Information Sciences 282 (2014) 388–400

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Fuzzy rough regression with application to wind speed prediction

Shuang An [a,b], Hong Shi [b], Qinghua Hu [b,*], Xiaoqi Li [a], Jianwu Dang [b]

[a] Northeastern University, Shenyang 110004, PR China
[b] Tianjin University, Tianjin 300072, PR China

ARTICLE INFO

Article history:
Received 19 January 2013
Received in revised form 11 March 2014
Accepted 15 March 2014
Available online 16 April 2014

Keywords:
Fuzzy rough set
Regression analysis
Fuzzy partition
Fuzzy approximation
Wind speed prediction

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.

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. $\underline{R}F$ and $\overline{R}F$. 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 $\underline{R}F(x) \leqslant F(x)$ for any samples, and $F(x) \leqslant \overline{R}F(x)$ for

* 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/

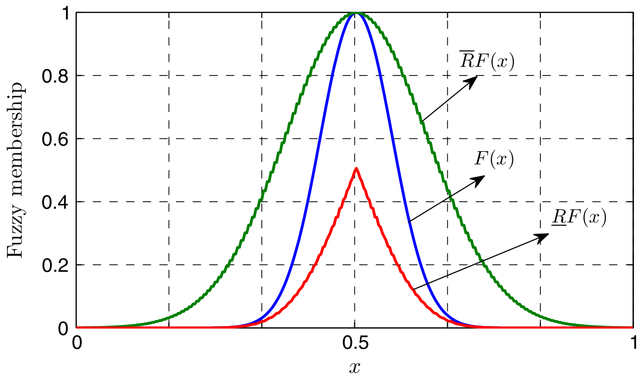

Fig. 1. Fuzzy rough approximations.

all samples. Consequently, $F(x) \in [\underline{R}F(x), \overline{R}F(x)]$. 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. The new method consists of three modules: fuzzy partition, fuzzy approximation and regression attribute value estimation. Firstly, we compute the fuzzy partition of training samples with k fuzzy classes on regression attribute. Secondly, for $\forall \mathbf{z}_0 \in Te$ (Te denotes test sample set), the membership of regression attribute value y_0 of the test sample $\mathbf{z}_0$ to the determined fuzzy set F_t can be fixed in a finite interval, i.e. $F_t(y_0) \in [\underline{R}F_t(\mathbf{z}_0), \overline{R}F_t(\mathbf{z}_0)]$, where F_t satisfies $\underline{R}F_t(\mathbf{z}_0) = \max\{\underline{R}F_i(\mathbf{z}_0) | i = 1, \ldots, k\}$. Finally, the regression attribute value y_0 of the test sample $\mathbf{z}_0$ is estimated with the weighted average of the lower and upper limits of the interval.

With the above analysis, a fuzzy rough regression prediction algorithm (FRRPA) is established by combining regression rule extraction and fuzzy rough prediction. We conduct some numerical experiments on wind speed prediction and other open data sets [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: $R(x,y) \geqslant \text{supmin}_{z \in U}\{R(x,z), R(z,y)\}$, 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

$$\begin{cases} \underline{R}F(x) = \inf\limits_{y \in U} \max\{1 - R(x,y), F(y)\}, \\ \overline{R}F(x) = \sup\limits_{y \in U} \min\{R(x,y), F(y)\}. \end{cases} \tag{1}$$

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 i.e. $\underline{R}F$ and $\overline{R}F$, 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. $F(x) \in [\underline{R}F(x), \overline{R}F(x)]$ 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 $U = \{(\mathbf{x}_i, y_i) | i = 1, 2, \ldots, n\}$. A is the set of conditional attributes, and Y is the regression attribute. $\mathbf{x}_i$ is p-dimensional vector. It means the objects are described with p conditional attributes, i.e. $(\mathbf{x}_i, y_i) = (x_{i1}, x_{i2}, \ldots, x_{ip}, y_i)$, 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

$$\mathrm{POS}_B(Y) = \bigcup_{i=1,\ldots,k} \underline{R_B} F_i. \tag{2}$$

The fuzzy dependency of Y on B is defined as

$$FDR_B(Y) = \frac{|\mathrm{POS}_B(Y)|}{|U|}, \tag{3}$$

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

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.

$$FDR_B(Y) = \frac{\sum_{i=1}^{5} \mathrm{POS}_B(Y)(\mathbf{x}_i)}{5}, \tag{4}$$

where $\mathrm{POS}_B(Y)(\mathbf{x}_i) = \max\{\underline{R_B}Low(\mathbf{x}_i), \underline{R_B}Med(\mathbf{x}_i), \underline{R_B}High(\mathbf{x}_i)\}$.

