

Analysis on Nutritious Balance in Recipes Based on Diversity Indices and Cluster Analysis of Food Ingredients

Jiale Zhuang¹, Junhui Gao²

¹Wuxi Big Bridge Academy, Wuxi, China

²American and European International Study Center, Wuxi, China

Email: jhgao68@163.com

How to cite this paper: Zhuang, J.L. and Gao, J.H. (2018) Analysis on Nutritious Balance in Recipes Based on Diversity Indices and Cluster Analysis of Food Ingredients. *Food and Nutrition Sciences*, 9, 834-839.

<https://doi.org/10.4236/fns.2018.97062>

Received: June 18, 2018

Accepted: July 15, 2018

Published: July 18, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This thesis research on nutritious balance of specific examples in Chinese recipes applying diversity indices and cluster analysis. Initially, based on data of the nutritional ingredients of food, such as proteins, fat and vitamins, we categorize 1200 kinds of specific food using cluster analysis; then, according to a recipe given by a local restaurant, we calculate and compare the diversity indices based on the components of 25 single dishes in the recipe and analyze the nutritious balance of each dish.

Keywords

Components of Food, Cluster Analysis, Recipe, Nutritious Balance, Diversity Indices

1. Introduction

The nutritional components of food include protein, fat, carbohydrates, calorie and minerals. All kinds of foods contain different components and proportions. People eat different kinds of food to satisfy the body's needs for various nutrients.

Liang Jia (2017) proposed that the human body needs a variety of nutrients to maintain the balance of the diet. The catering industry should pay attention to the dietary balance when providing food services for different consumer groups [1].

Zhang Fan, Guo Wenwei and Gao Zhihui (2017) also expressed that as the concept of healthy and green food widely spreads, the nutritious balance has become a problem that the catering industry must face in the process of promoting

the service level and promoting the figure of its own. Therefore, for the catering industry, because of the different regions have different dietary habits, it is necessary to strengthen the scientific dietary combination according to the people of different consumer groups and of different constitutions [2].

Song Ce, Li Hao and Wu Libin (2017) study the proper intake of fruit based on the health of human body. First, the fruits are classified by systematic cluster analysis, according to the nutritional components. Then the annual per capita consumption of representative fruits is predicted with the trend moving average model, comparing the various nutrients needed to protect human health to evaluate the rationality of Chinese residents' annual intake level of minerals, vitamins, dietary fiber and other nutrients. Finally, a single objective linear programming model is established to calculate the annual reasonable per capita consumption of the main fruit products under the condition of nutritional balance to meet the health needs, so as to meet the nutritional health needs with a lower purchase cost [3].

In this paper, a new method for calculating the nutritional balance of recipes is proposed by combining cluster analysis with diversity indices.

2. Materials and Methods

2.1. Data Sources

Our original data mainly consists of food composition data and 25 dishes of food.

Food composition data from [4], containing more than 1200 records, including the sequence number, name, test ingredients, energy, water, protein, fat, vitamins, trace elements, and so on. **Figure 1** shows 6 of them.

The 25 dishes of food are from the Sijinantang restaurant in Wuxi. It contains double fresh with green pepper, original flavor chicken brine, broccoli with mashed garlic, tomato and egg soup and so on. Taking tomato and egg soup as an example, its food composition is shown in **Table 1**.

2.2. The Principle or Concept of Cluster Analysis

Cluster analysis refers to the grouping of physical or abstract objects into multiple classes consisting of similar objects. The goal of cluster analysis is to collect and classify data on similar basis. In different applications, many clustering techniques have been developed, which are used to describe data, to measure the

Name	Edible part	Energy	Water	Protein	Fat	Dietary Fiber	Carbohydr	Vitamin A	Vitamin B1	Vitamin B2	Niacin	Vitamin E	Sodium	Calcium	Iron
millet	100.0	349.0	11.3	13.6	2.7	3.5	67.6		0.3	0.1	1.4	1.8	1.7	30.0	5.7
barley	100.0	307.0	13.1	10.2	1.4	9.9	63.4		0.1	0.1	5.0	0.3	1.6	13.0	5.1
paddy	64.0	359.0	10.2	9.9	2.2	1.4	74.8		0.1	0.1	5.0	0.3	1.6	13.0	5.1
rice	100.0	346.0	13.3	7.4	0.8	0.7	77.2		0.1	0.1	1.9	0.5	3.8	13.0	2.3
instant noodle	100.0	472.0	3.6	9.5	21.1	0.7	60.9		0.1	0.1	0.9	2.3	1144.0	25.0	4.1
bran	100.0	220.0	14.5	15.8	4.0	31.3	30.1	20.0	0.3	0.3	12.5	4.5	12.2	206.0	9.9

Figure 1. Food composition table (part).

