Fuzzy rough regression with application to wind speed prediction

Shuang An\textsuperscript{a,b}, Hong Shi\textsuperscript{b}, Qinghua Hu\textsuperscript{b,⇑}, Xiaoqi Li\textsuperscript{a}, Jianwu Dang\textsuperscript{b}

\textsuperscript{a}Northeastern University, Shenyang 110004, PR China
\textsuperscript{b}Tianjin University, Tianjin 300072, PR China

\textbf{Abstract}

Accurate wind speed prediction is a prerequisite of large-scale wind power generation. There are several uncertain factors which degrade the performance of the current wind speed prediction systems. Fuzzy rough sets are considered as a powerful tool to deal with uncertainty, and have been widely discussed and applied in classification learning. In this work we describe a regression algorithm based on fuzzy rough sets, consisting of fuzzy partition, fuzzy approximation and estimation of regression values. In this algorithm, the training set is divided into $k$ fuzzy classes with fuzzy partition, and then the predicted values of test samples are determined in the finite intervals with fuzzy rough approximation, finally they are estimated with lower and upper limits of the intervals. Numerical experiments on UCI data sets and wind speed prediction show the effectiveness of the proposed algorithm.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig1.png}
\caption{Illustration of fuzzy rough set approximation.}
\end{figure}

1. Introduction

Wind power, as one of the alternatives to fossil fuels, is plentiful, renewable and clean. However power plant scheduling, power trading and grid operations can only be effectively carried out if accurate and reliable prediction of wind power is available. There are several uncertain factors in building an accurate prediction system, such as uncertain weather, noise and sampling errors [31,43]. It is highly desirable to introduce new techniques for modeling the uncertainty in wind speed prediction.

Rough sets [30] and fuzzy sets [44] are two different and complementary mathematical tools to deal with uncertainty. In 1990, these theories were combined, and led to a new direction of fuzzy rough set (FRS) theory [10], where the indiscernibility and the fuzziness are both considered. Since the definition of fuzzy rough sets was introduced, it has been widely applied in various domains for handling uncertainty [4,22,35]. This theory is recognized as one of important mathematical tools for granular computing and uncertainty reasoning in the past decade [23,41]. Recently, this theory was further integrated with kernel trick [19], and some robust models were developed to deal with noise [16,17,20].

Given a fuzzy set $F$, the fuzzy rough set theory approximates it with a pair of fuzzy sets, called lower approximate and upper approximate, respectively, illustrated in Fig. 1, where $F$ is a given fuzzy set. It is approximated by two fuzzy sets, lower and upper approximations, i.e. $\underline{RF}$ and $\overline{RF}$. This inspires us that if we cannot compute the accurate function value $F(x)$ for Sample $x$, we can estimate an interval that $F(x)$ belongs to. Fig. 1 shows $\underline{RF}(x) \leq F(x)$ for any samples, and $F(x) \leq \overline{RF}(x)$ for

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig2.png}
\caption{Illustration of fuzzy rough set approximation.}
\end{figure}

\textsuperscript{⇑}Corresponding author. Tel.: +86 22 27401839.
E-mail address: huqinghua@tju.edu.cn (Q. Hu).

http://dx.doi.org/10.1016/j.ins.2014.03.090
0020-0255/© 2014 Elsevier Inc. All rights reserved.
all samples. Consequently, \( F(x) \in [RF(x), \overline{RF}(x)] \). This information is useful for power scheduling, which indicates that the fuzzy rough set theory has a great advantage in interval analysis.

In real-world applications, available information is often uncertain. So this information usually is represented by fuzzy sets or interval data \([14]\). Fuzzy regression analysis is an important tool to handle interval data. It has been successfully applied in different applications \([7,15]\). Fuzzy regression was first developed for linear systems \([34]\), and many researchers have been focusing on the study of fuzzy linear regression \([26,27,29,32,33,42]\). Fuzzy nonlinear regression usually assumes the underlying functions \([5,9,14]\), and treats the estimation procedures of some particular models, such as polynomial model, exponential model, and logarithmic model.

Summarizing the recent studies on fuzzy rough sets, it is easy to derive that the fuzzy rough set theory is mainly used in classification problems, including feature selection \([2,6,16,18,24,25,36]\), classifier designing \([1,17,37,46]\) and rule learning \([38,45]\). In addition, the FRS theory also performs its superiority in regression analysis. In \([21]\), fuzzy rough set theory was used in predicting real values by combining nearest neighbor rule. The dependency of FRS has been considered for regression-type data in \([8]\).

In wind speed data, some uncertain factors seriously affect the performance of the built prediction model. Luckily, fuzzy rough set theory is the powerful tool for dealing with the uncertainty of data in practice. In this work, we extend the fuzzy rough set theory to regression tasks, and present a fuzzy rough based regression analysis technique for wind speed prediction. We conduct some numerical experiments on wind speed prediction and other open data sets. Firstly, we compute the fuzzy partition of training samples with fuzzy equivalence class \([3]\) to test the prediction performance of the new algorithm via comparing FRRPA with some state-of-the-art algorithms.

The remainder of the paper is organized as follows. Section 2 reviews the basic theory of fuzzy rough sets and gives a new definition of fuzzy dependency for regression data set. In Section 3, we present the fuzzy rough regression principle, and construct a fuzzy rough regression prediction algorithm in Section 4. The experimental results and analysis are provided in Section 5. Finally, Section 6 shows some conclusions.

