A Dynamic Pattern Recognition Approach Based on Neural Network for Stock Time-Series

Bo ZHOU and Jinglu HU
Graduate School of Information, Production and System, Waseda University
Hibikino 2-7, Wakamatsu-ku, Kitakyushu-shi, Fukuoka, JAPAN
Email: zhoubo_bird@toki.waseda.jp, jinglu@waseda.jp

Abstract—**Pattern theorem in financial time-series is one of the most important technical analysis methods in financial prediction. Recent researches have achieved big progresses in identifying and recognizing time-series patterns. And most of the recent works on time-series deal with this task by using static approaches and mainly focus on the recognition accuracy, but considering that recognition of patterns in financial time-series, especially for stock time-series, are always time-consuming rather than pattern recognition in other fields, a dynamic recognition approach is more preferable so that investment on stock pattern become executable. In this paper we propose a dynamic approach for extracting and recognizing the patterns in stock-series. In our approach a slide window with flexible length is defined for extracting feature vertexes in stock-series, and in addition, a dynamic perceptual important point (PIP) locating method is proposed based on the PIP locating method for avoiding the computation expense problem and an artificial neural network (ANN) approach is involved for pattern recognition and window length identification.**

Keywords-finacial time-series; stock series; dynamic approach; pattern recognition

I. INTRODUCTION

The pattern theorem in stock-series prediction is one of the most important technical analysis approaches and major related with the time and price domain. A stock pattern means an especial shape formed in stock time-series chart which indicates a prospective trend in price. In [1] over 50 widely authorized stock patterns have been studied and the research results indicate that each stock pattern will have a rather high probability to reach a prospective price.

Based on the pattern theorem in stock-series, scientists began to build computer-based recognition by the rule-based approaches [2,5] and the template-based approaches [3,4], all these researches always require highly professional skills as well as involves considerable risks thus posing an enormous challenge for technical analysis.

With the increasing popularity in using artificial neural network (ANN) approach in stock-series prediction, scientists began to study the usage of ANN in the stock pattern recognition field. Reference [6] proposed a recognition algorithm for triangle patterns based upon a recurrent neural network, and in [7] X. Guo, et al.(2007) proposed a feature vertexes extraction algorithm and used a multi-output ANN trying to classify the stock patterns.

Another direction of the researches is the segmenting of the time-series data, Reference [8] suggested dividing time-series sequence into meaningful sub-sequence and in [9] a fixed length window was used to segment time series into subsequences and a time series was then represented by the primitive shape patterns that were formed. Because the window length became one of the most important factors affecting the segmenting performance here, Reference [10] proposed a conception of flexible window length. And in [12] a genetic algorithm was involved in segmenting the stock time-series and in [11] a perceptual important point locating method was introduced in detecting patterns from the segmenting results.

Unfortunately, all the researches above took static approaches and the input data were long term sequences. The static approaches are not suitable for stock investment because the stock time-series are always strong time-consuming and long term sequences will bring in huge computation cost in a dynamic method. So, in this paper we propose a dynamic approach with purpose of recognizing the recent-formed stock patterns, by using which we can make the investment based on stock-series patterns executable.

This paper is organized into five sections. First we give a brief introduction to our approach in Section II and in Section III how our proposed system been designed is described in detail and In Section IV the simulation results demonstrate the effectiveness of our approach and finally we give conclusions in Section V.

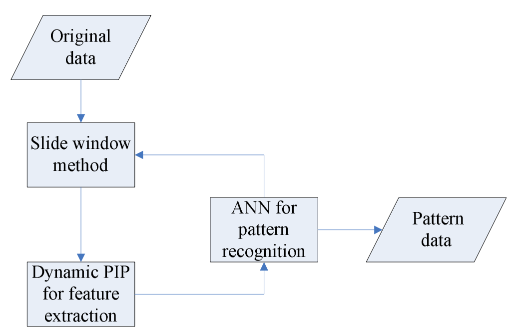

Figure 1. proposed approach

II. BRIEF INTRODUCTION TO THE PROPOSED APPROACH

This section gives a brief introduction to the proposed approach. A recurrent dynamic model is taken in proposed approach. As shown in Fig.1 the system has three components, a slide window with flexible length and a dynamic PIP locating algorithm are designed for the dynamic feature extraction and also with the focus on avoiding the computation expense problem, and an ANN is designed for the pattern recognition work by tracking the procedure of a pattern formed.

III. DETAIL SYSTEM DESIGN

A. Pattern Description

Usually a time-series patterns is composed of a group of segments with fixed numbers. Fig.2 gives an example of a head and shoulder pattern expressed by 7 vertexes and 6 segments. 11 classic patterns with this description model are involved in our researches.

