Wrapper ANFIS-ICA method to do stock market timing and feature selection on the basis of Japanese Candlestick

Sasan Barak\textsuperscript{a,}\textsuperscript{*}, Jalil Heidary Dahooie\textsuperscript{b}, Tomáš Tichý\textsuperscript{a}

\textsuperscript{a} Faculty of Economics, Technical University of Ostrava, Ostrava, Czech Republic
\textsuperscript{b} Faculty of Management, University of Tehran, Tehran, Iran

A R T I C L E   I N F O

Keywords:
Finance
Stock market forecasting
Technical analysis
Feature selection
Wrapper ANFIS-ICA

A B S T R A C T

Predicting stock prices is an important objective in the financial world. This paper presents a novel forecasting model for stock markets on the basis of the wrapper ANFIS (Adaptive Neural Fuzzy Inference System)-ICA (Imperialist Competitive Algorithm) and technical analysis of Japanese Candlestick. Two approaches of Raw-based and Signal-based are devised to extract the model’s input variables with 15 and 24 features, respectively. The correct predictions percentages for periods of 1–6 days with the total number of buy and sell signals are considered as output variables. In proposed model, the ANFIS prediction results are used as a cost function of wrapper model and ICA is used to select the most appropriate features. This novel combination of feature selection not only takes advantage of ICA optimization swiftness, but also the ANFIS prediction accuracy. The emitted buy and sell signals of the model revealed that Signal databases approach gets better result with 87% prediction accuracy and the wrapper features selection obtains 12% improvement in predictive performance regarding to the base study. In addition, since the wrapper-based feature selection models are considerably more time-consuming, our presented wrapper ANFIS-ICA algorithm’s results have superiority in time decreasing as well as increasing prediction accuracy as compared to other algorithms such as wrapper Genetic algorithm (GA).

\textcopyright 2015 Elsevier Ltd. All rights reserved.

1. Introduction

A good prediction contributes to better decision-making and planning for the future while it allows the current variables to create a considerable value. Predictions accuracy affects the organizations financial stability (Asadi, Hadavandi, Mehmankazir, & Nakhostin, 2012). In the industry of information processing, data mining is an ever growing technology and has been applied in engineering, business and management. Prediction of stock prices, credit scores, and even bankruptcy potentials are examples of significant applicability of data mining in the field of finance. Since financial markets are complex and non-linear dynamic systems, their predictions are really challenging (Huang & Tsai, 2009). There are different forecasting approaches such as statistical models, technical analysis (TA), and econometric methods (Atsalakis, Dimitrakakis, & Zopounidis, 2011; Kar, Das, & Ghosh, 2014) while in this study the focus is on wrapper approach based on TA.

The most crucial points for trading are peaks and bottoms of the price trend in the way that a professional and lucky practitioner sells at peaks and buys at bottoms. In real world, TA gives a considerable assistance in such decisions (Lin, Yang, & Song, 2011) and its application has been well conceived by the financial market experts (Menkhoff, 2010). It should be noted that for short term period prediction, TA is the best tool although lately it has gone beyond the intention of short-term trading (Cervelló-Royo, Guijarro, & Michniuk, 2015; Zhu & Zhou, 2009).

Duetothediscussedimportance,overthelasttwodecadesalotof researches have focused on development of intelligent soft computing models (Bisoi & Dash, 2014; Hafezi, Shahrabi, & Hadavandi, 2015; Majhi, Panda, & Sahoo, 2009). Most of them combine soft computing techniques and TA with stock analysis such as (Chen, Mabu, Shimada, & Hirasawa, 2009; Patel, Shah, Thakkar, & Kotecha, 2015; Wen, Yang, Song, & Jia, 2010).

Due to the discussed importance, over the last two decades a lot of researches have focused on development of intelligent soft computing models (Bisoi & Dash, 2014; Hafezi, Shahrabi, & Hadavandi, 2015; Majhi, Panda, & Sahoo, 2009). Most of them combine soft computing techniques and TA with stock analysis such as (Chen, Mabu, Shimada, & Hirasawa, 2009; Patel, Shah, Thakkar, & Kotecha, 2015; Wen, Yang, Song, & Jia, 2010).

Since the end of the 1980s, some financial scholars applied artificial neural networks (ANNs) to predict the stock market (Lee & Jo, 1999). However, the ANNs have some shortcomings in patterns learning because stock market data are noisy and with complex dimensions (Huang, Yang, & Chuang, 2008). Therefore, to deal with these challenges, scholars proposed combination of fuzzy theory and ANNs.