2. Review of fuzzy rough sets

Given a nonempty universe \( U \), \( R \) is a fuzzy binary relation on \( U \). If \( R \) satisfies (1) reflexivity: \( R(x,x) = 1 \); (2) symmetry: \( R(x,y) = R(y,x) \); (3) sup-min transitivity: \( R(x,y) \geq \sup_{z \in U} (R(x,z).R(z,y)) \), we say \( R \) is a fuzzy equivalence relation. The fuzzy equivalence class \( [x]_R \) associated with \( x \) and \( R \) is a fuzzy set on \( U \), where \( [x]_R(y) = R(x,y) \) for all \( y \in U \). Based on fuzzy equivalence relations fuzzy rough sets were first introduced in \([10]\).

Definition 1. Let \( U \) be a nonempty universe, \( R \) be a fuzzy equivalence relation on \( U \) and \( F(U) \) be the fuzzy power set of \( U \). Given a fuzzy set \( F \in F(U) \), the lower and upper approximations are defined as

\[
\begin{align*}
RF(x) &= \inf_{y \in U} \max\{1 - R(x,y), F(y)\}, \\
\overline{RF}(x) &= \sup_{y \in U} \min\{R(x,y), F(y)\}.
\end{align*}
\]
The definition of fuzzy rough sets shows that, given a fuzzy set $F \in F(U)$, it will be approximated by a pair of fuzzy sets $RF$ and $R_F$, which is illustrated by Fig. 1. Here, the pair of the fuzzy sets are estimated with finite samples. Given a sample $x$, the fuzzy membership $F(x)$ can be approximated into an interval i.e. $F(x) \in [RF(x), R_F(x)]$ with finite samples, which denotes fuzzy rough set theory can represent some uncertain information with interval data. This reveals the technique of fuzzy rough set theory on the research of handling uncertainty of data, and this technique is used to present fuzzy rough regression analysis principle in Section 3.

Dependency function in fuzzy rough set theory is an important measure to evaluate attributes in dimension reduction. In multidimensional regression tasks, a new fuzzy dependency function for selecting related attributes is proposed in this section.

Given a set of $n$ objects $< U, A \cup Y >$, $U$ is a nonempty universe, and $U = \{(x_i, y_i) | i = 1, 2, \ldots, n\}$. $A$ is the set of conditional attributes, and $Y$ is the regression attribute. Let $x_i$ is $p$-dimensional vector. It means the objects are described with $p$ conditional attributes, i.e. $(x_i, y_i) = (x_{i1}, x_{i2}, \ldots, x_{ip}, y_i)$, where $y_i$ is the output (regression attribute value). The fuzzy dependency function for regression data is defined as follows.

**Definition 2.** Given a regression data set $< U, A \cup Y >$, $U$ is a nonempty universe, $A$ is a set of conditional attributes, and $Y$ is regression attribute. Let $R$ be a fuzzy equivalence relation on $U$ and $F(U)$ be the fuzzy power set of $U$. Given $k$ fuzzy sets $F_1, \ldots, F_k \in F(U)$, for $\forall B \subseteq A$, the fuzzy positive region of regression attribute $Y$ on $B$ is defined as

$$\text{POS}_B(Y) = \bigcup_{i=1}^{k} R_{F_i}.$$  \hfill (2)

The fuzzy dependency of $Y$ on $B$ is defined as

$$\text{FDR}_B(Y) = \frac{|\text{POS}_B(Y)|}{|U|},$$  \hfill (3)

where $|\text{POS}_B(Y)| = \sum_{x \in U} \text{POS}_B(Y)(x)$.

An example of regression data is shown in **Table 1**, where $n = 5$ and $A = \{a_1, a_2, a_3, a_4\}$.

In order to compute the fuzzy dependency with regression data, we first compute the fuzzy partition of data with $k$ fuzzy sets on the regression attribute, shown as **Fig. 2**. Here $k = 3$, and Low, Med and High are three fuzzy sets. Note that all attribute values of data sets are normalized into the interval $[0, 1]$ in the preprocessing step. **Table 2** presents the sub fuzzified regression data set.

For instance, the fuzzy dependency of regression attribute $Y$ on conditional attribute set $B = \{a_1, a_2\}$ is computed as follows.

$$\text{FDR}_B(Y) = \frac{\sum_{i=1}^{5} \text{POS}_B(Y)(x_i)}{5},$$  \hfill (4)

where $\text{POS}_B(Y)(x_i) = \max \{R_{B\text{Low}}(x_i), R_{B\text{Med}}(x_i), R_{B\text{High}}(x_i)\}$.