Table 1
Example of regression task.

X	a_1	a_2	a_3	a_4	Y
$\mathbf{x}_1$	x_{11}	x_{12}	x_{13}	x_{14}	y_1
$\mathbf{x}_2$	x_{21}	x_{22}	x_{23}	x_{24}	y_2
$\mathbf{x}_3$	x_{31}	x_{32}	x_{33}	x_{34}	y_3
$\mathbf{x}_4$	x_{41}	x_{42}	x_{43}	x_{44}	y_4
$\mathbf{x}_5$	x_{51}	x_{52}	x_{53}	x_{54}	y_5

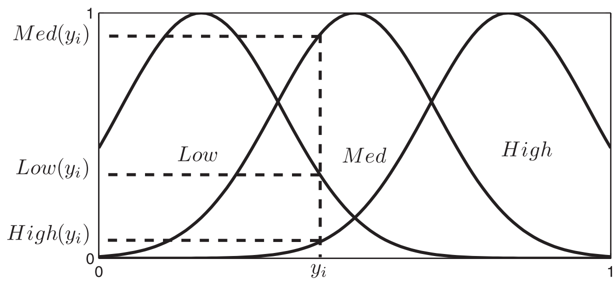

Fig. 2. Fuzzification.

Table 2
Example of fuzzified regression data set.

X	a_1	a_2	a_3	a_4	$Low(y_i)$	$Med(y_i)$	$High(y_i)$
$\mathbf{x}_1$	x_{11}	x_{12}	x_{13}	x_{14}	$Low(y_1)$	$Med(y_1)$	$High(y_1)$
$\mathbf{x}_2$	x_{21}	x_{22}	x_{23}	x_{24}	$Low(y_2)$	$Med(y_2)$	$High(y_2)$
$\mathbf{x}_3$	x_{31}	x_{32}	x_{33}	x_{34}	$Low(y_3)$	$Med(y_3)$	$High(y_3)$
$\mathbf{x}_4$	x_{41}	x_{42}	x_{43}	x_{44}	$Low(y_4)$	$Med(y_4)$	$High(y_4)$
$\mathbf{x}_5$	x_{51}	x_{52}	x_{53}	x_{54}	$Low(y_5)$	$Med(y_5)$	$High(y_5)$

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 $Tr = \{(\mathbf{x}_i, y_i) | i = 1, 2, \ldots, n\}$, 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 $Te = \{\mathbf{z}_i | i = 1, 2, \ldots, m\}$, described by a set of conditional attributes. Let $y_{\mathbf{z}_i}$ be the unknown regression attribute value of $\mathbf{z}_i$. In this step, the regression attribute value $y_{\mathbf{z}_i}$ is approximated into a finite interval. The detail process is illustrated as follows.

We first compute the lower approximation memberships $\underline{R}F_1(\mathbf{z}_i), \ldots, \underline{R}F_k(\mathbf{z}_i)$ of the test sample $\mathbf{z}_i \in Te$ 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_t which satisfies

$$\underline{R}F_t(\mathbf{z}_i) = \max\{\underline{R}F_1(\mathbf{z}_i), \ldots, \underline{R}F_k(\mathbf{z}_i)\}, F_t \in \{F_1, \ldots, F_k\}. \tag{5}$$

Finally, the interval that $y_{\mathbf{z}_i}$ belongs to is determined. That is

$$F_t(y_{\mathbf{z}_i}) \in \left[\underline{R}F_t(\mathbf{z}_i), \overline{R}F_t(\mathbf{z}_i)\right]. \tag{6}$$

Actually, we predict the interval that the membership of $y_{\mathbf{z}_i}$ to the determined fuzzy class belongs to.

