

Fuzzy transfer learning in time series forecasting for stock market prices

Shanoli Samui Pal*, Samarjit Kar†

Abstract

Transfer learning involves transferring prior knowledge of solving similar problems in order to achieve quick and efficient solution. The aim of fuzzy transfer learning is to transfer prior knowledge in an imprecise environment. Time series like stock market data are non-linear in nature and movement of stock is uncertain, so it is quite difficult following the stock market and in decision making. In this study, we propose a method to forecast stock market time series in the situation when we can use prior experience to make decisions. Fuzzy transfer learning (FuzzyTL) is based on knowledge transfer in that and adapting rules obtained domain. Three different stock market time series data sets are used for comparative study. It is observed that the effect of knowledge transferring works well together with smoothing of dependent attributes as the stock market data fluctuate with time. Finally, we give an empirical application in Shenzhen stock market with larger data sets to demonstrate the performance of the model. We have explored FuzzyTL in time series prediction to understand the essence of FuzzyTL. We were working on the question of the capability of FuzzyTL in improving prediction accuracy. From the comparisons, it can be said fuzzy transfer learning with smoothing improves prediction accuracy efficiently.

Keywords: Fuzzy transfer learning, time series forecasting, stock market data

1 Introduction

Stock market is unpredictable as there are several complex factors influencing its movements. Therefore the trend of the series is also affected by those factors and by their non-linear relationship. In stock market forecasting, technical analysis is one of the traditional methods applied by investors for decision making. There are some other statistical methods such as autoregressive conditional heteroscedasticity (ARCH) model [1], generalized ARCH (GARCH) model [2], autoregressive moving average (ARMA) model [3], autoregressive integrated moving average (ARIMA) model [3]. All these

models are different types of regression models assuming some mathematical distribution, those distributions are not always followed by realistic stock market time series data.

Nowadays, several data mining approaches like evolutionary algorithms, artificial neural networks, fuzzy logic, rough set theory, and their hybridization have been developed. All these approaches perform well in forecasting of stock markets. Back propagation neural network have been used to find the fuzzy relationship in fuzzy time series [4]. A hybridized genetic algorithm and neural network model has been developed to predict stock price index [5]. Caia [6] proposed a hybrid GA model based on fuzzy time series together with genetic algorithm (FTSGA) working on TAIEX as experimental data set and concluded that the model improved the accuracy. Teoh *et al.*[7] proposed a hybrid model based on multi-order fuzzy time series by using rough sets theory to mine fuzzy logical relationships from time series and an adaptive expectation model to improve forecasting accuracy on TAIEX and National Association of Securities Dealers Automated Quotations (NASDAQ) experimental data sets. Pai and Lin [8] developed a hybrid ARIMA and support vector machine model in stock price forecasting.

There are several pieces of work on fuzzy time series. A novel method on fuzzy time series had been proposed by Wang and Mendel [12] involving fuzzification of real data set, rule generation, rule reduction to reduce rule redundancy, defuzzification to real value, and finally forecasting from the analysis. The model had been applied on truck backer-upper control, Mackey-Glass time series prediction.

Traditional data mining technologies are not capable of handling the information when there is a time gap in their collection periods. Traditional data mining does not consider the domain transfer concept when inferring from information hidden in data. Transfer learning is the technology which works in situations like domain difference. It has been successfully applied to several areas of application like classification problems [13],[14], collaborative filtering problems [15], graph-based method on identifying games [16]. Fuzzy logic based transfer learning is the new technique which works on information with uncertainty, or, in imprecise environment. Shell and Coupland [17] proposed a fuzzy logic based transfer learning (TL) to form a prediction model in Intelligent Environments (IEs). They compared their re-

*Department of Mathematics, NIT Durgapur, Durgapur 713209, West Bengal, India, (email: shanoli.pal@gmail.com)

†Department of Mathematics, NIT Durgapur, Durgapur 713209, West Bengal, India, (email: kar_s.k@yahoo.com)

sult with the method proposed by Wang and Mendel [12].

We are motivated by the concept of transferring knowledge in a similar application domain. We have learnt from the application of FuzzyTL prediction model by [17] in Intelligent Environments, and applied our obtained knowledge in the stock market time series, which is a bit different from IEs where the data sources are sensors. But, in both cases FuzzyTL has been applied for prediction purpose. So, we adapted this concept as the purpose of both the use cases is prediction. The main contributions of this paper are : (1) multi-attributes of stock market data are represented as labeled and unlabeled data, (2) fuzzy transfer learning (FuzzyTL) is yet not explored in the time series forecasting, so it has been explored in time series of stock market price with modification, (3) FuzzyTL has been applied on three different stock market time series for results comparison with the model proposed by Wang and Mendel [12], Chen[18], and Cheng [19], and (4) a case study on Shenzhen stock price with a larger data set.

The remainder of this paper is organized as follows. First, there is a brief discussion on transfer learning in Section 2. In Section 3, we introduce fuzzy transfer learning. Our proposed approach is presented in Section 4. Data sets descriptions are given in Section 5. Results and discussions have been presented in Section 6. A case study on Shenzhen stock price is elaborated in Section 7. In section 8, we have summerized our work and discussed limitations of it. Finally, our conclusions and future research directions are presented in Section 9. The appendix is given thereafter.

2 Basic of transfer learning (TL)

Transfer learning (TL) contains two principle elements, a Domain and a Task. Pan and Yang [9] defined Domain as a pair of two components : a feature space X and a marginal probability distribution $P(X)$ where $X = x_1, x_2, \dots, x_n$ and Task as a pair of two components : a label space $Y = y_1, y_2, \dots, y_n$ and a predictive function $f(\cdot)$. The predictive function can be learned from the training data which is a pair (x_i, y_i) , $x_i \in X$ and $y_i \in Y$. The source domain can be defined as $D_s = (x_{s_1}, y_{s_1}), (x_{s_2}, y_{s_2}), \dots, (x_{s_n}, y_{s_n})$, where $x_s \in X$ is the input data point and $y_s \in Y$ is the corresponding label. The task domain can be defined as $D_t = (x_{t_1}, y_{t_1}), (x_{t_2}, y_{t_2}), \dots, (x_{t_n}, y_{t_n})$, where $x_t \in X$ is the input data point and $y_t \in Y$ is the corresponding output.

