Online Evolving Fuzzy Rule-Based Prediction Model For High Frequency Trading Financial Data Stream

Xiaowei Gu, Plamen P. Angelov, *Fellow, IEEE*,
and Azliza Mohd Ali
School of Computing and Communications
InfoLab21, Lancaster University,
Lancaster, LA1 4WA, UK
E-mail: p.angelov@lancarster.ac.uk

William A. Gruver[†], *Fellow, IEEE*
Simon Fraser University, Burnaby, BC, Canada
E-mail: gruver@sfu.ca

Georgi Gaydadjiev[1,2], *Senior Member, IEEE,*
[1]Maxeler Ltd, 3&4, Albion Place, London, W6 0QT, UK
[2]Department of Computing, Imperial College, UK
E-mail: georgi@maxeler.com

***Abstract*— Analyzing and predicting the high frequency trading (HFT) financial data stream is very challenging due to the fast arrival times and large amount of the data samples. Aiming at solving this problem, an online evolving fuzzy rule-based prediction model is proposed in this paper. Because this prediction model is based on evolving fuzzy rule-based systems and a novel, simpler form of data density, it can autonomously learn from the live data stream, automatically build/remove its rules and recursively update the parameters. This model responds quickly to all unpredictable sudden changes of financial data and re-adjusts itself to follow the new data pattern. Experimental results show the excellent prediction performance of the proposed approach with real financial data stream regardless of quick shifts of data patterns and frequent appearances of abnormal data samples.**

Keywords*— *online learning; online prediction; fuzzy rule based systems; high frequency financial data stream; recursively updating; data density

I. Introduction

High frequency trading (HFT) systems employ computer algorithms to analyze market data, determine buy and sell recommendations, submit trades to exchanges for execution, and manage trades after execution [1]. Although firms are secretive about their algorithms, many HFT algorithms are based on technical analysis that uses historical prices and indicators to identify zones of supply/demand where sellers/buyers are likely to change the price of the product. A complementary technique, known as quantitative analysis, uses mathematical and probabilistic models to make predictions.

Due to the fast changes and the massive of the global markets, individual agents and companies have to deal with a huge amount of high frequency financial data. Although, modern transportation and communication technologies have already minimized the distance between domestic markets of different countries, traditional methods of collecting and analyzing financial information, which depends mainly on manual operation and low accuracy methods, largely lower the efficiency of global business and inevitably increase errors in decisions. These errors can be very costly.

Recently, autonomous learning techniques have been developed in the framework of fuzzy rule-based systems [2,3] that adapt the system structure with new data samples arriving online, which means the structure can grow and shrink according to the shift/drift of the data stream [4,5]. Fuzzy rule-based systems now have been applied to classification [6,7], clustering [8], identification [9], fault detection [9,10], data prediction [11] and control problems [12] in the fields of machine learning, control theory, anomaly detection and data analytics successfully. Nonetheless, developing suitable self-learning predictive models for high frequency trading financial data stream is still challenging because of the continuous arrival of high frequency data samples.

In this paper, we propose an online evolving fuzzy rule-based model for predicting the trend of HFT financial data and a new, simpler form of data density. This model can automatically build/remove its rules, self-adjust the parameters as well as select inputs by learning from the HFT financial data accumulated from the past and present. Although sudden changes often present within HFT financial data, the proposed model can adapt quickly to follow the new trend.

In addition, the proposed approach also has the following advantages: it is recursive, incremental and memory efficient. This means that it is non-iterative, so that all parameters can be updated recursively, and the model sequentially learns the current input data and discards all the previous data at the same time and only retains important.

The rest of the paper is organized as follows. Theoretical basis of the proposed approach is provided in section II. Section III describes the details of the proposed approach and the experimental results are presented in section IV. Section V provides the conclusions.

II. Theoretical Basis of The Proposed Approach

In this section, the theoretical basis of the proposed approach will be introduced.

† Deceased.

This work was supported by The Royal Society (Grant number IE141329/2014).

We assume the high frequency trading financial data stream denoted as $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,...,\boldsymbol{x}_k,...\}$, where the subscript k denotes the time instance when that the data point $\boldsymbol{x}_k$ in the stream arrives. Each data point has d dimensions, denoted by $\boldsymbol{x}_k=\left[x_{k,1},x_{k,2},...,x_{k,d}\right]^{\mathrm{T}}$; $\boldsymbol{Y}=\{y_1,y_2,y_3,...,y_k,...\}$ defines the output of fuzzy rule-based model; $k=1,2,3,4,...$.

