HASIL

CEK_IJFS_RS+TSK+MBGDUR

by ljfs_rs +tsk+mbgdur

Submission date: 13-Dec-2021 01:07PM (UTC+0700)
Submission ID: 1728873815
File name: IJFS_RS_TSK_MBGDUR.docx (77.65K)
Word count: 2803
Character count: 15098
Implementation of Takagi Sugeno Kang (TSK) Fuzzy with Rough Set Theory and Mini-Batch Gradient Descent Uniform Regularization (MBGD-UR)

Sugiyanto Sugono¹, Zani Anjani Rafsanjani HSM², Annisa Eka Haryati²
¹Department of Mathematics, Ahmad Dahlan University, Yogyakarta, Indonesia
²Master of Mathematics Education, Ahmad Dahlan University, Yogyakarta, Indonesia

Abstract Takagi Sugeno Kang (TSK) is a fuzzy method often used because its output is in the form of a constant or a function. The input used in TSK fuzzy usually affects the number of rules generated such that the use of larger dimensions of data normally leads to more rules, thereby causing complexity for the rule. It is possible to overcome this problem using dimension reduction methods which minimize the number of existing dimensions in the data. An example of this method is rough set theory introduced by Pawlak in 1982. Moreover, Mini-Batch Gradient Descent (MBGD) which is an optimization method modified using Uniform Regularization (UR) was also applied. The body fat data of 252 respondents were used as input while the output was evaluated using Mean Absolute Percentage Error (MAPE). The results showed MAPE value was 37% and this is included in the reasonable category.

Keywords Rough set, Takagi Sugeno Kang fuzzy, Mini batch gradient descent, uniform regularization

1. Introduction

The dimensions in each data set have different sizes with those in large data often observed to have a high level of complexity which affects the number of rules generated when used in a fuzzy inference system. This is indicated by the fact that the use of large data dimensions as input normally leads to the generation of more rules.

One of the studies conducted by [1] uses Takagi Sugeno Kang (TSK) fuzzy which was observed to be limited when the input used was large. This TSK is a type of inference fuzzy introduced by Takagi and Sugeno in 1985 [2] which is often used for classification or prediction [3]–[5].

TSK was used by [6] to perform the AP-TSK-PID scheme in dealing with stochastic and non-stochastic uncertainties of nonlinear dynamic systems. It has also been applied in the health sector by [7] to predict the adequacy of dialysis in hemodialysis.
patients [8] to treat the problem associated with the control of glucose levels for type-1 diabetes.

TSK fuzzy can be optimized using Gradient Descent (GD) to ensure better operating time performance [9]. This GD is one of the optimization algorithms normally used to minimize the cost function in machine learning and usually conducted by updating each parameter based on previous step [10], [11]. The three types of GD are batch, stochastic, and Mini-Batch. The batch type has been previously used for classification by [12], [13] and the stochastic type by [14]–[16]. This present study used Mini-Batch Gradient Descent (MBGD) because it tends to have a smaller computational load and a faster convergence due to the involvement of only the data in a batch in each iteration [17], [18].

Several studies have been conducted on MBGD such as its use by [19] to efficiently train the ANN equalizer, [20] to optimize the IoT 4.0 industry while [21] combined it with MMLDA to predict IncRNA disease association. Moreover, the regularization technique was used to avoid overfitting and increase generalization due to its ability to promote the generalization of the algorithm by avoiding coefficients in order to fit the training sample [22], [23]. According to Goodfellow [23], regularization is defined as “any modification made to a learning algorithm intended to reduce generalization errors and not training errors”. This technique is believed to be needed to stabilize numerical calculations [24].

Rough set is a dimension reduction method introduced by Pawlak in 1982. It was conducted by [25] to perform feature selection in genetic algorithms and observed to have provided good results for the selected features. Another study by [26] also used this method for decision-making.

Previous explanation showed that [1] has limitations related to the large data dimensions. Therefore, this present study used rough set to reduce the dimension of the large data.

2. Preliminaries

2.1 Rough Set

Rough set is one of the dimension reduction techniques developed by Pawlak in 1982 with its principle associated with a reflexive, symmetrical, and transitive equivalence relation [27]. It was applied to analyze the data in this present study in accordance with a study conducted by [28].

