

Survey on Cooking using Algorithm Based on Data Mining

D.Nithyadevi¹, J.Divya², S.Monika³ and J.Kalaivani⁴

^{1,2,3} Student of Final Year, Information Technology, IFET College of Engineering, Villupuram. India.

⁴ Associate Professor, Information Technology, IFET College of Engineering, Villupuram. India.

Email: nithyadayalanir@gmail.com, divya96vijay@gmail.com, monikasakthi1997@gmail.com, jkvani2k7@gmail.com

Article Received: 24 January 2018

Article Accepted: 27 February 2018

Article Published: 08 April 2018

ABSTRACT

Data mining is the process of discovering new patterns in large data sets involving various methods with the intersection of machine learning, statistics, and database systems. New data patterns can be extracted by applying the methods of data mining. Data mining can be used in the field of cooking to extract new cooking recipes. There are so much of recipe sites and web applications available today. They help the users to find new procedures for cooking various recipes using the ingredients they have. In this paper, we survey the various techniques used to extract the characteristics of the ingredients. We summarize the general features and summarize them as a research in this area.

Keywords: KEGG, miRNA, TCGA, TF- DF, SNCoNMF.

1. INTRODUCTION

In this modern world people are very health conscious, their increasing concerns regarding the intake and safety of food and food related information has accumulated on the internet in an explosive manner, which has made the social networking sites on cooking to become very popular. Information regarding ingredients, cooking methods and new recipes are being added every day. Huge number of user generated recipes can also be posted in community web sites for cooking such as Cookpad and Rakuten Recipe. The posted recipes have their own titles representing the overview of the contents given by users. They sometimes contain cuisine types and words for arrangements. Cuisine type plays central roles in providing more information to make decisions, recipe recommendations and menu planning. Currently recipes are searchable and shareable on famous cooking websites and as a result, number of websites have grown exponentially. This opens possibilities for the direct analysis of recipes without the need for transferring information from books into electronic data.

2. LITERATURE SURVEY

In this section, we focus on various techniques used to find ingredients specific to a particular cuisine and their cooking actions.

1. Shengyu Liu, Qingcai Chen, Shanshan Guan, Xiaolong Wang, Huimiao Shidefines Mining Recipes in Microblog. A cross-data method to mine recipes in Microblog was proposed. In this method, snippets of text relevant to recipes are firstly extracted from Baidu Encyclopaedia. Then the snippets of text which was extracted already are used for training a domain-specific unigram language model and the candidate recipes are mined from the microblog based on the unigram language model. Finally, some heuristic rules are used to identify real recipes from the candidate recipes. It automatically extract training corpus and produce the domain-specific language model but most of tweets are ungrammatical and often contain noisy texts such as abbreviation, emoticons and spelling errors.

2. Kyung-Joong Kim and Chang-Ho Chung describes about a Data science Approach for Global Recipe Data on the Web. In this technique, thousands of recipes, representing different national cuisines, from the site recipessource.com were analysed with the aim of understanding the food culture of various countries by comparing the ingredients used in their food. The recipe analysis identified ingredients and their frequencies of use in the creation of unique recipes. Food Analyser, the program developed in this study, used data from recipes to examine correlations between individual food ingredients in recipes. This study found that each country or ethnic group used authentic ingredients that differed from others and the groupings of these authentic ingredients were essentially location-dependent. The boundaries of food-relevant areas were closely related to levels of precipitation. Meaningful correlations characterizing the food culture of each area can be explained by these authentic ingredients in recipes. It infer the identity of national cuisines based on thousands of internet recipes and, more specifically, their ingredient but the classification of ingredients (vegetarian, meat, fish, diary, beans and so on) could enrich the analysis but it requires additional work to categorize the ingredients.

3. Tomonobu Ozak, Xia Gao and Mako Mizutani defines the Extraction of ingredients characteristics set and the cooking actions on particular cuisine type. In this method, characterization of the cuisine type and the words for arrangement from the aspect of ingredients and cooking actions have to be done. For that purpose, a series of analyses by using a dataset in a Japanese recipe site was conducted. First, statistical methods are applied to explore the relationship among cuisine types. Then, an extended TF-IDF (term frequency and inverse document frequency) is employed to extract important pairs of cooking actions and ingredients. Furthermore, to identify characteristic sets of ingredients for each cuisine type, association rules on ingredients and cuisine types are extracted and evaluated by various evaluation measures. Through the analysis certain sets of ingredients and cooking actions they relate cuisine type deeply were discovered. It tackled the problem of extracting, characteristic sets of ingredients and cooking actions on cuisine types by using statistical analysis and an extended TF-IDF for handling cuisine hierarchy and association analysis with modified evaluation criteria but the results were not sufficient and the Ingredients specific to some cuisines only were identified.

4. Jiawei Luo, Gen Xiang and Chu Pan described about A novel method called SNetCoNMF (Sparse Network regularized non-negative matrix factorization for Co-regulatory modules identification) was proposed which adopts multiple non-negative matrix factorization framework to identify co-regulatory modules including miRNAs, TFs and genes. This method jointly integrates miRNA, TF and gene expression profiles, and additional priori networks were added in a regularized manner. In addition, to avoid the sparsity of these networks, the sparsity penalties were applied to the variables to achieve modular solutions. Using iterative multiplicative updating algorithm the mathematical formulations can be solved effectively. This method can be applied to multiple genomic data including the expression profiles of miRNAs, TFs and genes on breast cancer obtained from TCGA, priori miRNA-gene regulations, TF-gene regulations and

gene-gene interactions. The results show that the miRNAs, TFs and genes of the co-regulatory modules are significantly associated and modules have a reasonable size distribution. Furthermore, the co-regulatory modules are significantly enriched in GO biological processes and KEGG pathways, respectively. It is a novel computational approach to discover miRNA-TF-gene co-regulatory modules by integrating the miRNA/TF/gene expression profiles, target-site information (miRNA-gene and TF-gene regulations) as well as the protein-protein interaction and achieves great performance in identifying miRNA-TF-gene module, especially for TFs quantity.