Each data sample, both the input and the output, are standardized online by recursively updating the mean and standard deviation [1]:

$$x_{k,j}^{st}=\frac{x_{k,j}-\bar{x}_{k,j}}{\sigma_{k,j}},\bar{x}_{k,j}=\frac{k-1}{k}\bar{x}_{k-1,j}+\frac{1}{k}x_{k,j},\bar{x}_{1,j}=x_{1,j} \tag{1a}$$

$$\sigma_{k,j}^2=\frac{k-1}{k}\sigma_{k-1,j}^2+\frac{1}{k-1}\left(x_{k,j}-\bar{x}_{k,j}\right)^2,\sigma_{1,j}^2=0 \tag{1b}$$

$$y_k^{st}=\frac{y_k-\bar{y}_k}{\delta_k},\bar{y}_k=\frac{k-1}{k}\bar{y}_{k-1}+\frac{1}{k}y_k,\bar{y}_1=y_1 \tag{2a}$$

$$\delta_k^2=\frac{k-1}{k}\delta_k^2+\frac{1}{k-1}\left(y_k-\bar{y}_k\right)^2,\delta_1^2=0 \tag{2b}$$

where $\bar{x}_{k,j}$, $\sigma_{k,j}$ and $x_{k,j}^{st}$ denote the mean, standard deviation and standardized form of input data sample in the j^{th} dimension, respectively; $\bar{y}_k$, δ_k and y_k^{st} are the mean, standard deviation and standardized form of output data; $j=1,2,...,d$.

The cumulative proximity of every new input data sample in the data space is defined as:

$$\pi_k\left(\boldsymbol{x}_k\right)=\sum_{i=1}^{k-1}\left\|\boldsymbol{x}_k-\boldsymbol{x}_i\right\|^2 \tag{3}$$

The density of the k^{th} input data is as following [13]:

$$D_k\left(\boldsymbol{x}_k\right)=\frac{\sum_{i=1}^{k}\pi_k\left(\boldsymbol{x}_i\right)}{2k\pi_k\left(\boldsymbol{x}_k\right)} \tag{4}$$

Meanwhile, the online recursive form of the cumulative proximity π_k plays a very significant role in the proposed model because of the extremely huge amount of the arriving data samples. The recursive expression is as follows [1]:

$$\pi_k\left(\boldsymbol{x}_k\right)=k\left(\left\|\boldsymbol{x}_k-\boldsymbol{\mu}_k\right\|^2+X_k-\left\|\boldsymbol{\mu}_k\right\|^2\right) \tag{5a}$$

$$\pi_k\left(\boldsymbol{x}_i\right)=\pi_{k-1}\left(\boldsymbol{x}_i\right)+\left\|\boldsymbol{x}_i-\boldsymbol{x}_k\right\|^2,i=1,2,...,k-1 \tag{5b}$$

Where $\boldsymbol{\mu}_k=\frac{k-1}{k}\boldsymbol{\mu}_{k-1}+\frac{1}{k}\boldsymbol{x}_k,\boldsymbol{\mu}_1=\boldsymbol{x}_1$ (6a)

$$X_k=\frac{k-1}{k}X_{k-1}+\frac{1}{k}\left\|\boldsymbol{x}_k\right\|^2,X_1=\left\|\boldsymbol{x}_1\right\|^2. \tag{6b}$$

Additionally, the sum of cumulative proximity π_k of all the previously existing data samples can also be updated recursively [1, 13]:

$$\sum_{i=1}^{k}\pi_k\left(\boldsymbol{x}_i\right)=\sum_{i=1}^{k-1}\pi_{k-1}\left(\boldsymbol{x}_i\right)+2\pi_k\left(\boldsymbol{x}_k\right) \tag{7}$$

Therefore, the density $D\left(\boldsymbol{x}_k\right)$ of every arriving input data sample can be calculated recursively using (5)-(7).

The structure of the proposed evolving fuzzy rule-based model can be interpreted as an evolving fuzzy set of linguistic fuzzy rules of the multi-input-single-output (MISO) type:

$$Rule^i:IF\left(x_{k,1}\sim x_{i,1}^*\right)AND\left(x_{k,2}\sim x_{i,2}^*\right)AND...AND\left(x_{k,d}\sim x_{i,d}^*\right)$$
$$THEN:\left(y_{k,i}^*=a_{k,i,0}+\sum_{j=1}^{d}a_{k,i,j}x_{k,j}\right) \tag{8}$$

where $\left(x_{k,j}\sim x_{i,j}^*\right)$ is the j^{th} fuzzy set of the i^{th} fuzzy rule; $a_{k,i,j}$ is the corresponding parameter, $j=1,2,...,d$; $\boldsymbol{x}^*=\left[x_{i,1}^*,x_{i,2}^*,...,x_{i,d}^*\right]^{\mathrm{T}}$ is the center of the i^{th} fuzzy rule, $i^*=1,2,...,N_k$; N_k is the number of the rules at the time-instant *k*. The initial radius of each rule is $\boldsymbol{r}_{1,i}=\left[0.5,...,0.5\right]_{1\times d}$ assuming normalized data, and it is leaving updated later.

The rules can also be transformed into the following type:

$$y_{k,i}^*=\bar{\boldsymbol{x}}_k^{\mathrm{T}}\boldsymbol{a}_{k,i} \tag{9}$$

Here $\bar{\boldsymbol{x}}_k^{\mathrm{T}}=\left[1,\boldsymbol{x}_k^{\mathrm{T}}\right]$; $\boldsymbol{a}_{k,i}=\left[a_{k,i,0},a_{k,i,1},a_{k,i,2},...,a_{k,i,d}\right]^{\mathrm{T}}$ is the antecedent parameter of the i^{th} rule at the time-instant *k*; $y_{k,i}^*$ is the corresponding output.