A Takagi–Sugeno–Kang-type Neuro-fuzzy rule-based system to forecast Taiwan Stock Exchange price deviation is developed by Chang and Liu (2008) with accuracies of 97.6% and 98.08% in...
A hybrid Neuro-fuzzy with ANFIS (Adaptive Neural Fuzzy Inference System) is developed by Yunos, Shamsuddin, and Sallehuddin (2008) to predict daily movements of the Kuala Lumpur Composite Index (KLCI) and the results show superiority of the ANFIS to ANN. In comparison with the others, ANFIS has high speed of training, the most effective learning algorithm, and is simple in terms of the structure (Sarkheyli, Zain, & Sharif, 2015). ANFIS provides better results when applied without any pre-training (Vairappan, Tamura, Gao, & Tang, 2009), Bagheri, Mohammadi Peyhani, and Akbari, (2014) used an ANFIS with QPSO (Quantum-behaved Particle Swarm Optimization) hybrid method for financial forecasting. For the scholars from this field the review paper sof Atsalakis and Valavanis (2009) and Kar et al. (2014) can be useful.

The contribution of the paper is summarized as follows:

- Gathering a comprehensive Japanese Candlestick database regarding to new adaptive technical analysis.
- Developing a hybrid feature selection algorithm on the basis of wrapper ICA-ANFIS model.
- Exploring how much efficient an ANFIS would be to infer the TA clues if it is provided with important indices of the TA (based on wrapper feature selection) and actual trends.

The rest of the paper is organized as follows: Section 2 reviews the literature of wrapper, ANFIS, and ICA; Section 3 introduces the model; Section 4 presents the experimental results and finally Section 5 focuses on the conclusions.

2. The background

2.1. Wrapper feature selection method

In the wrapper method the goal is to find a subset of size $r$ from $n$ variables ($r < n$) that maximizes the predictor performance (Maldonado & Weber, 2009). The method utilizes the learning mechanism as the fitness function and seeks the best subset of the features while standard optimization techniques with learning mechanisms for ranking of the subsets are possible. Kohavi and John (1997) have a leading role in popularization of the wrapper approach that is really powerful in feature selection, but it has its computational complexities and it is more time consuming than Filter method (Huang, Cai, & Xu, 2007). Fig. 1 shows the approach concept.

For a given dataset of $G$ with $N$ features, the wrapper approach starts from a subset of $F_0$ (an empty set, a full set, or any randomly selected set) and with a particular strategy searches the features space. It evaluates each generated subset of $F_i$ by applying a learning model.
that considers \( F_i \) as an input and if the learning model performance improves with \( F_i \); \( F_i \) is regarded as the best current subset. Then the wrapper modifies \( F_i \) by adding or eliminating features and performing the search until coming to the predefined stopping criterion (Kabir & Islam, 2010). The above mentioned feature searching problem is NP-hard and the number of local minima can be quite large and naturally a wide range of heuristic search strategies including forward selection, backward elimination, hill-climbing, branch and bound algorithms, and metaheuristics algorithms like simulated annealing and genetic algorithms (GAs) have been used (Huang et al., 2008).

2.2. Adaptive Neural Fuzzy Inference System (ANFIS)

ANFIS is a multi-layer adaptive network-based fuzzy inference system proposed by Jang (1993). An ANFIS consists of total five layers to implement different node functions to learn and tune parameters in a Takagi–Sugeno fuzzy inference system (FIS) using a hybrid learning mode.

The first layer executes a fuzzification process, the second layer executes the fuzzy AND of the antecedent part of the fuzzy rules, the third layer normalizes the membership functions, the fourth layer executes the conclusion part of the fuzzy rules, and the last layer computes the output of the fuzzy system by summing up the outputs of the four layers. The feed forward equations of the ANFIS structure with two inputs and two labels for each input are presented as follows:

\[
W_i = \mu_A(x) \times \mu_B(x) \quad i = 1, 2 \tag{1}
\]

\[
\bar{W}_i = \frac{W_i}{W_1 + W_2} \quad i = 1, 2 \tag{2}
\]

\[
\begin{align*}
F_1 &= p_1x + q_1y + r_1z \\
F_2 &= p_2x + q_2y + r_2z
\end{align*}
\Rightarrow f = \frac{W_1f_1 + W_2f_2}{W_1 + W_2} = \bar{W}_1f_1 + \bar{W}_2f_2 \tag{3}
\]

where \( x \) and \( y \) are inputs to node \( i \), \( A_i \) and \( B_j \) are linguistic labels for inputs, \( W_i \) is the output of layer 3, and \( \{p_1, q_1, r_1\} \) are the parameters setting.

ANFIS model design consists of two sections: constructing and training. In the construction section, the number and type of FIS structure are defined. Construction of the ANFIS model requires the division of the input/output data into rule patches. This can be achieved by using a number of methods such as grid partitioning, subtractive clustering method and fuzzy c-means (FCM).

