Towards a Fully Personalized Food Recommendation Tool

Angelos Nezis
Athens University of Business and Economics
Greece
angelosnezis@gmail.com

Haris Papageorgiou
Athena Research and Innovation Center
Artemidos 6 & Epidavrou, Marousi
Greece
xaris@ilsp.gr

Pavlos Georgiadis
We Deliver Taste
Moldavská 11, Prague 10 10100
Czech Republic
pavlos@wedelivertaste.com

Petr Jiskra
We Deliver Taste
Moldavská 11, Prague 10 10100
Czech Republic
petr@wedelivertaste.com

Dimitris Pappas
Athena Research and Innovation Center
Artemidos 6 & Epidavrou, Marousi
Greece
dpappas@ilsp.gr

Maria Pontiki
Athena Research and Innovation Center
Artemidos 6 & Epidavrou, Marousi
Greece
mpontiki@ilsp.gr

ABSTRACT

We present [1] a personalized ingredient-based Deep Learning recommender on the food domain that exploits ingredients and nutrition information to create recipe representations and propose to every user a more personalized and healthier meal. The recommender will be a critical component in our Meal Prediction Tool (MPT) designed with a focus on the personalization of services, increasing business efficiency and sustainability in the hospitality, restaurant and catering (HoReCa) industry.

CCS CONCEPTS

• **Information Systems → Information Retrieval**; Recommender Systems • **Information Systems Applications→** World Wide Web

KEYWORDS

Food Recommender, Deep Learning, HoReCa industry

1 INTRODUCTION

Social media like TripAdvisor and Yelp are actively changing the global dining landscape, offering the possibility for consumers to interact online with data; social users can generate ratings, recommendations or upload a variety of photos and location data. The exploitation of such data now enters into the mainstream of HoReCa industry, since data-driven services make it possible to predict the needs of diverse and demanding consumer audiences influenced by rapidly evolving food trends and consumption patterns. Analyzing such trends and patterns can help both food consumers and providers in making informed decisions on their activities (e.g. design personalized meal recommenders, empower businesses to consider various types of external influences). At the same time, detailed analysis of recipes and cookbooks available online reveals a wealth of data that may lead to a number of new cultural correlations and findings. Such data can be exploited in order to integrate social factors into the design of food supply chains, putting emphasis on blendings of flavours and popular ingredients, perceptions of different tastes, food identities or healthiness factors. This paper presents a personalized ingredient-based Deep Learning recommender on the recipe domain designed to propose to each user a more personalized and healthier meal based on ingredients and nutrition information (Section 3). The recommender will be a critical component in our MPT [2] developed in the context of the CAPSELLA [3] project (Section 2). The integration of the recommender will offer the additional possibility to filter the MPT personalized recommendations based on criteria such as food healthiness, considering new aspects previously not addressed by mainstream food services.

2 THE CAPSELLA MEAL PREDICTION TOOL

The MPT combines different sets of data coming from actual consumers who are searching for dining options. This data provides insights on the demographics of demand and locations of where these searches are taking place around a city. By sharing data on their specific demands such as cuisine preferences (e.g. Chinese or vegetarian), occasion (e.g. breakfast, lunch), and specific meals (i.e. by submitting queries in the respective user interface), consumers receive personalized recommendations on different restaurants which match their food preferences. The results are ranked based on the proximity of a restaurant to the user location and sentiment ratings (Fig. 1).

[2] http://meal-pilot.herokuapp.com/#!/search
[3] http://www.capsella.eu/

Figure 1: Meal Prediction Tool Customer Interface.

This data then becomes useful for restaurant owners, who now have a way to predict demand, identify food trends in real time and understand the demographics of their audiences. This user generated data is then combined with other data coming from restaurant menus and social media, which are further analysed to deliver enhanced recommendations that are accurate, personalized and on real time.

3 INGREDIENT-BASED RECOMMENDER

3.1 Related Work

Recommendation of recipes can be tackled by breaking recipes to ingredients and scoring them based on the ingredients that a user has rated positively [1], or by taking also into account the ingredients that a user dislikes [2]. The use of ingredient complements and networks of substitutes that identify which ingredients fit well together and which can be substituted has been proposed [3] as a means to get accurate results. Other approaches [4-5] focus on recommending not only desired recipes, but also healthy ones. For example, the authors in [5] apply a post-filtering approach and re-rank the recommendations according to their healthiness scores. Given that user preferences variations are mainly context-dependent, in this paper we propose a content-oriented approach that exploits ingredients as well as nutrition information in order to create recipe embeddings, and to propose to each user a more personalized and healthier meal.