And during the research procedure, another discovery is that these time-series patterns have some interior relationships between their vertexes. First, many patterns are symmetrical, such as head and shoulder pattern, two top/bottom pattern, triple top/bottom pattern, diamond pattern; and for the other patterns such as the triangle pattern and the bump and run patterns, their upside reversal points and the downside reversal points seem satisfied some linear relationships.

B. Slide window method

A slide window method is involved with purpose of snatching the recently formed, invest-able patterns dynamically. To achieve such purpose, the input data, slide window length and sliding strategy should be taken into concern. Each point in the stock time-series represents the daily ending price of one single exchange day and the slide window will slide forward each exchange day. The window length is a factor difficulty to be decided because it will directly affect the performance and efficient of the extraction and recognition work. Considering that the length of pattern differs greatly, even between the patterns with the same category, a flexible length approach will be preferable than fixed length approach, and we will describe how to design the window with flexible length later.

C. Dynamic Feature Extraction

In [11, 12], a Perceptual Important Point (PIP) locating method was proposed and proved its capacity in detecting patterns from the time-series segments.

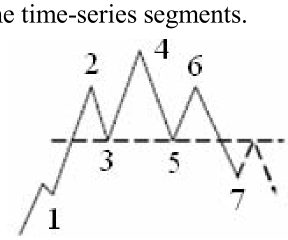

Figure 2. A head and shoulder pattern

```
Int PIP_tree(int Q(m), int start, int end){
    Tree_node t;
    start=0;
    end=m-1;  // initialize the sequence Q,
    for i=start:end;
    Dist(i)=Q(i)-(Q(end)-Q(start))*(i-start)/(end-start);
    t=Q(m);   //Q(m) is with the maximum Dist(i)
    t.left= PIP_tree(int Q(m), start, m)
    t.right=PIP_tree(int Q(m), m, end)
    return t;}
```

Figure 3. PIP locating with binary tree

```
Dynamic PIP(int Q[2:m+1], int start, int end, tree T){
    // T is the binary-tree constructed in searching Q[1:m]
    Check density variable//if Q[start] and Q[end] are near to
    each other the function break
    Define searching level;
    Tree_node T2,
    Searching Q[start], Q[end] in T;
    if find
        copy T;
    else {
    P={Q[start]; Q[end]; T with search level}
        T2=PIP_tree(Q);
    T2.left=Dynamic PIP(Q[], start, T2.index, T);
    T2.right=Dynamic PIP(Q[], T2.index, end, T); }
    Return T2;}
```

Figure 4. Dynamic PIP algorithm

Though the PIP locating method performs well in detecting feature vertexes, it’s not suitable for implementing the feature vertexes extraction in our dynamic approaches because its computation cost is huge. In PIP locating method the perceptual importance of the vertexes is related to the starting and ending vertexes of the time-series sequence, but the starting and ending vertexes are daily changing in our dynamic system, so if just recurrently using the PIP method a huge computation cost will be involved and the system efficiency will be reduced.

To satisfy the demand of the dynamic recognition system, we proposed a dynamic algorithm based on PIP locating:

First, a binary tree structure is used to implement the PIP locating method and organize the located PIP point. (Fig.3)

Then, the dynamic PIP locating algorithm (Fig.4) is proposed and ameliorations are made to avoid the computation expense problem.

a) Density variable: A density variable is set with the purpose of avoiding the dense distribution of vertexes. Usually the distance between the pattern vertexes is larger than 6, it’s meaningless to traverse a sequence a[k] to a[k+5], a density value will check and avoid such situations. In this research the density value is set as 5.

b) Shortening sequence length: An important fact is that the tree structure changed slowly and tiny when the slide window move forward, that means if a vertex get a higher position in the prior tree Q[*1*:*m*], the height will not change greatly in tree Q[*2*:*m*+*1*]. The design of shortening sequence length is based on such property. Assuming we want to search the root node of tree Q[*2*:*m*+*1*], a traverse to the first 3 levels of the nodes in tree Q[*1*:*m*] will guarantee that we can achieve the goal.

c) Vertex validation: The vertex validation aims to remove the overlapping computation. When a PIP method works on a sequence, first the sequence starting and ending vertexes are searched in the prior tree. Once the two vertexes been found constructing a parent-child relationship, which means the same sequence has been detected by a PIP method in constructing the prior binary tree, it is unnecessary to compute the sequence again and just need to use the part of the prior tree to construct the new tree.