**Table 1**

<table>
<thead>
<tr>
<th>$x$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
<th>$a_4$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_{11}$</td>
<td>$x_{12}$</td>
<td>$x_{13}$</td>
<td>$x_{14}$</td>
<td>$y_1$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>$x_{21}$</td>
<td>$x_{22}$</td>
<td>$x_{23}$</td>
<td>$x_{24}$</td>
<td>$y_2$</td>
</tr>
<tr>
<td>$x_3$</td>
<td>$x_{31}$</td>
<td>$x_{32}$</td>
<td>$x_{33}$</td>
<td>$x_{34}$</td>
<td>$y_3$</td>
</tr>
<tr>
<td>$x_4$</td>
<td>$x_{41}$</td>
<td>$x_{42}$</td>
<td>$x_{43}$</td>
<td>$x_{44}$</td>
<td>$y_4$</td>
</tr>
<tr>
<td>$x_5$</td>
<td>$x_{51}$</td>
<td>$x_{52}$</td>
<td>$x_{53}$</td>
<td>$x_{54}$</td>
<td>$y_5$</td>
</tr>
</tbody>
</table>

![Fig. 2. Fuzzification.](image-url)
Dependency of fuzzy rough sets is an important evaluation measure of attribute in dimension reduction. In Section 4, we will take the new dependency function as evaluation measure to construct a feature selection algorithm for dimension reduction of multidimensional regression.

3. Fuzzy rough regression analysis

In this section, the principle of regression analysis based on fuzzy rough set theory is presented. It consists of three steps: fuzzy partition, fuzzy approximation and regression value estimation.

3.1. Fuzzy partition

This step is to obtain the fuzzy partition of training set with \( k \) selected fuzzy membership functions on regression attribute.

Given a set of \( n \) input–output training samples \( Tr = \{(x_i, y_i)|i = 1, 2, \ldots, n\} \), we first compute the fuzzy partition of training samples with \( k \) selected fuzzy membership functions \( F_1(x), \ldots, F_k(x) \) on regression attribute. In this way, we can obtain a fuzzy partition with \( k \) fuzzy classes \( F_1, \ldots, F_k \). For instance, in Section 2, Low, Med and High are the three fuzzy classes obtained by fuzzy partition when \( k = 3 \). In this work, the \( k \) selected membership functions are Gaussian functions, And the mean \( \mu_1, \ldots, \mu_k \) of selected functions \( F_1(x), \ldots, F_k(x) \) satisfy \( \mu_1 < \cdots < \mu_k \).

3.2. Fuzzy approximation

Given a set of test samples \( Te = \{z_i|i = 1, 2, \ldots, m\} \), described by a set of conditional attributes. Let \( y_x \) be the unknown regression attribute value of \( z_x \). In this step, the regression attribute value \( y_x \) is approximated into a finite interval. The detail process is illustrated as follows.

We first compute the lower approximation memberships \( RF_1(z_i), \ldots, RF_k(z_i) \) of the test sample \( z_i \in Te \) to fuzzy classes \( F_1, \ldots, F_k \) respectively. Here, \( F_1, \ldots, F_k \) are obtained by fuzzy partition. And then find the fuzzy class \( F_i \) which satisfies

\[
RF_i(z_i) = \max \{ RF_1(z_i), \ldots, RF_k(z_i) \}, F_i \in \{F_1, \ldots, F_k\}. \tag{5}
\]

Finally, the interval that \( y_x \) belongs to is determined. That is

\[
F_i(y_x) \in [RF_i(z_i), RF_i(z_i)]. \tag{6}
\]

Actually, we predict the interval that the membership of \( y_x \) to the determined fuzzy class belongs to.

For example, let Low, Med and High be three fuzzy membership functions. And training set is divided into three fuzzy classes Low, Med, High by fuzzy partition with above three membership functions. Given a test sample \( z_i, y_x \) is supposed to be the unknown value of regression attribute. Now, \( R\text{Low}(z_i), R\text{Med}(z_i), R\text{High}(z_i) \) can be computed with training set. Suppose

\[
R\text{Med}(z_i) = \max \{ R\text{Low}(z_i), R\text{Med}(z_i), R\text{High}(z_i) \}. \tag{7}
\]

the fuzzy membership of \( y_x \) to Med should satisfy

\[
\text{Med}(y_x) = \max \{ \text{Low}(y_x), \text{Med}(y_x), \text{High}(y_x) \}. \tag{8}
\]

Consequently, the possibility of \( y_x \) valued in the interval \( [c, d] \) is the maximal (Fig. 3).

The above computation is the same as classification analysis with the fuzzy rough theory, which has been successfully applied in classification problems in [17].

Clearly, \( y_x \) belongs to the interval \( [c, d] \), and the fuzzy membership of \( y_x \) to the fuzzy set Med is approximated into a certain interval, which is denoted by

\[
\text{Med}(y_x) \in [R\text{Med}(z_i), R\text{Med}(z_i)]. \tag{9}
\]

In this way, the regression attribute values are approximated with intervals according to the fuzzy rough theory.
3.3. Regression attribute value estimation

Now, the regression attribute values of test samples are estimated with the lower and upper limits of intervals which are determined by fuzzy approximation.

We adopt weighted average method to compute the regression attribute value \( y_z \) with \( \overline{RF}_t(z) \) and \( \overline{RF}_t(z) \). The formula is shown as (10).

\[
F_t(y_z) = a \overline{RF}_t(z) + b \overline{RF}_t(z),
\]

(10)

where \( a \) and \( b \) are the weighting factors. They are related to the values of \( \overline{RF}_t(z) \) and \( \overline{RF}_t(z) \). Fig. 1 shows that the larger the \( \overline{RF}_t(x) \) is, the smaller the error \( \overline{RF}_t(x)/C_0 \overline{F}_t(x) \) is, and the larger the error \( \overline{F}_t(x)/C_0 \overline{RF}_t(x) \) is. If \( \overline{RF}_t(x) = 1 \), \( \overline{RF}_t(x) = 0 \), when \( a = 0 \) and \( b = 1 \). Consequently, it is easy to see if \( \overline{RF}_t(z) \) is larger, \( b \) should also be larger, while \( a \) should be smaller. According to the above analysis, we let \( a = \frac{1}{C_0 \overline{RF}_t(z)} \) and \( b = \overline{RF}_t(z) \) in this work.