For example, let $Low(y), Med(y)$ and $High(y)$ 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 $\mathbf{z}_i, y_{\mathbf{z}_i}$ is supposed to be the unknown value of regression attribute. Now, $\underline{R}Low(\mathbf{z}_i), \underline{R}Med(\mathbf{z}_i), \underline{R}High(\mathbf{z}_i)$ can be computed with training set.

Suppose

$$\underline{R}Med(\mathbf{z}_i) = \max\{\underline{R}Low(\mathbf{z}_i), \underline{R}Med(\mathbf{z}_i), \underline{R}High(\mathbf{z}_i)\}, \tag{7}$$

the fuzzy membership of $y_{\mathbf{z}_i}$ to Med should satisfy

$$Med(y_{\mathbf{z}_i}) = \max\left\{Low(y_{\mathbf{z}_i}), Med(y_{\mathbf{z}_i}), High(y_{\mathbf{z}_i})\right\}. \tag{8}$$

Consequently, the possibility of $y_{\mathbf{z}_i}$ 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, $y_{\mathbf{z}_i}$ belongs to the interval $[c, d]$, and the fuzzy membership of $y_{\mathbf{z}_i}$ to the fuzzy set Med is approximated into a certain interval, which is denoted by

$$Med(y_{\mathbf{z}_i}) \in \left[\underline{R}Med(\mathbf{z}_i), \overline{R}Med(\mathbf{z}_i)\right]. \tag{9}$$

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

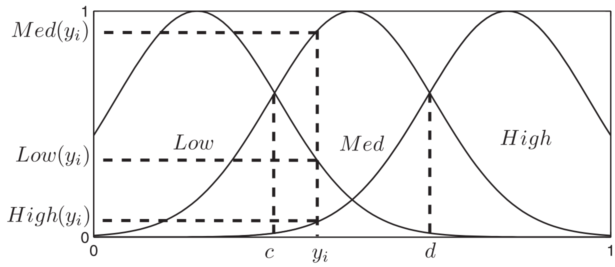

Fig. 3. Interval estimation.

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 $y_{\mathbf{z}_i}$ with $\underline{R}F_t(\mathbf{z}_i)$ and $\overline{R}F_t(\mathbf{z}_i)$. The formula is shown as (10).

$$F_t(y_{\mathbf{z}_i}) = a\underline{R}F_t(\mathbf{z}_i) + b\overline{R}F_t(\mathbf{z}_i), \tag{10}$$

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

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

$$F_t(x) = \exp\left(-\frac{(x-\mu_t)^2}{2\sigma_t^2}\right), \tag{11}$$

where μ_t is the mean, and σ_t is the standard error. Thus, the regression value of $y_{\mathbf{z}_i}$ can be computed with inverse function of Gaussian function, i.e.

$$y_{\mathbf{z}_i} = \mu_t \pm \sqrt{-2\sigma_t^2 \ln F_t(y_{\mathbf{z}_i})}, t \in \{1, \ldots, k\}. \tag{12}$$

And the value of $y_{\mathbf{z}_i}$ is determined by

$$y_{\mathbf{z}_i} = \begin{cases} \mu_t - \sqrt{-2\sigma_t^2 \ln F_t(y_{\mathbf{z}_i})}, \underline{R}F_{t-1}(\mathbf{z}_i) > \underline{R}F_{t+1}(\mathbf{z}_i), \\ \mu_t + \sqrt{-2\sigma_t^2 \ln F_t(y_{\mathbf{z}_i})}, \underline{R}F_{t-1}(\mathbf{z}_i) < \underline{R}F_{t+1}(\mathbf{z}_i). \end{cases} \tag{13}$$

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

Assume $Med(y_{\mathbf{z}_i}) \in [\underline{R}Med(\mathbf{z}_i), \overline{R}Med(\mathbf{z}_i)], Med(y_{\mathbf{z}_i})$ can be estimated with formula (10) i.e.

$$Med(y_{\mathbf{z}_i}) = a\underline{R}Med(\mathbf{z}_i) + b\overline{R}Med(\mathbf{z}_i). \tag{14}$$

