PROCEEDINGS OF SCIENCE

Price Association Analysis of Agricultural Products based on Apriori Algorithm

Lang Qiao[1]

Capital Normal University Information Engineering College; National Engineering Research Center for Information Technology in Agriculture;Key Laboratory of Agri-informatics, Ministry of Agriculture Beijing Engineering Research Center of Agricultural Internet of Things
Beijing, 100000, China
E-mail: 471831999@qq.com

Cheng Peng[a]**; Xinyu Guo**[2b]**; Yuansheng Wang**[c]

National Engineering Research Center for Information Technology in Agriculture;Key Laboratory of Agri-informatics, Ministry of Agriculture;Beijing Engineering Research Center of Agricultural Internet of Things
Beijing, 100000, China
E-mail: [a]pengc@nercita.org.cn;[b]guoxy@nercita.org.cn;[c]wangys@nercita.org.cn

In this era of big data, it is difficult to extract the information law of agricultural products prices from the vast amount of agricultural products. In this paper, we select 6 kinds of agricultural products prices in 2014, and use the association rule algorithm to analyze the correlation between them, and then get the strong association rules among the prices of agricultural products. Empirical results show that the price of corn in 2014 had a strong correlation with the price of soybean. The price trend of these two is basically the same. In the study of corn or soybean prices, we can refer to either price chart. The experimental results demonstrate the correctness of the algorithm. The application of the algorithm in the analysis of the price of agricultural products can play a positive role in guiding the analysis and prediction of the price of agricultural products.

ISCC2017
16-17 December 2017
Guangzhou, China

[1] Speaker
This work was supported by Beijing Municipal Natural Science Foundation (4151001) and the Natural Science Foundation of China (41501418).
[2] Corresponding author

1. Introduction

In recent years, with the rapid development of information technology, the agricultural informatization has also been developed rapidly in China. In the process, it produces huge amounts of agricultural data, forming big data in the field of agriculture. With the emergence of massive data, it becomes more difficult to find artificial ways of finding new laws[1]. With the Internet technology based on Internet of things and big data becoming more mature, agricultural informatization has reached a climax. The significance of big data is not only to grasp large amounts of data, but also analyze data and other professional methods to realize the value and significance of data [2]. Combining data mining technology with massive agricultural data and applying it to agricultural informatization can play a positive guiding role in the development of agriculture. The methods used in data mining involves feature and comparison, association rules, classification, prediction and clustering analysis etc. In this paper, association rules mining is used to analyze the relevance of prices of agricultural products.

At present, many domestic and foreign scholars have analyzed the price of agricultural products from various aspects, Peng Cheng and so on [3] carried on the analysis to the agricultural product price based on the spatial statistical analysis; Li Si and so on [4] used multi-dimensional scaling method to carry on the analysis on the agricultural product price; Liu Jinshan and Wang Jing analyzed the price of agricultural products based on the GED-ARCH model, from the perspective of finance; Huang Feng and so on [6], analyzed the spatial autocorrelation of the prices of agricultural products during the fluctuation period; Li Zhiqiang and Zhang Yumei [7]compared prices of other agricultural products with vegetables and analyzed them; Wang Dongjie and others [8] analyzed the price change of China's major agricultural products in 2013; Lv Xiaodong [9] analyzed the price of agricultural products in Heilongjiang by summarizing the data, and analyzed the changing trend and fluctuation factors of the prices of agricultural products; Liu Hui and Li Ninghui [10] analyzed the price fluctuation trend of our small farm products by means of HP and BP filtering; Lu Jun and others [11] analyzed the distribution of agricultural prices with time and space based on the time-series model. Many scholars pay more attention to the factors such as time, space and agricultural products themselves when they analyze the price of agricultural products. Few scholars pay attention to the interact of the prices of different agricultural products. This paper uses Apriori algorithm for mining association rules in the analysis of the relationship amongthe prices of agricultural products, mining the association rules among prices of different kinds of agricultural products, and provides a method for mining the rule of the price of agricultural products to show a new direction for the analysis and forecast of the price of agricultural products.

2. Data And Methods

2.1 Association Rules

The application of association rules mining is an important research topic in the field of data mining. As an important branch of data mining, association rule plays an important role in discovering the possible association between objects and objects behind the data. Let I = {I1, I2…Im} is a collection of items. Let T be the transaction table, and Ti is a collection of items and Ti ⊆ I. Let A be a set, the transaction Ti contains A if and only if A ⊆ Ti. Association rules are like the formula that $A \Rightarrow B$, in this formula, A ⊂ I, B ⊂ I, also A ∩ B = Ø [12], and A is called the premise of the rule, and B is called the result of the rule. The intensity of association rules can be measured with support and confidence [12].