Now \( F_t(y_z) \) is known. In this work \( F_t(x) (t \in \{1, 2, \ldots, k\}) \) is Gaussian function computed as

\[
F_t(x) = \exp \left( -\frac{(x - \mu_t)^2}{2\sigma_t^2} \right),
\]

(11)

where \( \mu_t \) is the mean, and \( \sigma_t \) is the standard error. Thus, the regression value of \( y_z \) can be computed with inverse function of Gaussian function, i.e.

\[
y_z = \mu_t \pm \sqrt{-2\sigma_t^2 \ln F_t(y_z)}, t \in \{1, \ldots, k\}.
\]

(12)

And the value of \( y_z \) is determined by

\[
y_z = \begin{cases} 
\mu_t - \sqrt{-2\sigma_t^2 \ln F_t(y_z)}, & \overline{RF}_{t-1}(z) > \overline{RF}_{t+1}(z), \\
\mu_t + \sqrt{-2\sigma_t^2 \ln F_t(y_z)}, & \overline{RF}_{t-1}(z) < \overline{RF}_{t+1}(z).
\end{cases}
\]

(13)

Now we use an example to illustrate the effectiveness of Formula (10) with Fig. 4.

Assume \( Med(y_z) \in [RMed(z), \overline{RMed}(z)] \). \( Med(y_z) \) can be estimated with formula (10) i.e.

\[
Med(y_z) = aRMed(z) + b\overline{RMed}(z).
\]

(14)

In Fig. 4, horizontal axis is the distribution of regression attribute value. The \( \text{AVE}(\overline{RMed}, RMed) \) curve is obtained with Formula (14) while \( a = b \). The \( \text{AVE}(aRMed, b\overline{RMed}) \) curve is computed with the weighted average of \( RMed \) and \( \overline{RMed} \), where
weights are $a = 1 - \overline{\text{RMed}}(z_i)$ and $b = \overline{\text{RMed}}(z_i)$, respectively. Fig. 4 shows the AVE$(a\overline{\text{RMed}}, b\overline{\text{RMed}})$ curve is closer to Med than the AVE$(\overline{\text{RMed}}, \overline{\text{RMed}})$ curve, which illustrates Formula (10) is reasonable in predicting fuzzy degree with the lower and upper approximations.

The detailed process of fuzzy rough regression analysis is shown in Table 3.

### 4. Fast fuzzy rough prediction algorithm

If the number of fuzzy subsets $k$ is very large, the time complexity of the prediction algorithm becomes high. In order to address this problem, we combine the technique of rule extraction with fuzzy rough regression analysis for introducing an efficient regression prediction algorithm.

#### 4.1. Regression rule extraction

First, attribute selection is used to reduce dimensions of regression data. The fuzzy dependency $FDR$ of regression attribute on conditional attributes is taken as the attribute evaluation measure, and the sequential forward selection as the search strategy. If the increment $\Delta FDR$ of fuzzy dependency is smaller than $\delta$ ($\delta$ is a given threshold), the attribute selection algorithm stops. The pseudocode of attribute selection is given in Table 4.

#### Attribute selection algorithm for regression data sets.

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Tr = {(x_i, y_i)</td>
<td>i = 1, 2, \ldots, n}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attr</th>
<th>$\delta$-threshold of stopping criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Attr}_1 = {a_1, \ldots, a_q}$ ($q \leq</td>
<td>\text{Attr}</td>
</tr>
<tr>
<td>$FDR_{\text{Attr}<em>1}(Y) = \max</em>{a \in \text{Attr}<em>1} {FDR</em>{\text{Attr}_1}(a)(Y)}$</td>
<td>$\text{FDR}_{\text{Attr}<em>1}(a)(Y) = \max {FDR</em>{\text{Attr}_1}(a)(Y)}$</td>
</tr>
<tr>
<td>$\Delta FDR = FDR_{\text{Attr}<em>1}(a)(Y) - FDR</em>{\text{Attr}_1}(Y)$</td>
<td>$\Delta FDR &gt; \delta$</td>
</tr>
<tr>
<td>$\text{Attr}_1 = \text{Attr}_1 \cup {a_0}$</td>
<td>$\text{Attr}_1 \leftarrow \text{Attr}_1 \cup {a_0}$</td>
</tr>
<tr>
<td>$\text{Attr} = \text{Attr} - {a_0}$</td>
<td>$\text{Attr} = \text{Attr} - {a_0}$</td>
</tr>
<tr>
<td>End</td>
<td>End</td>
</tr>
</tbody>
</table>

---

Table 3

**Fuzzy rough regression analysis.**

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training sets: $Tr = {(x_i, y_i)</td>
<td>i = 1, 2, \ldots, n}$</td>
</tr>
<tr>
<td>Test sets: $Te = {x_i</td>
<td>i = 1, 2, \ldots, m}$</td>
</tr>
<tr>
<td>$k$ fuzzy sets $F_1, F_2, \ldots, F_k$</td>
<td></td>
</tr>
</tbody>
</table>