In Fig. 4, horizontal axis is the distribution of regression attribute value. The AVE($\underline{R}Med$, $\overline{R}Med$) curve is obtained with Formula (14) while $a = b$. The AVE($a\underline{R}Med, b\overline{R}Med$) curve is computed with the weighted average of $\underline{R}Med$ and $\overline{R}Med$, where

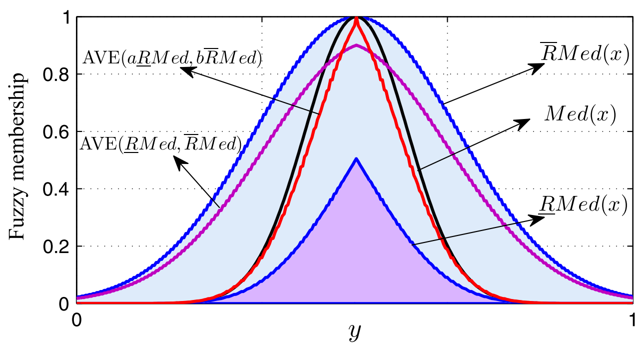

Fig. 4. Regression attribute value estimation.

Table 3
Fuzzy rough regression analysis.

Input	Training sets: $Tr = \{(\mathbf{x}_i, y_i) \mid i = 1, 2, \ldots, n\}$ Test sets: $Te = \{\mathbf{z}_i \mid i = 1, 2, \ldots, m\}$ k fuzzy sets $F_1, F_2, \ldots, F_k$
	For each test sample $\mathbf{z}_i$
	Compute $\underline{R}F_j(\mathbf{z}_i)(j = 1, \ldots, k)$;
	$F_t(y_{\mathbf{z}_i}) = \max\{\underline{R}F_1(\mathbf{z}_i), \underline{R}F_2(\mathbf{z}_i), \ldots, \underline{R}F_k(\mathbf{z}_i)\}$;
	Compute $\overline{R}F_t(\mathbf{z}_i)$;
	$F_t(y_{\mathbf{z}_i}) = a\underline{R}F_t(\mathbf{z}_i) + b\overline{R}F_t(\mathbf{z}_i)$;
	$y_{\mathbf{z}_i} = \begin{cases} \mu_{F_t} - \sqrt{-2\sigma_{F_t}^2 \ln F_t(y_{\mathbf{z}_i})}, \underline{R}F_{t-1}(\mathbf{z}_i) > \underline{R}F_{t+1}(\mathbf{z}_i); \\ \mu_{F_t} + \sqrt{-2\sigma_{F_t}^2 \ln F_t(y_{\mathbf{z}_i})}, \underline{R}F_{t-1}(\mathbf{z}_i) < \underline{R}F_{t+1}(\mathbf{z}_i). \end{cases}$
	end
Output	Regression attribute values of test samples: $Y = \{y_{\mathbf{z}_1}, y_{\mathbf{z}_2}, \ldots, y_{\mathbf{z}_m}\}$

weights are $a = 1 - \overline{R}Med(\mathbf{z}_i)$ and $b = \overline{R}Med(\mathbf{z}_i)$, respectively. Fig. 4 shows the AVE$(a\underline{R}Med, b\overline{R}Med)$ curve is closer to Med than the AVE$(\underline{R}Med, \overline{R}Med)$ 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 ΔFDR of fuzzy dependency is smaller than δ (δ is a given threshold), the attribute selection algorithm stops. The pseudocode of attribute selection is given in Table 4.

Assume $Attr_1 = \{a_1, .., a_q\}$ $(q \leqslant |Attr|)$ is the selected attribute subset. We first divide the domain of each attribute into k intervals. That is to say for each training sample $(\mathbf{x}_i, y_i) = (x_{i1}, x_{i2}, \ldots, x_{iq}, y_i)$, we first determine the intervals that each attribute value belongs to. The method is illustrated as follows.