2.2 Fuzzy Set

Definition 1: Let $X$ represent the universe of discourse, $x$ is a member of the universe while $X$ and $A$ represent fuzzy sets. Therefore, fuzzy set with membership function of $\mu_A(x)$ is:

$$\mu_A(x): X \rightarrow [0,1]$$

Definition 2: The fuzzy set $A$ in universe $X$ can be defined as a set of ordered pairs as indicated in the following equation:

$$A = \{(x, \mu_A(x))| x \in X\}$$

where, $\mu_A(x)$ is the membership function $x$ in fuzzy set $A$ which lies on the interval [0,1] [29].

2.3 Takagi Sugeno Kang (TSK) Fuzzy

TSK fuzzy system with one input $x_1$ and $x_2$ as well as output $y$ is described by fuzzy inference rules as follows [30]:
\[ R_j = IF \ x_1 \in G_j(x_1) \ AND \ x_2 \in G_j(x_2) \ THEN \ y = P_j(x_1, x_2) \]  

Where, \( j = 1, 2, \ldots, r, P_j(x_1), G_j(x_2) \) is a fuzzy set and \( P_j(x_1, x_2) \) is a degree polynomial \( d \).

Definition 3: TSK system in line with rules (2.3) is defined as follows [30]:

- The order is zero if \( P_j(x_1, x_2) = b_j \), where, \( b_j \in \mathbb{R} \), and this means the consequent function is a constant (degree polynomial \( d \) is equal to zero).
- The order is one if \( P_j(x_1, x_2) = w_{j1}x_1 + v_{j1}x_2 + b_j \), where, \( w_{j1}, v_{j1}, b_j \in \mathbb{R} \), and this means the consequent functions are linear (a degree polynomial \( d \) is equal to one).
- The order is high if \( P_j(x_1, x_2) = w_{m_j}x_1^{m} + \ldots + w_{1j}x_1 + v_{m_j}x_2^{m} + \ldots + v_{1j}x_2 + b_j \), where, \( m \geq 2, w_{k_j}, v_{k_j} \in \mathbb{R} \) and \( k = 2, 3, \ldots, m \) and this means the consequent function is nonlinear (a degree polynomial \( d \) is greater than one).

Defuzzification is a fuzzy process aimed at converting fuzzy numbers to crisp numbers. Therefore, the defuzzification value \( (\hat{y}^*) \) was calculated using the following equation:

\[ \hat{y}^* = \frac{\sum_{i=1}^{N} a_i y_i}{\sum_{i=1}^{N} a_i}, \quad i = 1, 2, \ldots, N \]  

Where:
- \( y_i \) = output value in the \( i \)-th rule
- \( \hat{y}_i \) = output value in the \( i \)-th rule

### 2.4 Mini Batch Gradient Descent (MBGD)

MBGD is GD method which uses the concept of Mini-Batch to update parameters. Meanwhile, the updated parameter can be defined as follows [17]:

\[ \theta = \theta - \eta \cdot \nabla f(\theta, x_{(i+n)}, y_{(i+n)}) \]  

Where, \( \eta > 0 \) is learning rate (step size) [31].

#### 2.5 Uniform Regularization (UR)

UR is a regularization method which forces the rules to have firing levels by minimizing losses [1]. It can be calculated as follows:

\[ \bar{\xi}_{UR} = \sum_{i=1}^{R} \left( \frac{1}{N} \sum_{n=1}^{N} \xi_{f}(x_n) - \bar{r} \right)^2 \]

Where, \( N \) is the number of training samples and \( r \) is the firing level of each rule. Furthermore, \( \bar{\xi}_{UR} \) is added to the loss function in MBGD-based TSK classification training using Mini-Batch with \( N \) training samples and this is represented as follows

\[ \mathcal{L} = \mathcal{L} + \alpha \bar{\xi}_{UR} + \lambda \sum_{i=1}^{R} \left( \frac{1}{N} \sum_{n=1}^{N} f_{r}(x_n) - \bar{r} \right)^2 \]

### 2.6 Mean Absolute Percentage Error (MAPE)

MAPE is one of the methods normally used to evaluate a model and its value can be determined using the following equation [32]:

\[ MAPE = \frac{\sum_{i=1}^{n} |y_i - y'_i|}{\sum_{i=1}^{n} y_i} \times 100\% \]

Where, \( y_i \) is the \( i \)-th data, \( y'_i \) is the \( i \)-th data for forecasting, and \( n \) is the total data. The prediction criteria for MAPE as indicated by [32] are as follows:
3. Results

The body fat data from a database known as Kaggle was used in this study. It consists of data for 252 respondents with 14 independent variables and 1 dependent variable as indicated in the following Table 2.