Table 4

**Attribute selection algorithm for regression data sets.**

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Tr = {(x_i, y_i)</td>
<td>i = 1, 2, \ldots, n}$</td>
</tr>
<tr>
<td>$\text{Attr}$-conditional attribute set</td>
<td></td>
</tr>
<tr>
<td>$\delta$-threshold of stopping criterion</td>
<td></td>
</tr>
</tbody>
</table>
In this way, we can extract a regression rule $R_i$ with each training sample $(x_i, y_i)$ $(i \in \{1, \ldots, n\})$. And the regression rules are formulated as

$$\begin{align*}
\text{IF} & \quad (x_1 \in \text{INT}_{F_1}) \land \ldots \land (x_q \in \text{INT}_{F_q}) \land \ldots \land (x_n \in \text{INT}_{F_n}), \quad \text{THEN} \quad (y_i \in \text{INT}_{F_i}),
\end{align*}$$

where $r, s, p, t \in \{1, 2, \ldots, k\}$.

For each training sample, we can build a regression rule. Finally, we reduce these rules by removing the same rule, and obtain the final rule set $R = \{R_1, \ldots, R_m\}$.

Given a test sample $z_i \in T_e = \{z_i|i = 1, \ldots, m\}$, we first compute the fuzzy memberships $F_t(z_i)$ $(t = 1, \ldots, k, j = 1, \ldots, q)$, and get the following result.

$$(z_{i1} \in \text{INT}_{F_{i1}}) \land \ldots \land (z_{iq} \in \text{INT}_{F_{iq}}).$$

According to regression rules extracted from training set, we can estimate the interval that $y_{z_i}$ belongs to and the corresponding fuzzy subset $F_t$. For instance,

$$y_{z_i} \in \text{INT}_{F_t}.$$ 

The purpose of the above process is to determine the intervals that regression attribute values of test samples belong to and the corresponding fuzzy subsets. In this way, we only need to compute the $RF_t(z_i)$ and $RF_t(z_i)$ for estimating the regression attribute value $y_{z_i}$.

### 4.2 Regression prediction algorithm

Now we present the regression algorithm for fuzzy rough prediction by combining the fuzzy rough regression principle with fuzzy rough regression rule extraction. Given a test sample $z_i \in T_e, y_{z_i}$ is supposed to be the regression attribute value of $z_i$. The prediction process of $y_{z_i}$ is summarized as follows.

Firstly, extract regression rules, denoted by $R = \{R_1, \ldots, R_m\}$. Secondly, determine the interval of regression attribute value of each test sample $z_i$ and corresponding fuzzy subset $F_t$. And then compute the memberships of lower and upper approximations of $z_i$ to the fuzzy set $F_t$ i.e. $RF_t(z_i)$ and $RF_t(z_i)$. Finally, estimate $y_{z_i}$ with method in Section 3.3.

The pseudocode of the fuzzy rough regression prediction algorithm is shown in Table 5.

According to introducing regression rule extraction, the complexity of fuzzy rough prediction is reduced. If we predict the regression attribute value without regression rules, we need to compute $k$ lower approximation memberships of the test sample to $k$ fuzzy subsets. And the time complexity of fuzzy rough prediction is $O(kn^3)$. If $k \rightarrow n, O(kn^3) \rightarrow O(n^3)$. By introducing regression rules, we only need to compute the lower and upper approximation memberships of the test sample to the fuzzy subset determined by regression rules. Now the complexity of fuzzy rough prediction is $O(n^2)$.

### 5. Experimental analysis

In this section, we conduct some experiments to test fuzzy rough regression principle and algorithm on several regression data sets, including wind speed prediction and other public tasks.

#### 5.1 Experiments on UCI data

The summary of data sets is shown in Table 6.

Our purpose is to test the correlation between real regression attribute values and test regression attribute values with FRRPA. The experiments adopt leave-one-out cross-validation [47] method. Given a sample set containing $n$ samples...
described by $p$ attributes, leave-one-out cross-validation is to predict the regression value of a sample with the other $n-1$ samples using the proposed fuzzy rough regression prediction algorithm.

We first give an experiment to illustrate the basic idea of the fuzzy rough approximation principle on data machinecpu. The experimental results are shown in Fig. 5. In each subfigure, there are three lines, RM, LAM and UAM. RM denotes fuzzy membership curve of the real regression attribute values of test samples. LAM and UAM denote the lower and upper approximation curves of the real fuzzy membership curve, respectively. The first subfigure shows the LAM, RM and UAM of the first 100 samples in machinecpu data set, and second subfigure shows the LAM, RM and UAM of the rest 109 samples in machinecpu data set.

Fig. 5 illustrates that most of the real memberships of test samples are approximated into the area surrounded by LAM and UAM, which indicates fuzzy lower and upper approximation operators can be used to approximate the real values of regression values, and give intervals that real values belong to. Furthermore, there are several real fuzzy degrees of regression attribute values of test samples are out of the intervals that are approximated by lower and upper approximations. These instances appear due to the existing of noise in data sets, to which fuzzy lower and upper approximations are sensitive. In order to address this problem, we are going to introduce the robust fuzzy rough set models to establish the robust fuzzy rough regression models in the future.