The formula for support is: $\mathrm{Support}(A \Rightarrow B) = P(A \cup B)$. The support reveals the probability of simultaneous occurrence of A and B. If A and B occur at the same time, the probability is small, indicating that A has little to do with B; if A and B appear very frequently at the same time, then A and B are always related.

The confidence formula is: $\mathrm{Confidence}(A \Rightarrow B) = P(A \mid B)$. Confidence reveals the probability that B will occur when A appears. If the confidence level is 100%, then A and B can be bundled. If the confidence level is too low, it shows that the occurrence of A has little to do with that of B.

The rules that satisfy both the minimum support threshold (min_sup) and the minimum confidence threshold (min_conf) are called strong rules [12].

The occurrence frequency of an item set is the number of transactions that contain an item set, referred to as the support count of an item set, if the set frequency is greater than or equal to a given minimum support threshold, it can be said a set of frequent itemsets. The frequent K itemsets are used to explore the frequent (k+1) itemsets. The subsets of frequent itemsets are also frequent itemsets. The superset of non frequent itemsets is also non frequent itemsets.

2.2 Apriori Algorithm

Apriori algorithm is a classical algorithm of association rules in data mining proposed by Agrawal et al in 1994 [13]. Apriori algorithm is one of the most influential algorithms for mining frequent itemsets of Boolean association rules [12]. The algorithm uses an iterative method called layer by layer search, and the K itemsets are used to explore (k+1) itemsets [12]. The basic idea of the Apriori algorithm is: first, find the set of frequent 1 itemsets which are denoted as L1, and L1 is used to find the set of 2 frequent itemsets L2, while L2 is used to find L3, so go on until you can't find the K itemsets. Each frequent entry needs a database scan until all the frequent itemsets are found. The frequent itemsets must satisfy the minimum support, and all the nonempty subset of frequent itemsets must be frequent itemsets. Then, strong association rules are generated from frequent itemsets, and these rules must satisfy minimum support and minimum confidence.

2.3 Raw Data

The 2014 data released by the "Chinese Agricultural Statistics Yearbook" and " Compilation of national agricultural products data" compilation of information based on the monthly price of wheat, corn, soybean, peanut meat, rapeseed and Chinese cabbage as the object of analysis on the correlation between the prices of agricultural products were analyzed.

2.4 Data Preparation

In the vast amounts of raw data, there is a large number of messy, repetitive and incomplete datawhich seriously affects the efficiency of data mining algorithms, and may lead to the deviation of mining results. Therefore, in the early stage of data mining, data preprocessing is necesitated For the purpose of data mining, data preprocessing mainly aims at cleaning, integrating, transforming, and protocoling the target data, so as to prepare for the next data analysis work [14].

(1) Data cleansing: removing noise data and extraneous data from raw data sets, dealing with missing data and cleaning dirty data, missing values, identifying and deleting isolating points etc.

(2) Data integration: to combine and store data in multiple data sources in a consistent data store, such as a data warehouse or a data cube. These data sources may include multiple databases, data parties, or general files.

(3) Data protocol: you can use data aggregation, dimensional specification, data compression and numerical compression to obtain a specification representation of the data set, which is much smaller, but still close to the integrity of the original data.

(4) Data transformation: data is transformed into a form suitable for mining by smoothing, aggregation, data generalization, normalization, and attribute construction. In this paper, the variables that need to be analyzed are transformed into Boolean, which is convenient for association analysis.

In these ten transactions, assume that each price increase of agricultural products is an item in a transaction. Set the corresponding numbers for each breed: I1: wheat, I2: corn, I3: soybean, I4: peanut meat, I5: rapeseed, I6: Chinese cabbage.

The transaction table T is as follows:

T	ITEMS
T1	I1, I2
T2	I4, I6
T3	I2, I3, I6
T4	I2, I3, I4
T5	I2, I3, I6
T6	I1, I2, I3, I5, I6
T7	I1, I2, I3, I4, I5
T8	I1, I3, I4, I5
T9	I1, I4
T10	I4, I5

Table 1: Transaction table

3. Introduction Result

We apply the Apriori algorithm to analyze the price information of agricultural products. According to the price information of agricultural products in different periods of the year, the correlation analysis of the change of prices of different varieties of agricultural products was carried out, and according to the analysis results, the association rules can be used to analyze and predict the price of the target agricultural products. This article assumes that the minimum transaction support count is 3(min_sup=3/10=30%), and the minimum confidence threshold is 80% (min_conf=80%).