Table 1. Food Composition of tomato and egg soup.

Sub item	Mass(grams)	Type	Sub item	Mass(grams)	Type
Tomato	150	Main Ingredient	Salt	5	Accessory
Egg	100	Main Ingredient	MSG	2	Accessory
Agaric	50	Main Ingredient	Sesame Oil	1	Accessory
Green Onion	2	Accessory			

similarity between different data sources, and to classify data sources into different clusters [5].

Chen Jian (2007) briefly introduced the concept and principle of cluster analysis [6]. Cluster analysis, also known as group analysis and point group analysis, is an ideal multivariate statistical technique for classification, mainly including partitioning clustering and hierarchical clustering. Partition clustering is also called segmentation clustering. Given a database of M data objects or tuples, the partition clustering is to construct K partition of data, each partition representing a cluster, and $k \leq m$. That is, it divides data into K groups and meets the following requirements: 1) each group contains at least one object; 2) each object must belong to and only belong to one group, while second requirements in some fuzzy partitioning techniques can be loosened. The hierarchical clustering divides the database into several levels, then divides and clusters the data of different levels, and outputs a hierarchical classification tree. Hierarchical methods can be divided into cohesive and divisive ones. The method of condensation, also known as a bottom-to-head approach, first takes each object as a cluster, and then merges a similar cluster into a growing cluster until all the objects are in a cluster, or a certain terminative condition is satisfied. The splitting method, also known as the head-to-bottom method, first places all the objects in a cluster and gradually subdivides them into smaller clusters, until each object becomes a cluster, or a certain terminative condition is satisfied.

This thesis uses K-MEANS method [7], which is one type of partitioning clustering. The idea of the method is: given the number of the classes K , divide N objects into K classes, making the similarity among the objects in the class the most, and the similarity among the classes is the smallest. The calculation of similarity is carried out according to the average value of the objects in a cluster (seen as the center of clustering); that is, each cluster is represented by the average value of the objects in the cluster. The greatest advantage of the method is simplicity and efficiency. The key of the method is the initial center selection and the formula for calculation of the similarity among classes.

2.3. How to Implement Clustering Analysis

We used Python3.6 and sklearn.cluster's KMeans to cluster analysis, and carried out 20 types and 30 types of cluster analysis respectively. Sklearn is a commonly used Python third party module in machine learning, which can be installed through pip.

2.4. Diversity Induces

The Shannon Wiener diversity index [8] is an index for investigating local diversity in plant communities (alpha diversity), often used in conjunction with the Simpson diversity index. The Shannon Wiener index formula is:

$$H = -\sum (P_i)(\ln P_i) \quad (1)$$

\ln was also replaced by \log_2 , P_i for the proportion of individual species in the total number of individuals. In this formula, H = the information content of samples = the diversity index of the community, S = species number, P_i = the proportion of the individuals belonging to the i species in the sample. If the total number of samples of the sample is N , and the number of i species is n_i , then $P_i = n_i/N$.

In the Shannon Wiener index, there are two components: species number and distribution uniformity among individuals (equability or evenness). The more evenly distributed individuals are, the greater the H value. If each individual belonged to different species, the diversity index would be the largest; if each individual belonged to the same species, the diversity index would be the smallest.

2.5. Calculation of Nutritious Balance

We use formula (1) to calculate the nutritional balance. In the formula, P_i refers to the total mass of each dish in group i and the percentage of the total mass of all ingredients. A total of 20 groups are generated by clustering.

Take salad ribs as an example. The mass percentages of species number 3, 0, 17, and 19 were 0.9597, 0.0192, 0.0154, 0.0058 respectively. The equilibrium index calculated by formula (1) was 0.3017.

In actual calculation, the quality data (content) of each component of the dish is given by the chef of Wuxi Four Seasons restaurant.

3. Results

3.1. Clustering Results in the Food List

Observing the 20 species, the distribution of 1200 foods in 20 species is shown in the **Table 2**.

3.2. Clustering Information of Twenty-Five Dishes

Taking tomato egg soup as an example, on the basis of **Table 2**, ID (identification number of more than 1200 foods) of the subitems, species number of 20 clusters and species number of 30 clusters were added. For example, in **Chart 3** the species number of 20 clusters of eggs is 3.

From **Table 3**, we can see that the subterm salt and monosodium glutamate of the tomato and egg soup belong to the 17 category of the 20 classes, and the quality is added when the formula (1) is used to calculate the nutritional balance.