In the proposed model, all the rules cooperate to make inference (output) through the *center of gravity* aggression mechanism. For each data sample, the membership function of each fuzzy set is defined in the form of a bell function due to their generalization capabilities [7]:

$$\phi_{k,i,j}=e^{\left\{-\frac{\left(x_{k,j}-x_{i,j}^*\right)^2}{2r_{k,i,j}^2}\right\}};j=1,2,...,d;i=1,2,...,N_k \tag{10}$$

The corresponding activation level of each rule is expressed as the product of the respective membership values of all the fuzzy sets of this rule [3]:

$$\tau_{k,i}=\prod_{j=1}^{d}\phi_{k,i,j} \tag{11}$$

The degree of activation of each rule is proportional to the level of its contribution to the overall output of the evolving fuzzy rule-based model. Therefore, the overall output is expressed as a fuzzy combination of the outputs of all the existing fuzzy rules:

$$y_k = \sum_{i=1}^{N_k} \lambda_{k,i} y_{k,i}^* = \bar{\boldsymbol{x}}_k^{\mathrm{T}} \boldsymbol{A}_k \boldsymbol{\lambda}_k \tag{12}$$

where $\boldsymbol{A}_k = \left[\boldsymbol{a}_{k,1}, \boldsymbol{a}_{k,2}, \ldots, \boldsymbol{a}_{k,N_k}\right]$; $\boldsymbol{\lambda}_k = \left[\lambda_{k,1}, \lambda_{k,2}, \ldots, \lambda_{k,N_k}\right]^{\mathrm{T}}$; $\lambda_{k,i}$ is the normalized activation level of the i^{th} rule at time-instant *k*, defined as follows [3]:

$$\lambda_{k,i} = \frac{\tau_{k,i}}{\sum_{j=1}^{N_k} \tau_{k,j}} \tag{13}$$

Each rule will additionally have a degree called '*Utility*' indicating how much it has been used since it was generated [2]. *Utility* is the accumulated normalized activation level, defined as:

$$U_{k,i} = \frac{\sum_{j=1}^{k} \lambda_{j,i}}{k - T_i} \tag{14}$$

where T_i is the time-instant indicating when the i^{th} rule was formed.

Along with the sequential arrival of the data samples, the density of every rule center should be updated online to follow the drift/shift of data stream. Because the cumulative proximity of the centers can be updated recursively as well using (5b), we can update the densities of centers in the similar way as recursive calculation of $D_k\left(\boldsymbol{x}_k\right)$ using (4)-(7) as:

$$D_k\left(\boldsymbol{x}_i^*\right) = \frac{\sum_{i=1}^{k} \pi_k(\boldsymbol{x}_i)}{2k\pi_k(\boldsymbol{x}_i^*)} = \frac{\sum_{i=1}^{k-1} \pi_{k-1}(\boldsymbol{x}_i) + 2\pi_k(\boldsymbol{x}_k)}{2k\left(\pi_{k-1}(\boldsymbol{x}_i^*) + \left\|\boldsymbol{x}_i^* - \boldsymbol{x}_k\right\|^2\right)} \tag{15}$$

The radii of influence of the fuzzy rules are updated as [1]:

$$\begin{cases} \boldsymbol{r}_{k,i} = \dfrac{\boldsymbol{r}_{k-1,i} + \boldsymbol{\sigma}_k^2}{2} & i = \arg\min_{j=1}^{N_k}\left(\left\|\boldsymbol{x}_k - \boldsymbol{x}_j^*\right\|^2\right) \\ \boldsymbol{r}_{k,i} = \boldsymbol{r}_{k-1,i} & else \end{cases} \tag{16}$$

where $\boldsymbol{\sigma}_k^2 = \left[\sigma_{k,1}^2, \sigma_{k,2}^2, \ldots, \sigma_{k,d}^2\right]$.

The antecedent parameters of the fuzzy rules need to be updated online as well; in this approach we update them locally [2,3]:

$$\boldsymbol{a}_{k+1,i} = \boldsymbol{a}_{k,i} + \boldsymbol{C}_{k,i} \bar{\boldsymbol{x}}_k \lambda_{k,i}\left(y_k - \bar{\boldsymbol{x}}_k^T \boldsymbol{a}_{k,i}\right), \boldsymbol{a}_{1,i} = \boldsymbol{0} \tag{17a}$$

$$\boldsymbol{C}_{k+1,i} = \boldsymbol{C}_{k,i} - \frac{\lambda_{k,i} \boldsymbol{C}_{k,i} \bar{\boldsymbol{x}}_k \bar{\boldsymbol{x}}_k^T \boldsymbol{C}_{k,i}}{1 + \lambda_{k,i} \bar{\boldsymbol{x}}_k \boldsymbol{C}_{k,i} \bar{\boldsymbol{x}}_k^T}, \boldsymbol{C}_{1,i} = \Omega \boldsymbol{I}_{(d+1)\times(d+1)} \tag{17b}$$

where $\boldsymbol{I}_{(d+1)\times(d+1)}$ is a $(d+1)\times(d+1)$ unitary matrix, and Ω is a large constant number, usually $\Omega = 500$.