First, the algorithm scans all transactions in the transaction table, and counts the number of times each item appears, and satisfies the minimum number of transaction support counts, and obtains the frequent 1 item set L1, such as table 2. The Apriori algorithm generates C2 a set of candidates 2 itemset by L1 and records each transaction that is scanned in the process of generating C2. In C2, the items that satisfy the minimum transaction support count will form frequent 2 itemsets L2, such as table 3. In this loop, we'll get L3, such as table 4. C4 cannot be generated by L3, therefore C4=Ø, and the algorithm terminates, and all frequent itemsets have been found.

L1	
Itemset	Support
I1	5
I2	6
I3	6
I4	6
I5	4
I6	4

Table 2: Frequent 1 Item Sets

L2	
Itemset	Support
I1, I2	3
I1, I3	3
I1, I4	3
I1, I5	3
I2, I3	5
I2, I6	3
I3, I4	3
I3, I5	3
I3, I6	3

Table 3: Frequent 2 Item Sets

L3	
Itemset	Support
I1, I3, I5	3
I2, I3, I6	3

Table 4: Frequent 3 Item Sets

According to the set minimum support count and minimum confidence threshold, it is easy to generate strong association rules by frequent itemsets.Strong association rules table is as follows.

Rule	Support	Confidence
I2=>I3	83%	80%
I3=>I2	83%	80%
I1, I3=>I5	100%	100%
I1, I5=>I3	100%	100%
I3, I5=>I1	100%	100%
I2, I6=>I3	100%	100%
I3, I6=>I2	100%	100%

Table 5: Strong Association Tule

According to the strong association rules table obtained, higher prices for wheat and soybean could lead to higher rapeseed prices; higher prices for wheat and rapeseed could lead to higher soybean prices; higher prices for soybean and rapeseed could lead to higher wheat prices ; higher prices for corn and Chinese cabbage could lead to higher soybean prices; higher prices for soybean and Chinese cabbage could lead to higher corn prices; the rise in prices for corn and soybeans is relevant. In this paper, the association rules are analyzed with corn and soybean as examples: I2 represents corn and I3 stands for soybeans. In the association rules

I2=>I3, support degree is 83% said that the probability is 83% that corn and soybeans are rising at the same time, confidence degree is 80% said that the probability is 83% that soybean prices also rise when corn prices rise. In the association rules I3=>I2, support degree is 83% said that the probability is 83% that corn and soybeans are rising at the same time, confidence degree is 80% said that the probability is 83% that corn prices also rise when soybean prices rise. Rule I2=>I3 and I3=>I2 exist simultaneously, indicating that the higher corn price is closely related to the higher soybean price. In Figure 1, the horizontal axis represents the ten times points, the left vertical axis is the soybean price and the right vertical axis is corn price. To a certain extent, the price line of corn is consistent with the price trend of soybean, that is, when studying the price trend of corn, we can refer to the price trend of soybeans. We can refer to the price trend of corn when we study the price trend of soybeans.