When there is only one output, FCM can be applied for making a primary FIS to train the ANFIS (Bezdek, 1981). FCM is done with minimizing a goal function that represents the distance of each data point to data center that has been weighted by membership degree of the data point (Eq. (4)).

\[
\min J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} \frac{U_{ij}^{m} \|X_i - C_j \|^2}{\sum_{i=1}^{N} U_{ij}^{m}} \quad 1 \leq m \leq \infty \tag{4}
\]

where \( m \) is a real number greater than 1 and each of \( U_{ij} \), \( X_i \) and \( C_j \) shows the degree of membership of \( X_i \) in the jth cluster, the \( p \)-dimensional data and the \( p \)-dimensional center of the cluster, respectively, and \( \|x\| \) is any norm that shows the similarity between each measured data and the center. With iterative optimization of the above objective function, fuzzy partitioning is done, by updating membership \( U_{ij} \) and cluster centers \( C_j \), as follows:

\[
U_{ij} = \frac{1}{\sum_{k=1}^{C} \left( \frac{\|X_i - C_k\|^2}{\|X_i - C_j\|^2} \right)^{1-\frac{1}{m}}} \tag{5}
\]

when \( \max_{i,j} \left( \|U_{ij}^{(k+1)} - U_{ij}^{(k)}\| \right) < \varepsilon \) is satisfied, iteration will stop, where \( \varepsilon \) is a number between 0 and 1, and \( k \) is the step’s number of iteration (Esfahanipour & Aghamiri, 2010; Sarkheyli et al., 2015).

2.3. Imperialist Competitive Algorithm (ICA)

ICA is a new socio-politically motivated global search strategy that has been introduced for dealing with different optimization tasks (Atashpaz-Gargari & Lucas, 2007). This evolutionary optimization strategy has shown great performance in both convergence rate and better global optima achievement (Atashpaz-Gargari & Lucas, 2007; Biabangard-Oskouyi, Atashpaz-Gargari, Soltani, & Lucas, 2009; Gargari, Hashemzadeh, Rajabioun, & Lucas, 2008; Khoshnevisan et al., 2015; Rajabioun, Atashpaz-Gargari, & Lucas, 2008; Sepehri Rad & Lucas, 2008). Nevertheless, its effectiveness, limitations and applicability in various domains are currently being extensively investigated.

Fig. 2 (Atashpaz-Gargari & Lucas, 2007) shows the flowchart of the ICA. Similar to other evolutionary algorithms, this algorithm starts with an initial population. Each individual of the population is called a country. Some of the best countries (in optimization terminology, countries with the least cost) are selected to be the imperialist states and the rest form the colonies of these imperialists. All the colonies of initial countries are divided among the mentioned imperialists based on their power. The power of each country, the counterpart of fitness value in the GA, is inversely proportional to its cost. Then an imperialist with its colonies can found an empire. Fig. 3 shows the primary empires while the most powerful imperialist 1 has more countries than the others.
After forming initial empires, the colonies in each of them start moving toward their relevant imperialist country. This movement is a simple model of assimilation policy which was pursued by some of the imperialist states. The Assimilation makes the colonies of each empire in searching space, closer to the imperialist position while the revolution covers random changes in position of some of the countries. During Assimilation and revolution a colony may advance to a better position and has the probability of controlling the entire empire and replace the imperialist.

If after this move, one of the colonies has more power than the imperialist, they will be replaced. To initiate the competition among the empires, the objective function of each empire is calculated. This function depends on the imperialist objective function and its colonies. Then the competition starts and the weakest empire lose its assets and the powerful empire tries to achieve them. The empire that loses all of its colonies will vanish. Finally the most powerful empire possesses the other empires assets and wins the competition. From optimization point of view this leads to finding the optimal solution of the problem, i.e. solution with least cost value.

3. The proposed model

3.1. Adaptive TAs and ANFIS

Technical analysis normally has two general approaches to evaluate the stock prices while the first uses technical indicators and oscillators, and the second uses charts (Chavarnakul & Enke, 2008). Whether mathematical or pattern charts, the mechanism of a typical TA is based on signs and (their associated) signals, while signs are generated by the stock price alteration and signals are restricted to three states of ascending, descending and neutral (Lu, 2014).

The concept of adaptive TAs is not much different from the old one. The new generation of TA models is again categorized into two categories of mathematical and pattern charts and it is again a system of signs and signals but unlike the static old one, it is dynamic.

In traditional combination, target data are results of applying the input data to some pre-determined rules of the selected TA. That is the rules are constant and the concern of model developer is to better educate the network. In traditional combination the target data is determined according to the input data and the predetermined rules. In statistic words the independent and dependent variables are input
and target data, respectively. The following rules (Rules 1–6) are some of predetermined rules in TA which are illustrated in Fig. 4 (Lu, Shiu, & Liu, 2012). The focus is on open ($O_i$), high ($H_i$), low ($L_i$) and close ($C_i$) prices of the stock on the $i$th day due to the Japanese Candlestick (do Prado, Ferneda, Morais, Luiz, & Matsura, 2013).

