Mining Time Series Data with Two Dimensional Fuzzy Pattern Rules

Haifeng Xia
College of Science
Huazhong Agricultural University
Wuhan, China

Bing Chen
College of Science
Huazhong Agricultural University
Wuhan, China

Jiawei Fan
College of Science
Huazhong Agricultural University
Wuhan, China

Zhi Li[*]
College of Science
Huazhong Agricultural University
Wuhan, China
*Corresponding author

Dan Gao
College of Public Administratio
Huazhong Agricultural University
Wuhan, China

Abstract—**Based on association rules and fuzzy pattern theory, this paper presents a data mining model to extract patterns using two-dimensional fuzzy pattern rules, focusing on the discretization of the time series of dependent and independent variables. The referential vector distance is established as a similarity function, and K-means clustering method is used to acquire the basic trend features. The silhouette coefficient is used to test and analyze the clustering result. Due to the potential delay of the influence of the independent variable on the dependent variable, we determine the optimal lag phase with the sort rule of J-measure based on the principle of maximum entropy. The fuzzy pattern rules of the influence of the independent variable on the dependent variable are determined and interpreted according to the trend features and the fuzzy pattern definitions. The obtained character sets are adjusted and iterated several times based on the definitions of the stage characteristics to locate and interpret the inherent periodic characteristics of the couple of time series. The reliability of the proposed model is tested with an empirical analysis of the effect of the price of corn on that of the pork.More importantly, the analysis unit of this model can be as accurate as the time of a day, suggesting its potential advantages in analyzing the short-term characteristics of time series.**

Keywords-Fuzzy Pattern Rule; Referential Vector Distance; K-means Cluster; J-measure;Stage characteristics

I. INTRODUCTION

Time series is a particular temporal data object widely existing in the real life. With the extension of time, this data object will experience an explosive growth, suggesting the necessity to establish a reasonable model for knowledge discovery and pattern mining. The conventional mining method is focused on building a strict mathematical model and testing the model by hypothesis verification and parameter estimation[1,2,3], which needs to be performed under strict assumptions, resulting in a poor effect in solving practical problems. The data mining method, however, is concentrated on the construction of a model with big data[4], which can produce a result similar to the actual situation due to the effective Reduction of the noise and redundant data.

Time series data mining is usually performed as follows: The target sequence is discretized first, followed by analyzing the discretized sequence with measuring similarity and clustering[5,6]. Finally, the pattern is mined and extracted based on the clustering results to predict the developing trend in the future.

It is difficult to describe the similarity and the pattern in time series data mining. The similarity of two time series is usually described directly by Euclidean Distance, namely if $\| Y - X \| < \varepsilon$, then two sequences of this group is considered similar, and vice versa[7]. However, ALCOCK[8] pointed out that, when measuring two similar curves, this method may cause serious mistakes in clustering results due to the large distance of the two curves resulting from different starting points in calculation, and the calculation can also be expensive. Rafiei[9] proposed the adoption of the Discrete Fourier Transformation to compress the time series before measuring the similarity of signal with the Euclidean Distance. This method can effectively compress the information redundancy and enhance the calculation speed, but it sometimes may fail to extract the most important features of time series. Jun[10] and Z. Yu[11]proposed the symbolization of time series by endowing them the status of down, smooth or up with -1, 0 and 1 before measuring the similarity. However, this method fails to reveal the speed and time of the transformation of local trends. Based on the theory of Meta-Model,and Nanopoulos[12]

proposed a feature-based method for the similarity, but this description is not suitable for the time series with unequal length. Currently, Dynamic Time Warping (DTW) is frequently used to handle time series data with unequal length[13,14]. But this method is not suitable for multidimensional time series because of the huge consumption of time and space. For pattern induction and extraction, Gautam Dans[15]presented a pattern to discretize time series with sliding window[16]first, followed by clustering the subsequence space with similarity function defined on the set of the subsequences, then symbolizing the sequence and finally obtaining the strong temporal association rules from the symbol sequence. Additionaly,[17]proposed to solve the problem of association rules with rough set techniques, which could produce a good effect for practical questions but not for the extraction of multivariate time series data.

In light of the aforementioned research results and theories, this paper introduces the concept of referential vector distance into similarity measuring based on the idea of meta-model and the principal component[18], and sets the similarity function as the basis of clustering. In the process of extraction and induction of sequential patterns, this paper proposes the concept of fuzzy pattern to facilitate sequence analysis and the concept of classification trees in the prediction of time series to improve the accuracy of the prediction result.

