

A Nutritional Recommender System for Rehabilitation of NCD's by Using Data Mining Techniques

B. Sri Surya

M. Tech,

Department of CSE,

Shri Vishnu Engineering College for Women (A),
Vishnupur,

Bhimavaram, West Godavari District, Andhra
Pradesh.

Mrs. P. R. Sudha Rani

Associate Professor,

Department of CSE,

Shri Vishnu Engineering College for Women (A),
Vishnupur,

Bhimavaram, West Godavari District, Andhra
Pradesh.

Abstract- Non Communicable or Chronic diseases tend to be long duration and are the results of a combination of genetic, physiological, environmental and behaviour factors. About 41 million people died from NCD's each year, which is equivalent to 70% of the global death toll. In many high-income countries, standard care for the long term management of NCD's include Rehabilitation. So, in this paper I proposed that suitable Nutritional Diets play an important role in maintaining health, rehabilitating the NCD's and preventing the occurrence of NCD's. For this purpose, I worked on a novel frame work named NutIngredientFood, which models the relationship between the Ingredients and their Proportions within food for the purpose of offering healthy recommendations. Specifically, NutIngredientFood consists of three main components: 1) using an embedding based ingredient predictor to predict the relevant ingredients with user-defined initial ingredients(Ingredient Predictor), 2) predicting the amounts of the relevant ingredients(Amount Predictor), 3) creating a healthy pseudo-NutIngredient with a list of ingredients and their amounts according to the nutritional information and recommending the top similar nutritional ingredients with the pseudo-Nut Ingredient.

Keywords: NCD, pseudo-NutIngredient

1. Introduction

The presence of food related content on the web has become prominent in recent years. With the large quantity of available data, finding the right recipe becomes a difficult task. Food recommender systems, a domain specific subclass of recommender systems, have been designed to assist in providing users with Nutritional Ingredients that accommodate their preferences.

Positive Nutritional Ingredients Recommenders, which usually rely on a tally to indicate people's aggregated opinion; however are inevitably facing the following challenges that cannot be easily addressed by adopting existing methods. First, the majority of online positive nutritional do not adhere to the nutritional guidelines set by international food agencies and are therefore regarded as unhealthy. Second, a negative correlation is displayed between rating scores and healthiness measures, meaning highly rated nutritional ingredients tend to be even less healthy.

I propose NutIngredientFood, a framework that tackles the completion task whilst striving to retrieve healthy Positive Nutritional Ingredients. While in previous studies the relationships between ingredients are investigated and result in Positive Nutritional Ingredients recommendations for the purpose of list-completion or substitution, here I aim to generate full healthy Positive Nutritional Ingredients recommendations.

Specifically, I first propose an embedding-based ingredient predictor to predict the relevant Positive Nutritional Ingredients with some given ingredients, and then utilize a multi-layer perception-based predictor to capture the quantities of supplemented ingredients. Using an iterative process, with nutritional compatibility considered at each step, I generate a healthy pseudo-NutIngredient with a set of ingredients and the corresponding amounts. This

pseudo-NutIngredient is then matched against the nutritional ingredients in the dataset to find similar nutritional ingredients recommendations.

I propose a Positive Nutritional Ingredients Recommendation framework (NutIngredientFood), which first builds a positive pseudo-NutIngredient considering the nutritional values and then scans the nutritional ingredients dataset for items resembling the pseudo-NutIngredient. Our proposed NutIngredientFood relies not only on the relationship between the ingredients and themselves, but also on those of their quantities, which ultimately dictate the healthiness of nutritional ingredients. To the best of our knowledge, no prior study has incorporated these features.

The pseudo-NutIngredient is a list of ingredients with their quantities, and the nutritional values of the pseudo-NutIngredient should match the predefined targets as best as possible. To generate the pseudo-NutIngredient, I first propose an embedding-based ingredient predictor, which embeds all the ingredients into a latent space and predicts the supplemented ingredients based on the distances of ingredient representations; I propose an amount predictor to compute the quantities of the supplemented ingredients.