III. Evolving Fuzzy Rule-based HFT Data Prediction Approach

In this section, we will introduce the proposed approach in detail.

A. The Proposed Prediction Model

Firstly, let us assume that there is no any existing rule or structure, namely 'start from scratch', as it is the most general and challenging case.

For every input data sample, it will always lead to two possible results:

a) It is associated with the existing rule;

b) It is the first member of a new rule.

For every new data sample, after the density is calculated by (4) and all the centers' densities are updated using (15), we check the Condition A as follows [2]:

Condition A:

$$\begin{gathered} IF\left(D_k(\boldsymbol{x}_k) > \max_{i=1}^{N_k}\left(D_k\left(\boldsymbol{x}_i^*\right)\right) OR\ D_k(\boldsymbol{x}_k) < \min_{i=1}^{N_k}\left(D_k\left(\boldsymbol{x}_i^*\right)\right)\right) \\ THEN\left(N_k \leftarrow N_k + 1\right) \end{gathered} \tag{18}$$

If Condition A is not satisfied, the new data sample belongs to one of the existing rule, namely this data sample leads to the result a).

In contrast, when Condition A is met, it means that the new coming data sample is initiating a new rule, namely it leads to the result b), and the number of the rules should be updated to N_k*+1*. To be more general, we always use N_k as the number of rules at the current time-instant.

There are always outliers in data streams, and the outliers will create new rules as they normally will have the minimum densities. However, due to the degree of *Utility*, the rules created by outliers will be removed shortly through the structure evolution, and we will introduce this later in Condition C.

Then Condition B is trigged to avoid overlap of the existing rule with the new rule [1]:

Condition B:

$$\begin{gathered} IF\left(\exists i, i = [1, N_k] : \phi_{k,i,j} > e^{-1}, \forall j, j = [1,d], k = 2,3,\ldots\right) \\ THEN\left(N_k \leftarrow N_k - 1\right) \end{gathered} \tag{19}$$

When Condition B is satisfied, it means the new center describes any of the previously existing rule centers to a relatively high degree, and the existing overlapping rule should be removed.

The antecedent parameters of the new added rule are obtained in two ways:

a) Condition B is not met: $\boldsymbol{a}_{k,N_k+1} = \sum_{i=1}^{N_k} \lambda_{k,i} \boldsymbol{a}_{k,i}$; or

b) Condition B is met: $\boldsymbol{a}_{k,N_k+1}=\boldsymbol{a}_{removed}$; $\boldsymbol{a}_{removed}$ is the parameter vector of the removed rule;

and $\boldsymbol{C}_{k,N_k}=\Omega\boldsymbol{I}_{(d+1)\times(d+1)}$, $\boldsymbol{r}_{1,N_k}=\left[0.5,...,0.5\right]_{1\times d}$.

After Condition A and Condition B are checked, the *Utilities* of all the rules are updated using (14).

Since the structure of the evolving model is flexible, that is to say, new rule can be added for better coverage of the data space, while old rule should also be removed if the data pattern shifted away from this rule. Here Condition C is used to discover and disable the old rule [1].

Condition C:

$$\begin{aligned}&IF\left(\exists U_{k,i}<\eta, i=1,2,...,N_k\right)\\&\quad THEN\left(N_k\leftarrow N_k+1\right)\end{aligned}\tag{20}$$

where η is a constant parameter, $\eta\in\left[0.03,0.1\right]$. Once Condition C is satisfied, the rule is removed.

In section II, we have assumed the dimension of input data samples as d. However, in practical cases, often there is a large variety of possible inputs, which are inner correlated. Therefore, it is very important to introduce the input selection operation. In this section, besides the three conditions above, we additionally add Condition D [1]:

Condition D:

$$\begin{aligned}&IF\left(\exists j\mid\omega_{k,i,j}<\varepsilon\bar{\omega}_{k,j}, j=1,2,...,d,\ i=1,2,...,N_k\right)\\&\quad THEN\left(remove\ the\ j^{th}\ set\ from\ the\ i^{th}\ rule\right)\end{aligned}\tag{21}$$

where $\omega_{k,i,j}$ is the normalized accumulated sum of parameter values at the time-instant k, $\omega_{k,i,j}=\dfrac{\pi_{k,i,j}}{\sum_{j=1}^{d}\pi_{k,i,j}}$; $\bar{\omega}_{k,j}$ is the average value of the normalized accumulated sum of the j^{th} fuzzy set of all the rules, $\bar{\omega}_{k,j}=\dfrac{\sum_{i=1}^{N_k}\omega_{k,i,j}}{N_k}$; ε is a constant parameter, $\varepsilon\in[0.03,0.05]$. The accumulated sum of parameter values, $\pi_{k,i,j}$, is expressed as [1]:

$$\pi_{k,i,j}=\sum_{t=1}^{k}\left|a_{t,i,j}\right|, j=1,2,...,d\tag{22}$$

If Condition D is met, we remove the corresponding fuzzy set of the rule, and set the radius of this fuzzy set as the initial radius.

