Research on Food Consumer Price Index Based on Association Rules and Clustering

Yuqi Luo

School of Statistic, Shanxi University of Finance and Economics, Wucheng Road, Taiyuan, China

Abstract: A variety of data mining methods in their daily lives more and more attention, this article through the association rules and clustering, on China's 31 provinces food consumer price index of residents were analyzed. The association rules mainly discover the correlation between the food consumer price classification index. Clustering is based on the food consumer price index of each region and combines its own information to classify the highest similarity into a group. From the difference analysis of consumer price index, we can explore the influencing factors of different regions, so as to provide ideas for regional policy adjustment and benefit the people.

1. Introduction

At present, a large amount of data is generated in every field at every moment, which contains rich information. People have made a lot of explorations on how to use these data to obtain useful information. Data mining can discover useful information from massive data, help businesses understand customer needs, thus creating strong competitiveness for businesses.

Association rules are one of the most popular technologies in data mining. The focus of association rules is to find out whether there is some association between different categories. By mining these rules, we can find strong and weak links between different categories, which can provide the basis for the next adjustment.

Consumer price index (CPI) is a statistical index reflecting the macroeconomic operation of a country or region, and an important basis for analyzing economic operation, monitoring and regulating the overall level of prices. At present, China's CPI index system is not perfect, such as the representative sample survey, the rationality of weight determination, consumer price index settings and other issues to be further discussed, so it is necessary for us to conduct a more in-depth study of CPI index.

2. Introduction of Model and Algorithms

A few years ago, with the classic story of "beer and diapers", data warehouse came into people's vision. "Beer and diapers are sold together" is a classic story of association rule mining. It tells people to use the existing data to mine the law between the data.

2.1 Association Rules

Association rule mining is one of the most important branches of data mining. The first step is to find all frequent itemsets, and then generate strong association rules from frequent itemsets, that is, association rules satisfying minimum support and confidence at the same time.

2.2 Apriori Algorithm

Apriori algorithm is the first association rule mining algorithm and the most classical algorithm. It uses the iteration method of layer-by-layer search to find out the relationship between itemsets in the database, so as to form rules. The process consists of connection (class matrix operation) and pruning (eliminating unnecessary intermediate results). The set containing K items is a set of K items. If a set satisfies the minimum support, it is called a frequent itemset.

2.3 Clustering

Clustering is an unsupervised data mining task and a data mining method to discover interesting patterns among data. The main concept of clustering is to divide the given irregular data into several regular groups. Clustering is an unsupervised classification, which divides data into several clusters. The similarity of data within each cluster is the greatest, and the similarity between clusters is the smallest.

3. Empirical Analysis

Over the past decade, the consumption level and structure of Chinese residents have had tremendous changes. The consumption level of urban residents has been greatly improved, but at the same time, there are huge differences in consumption level and structure in different regions of the country. Facing these changes, we can understand the regional consumption characteristics of Urban Households by studying the regional differences of the national consumer price index, which provides a favorable basis for the formulation of regional macroeconomic policies and targeted development of consumer markets.

By processing the data describing the food consumer price classification index of 31 provinces and municipalities in China, this example shows the association rule modeling and a graph showing the relationship between the food consumer price index. Finally, 31 regions are divided into four categories.

The data source is the consumer price classification index of the regions in the statistical yearbook of 2018. On this basis, the data are modified. The value of each kind of consumer price index in each region is recorded as 0, higher or equal to 1, and the numerical data is changed to classified data.

3.1 Association Rules

First, connect the Type node to the data source, and then connect the node to the Table node. Set the area field identified by the field card to untyped.

Now, the operation flow to "type" node is instantiated and display table. The data set contains 14 fields, each of which represents the consumer price classification index of a region.

Fourteen fields are displayed in the title:

Grain, potatoes, beans, edible oil, vegetables, livestock, poultry, aquatic products, eggs, milk,dried fresh melons and fruits, confectioneries and pastries, condiments, else.

First, we need to use Apriori to get a general idea of the relationship (association) in food to generate Association rules. Select the fields to be used in this modeling process: edit the Type node, set the roles of all food categories to arbitrary, and set all other roles to none.

After specifying the field for modeling, attach the Apriori node to the Type node, edit it, select the option to display only the flag variable whose value is real, and then click Run on the Apriori node.

Table 1 A part of asocciation rules

the latter	the former	support degree	confidence coefficient
edible oil	aquatic products and dried fresh melons and fruits	16.129	100
else	aquatic products and dried fresh melons and fruits	16.129	100
else	potatoes and poultry	22.581	100
eggs	Potatoes and poultry	22.581	100
vegetables	grain and poultry	19.355	100
dried fresh melons and fruits	condiments and vegetables	19.355	100

These rules show that there are many correlations among consumer price indices of various food

categories. Some have two-way Association rules.