When mining the association between two time series, we consider the pattern features of time series separately, calculate their local similarity indirectly and extract the rules by using the Bayes principle to enhance the credibility of the extracted rules.

II. DISCRETIZATION OF TIME SERIES

In the process of time series discretization,Lin[19], Keogh[20], and Liu[21] discretized the sequence to vector by clustering discretization. Based on clustering discretization and sliding window algorithm, we slide the time series to discretize the time sequence to a series of sub-sequences with the length of w, which is described as follows.

First, we suppose a time series S and a given slide window, the length of which is w. For S, every window is composed of a set of vectors $(a_i,\ldots,a_n)$, and the sub-sequences of these windows are designated as $s_i,\ldots,s_{i+w-1}$, and expressed as

$$W(s) = \{s_i \mid i = 1,\ldots,n-w+1\} \tag{1}$$

This method[22] depends to some extent on the selected length and footstep of window, but to a large extent on the scope of local characteristics considered by the researchers. The larger is the window length, the closer will be the sequence characteristics to the macro characteristics. Otherwise, the sequence characteristics will be closer to the micro characteristics.

III. SIMILARITY OF TIME SERIES AND CLUSTERING

Definition 1: Fluctuating Value. This value can be obtained by

$$\sigma^2 = \sum_{i=1}^{k}(x_i - u)^2 \tag{2}$$

and it is used to describe the fluctuation of time series around the average. If a vector of time series is shown as $X = (x_1, x_2,\ldots,x_k)$, then

$$u = \frac{x_1 + x_2 + \cdots + x_k}{k} \tag{3}$$

where u indicates the average value of X and k is the number of X.

Definition 2: Trend Value. This value can be obtained by

$$tr = \frac{x_k - x_1}{T} \tag{4}$$

and it is used to describe the change of time series from start to end as well as the trend of the research object. Here a vector of time series is shown as $X = (x_1, x_2,\ldots,x_k)$, x_k indicates the end value of the time series, x_1 indicates the beginning value of the time series, and T the time span from x_1 to x_k.

Definition 3: Range. This value can be obtained by

$$d = \max\{x_i\} - \min\{x_i\} \tag{5}$$

and it is used to directly describe the special value of time series.

Definition 4: Referential Vector.

This value is expressed as $f : X \to C$, where $X = \{x_1,\ldots,x_k\}$ and $C = \{\sigma^2, tr, d\}$. The vectorization of time series indicates that X maps to reference vector C through the rule f, which can prove that the mapping is injective.

Definition 5: Referential Vector Distance. This value can be obtained by

$$d(C_1, C_2) = w_1 \mid \sigma_1^2 - \sigma_2^2 \mid + w_2 \mid tr_1 - tr_2 \mid + w_3 \mid d_1 - d_2 \mid \tag{6}$$

where w_1, w_2, w_3 indicate the weight of fluctuating value, trend value and range, respectively.

Nature 1: $d(C_1, C_2) = d(C_2, C_1)$;

Nature 2: $d(C_1, C_2) + d(C_2, C_3) \le d(C_1, C_3)$;

Nature 3: $d(C_1, C_2) \ge 0$, only if $C_1 = C_2$ will Equation be set up.

When constructing the similarity function, we take the fluctuating value, trend value and range into consideration based on the idea of the meta-model, which can effectively reduce the information redundancy and reserve useful information.

In the process of clustering, we use referential vector distance as the basis of clustering, i.e., if $|d(C_1,C_2)| < \varepsilon$, then the time series C_1, C_2 will be clustered into the same class, and if not, they will be clustered into two different classes (Figure 1). Here, the K-means clustering method[23,24] is adopted to reduce time complexity.
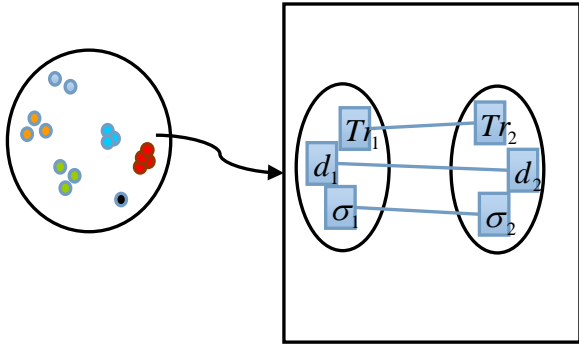