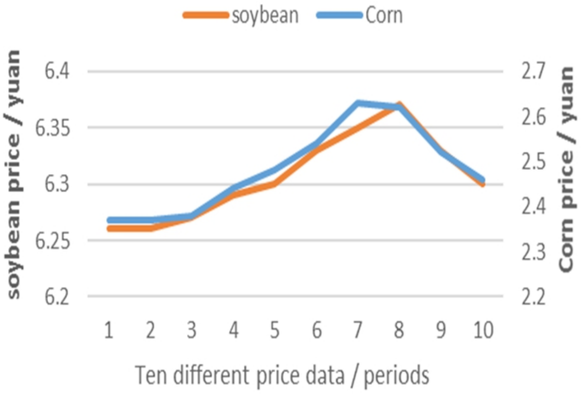

Figure 1: Price Chart for Soybeans and Corn

4. Introduction Conclusion

Traditional methods turns out to be ineffective in analyzing the price of agricultural products from the correlation among prices of agricultural products. According to the experimental results, when the government strengthens and improves market regulation of Corn prices, we should pay close attention to the fluctuation of soybean and Chinese cabbage prices; when the government strengthens and improves market regulation of soybean prices, we should pay close attention to the changing trend of Corn and Chinese cabbage prices; when the government is monitoring the price of soybeans, it is important to focus on the changes in prices of wheat and rapeseed; when the government is monitoring the price of wheat, it is important to focus on the changes in prices of soybeans and rapeseed; when the government is monitoring the price of rapeseed, it is important to focus on the changes in prices of wheat and soybeans; the corn and soybean industry is dependent on each other in China. In this paper, the association rules mining method in data mining is applied to analyze the price of agricultural products, and the Apriori algorithm is used to analyze the price of agricultural products from the correlation of prices of different agricultural products. In this paper, we analyze the correlation of 6 agricultural price data and calculate the strong association rulesaccording to the comparison between the calculation result and the price chart. It is proved that the Apriori algorithm is correct and feasible for the analysis of the correlation between the prices of the agricultural products. After the process of studying the prices of agricultural products, we can do a more comprehensive analysis by referring to the prices of agricultural products that are strongly associated with them.

Experiments have proved that the Apriori algorithm is applied to agriculture, and it is a correct and effective method to analyze the correlation between the prices of agricultural products and to analyze and predict the prices of agricultural products. However, this research used data fromthe national average data because data of agricultural products in different seasons in different regions are very different.Next we will combinethe two dimensions of time and space in the study of agricultural product data,to analyze and process the data from many aspects at the same time.

References

[1] LIU Hui-min. *Research on the Application of Apriori Algorithm in Association Analysis of the price of goods* [J]. Information & Communications , 2012 (4) :29-31

[2] JIA Kebin, LI Hanjing, YUAN Ye. *Application of Data Mining in Mobile Health System Based on Apriori Algorithm* [J]. Journal of Beijing University of Technology , 2017, 43 (3) :394-401

[3] PENG Cheng, WU Hua-rui, HUANG Feng, QIN Xiang-yang, WANG Yi-hong, LIU Yan-ping. *Data Mining on Agricultural Products'Price Based on Spatial Statistics* [J]. Research of Agricultural Modernization, 2014, 35 (01):000029-32

[4] LI Si, CHANG An-ding, ZHANG Meng-qian, WANG Lin-ru. *Price analysis of agricultural products based on multidimensional scaling* [J]. Science & Technology Ecnony Market,2017(5)

[5] LIU Jin-shan, WANG Jing. *Price analysis of agricultural products from the perspective of Finance -- Based on GED-ARCH model* [J]. Journal of Finance and Economics, 2015(01):30-34

[6] HUANG Feng, ZHA0 Chun-jiang, PENG Cheng, WU Hua-rui. *Analysis of spatial autocorrelation during fluctuation of agricultural products' price* [J]. Computer Engineering and Design, 2016, 37 (01) :275-280

[7] LI Zhi-qiang, ZHANG Yu-mei. *A comparative analysis of price changes of vegetables and other agricultural products* [J]. China Vegetables, 2013, 1 (09) :1-6

[8] WANG Dong-jie, DONG Xiao-xia, LI Zhe-min. *Analysis of the characteristics and causes of the price change of China's major agricultural products in 2013* [J]. Prices Monthly,2014(9):11-16

[9] Lv Xiaodong. *Price analysis of agricultural products in Heilongjiang- Based on Provincial fixed observation points in rural areas in 2011 aggregated data*[J]. Modernizing Agriculture, 2012 (4): 50-52

[10] Liu Hui, Li Ninghui. *Price fluctuation trend of small farm products in China and its prediction- An analysis of mung beans*[J]. Price:Theory & Practice, 2012 (6) :57-58

[11] Lu Jun, Song Junhui. *Research on agricultural price data mining based on time series model*[J]. China CIO News , 2011 (8) :29-30

[12] Han Jiawei, Mieheline Kamber, Fan Ming, YU Xiao-feng. Translation. *Data mining: concepts and techniques* [M]. Beijing: Machinery Industry Press, 2001.

[13] AGRAWAL R, RAKESH, SRIKANT R, et al. *Fast algorithms for mining association rules in large databases*[J]. Journal of Computer Science & Technology, 2000,15(6):619-624.

[14] CHENG Yuan, ZENG Xi-fang. *Application of Apriori Algorithm for Type 2 diabetes and its complications* [J]. Laser Journal, 2011, 32 (1) :82-83.