RESEARCH ARTICLE OPEN ACCESS

Hybrid Food Recommender System based on cancer beating score of ingredients

Rajendra Pawar[1], Shruti Gupta[2], Harshaan Arora[3], Jash Mehta[4], Arya Patil[5]
School of Computer Engineering & Technology, MIT World Peace University, Pune, Maharashtra, India.

Abstract
Recent studies have indicated that about 35% of cancer development is linked to a poor diet. Thus, we introduce an architecture of a food recommendation system that is based on the anti-cancer molecules present in the set of ingredients of a recipe belonging to a cuisine. A hybrid recommender system is chosen to cope with the limitations of a single methodology. The system recommends a list of recipes to users whilst considering their preferences, their recipe ratings as well as cancer beating properties of each recipe. This will provide a way to determine a strategy for cancer-preventing measures, keeping nutrition at the forefront.

Keywords – Cancer Beating Molecules, Food Recommendation System, Healthy Food, User Preferences

---

Date of Submission: 15-06-2021 Date of Acceptance: 30-06-2021

I. INTRODUCTION

A healthy and nutritional diet plays a massive role in preventing illnesses like cancer among people, especially recovering patients. Because of this, many dietary food recommender systems have emerged to keep in check one's day-to-day intakes and eating habits. Despite this, there are no specific and niche food recommender systems based solely on how food interacts with cancer-beating compounds.

One such idea was for a recommender system for Hyperfoods [1], which shows how different cuisines have different anti-cancer molecules based on cancer-beating properties of the ingredients involved in the recipes of those cuisines. However, it focuses solely on suggesting alternative ingredients for a recipe rather than suggesting similar recipes. Recommending a curated recipe list based on user history and preferences makes the system more personable. It gives a choice to the user to try new recipes rather than eating a healthier version of the recipes that they already know.

We propose an extended architecture to this recommender system which recommends not only alternate ingredients but also similar recipes. The system is then a personalized recommender system designed to help user broaden their palate while keeping their preferences at the forefront.

The next section gives an overview of related work, after that Section 3 focuses on our proposed architecture based on research on various recommender systems in the domain of food and restaurants and Section 4 is the conclusion and talks briefly about future work.

II. RELATED WORK

Foods that can help in preventing cancer need to be identified and obtained so that they can also satisfy nutritional and caloric needs. 35% of cancer development is associated with a poor diet [2]. Because of their undergoing treatment, cancer patients are suggested diets rich in calcium, antioxidants, fiber, etc. According to [3], immunomodulating and antineoplastic agents, along with drugs targeting the respiratory system, alimentary tract and metabolism, cardiovascular system, and nervous system make up for the contents of the highest-selling cancer beating drugs.

It has already been established that certain foods affect the working of drugs, making meal planning and recipe selection for a patient a difficult task. Climate is a characteristic of a region that impacts the common ingredients found in recipes. These recipes are tied to cultural backgrounds resulting in different cuisines.

As varied cuisines are on the rise, it is vital to understand the interaction of these cuisines with the drugs mentioned above. For that, a quantitative factor that aids in making a well-formed comparison has been proposed in [3]. Analyzing the negative interaction of a drug with a cuisine results in a threshold value or premise levels of interaction of drugs and cuisines; the value obtained is the chance of a negative food interaction with the drug.

Taking this into consideration, [1] presented an idea to map cancer-beating molecules in food using Machine Learning. This resulted in a recommender system that can predict ingredients of the recipe using an image inputted to the system by a URL and suggests healthy alternatives to the recipe's ingredients based on their proximity in the embedded vectorial space to the ones identified by the Inverse Cooking Algorithm [4].

The algorithm predicts which cuisine the recipe belongs to from the resultant set of ingredients of that recipe and estimates the probability of negative drug interaction of the recipe with antineoplastic drug based on the predicted cuisine.

This was achieved by taking ingredients from the Recipe 1M+ dataset [5] and K&N datasets to create a vocabulary that was then embedded into a vectorial space to reduce the dimensionality. This was used by an SVC classifier to predict the cuisines from each set of ingredients that made a recipe, taking into consideration the parameters and functions, accounting for the dimensions of the training set. Next, the recipes and cuisines were classified based on cancer-beating molecules present to predict their negative drug interactions.