<table>
<thead>
<tr>
<th>X₁</th>
<th>X₂</th>
<th>X₃</th>
<th>...</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>154.25</td>
<td>67.75</td>
<td>12.3</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>173.25</td>
<td>72.25</td>
<td>6.1</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>190.75</td>
<td>70.5</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>207.5</td>
<td>70</td>
<td>31.9</td>
<td></td>
</tr>
</tbody>
</table>

The dimensions of the data set were reduced using a rough set and the results are presented in Table 3.

<table>
<thead>
<tr>
<th>X₁</th>
<th>X₂</th>
<th>X₃</th>
<th>...</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>154.25</td>
<td>67.75</td>
<td>1.0708</td>
<td>12.3</td>
<td></td>
</tr>
<tr>
<td>173.25</td>
<td>72.25</td>
<td>1.0853</td>
<td>6.1</td>
<td></td>
</tr>
<tr>
<td>190.75</td>
<td>70.5</td>
<td>1.0399</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>207.5</td>
<td>70</td>
<td>1.0271</td>
<td>31.9</td>
<td></td>
</tr>
</tbody>
</table>

The data in Table 3 were further converted into fuzzy numbers, and the results are presented in the following table.

<table>
<thead>
<tr>
<th>X₁</th>
<th>X₂</th>
<th>X₃</th>
<th>...</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.2713</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The input in Table 4 was used to obtain the following rules:

[R1] If X₁ is nonstandard, X₂ is high, then Y is athletic.
[R2] If X₁ is standard, X₂ is high, then Y is athletic.
[R3] If X₁ is nonstandard, X₂ is high, then Y is good.
[R4] If X₁ is nonstandard, X₂ is high, then Y is normal.
[R5] If X₁ is nonstandard, X₂ is high, then Y is overweight.
[R6] If X₁ is standard, X₂ is high, then Y is good.
[R7] If X₁ is nonstandard, X₂ is high, then Y is athletic.

The similarities in each rule were later determined using MBGD-UR, and the results are indicated as follows:

\[
\begin{align*}
    y₁ &= -0.1515 + 0.9664X₁ + 0.9201X₂ + 0.6203X₃ \\
    y₂ &= -0.2981 + 0.8369X₁ + 0.8939X₂ + 0.7895X₃ \\
    y₃ &= -0.5750 + 0.6356X₁ + 0.6385X₂ + 0.6492X₃ \\
    y₄ &= -0.6431 + 0.8585X₁ + 0.6853X₂ + 0.5158X₃ \\
    y₅ &= -0.5017 + 0.7990X₁ + 0.4268X₂ + 0.6014X₃ \\
    y₆ &= -0.3812 + 0.7074X₁ + 1.1365X₂ + 0.4482X₃ \\
    y₇ &= -0.3093 + 0.9108X₁ + 0.6590X₂ + 0.7610X₃ \\
\end{align*}
\] (9)
Defuzzification was conducted on the rules obtained, and the results are indicated in Table 5.

<table>
<thead>
<tr>
<th>$Y'$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.3</td>
<td>18.2045</td>
</tr>
<tr>
<td>6.1</td>
<td>19.7421</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>20.0780</td>
</tr>
<tr>
<td>31.9</td>
<td>20.6161</td>
</tr>
</tbody>
</table>

MAPE value was determined as follows:

$$MAPE = \frac{\sum_{i=1}^{n} |y_i - y'_i|}{\sum_{i=1}^{n} y_i} \times 100\%$$

\[= \frac{12.3-18.2045}{12.3} + \frac{6.1-19.7421}{6.1} + \frac{31.9-20.6161}{31.9} \times 100\% \approx 36.5510\% \approx 37\%\]

4. Discussion

This study used 252 data with 14 independent variables and 1 dependent variable, and the results of the dimension reduction using rough set method are presented in Table 3. The 14 variables were discovered to be reduced to 4 variables including weight ($X_1$), height ($X_2$), density ($X_3$), and body fat ($Y$).

Rough set results were subsequently used as input in TSK with each variable subjected to a fuzzification process to convert the data to fuzzy numbers using membership functions. This led to the generation of 7 rules which are in the form of IF-THEN as in Equation (3). Moreover, the consequences for each rule were optimized using MBGD-UR, and the constants generated for each rule were arranged into Equation (8).