Transfer learning (TL) can be defined as : ([9], [10]) Given a source domain D_s and a learning task T_s , a target domain D_t and a learning task T_t , objective of the TL process is to improve the learning of a new task T_t through the transfer of knowledge from previously-acquired knowledge or task T_s by the learning of the predictive function $f_t(\cdot)$ in the target domain D_t where,

$D_s \neq D_t$ or $T_s \neq T_t$.

When the source and target domains are the same ($D_s = D_t$) and their learning tasks are also same ($T_s = T_t$), then the learning problem becomes a traditional machine learning problem. Based on the application point of view, TL techniques can be classified into four categories such as (a) Neural network in transfer learning, (b) Bayesian techniques in transfer learning, (c) fuzzy logic in transfer learning, and (d) transfer learning with evolutionary algorithms [11].

In this work we are focusing on fuzzy transfer learning in stock market time series analysis and prediction. It is discussed briefly in the following section.

3 Fuzzy transfer learning (FuzzyTL)

FuzzyTL is the combination of fuzzy logic (FL) and TL to bridge the knowledge gap by the learning process and an adaptation from the learning process of one context to another. The change of context may be due to change of domain, missing information, change of situation etc. TL has the ability to transfer knowledge from one situation to another. Here, knowledge can be in the form of information.

There are two distinct processes : one is transferring the fuzzy concepts along with their inter relationship and another is the adaptation of the fuzzy concepts.

In the first process, source data are used to develop Fuzzy Inference System (FIS), which consists of fuzzy sets and fuzzy rules. FIS captures the knowledge from the source and is used to transfer it to the target task.

The second process is the adaptation of the FIS. The adaptation process uses the knowledge learned from previous information to the unlabeled task data set. This process adapts the individual components of the FIS to capture the variations in the data. Alterations and variations from situation to situation are absorbed through the changes made within the domains of the fuzzy sets and adaptations of the rulebase.

• Fuzzy framework for producing output:

1. In FuzzyTL, a source domain D_s can be defined as

$$D_s = \{(x_1^s, x_2^s, y^s)\}_s^N \quad (1)$$

where $x_1, x_2 \in X$ are inputs, $y \in Y$ is an output, and N is the number of data tuples in the source domain. The domains of the each input and output are the intervals, i.e., the universe of discourse of that particular domain. Based on this definition, a source domain can be defined using the interval denoted as $D_s^I = \{[x1_L^s, x1_R^s], [x2_L^s, x2_R^s], [y_L^s, y_R^s]\}$.

A domain with fuzzy sets of two inputs and a output is defined as

$$D_s = \{(^f X_1^s, ^f X_2^s, ^f Y^s)\}_s^N \quad (2)$$

where $^f X_1^s$ and $^f X_2^s$ are fuzzy inputs and $^f Y^s$ is the fuzzy output.

The universe of discourse of each input and output is partitioned into equal number of partitions or unequal number of partitions as per the requirement. Membership function of the fuzzy sets can be Gaussian, Triangular etc.

2. Fuzzy rulebase :

Rulebase containing two antecedent and one consequent sets can be constructed as :

$$R = \{^f X_1^r, ^f X_2^r, ^f Y^r\}^P \quad (3)$$

where X_1 and X_2 are the inputs, Y is the output, and P is the number of rules. This rulebase is called exhaustive rulebase when $P = N$.

If N is large, then the number of rules is also large. So, the rulebase is reduced by the following mechanism proposed by Shell and Coupland [17], which will be discussed in next step **Transferring fuzzy concepts**.

After reducing the rulebase, fuzzy sets are defuzzified to get the output of the target domain by using the method proposed by Wang and Mendel [12], given in Appendix.

- 3. Transferring fuzzy concepts :** This is the first stage of generating the FIS (Fuzzy Inference System) of FuzzyTL. This method uses numerical data of labeled data i.e., the pair of input and output to produce the sets and rules which are previously discussed. A rule reduction method is used to reduce the repetition of rules and the impact of anomalous data, and increases the use of information mostly supported by the numerical data. Fuzzy frequency measure [17] is used to form the reduced rulebase. This concept extends the Wang-Mendel method.

In standard Wang-Mendel method [12], the membership values of each data instance are used to decide the strength of a rule. In [17], the FuzzyTL framework uses a frequency value for each rule according to the number of the repetition of their antecedents to capture more information from the exhaustive rulebase. FuzzyTL uses the following fuzzy measure function for each rule :

$$\mu_{Freq}(F) = e^{-\frac{(F_r - c)^2}{2\sigma^2}} \quad (4)$$

where F_r is the frequency of each rule which is the input to the function, c is the value F_{max} that denotes the quantity of highest occurring rule, σ is $\{(F_{max} - F_{min}) \times 0.5\}$, and F_{min} denotes the quantity of lowest occurring rules.

• Adaptation of the fuzzy concepts :

The adaptation of the fuzzy concept consists of the following five stages :

1. External input domain adjustment :

Input domains of the source are adapted according to the target task when a knowledge gap occurs due to the target inputs outside the respective intervals. It is adapted if it is needed as an input domain of the target inputs. The input intervals are adapted if the value extends beyond the left or right boundaries. The boundaries are adjusted based on the data from target domain. Each input is compared with the respective interval, if it is less than the left boundary, then the left boundary is decreased to that input value, if it is greater than the right boundary, then right boundary is increased to that input value.

2. Internal input domain adjustment :

This work also focuses on input domain adaptation when the knowledge gap occurs due to fully or partially overlapped intervals of source and target tasks. It also works on input domains. Source input domains are transferred to adapt the knowledge gap according to target the input domains. The whole process are described below in few steps.

Step 1: The target input intervals are compared to the source input intervals $[x_l, x_r]$.

Step 2: The adaptation procedure uses the local minimum and maximum of the target values to compare with the source values.

Step 3: If one, or both of these values lie within the interval that is represented by the source values x_l and x_r , a proximity measure [17] is calculated to decide whether the domain is able to adapt or not. The measure taken is assigning a membership function based on the source input domain interval.

Step 4: Here, in this procedure, a threshold value is considered above which the membership function of those intervals is adapted. Otherwise, the local minimum and maximum are accepted as lower and upper end points of target values.

3. Output domain adaptation :

The adaptation of the output generated from

the framework is itself based upon the target domain. Gradient control [17] mechanism is used to adapt the knowledge gap in target consequent sets. There are few steps described below.

Step 1: m sized sliding window data of source and target domain are collected for each input $x \in X$ and output $y \in Y$. The output values for target domain are generated beforehand by the framework of the fuzzy set itself using Wang-Mendel [12] defuzzification procedure.