3.3. Comparison of Nutritious Balance between Twenty-Five Dishes

Using formula (1) and the method mentioned in the preceding 2.5 sections, the

data of nutrient balance of 25 dishes are calculated, as shown in **Table 4**.

4. Discussion

We clustered basic foods into 20 categories, and then analyzed and calculated the nutritional balance of vegetables. Well, is the 20 category reasonable? In fact, we have clustered in 30 categories. The results are different from those of 20 categories, but there is little difference. From **Table 3**, we found that there are 20

Table 2. Distribution and examples of various kinds of food.

Species Number	Number of Food	Examples	Species Number	Number of Food	Examples
0	163	Cabbage, Pear	10	260	Peach, Radish
1	70	Chicken	11	273	Rice, Flour
2	3	Ox liver	12	6	Gomphidius rutilus
3	164	Egg, Intestine	13	1	Sweet pepper
4	139	Walnut, Fish	14	9	Yolk
5	1	Starch	15	3	Mushroom
6	10	Sesame	16	2	Russula Sanguinaria
7	23	Milk powder	17	2	Salt
8	1	Preserved cabbage	18	16	Bamboo shoot
9	29	Salad oil	19	48	Soy sauce

Table 3. Cluster information of tomato and egg soup.

Subitem	Mass (grams)	Type	ID	20 clusters	30 clusters
Tomato	150	Main Ingredient	937	10	29
Egg	100	Main Ingredient	654	3	24
Agaric	50	Main Ingredient	238	18	17
Green Onion	2	Accessory	891	0	7
Salt	5	Accessory	1208	17	16
MSG	2	Accessory	1207	17	16
Sesame Oil	1	Accessory	1109	9	19
				6 species	6 species

Table 4. Comparison of nutritious balance between 25 dishes.

Name	Index	Name	Index	Name	Index
double fresh with green pepper	2.472	Field flavors	1.836	Stinky mandarin fish	1.506
Assorted gluten	2.367	Celery and yam and black fungus	1.713	fried whole fish that looks like a squirrel	1.444
Fat beef in sour soup	2.191	Fresh shellfish with nuts and mushrooms	1.701	baby cabbage in chicken soup	1.373
Fish with Japanese pepper	2.145	Tomato and egg soup	1.655	mushroom with Goose liver sauce	1.333
ring oil eel paste	1.920	Black tofu	1.630	Fresh shellfish with egg white	1.329

and 30 two species of clustering, two subitems of salt and monosodium glutamate, which belong to seventeenth in 20 of the 20, and all belong to the 16 class at the time of 30, and still belong to the same category. In addition, after two clustering methods, the number of subitems is 6. This shows that clustering is changed from 20 to 30, and the result is stable.

In general, the distance between different centers is not equal, which will affect the accuracy of nutrition balance calculation.

The deficiency of this article is that the food ingredients and recipes in China are analyzed only, but not in other countries.

5. Conclusions

In this paper, a new method for calculating the nutritional balance of recipes is proposed by combining cluster analysis with diversity index for the first time. In this paper, 25 Chinese dishes were calculated and compared. Among them, green pepper had the highest nutrition balance index, while the salad had the lowest nutrition balance index.

This method can be applied to food composition and menu analysis in all countries, so it is easy to popularize.

Acknowledgements

The chef of Wuxi four seasons warm hall restaurant provides us with the detailed data of 25 dishes, including the composition of each dish, the quality of the component, and thank you here.

References

- [1] Liang, J. (2017) Principles and Necessities of Balanced Diet and Nutrition in Catering Industry. *Food Safety Weekly*, No. 12, 57-57.
- [2] Zhang, F., Guo, W.W. and Gao, Z.H. (2017) Principles and Necessity of Balanced Diet and Nutrition in Catering Industry. *Consumer Guide*, No. 26.
- [3] Song, C., Li, H. and Wu, L.B. (2017) Analysis of Reasonable Intake of Fruit Based on Nutritional Health Perspective. *Journal of Qiqihar University (Natural Science)*, **4**.
- [4] Wang, G.Y. (2009) Chinese Food Composition List. The Medical Press of Peking University, Beijing.
- [5] <https://baike.baidu.com/item/%E8%81%9A%E7%B1%BB%E5%88%86%E6%9E%90/3450227?fr=aladdin>
- [6] Chen, J. (2007) Analysis of Common Cluster Analysis Algorithms. *Journal of Anhui Electronic Information Vocational and Technical College*, **1**, No. 6.
- [7] MacQueen, J. (1967) Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 281-297.
- [8] Pielou, E.C. (1975) Ecological Diversity. Wiley & Sons, New York.