FIGURE I. THE PROCESS OF CLUSTERING

IV. EFFECTIVENESS OF CLUSTERING

Definition 1: The silhouette coefficient of a single time series sample. This value can be obtained by

$$s_i = \frac{\min\{b_{ij}\} - a_i}{\max\{\min\{b_{ij}\}, a_i\}} \quad (7)$$

This equation indicates that the sample s_i is clustered into class A, where a_i means the average vector distance between s_i and other samples in the same class, b_{ij} means the vector distance between sample s_i and the centers of other classes.

Definition 2: The silhouette coefficient of a single time series class. This value can be obtained by

$$s_k = \frac{1}{|s_k|} \sum s_i \quad (8)$$

where $|s_k|$ means the sample distance in class k.

The silhouette coefficient is one of the important indexes on estimating the clustering results[25,26]. In general, the silhouette coefficient value of the sample will fluctuate in the range of [-1, 1]. The value of close to 1 indicates that the sample is correctly classified, the value of close to 0 suggests the sample should be set between this class and another class, and the value of close to -1 means that the sample should belong to another class. The silhouette coefficient values in class and in sample have the same properties. However, in the real experimentation, we define a threshold value of 0.5 as the silhouette coefficient. If most silhouette coefficient values of a class are higher than 0.5, the sample is believed to be well clustered, and vice versa.

V. PATTERN EXTRACTION AND KNOWLEDGE MINING

A. Determination of Optimal Lag Phase

While extracting the time series, the sequences of dependent variables are first defined as A. After clustering, the sub-sequences of A are divided into m classes, denoted as $A \propto \{A_1, A_2, \ldots, A_m\}$, and the sub-sequences of each class are denoted as $\{a_1, a_2, \ldots a_m\}$ and $n = \sum_{i=1}^{m} |A_i|$. Each subsequence a_i is a set with four members, shown as $a_i = \{tr_i, \sigma_i^2, d_i, t_i\}$, where t_i means the time node for a_i. By following the same procedure, we can obtain the classified results of independent variables shown by $B \propto \{B_1, B_2, \ldots, B_m\}$, and the sub-sequences of each class $\{b_1, b_2, \ldots, b_m\}$, denoted as $b_i = \{tr_i, \sigma_i^2, d_i, t_i\}$. For the convenience of description, each clustering result is shown as a character.

Because of the ordering of time series data, the lag phase of the effect of independent variables on dependent variables during knowledge mining should be confirmed before locking the inherent value of each character. Each B_i will influence the generation of A_j in the space of T. Here these effects are converted into a form of probability as shown by the following equation:

$$\text{influence}(B_i \xrightarrow{T} A_j) = P(B_i \xrightarrow{T} A_j) = P(A_j | B_i), \, while \, T = t_j - t_i \quad (9)$$

Based on the definition and transformation above, we can obtain the probability distributions of the influence of A on B, which is expressed as $P(Y|X)$, where $X \in B, Y \in A$ and $P(Y|X)$ means the probability between A and B under the lag phase T.

Now we can calculate the information content by using the obtained probability distributions between A and B, and choose the proper lag with the maximum information content. When calculating the information under every lag phase, we adopt the sort rule of J-measure [27], which is defined as follows.

$$J(B_T; A) = p(A)(p(B_T|A)\log(\frac{p(B_T|A)}{p(B_T)}) +$$
$$(1 - p(B_T|A))\log(\frac{p(B_T|A)}{p(B_T)}) \quad (10)$$

where $J(B_T; A)$ means the information content of B that appears after A in the lag phase of T, $P(A)$ means the probability of the appearance of A, $P(B_T|A)$ means the probability of the appearance of B after A in the lag phase of T.

B. Knowledge Mining

1) Trend Characteristics

Definition 1: Character variation tendency. For a character, the variation of its beginning and end values can objectively reflect its variation tendency, which is denoted as

$$\theta = \frac{A_t - A_0}{t}.$$
(11)

where A_t means the end values and A_0 means the beginning values of this character, t means the length of this character.

Definition 2: Membership function of θ_i. The membership function of θ_i is expressed as

$$\gamma_i = \frac{\theta_i - \theta_0}{|\theta|}$$
(12)

where θ_i means the trend values of the $i-th$ class, θ_0 means the beginning value of a fuzzy set, $|\theta|$ means the length of this fuzzy set.