Step 2: Gradients are calculated for each input and output of source data set, taking the mean and standard deviation of those elements within n window size. The gradients are obtained by normalization based on the standard score method defined as

$$z = \frac{x - \bar{x}}{\sigma} \quad (5)$$

where z is the gradient of input or output particular x , \bar{x} is the mean, and σ is the standard deviation of the sliding window.

Step 3: Gradients are compared with each other at each individual input value.

Step 4: The differences within the source and target gradients show the knowledge gaps which are required to adapt for improvising the unlabeled data. The consequent adaptation can be expressed as

$$dD_a = \phi \sum_{i=1}^n (g_{s_i} - g_{t_i}) \quad (6)$$

where dD_a denotes the delta to adapt the consequent sets, ϕ is a learning parameter that can be user defined [17], g_s and g_t are the gradients of source and target data for sliding window size n respectively.

Positive differences between source target output gradients produce a reduction in the domain, whereas negative differences initiate the enlargement of the domain.

The previous three stages consider the domain adaptation of the fuzzy sets. The following stages describe the adaptation of the fuzzy rulebase.

4. Rulebase modification according to the adaptation rulebase :

Exhaustive rulebase is used to produce an adaptive rulebase. Reduced rulebase and adaptive rulebase are compared to get the final rulebase. Reduced rules are examined and applied to the target domain data to

check their applicability. The exhaustive rulebase is evaluated iteratively to find the most used rules with greater weighting, which shows greater applicability of those rules within the target domain. The steps are listed given below.

Step 1: Examine the exhaustive rulebase to identify the rules those fire using the target input data.

Step 2: Rules those fire with the highest membership value from each data point are kept as adaptive rulebase.

Step 3: The adaptive rulebase is compared to the reduced rulebase. The better is kept in adaptive rulebase. Any rule which is in reduced rulebase but not in adaptive rulebase, then that rule is added in the adaptive rulebase.

5. Rule adaptation using Euclidean distance measure :

The knowledge gap is not fully covered by previously learned information, it needs new information to remove incompleteness. To do so, the formation of antecedent sets based on domain adaptation, described in previous stages, is required to move input domains of the target towards the true state. In the case of the consequent set, the euclidean distance based on the source input values against the target input values is evaluated. The source output corresponding to the closest source input values represent the target output of the corresponding target inputs, for which Euclidean distance measure is calculated.

4 Proposed approach

Transfer learning (TL) can transfer knowledge in the form of information gained through past experience to utilize the same in the similar or sort of similar domains of problem. In this paper, FuzzyTL is applied with modification to study the stock market data where we use a set of labeled data to acquire knowledge and the gained knowledge is used on unlabeled data prediction. Stock market data has basic indexes high price, low price, opening price, closing price, volume etc., among them, high and low price represents input pair whereas closing price represents output. In case of labeled data, the mentioned inputs and output are known. As example, 241 data points are used as labeled data and 84 data points are used as unlabeled data of the BSE stock market where inputs high and low price are known, but it is assumed that output closing prices are unknown. Another important thing is that, as current time point data are highly dependent on just the previous histori-

cal data points, we have used here adaptive forecasting. Adaptive forecasting is a smoothing technique with the just the previous history along with a smoothing factor.

Labeled data set is fuzzified to form the framework of FuzzyTL which is previously mentioned. Then n number of rules in an exhaustive rulebase are formed based on maximum membership in their respective input or output region where there are n number data points are considered in labeled data. To reduce the exhaustive rulebase, the method in section **Transferring fuzzy concepts** is used. Then the Wang-Mendel defuzzification method is used to get the output for unlabeled data set.

To transfer the knowledge present in the domain as well as in rulebase all the steps like, External input domain adjustment, Internal input domain adjustment, Output domain adaptation according to the adaptation rulebase, discussed previously, are utilized. To transfer knowledge from the rulebase, we have followed separate steps, which are not similar to the forth and fifth stages described in previous section.

4.1 Adapted rule reduction :

In this section, we describe the modification of FuzzyTL for our approach. The output values for the target domain are generated beforehand by the fuzzy sets framework itself using Wang-Mendel [12] defuzzification procedure. Now, using the target inputs we find the universe of discourse of the target output and we partition the universe into some region. Gaussian memberships are calculated in each region for each target output. Maximum membership gives the inclusion of each target value in a particular region and the rules are generated. There are some repetitions in the rules. So, rules with the same antecedent and consequent fuzzy sets form a group and represented by one rule. Using those combined rules and Wang-Mendel defuzzification method, we get the final target output values.

Another methodology using smoothing factor on just previous output, the current output is calculated. This is called adaptive forecasting. The equation is given below.

$$F_t = O_{t-1} + \alpha \times (defuzzF_t - O_{t-1}) \quad (7)$$

where F_t is the predicted value at t^{th} time point, O_{t-1} represents closing price value at $(t-1)^{\text{th}}$ time point as it is output in our work, α is the smoothing factor where $0 < \alpha < 1$, and $defuzzF_t$ is the defuzzified value of t^{th} data point.

5 Data sets descriptions

Stock market data are multi-attribute in nature and stock market prediction depends on several factors so it

Table 1: Maximum and minimum of labeled BSE data

	Maximum	Minimum
input1	28822.369141	20255.519531
input2	28538.439453	19963.119141
output	28693.990234	20193.349609

is quite difficult to build forecasting models for itself. As there are several factors for stock exchange ups and downs, it is non-linear in pattern. In our work, basic indices of stock exchange time series data such as high price, low price and closing price are used to build up the models. Time series data from three different stock exchanges are collected as labeled and unlabeled data sets.

- **Bombay stock exchange (BSE):**

BSE's popular equity index - the S&P BSE SENSEX- is India's mostly tracked stock market benchmark index. There are several types of work i.e., stock market analysis, forecasting models on BSE data [20], [21]. In our work, labeled data sets are collected daily from Jan to Dec, 2014 (241 data points) and unlabeled data sets are collected daily from Jan to April, 2015 (84 data points) [24].

- **New York stock exchange (NYSE):**

NYSE is American stock exchange, hugely analysed in several research works [22], [23]. In our work, labeled data set is collected daily from Jan to Dec, 2014 (252 data points) and unlabeled data set collected daily from Jan to April, 2015 (82 data points) are used [25].

