Advance Journal of Food Science and Technology 11(6): 439-443, 2016
DOI:10.19026/ajfst.11.2657
ISSN: 2042-4868; e-ISSN: 2042-4876

Submitted: June 24, 2015 Accepted: August 5, 2015 Published: June 25, 2016

Research Article

Study on Retail Price Classification of Food Based on Data Mining Technology

[1]Han Tian-Peng, [2]Bai Ling-Ling and [1]Wang Dai-Mu
[1]School of Computer and Information Engineering, Fuyang Teachers College,
[2]Information Management Center, Fuyang Party Institute of CCP, Fuyang Anhui 236037, China

Abstract: In order to carry out classification of food retail price correctly, the data mining technology is applied in it. Firstly, the real situation of Chinese food retail price is analyzed. Secondly, the basic theory of data mining is studied in depth. Thirdly, the corresponding algorithm procedure is design. Finally, Chinese food retail price fluctuating and classification rules are obtained based on data mining technology, results show that the food can be divided into six classification, the data mining technology is an effective means for analyzing the food retail price classification.

Keywords: Classification, data mining technology, food retail price

INTRODUCTION

With increasing of economics, the concerning degree on economical development by people increases constantly, from 2009 to now, the consumer consumption price index has increased constantly. The price of goods relating with living of people is increasing, the living level of people is shrinking constantly with increasing of goods price. The food is main part of goods for people living, therefore the food retail price is used as researching object, the relating information of consumption price index is analyzed. Because the inner characteristic of food manifests in the price fluctuation through supply and demand of market, therefore the common inner characters of food are found out starting from price fluctuation and the proper classification is carried out for food, in order to improve the effectiveness of food retail price classification, an advanced technology should be established (Tres *et al.*, 2012).

Data mining technology can extract potential and valued information and knowledge from incomplete, noisy, fuzzy and random application data, which solve complex problem through analyzing the data in depth. During the procession of data mining, data can be stored in the computer in electronic form, the data searching can be achieved automatically, then the data can be processed in further through association rules, categorical Regression and cluster analysis, the structural style of data can be described in depth, the valued information hidden in the data can be mined finally, therefore the data mining technology is a good method for food retail price classification (Li *et al.*, 2012). This research aims at carry out food retail price classification based on data mining technology.

SITUATION OF FOOD RETAIL PRICE

Consumer price index is a goods price changing index, which can reflect the price statistical results of production and labor service relating with people living, it is the important index of observing the mounting inflation. Food price may increase in further and CPI also increases accordingly, which show that inflationary pressures is increasing. The food retail price of city dweller is an important part of goods price index for consumer, according to relating research, the increasing of grain production and currency cost will lead to the increasing of agricultural goods price, especially the exceptional weather happens, the production cost will increases greatly, the international grain price will affect the domestic supply and demand, the food price will increase in further (Seck *et al.*, 2013). In recent years, the CPI has bigger rising amplitude, then the mounting inflation will increase, the city resident food retail price is an import part of consumer goods price index, researches on the changing trends of food retail price is an effective tool for studying the changes of CPI, the classification of food retail index is better method for studying the different kind of food price and finally the changing trends of good can be obtained.

Food of city resident can be classified into five types, the first class food concludes fresh pork, fresh beef and fresh lamb. The second class food concludes peanut oil, chicken and fish. The third class food concludes beans, celery, white sugar, brown sugar. The forth class food concludes blend of rapeseed oil,

Corresponding Author: Han Tian-Peng, School of Computer and Information Engineering, Fuyang Teachers College, Fuyang Anhui 236037, China

soybean oil, soybean oil, eggs, grass carp, carp, rape, cucumber, eggplant, tomatoes, green peppers, leeks, apples, watermelon, soy sauce, vinegar, fresh milk. The fifth class food concludes cabbage, radish, potatoes, carrots, cabbage, bananas, tofu and edible salt (Suchomel *et al*., 2012).

Food retail price changes can affect the economic stability of Chinese market economics and also affect stability of Chinese state of the union, in order to control the increasing of food retail price effectively, the corresponding subsidy measures should be taken for agriculture, for example, the subsidies of famer is added, the new technology and new agricultural production should be popularized for famer, then the enthusiasm of the farm for the famer can be improved quickly, the supply of grain production can be ensured. The price of chemical fertilizer and agricultural chemicals should be controlled strictly, then production cost of agriculture can be improved. The famer should grasp the whole market information and the distribution chain can shortened, then the distributing cost can reduce, the monitoring of market price and cost investigation should be carried out effectively, the monitoring analysis of supply and demand, price changes should be strengthened. The grain, edible vegetable oil, meat, eggs, vegetables, milk and theirs food products should be monitored, the problem of orientation, signs should be early warned, the supply of grain can be ensured and the grain market order can be kept and the food retail price can be smooth for a long time (Kato *et al*., 2012).