Fig. 6 gives the experimental results produced with FRRPA on machinecpu, pyrim, stock and housing data sets. There are two curves of regression attribute values, and they are the real attribute values and the predicted values of regression attribute. Obviously, the better the consistency of the two curves is, the higher the prediction precision of FRRPA is. It is shown that the predictions are accurate on machinecpu and stock.

Besides, we also conduct some experiments to compare FRRPA with GP (Gaussian Processes) [39], LR (Linear Regression), MLP (Multilayer Perceptron) [28], SLR (Simple Linear Regression) [13], AR (Additive Regression) [12], RBD (Regression By Discretization) [11], RT (REP Tree) [40] and FRNN [21] algorithms.

<table>
<thead>
<tr>
<th>Data</th>
<th>Samples</th>
<th>Conditional attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>abalone</td>
<td>4177</td>
<td>9</td>
</tr>
<tr>
<td>autoMPG8</td>
<td>9517</td>
<td>7</td>
</tr>
<tr>
<td>bank8FM</td>
<td>8192</td>
<td>9</td>
</tr>
<tr>
<td>bank32nH</td>
<td>8192</td>
<td>33</td>
</tr>
<tr>
<td>elevators9</td>
<td>16,599</td>
<td>19</td>
</tr>
<tr>
<td>housing</td>
<td>506</td>
<td>14</td>
</tr>
<tr>
<td>machinecpu</td>
<td>209</td>
<td>7</td>
</tr>
<tr>
<td>puma32H</td>
<td>8192</td>
<td>33</td>
</tr>
<tr>
<td>pyrim</td>
<td>74</td>
<td>28</td>
</tr>
<tr>
<td>stock</td>
<td>950</td>
<td>10</td>
</tr>
<tr>
<td>triazines</td>
<td>186</td>
<td>61</td>
</tr>
</tbody>
</table>
GP: Implements Gaussian processes for regression without hyperparameter-tuning.
LR: Linear regression prediction algorithm.
MLP: A regression algorithm that uses backpropagation to predict the regression values of instances.
SLR: A simple Linear Regression.
AR: A regression algorithm that selects regression coefficient by iterativeness.
RBD: A regression scheme that employs any classifier on a copy of the data that has the class attribute discretized.
RT: A fast regression tree learner using reduced-error pruning (with backfitting).
FRNN: A fuzzy rough nearest neighbors for classification and prediction.

The results are shown in Table 7. For each regression algorithm, we take leave-one-out cross-validation method to predict the regression values, and compute the correlation coefficient between test values and real values. Correlation coefficient is computed with the following formula.

\[ \text{Corr}_\text{coeff}(W, W') = \frac{\sum (w_i - \bar{W})(w'_i - \bar{W}')} {\sqrt{\sum (w_i - \bar{W})^2 \sum (w'_i - \bar{W}')^2}}, \]  

(19)
where \( W = \{w_1, w_2, \ldots, w_m\} \) and \( W' = \{w'_1, w'_2, \ldots, w'_m\} \) are respectively real regression values and test regression values. \( \text{CorrCoeff}(W, W') \) takes values in \([-1, 1]\). Large correlation coefficients means the predicted values are highly correlated with the real values.

Table 7 shows that FRRPA produces the best performance among all the algorithms on bank32NH, machinecpu, pyrim, triazines, and has the largest average of correlation coefficient. Furthermore, FRRPA has better performance than GP on ten data sets, LR on seven data sets, MLP on seven data sets, SLR on nine data sets, AR on eight data sets, RBD on seven data sets, RT on six data sets, FRNN on seven data sets. The above results show FRRPA is effective.

Besides, apart from using the correlation coefficient for evaluating the performance of FRRPA, we consider another evaluation measure, i.e. root mean square error. The results are shown in Table 8. Table 8 shows FRRPA performs better than GP on eight data sets, LR on eight data sets, MLP on three data sets, SLR on nine data sets, AR on six data sets, RBD on four data sets, RT on five data sets, FRNN on four data sets.

Furthermore, rank-sum test method is used for testing differences between the above prediction methods, where confidence level is 0.05. Here, the eleven results computed with a prediction algorithm compose a sample. Due to the sample size equals 11, so the test uses the following formula.

\[
Z = \frac{T - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}}}
\]  

(20)

where \( n_1 \) and \( n_2 \) are the sample size, and \( T \) is the rank-sum of the smaller sample.

With the rank-sum test we test the differences between FRRPA and other methods respectively using the results in Tables 7 and 8. The confidence degree is 0.05, and degree of freedom is \( n_1 + n_2 - 2 = 20 \). The test results (values of \( Z \)) are shown in Table 9.
Table 9, |Z| < t_{20}(0.05) = 1.725 means there is no significance difference between FRRPA and a method, else there are significance difference between FRRPA and the method. Table 9 shows there are not any significance differences between FRRPA and LR, MLP, AR, RBD, RT and FRNN with results of correlation coefficient. And there are not any significance differences between FRRPA and other methods with results of root mean square error. The above analysis indicates FRRPA is effective.

5.2. Wind speed prediction

We collect a set of wind-speed data from a Chinese ind farm. The data set contains 62,466 samples collected during 434 days, and each sample is the average of wind-speed during 10 min.