- **Taiwan stock exchange corporation (TAIEX):**

TAIEX is Taiwan stock exchange benchmark index. It is also hugely used in several analyses [6], [7]. In our work, labeled data set is collected daily from Jan to Dec, 2014 (247 data points) and unlabeled data set collected daily from Jan to April, 2015 (81 data points) are used [26].

6 Result and discussion

Stock indices of high and low price are used as the pair of input and closing price as output. Number of labeled data and unlabeled data are described in section 5 for each BSE, NYSE and TAIEX time series.

Here, each input and output data set are partitioned into 10 overlapping regions, which are represented by 10 linguistic variables. The maximum and minimum

are used to decide the universe of discourse (UOD). Table 1, Table 2, Table 3, and Table 4 show the details for the BSE time series. Similar details have been obtained for NYSE and TAIEX time series. Table 1 shows the maximum and minimum for high price (input1), low price (input2), and closing price (output) of BSE. UOD is [19000, 29000] for each input1, input2, and output. Gaussian membership is calculated for each data point in each region. Based on the maximum value linguistic is assigned to the respective data point. Then, an exhaustive rulebase is obtained.

To reduce repeating the rules, the method proposed by Shell and Coupland [17] is followed and we get the number of groups and the members in each group which is shown in Table 2. Table 3 depicts the reduced rulebase. Input domains are adopted using the first three stages of adaptation of the fuzzy concepts method described in section 3.

Then, Wang-Mendel defuzzification method is used to get the target output from FuzzyTL framework. The adapted target input domain and the target output originated from FuzzyTL are used to produce a new set of rulebase for target data set. First, UODs are found out, then those are partitioned into 10 regions. A Gaussian membership is assigned to each data point with respect to each region. The rulebase is generated based on the maximum membership. Then, the repeated rules are reduced to get the final set of rulebase. Table 4 shows the number of groups and the members of each group for the target data set. Finally, Wang-Mendel defuzzification is used to get the final target outputs.

It is well known that, the movement of current stock market data highly depends on its previous historical data. So, we have tried smoothing techniques, the expressions which is given in equation (7). The result comparisons of three stock market data are given in Table 5. Root mean square errors (RMSEs) of proposed model together with Wang-Mendel method [12], Chen [18] model, and Cheng [19] models have been presented. In Table 5, the columns with Wang-Mendel (WM) method and FuzzyTL with WM are the existing approaches. Our novelty lies in applying the existing FuzzyTL with WM existing method in stock market data and adding smoothing coefficients with the existing method. Comparison exhibiting an improved accuracy as compared to the results obtained in existing methods given in Table 5. RMSE on FuzzyTL with smoothing factors $\alpha = 0.001$ and $\alpha = 0.01$ are also depicted in Table 5. From the result, it is observed that FuzzyTL performs better than the Wang-Mendel [12], Chen [18], and Cheng [19] methods in learning and transferring experience from past information to uncover better future inferences.

Figure 1, Figure 2, and Figure 3 respectively show the actual and forecasted stock market time series of BSE, NYSE and TAIEX for unlabeled data.

The time complexity of an algorithm depends on

the execution time of each statement of the code. In FuzzyTL code, as well as, the Wang-Mendel code, for time series prediction, we have used a block of codes where the exhaustive rulebase is generated for source domain. In case of transfer learning, number of unlabeled data is less than number of labeled data. Labeled data are considered as source domain and unlabeled data as target domain. In that block of codes, we need to use the for-loop N^2 -times, where N is the number of source data points. So, time complexity of the Wang-Mendel code is $\mathcal{O}(N^2)$. But, FuzzyTL takes little bit more time comparison to Wang-Mendel method as the same block of code is used for target domain. If M is the number of target data points, then time complexity of FuzzyTL is $\mathcal{O}(N^2 + M^2)$ where $N > M$.

7 Case study on Shenzhen stock prices

Economic conditions in China hugely depend on stock price movements. Shenzhen stock exchange is one of the three stock exchanges in China. There are several existing works on Shenzhen stock exchange [27], [28]. Opening, high, low, and closing prices of 30 stocks are collected and analyzed. There are 573 data points of each stock from 11th April, 2004 to 3rd Dec, 2015. Among these long period time series data points, we have considered 80% as labeled data and remaining as unlabeled data. Deciding the domains or UODs of 20% target data points are difficult in case of some Shenzhen stock prices because the time series have sudden ups and downs. The domains are adapted using first step **External input domain adjustment** discussed in adaptation of the fuzzy concepts method described in section 3. The output values for target domain are obtained by adapted Wang-Mendel method [12] after generating the set of reduced rules from labeled data. Some of the stock's original and predicted movement on unlabeled data are shown in Figure 4. RMSE comparison between Wang-Mendel method [12] and FuzzyTL are shown in Table 6. It is observed that the forecasting model by using FuzzyTL performs better in most of the cases in comparison to the model by Wang-Mendel [12]. Bold face fonts shows the better results in Table 6.

7.1 Statistical significant test

In this subsection, we elaborate the non-parametric statistical test Wilcoxon signed-ranks test to test the significant difference between two methods: (1) WM method, and (2) FuzzyTL in time series forecasting. This statistical test examines the null-hypothesis that both the methodologies perform equally well. Data sets on 29 stocks of the Shenzhen stock exchange are used to observe the performance of the two mentioned methodologies. RMSEs are tabulated in Table 6 for the actual and

Table 2: Exhaustive rulebase for labeled BSE data (Total no of group 25)

Gr.	Members
0	0, 9, 10, 11, 12, 13, 14, 15, 16, 17, 44, 45, 46, 49, 50, 51, 53, 54, 55
1	1, 40, 42
2	2, 3, 4, 5, 6, 7, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39
3	8, 41, 43
4	24
5	47, 48, 52
6	56, 57
7	58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84
8	85
9	86, 88, 89
10	87
11	90
12	91, 92, 93, 94, 95, 97, 98, 99, 100, 101, 102, 103
13	96
14	104, 116, 130, 131
15	105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 117, 118, 119, 120, 121, 122, 123, 124, 125, 128, 129, 132, 133, 134, 135, 136, 144, 145, 146, 147, 148, 149, 150, 151, 152
16	126, 137, 142, 153, 193
17	127, 141, 143, 192
18	138, 139, 140, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 173, 174, 175, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 194, 195, 196, 197, 198, 199, 233
19	164, 166, 167, 172, 176, 178, 200, 234
20	165, 168, 169, 170, 177, 201, 202, 203, 205, 207, 228, 229, 230, 231, 235, 236, 237, 238, 239, 240
21	171, 179, 232
22	204, 206, 209, 227
23	208, 210, 211, 213, 214
24	212, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226