Finally, the system thus displays which cuisine the recipe belongs to highlighted by green or red based on the cancer beating score calculated. Green indicates that the cuisine to which the recipe is a good choice and red means it is not. Substitutes to the ingredients of the recipe are also recommended (e.g. - if the recipe has rice, the recommendation will be to use brown rice or Bok choy and so on).

What we propose is an architecture that is an extension to the above recommender system which recommends recipes rather than alternative substitutes to the ingredients of a recipe, based on a user's cuisine preferences and if they liked or disliked the recipes which were recommended to them. To deal with the limitations of a single technique, we propose a Hybrid Recommender System combining both collaborative and content-based recommender techniques. The most widely used rating prediction technique for CF is matrix factorization [6].

In matrix factorization, a given matrix of numerical ratings (e.g., Likert scale) of users for items $\{r_{ui}\}_{m\times n}$ (m users and n items) is decomposed into two lower dimensional matrices, so that a rating prediction for all the unknown entries in the original matrix $\{r_{ui}\}$ can be computed [7]. Vectors of Abstract factors are learned by mining these ratings and both users and items are modelled with them. The relationships between items and users are used for predicting missing ratings and generating recommendations [6].

Some variants of this algorithm exploit additional data for improving the accuracy of prediction. SVD++ [7] is one such popular variant. It considers implicit feedback provided by the user such as the user's eating history, using which proved highly valuable for the prediction algorithm. gSVD++, an extension to SVD++, was proposed that uses explicitly stated preferences, e.g., user ratings, besides the user's implicit feedback to train the recommendation model [8].

To create a food recommendation system, it is important to consider the sequencing and combination of food recommendation components.

The interaction between the user and system includes:

1. Preference Elicitation Step: user rates and tags recipes.
2. Recommendation Step: user is presented with recommendations.

The first step, preference elicitation, focuses on collecting the long term (stable) user preferences, i.e., their general taste. In our proposed system, it is completed in two stages:

(1) asking the users to choose cuisines they like
(2) rating and tagging some recipes that the user has experienced

The system needs to explore a bit more of the user's preferences by presenting additional recipes for the user to rate. These recipes are determined by speculating what the user may have eaten but has not yet marked in the first step. Active learning techniques such as Binary Prediction technique [9, 10] can be used for finding such recipes.

This implicit feedback of the user's level of satisfaction with the rated recipe was then converted to confidence levels. [11]. Three CF algorithms namely Alternating Least Squares (ALS) [12], Bayesian Personalized Ranking (BPR) [13] and Logistic Matrix Factorization (LMF) [14] were tested. The best performing algorithm is ALS [11].

In [11], the system relies on ALS to output a preference score $s(r, u)_p \in [0, 1]$, for each recipe and user combination. A health score independent of users' preferences, is also assigned to the recipes $s(r)_h \in [0, 1]$.

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

$$s(r,u) = \frac{w_p \times s(r,u)_p + w_h \times (1 - s(r)_h)}{w_p + w_h}$$

where w_p is the weight assigned to the preference factor and w_h to the health factor. Recipes are ranked and recommended to the user based on this score . The users can select values for w_h and w_p to define if the recommendations should be more taste-oriented or more health-oriented.

(1) Preference-based recommender. ($w_h = 0$)
(2) Healthy recommender ($w_p = 0$)
(3) Hybrid recommender: ($w_{p,} w_h \neq 0$)

III. PROPOSED ARCHITECTURE

Our hybrid recommendation system comprises of both content-based and collaborative filtering recommendation systems. After a user logs into the system, preferences are collected based on [8].

In the first layer we have two blocks – the content-based(CB) block and the collaborative filtering(CF) block as shown in Fig. 1.

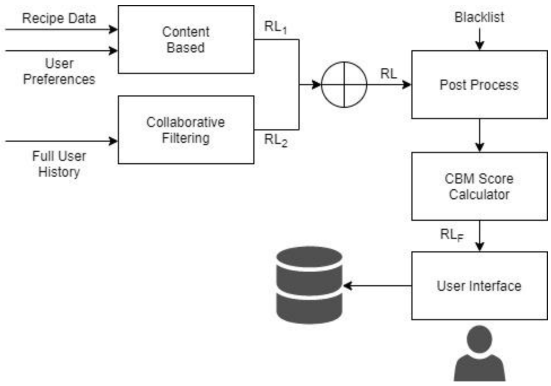

Figure 1. architecture of the hybrid recommender system

In the CB block, we input the recipe preference data collected by the system, along with the recipe data. The system filters recipes according to user preference and returns a set of recipes (RL1), whose preference score $\mathrm{s(r)_p}$ is defined as (2):