Basic theory of data mining: The food price data is used the classification of source and the food retail price changing history is recorded and the relationship between the characteristics and data of food retail price can be found out, then the food retail price can be classified finally. The food retail price classification procedure based on data mining is listed as follows: data preparation and classification object confirmation, data clean, data standardization, data discrete and minimalism of data, the final classification of food retail price can be obtained.

Data clean: This step is to delete the abnormal data of data collected and the vacancy value is added, then invalid data coming from the error of sensor can be deleted, or the data at any moment with collection during the procession of data transmission can made up.

Data standardization: The step is to compare the data collected by sensor based on united standard, when the food retail price changes unusually, the abnormity of data collected can not be confirmed, the maximum, minimum and average values of data collected in real-time has no obvious effect on the classification of food retail price, in order to improve the classification effect of food retail price, the following data standardization method can used (Lone and Khan, 2014):

$$Z_{i,k} = \frac{x_{i,k} - f_k}{S_k}, \quad i = 1,2,\cdots,n \tag{1}$$

$$f_k = \frac{1}{2}(x_{\max,k} + x_{\min,k}) \tag{2}$$

$$S_k = \frac{1}{2}(|x_{\max,k} - f_k) + |x_{\min,k} - f_k|) \tag{3}$$

where, $Z_{i,k}$ denotes the abnormal factor of data collected, when $-1 < Z_{i,k} < 1$, data collected is in the normal condition, f_k denotes average valued of property k, S_k denotes average absolute deviation of property k, $x_{\max,k}$ and $x_{\min,k}$ denotes the maximum and minimum value of property k.

Data discretization: This step can divide the condition proper interval into several limited intervals, the element of different interval can be divided in to several limited intervals and decision value relating to different interval element is same. The procession of data discretization should consider the characteristic of data collected, when the changes of data collected is not big, while has big effect on the classification of food retail price, the small discrete interval can chosen, then the classification of food retail price can be obtained correctly, otherwise, the big discrete interval can be chosen.

Data mining method of fuzzy rough set: The fuzzy rough set is the extension of rough set theory, discourse domain U is known, the fuzzy collection is defined by $F(U)$, for a weak fuzzy division, the following definition is listed as follows (Nishimura *et al*., 2015):

$A_i \in F((U)$, $1 \le i \le k$, then $\varepsilon = \{A_1, A_2, \cdots, A_k\}$ denotes a weak fuzzy division and the following condition can be satisfied:

- $\bigcup_{1 \le i \le k} \operatorname{supp} A_i = U$
- A_i is the regular fuzzy sets on U

Two fuzzy collections $A, B \in F((U)$ are known, the relationship of the two fuzzy collections can be denoted by inclusion degree and the expression of inclusion degree is listed as follows:

$$D(A/B) = \frac{|n(A,B)|}{|\operatorname{supp} B|} B \ne \Phi \tag{4}$$

where $n(A,B) = \{x \in \operatorname{supp} B \mid B(x) \le A(x)\}$ 。

Upper approximate set and bottom approximate set of fuzzy collection X to ε are defined as follows:

$$\overline{A}_\sigma(X) = \bigcup\{A_i \mid D(X/A_i) \ge \sigma\} \tag{5}$$

$$\underline{A}_\tau(X) = \bigcup\{A_i \mid D(X/A_i) \ge \tau\} \tag{6}$$

where,

$\overline{A}_\sigma(X)$ = The upper approximate set

$\underline{A}_\tau(X)$ = Bottom approximate set

Condition property and decision property of decision system belong to fuzzy collection and the inclusion degree can reflect the ratio belonging to or not the fuzzy collection according to the different property value in the property collection, which can offer the benefit basis for rule reduction.

The knowledge reasoning is an important part of decision system, the decision table can express the knowledge effectively. The object and decision rule of decision table maps one-for-one, the decision can defined as follows (Wang *et al*., 2014):

The information system $S = (U, A)$ is known, C is the condition property, D is the decision property, which belong to two sub collections of A, $C \bigcup D = A$, $C \bigcap D = \Phi$, then S is the decision table.

During the procession of decision, based on the reduction of decision table the condition property is fewer, the consistent decision table has the following reduction procedure:

Firstly, on column is deleted in decision table.
Secondly, the redundant lines are deleted.
Finally, the redundant number of properties is deleted in the decision rule.