D. Neural network(NN) design

For pattern recognition, three-layer NNs are designed to detect the interior relationships between the patterns vertexes. Assuming a pattern A with n vertexes and the first mth vertexes are already known, n-m multilayer neural network structures will be designed, for the kth network there will has $m+k-1$ input node, each node represents a vertex and the input will be both the time and price, and the output node will produce the $(m+k)$th vertex. Usually the variable m differs with the patterns. For a symmetrical pattern the m equals at least $n/2+1$ and for the triangle and broadening patterns the $m=4$.

For the 11 classic patterns in this research 4 Neural Networks are designed.

- NN1 has 3 input nodes and 2 output nodes and each output node corresponds to a predefined pattern. The corresponding patterns are two tops/bottoms patterns.
- NN2 has 4 input nodes and 7 output nodes and each output node corresponds to a predefined pattern. The corresponding patterns are two tops/bottoms, triple bottom, diamond top/bottom, head and shoulder top/bottom patterns.
- NN3 has 5 input nodes and 9 output nodes and each output node corresponds to a predefined pattern. The corresponding patterns are triple bottom, diamond top/bottom, head and shoulder top/bottom, symmetric triangle upside/downside, bump and run upside /downside patterns.
- NN4 has 6 input nodes and 9 output nodes and each output node corresponds to a predefined pattern. The corresponding patterns are triple bottom, diamond top/bottom, head and shoulder top/bottom, symmetric triangle upside/downside, bump and run upside /downside patterns.

E. Flexible length of slide window

Based on the description above, a flexible length slide window approach (Fig.5) is proposed. First ANN is trained, and dynamic PIP method is done to the slide window and the result is checked, if the output binary tree satisfies the input condition of the ANN, a matching is done, and the output of the neural network is checked, we will get the pattern or a vertex of the pattern, if we get vertex we decide whether we have to enlarge the window or not, and then the length changing is done to the slide window when the window sliding forward and redo this procedure.

IV. SIMULATIONS AND RESULTS

30 stocks from Hong Kong market, 37 stocks from US market and 200 stocks from Chinese Shenzhen and Shanghai stock exchange market are investigated in experiments, all of which stocks must have at least 5 years' price data, 378 patterns with 11 categories are taken into experiments.

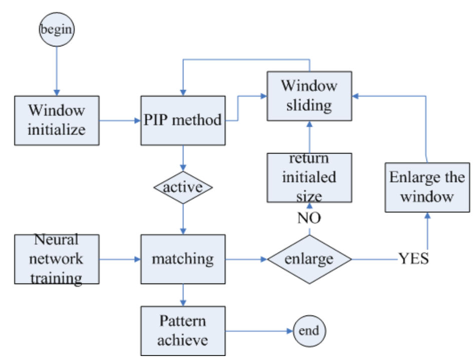

Figure 5. Flexible length window design

TABLE I. RESULT OF EXPERIMENT 1

Pattern	Expect	Static approach		Proposed approach			
		Find	Missed or error	Find with no delay	Acceptable delay (3-5 days)	Long delay (6-10 days)	Missed or error
two bottom	38	36	2	24	5	6	3
triple bottom	25	22	3	13	6	3	3
two top	29	28	1	13	11	2	3
diamond top	43	22	21	23	11	1	8
diamond bottom	45	26	19	11	24	1	10
head and shoulder top	29	29	0	10	14	3	2
head and shoulder bottom	33	31	2	8	17	3	5
bump&run top	29	18	11	12	7	1	6
bump&run bottom	26	14	12	8	11	0	7
triangle upside	39	36	3	20	11	0	2
triangle downside	42	41	1	18	19	1	4

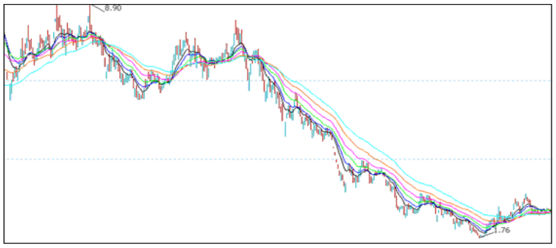

Figure 6. Series 1

TABLE II. RESULT FOR EXPERIMENT 2

Method	expected	Actual	Time for 1 sliding
Only PIP	70104	56171	0.49 second
PIP with density function	70104	37644	0.25 second
Dynamic PIP		10062	0.166 second
Dynamic PIP with density function		9759	0.15 second

A. *Experiment 1*

This experiment shows the performance of our proposed method in the pattern recognition works. For each pattern, dataset is divided into 3 groups, each time we take 2 groups for training and 1 group for testing. We summarize the results together in Tab.I and make a contrast with the static approach in Ref.[7], which using a clustering algorithm for feature extraction and ANN classifier for recognition. We can see that the proposed approach makes a good performance in recognizing most of the patterns and performs better in recognizing diamond and bump&run patterns whose feature are difficulty to identify.