For example: edible oil - aquatic products, dried fresh melons and fruits dried and fresh melons and fruits - quatic products, edible oils, else

Then, attach the network node to the "type" node, edit the network node, select all content fields, select only the true flag, and then click "Run" on the network node.
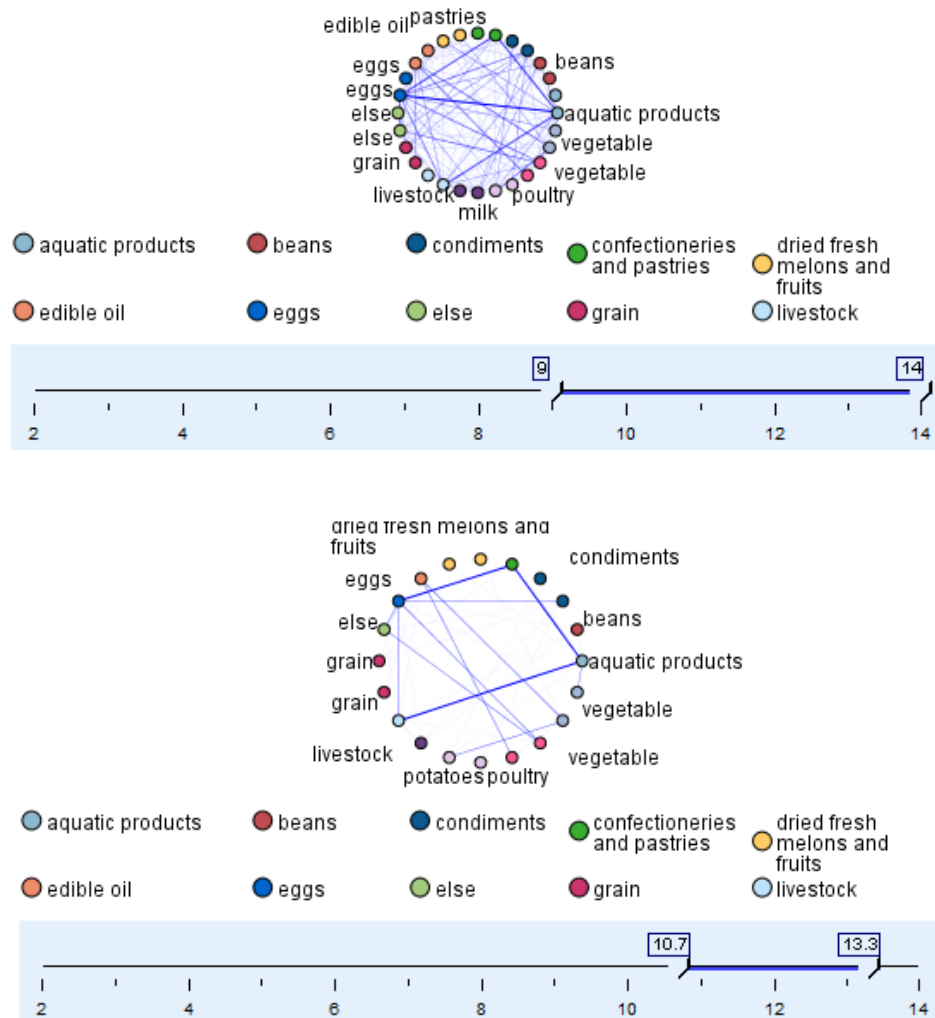

Figure 1 Meshy figures

There are too many links in the graph because there are many classifications of consumer food price index. By adjusting the upper and lower bounds of links below, more strongly related categories can be highlighted.

Three groups will eventually be highlighted:
➢ Egg, confectioneries and pastries and aquatic product
➢ Livestock and aquatic product
➢ Vegetable and edible oil

3.2 Clustering

In addition to finding out the correlation of consumer food price index, 31 regions can be clustered according to these aspects.

First, connect the Type node to the data source, and then connect the node to the Table node. The data are the same as those used in the above methods. Instead of changing the roles of each field, all fields are set as input, and only the area fields are set as targets.

Then, K-Means in "Modeling" is added to the type node to model, and the number of clusters is set to 4. Run.

Then connect the node to the "table" node to get the category of each region.