$$\mathrm{s(r)_p} = 1$$

In the CF block, we input the full user history data, to generate a distance graph of user preferences, using cosine similarity. This creates clusters of users with similar taste. We then filter 15 neighbours closest to the current user and retrieve recipes that have been liked by these users. The final preference scores for these recipes are calculated using (3):

$$\mathrm{s(r)}_p = \frac{\mathrm{d_{max}} - \mathrm{d_n}}{\mathrm{d_{max}}} * \mathrm{r_n}$$

where d_{max} is the maximum distance of a neighbour from the user, d_n is the distance of user from neighbour who liked the recipe, r_n is the neighbours rating. The preference scores are then scaled, and a set of recipes(RL2) with preference scores is generated.

The union of recipe sets RL1 and RL2 is passed through a post process block which removes any recipes previously liked/disliked by the user. The resulting set is inputted to the CBM Block to get their associated health scores $s(r)_h \in [0, 1]$. We use the architecture proposed in [1], to determine the health scores of each recipe. These health scores are independent of the preference scores. The final score for a recipe is then calculated using (4):

$$s(r) = \frac{w_p \times \mathrm{s(r)}_p + w_h \times \mathrm{s(r)}_h}{w_p + w_h}$$

The scores are then used to rank and display them. The user can browse and rate these recipes, which will be added to user history.

IV. CONCLUSION

In this paper, we proposed the architecture for a Hybrid food recommendation system that can be used at health care centers to prepare diet charts accordingly or by individuals on their own devices. The approach consisted of a content-based system that filtered the recipes according to user preference and returned a set of recipes along with a Collaborative Filter that considered the user history, thus introducing subtle changes to the user's palate which avoids a drastic shift to a rigorous diet.

In the future, implementation of the proposal with a different algorithm with more features pertaining to people with different needs and issues such as diabetes, obesity, etc. might result in a recipe list with varied performance scores and results, the differences to which might be interesting to observe.

REFERENCES

[1]. Veselkov, K., Gonzalez, G., Aljifri, S. et al., HyperFoods: Machine intelligent mapping of cancer-beating molecules in foods, *Sci Rep 9, 9237*, 2019.

[2]. H. Arem and E. Loftfield, Cancer Epidemiology: A Survey of Modifiable Risk Factors for Prevention and Survivorship, *American Journal of Lifestyle Medicine, vol. 12, no. 3*, 2018.

[3]. M. Jovanovik, A. Bogojeska and D. e. a. Trajanov, Inferring Cuisine — Drug Interactions Using the Linked Data Approach, *Scientific Reports, vol. 5, no. 9346*, 2015.

[4]. A. Salvador, M. Drozdzal, X. Giro-i-Nieto and A. Romero, Inverse Cooking: Recipe Generation from Food Images, *Computer Vision and Pattern Recognition*, 2018.

[5]. J. Marín et al., Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1*, 2021.

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

[7]. Y. Koren and R. Bell. Advances in collaborative filtering. In F. Ricci, L. Rokach, B. Shapira, and P. Kantor, editors, Recommender Systems Handbook, *Springer Verlag*, 2011.
[8]. J. Freyne and S. Berkovsky, Intelligent food planning: personalized recipe recommendation, *IUI*, 2010.
[9]. M. Elahi, F. Ricci, and N. Rubens, Active learning strategies for rating elicitation in collaborative filtering: A system-wide perspective, *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2013.
[10]. M. Elahi, F. Ricci, and N. Rubens, Active learning in collaborative filtering recommender systems : E-Commerce and Web Technologies, *Springer International Publishing*, 2014.
[11]. Pecune Florian & Callebert Lucile & Marsella Stacy, A Recommender System for Healthy and Personalized Recipe Recommendations, 2020.
[12]. Morgan Harvey, Bernd Ludwig, and David Elsweiler, You are what you eat: Learning user tastes for rating prediction, *International Symposium on String Processing and Information Retrieval*. *Springer*, 2013.
[13]. Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme, BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618*, 2012.
[14]. Christopher C Johnson, Logistic matrix factorization for implicit feedback data, 2014.

Rajendra Pawar, et. al. "Hybrid Food Recommender System based on cancer beating score of ingredients." *International Journal of Engineering Research and Applications (IJERA)*, vol.11 (6), 2021, pp 31-34.