According to the above two definitions, the trend characteristics of the couple of time series can roughly be divided into 9 different fuzzy patterns as shown in Figure 2.
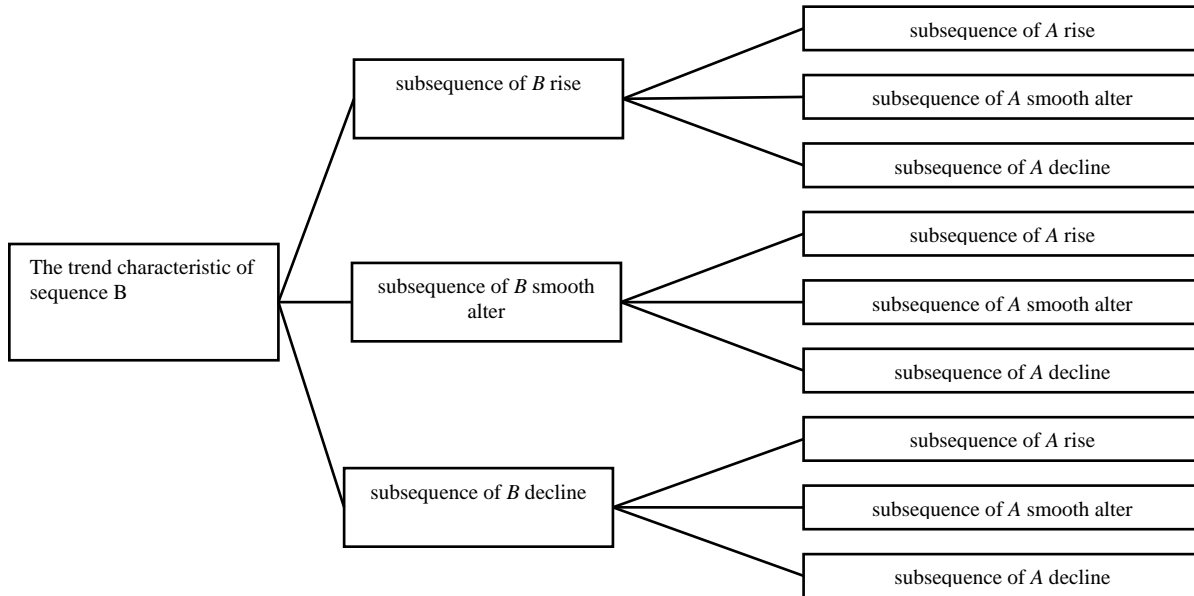
$$r_{ij} = \frac{(\theta_i - \theta_0)(\theta_j - \theta_0^{'})}{|\theta||\theta^{'}|}$$
(13)

where $\theta_i \in A$ and $\theta_j \in B$, θ_0 and $\theta_0^{'}$ means the beginning value of a certain fuzzy set respectively, $|\theta|$ and $|\theta^{'}|$ means the length of a certain fuzzy set [28] respectively.

Definition 4: The probability of the appearance of each fuzzy pattern.

If the fuzzy pattern of every sequence group is divided into five classes (Table 1), we need to discuss 25 solutions for describing the trend characteristics of two time series. According to definition 3, C_{ij} can be obtained and C_{ij} is used to indicate the probability of the appearance of each fuzzy pattern, where i means the $i-th$ trend characteristic of sequence B, and j means the $j-th$ trend characteristic of sequence A.

TABLE I. FIVE DIFFERENT FUZZY PATTERNS

Level	The Section of Dip Angle	Meaning
1	[-90, -40]	Sharp Decline
2	[-40, -10]	Slow Decline
3	[-10, 10]	Smooth Fluctuation
4	[10, 40]	Slow Rise
5	[40, 90]	Sharp Rise

FIGURE II. THE TREND CHARACTERISTIC

Definition 3：Membership function of θ_i and θ_j. The membership function of θ_i and θ_j is expressed as

2) Stage Characteristics

It is known to all that, in the long time series, a feature formed by the two sequences is likely to concentrate on several

specific moments, and may appear again after a phase to form a regular fluctuation. Through learning this knowledge, we can better understand the inherent volatility law, which is called phase characteristics.