I conduct extensive experiments with two nutritional ingredients datasets, and the experimental results confirm the superiority of our methods over the baselines.

2. Related work

Recently, Positive Nutritional Ingredients Recommendation has been a hot research problem. In this project, I focus on Positive Nutritional Ingredients Recommendation and discuss the recent progress of the related methods and applications.

Positive Nutritional Ingredients Recommendation:

Positive Nutritional Ingredients Recommendation is a sub-domain of the larger nutritional ingredients recommendation task, and the problem of recommending healthy nutritional ingredients has been tackled in numerous studies. In literature, healthiness is measured by comparing Macro-Nutrient values against several International Guidelines.

Recommendation with Pre-Existing Positive Nutritional Ingredients:

Traditionally, Positive Nutritional Ingredients Recommendation methods rely on ratings and strive to suggest nutritional ingredients the user would rate high. For example, Positive Nutritional Ingredients ratings to explore the relationship between healthy Positive nutritional ingredients and their ingredients. They devise a break-down technique and examine the feasibility of extracting the users' ingredient preferences from their ratings on nutritional ingredients.

Recommendation with Online Positive Nutritional Ingredients Datasets:

To improve the healthiness of the recommendations, multiple methods are presented with online Positive Nutritional Ingredients Datasets. For instance, post-filtering on the generated recommendations and re-rank them to make the healthier Positive nutritional ingredients appear on the top of the prediction list. The research contributes to a better understanding in online food behavior and benefits healthy Positive Nutritional Ingredients Recommendation. In this project, I focus on Positive Nutritional Ingredients and try to tackle the challenge from another angle. Our method first generates a healthy nutritional ingredients-like draft and then searches the dataset for matching's. Unlike other papers, it does not rely on ratings or on the availability of nutritional information for the Positive nutritional ingredients.

3. Proposed framework: NutIngredientFood

General Framework

The proposed framework NutIngredientFood is inspired by a human-like deliberation and seeks to generate healthy Positive nutritional ingredients recommendations over an online non-healthy dataset. The user's preferences are accommodated in the form of an ingredient-list and are passed to the framework as input. NutIngredientFood consists of three main stages: 1) predicting the relevant and Positive nutritional ingredients with an ingredient-list(Ingredient Predictor); 2) predicting the amounts of the relevant ingredients(Amount Predictor); 3) creating a healthy pseudo-NutIngredient and searching for similar nutritional ingredients in the dataset. The working process of NutIngredientFood is presented in full in Algorithm.

Ingredient Predictor: I propose an embedding-based predictor to find candidate relevant ingredients by computing top compatible ingredient lists with the current ingredient-set prSet (line 3).