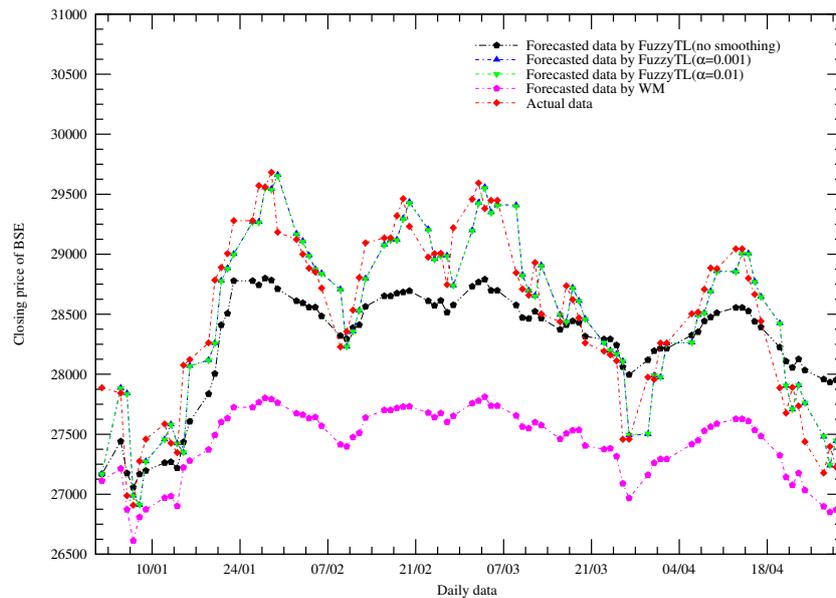


Figure 1: BSE unlabeled data from Jan to April, 2015

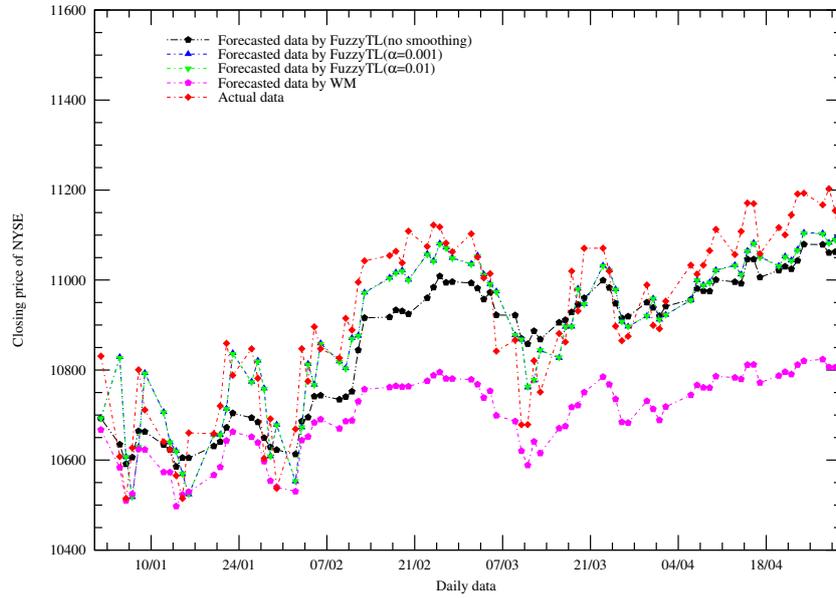


Figure 2: NYSE unlabeled data from Jan to April, 2015

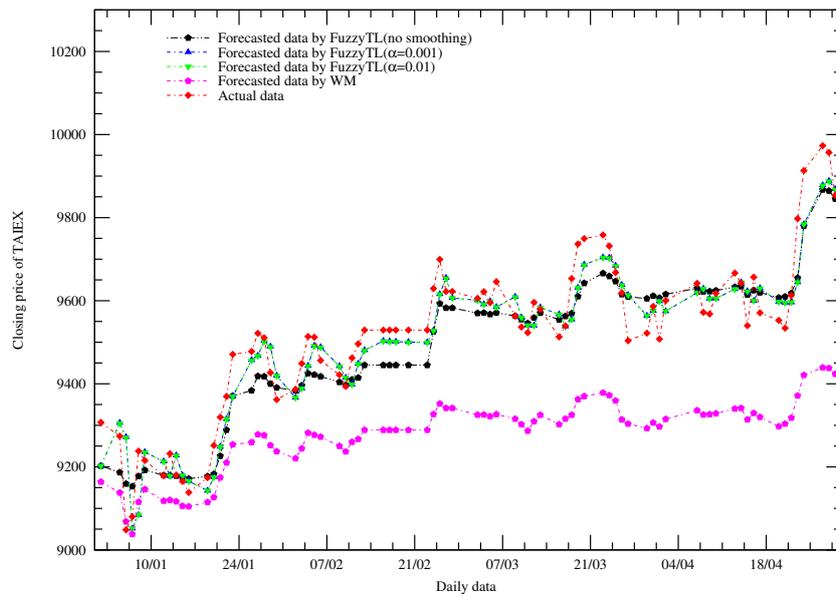


Figure 3: TAIEX unlabeled data from Jan to April, 2015

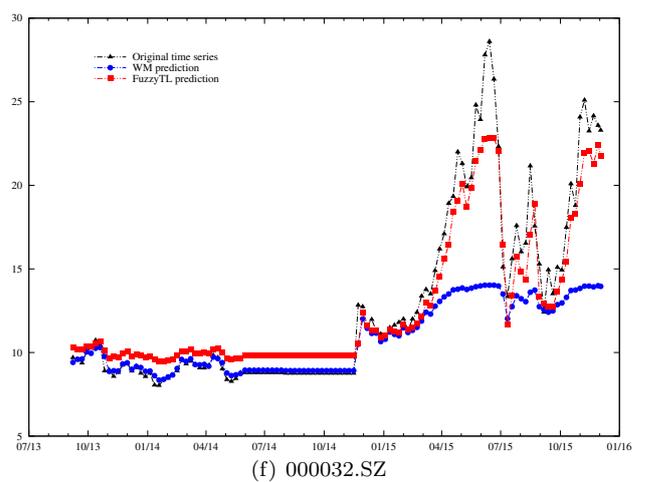
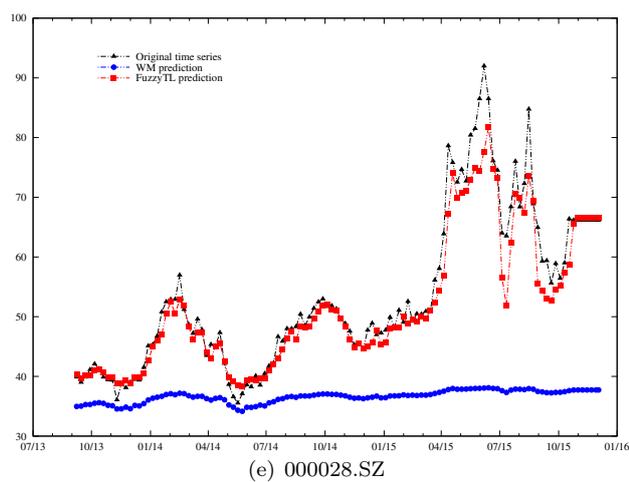
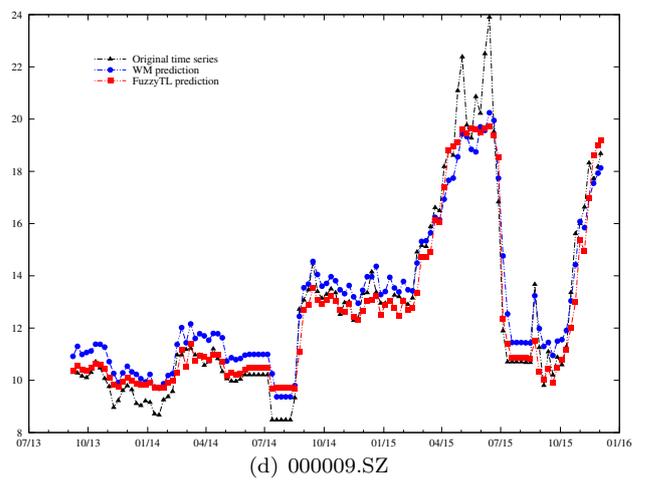
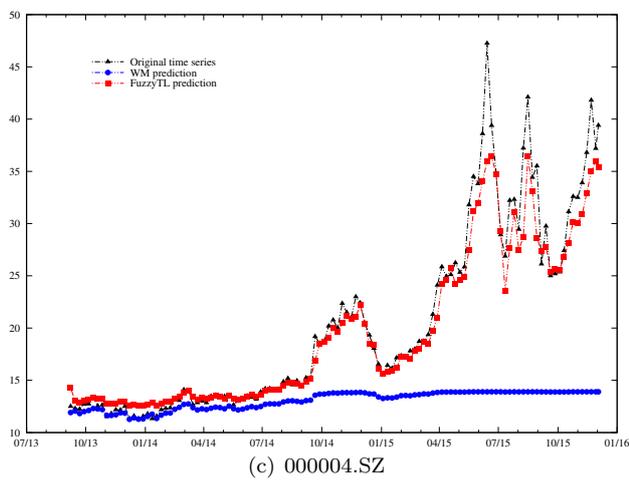
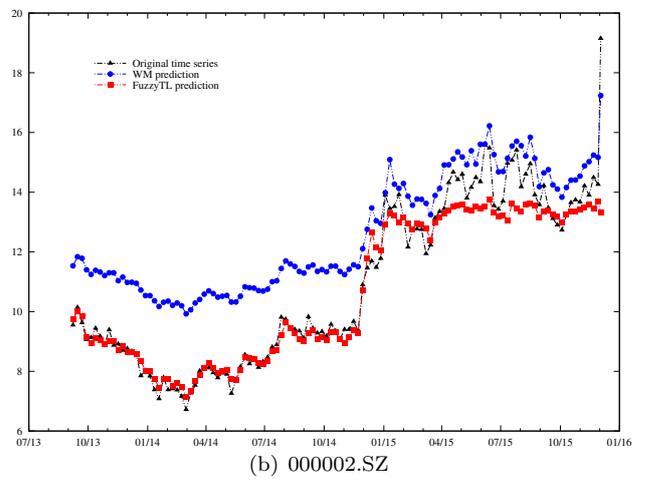
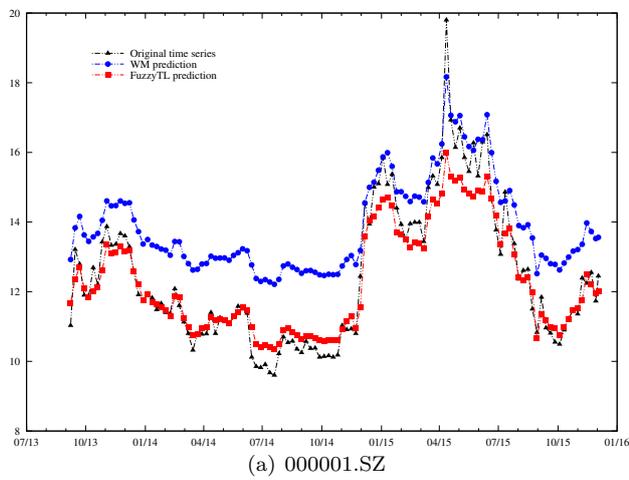


Figure 4: Original and forecasted unlabeled closing price data of Shenzhen stock exchange from 8th Sept, 2013 to 3rd Dec, 2015

Table 3: Reduced rulebase of labeled BSE data

If x_1 is l_3 and x_2 is l_3 then y is l_3
If x_1 is l_3 and x_2 is l_2 then y is l_2
If x_1 is l_2 and x_2 is l_2 then y is l_2
If x_1 is l_3 and x_2 is l_2 then y is l_3