1. The Piercing: $O_1 > C_1$, $O_2 < C_2$, $O_2 \leq C_1$, and $C_2 > C_1 + 0.5(O_1 - C_1)$.
2. The Bullish Engulfing: $O_1 > C_1$, $O_2 < C_2$, $O_2 \leq C_1$, and $C_2 \geq O_1$.
3. The Bullish Harami: $O_1 > C_1$, $O_2 < C_2$, $O_2 \geq C_1$, and $C_2 < C_1 - 0.5(C_1 - O_1)$.
4. The Dark-cloud Cover: $O_1 < C_1$, $O_2 > C_2$, $O_2 \geq C_1$, and $C_2 < C_1 - 0.5(C_1 - O_1)$.
5. The Bearish Engulfing: $O_1 < C_1$, $O_2 > C_2$, $O_2 \geq C_1$, and $C_2 \leq O_1$.
6. The Bearish Harami: $O_1 < C_1$, $O_2 > C_2$, $O_2 < C_1$, and $C_2 > O_1$.

As we can see in the upper rules, they are some predetermined rules; however, in this present paper, our approach is completely dependent on input and target data in the way that they may be completely different for two distinct periods of time. In other words, there is no pre-determined rule in the proposed synthetic model but include only some inputs and their associated target data on the basis of which new rules are extracted. The most important task in the new model is with the analyzer that confirms the old rules or extracts the new ones. Fig. 5 shows how the analyzer relates input data, target data, and the new technical rules together. In addition, the new approach inserts some kind of continuity to the selected rules.

In this study an ANFIS like in many similar works in the literature plays the analyzer role. But unlike the previous works, in the new model the ANFIS knowledge changes over time according to new training data.

Therefore in this study a modern combination of Japanese Candlestick charting and ANFIS is presented. In fact, it is explored how much efficient an ANFIS would be to infer the TA clues if it is provided with important indices of the TA (based on wrapper feature selection) and actual trends. This model created for presented research, if successful, could be beneficial for stock traders, much in allowing them to make better decision making using the most up-to-date investment techniques.

3.2. Model

The model proposes an adaptive TA based novel methodology for signal prediction and feature selection with a wrapper approach called ANFIS-ICA, which results from the combination of ANFIS as a signal analyzer and ICA, as feature finder and subset evaluator. In the other words, this model uses ANFIS as a signal analyzer tool while a novel wrapper feature selection algorithm selects important
Candlestick based features to feed the model. Fig. 6 shows the general structure of the model while it will be discussed in details in the following parts.

The following parameters are used throughout the study:

\[N_{attrib} : \text{number of features}\]

\[N_{pop} : \text{number of initial population}\]

\[N_{imp} : \text{number of imperialists}\]

\[\text{zeta} : \text{the effect coefficient of the colonies cost on the empire cost}\]

\[\text{Prevolution} : \text{probability of revolution}\]

\[\text{Imp Colonies Costs (i)} : \text{colonies costs of ith imperialist}\]

\[\text{Imp costs (i)} : \text{cost of ith imperialist}\]

\[\text{Imp fitness (i)} : \text{fitness of ith imperialist}\]

\[\text{Max Decades} : \text{maximum periods as stopping criterion in the ICA}\]

\[\text{Num MFs} : \text{number of membership functions for FCM}\]

\[\text{Epoch}_n : \text{number of train epochs in the FCM}\]

### 3.2.1. Development of the initial population in ICA

To generate the initial population, \(N_{pop}\) random permutation of integers from 1 to \(N_{attrib}\) in a matrix structure \((N_{pop} \times N_{attrib})\) as it is shown in Fig. 7 is developed.

The colonies are organized on the basis of the initial population. The \(i\)th population with a position that includes \(N_{attrib}\) features makes the \(i\)th colony, while \(j\)th number in the colony is the number of the feature. Then cost of each colony is calculated and the colonies are sorted on the basis of their cost ascend and the most powerful colonies are selected as imperialists and the others \((N_{pop} - N_{imp})\) are colonies of these imperialists (see Atashpaz-Gargari & Lucas, 2007).

### 3.2.2. Cost function

ANFIS model is applied to calculate the cost function in ICA. Firstly the FCM function is applied for appointing the number of rules and membership functions. FCM uses Gaussian and Linear functions for input and output membership functions respectively. After making the initial FIS structure, ANFIS function is applied to train the system by the initial FIS structure that is made by FCM and training data. Then training RMSE of fuzzy output is calculated by Eq. (6).

\[
RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y'_t - y_t)^2}{n}}
\]  \quad (6)

where \(y'_t\) is the result of prediction and \(y_t\) is the real amount of \(t\)th data.

After the training phase, checking dataset is used to check the ANFIS. Furthermore, the checking RMSE is calculated. The check RMSE is used as the performance measure and cost function in ICA.

### 3.2.3. Assimilation policy

To draw the colony closer to the imperialist (Assimilation), 2 random numbers from 1 to \(N_{attrib}\) are selected as \(R_1\) and \(R_2\), and then \(C_1\) and \(C_2\) are considered as minimum and maximum of them respectively. Then from 1 to \(C_1 - 1\) features of the imperialist are considered as 1 to \(C_1 - 1\) features of the child country; from \(C_1\) to \(C_2\) features of colony are considered as \(C_1\) to \(C_2\) features of the child country; from \(C_2 + 1\) to \(n\) features of the imperialist are considered as \(C_2 + 1\) to \(n\) features of the child country and so the features of the child country is achieved. Fig. 8 shows an example of Assimilation with \(R_1 = 3\) and \(R_2 = 6\); \(C_1 = 3\) and \(C_2 = 6\).