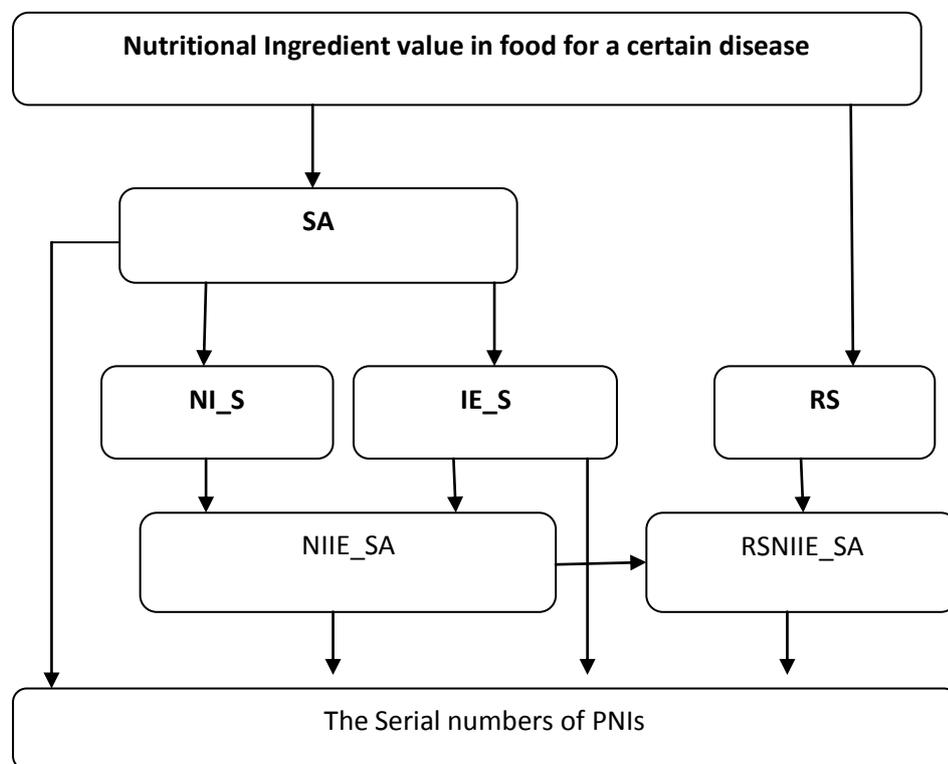


Fig :- Architecture

Amount Predictor: A list of ingredients does not suffice to calculate nutritional values, to accomplish this task i demand the rations. I develop an amount predictor which enables us to predict the quantities of a given ingredient-set (line 4).

Healthy Positive Nutritional Ingredients Recommendation: By using the ingredient predictor and the amount predictor, the FindBestIngredientToAdd assess how each candidate ingredient impacts the nutritional values and returns the best ingredient to append to the working set.

Algorithm:- Best Nutrition Predicting

Input:

s: the initial ingredient-set
k: the number of recommendation

Output:

l: the list of recommended Nutritional Ingredients

Tuning params:

n: the number of ingredients adding to the pseudo-Disease
cos: cosine-weight for similar Disease

```

1: prSet = s
2: for i = 0 to n do
3: candidateIngredients = IngredientPredictor(prSet)
4: amountofCandidateIngredients = AmountPredictor(prSet)
5: bestIngredient = FindBestIngredientToAdd(candidateIngredients, amountofCandidateIngredients)
6: prSet += bestIngredient
7: end for
8: pseudoDisease = CreateDiseaseFromSet(prSet)
9: l = FindSimilarDisease (pseudoDisease, cos, k)
10: return l
    
```

I iteratively execute this process for n times, and generate a healthy pseudo-NutIngredient, i.e. a data-structure containing the ingredients along with their respective amounts (line 9). Finally, with the healthy pseudo-NutIngredient, I search the dataset for the top-k most similar nutritional ingredients (line 10).

The Ingredient Predictor

Model Description :

In order to retrieve the most befitting ingredients, I propose an embedding-based ingredient predictor (IP-embedding), in which the ingredients are projected into a latent space and the ones that usually occur in a positive nutritional are close to each other.

I propose to use a distributed representation method to model the generation of the given nutritional. Given a nutritional: $\{i_1, i_2 \dots i_{N_r}\}$ containing N_r ingredients, the objective function is to maximize the probability of each target ingredient i_a given its corresponding context information:

$$g(i_a) = \Pr(i_a | \text{Context}(i_a)) - \text{eq}(1)$$

Where I consider the ingredients except i_a in the nutritional r as its context. Formally, I model each target ingredient i_a with a d-dimensional embedding vector $v_i \in d$ and embed each contextual ingredient into the same latent vector space. Given an ingredient i_a , I average the embedding vectors of multiple contextual ingredients to compute a contextual vector

$$c_{ia} = 1 / N_r - 1 \sum_{t=1, t \neq a}^{N_r} V'_{it}, - \text{eq}(2)$$

Where V'_{it} is the embedding vector of the contextual ingredient it. Note that, I distinguish between the role of a target ingredient and a contextual ingredient, and represent the same ingredient i_a using two different vectors (v_{ia} and v'_{ia}) depending on which role it takes. I then apply a softmax function to compute the probability of a target ingredient i_a given its contextual vector as follows:

$$\Pr(i_a | \text{Context}(i_a)) = \exp(c_{ia}^T \cdot v_{ia}) / \sum \exp(c_{ia}^T \cdot v_i) - \text{eq}(3)$$