The defuzzification process was later used to obtain output in the form of firm numbers. It is important to know that the defuzzification value was determined by multiplying the $y$ value with the predicate alpha in each rule and dividing it by the total predicate alpha. The defuzzification ($Y'$) was calculated using Equation (4) and the results are shown in Table 5.

![Real Data and Predicted Results](image)

Figure 1 was used to compare the data from the predicted results and the real data. Moreover, MAPE value was calculated using Equation (8) to determine the accuracy of the model obtained and the value was found to be 37% which is classified as a reasonable category as indicated in Table 1.

5. Conclusion

This study applied rough test dimension reduction method TSK fuzzy to body fat data after which the rule from TSK fuzzy was optimized through MBGD modified using UR. Moreover, MAPE was used for evaluation and its value was recorded to be 37%, thereby indicating the model is included in the reasonable category. Therefore, it is recommended that further studies use other dimension optimization and reduction methods.
Acknowledgments

The authors are grateful to Ahmad Dahlan University research institute for supporting this study.

References


15. B. Y. Hsueh, W. Li, and I. C. Wu,


32. Minin F. Rohlman and L. Ardiantoro,
### Originality Report

<table>
<thead>
<tr>
<th>Similarity Index</th>
<th>Internet Sources</th>
<th>Publications</th>
<th>Student Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>14%</td>
<td>8%</td>
<td>11%</td>
<td>1%</td>
</tr>
</tbody>
</table>

#### Primary Sources

1. **export.arxiv.org**
   - Internet Source
   - 1%

2. **Krzysztof Wiktorowicz, Tomasz Krzeszowski.**
   - Publication
   - 1%

3. **link.springer.com**
   - Internet Source
   - 1%

4. **ebin.pub**
   - Internet Source
   - 1%

5. **Mohd Omar, Pradeep Kumar.**
   - "Detection of Roads Potholes using YOLOv4", 2020 International Conference on Information Science and Communications Technologies (ICISCT), 2020
   - Publication
   - 1%
<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Authors</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Bearing fault diagnosis based on Natural Adaptive Moment Estimation algorithm and improved octave convolution</td>
<td>Mei Ying Qiao, Xia Xia Tang, Jian Ke Shi, Jian Yi Lan</td>
<td>IEEE Access, 2020</td>
</tr>
<tr>
<td>7</td>
<td>An Optimization of Several Distance Function on Fuzzy Subtractive Clustering</td>
<td>Sugiyarto Surono, Annisa Eka Haryati, Joko Eliyanto</td>
<td>IOS Press, 2021</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>M. Zeinali</td>
<td>docplayer.net</td>
</tr>
<tr>
<td>9</td>
<td>Adaptive chattering-free sliding mode control design using fuzzy model of the system and estimated uncertainties and its application to robot manipulators</td>
<td>M. Zeinali</td>
<td><a href="http://www.diat.ac.in">www.diat.ac.in</a></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>file.scirp.org</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td><a href="http://www.mdpi.com">www.mdpi.com</a></td>
</tr>
<tr>
<td></td>
<td>Title</td>
<td>Author(s)</td>
<td>Source</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>14</td>
<td>Hybrid Artificial Intelligent Systems</td>
<td>arxiv.org</td>
<td>Internet Source</td>
</tr>
<tr>
<td>#</td>
<td>Reference</td>
<td>Publication</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>--------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;Random reselection particle swarm optimization for optimal design of solar photovoltaic modules&quot;, Energy, 2022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>&quot;A self-organizing deep belief network based on information relevance strategy&quot;, Neurocomputing, 2020</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ying-Xu Wang, Hong-Gui Han, Min Guo, Jun-Fei Qiao.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>doctorpenguin.com</td>
<td>Internet Source</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>eprints.uad.ac.id</td>
<td>Internet Source</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td><a href="http://www.ccdc.neu.edu.cn">www.ccdc.neu.edu.cn</a></td>
<td>Internet Source</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Title</td>
<td>Authors</td>
<td>Source</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>27</td>
<td>&quot;Intelligent Computing Theories and Application&quot;, Springer Science and Business Media LLC, 2021</td>
<td></td>
<td>Publication</td>
</tr>
<tr>
<td>29</td>
<td>jyx.jyu.fi</td>
<td></td>
<td>Internet Source</td>
</tr>
</tbody>
</table>

Exclude quotes: On
Exclude bibliography: On
Exclude matches: Off