5. At present the researchers in IBM are doing research on food industry analytics by creating a program which generates original recipes. It ensures that the recipes are innovative, unusual and pleasing to eat. The program works in five steps. In the first step of the program the users are asked to set some parameters for the type of recipe they like to create by selecting a single ingredient, a regional cuisine and then the type of dish. In the next step, the molecular structure, chemical compounds present in each ingredient was analysed according to the user's flavour preferences. Then, based on the parameters which were set on the step one, the computer suggests new cooking ideas based on traditional recipes. At last, based on the novelty and quality the computer recommends the best idea to the user. It will create a revolution in the food industry. Chefs are also provided with innovative cooking ideas. It does not suggest a boring cooking recipe.

3. CONCLUSION AND FUTURE SCOPE

The new innovative recipes can be generated by analyzing the characteristics of existing recipes which is similar to the user search. The recipes for rare ingredients can also generated. More ingredients can be added in the ingredients list. More number of regional cuisines and the type of dishes will be included. Various languages can also be included for better search by various people.

Thus our proposed system will helps the person who doesn't know how to cook a particular recipe. The standard of cooking will be improved and it ensures the novelty and the quality of the recipe. The innovative new recipes can be found which is not common. The incompatible ingredients with the main ingredient of the recipe will be ignored. Most of the applications available today will suggest the user with the ingredients, quantity of ingredients and the procedure for cooking a particular recipe. But it does not ensure whether the suggested ingredients maps with other ingredients or not so this system helps the user by ignoring the incompatible ingredients.

REFERENCES

- [1] Shengyu Liu, Qingcai Chen, Shanshan Guan, Xiaolong Wang, Huimiao Shi "Mining Recipes in Microblog". Asian Language Processing (IALP), 2013 International Conference on IEEE, 17-19 Aug. 2013.

- [2] Muthukumar. N and Ravi. R, 'Hardware Implementation of Architecture Techniques for Fast Efficient loss less Image Compression System', Wireless Personal Communications, Volume. 90, No. 3, pp. 1291-1315, October 2016, SPRINGER.
- [3] Muthukumar. N and Ravi. R, 'The Performance Analysis of Fast Efficient Lossless Satellite Image Compression and Decompression for Wavelet Based Algorithm', Wireless Personal Communications, Volume. 81, No. 2, pp. 839-859, March 2015, SPRINGER.
- [4] Muthukumar. N and Ravi. R, 'VLSI Implementations of Compressive Image Acquisition using Block Based Compression Algorithm', The International Arab Journal of Information Technology, vol. 12, no. 4, pp. 333-339, July 2015.
- [5] Muthukumar. N and Ravi. R, 'Simulation Based VLSI Implementation of Fast Efficient Lossless Image Compression System using Simplified Adjusted Binary Code & Golomb Rice Code', World Academy of Science, Engineering and Technology, Volume. 8, No. 9, pp.1603-1606, 2014.
- [6] Ruban Kingston. M, Muthukumar. N, Ravi. R, 'A Novel Scheme of CMOS VCO Design with reduce number of Transistors using 180nm CAD Tool', International Journal of Applied Engineering Research, Volume. 10, No. 14, pp. 11934-11938, 2015.
- [7] Kyung-Joong Kim and Chang-Ho Chung "Tell Me What You Eat, and I Will Tell You Where You Come From: A Data Science Approach for Global Recipe Data on the Web", : IEEE Access, Volume: 4, pp. 8199 - 8211, 16 August 2016.
- [9] Muthukumar. N and Ravi. R, 'Design and analysis of VLSI based FELICS Algorithm for lossless Image Compression', International Journal of Advanced Research in Technology, Vol. 2, No. 3, pp. 115-119, March 2012.
- [10] Manoj Kumar. B and Muthukumar. N, 'Design of Low power high Speed CASCADED Double Tail Comparator', International Journal of Advanced Research in Biology Engineering Science and Technology, Vol. 2, No. 4, pp.18-22, June 2016.
- [11] N. Muthukumar, 'Analyzing Throughput of MANET with Reduced Packet Loss', Wireless Personal Communications, Vol. 97, No. 1, pp. 565-578, November 2017, SPRINGER.
- [12] Tomonobu Ozak, Xia Gao and Mako Mizutani "Extraction of characteristic sets of ingredients and cooking actions on cuisine type" , Advanced Information Networking and Applications Workshops (WAINA), 2017 31st International Conference on IEEE, 27-29 March 2017
- [13] P.Venkateswari, E.Jebitha Steffy, Dr. N. Muthukumar, 'License Plate cognizance by Ocular Character Perception', International Research Journal of Engineering and Technology, Vol. 5, No. 2, pp. 536-542, February 2018.
- [14] N. Muthukumar, Mrs R.Sonya, Dr.Rajashekhara and Chitra V, 'Computation of Optimum ATC Using Generator Participation Factor in Deregulated System', International Journal of Advanced Research Trends in Engineering and Technology, Vol. 4, No. 1, pp. 8-11, January 2017.

- [15] JiaweiLuo, Gen Xiang and Chu Pan “Discovery of microRNAs and the transcription factors of co-regulatory modules by integrating various types of genomic data”, IEEE Transactions on NanoBioscience, Volume: 16, Issue: 1, Jan. 2017.