Given a training sample $(\mathbf{x}_i, y_i) = (x_{i1}, \ldots, x_{ij}, \ldots, x_{iq}, y_i)$, x_{ij} is the attribute value of $\mathbf{x}_i$ on the attribute a_j, and y_i is the regression attribute value. With a_j, we first compute the fuzzy partition of training samples with k fuzzy subsets $F_1, \ldots, F_k$. And x_{ij} can be partitioned into the interval $\mathrm{INT}_{F_s} (s \in \{1, 2, \ldots, k\})$, where F_s satisfies $F_s(x_{ij}) = \max\{F_1(x_{ij}), \ldots, F_k(x_{ij})\}$. In this way, $(\mathbf{x}_i, y_i)$ can be described as

$$\left(x_{i1} \in \mathrm{INT}_{F_r}, \ldots, x_{ij} \in \mathrm{INT}_{F_s}, \ldots, x_{iq} \in \mathrm{INT}_{F_p}, y_i \in \mathrm{INT}_{F_t}\right), \tag{15}$$

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

Table 4
Attribute selection algorithm for regression data sets.

Input	$Tr = \{(\mathbf{x}_i, y_i) \mid i = 1, 2, \ldots, n\}$ $Attr$-conditional attribute set δ-threshold of stopping criterion
	Initialize $Attr_1 = \phi$
	While $\Delta FDR > \delta$
	$FDR_{Attr_1 \cup \{a_0\}}(Y) = \max_{a_i \in Attr}\{FDR_{Attr_1 \cup \{a_i\}}(Y)\}$
	$\Delta FDR = FDR_{Attr_1 \cup \{a_0\}}(Y) - FDR_{Attr_1}(Y)$
	If $\Delta FDR > \delta$
	$Attr_1 = Attr_1 \cup \{a_0\}$,
	$Attr = Attr - \{a_0\}$
	End
	End
Output	Attribute subset $Attr_1$

Table 5
Fuzzy rough regression prediction algorithm (FRRPA).

Input	Training sets: $Tr = \{(\mathbf{x}_i, y_i) \mid i = 1, 2, \ldots, n\}$ Test sets: $Te = \{\mathbf{z}_i \mid i = 1, 2, \ldots, m\}$ k fuzzy sets $F_1, F_2, \ldots, F_k$
1	Extract regression rules $R_1, \ldots, R_{n_1}$ with Tr;
2	For each test sample $\mathbf{z}_i$
3	Determine the fuzzy set F_t with regression rules;
4	Compute $\underline{R}F_t(\mathbf{z}_i)$ and $\overline{R}F_t(\mathbf{z}_i)$;
5	$F_t(y_{\mathbf{z}_i}) = a\underline{R}F_t(\mathbf{z}_i) + b\overline{R}F_t(\mathbf{z}_i)$;
6	$y_{\mathbf{z}_i} = \begin{cases} \mu_{F_t} - \sqrt{-2\sigma_{F_t}^2 \ln F_t(y_{\mathbf{z}_i})}, \underline{R}F_{t-1}(\mathbf{z}_i) > \underline{R}F_{t+1}(\mathbf{z}_i); \\ \mu_{F_t} + \sqrt{-2\sigma_{F_t}^2 \ln F_t(y_{\mathbf{z}_i})}, \underline{R}F_{t-1}(\mathbf{z}_i) < \underline{R}F_{t+1}(\mathbf{z}_i). \end{cases}$
7	end
Output	Regression values of test samples: $Y = \{y_{\mathbf{z}_1}, y_{\mathbf{z}_2}, \ldots, y_{\mathbf{z}_m}\}$

In this way, we can extract a regression rule R_i with each training sample $(\mathbf{x}_i, y_i)$ $(i \in \{1, \ldots, n\})$. And the regression rules are formulated as

$$\text{IF} \quad (x_{i1} \in \text{INT}_{F_r}) \wedge \ldots \wedge (x_{ij} \in \text{INT}_{F_s}) \wedge \ldots \wedge (x_{iq} \in \text{INT}_{F_p}), \ \text{THEN} \quad (y_i \in \text{INT}_{F_t}), \tag{16}$$

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 $R = \{R_1, \ldots, R_{n_1}\}$.

Given a test sample $\mathbf{z}_i \in Te = \{\mathbf{z}_i \mid i = 1, \ldots, m\}$, we first compute the fuzzy memberships $F_t(z_{ij})$ $(t = 1, \ldots, k, j = 1, \ldots, q)$, and get the following result.

$$(\mathbf{z}_{i1} \in \text{INT}_{F_r}) \wedge \ldots \wedge (\mathbf{z}_{iq} \in \text{INT}_{F_s}). \tag{17}$$

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

$$y_{\mathbf{z}_i} \in \text{INT}_{F_t}. \tag{18}$$

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 $\underline{R}F_t(\mathbf{z}_i)$ and $\overline{R}F_t(\mathbf{z}_i)$ for estimating the regression attribute value $y_{\mathbf{z}_i}$.