When certain windows of the dependent variable and the independent variable are noted as A_j and B_i, $P(B_i, t, A_j)$ represents the probability that A_j appears after B_i in the time of t. By counting the frequency of the appearance of B_i and A_j in the data set N, we can obtain $P(B_i, t, A_j)$, which makes it possible to obtain the frequency of the appearance of y in the period of T for every word x in N by $P(B_i = x, t, A_j = y)$. When $|N| \to \infty$, $P(B_i = x, t, A_j = y)$ is stable. Given a larger data set, we can obtain an approximate value to reveal the probability of the appearance of A_j.

In practice, we can assume that stage characteristics emerge with the changes of seasons and set a threshold value ε. If $P(A_j \mid B_i) \geq \varepsilon$ in one season, the knowledge can be considered as a significant character, defined as the inherent characteristics of the season, and recorded as η_1. Considering the retardance in practice, we should study every interval which is determined on the basis of season after an extended period of T. If $P(A_j \mid B_i) \geq \varepsilon$ appears in one interval, the law can be considered as a significant character, which is defined as the inherent characteristics of the season and recorded as η_2. By repeated adjustment and multiple iterations, we can find the interval with the strongest knowledge learning, which is defined as the phase interval, and the characteristics in that interval are determined as phase characteristics.

VI. EMPIRICAL ANALYSIS

A. Data Sources and Pretreatment

For empirical analysis, we have chosen a set of important indicators such as the prices of pork and corn in time series of agricultural production from the published data of Chinese animal husbandry website. The time interval is from Jan. 8, 2010 to March 31, 2014.

All the data are pretreated to meet the requirements of experiment. The missing data are filled by the average value or piecewise fitting. Additionally, the original data are interpolated for short-term pattern mining because they are recorded in the unit of week.
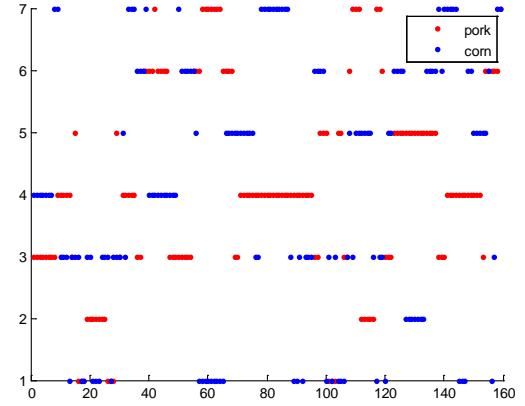

FIGURE III. CLUSTERING RENDERING OF PORK AND CORN PRICE TIME SERIES

B. Discretization of Time Series and Cluster Subsequences of Pork and Corn Prices

Figure 3 shows the clustering result of the couple of time series of pork and corn prices through discretization and clustering.

Based on the aforementioned clustering validity analysis method, we tested the clustering effectiveness of the two time series and obtained the silhouette coefficient values shown in Figure 4. From the values, it can be concluded that the clustering effects based on reference distance are more significant than those based on Euclidean distance.
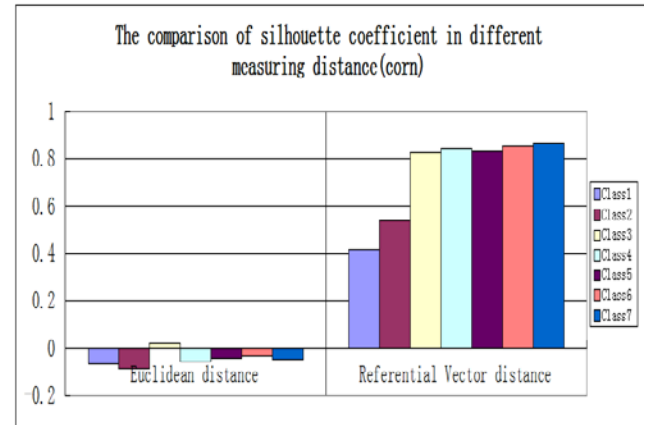