The new country (child) is a combination of the previous colony that converges to the imperialist country.

### 3.2.4. Revolution

For the \(j\)th colony of the \(i\)th imperialist if revolution chance (random number) is smaller than \(P\), revolution is done by random changes in some of the colonies. In this regard two random numbers of \(k_1\) and \(k_2\) from 1 to \(N_{attrib}\) are selected and then feature \(k_1\) from the \(j\)th colony and the \(i\)th empire is replaced with feature \(k_2\) from that exact colony and so the revolution function in some of the countries is done. Fig. 9 shows an example of the revolution.

### 3.2.5. Replacement of the imperialist with the best colony

If one of the colonies possesses more power than its relevant imperial, they will exchange their positions.

### 3.2.6. Cost calculation of the entire empire

The cost of each empire depends on the costs of imperialist and colonies as follows:

\[
\text{Imp fitness (i)} = \text{Imp costs (i)} - \text{zeta} \times \frac{\sum_{t=1}^{n} \text{Imp Colonies Costs (i)}}{n}
\]  \quad (7)

where \(n\) is the number of colonies in \(i\)th empire.

### 3.2.7. Imperialist competition

In this regard, first of all the empire fitness is updated by Eq. (8) and then the intended probability is obtained by Eq. (9).

\[
\text{Update Imp fitness (i)} = \max_{i=1, \ldots, N_{imp}} \text{Imp fitness (i)} - \text{Imp fitness (i)}
\]  \quad (8)

\[
p_i = \frac{\text{update Imp fitness (i)}}{\sum_{i=1}^{N_{imp}} \text{Imp fitness (i)}}
\]  \quad (9)

After calculating the probabilities, the vector of \(P\) is organized as follows:

\[
P = [p_1, p_2, p_3, \ldots, p_{N_{pop}}]
\]  \quad (10)

To distribute the colonies among the empires according to their possession probabilities, the approach of roulette is applied. The Empire with more fitness is has higher probability to be selected. After this stage it should be checked that if the weakest empire does not have any colony, and then it must be allocated as a colony to other empires on the basis of fitness probability of the empires and roulette wheel.
Table 1
Raw approach.

<table>
<thead>
<tr>
<th>No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1/C1</td>
<td>C2/C1</td>
<td>C3/C1</td>
<td>C4/C1</td>
<td>C5/C1</td>
<td>C6/C1</td>
<td>C7/C1</td>
<td>C8/C1</td>
<td>D1/H1</td>
<td>H1/C1</td>
<td>L1/S1</td>
<td>O1/C1</td>
<td>H2/C1</td>
<td>L2/C1</td>
<td>O2/C1</td>
<td>H3/C1</td>
</tr>
</tbody>
</table>

Table 2
Signal approach.

<table>
<thead>
<tr>
<th>No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1/C1</td>
<td>C2/C1</td>
<td>C3/C1</td>
<td>C4/C1</td>
<td>C5/C1</td>
<td>C6/C1</td>
<td>C7/C1</td>
<td>C8/C1</td>
<td>D1/H1</td>
<td>H1/C1</td>
<td>L1/S1</td>
<td>O1/C1</td>
<td></td>
</tr>
</tbody>
</table>

Table 3
Details of the applied data sets.

<table>
<thead>
<tr>
<th>No.</th>
<th>Period of training data (year)</th>
<th>Period of testing data (year)</th>
<th>No.</th>
<th>Period of training data (year)</th>
<th>Period of testing data (year)</th>
<th>No.</th>
<th>Period of training data (year)</th>
<th>Period of testing data (year)</th>
</tr>
</thead>
</table>

Table 4
The correct prediction percentage for 1 and 6 day periods with total number of buy and sell signals.

<table>
<thead>
<tr>
<th>Raw database</th>
<th>Signal database</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>1d (%)</td>
</tr>
<tr>
<td>1</td>
<td>0.33</td>
</tr>
<tr>
<td>2</td>
<td>0.32</td>
</tr>
<tr>
<td>3</td>
<td>0.36</td>
</tr>
<tr>
<td>4</td>
<td>0.35</td>
</tr>
<tr>
<td>5</td>
<td>0.29</td>
</tr>
<tr>
<td>6</td>
<td>0.33</td>
</tr>
<tr>
<td>7</td>
<td>0.33</td>
</tr>
<tr>
<td>8</td>
<td>0.33</td>
</tr>
<tr>
<td>9</td>
<td>0.33</td>
</tr>
<tr>
<td>10</td>
<td>0.33</td>
</tr>
<tr>
<td>11</td>
<td>0.33</td>
</tr>
<tr>
<td>12</td>
<td>0.33</td>
</tr>
<tr>
<td>13</td>
<td>0.33</td>
</tr>
<tr>
<td>14</td>
<td>0.33</td>
</tr>
</tbody>
</table>

3.2.8. Stopping measure
The algorithm is repeated until it comes to a pre-determined number of repetitions and finally the best answer is selected from them.