Where is the set of ingredients? Such a model connects the target ingredient and its contexts v_{ia} the embedding representations. For parameter learning, I need to maximize the log probability defined in Eq. (1) over all the nutritional. However, directly optimizing this objective function is impractical because the cost of computing the full softmax for the multi-

classifier to predict the target ingredient is extremely high. Therefore, i adopt the efficient and effective negative sampling strategy to approximate the full softmax. When i train the vector of i_a , i first obtain a corresponding negative sample set $NEG(i_a)$, in which i_x is not the same as i_a if $i_x \in NEG(i_a)$. Then i define

$$L^{i_a}(i) = \begin{cases} 1 & i = i_a; \\ 0 & i' = i_a; \end{cases} \quad (4)$$

Where $L^{i_a}(i)$ is the label of the ingredient i . The labels of positive instances are 1, and 0 otherwise.

Given the contextual ingredients of a target ingredient i_a , i want to maximize the occurrence probability of i_a and meanwhile minimize the occurrence probability of negative samples $i_x \in NEG(i_a)$. The objective function for the target ingredient i_a therefore becomes

$$g(i_a) = \prod \Pr(i | Context(i_a)) \quad (5)$$

$$\text{where } \Pr(i | Context(i_a)) = \begin{cases} \sigma(Cia Vi) & Lia(i) \\ 1 - \sigma(Cia Vi) & Lia(i) \end{cases} \quad (6)$$

in which $\sigma(z) = (1 + \exp(-z))^{-1}$ is the sigmoid function.

Further,

$$\begin{aligned} g(i_a) &= \prod_{i \in i_a \cup NEG(i_a)} \Pr(i | Context(i_a)), \\ &= \prod_{i \in i_a \cup NEG(i_a)} [\sigma(c_{ia}^T v_i)]^{Lia(i_a)} \cdot [1 - \sigma(Cia Vi) Lia(i)], \\ &= \sigma(c_{ia}^T v_{ia}) \prod_{i \in NEG(i_a)} [1 - \sigma(Cia Vi)], \quad (7) \end{aligned}$$

Finally, i define the object function for all the ingredients,

$$\begin{aligned} \ell &= \sum_{r \in R} \sum_{a=1}^{Nr} \log g(i_a), \\ &= \sum_{r \in R} \sum_{a=1}^{Nr} [\log[\sigma(c_{ia}^T v_{ia})] + \sum_{i \in NEG(i_a)} \log[1 - \sigma(c_{ia}^T v_{ia})]], \\ &= \sum_{r \in R} \sum_{a=1}^{Nr} [\log[\sigma(c_{ia}^T v_{ia})] + \sum_{i \in NEG(i_a)} \log[\sigma(-c_{ia}^T v_{ia})]], \quad (8) \end{aligned}$$

Given an incomplete nutritional, i regard all its ingredients as the context, and compute $\Pr(i | Context(i))$ based on Eq. (3) for each candidate ingredient i , and return the top-Ni ingredients as outputs.

Parameter Learning:

IP-embedding needs to learn the embedding vectors of all the ingredients. I represent all the parameters with Θ , and learn the IP-embedding model by using maximum a posterior (MAP): $\hat{\Theta} = \text{argmax} \sum_{r \in R} \sum_{a=1}^{Nr} \log g(i_a) - \frac{1}{2} \lambda \|\Theta\|^2 \quad (9)$

Where $\lambda \|\Theta\|^2$ is the regularization term.