In this work, we transform the wind-speed data set into the format shown in Table 10, where y(t_i) is the wind-speed on current second t_i, y(t_i – Δt) is the wind-speed before Δt minutes, and y(t_i + kΔt) is the wind-speed in kΔt minutes. In this work, Δt = 10 minutes.

Here, we also take leave-one-out cross-validation method to test FRRPA algorithm in predicting the wind-speed on the next second, i.e. y(t_i + Δt) with y(t_i – sΔt), …, y(t_i – Δt), y(t_i). Here, 1000 seriate samples are selected to be used in the demonstration experiment, where s = 10 and k = 1. The prediction performance of FRRPA is shown in Fig. 7.

The two curves are the real wind-speed values and predicted wind-speed values. The figure shows that the test value curve is consistent with the real value curve, where the root mean square error between the two curves is 0.0449. This illustrates FRRPA is effective in wind-speed forecast.

Table 10
Illustration of wind-speed data set.

<table>
<thead>
<tr>
<th>y(t_i – sΔt)</th>
<th>…</th>
<th>y(t_i – Δt)</th>
<th>y(t_i)</th>
<th>y(t_i + Δt)</th>
<th>…</th>
<th>y(t_i + kΔt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>y(t_i – sΔt)</td>
<td>…</td>
<td>y(t_i – Δt)</td>
<td>y(t_i)</td>
<td>y(t_i + Δt)</td>
<td>…</td>
<td>y(t_i + kΔt)</td>
</tr>
<tr>
<td>y(t_i – sΔt)</td>
<td>…</td>
<td>y(t_i – Δt)</td>
<td>y(t_i)</td>
<td>y(t_i + Δt)</td>
<td>…</td>
<td>y(t_i + kΔt)</td>
</tr>
<tr>
<td>…</td>
<td></td>
<td>…</td>
<td>…</td>
<td>…</td>
<td></td>
<td>…</td>
</tr>
<tr>
<td>y(t_{i-1} – sΔt)</td>
<td>…</td>
<td>y(t_{i-1} – Δt)</td>
<td>y(t_{i-1})</td>
<td>y(t_{i-1} + Δt)</td>
<td>…</td>
<td>y(t_{i-1} + kΔt)</td>
</tr>
<tr>
<td>y(t_{i-2} – sΔt)</td>
<td>…</td>
<td>y(t_{i-2} – Δt)</td>
<td>y(t_{i-2})</td>
<td>y(t_{i-2} + Δt)</td>
<td>…</td>
<td>y(t_{i-2} + kΔt)</td>
</tr>
</tbody>
</table>

![Fig. 7. Wind-speed prediction.](image-url)

Table 11
Root mean square error between real values and test values (Wind-speed data).