To sum up these steps, the semi-codes of the model are as follows:

1. Set up ICA settings:
   - Max Decades; N_{pop}; N_{imp}; Prevolution; zeta;
   - N_{col} = N_{pop} - N_{imp};

2. Set up training and checking data and epoch_n
(3) \( \theta = 1 \)
(4) Generate the initial population
(5) Calculate the cost of each colony:
1. Develop the FIS structure by FCM.
2. Train the ANFIS by the initial FIS structure.
3. Validate the model by RMSE and compute the train RMSE.
4. Validate the model by the checking data.
5. Compute the check RMSE and set as cost function.
(6) Select the most powerful colonies as imperialists.
(7) Allocate rest of the colonies to the imperialist on the basis of the imperialist power.
(8) Generate empires by imperialists and their colonies.
(9) Move the colonies toward the relevant imperialist.
(10) Revolve in some of the colonies.
(11) Exchange the position of a colony and the imperialist if its cost is lower.
(12) Compute the objective function (the total cost) of all empires.
(13) Pick the weakest colony and give it to the best empire.
(14) Eliminate the powerless empires.
(15) If the stop condition is satisfied stop, and if not go to step 10.

(16) Select the optimum subset of the features with the minimum check error (Check RMSE).
(17) If $\theta = \text{Max Decades}$ then stop, and if not $\theta = \theta + 1$ and go to step 5.

4. Empirical results

4.1. Input data

In this study the training data are based on the two applied approaches of Jasemi et al. (2011). The first approach (Raw database) is based on Raw input features including 15 items and 1 output. In this
approach the focus is on open ($O_i$), high ($H_i$), low ($L_i$) and close ($C_i$) prices of the stock in the $i$th day due to the Japanese Candlestick during last 3 days while to cover the stock price trend the close prices of the stock during the last 7 days are also included. Totally this approach comes to 15 normalized indices of $C_i = 2, 3, 4; O_i, H_i, L_i$ and $C_i = 5, 6, 7$.

The 15 indices of this approach are shown in Table 1.

The second approach (Signal database) is based on the reverse signals of Japanese Candlestick technique including 24 input features and 1 output. This package covers the important factors of decision making in the technique. In this approach the focus is on the reversal signals of Japanese Candlestick like Morning star, Inverted hammer, Harami and many others. In fact this set of inputs is more advanced than the previous and covers important pattern charts’ clues of the TA which are defined as shown by Table 2.

These approaches in 48 data sets, according to Table 3, are applied to train and test the introduced ANFIS model. This structure gains the advantage of adaptive TA in which the input data and the target data are dependent and change over the time. Also, this structure is enjoyed from sliding window function in time series prediction (Mozaffari, Mozaffari, & Azad, 2015).

4.2. Results

The input data of our experiment belong to daily stock prices of General Motors Company at New York Stock Exchange from 2000 to 2009. 48 data sets according to Table 3 are applied for learning and checking.

After the experiments is it examined that the proposed model is efficient? And which approach has a better performance? Does the quality of forecasting increase when the training data covers a longer period? Does the quality of forecasting decline when the distance between the training and testing data increases?

To optimize the ANFIS parameters, the pre-assumed hybrid method of pre-propagation and least squares is applied. After different evaluation, 20 membership functions were selected as optimal number of FCM rules. In ICA on the basis of different runs, the total number of colonies, number of empires, rate of colonies revolution ($P$), effect coefficient of colonies power on empire power ($\zeta$) and maximum number of repetition as the stop condition are 20, 5, 0.5, 0.1 and 30, respectively.

The following figures show the outputs of the wrapper ANFIS-ICA model with the Raw input data of 2000–2001 for train and 2003 for checking. Fig. 10a shows the features attitudes of the data approach. The upper trend shows the features trend and the bottom represents the outputs for one sample of normalized data. Fig. 10b shows the empires costs and the stages of empires elimination in ICA in the way that the weak empires are eliminated and the strong ones remain. Fig. 10c shows the decrease of RMSE (cost) in repetition of the algorithm and represents two charts of the costs mean and the minimum cost in each repetition. Fig. 10d shows checking output and ANFIS output with root of mean square error for best position on top of the figure.
Lee and Jo (1999) believe that if the hit ratio defined by
\[
\frac{\text{Number of successes}}{\text{Total number of signals}} \times 100\% > 51\%,
\]
the model is regarded useful. The hit ratios of the new model for the first and second approaches are 85\% and 87\%, respectively while the second approach gives more number of buy and sell signals.