FIGURE IV. COMPARISON OF SILHOUETTE COEFFICIENTS

C. Selection of Optimal Lag Phase

TABLE II. PROBABILITY DISTRIBUTIONS OF FLUCTUATION TRENDS

Corn \ Pork	First Class	Second Class	Third Class	Fourth Class	Fifth Class	Sixth Class	Seventh Class
First Class	0. 19747899	0. 06302521	0. 273109244	0. 109243697	0. 088235294	0. 029411765	0. 239495798
Second Class	0. 02941176	0. 088235294	0. 058823529	0. 058823529	0. 117647059	0	0. 647058824
Third Class	0. 18072289	0. 012048193	0. 196787149	0. 13253012	0. 084337349	0. 044176707	0. 34939759
Fourth Class	0. 08910891	0. 069306931	0. 207920792	0. 089108911	0. 108910891	0	0. 435643564
Fifth Class	0. 23529411	0. 029411765	0. 147058824	0. 147058824	0. 161764706	0. 058823529	0. 220588235
Sixth Class	0. 10676156	0. 014234875	0. 213523132	0. 117437722	0. 028469751	0. 078291815	0. 441281139
Seventh Class	0. 13580246	0. 012345679	0. 172839506	0. 086419753	0. 24691358	0	0. 345679012

In practice, we selected 42 days as the optimal lag phase due to its coverage of the largest amount of information. Under this optimum lag phase, we obtained the probability distributions about the fluctuation trends of pork prices relative to different corn prices (Table 2). Meanwhile, the chart of corn price fluctuation trends is shown in Figure 5. As shown in Figure 5, the first class shows a slow-decline trend, the second class a rapid rise, the third and the fifth classes a slow rise, the fourth and the sixth classes a smooth fluctuation, and the seventh class a sharp decline.

From Table 4 and Figure 5, it can be seen that, under the lag phase of 42 days, when the corn price presents a slow-decline trend, the pork price shows a slow downward trend 42 days later with a possibility of 27%. When the corn price shows a sharp rise, the pork price will experience a smooth fluctuation with a possibility of 64%. When the corn price shows a slow rising trend, the pork price will fluctuate stably under the probability of 34% and rise slowly under the probability of 24%. When the corn price is in a steady state, the pork price

FIGURE VI. THE INFLUENCE LAW BETWEEN CORN AND PORK PRICES

D. Extraction of Fuzzy Pattern and Mining on Pork and Corn Prices

We extracted and mined the fuzzy patterns of the pork and corn prices using the methods described above.

With the concepts defined above, we can obtain the short-term(42 days) and long-term(140 days) state transition probability matrixes of pork prices (Table 3 and Table 4).

According to the classification criteria, the first and the fifth classes belong to the slow-rise pattern, the second class the sharp-rise pattern, the third and the fourth classes the slow-decline pattern, and the seventh class the smooth and steady pattern.

In the short term matrix, the pork price has a great possibility to maintain the original state. With a slow rise of the pork price, there is a 50% possibility for the price to enter a state of sharp decline. With a sharp rise of the pork price, there is a 34.375% possibility of a slow rise state for the pork price.
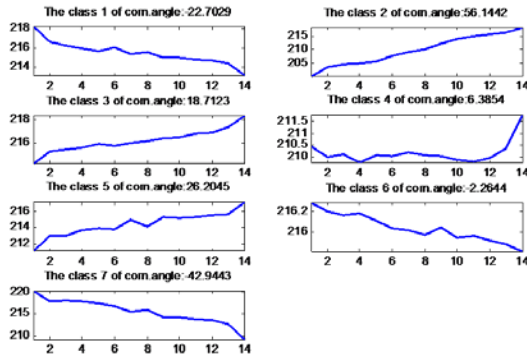

FIGURE V. FLUCTUATION TRENDS OF CORN PRICES

will fluctuate stably under the probability of 44%. When the corn price sharply declines, the pork price will have a stationary fluctuation under the probability of 34%. All of the above trends are summarized in Figure 6.
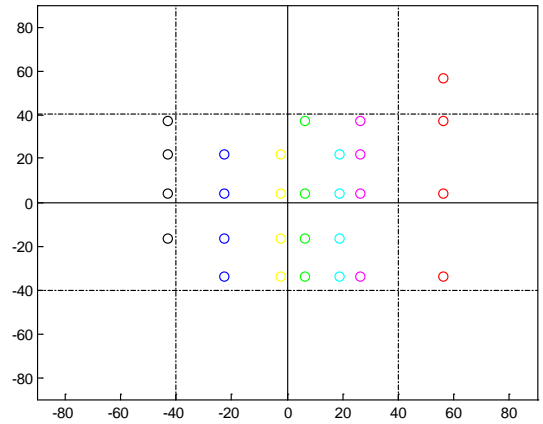
TABLE III. SHORT-TERM STATE TRANSITION PROBABILITY MATRIX OF PORK PRICES