ALGORITHM DESIGN OF DATA MINING

For a lot of data formed in the food retail price database, the correctness and effectiveness of conventional relational measure algorithm can not be ensured, the main reason is listed as follows: a lot of data and infinite data is coped with by a processer, then the processing efficiency decreases, which is not disposable processed, then the effective data mining algorithm of the massive data is established through cloud computing resource interactive conception, a quick and effective relational measure algorithm is constructed based on the feature of the result convergent. The connection expressed is listed by:

$$c(u,v) = (u^{-\alpha} + v^{-\alpha} - 1)^{1/\alpha} \tag{7}$$

The expression is simple, symmetrical and combinable. According to the objective facts of sample capacity and comprehensive information content, the weight information is added into formulation (7) and the scale of relational coefficient is considered comprehensively, therefore the design of the connection function is expressed as follows:

$$c_k(u,v) = 2\left(\frac{N^*}{N}\left(\frac{u+1}{2}\right)^\alpha + \frac{n_k}{N}\left(\frac{v+1}{2}\right)^\alpha\right)^{1/\alpha} - 1 \tag{8}$$

where, n_i is ith data sample, $N^* = n_1 + n_2 + \cdots + n_{k-1}$; $N = N^* + n_k$, $-1 \le u, v \le 1$, $0 < \alpha < 1$, $-1 \le c_k \le 1$; α is the set parameter.

Based on the basic thought of data diving into block, a small sample collection s_i the massive data is transferred into end to be coped with, the effective computing means of relational coefficient is confirmed based on the especial feature of data and the relational effect level r_i of x to y is evaluated by the following formulation:

$$\widehat{r}_i = \frac{\sum_{j=1}^{n_i}(x_j - \overline{x})(y_j - \overline{y})}{\sqrt{\sum_{j=1}^{n_i}(x_j - \overline{x})^2}\sqrt{\sum_{j=1}^{n_i}(y_j - \overline{y})^2}} \tag{9}$$

where, x_j and y_j denote observed values in s_i, $\bar{x}$ and $\bar{y}$ denote observed values, n_i is the sample capacity of s_i.

Other independent sample collection s_{i+1} happens in the massive data and the relational effect level $\widehat{r}_{i+1}$ is computed by Kyunglag *et al*. (2014):

$$\widehat{r}_{i+1} = \frac{\sum_{j=1}^{n_i+1}(x_j - \overline{x})(y_j - \overline{y})}{\sqrt{\sum_{j=1}^{n_i+1}(x_j - \overline{x})^2}\sqrt{\sum_{j=1}^{n_i+1}(y_j - \overline{y})^2}} \tag{10}$$

The following sampling collection is confirmed after the two small sample collections are merged:

$$s_d = s_i + s_{i+1} \tag{11}$$

The sample capacity is defined by:

$$n_d = n_i + n_{i+1} \tag{12}$$

The relational measure evaluated of the two small samples can be connected by the establishing connection function and the relational degree of the compound sample merged is defined by:

$$\widehat{\rho}_d = 2\left(\frac{n_i}{n_d}\left(\frac{\overline{r}_i + 1}{2}\right)^\alpha + \frac{n_{i+1}}{n_d}\left(\frac{\overline{r}_{i+1} + 1}{2}\right)^\alpha\right)^{1/\alpha} - 1 \tag{13}$$

After $\widehat{\rho}_d$ is assessed, the end assess the relation measure $\overline{r}_{i+2}$ of x and y in s_{i+2} with n_{i+1} sample capacity. s_d and s_{i+2} are merged and the sample collection s_{d+1} is confirmed, the sample capacity of it can be defined by:

$$n_{d+1} = n_i + n_{i+2} \tag{14}$$

Table 1: Amplitude classification of increasing ratio of food retail price

One order difference of food retail price/%	Number	Average value	Minimum value	Maximum value	Variance
(-76.03, -11.53)	2	-70.54	-70.54	-70.54	NA
(-11.53, -6.42)	3	-5.8	-5.9	-5.3	NA
(-6.42, 6.42)	49	0.342	-4.32	4.83	1.93
(6.42, 11.43)	2	5.94	5.94	5.94	NA
(11.43, 75.32)	2	53.4	53.4	53.4	NA
Grand total	58	-0.054	-65.43	58.4	10.43