Once Condition C and Condition D have been checked, the updating operation of the structures of all the existing fuzzy rules is finished.

Finally, the antecedent parameters of all the rules are updated through the utilization of (17a) and (17b), and the inference (output of the prediction model) is made using (12).

B. Approach Summary

The overall approach process is summarized in this section in a form of a pseudo-code.

```
Stage A. Initialize the structure of the fuzzy rule-based prediction model with the first data sample.
Stage B. Read and standardize the next data sample using (1a), (1b), (2a) and (2b).
Stage C. Calculate the density of the new data sample by (5a), (6) and (7).
Stage D. Update the densities of the centers of the existing rules using (15).
Stage E. Check Condition A.
    *If Condition A is met, set the new data sample as the center of a new rule, and go to Stage F.
    *If Condition A is not met, go to Stage G directly.
Stage F. Check Condition B.
    *If Condition B is met, remove the overlapping rule.
Stage G. Update the Utility of each rule by (14) and check Condition C.
    *If Condition C is met, remove the old rule.
Stage H. Check Condition D.
    *If Condition D is met, remove the useless input for the rule.
Stage I. Update the radii and normalized activation levels of all the existing rules using (16) and (13).
Stage J. Update the antecedent parameters of all the fuzzy rules using (17a) and (17b).
Stage K. Make inference using (12) and go back to Stage B if new input data is available.
```

IV. Numerical Experiments

In this section, several numerical experiments with real HFT data are conducted to testify the out of sample prediction performance of the proposed approach.

The data stream that we use in the experiments is the QuantQuote Second Resolution Market Database [15], which contains tick-by-tick data on all NASDAQ, NYSE, and AMEX securities from 1998 to the present moment in time. Survivorship bias free lists of major indexes are maintained. This dataset contains 19144 data samples, and each sample has seven attributes: 1) Time; 2) Open price; 3) High price; 4) Low price; 5) Close price; 6) Trading volume; 7) Suspicious price. Here we only use the four of them (2)-5)) for prediction of the future values of 3) 5, 30, 60, 75, 90 steps ahead. The frequency of tick data varies from one second to few minutes.

The algorithm was developed using MATLAB R2015a, performance was evaluated on a PC with processor 3.60 GHz×2, and 8 GB RAM.