In the base study of Jasemi et al. (2011) the hit ratios for the first and second approaches are 75\% and 74\%, respectively. Unlike the base study our results show the superiority of the second approach.

The percentages of correct signals for 1-day period are 40\% and 43\% for the first and second approaches, respectively. It is to be noted that increasing the learning period’s length does not have a sensible effect on the results.

Table 4 shows the prediction results for 1-day and 6-day periods and also the total number of buy and sell signals. The correct signals are achieved from comparing the predictions with real happenings of the stock market. The complete list of the results is presented in Appendix A.

Fig. 11 shows the RMSE changes for the datasets of 1, 2, 3 and 4 in the first approach for different number of features from 3 to 15. Figs. 12 and 13 show the RMSE changes for different number of features and datasets. The first approach results show that all of the 15 features are important and have the best prediction; while by increasing the number of features in any of the 48 data sets, RMSE is decreased.

Fig. 14 shows RMSEs in different numbers of features for the first four datasets in the second approach. In datasets 1, 2 and 4, 15, 10 and 20 features respectively have the minimum RMSE. The RMSE changes in all datasets are shown in Fig. 15. Also, Fig. 16 shows the RMSE changes with different number of features for all datasets. Results of this approach show that the feature selection is very important and affect the model efficiency considerably. In most of the second approach data sets, RMSE decreases by increasing the number of features and gets its minimum for a particular number of features and then increases. The best result in almost all of the data sets occurred in an interval from 15 to 20 features.

It should be noted that, the main aim of feature selection is to obtain fewer but more effective numbers of features and simultaneously minimizing the loss of the model’s prediction accuracy. The results indicated that, by using 15 features, our obtaining accuracy is less than using the second approach with 24 features. However, in the second approach, we find the optimized number of features based on wrapper algorithm. Our results in finding the most effective features can enable the investors to analyze the market with fewer features, and not getting confused in the market by many features which are not necessarily effective.

It is to be noted that the features are not the same in different datasets. Table 5 shows the optimum features and their best and mean costs for each dataset.

The wrapper ANFIS-ICA output of this study shows that feature selection by ICA has a good impact on the accuracy of ANFIS for different datasets. We also compare our method with Jasemi et al. (2011) which used a feed forward neural network (NN) optimized by a

![Fig. 12. RMSE changing for the first approach in the 48 datasets.](image-url)
Fig. 13. RMSE changing for the first approach with different number of features (Nattrib) for each dataset.

Fig. 14. RMSE changing for the second approach in the first four datasets.

back propagate algorithm and ANFIS-GA wrapper model which used genetic algorithms (GA) instead of ICA for feature selection. The results accuracy shows that wrapper ICA-ANFIS algorithm dominates the NN and also wrapper ANFIS-GA algorithm. In more detail discussion, the ANFIS model has solely better results than NN algorithm. Forecasting accuracy of NN depends on learning data set and their adequacy. Moreover, NN methods sometimes get stuck in local minimum, so choosing proper dataset is too critical in neural network models. In the other side, the ANFIS has better result when it deals with noisy and inadequate data set. Based on the fact that, in stock exchange dataset we always encounter with noisy data, enjoying from ANFIS model is preferable than NN models. Also, the ICA can diversify the dataset for learning issues better than other models such as GA model (Table 6).
It is considerable to mention that ICA algorithm gets the final response more quickly than GA and because of the fact that wrapper models are time-consuming and need a large amount of time to convergence, this algorithm proves its ability in other point of view.

5. Conclusion

The paper proposes a novel methodology for feature selection and time series estimation based on a wrapper approach called ANFIS-ICA, which results from the combination of ANFIS, a fuzzy logic based
function estimator that adopts the adaptive systems framework to facilitate learning and adaptation, and ICA, an evolutionary algorithm used for subset evaluation. Additionally, the proposed model is applied on the basis of the technical analysis of Japanese Candlestick by two approaches of Signal and Raw databases. In fact, the paper incorporates the technique of Japanese Candlestick charts for data transformation and attribute generation. Finally, daily stock prices for the General Motors Company at New York Stock Exchange from 2000 to 2009 are applied for benchmarking.

The root of mean square errors, the number and percentage of the signals that are predicted correctly are considered as the assessment measures while the results show a good reliability of the model that is even better than the base study and model of Jasemi et al. (2011).

In the first approach, all the 15 features together have the best approximation but in the second approach, feature selection shows its effect on the results very well, while in different datasets almost 20 numbers of features are more appropriate. As a matter of fact it is believed that the analysis of different features or different subsets of them has significant influence on prediction accuracy. The second approach has a better performance than the first one. Also, it should be noted that the prediction accuracy of this model is better than some other time series models and shows great abilities for decision making about stock trading.
Futureresearchdirectionsofthepaperincludebutarenotlimited
to

1. Combiningpredictionmethodsintheframeworkoffusionmodels
or optimize the classification algorithms by applying some meta-
heuristics algorithms to improve the prediction results.

