity of the retrieved recipes and use the count of unique words per title in the search results as a measure of title diversity. Stop words and punctuation are removed from title for unique word counting. The overall results are shown in Table III. In addition to the quantitative measurements, we also conducted a human evaluation. We recruited 5 users to evaluate the 50 search results with regard to healthiness, nutritional diversity, and title diversity. For the user profiles, the 5 users had completed a bachelor's degree or higher. Two of them have bachelor's degrees in computer science and the other three got master's degrees in finance, architecture, and mathematics respectively. The FSA ratings along with nutrition information are both provided for evaluators. Each property is given five grades: excellent (4), good (3), fair (2), poor (1), bad (0). The averaged human evaluation scores are also reported in Table III.

[1] https://www.food.com/search/Roasted+Brussel+Sprouts

As demonstrated in Table III, the overall healthiness of search results from food.com's keyword search is 2.41, compared to 2.42 from the pre-trained and 2.41 from the sentence embedding-based retrieval. In view of nutrition information, results from embeddings are healthier, though the difference is minor. Human evaluation shows the same tendency (2.19 vs. 2.22 vs. 2.24),with results not that significantly different. The dessert category gets the lowest healthiness score in all the three measures. It makes sense as most desserts containing excessive sugar and fat are unhealthy according to the FSA scores. In healthiness evaluation, search results retrieved from the pre-trained embeddings get the highest score in quantitative measure, while the human evaluation is in favor of the performance of sentence embeddings. One potential reason might be that the FSA ratings are also provided for users, the colored ratings have more impact on the judgement of users than the numerical nutrition information.

The nutritional diversity of search results from food.com's keyword search is 0.42, larger than the diversity score from the embedding-based retrievals which are 0.38 and 0.41. Human evaluation demonstrates the same performance (2.62 vs. 2.35 vs. 2.45). Intuitively, for a recipe search, the higher healthiness the result has, the lower nutritional diversity it shows.

For title diversity, the unique word count of the keyword search on food.com is 10.8, compared to 12.38 from the embedding-based retrievals, and 11.04 from the sentence embeddings. This indicates that food recommendation based on the pre-trained recipe embeddings is more diverse than food retrieved from sentence embeddings and food.com, and the comparison is also consistent with human evaluation. This is due to the fact that the recipe embeddings were trained on recipe ingredients and cooking directions in Recipe1M and yield recommendations with similar content, whereas the recommendations from sentence embeddings tend to return titles semantically related and food.com only provide recipes containing similar keywords in the title. We also observe that the main-dish category achieves the highest title diversity scores with regards to all the three measures in both quantitative and human evaluations. It might be because recipes categorized to main-dish are more diverse in nature and comprise a large portion ($> 60\%$) of the Recipe1M dataset.

Overall, the quantitative evaluation is mostly consistent with human evaluation except for the judgement on healthiness between the pre-trained and sentence embeddings. An interesting finding is that over two thirds of the queries which have diverse search results also have varied nutrition which is consistent with human intuition. Regarding food recommendation, a user can either choose similar recipes based on the results from food.com or more diverse choices from two types of embeddings. The limitation of food.com recommendations is that it cannot always return multiple search results with a given query while recipe embeddings can always return top k similar recipes. For the embedding-based search, we adopt the ANN search to facilitate the search speed with the sacrifice of accuracy.

V. Conclusion And Future Work

In this paper, we studied the problem of nutrition guided recipe search through pre-trained recipe embeddings. We expand the largest recipe dataset, Recipe1M, with better nutritional information and perform food recommendation for users via pertained embeddings and nutrition facts. We compare the characteristics of recipes retrieved using recipe embeddings to those returned by a standard keyword search. In the examination of search results properties, we observe that both approaches are on par with respect to the healthiness and nutritional diversity of the results. On the other hand, the embedding-based results yield notably more diverse recipe titles. This combination of findings suggests that embedding-based retrieval may offer more meaningful meal alternatives without necessarily affecting meal nutrition. Additionally, since embedding-based retrieval still provides sufficient nutritional diversity, we see promise in using this approach to support efficient nutrition guided recipe retrieval.