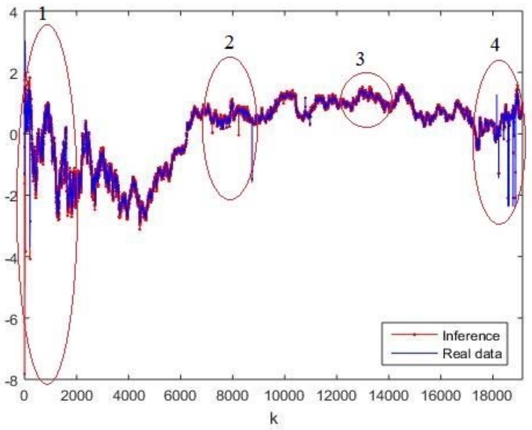


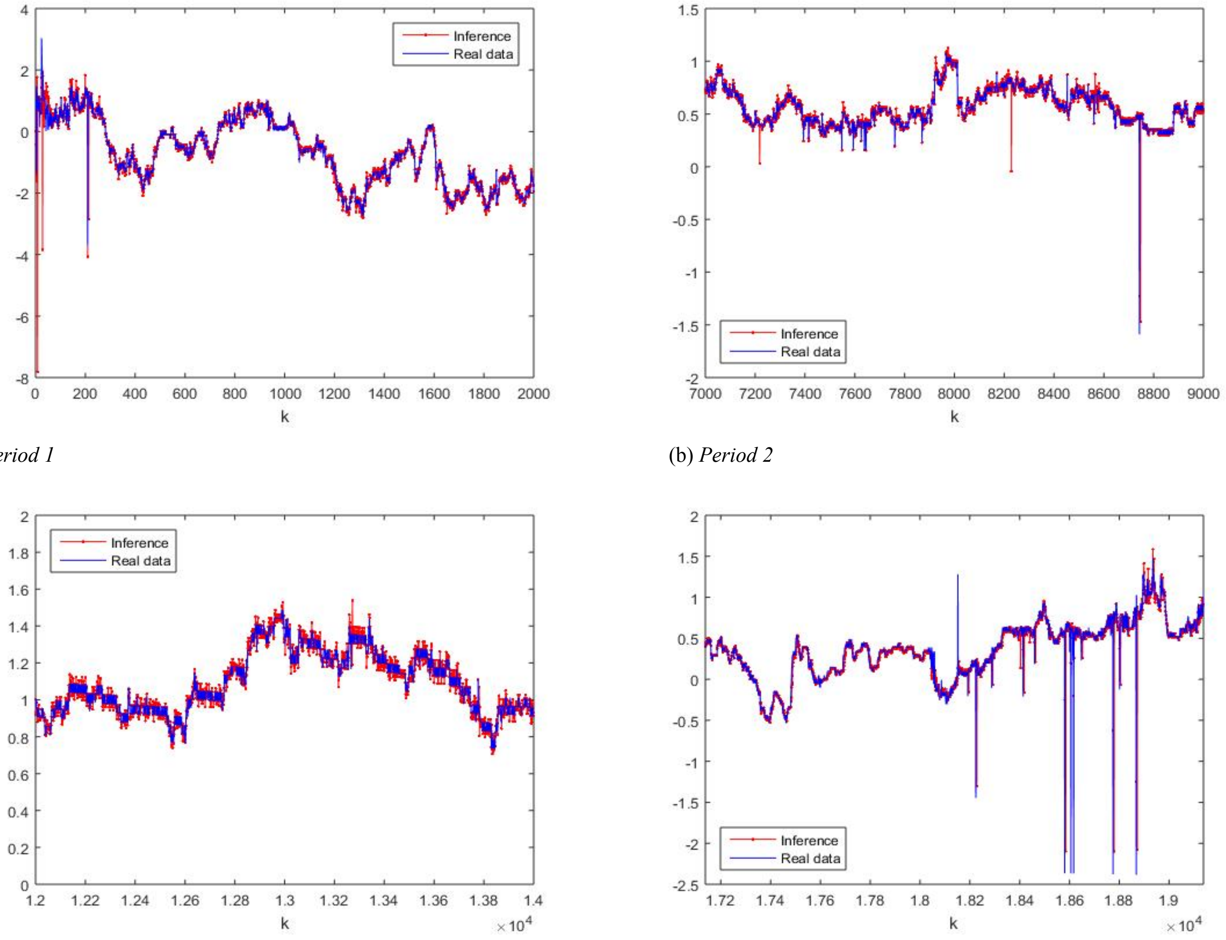
Fig. 1. Overall Prediction Result

Firstly, data sample $\boldsymbol{X}_k(O_k, H_k, L_k, C_k)$ is used as the input of the prediction model to predict the high price five steps ahead H_{k+5}. The prediction result is shown in Fig. 1. Four periods in Fig. 1 are zoomed in to further illustrate the prediction result, and the figures are shown in Fig. 2.

As we can see from Fig. 1, there are many abnormal data samples and random fluctuations in the data stream. At the beginning of the data stream, large fluctuations keep appearing, which can be clearly seen in Fig. 2(a), while most of the abnormal data samples appear together in the last part of the data stream as shown in Fig. 2(d). In the second zoom-in period of the data stream (shown in Fig. 2(b)), the change of data pattern is relatively smooth with slightly abnormal data samples and in Fig. 2(c), there is no obvious abnormal data samples and only normal fluctuations. Although, the trend of the data is unpredictable, the proposed approach can always adapt itself to follow the shift/drift of the pattern of the data stream [4].

Three fuzzy rules of the final rule base (12 rules in total) are given as examples:

$$\textit{Rule 1:}\quad \begin{aligned} &IF\left(O_k \sim -0.0892\right) AND\left(H_k \sim 0.7293\right)\\ &\quad AND\left(C_k \sim 0.7441\right)\\ &THEN\begin{pmatrix} H_{k+5} = 0.0113 + 0.2324O_k + 0.932H_k \\ \qquad + 0.4526C_k \end{pmatrix}\end{aligned};$$



(a) *Period 1*

(b) *Period 2*

(c) *Period 3*

(d) *Period 4*

Fig. 2. Zoom-in results

$$\textit{Rule 2}\quad \begin{array}{c} IF\left(H_k \sim 0.1421\right) AND\left(C_k \sim 0.1557\right) \\ THEN\left(H_{k+5} = 0.0956 + 0.8615H_k - 0.0414C_k\right) \end{array};$$

$$\textit{Rule 3:}\quad \begin{array}{c} IF\left(O_k \sim 0.9294\right) AND\left(H_k \sim 0.9188\right) \\ AND\left(L_k \sim 0.7908\right) AND\left(C_k \sim 0.7797\right) \\ THEN\left(\begin{array}{c} H_{k+5} = 0.2527 + 0.2341O_k + 0.2314H_k \\ +0.2151L_k + 0.2123C_k \end{array}\right) \end{array};$$

The evolution of the modal structure (number of fuzzy rules) in time is presented in Fig. 3.

The model structure evolution shown in Fig.3 also exhibits the corresponding relationship between the evolution of the data pattern and the changes of the structure of the prediction model.

During the period 1, the initial model is being formed “from scratch”. There are a few abnormal data samples happened in period 2, and several rules are formed because of them and quickly disabled due to the low Utility. There are no significant changes in period 3. Finally, in period 4, the market becomes volatile, and the modal structure evolves to follow a new trend.