All parameters are trained using the stochastic gradient ascent method. During the training process, the algorithm iterates over the target ingredients of all the nutritional. At each time, a target ingredient i_a with its contexts is used for update. For a certain context (i_a), the target ingredient i_a is a positive sample, and we randomly sample N_e unobserved ingredients as negative samples. After computing the loss function, the error gradient is obtained v_{ia} back propagation and we use the gradient to update the parameters in our model.

$$\begin{aligned} V_{ia} &= V_{ia} + \eta [[1 - \sigma(Cia TVia)]. c_{ia} - \lambda V_{ia}] \\ V_{ix} &= V_{ix} + \eta [[-\sigma(Cia TVix)]. c_{ia} - \lambda V_{ix}] - i_x \in NEG(i_a), \\ V'_{ic} &= V'_{ic} + \eta [[1 - \sigma(Cia TVix)]. v_{ia} - \sum_{i_x \in NEG(i_a)} [\sigma(c_{ia}^T v_{ix}) \cdot v_{ix}] - \lambda V'_{ic} \\ I_c &\in Context(i_a), \quad (10) \end{aligned}$$

where η is the learning rate

Healthy Positive Nutritional Recommendation:

Finding the best supplemented ingredients and generating the pseudo-NutIngredient In order to effectively capture the user's preference and maintain the healthiness, i evaluate the content similarity between the target nutrients and each candidate set. In particular, Mean

Squared Errors (MSE) measure is adopted as the distance measure. The error in this scenario is the distance between the target macro-nutrient values and the predicted ones. Formally:

$$MSE = 1/|N| \sum_{i \in N} (t_i - P_i)^2 \quad (11)$$

Where N is the set of macro-nutrients (protein, carbohydrates, sugars, fat, saturated fat, sodium fiber, etc...), t is the macro-nutrients target values and p is the predicted macro-nutrients values of the current set extended by the inspected candidate. The ingredient which results in the lowest MSE score is chosen. The process halts after n iterations or when no new ingredient can improve upon the error. In this work i have limited the number of iteration to 5, so as to not exceed the average number of positive nutritional ingredients. Additionally, i do not wish to encourage the algorithm to pick ingredients which typically appear in small amounts (and thus do not meaningfully influence the MSE) just for the sake of adding more ingredients. The final product of this step is our pseudo-NutIngredient.

Retrieving Similar Nutritionals:

I point out our aspiration is to find positive nutritionals which resemble the pseudo-NutIngredient both in the ingredients themselves but also their quantities. Unfortunately, these two properties do not necessarily coincide. To mitigate this issue, i prescribe similarity as a weighted average of the Jaccard and cosine metrics:

$$sim = COS_{weight} \cdot sim_{cos} + (1 - COS_{weight}) \cdot sim_{jaccard} \quad (12)$$

Cosine similarity captures the likeness in terms of amounts:

$$sim_{cos} = \mathbf{a}' \cdot \mathbf{b}' / |\mathbf{a}'| \cdot |\mathbf{b}'| = \sum_{j=1}^m \mathbf{a}' \cdot \mathbf{b}' / [\sum_{j=1}^m \mathbf{a}' \cdot \sum_{j=1}^m \mathbf{b}'] \quad (13)$$

where a' is the pseudo-NutIngredient amounts vector and b' is a positive nutritional amounts vector, both of length m (the number of unique ingredients in the dataset).

4. Conclusion

In this project, I have proposed a framework named NutIngredientFood to tackle the healthy Positive Nutritional Ingredients Recommendation problem. To solve this problem, I used the classification techniques between two fact contents, one is the behavioural factors of a person and another one is Nutritional Ingredients. NutIngredientFood first predicts the relevant ingredients and their amounts with some user-defined initial ingredients, and then creates a healthy pseudo-NutIngredient considering the nutritional values. Finally, NutIngredientFood searches the top similar healthy Positive Nutritional Ingredients based on the pseudo-NutIngredient in the dataset. As evident from the presented results, NutIngredientFood is able to improve the average healthiness of a person affected by NCD's without requiring any pre-computed nutritional information for the nutritional ingredients measures, meaning Positive nutritional ingredients tend to be even less healthy. The Mining of data is somewhat difficult to better select the core ingredient, if our perspective is combined with better attribute reduction algorithms the result would be likely be better and in line with reality, which will be our future work direction.