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 $\mathbf{z}_i \in Te, y_{\mathbf{z}_i}$ is supposed to be the regression attribute value of $\mathbf{z}_i$. The prediction process of $y_{\mathbf{z}_i}$ is summarized as follows.

Firstly, extract regression rules, denoted by $R = \{R_1, \ldots, R_{n_1}\}$. Secondly, determine the interval of regression attribute value of each test sample $\mathbf{z}_i$ and corresponding fuzzy subset F_t. And then compute the memberships of lower and upper approximations of $\mathbf{z}_i$ to the fuzzy set F_t i.e. $\underline{R}F_t(\mathbf{z}_i)$ and $\overline{R}F_t(\mathbf{z}_i)$. Finally, estimate $y_{\mathbf{z}_i}$ 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 $O(kn^2)$. If $k \to n, O(kn^2) \to O(n^3)$. 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

Table 6
Summary of data sets.

Data	Samples	Conditional attributes
abalone	4177	9
autoMPG8	9517	7
bank8FM	8192	9
bank32nH	8192	33
elevators9	16,599	19
housing	506	14
machinecpu	209	7
puma32H	8192	33
pyrim	74	28
stock	950	10
triazines	186	61

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.

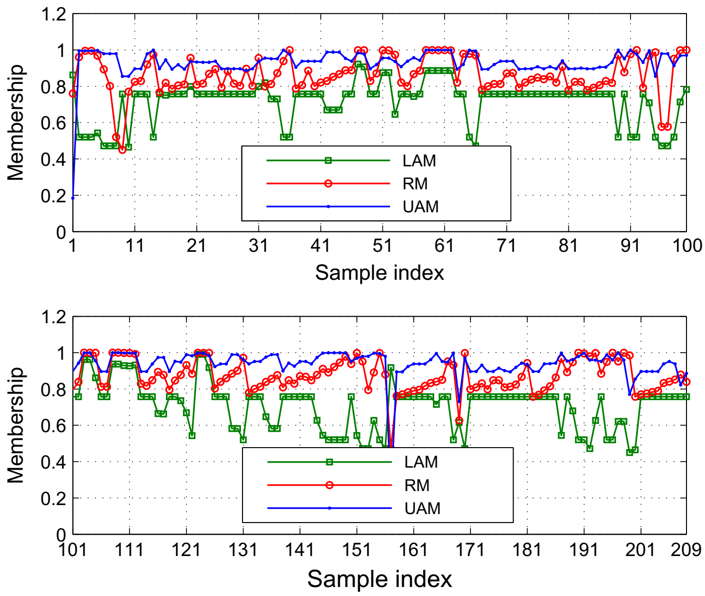

Fig. 5. Fuzzy rough approximation.