In future work, we plan to examine the performance of other types of recipe embeddings (i.e., knowledge-graph based embeddings and cross-modal embeddings) on recipe search and thereby to see how "differently structured KGs/modalities" might lead to embeddings that are more useful for different types of searches. In this paper, searching first then choosing with additional nutrition information is a naive way to do nutrition guided recipe search; more skilful approaches are worth exploring in the future.

References

[1] W. Min, S. Jiang, L. Liu, Yong Rui, R. Jain. "A survey on food computing." ACM Computing Surveys (CSUR) 52, no. 5 (2019): 1-36.

[2] O. Mouritsen, R. Edwards-Stuart, Y.-Y. Ahn, S. Ahnert. "Data-driven methods for the study of food perception, preparation, consumption, and culture." Frontiers in ICT 4 (2017): 15.

[3] S. Sajadmanesh et al. "Kissing cuisines: Exploring worldwide culinary habits on the web." WWW Conf. 2017.

[4] M. Rokicki, C. Trattner, E. Herder. "The Impact of Recipe Features, Social Cues and Demographics on Estimating the Healthiness of Online Recipes." In ICWSM, pp. 310-319. 2018.

[5] S. Yagcioglu, A. Erdem, E. Erdem, N. Ikizler-Cinbis. "RecipeQA: A challenge dataset for multimodal comprehension of cooking recipes." In EMNLP, 2018.

[6] D. Batra et al. "RecipeDB: a resource for exploring recipes." Database (Oxford). Nov 25, baaa077, 2020.

[7] J. Marin et al. "Recipe1M+: A dataset for learning cross-modal embeddings for cooking recipes and food images". IEEE PAMI, 2019.

[8] C. Trattner, D. Elsweiler."Food recommender systems: important contributions, challenges and future research directions". arXiv preprint arXiv:1711.02760. 2017.

[9] A. Salvador, N. Hynes, Y. Aytar, J. Marin, F. Ofli, I. Weber, A. Torralba. "Learning cross-modal embeddings for cooking recipes and food images." In CVPR, 2017.

[10] M. Carvalho, R. Cadene, D. Picard, L. Soulier, N. Thome, M. Cord. "Cross-modal retrieval in the cooking context: Learning semantic text-image embeddings." In SIGIR, 2018.

[11] J-J. Chen, C-W. Ngo, F-L. Feng, T-S. Chua. "Deep understanding of cooking procedure for cross-modal recipe retrieval." In Multimedia Conf, 2018.

[12] M. Fontanellaz, S. Christodoulidis, S. Mougiakakou. "Self-Attention and Ingredient-Attention Based Model for Recipe Retrieval from Image Queries." In Wkshp on Multimedia Assisted Dietary Management, 2019.

[13] H. Wang, D. Sahoo, C. Liu, E-P. Lim, S. Hoi. "Learning cross-modal embeddings with adversarial networks for cooking recipes and food images." In CVPR, 2019.

[14] D. Li, M. J. Zaki. "Reciptor: An effective pretrained model for recipe representation learning." In SIGKDD, 2020.

[15] J. Johnson, M. Douze, H. Jegou. "Billion-scale similarity search with GPUs". IEEE Transactions on Big Data, 2019.

[16] J. Devlin, M-W. Chang, K. Lee, K. Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." In NAACL, 2019.

[17] H. Xie, L. Yu, Q. Li. "A hybrid semantic item model for recipe search by example." IEEE Multimedia Symp., 2010.

[18] M. Yasukawa, F. Diaz, G. Druck, N. Tsukada. "Overview of the NTCIR-11 Cooking Recipe Search Task." In NTCIR, 2014.