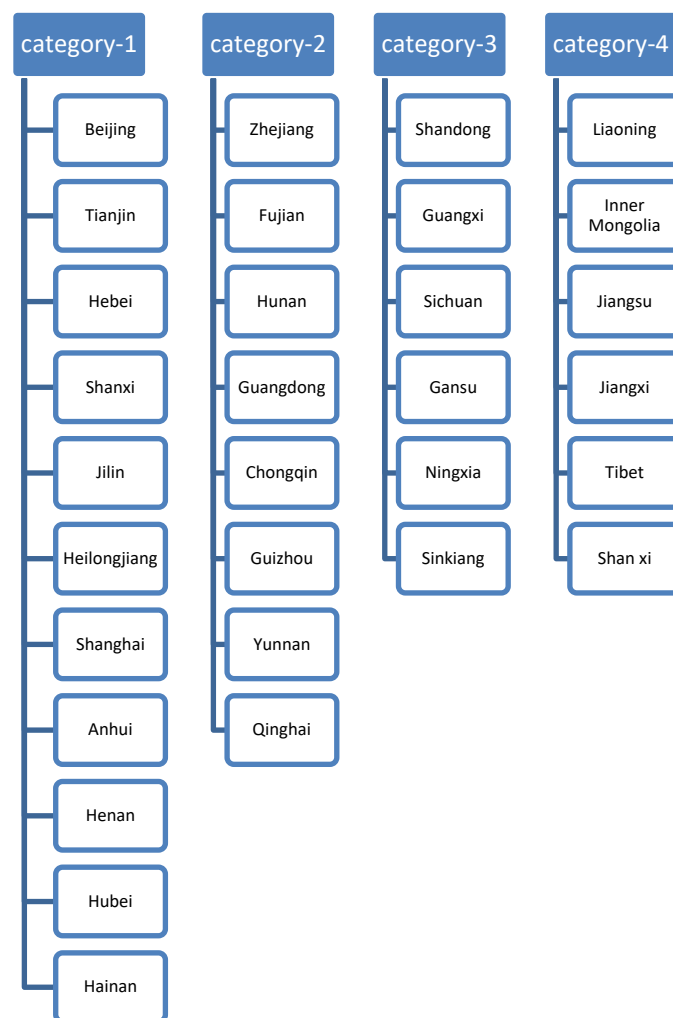| category-1	category-2	category-3	category-4
Beijing	Zhejiang	Shandong	Liaoning
Tianjin	Fujian	Guangxi	Inner Mongolia
Hebei	Hunan	Sichuan	Jiangsu
Shanxi	Guangdong	Gansu	Jiangxi
Jilin	Chongqin	Ningxia	Tibet
Heilongjiang	Guizhou	Sinkiang	Shan xi
Shanghai	Yunnan		
Anhui	Qinghai		
Henan			
Hubei			
Hainan			

Figure 2 Urban cluster

It can be seen from the results that Inner Mongolia, Liaoning, Jiangsu, Jiangxi, Tibet and Shanxi are in one category, Zhejiang, Fujian, Hunan, Guangdong, Chongqing, Guizhou, Yunnan and Qinghai are in one category, Shandong, Guangxi, Sichuan, Gansu, Ningxia and Xinjiang are in one category, and the remaining cities are in one category.

Among them, Inner Mongolia, Liaoning, Jiangsu, Tibet, Shanxi and other basic food consumption price indices are basically higher than the national average level, which shows that these cities spend more on food and clothing. However, more developed cities such as Zhejiang and Hunan pay more attention to meat and dairy, which are highly nutritious.

Further analysis shows that the main positive factors affecting the CPI are the sustained and stable development of economy and the rise of wage level, while the negative factors such as slow growth of income and insufficient effective demand will restrain the growth of consumption.

4. Conclusion

This paper mainly studies the classification index of food consumer price in 31 provinces and municipalities in China. Through clustering and association rules, the potential correlation of each consumer price index is mined, and three indexes with higher degree of correlation are obtained, namely, eggs, confectioneries and pastries, and aquatic products, livestock and aquatic product, vegetable and edible oil.

After that, the regions were clustered into four groups. Because of the regional differences in different regions, the development model and consumption situation are different, and the consumer price index is also different in different regions. Here only the food consumer price index is analyzed, and the regional differences are significant. After that, we can study the CPI from other

aspects, fully excavate the relevant factors in different regions, and provide ideas for the reform of regional governance.

References

[1] Consumer Price Index (CPI)[J]. Rivers and seas, 2007,000 (5):60.

[2] Consumer Price Index (CPI) by Category (2012. 10)[J]. China Textile, 2012,(12):59-59.

[3] TWARDOWSKI PETER (CA).METHOD OF CALCULATING A CONSUMER PRICE INDEX: World Patent Organization, WO0125978 [P].2001-04-12.

[4] Gonglin Yuan, Xiangrong Li. A Modified Consumer Price Index [J].,2010,(2):112.

[5] CONSUMER PRICE INDEX IN 35 MAJOR CITIES IN CHINA [J]. World Economy and China, 1997, (2):60.

[6] McCully, C. P..Reconciliation between the Consumer Price Index and the Personal Consumption Expenditures Price Index[R].

[7] Wei Wang,Tao Wang,Yanyan Shi.Factor Analysis on Consumer Price Index Rising in China from 2005 to 2008[C].//2009 International Conference on Management and Service Science (MASS 2009). [v.9].School of Basic Science, Harbin University of Commerce%School of Economics, Harbin University of Commerce% School of Economics, Harbin University of Commerce, 2009:6041-6044.

[8] Jeff A. Schnepper.Adjusting the Consumer Price Index [J].USA today, 1996, 124(2608):33-0.

[9] BundeyFreya.The Financialized Household and the Consumer Price Index [J].Review of Radical Political Economics, 2015, 47(4):625-640. DOI: 10.1177/0486613415574267.

[10] Yi Liu, Jiawen Peng, and Zhihao Yu. 2018. Big Data Platform Architecture under the Background of Financial Technology: In the Insurance Industry As An Example. In Proceedings of the 2018 International Conference on Big Data Engineering and Technology (BDET 2018). ACM, New York, NY, USA, 31-35. DOI: https://doi-org.ezlibproxy1.ntu.edu.sg/10.1145/3297730.3297743