Table 2: Classification results of food retail price based on data mining

Classification	Food name
Oil	Rapeseed oil, peanut oil, soybean oil, fresh pork (lean meat), fresh pork (rib meat)
High protein meat	Fresh beef, fresh mutton (boneless), fresh mutton (with bone), chicken, chicken, eggs, octopus, grass carp, carp
Single season vegetables,	Cucumber, cabbage, tomato, pepper, green pepper
Multi season vegetables	Cucumber, cabbage, beans, garlic, leek, banana, watermelon, apple
Non seasonal vegetables	Potatoes, carrots, celery, cabbage, rape
Accessories category	Salt, soy sauce, vinegar, cotton sugar, white sugar, brown sugar, fresh milk, tofu

The relational measure $\widehat{\rho}_{d+1}$ of variables *x* and *y* in s_{d+1} is assessed according to connection function, which is defined by:

$$\widehat{\rho}_{d+1} = (\frac{n_d}{n_{d+1}}(\frac{\widehat{\rho}_d + 1}{2})^{\alpha} + \frac{n_{i+2}}{n_{d+1}}(\frac{\overline{r}_{i+2} + 1}{2})^{\alpha})^{1/\alpha} \quad (15)$$

The end condition of algorithm is defined by:

$$|\widehat{\rho}_{d+k} - \widehat{\rho}_{d+k-1}| \leq \varepsilon \quad (16)$$

where,

$\varepsilon<0$

CHINESE FOOD RETAIL PRICEFLUCTUATING AND CLASSIFICATION RULES

In recent years, Chinese food retail fluctuation shows a certain rules, however in some years the food retail price will generate mutation, this phenomenon shows that the food retail price is affected by market strength, but the random factor can generate the concussion (Yang *et al*., 2012). Based on the quarterly data of food retail price, the increasing ratio of food retail price is shown in Table 1.

Based on the basic theory of data mining, the corresponding programmer is compiled by MATLAB software and the final classification results of food retail price are shown in Table 2.

From Table 2, the food can be divided into six kinds, which are oil, high protein meat, single season vegetables, multi season vegetables, non seasonal vegetables and accessories category. The classification of food retail price has strong economic signification meaning and the classification results are agreed with real situation and the inner and immeasurable property of food can be reflected effectively, such as the supply and demand changing situation, requiring price elastic, which can offer strong guidance signification.

CONCLUSION

Food retail price is an important economical index concerned by resident, which can fluctuate under all kinds of factors. The food retail price has something with living of people, the fluctuation of it will affect the living of people greatly, therefore the fluctuation should be controlled effectively and then the people can consume in stable economical environment. The data mining technology is applied in classification of food retail price and the numerical simulation results show that it is an effective means for analyzing the food retail price and the affecting factors of food retail price can be considered during the procession of analysis.

REFERENCES

Kato, T., D.T. Pham, H. Hoang, Y. Xue and Q.V. Tran, 2012. Food residue recycling by swine breeders in a developing economy: A case study in Da Nang, Viet Nam. Waste Manage., 32(12): 2431-2438.

Kyunglag, K., K. Daehyun, Y. Yeochang, S. Jong-Soo and C. In-Jeong, 2014. A real time process management system using RFID data mining. Comput. Ind., 65(4): 721-732.

Li, T., Z.G. Zhang and G.L. Liu, 2012. Sensitivity analysis model of food prices on SVR. J. Converg. Inform. Technol., 7(21): 205-211.

Lone, T.A. and R.A. Khan, 2014. Data mining: Competitive tool to digital library. DESIDOC J. Libr. Inform. Technol., 34(5): 401-406.

Nishimura, K., Y. Maehata and W. Sunayama, 2015. Improved inspection of facilities for high-voltage class using data mining. Electr. Eng. Jpn., 191(2): 47-54.

Seck, G.S., G. Guerassimoff and N. Maïzi, 2013. Heat recovery with heat pumps in non-energy intensive industry: A detailed bottom-up model analysis in the French food & drink industry. Appl. Energ., 111: 489-504.

Suchomel, J., M. Gejdoš, L. Ambrušová and R. Šulek, 2012. Analysis of price changes of selected roundwood assortments in some Central Europe countries. J. Forest Sci., 58(11): 483-491.

Tres, A., G. van der Veer, M.D. Perez-Marin, S.M. Van Ruth and A. Garrido-Varo, 2012. Authentication of organic feed by near-infrared spectroscopy combined with chemometrics: A feasibility study. J. Agr. Food Chem., 60(33): 8129-8133.

Wang, X.F., Y. Wang, H.B. Bi and R.N. Gao, 2014. Heat-supply network state prediction based on optimum combination model of data mining. J. Appl. Sci., 13(13): 2443-2449.

Yang, X.Q., C.F. Zou, L. Yue and R. Gao, 2012. Research on food complains document classification based-on topic. J. Softw., 7(8): 1687-1693.