If x_1 is l_2 and x_2 is l_1 then y is l_2
If x_1 is l_4 and x_2 is l_3 then y is l_3
If x_1 is l_4 and x_2 is l_3 then y is l_4
If x_1 is l_4 and x_2 is l_4 then y is l_4
If x_1 is l_5 and x_2 is l_4 then y is l_4
If x_1 is l_5 and x_2 is l_5 then y is l_5
If x_1 is l_6 and x_2 is l_5 then y is l_5
If x_1 is l_7 and x_2 is l_5 then y is l_6
If x_1 is l_6 and x_2 is l_6 then y is l_6
If x_1 is l_7 and x_2 is l_6 then y is l_6
If x_1 is l_7 and x_2 is l_6 then y is l_7
If x_1 is l_7 and x_2 is l_7 then y is l_7
If x_1 is l_8 and x_2 is l_7 then y is l_8
If x_1 is l_8 and x_2 is l_7 then y is l_7
If x_1 is l_8 and x_2 is l_8 then y is l_8
If x_1 is l_9 and x_2 is l_8 then y is l_9
If x_1 is l_9 and x_2 is l_9 then y is l_9
If x_1 is l_9 and x_2 is l_8 then y is l_8
If x_1 is l_{10} and x_2 is l_9 then y is l_9
If x_1 is l_{10} and x_2 is l_9 then y is l_{10}
If x_1 is l_{10} and x_2 is l_{10} then y is l_{10}