3.2 Data Collection

Our dataset encompasses 40.225 recipes along with the respective metadata (i.e. recipe name, description, category, preparation steps, ingredients list, nutrition information, user reviews & ratings) collected from allrecipes.com using a standard web-crawler. To determine the healthiness of a recipe we followed the international standards introduced by the Food Standards Agency (FSA[4]) [6], and we calculated a score for each of the following nutrients: fat, saturated fat, sugar and sodium measured in 100g of each recipe. The FSA score relates only to these four macro-nutrients. The scale is green for "healthy", amber for "less healthy" and red for "unhealthy". We derived a single metric following the work of [7] that first assigns an integer value to each color (green=1, amber=2, red=3), and then sums up the scores so

as to derive a total score for the recipe. The final score can vary from 4 (very healthy recipe) to 12 (very unhealthy recipe).

3.3 Methodology

We implemented a Paragraph Vector method [8] to identify similar recipes based on recipe embeddings similarity. We used an embedding size of 50 and the distributed Bag of Words (PV-DBOW) model instead of the distributed memory model (PV-DM), since it requires less data storage. The PV-DBOW model uses as input paragraph vectors and tries to predict other words in the paragraph. To get the recipe embeddings we used vectors with processed ingredients data (i.e. after removing blacklist words like cup, kg, etc.) as input for the Doc2Vec model, and for each user's preferred recipe we got the **n** most similar recipes (Fig 2). In this way, we created a list of recipes similar to the recipes rated by every user. We did the assumption that since a user has rated a recipe, she also likes its ingredients. Finally, we sorted the final

	MAP@5	FSA@5	MAP@10	FSA@10	MAP@15	FSA@15
Similarity	0.0016	10.483	0.0019	9.475	0.0021	9.213
FSA Score	0.0012	8.739	0.0015	7.906	0.0017	7.659

recipe list on similarity index and FSA score and recommended the top-k recipes. To validate the recommendation results, we have used a 5-fold cross validation and the MAP@k performance metric. Our task was to predict the top-k recipes that each user would probably rate higher. At second phase B we set n=2 (i.e. for each recipe a user has rated we selected the 2 most similar). We report the results after sorting on similarity index as well as on FSA Score. The mean FSA Scores for the top 5, 10 and 15 recommendations are also reported (Table 1). As expected, if we sort on similarity index after step B we get better accuracies and worst FSA Scores.

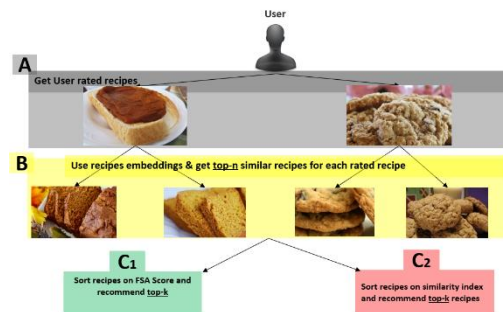**Table 1: Recommender accuracy for k=5, 10 and 15**

Figure 2: Personalized ingredient-based recommender.

4 FUTURE WORK

Future work will focus on the development of a fully personalized meal recommendation engine that will consider user preferences, nutritional needs or restrictions, allergy information and cooking habits in order to make a more robust recommendation.

---

[4] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/300886/2902158_FoP_Nutrition_2014.pdf

REFERENCES

[1] Jill Freyne and Shlomo Berkovsky, 2010. "Intelligent food planning: Personalized recipe recommendation" *In Proceedings of the 15th international conference on Intelligent User Interfaces (IUI),* New York, ACM, 321-324. DOI: https://doi.org/10.1145/1719970.1720021

[2] Morgan Harvey, Bernd Ludwig and David Elsweiler, 2013. "You are what you eat: Learning user tastes for rating prediction". *In Proceedings of the 20th International Symposium on String Processing and Information Retrieval* – Vol. 8214, NY, USA: Springer-Verlag New York, Inc. 153–164, DOI: https://doi.org/10.1007/978-3-319-02432-5_19, 2013.

[3] Chun-Yuen Teng, Yu-Ru Lin and Lada Adamic, 2012. "Recipe recommendation using ingredient networks," *In Proceedings of the 4th Annual ACM Web Science Conference, 298–307*. DOI: https://doi.org/10.1145/2380718.2380757

[4] Gijs Geleijnse, Peggy Nachtigall, Pim van Kaam and Luciënne Wijgergangs, 2011. "A personalized recipe advice system to promote healthful choises" *In Proceedings of the 16th international conference on Intelligent user interfaces (IUI), ACM ,* 437–438.

[5] 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, 2017.

[6] FSA guide to creating a front of pack (fop) nutrition label for pre-packed products sold through retail outlets available.

[7] Gary Sacks, Mike Rayner and Boyd Swinburn, 2009. "Impact of front-of-pack 'traffic-light' nutrition labelling on consumer food purchases in the UK" *Health promotion International,* vol. 24, Issue 4, 344-352. DOI: https://doi.org/10.1093/heapro/dap032

[8] Quoc Le and Tomas Mikolov, 2014. "Distributed Representations of Sentences and Documents" *In Proceedings of the 31st International Conference on Machine Learning (ICML),* 1188-1196.