TABLE I. PERFORMANCE DEMONSTRATION AND COMPARISION

Input and Output	**Performance**		
	Approach	***CORR***	***RMSE***
Input: $\boldsymbol{X}_k$ Output: H_{k+5}	The proposed approach	**0.9946**	**0.1128**
	LSLR agorithm	0.9919	0.1440
Input: $\boldsymbol{X}_k$ Output: H_{k+30}	The proposed approach	**0.9929**	**0.1303**
	LSLR agorithm	0.9723	0.2628
Input: $\boldsymbol{X}_k + \boldsymbol{X}_{k-1}$ Output: H_{k+5}	The proposed approach	**0.9977**	**0.0733**
	LSLR agorithm	0.9919	0.1443
Input: $\boldsymbol{X}_k + \boldsymbol{X}_{k-1}$ Output: H_{k+30}	The proposed approach	**0.9970**	**0.0841**
	LSLR agorithm	0.9712	0.2684
Input: $\boldsymbol{X}_k + ... + \boldsymbol{X}_{k-2}$ Output: H_{k+5}	The proposed approach	**0.9985**	**0.0582**
	LSLR agorithm	0.9917	0.1467
Input: $\boldsymbol{X}_k + ... + \boldsymbol{X}_{k-2}$ Output: H_{k+30}	The proposed approach	**0.9977**	**0.0736**
	LSLR agorithm	0.9711	0.2687
Input: $\boldsymbol{X}_k + ... + \boldsymbol{X}_{k-2}$ Output: H_{k+60}	The proposed approach	**0.9974**	**0.0788**
	LSLR agorithm	0.9621	0.3028
Input: $\boldsymbol{X}_k + ... + \boldsymbol{X}_{k-3}$ Output: H_{k+75}	The proposed approach	**0.9965**	**0.0920**
	LSLR agorithm	0.9620	0.3025
Input: $\boldsymbol{X}_k + ... + \boldsymbol{X}_{k-4}$ Output: H_{k+75}	The proposed approach	**0.9968**	**0.0878**
	LSLR agorithm	0.9614	0.3050
Input: $\boldsymbol{X}_k + ... + \boldsymbol{X}_{k-5}$ Output: H_{k+90}	The proposed approach	**0.9946**	**0.1132**
	LSLR agorithm	0.9601	0.3096

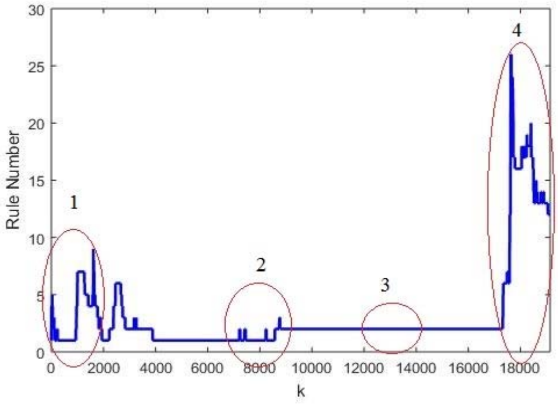
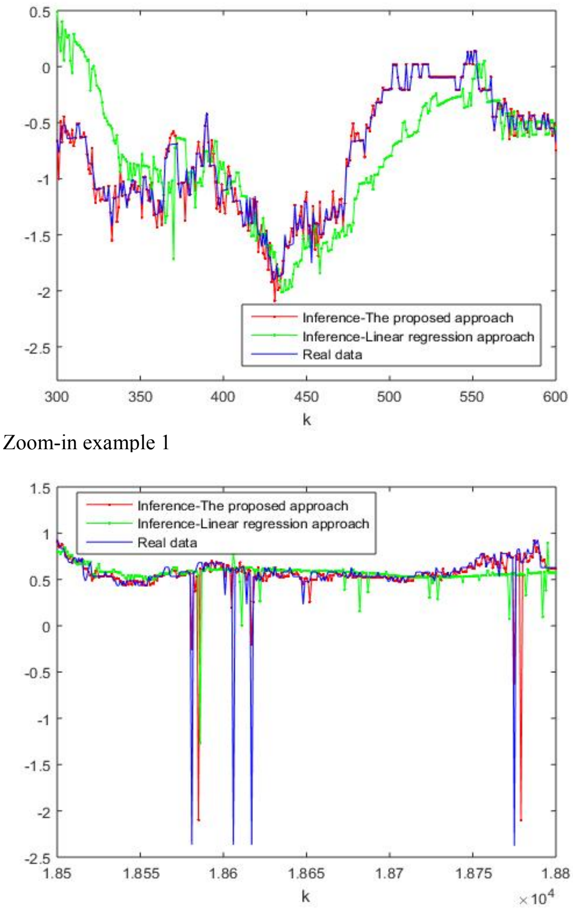


Fig. 3. Change of the number of rules

In order to further demonstrate the performance of the proposed approach, more experiments have been done and the results are shown in Table I. Additionally, we compared the proposed approach with least square linear regression (LSLR) algorithm that is often used in the fields of finance and



(a) Zoom-in example 1

(b) Zoom-in example 2

Fig. 4. Example of performance comparison (input is $\boldsymbol{X}_k$ and output is H_{k+5})

economy [15, 16], and the results are depicted in the Table I. Two zoom-in examples of the comparison result (input is $\boldsymbol{X}_k$ and output is H_{k+5}) are also presented in Fig. 4. The width of the sliding window for LSLR algorithm is 200.