predicted values using two methodologies. In Table 6, differences, absolute differences, and ranks based on differences are also tabulated. Exact critical value, say T is calculated as $\min(R^+, R^-)$, where R^+ is the sum of the ranks for the data sets on which second methodology outperformed the first, and R^- is the sum of the ranks for the others. For larger number of data sets, if the statistics

$$z = \frac{T - \frac{1}{4}N(N+1)}{\sqrt{\frac{1}{24}N(N+1)(2N+1)}} \quad (8)$$

is smaller than -1.96 with confidence level $\alpha = 0.05$, then null-hypothesis can be rejected where N is the number of data sets. Here, in our case, $T = \min(R^+, R^-) = \min(401, 34) = 34$, $N = 29$, $z = -3.96785$. So, we can conclude that null-hypothesis can be rejected and the two methodologies have significant difference.

8 Summarization and limitation

Our motivation for this work is to implement fuzzy transfer learning concepts in time series prediction. For domain transfer concept we have used multi-attributed time series. To adapt the transferred knowledge, domain of discourses are adjusted and then rule base have

been generated. The proposed approach has been implemented on stock market time series and result comparisons have been provided considering traditional fuzzy time series models. So, main advantage of this work is to implement transfer learning in time series prediction.

There are two main steps in FuzzyTL: domain adjustment and rule base generation. For our model building, we have considered the unlabeled data as labeled data. In creating an adapted rule base, we have used Wang-Mendel [12] defuzzification procedure as we are dealing with unlabeled data. If we apply the FuzzyTL in predicting proper labeled data, classification algorithms can be implemented to get the predicting attributes. Using proper labeled data and deciding predicted attributes for reduced rule base have not been considered in our work, which may be considered as limitations of this work.

9 Conclusion

A novel FuzzyTL for time series forecasting has been proposed in this paper. The concept of learning from labeled data and transferring the knowledge (or experience) to make a decision on unlabeled data is applied to some extent. Proposed approach has advantage as well as some limitations due to the use of unlabeled data. It was observed in previous pieces of work that the current time point value is highly dependent on just the previous or some of the previous time point values on time series analysis. So, here we have implemented this observation as it is true in case of stock market data. Wang-Mendel method [12] used this characteristic of time series analysis to predict Mackey-Glass time series. Here, we have considered the same concept on stock price time series prediction. It is observed that FuzzyTL with smoothing and without smoothing work better in comparison to Wang-Mendel method [12]. Also, it is performing well in comparison to some traditional fuzzy time series models, where transfer learning concept is not considered.

In future, we can apply all the five stages (discussed in section 3) of TL adaptation for better performance. Also rule reduction can be improved by using rough set theory.

Appendix

10 Wang-Mendel defuzzification method :

After finding the exhaustive rulebase, weighting is assigned to each rule, based on that and similar antecedents of rules, a reduced rulebase or combined fuzzy rulebase is generated. First, for a given input (x_1, x_2) , we combine the antecedents of the i^{th} fuzzy rule using product operation to decide the degree $d_{O_i}^i$ of the output

Table 4: Adapted reduced rulebase for unlabeled BSE data (Total no of group 13)

Group	Members
0	0
1	1, 11, 12
2	2, 27, 56, 57, 58, 59, 62, 63, 64, 65, 76, 78, 79
3	3, 4
4	5, 6, 7, 8, 60, 61, 77, 80, 81, 82
5	9
6	10
7	13, 14, 24, 30, 40, 70
8	15, 16, 21, 22, 23, 36, 37, 38, 39, 46, 50, 71, 72
9	17, 31, 32, 33, 34, 35, 41, 42, 44, 45
10	18, 19, 20, 43
11	25, 26, 28, 29, 47, 48, 49, 51, 52, 53, 54, 55, 66, 67, 68, 69, 73, 74, 75
12	83

Table 5: Root mean square error (RMSE) for unlabeled BSE, NYSE, TAIEX data

Time series	Wang-Mendel (WM) method	Chen Model	Cheng Model	FuzzyTL		
				$\alpha=0.001$	$\alpha =0.01$	No smoothing(FuzzyTL WM)
BSE	1150.3713	688.9020	387.6920	274.0837	273.4513	457.1822
NYSE	240.6533	103.7715	103.9815	86.0553	86.0666	98.6410
TAIEX	263.9279	88.8127	86.4081	62.7620	62.7070	68.2226

y corresponding to (x_1, x_2) , i.e.,

$$d_{O_i}^i = m_{I_1}^i(x_1)m_{I_2}^i(x_2) \quad (9)$$

where O^i denotes the output region of the i^{th} rule, $m_{I_1}^i(x_1)$ denotes the membership of the input x_1 at I_1^i region of the i^{th} rule and $m_{I_2}^i(x_2)$ denotes the membership of the input x_2 at I_2^i region of the i^{th} rule. The centroid defuzzification formula to map the output in crisp value is given following:

$$y = \frac{\sum_{i=1}^r d_{O_i}^i \bar{y}^i}{\sum_{i=1}^r d_{O_i}^i} \quad (10)$$

where y is the crisp output, \bar{y}^i is the center value of the region O^i of the i^{th} rule and r is the number of combined rules.

Acknowledgement

The first author would like to thank DST INSPIRE, India for their help and supports to sustain the work.

References

[1] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimator of the variance of

United Kingdom inflation," *Econometrica*, vol. 50, pp. 987-1008, 1982.