<table>
<thead>
<tr>
<th>Data</th>
<th>FRRPA</th>
<th>GP</th>
<th>LR</th>
<th>MLP</th>
<th>SLR</th>
<th>AR</th>
<th>RBD</th>
<th>RT</th>
<th>FRNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>10min[k=1]</td>
<td>0.0698</td>
<td>0.2251</td>
<td>0.0365</td>
<td>0.0400</td>
<td>0.0366</td>
<td>0.0972</td>
<td>0.1096</td>
<td>0.0795</td>
<td>0.0932</td>
</tr>
<tr>
<td>30min[k=3]</td>
<td>0.0850</td>
<td>0.2295</td>
<td>0.0619</td>
<td>0.0721</td>
<td>0.0614</td>
<td>0.1344</td>
<td>0.1248</td>
<td>0.1006</td>
<td>0.1129</td>
</tr>
<tr>
<td>60min[k=6]</td>
<td>0.1057</td>
<td>0.2355</td>
<td>0.0875</td>
<td>0.0894</td>
<td>0.0867</td>
<td>0.1593</td>
<td>0.2117</td>
<td>0.1245</td>
<td>0.1506</td>
</tr>
<tr>
<td>2hours[k=12]</td>
<td>0.1519</td>
<td>0.2486</td>
<td>0.1264</td>
<td>0.1180</td>
<td>0.1264</td>
<td>0.2250</td>
<td>0.2410</td>
<td>0.2301</td>
<td>0.1934</td>
</tr>
<tr>
<td>3hours[k=18]</td>
<td>0.1840</td>
<td>0.2613</td>
<td>0.1651</td>
<td>0.1534</td>
<td>0.1651</td>
<td>0.2316</td>
<td>0.2371</td>
<td>0.2037</td>
<td>0.2021</td>
</tr>
<tr>
<td>4hours[k=24]</td>
<td>0.1939</td>
<td>0.2681</td>
<td>0.1952</td>
<td>0.1937</td>
<td>0.1952</td>
<td>0.2311</td>
<td>0.2369</td>
<td>0.2183</td>
<td>0.2247</td>
</tr>
<tr>
<td>5hours[k=30]</td>
<td>0.2483</td>
<td>0.2862</td>
<td>0.2247</td>
<td>0.1822</td>
<td>0.2247</td>
<td>0.3049</td>
<td>0.3006</td>
<td>0.2703</td>
<td>0.2418</td>
</tr>
<tr>
<td>6hours[k=36]</td>
<td>0.2528</td>
<td>0.2982</td>
<td>0.2422</td>
<td>0.1637</td>
<td>0.2422</td>
<td>0.2887</td>
<td>0.3190</td>
<td>0.2893</td>
<td>0.2542</td>
</tr>
<tr>
<td>7hours[k=42]</td>
<td>0.2588</td>
<td>0.3077</td>
<td>0.2567</td>
<td>0.1885</td>
<td>0.2567</td>
<td>0.3040</td>
<td>0.3331</td>
<td>0.3024</td>
<td>0.2670</td>
</tr>
<tr>
<td>8hours[k=48]</td>
<td>0.2646</td>
<td>0.3162</td>
<td>0.2704</td>
<td>0.2038</td>
<td>0.2704</td>
<td>0.3293</td>
<td>0.3474</td>
<td>0.2748</td>
<td>0.2734</td>
</tr>
<tr>
<td>9hours[k=54]</td>
<td>0.2617</td>
<td>0.3253</td>
<td>0.2836</td>
<td>0.2438</td>
<td>0.2836</td>
<td>0.3541</td>
<td>0.3242</td>
<td>0.3138</td>
<td>0.2793</td>
</tr>
<tr>
<td>10hours[k=60]</td>
<td>0.2589</td>
<td>0.3329</td>
<td>0.3007</td>
<td>0.2442</td>
<td>0.2986</td>
<td>0.3558</td>
<td>0.3406</td>
<td>0.3317</td>
<td>0.2954</td>
</tr>
<tr>
<td>AVE</td>
<td>0.1947</td>
<td>0.2779</td>
<td>0.1876</td>
<td>0.1577</td>
<td>0.1873</td>
<td>0.2513</td>
<td>0.2605</td>
<td>0.2282</td>
<td>0.2157</td>
</tr>
<tr>
<td>Stat. test</td>
<td>0</td>
<td>−2.7135</td>
<td>0.0577</td>
<td>1.5011</td>
<td>0.0577</td>
<td>−1.5588</td>
<td>−1.7321</td>
<td>−1.3856</td>
<td>−0.8660</td>
</tr>
</tbody>
</table>
However, in the study of wind-speed prediction, the purpose is to predict the wind speed in the future with current and past wind-speed values. To test the performance of FRRPA further, we take \( n \) samples as training set to predict the wind-speed value in \( k \times 10 \) (where \( k = 1, 2, \ldots, 60 \)) minutes. For example, if we want to predict the wind-speed value in \( k \times 10 \) minutes, the format of training data set can be denoted as

\[
Tr = \{ (x_i, y_i) | x_i = [y(t_i - s \Delta t), \ldots, y(t_i)], y_i = y(t_i + k \Delta t), i = 1, \ldots, n \}. \tag{21}
\]

In this work, with above method we construct 12 data sets with \( s = 9, n = 1000, \) and \( k = 1, 3, 6, 12, 18, 24, 30, 36, 42, 48, 54, 60 \). For each data set, 800 samples are taken as training samples, and 200 samples are taken as test samples. The results of root mean square error between real values and test values of test samples are shown in Table 11.

AVE means the average of all root mean square error computed using the same method. The smaller the average is, the better the method is. The average shows FRRPA performs better performance than other methods except LR, MLP and SLR algorithm. According to the boundary value tables, \( t_{20}0.05 = 1.717 \). From Table 11, we can see the statistical test results for GP and RBD satisfy \( |Z| > 1.717 \), which shows there are not any significance differences between FRRPA and other algorithms except GP and RBD methods. It also shows FRRPA produces better performance than the GP and RBD methods. The above experimental analysis shows FRRPA is effective.

6. Conclusions

Fuzzy rough set theory has been widely applied in constructing classifiers, attribute selection and so on. In this work, this theory is used to establish fuzzy rough regression prediction algorithm due to its frame of reference of fuzzy lower and upper approximations. The main contributions of this paper are summarized as follows.

A fuzzy rough regression analysis principle is presented, and it consists of three steps which are fuzzy partition, fuzzy approximation and regression attribute value estimation. Fuzzy partition on regression attribute is used to compute the fuzzy partition of training sample set with \( k \) fuzzy classes. And each regression attribute value is determined in a finite interval in the step of fuzzy approximation. Finally, the regression attribute value is estimated with interval lower and upper limits.

A fuzzy rough regression prediction algorithm is proposed by combining fuzzy rough regression rule extraction and fuzzy rough regression prediction principle, where extraction of regression rules is used to reduce the complexity of the prediction algorithm.

The numerical experiments are conducted to test the fuzzy rough regression prediction algorithm both on eleven UCI data sets and twelve constructed wind-speed data sets. The results show the new algorithm has better average performance by taking correlation coefficient and root mean square error as evaluation measures. And statistical test results indicate there are not any significance differences between fuzzy rough regression prediction algorithm and the state-of-the-art regression algorithms compared. All the experimental results show the fuzzy rough regression prediction algorithm is effective.

Acknowledgments

This work is partially supported by National Program on Key Basic Research Project (2012CB215201), National Natural Science Foundation of China (61222210, 61202259 and 61071179), Natural Science Foundation of Hebei Province of China (F2013501052), Science Foundation of Ministry of Education of China (N120423006), and China Postdoctoral Science Foundation (2013M530874).

References