Character of pork price	First Class	Second Class	Third Class	Fourth Class	Fifth Class	Sixth Class	Seventh Class
First Class	0. 47368	0. 00000	0. 00000	0. 02632	0. 00000	0. 50000	0. 00000
Second Class	0. 00625	0. 30000	0. 11875	0. 01875	0. 34375	0. 21250	0. 00000
Third Class	0. 00000	0. 10638	0. 39574	0. 13617	0. 32766	0. 02553	0. 00851
Fourth Class	0. 00820	0. 03279	0. 27049	0. 49180	0. 00820	0. 07377	0. 11475
Fifth Class	0. 00000	0. 10891	0. 20050	0. 02475	0. 64109	0. 01980	0. 00495
Sixth Class	0. 12162	0. 26351	0. 04730	0. 02703	0. 05405	0. 48649	0. 00000
Seventh Class	0. 00000	0. 00000	0. 08333	0. 29167	0. 00000	0. 00000	0. 62500

TABLE IV. LONG-TERM STATE TRANSITION PROBABILITY MATRIX OF PORK PRICES

Character of pork price	First Class	Second Class	Third Class	Fourth Class	Fifth Class	Sixth Class	Seventh Class
First Class	0. 00000	0. 00000	0. 02632	0. 00000	0. 71053	0. 26316	0. 00000
Second Class	0. 01307	0. 13725	0. 22876	0. 15686	0. 22876	0. 05229	0. 18301
Third Class	0. 00000	0. 21154	0. 23558	0. 12500	0. 37500	0. 03846	0. 01442
Fourth Class	0. 03604	0. 12613	0. 33333	0. 05405	0. 25225	0. 19820	0. 00000
Fifth Class	0. 06582	0. 10886	0. 18481	0. 14177	0. 35190	0. 12658	0. 02025
Sixth Class	0. 00000	0. 12838	0. 27027	0. 04730	0. 37838	0. 11486	0. 06081
Seventh Class	0. 15385	0. 07692	0. 05128	0. 12821	0. 00000	0. 58974	0. 00000

Conversely, with a sharp drop in the pork price, there is a 26.351% possibility for the price to enter a state of sharp decline. With a slow decline, the possibility to entrer a slow-rise sate is 32.766%. Additionally, with a smooth fluctuation , there is a 37.5% probability for the price to enter a slow-decrease state or maintain a steady state.

But in the long-term matrix, if the pork price rises slowly, the probability for the price to maintain the upward trend is 71.8%. If the price sharply rises, the probability to keep stable is 22.876%. When the price declines slowly, the probability to be transformed into a slow-rise pattern is 37.5%. When the price decreases sharply, the probability to be turned into a slow-rise state is 37.838%. When the pork price remains stable, maintaining the price in a stable state will be very difficult, which is likely to account for a sharp decline under the possibility of 58.794%.

VII. CONCLUSIONS

This paper has presented a new method for extracting patterns in time series. First, the original time series is discretized, followed by clustering the subsequences of the independent variable and the dependent variable using the proper similar function and K-means Algorithm. Secondly, a proper lag phase is determined based on the hysteresis effect between the two variables and the sort rule of J-measure. Finally, we can extract patterns and analyze the patterns with the obscured variables under the proper lag phase. The design ideas can provide reference for other similar studies.This method gives sufficient consideration to the hysteresis effect and the coarse-grained pattern rules between the two variables. A large data set is the precondition of the proposed method. With a large data set, the influence of the noise and redundancy in the data is small, which helps to reduce the spatial expenditure of data pretreatment. Additionally, this method can obtain the probability distributions of samples approximately without making any assumption or estimation and thus the result is more reliable. However, there are still some limitations in this method. For example, the time and space will become expensive with the increase of the dimensions of time series and the calculation speed will become slow. Further studies should focus on how to balance the relationship between the accuracy and efficiency of pattern mining.

REFERENCES

[1] Wagner, A. K., Soumerai, S. B., Zhang, F., & Ross‐Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. Journal of clinical pharmacy and therapeutics,27(4), 299-309.

[2] McQuarrie, A. D., & Tsai, C. L. (1998). Regression and time series model selection (Vol. 43). Singapore: World Scientific.

[3] Hall, D. H. (1992). The science-industry interface: correlation of time series of indicators and their spectra, and growth models in the nuclear fuels industry. Scientometrics, 24(2), 237-280.