2. Predicting the other important variable (in addition to return)
such as liquidity (Barak, Abessi, & Modarres, 2013).

3. Using fundamental features and textual information, in addition
to technical features, in order to have a more comprehensive fea-
tures and to be able to predict long term situation of stocks.

4. Prediction of stocks in other popular stock markets such as DAX
or NASDAQ.

5. Applying statistical feature selection models such as Filter meth-
ods (Huang & Tsai, 2009) and CFS(Zhang et al., 2014), or using
other wrapper models to compare the results.

Acknowledgments

The research was supported by the Operational Programme Edu-
cation for Competitiveness – Project no. CZ.1.07/2.3.00/20.0296 and
also within the Czech Science Foundation (GACR) project No. 13-
13142S and SP2015/15, and SGS research project of VSB-TU Ostrava.

Appendix A

Table A1: The complete list of the results is presented here. The
columns 1–6, column 7 and column 8 show the correct signals in 1, 2,
3, 4, 5, and 6 day periods, the total number of correct signals and the
total number of the emitted signals by the system, respectively.

References

evolutionary Levenberg-Marquardt neural networks and data pre-processing for

rithm for optimization inspired by imperialistic competition. In Proceedings of IEEE
congress on evolutionary computation, CEC 2007 (pp. 4661–4667). IEEE.

and neuro-fuzzy systems, in stock market prediction: the WASP system. Expert Systems
with Applications, 38, 9106–9206.

evolutionary Levenberg-Marquardt neural networks and data pre-processing for

rithm for optimization inspired by imperialistic competition. In Proceedings of IEEE
congress on evolutionary computation, CEC 2007 (pp. 4661–4667). IEEE.

evolutionary Levenberg-Marquardt neural networks and data pre-processing for

rithm for optimization inspired by imperialistic competition. In Proceedings of IEEE
congress on evolutionary computation, CEC 2007 (pp. 4661–4667). IEEE.

evolutionary Levenberg-Marquardt neural networks and data pre-processing for

rithm for optimization inspired by imperialistic competition. In Proceedings of IEEE
congress on evolutionary computation, CEC 2007 (pp. 4661–4667). IEEE.

evolutionary Levenberg-Marquardt neural networks and data pre-processing for

rithm for optimization inspired by imperialistic competition. In Proceedings of IEEE
congress on evolutionary computation, CEC 2007 (pp. 4661–4667). IEEE.
Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting tech-
tems with Applications, 41, 6235–6250.
Barak, S., & Modarres, M. (2015). Developing an approach to evaluate stocks by forecast-
plication of imperialist competitive algorithm for materials property characterization from
market trend analysis and prediction using uncentered Kalman filter. Applied Soft
Computing, 19, 41–56.
on pattern recognition and technical analysis: forecasting the DJIA index with intraday data. Expert Systems with Applications, 42, 5963–5975.
with learning approach for enhanced stock trading model. Expert Systems with Ap-
plications, 36, 12537–12546.
Crone, S. F., & Koutentzes, N. (2010). Feature selection for time series prediction – a
ficacy of Candlestick Chart Analysis for the Brazilian Stock Market. Procedia
Computer Science, 22, 1136–1145.
algorithm: a novel approach for PID controller design in MIMO distillation column
tem (BNMNAS) for stock price prediction: case study of DAX stock price. Applied Soft
Huang, C.-L., & Tsai, C.-Y. (2009). A hybrid SOFM-SVR with a filter-based feature selec-
network. Neurocomputing, 73, 3273–3283.
Khoshnevisan, B., Bolandnazar, E., Barak, S., Shamshirband, S., Maghsoudlou, H.,
algorithm for better energy use in crop production. Stochastic Environmental Research and Risk Assessment.
gence, 97, 273–324.
Lawrence, S., Giles, C. L., & Tsoi, A. C. (1997). Lessons in neural network training: over-
fitting may be harder than expected. In Proceedings of AAAI/IAAI (pp. 540–545).
Lee, K., & Jo, G. (1999). Expert system for predicting stock market timing using a candle-
Lin, X., Yang, Z., & Song, Y. (2011). Intelligent stock trading system based on improved
port vector machines. Information Sciences, 179, 2208–2217.
Menkhoff, L. (2010). The use of technical analysis by fund managers: international evi-
window time series analysis and an evolutionary least learning machine: a case study on San Francisco urban roads. Engineering Science and Technology, 18, 150–162.
2172.
tional science and its applications, ICCSA 2008 (pp. 680–695). Springer.
erference (CSICC’08). Kish Island, Iran.
prediction: union, intersection, and multi-intersection approaches. Decision Sup-
port Systems, 50, 258–269.
Vairappan, C., Tamura, H., Gao, S., & Tang, Z. (2009). Batch type local search-based ad-
aptive neuro-fuzzy inference system (ANFIS) with self-feedbacks for time-series pre-
diction. Neurocomputing, 72, 1870–1877.
1022.