- [2] T. Bollerslev, "Generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, vol. 31, pp. 307-327, 1986.
- [3] G. Box and G. Jenkins, "Time Series Analysis: Forecasting and Control," Holden-Day, San Francisco, 2nd ed., 1976.
- [4] K. Hurang and H. K. Yu, "The application of neural networks to forecast fuzzy time series," *Physica A*, vol. 363, pp. 481-491, 2006.
- [5] C. Nikolopoulos and P. Fellrath, "A hybrid expert system for investment advising," *Expert Systems*, vol. 11, pp. 245-250, 1994.
- [6] Q. Caia, D. Zhanga, B. Wua, and S. C. H. Leung, "A novel stock forecasting model based on fuzzy time series and genetic algorithm," *Proc. Computer Science*, vol. 18, pp. 1155-1162, 2013.
- [7] H. J. Teoh, T. L. Chen, C. H. Cheng, and H. H. Chu, "A hybrid multi-order fuzzy time series for forecasting stock markets," *Expert Systems with Applications*, vol. 36, pp. 7888-7897, 2009.

Table 6: Root mean square error (RMSE) for unlabeled closing price data of Shenzhen stock exchange

Stocks	WM method	FuzzyTL	Differences	Absolute differences	Ranks
000001.SZ	1.611087	0.766559	+0.844528	0.844528	13
000002.SZ	1.897229	0.688697	+1.208532	1.208532	16
000004.SZ	10.978931	2.358990	+8.619941	8.619941	26
000005.SZ	1.305526	0.963013	+0.342513	0.342513	11
000006.SZ	3.287760	0.914095	+2.373665	2.373665	18
000007.SZ	3.005471	1.105950	+1.899521	1.899521	17
000008.SZ	7.185155	2.168828	+5.016327	5.016327	23
000009.SZ	1.004225	1.054493	-0.050268	0.050268	3
000010.SZ	0.771558	0.957016	-0.185458	0.185458	9
000011.SZ	1.104852	1.012296	+0.092556	0.092556	6
000012.SZ	1.111374	0.613068	+0.498306	0.498306	12
000014.SZ	1.971918	1.897595	+0.074323	0.074323	4
000016.SZ	6.724792	1.543654	+5.181138	5.181138	24
000017.SZ	4.210160	1.300020	+2.91014	2.91014	21
000018.SZ	12.644798	2.543516	+10.101282	10.101282	27
000019.SZ	1.002364	1.134859	-0.132495	0.132495	7
000020.SZ	5.634376	1.769913	+3.864463	3.864463	22
000021.SZ	0.798732	0.886530	-0.087798	0.087798	5
000022.SZ	0.955108	0.948929	+0.006179	0.006179	1
000023.SZ	9.151912	1.645808	+7.506104	7.506104	25
000024.SZ	1.397051	1.352104	+0.044947	0.044947	2
000025.SZ	16.454901	5.001511	+11.45339	11.45339	28
000026.SZ	1.268778	1.119854	+0.148924	0.148924	8
000027.SZ	2.058834	1.183786	+0.875048	0.875048	14
000028.SZ	20.620278	3.607171	+17.013107	17.013107	29
000029.SZ	1.864301	0.844695	+1.019606	1.019606	15
000030.SZ	0.518349	0.811959	-0.29361	0.29361	10
000031.SZ	3.684534	1.178407	+2.506127	2.506127	19
000032.SZ	4.094317	1.455326	+2.638991	2.638991	20

- [8] P. F. Pai and C. S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, vol. 33, pp. 497-505, 2005.
- [9] S. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1345-1359, 2009.
- [10] L. Torrey and J. Shavlik, "Transfer learning," *Handbook of Research on Machine Learning Applications*, vol. 3, IGI Global 3, 2009, pp. 17-35.
- [11] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: A survey," *Knowledge-Based Systems*, vol. 80, pp. 14-23, 2015.
- [12] L. X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, pp. 1414-1427, 1992.
- [13] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, Bollywood, Boom-Boxes and Bleders: Domain adaptation for sentiment classification," *Proc. 45th Ann. Meeting of the Assoc. Computational Linguistics*, pp. 432-439, 2007.
- [14] P. Wu and T. G. Dietterich, "Improving SVM accuracy by training on auxiliary data sources," *Proc. 21st Int'l Conf. Machine Learning*, pp.110, July 2004.
- [15] B. Li, Q. Yang, and X. Xue, "Transfer learning for collaborative filtering via a rating-matrix generative model," *Proc. 26th Int'l Conf. Machine Learning*, June 2009.
- [16] G. Kuhlmann and P. Stone, "Graph-based domain mapping for transfer learning in general games," *Proc. 18th European Conf. Machine Learning*, pp. 188-200, Sept. 2007.
- [17] J. Shell and S. Coupland, "Fuzzy transfer learning: Methodology and application," *Information Sciences*, vol. 293, pp. 59-79, 2015.
- [18] S. -M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets System*, vol. 81, no. 3, pp. 311-319, 1996.
- [19] C. -H. Cheng, Y.-S. Chen, and Y.-L. Wu, "Forecasting innovation diffusion of products using trend-weighted fuzzy time-series model," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1826-1832, 2009.
- [20] S. S. Gangwar and S. Kumar, "Probabilistic and Intuitionistic Fuzzy Sets-Based Method for Fuzzy Time Series Forecasting," *Cybernetics and Systems: An International Journal*, vol. 45, pp. 349-361, 2014.
- [21] B. P. Joshi and S. Kumar, "Intuitionistic Fuzzy Sets Based Method for Fuzzy Time Series Forecasting," *Cybernetics and Systems: An International Journal*, vol. 43, pp. 34-47, 2012.
- [22] W. Leigh, R. Purvis, and J. M. Ragusa, "Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support," *Decision Support Systems*, vol. 32, pp. 361-377, 2002.
- [23] C. W. J. Granger, "Forecasting stock market prices: Lessons for forecasters," *International Journal of Forecasting*, vol. 8, pp. 3-13, 1992.
- [24] "BSE data set",
<https://in.finance.yahoo.com/q/hp?s=BSESN>.
- [25] "NYSE data set",
<http://finance.yahoo.com/q/hp?s=N%26Y%26A%26Historical%26Prices>.
- [26] "TAIEX data set",
<http://finance.yahoo.com/q/hp?s=T%26W%26I%26Historical%26Prices>.
- [27] R. Zhou, Y. Zhan, R. Cai, and G. Tong, "A mean-variance hybrid-entropy model for portfolio selection with fuzzy returns," *Entropy*, vol. 17, pp. 3319-3331, 2015.
- [28] Y. Lin and Y. Yang, "Stock markets forecasting based on fuzzy time series model," *IEEE Int'l Conf. Intelligent Computing and Intelligent Systems*, pp. 782-786, Nov. 2009.