B. *Experiment 2*

Experiment 2 focuses on the performances of the feature extraction algorithms, in these experiments we set a counter in the PIP algorithm, once the dynamic system traverse a point the counter add up, we can see how many times the dynamic method will traverse a point using different approaches.

Series 1(Fig.6) is 354 points length from the stock Shiji Wanke in Chinese Shanghai stock market (the stock ID 000005) from July 13th, 2007 to Jan 19th, 2009. This experiment aim to compare the performances between different feature extraction approaches.

From Tab.II we can see a density function reduces one third of the PIP computation and the dynamic PIP method does a rather good job for avoiding computation expense and reduces the computation cost to 20% of the original PIP, but a density variable combined PIP doesn't bring in much improvement.

V. CONCLUSION

Pattern recognition in time series system with dynamic approach is a rather new direction in the researches of stock patterns. Considering strongly time-consuming of the patterns in stock series, dynamic approach is very necessary for investors to analyze and make investment. Our proposed approach serves this requirement.

In this paper we proposed a dynamic recognition approach which made the real-time pattern recognition applicable. By using the reformative dynamic PIP locating method for the feature extraction and ANNs for the pattern recognition, we were able to track the procedure of a pattern shaped and make the dynamic approach more efficient. The empirical results proved the effectiveness of our approach.

We also found that there are some patterns were not suitable by using the expression method in this paper, because the segments number cannot be fixed. In the future we will extend our approaches to deal with such kinds of patterns and improve the performance of proposed approach.

REFERENCES

[1] N. Bulkowski. Encyclopedia of Chart Patterns, 2nd Edition. John Wiley and Sons, 2005.

[2] C. L. Osler, and P. H. K. Chang. "Head and Shoulders: Not Just a Flaky Pattern". Staff Report No.4, Federal Reserve Bank of New York.

[3] W. Leigh, N. Paz, and R. Purvis. "Market timing: a test of a charting heuristic". Economics Letters, 77(1)(2002), 55-63.

[4] W. Leigh, N. Modani, and R. Hightower. "A Computational Implementation of Stock Charting: Abrupt Volume Increase As Signal for Movement in New York Stock Exchange Composite Index". Decision Support Systems, 37(4)(2004), 515-530.

[5] A. Lo, H. Mamaysky, and J. Wang. "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation". *Journal of Finance*, 55(4)(2000), 1705-1765.

[6] K. Kamijo, and T. Tanigawa. "Stock Price Pattern Recognition: A Recurrent Neural Network Approach". *Proc. Of the Int. Joint Conf. on Neural Networks*, vol. 1, 1990, 215-221.

[7] Xinyu Guo, Xun Liang and Xiang Li, "A Stock Pattern Recognition Algorithm Based on Neural Networks" ,*Third International Conference on Natural Computation* (ICNC 2007),0-7695-2875-9/07 IEEE

[8] H. Shatkay and S. B. Zdonik, "Approximate queries and representations for large data sequences," in *Proc. Int. Conf. Data Engineering*. Los Alamitos, CA: IEEE Computer Society Press, 1996, pp. 536–545.

[9] G. Das, K. I. Lin, and H. Mannila, "Rule discovery from time series,"in Proc. ACM SIGKDD *Int. Conf. Knowledge Discovery Data Mining*, 1998, pp. 16–22.

[10] T. C. Fu, F. L. Chung, V. Ng, and R. Luk, "Evolutionary segmentation of financial time series into subsequences," in *Proc. Congr. Evolutionary Computation*, Seoul, Korea, 2001, pp. 426–430.

[11] F. L. Chung, T. C. Fu, R. Luk, and V. Ng, "Flexible time series pattern matching based on perceptually important points," in *Proc. Int. Joint Conf. Artificial Intelligence Workshop (Learning From Temporal and Spatial Data)*, Seattle, WA, Aug., 4–10 2001, pp. 1–7.

[12] Fu-Lai Chung, Tak-Chung Fu, Vincent Ng, and Robert W. P. Luk, "An Evolutionary Approach to Pattern-Based Time Series Segmentation", *IEEE Trans. on evolution computation*, Vol. 8, Vol. 5, OCTOBER 2004