5. Reference

- [1] CNS, "2016 Global Nutrition Report," in Chinese Nutrition Society, 2016.
- [2] WHO, "Global Status Report on Noncommunicable Diseases (2014)," in World Health Organization, 2014.
- [3] S. Balsari, P. Vemulapalli, M. Gofine et al., "A Retrospective Analysis of Hypertension Screening at a Mass Gathering in India: Implications for Non-communicable Disease Control Strategies," *Journal of Human Hypertension*, vol. 31, no. 11, pp. 750–753, 2017.
- [4] DNHFPC of PRC, "Chinese Resident's Chronic Disease and Nutrition (2015)," in National Health and Family Planning Commission of the People's Republic of China, 2015.
- [5] S. Tellier, A. KiabyLars, P. Nissen et al., "Basic Concepts and Current Challenges of

- Public Health in Humanitarian Action,” *International Humanitarian Action*, pp. 229–317, 2017.
- [6] F. Ara1, F. Saleh, S. J. Mumu, F. Afnan and L. Ali, “Awareness Among Bangladeshi Type 2 Diabetic Subjects Regarding Diabetes and Risk Factors of Non-communicable Diseases,” *Diabetologia*, pp. S379, 2011. DOI:10.1007/s00125-011-2276-4.
- [7] QIANZHAN, “Report of Market Prospective and Investment Strategy Planning on China Intelligent medical construction industry (2017-2022),” in *Qianzhan Intelligence CO.LTD*, 2017.
- [8] W. H. Ling, “Progress of Nutritional Prevention and Control on Noncommunicable Chronic Diseases in China,” *China J Dis Control Prev*, vol. 21, no. 3, pp. 215–218, 2017.
- [9] M. B. Margaret, B. K. Barbara and D. Colette, “Developing Health Promotion Workforce Capacity for Addressing Non-communicable Diseases Globally,” *Global Handbook on Noncommunicable Diseases and Health Promotion*, pp. 417–439, 2013.
- [10] M. Williams and H. Moore, “Lumping Versus Splitting: the Need for Biological Data Mining in Precision Medicine,” *BioData Mining*, vol. 8, no. 16, pp. 1–3, 2015.
- [11] G. M. Oppenheimer, “Framingham Heart Study: The First 20 Years,” *Progress in Cardiovascular Diseases*, vol. 53, no. 1, pp. 55–61, 2010.
- [12] W. Y. Jiao, Y. Xue, T. C. He, Y. M. Zhang and P. Y. Wang, “Association Between South Korean Dietary Pattern and Health,” *Food and Nutrition in China*, vol. 23, no. 5, pp. 81–84, 2017.
- [13] K. W. Lee and M. S. Cho, “The Traditional Korean Dietary Pattern Is Associated with Decreased Risk of Metabolic Syndrome: Findings from the Korean National Health and Nutrition Examination Survey 1998- 2009,” *Journal of Medicinal Food*, vol. 17, no. 1, pp. 43–56, 2014.
- [14] H. Du, L. Li, D. Bennett et al., “Fresh Fruit Consumption and Major Cardiovascular Disease in China,” *N Engl J Med*, vol. 374, no. 14, pp. 1332–1343, 2016.
- [15] H. Li, W. Gan, L. Lu et al., “A genome-wide Association Study Identifies GRKS and RASGRP1 as Type 2 Diabetes Loci in Chinese Hans,” *Diabetes*, vol. 62, no. 1, pp. 291–298, 2013.