As it is clear from Table I as well as Fig. 4, the performance of the proposed evolving fuzzy rule-based prediction approach is better when more input data samples are involved and the fewer steps are taken before the inference (output). Compared with the least square linear regression (LSLR) algorithm, the proposed approach always exhibits a significantly better performance, which shows the potential of our approach.

Moreover, it has to be stressed that, the time interval of between two data ticks ranges from as little as a second to few minutes, however the proposed method only needs less than 0.001 second in average to process each tick in the case in which the inputs are $\boldsymbol{X}_k + ... + \boldsymbol{X}_{k-5}$ and output is H_{k+90} . Therefore, the proposed algorithm allows a frequency/time interval as high as milliseconds.

In addition, the proposed method offers a better insight and clear, transparent model unlike methods such as neural networks and, support vector machines.

V. Conclusion

In this paper, we presented a new online evolving fuzzy rule-based prediction model for high frequency trading data streams. This model will automatically follow unpredictable sudden changes of the financial data, and automatically update its parameters online. Numerical experiments show that this approach can predict the future trend of financial data with high accuracy based on historical data. The high prediction performance and the advantage of recursively updated parameters make this approach a suitable tool for handling large amounts of high speed financial data.

Acknowledgment

This paper is dedicated to the memory of William A. Gruver, who passed away on Feb. 29, 2016. Dr. Gruver was a pioneer of the study of robotics and intelligent systems, *IEEE* fellow, laboratory head of Intelligent/Distributed Enterprise Automation Laboratory and professor emeritus of School of Engineering Science at Simon Fraser University, Canada.

References

[1] G. Nuti, M. Mirghaemi, et al., "Algorithmic trading," IEEE Computer, vol. 44(11), 2011.

[2] P. Angelov, Autonomous Learning Systems from Data Stream to Knowledge in Real Time. West Sussex, United Kingdom: John Wiley & Sons, Ltd., 2012.

[3] P. Angelov and R. Buswell, "Identification of evolving fuzzy rule-based models," IEEE Transactions on Fuzzy Systems, 2002, vol.10(5), pp. 667-677.

[4] P.Angelov and E.Lughofer, "Data-driven evolving fuzzy systems using eTS and FLEXFIS: comparative analysis," International Journal of General Systems, 2008, vol.37(1), p.45-67

[5] E. Lughofer, P. Angelov, "Handling drifts and shifts in on-line data streams with evolving fuzzy systems," Applied Soft Computing Journal, 2011, Vol.11(2), pp. 2057-2068

[6] S. Ahmad and M. Richard, "Understanding consumer's brand categorization across three countries: Application of fuzzy rule-based classification," Journal of Business Research, 2014, vol. 67(3), pp. 278-287.

[7] P. Angelov and X. Zhou, "Evolving fuzzy-rule-based classifiers from data streams", IEEE Transactions on Fuzzy Systems, 2008, vol.16(6), pp. 1462-1475.

[8] E. Mansoori, "FRBC: A Fuzzy Rule-Based Clustering Algorithm," IEEE Transactions on Fuzzy Systems, 2011, vol.19(5), pp. 960-971.

[9] B. Costa, P. Angelov and L. Guedes, "Fully unsupervised fault detection and identification based on recursive density estimation and self-evolving cloud-based classifier," Neurocomputing, 2015, vol.150, pp.289-303.

[10] A. Lemos, W. Caminhas and F. Gomide, "Adaptive fault detection and diagnosis using an evolving fuzzy classifier." Information Sciences, 2013, vol.220, pp. 64-85.

[11] P. Angelov and A. Kordon, "Adaptive Inferential Sensors Based on Evolving Fuzzy Models", IEEE Transactions On Systems Man And Cybernetics Part B-Cybernetics, 2010, vol.40(2), pp.529-539

[12] P. Angelov, "A fuzzy controller with evolving structure," Information Sciences, 2004, vol.161(1), pp.21-35.

[13] P. Angelov, X. Gu, J. Principe and D. Kangin, "Empirical data analysis: A new tool for data analytics," *unpublished.*

[14] P. Angelov, "Typicality distribution function- a new density-based data analytics tool" in IEEE International Joint Conference on Neural Networks (IJCNN), Killarney, 2015, pp. 1-8.

[15] https://quantquote.com/historical-stock-data

[16] B. Vincenzo, O. Manca, and S. Nardini, "Electricity consumption forecasting in Italy using linear regression models," Energy, 2009, vol. 34(9), pp. 1413-1421.