[4] Keogh, E., & Kasetty, S. (2003). On the need for time series data mining benchmarks: a survey and empirical demonstration. Data Mining and knowledge discovery, 7(4), 349-371.

[5] Fu, T. C. (2011). A review on time series data mining. Engineering Applications of Artificial Intelligence, 24(1), 164-181.

[6] Esling, P., & Agon, C. (2012). Time-series data mining. ACM Computing Surveys (CSUR), 45(1), 12.

[7] Agrawal, R., Faloutsos, C., & Swami, A. (1993). Efficient similarity search in sequence databases (pp. 69-84). Springer Berlin Heidelberg.

[8] Alcock, R. J., & Manolopoulos, Y. (1999, August). Time-series similarity queries employing a feature-based approach. In 7th Hellenic conference on informatics, Ioannina, Greece (pp. 1-9).

[9] Rafiei, D., & Mendelzon, A. (1998). Efficient retrieval of similar time sequences using DFT. arXiv preprint cs/9809033.

[10] Jun Zhang, Hanwu Chen,Zhiming Ma.(2005).A Fast Algorithm for Time Series Similarity. Journal of Nanjing Normal University.

[11] Yu,Z.,Peng,H.,Zheng,Q. L.(2005).Pattern Distance of Time Series Based on Segmentation by Important Points. Machine Learning and Cybemetics(pp.1563-1567).

[12] Nanopoulos, A., Alcock, R., & Manolopoulos, Y. (2001). Feature-based classification of time-series data. International Journal of Computer Research, 10(3).

[13] Chu, S., Keogh, E. J., Hart, D. M., & Pazzani, M. J. (2002, April). Iterative Deepening Dynamic Time Warping for Time Series. In SDM (pp. 195-212).

[14] Keogh, E. J., & Pazzani, M. J. (2000, August). Scaling up dynamic time warping for datamining applications. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining(pp. 285-289). ACM.

[15] Das, G., Lin, K. I., Mannila, H., Renganathan, G., & Smyth, P. (1998, August). Rule Discovery from Time Series. In KDD (Vol. 98, pp. 16-22).

[16] Li, H. F., & Lee, S. Y. (2009). Mining frequent itemsets over data streams using efficient window sliding techniques. Expert Systems with Applications,36(2), 1466-1477.

[17] Xiaobin Wu.(2008). Rough Clustering of Financial Time Series in Data Ming.Journal of Xiamen University.

[18] Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis.Chemometrics and intelligent laboratory systems, 2(1), 37-52.

[19] Lin, J., Keogh, E., Lonardi, S., & Chiu, B. (2003, June). A symbolic representation of time series, with implications for streaming algorithms. InProceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery (pp. 2-11). ACM.

[20] Liu, H., & Setiono, R. (1995, November). Chi2: Feature selection and discretization of numeric attributes. In 2012 IEEE 24th International Conference on Tools with Artificial Intelligence (pp. 388-388). IEEE Computer Society.

[21] Keogh, E., Lonardi, S., & Chiu, B. Y. C. (2002, July). Finding surprising patterns in a time series database in linear time and space. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 550-556). ACM.

[22] Lonardi, J. L. E. K. S., & Patel, P. (2002). Finding motifs in time series. InProc. of the 2nd Workshop on Temporal Data Mining (pp. 53-68).

[23] Celebi, M. E., Kingravi, H. A., & Vela, P. A. (2013). A comparative study of efficient initialization methods for the k-means clustering algorithm. Expert Systems with Applications, 40(1), 200-210.

[24] Na, S., Xumin, L., & Yong, G. (2010, April). Research on k-means clustering algorithm: An improved k-means clustering algorithm. In Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on (pp. 63-67). IEEE.

[25] Aranganayagi, S., & Thangavel, K. (2007, December). Clustering categorical data using silhouette coefficient as a relocating measure. In Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on (Vol. 2, pp. 13-17). IEEE.

[26] Frahling, G., & Sohler, C. (2008). A fast k-means implementation using coresets. International Journal of Computational Geometry & Applications,18(06), 605-625.

[27] Smyth, P., & Goodman, R. M. (1991). Rule Induction Using Information Theory. Knowledge discovery in databases, 1991.

[28] Zadeh, L. A. (1965). Fuzzy sets. Information and control, 8(3), 338-353.