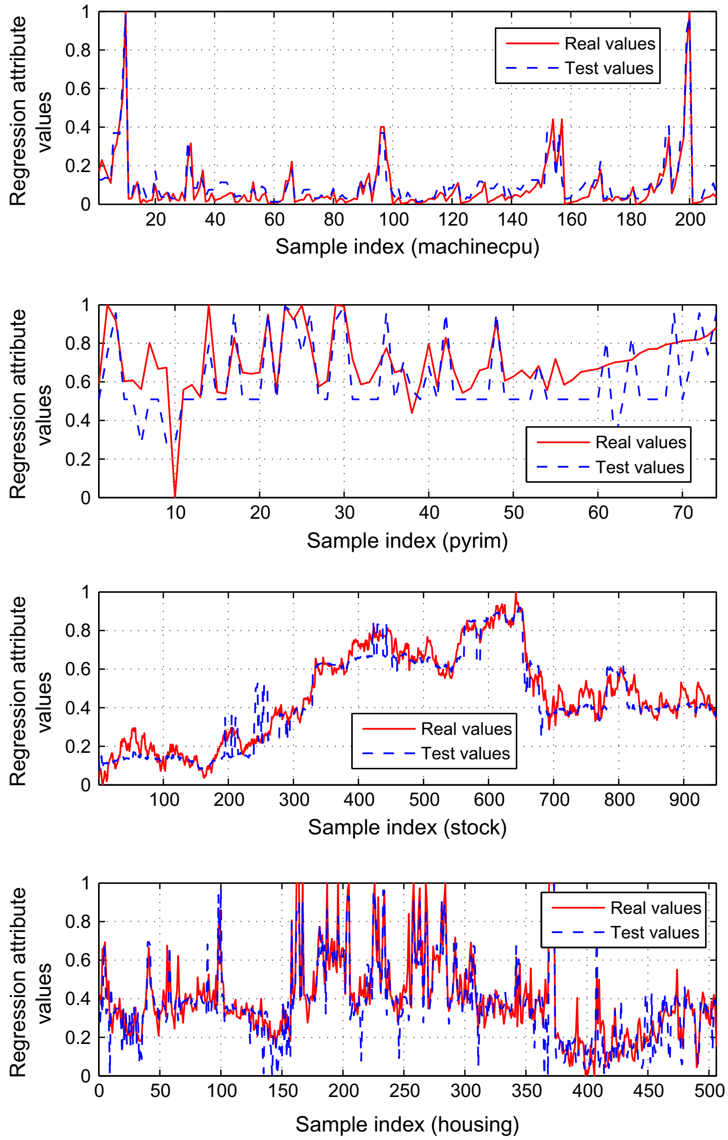

Fig. 6. Performance illustration of FRRPA.

- 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.

$$Corr\_coeff(W, W') = \frac{\sum (w_i - \overline{W})(w'_i - \overline{W'})}{\sqrt{\sum (w_i - \overline{W})^2 \sum (w'_i - \overline{W'})^2}}, \tag{19}$$

Table 7
Correlation coefficients between real values and test values (UCI data).

Data	FRRPA	GP	LR	MLP	SLR	AR	RBD	RT	FRNN
abalone	0.7001	0.6723	0.7622	0.7327	0.7054	0.7243	0.6452	0.7239	0.7373
autoMPG	0.7374	0.7046	0.8134	0.7719	0.6665	0.7662	0.7502	0.7754	0.8037
bank8FM	0.8806	0.6774	0.9657	0.9584	0.9665	0.9479	0.9492	0.9590	0.6553
bank32NH	0.8945	−0.2566	−0.1787	0.5056	0.0025	0.1803	0.1535	0.0800	0.0598
elevators	0.7772	0.9081	0.9134	0.9349	0.6462	0.7632	0.7965	0.7765	0.8006
housing	0.8606	0.8434	0.8427	0.8671	0.7315	0.8485	0.8720	0.8946	0.8402
machinecpu	0.9426	0.7731	0.8991	0.9412	0.7808	0.8899	0.9210	0.8509	0.8838
puma32H	0.9068	0.4303	0.4605	0.2498	0.4716	0.4008	0.8992	0.9375	0.8760
pyrim	0.7445	0.5115	0.5201	0.7062	0.3956	0.6771	0.4702	−0.2502	0.6409
stock	0.9868	0.8457	0.9335	0.9827	0.7103	0.9445	0.9831	0.9816	0.9920
triazines	0.6071	0.4391	0.4087	0.3819	0.1195	0.3935	0.3958	0.3481	0.5401
AVE	0.8217	0.5954	0.6673	0.7302	0.5633	0.6851	0.7124	0.6434	0.7118

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. $Corr\_coeff(W, W')$ 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 eitht 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.

$$Z = \frac{T - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}, \tag{20}$$

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 8
Root mean square error between real values and test values (UCI data).

Data	FRRPA	GP	LR	MLP	SLR	AR	RBD	RT	FRNN
abalone	0.1033	0.1080	0.0940	0.0996	0.1029	0.1001	0.1180	0.1005	0.0981
autoMPG	0.0659	0.0662	0.0541	0.0597	0.0694	0.0603	0.0624	0.0589	0.0554
bank8FM	0.1029	0.0835	0.0266	0.0297	0.0263	0.0330	0.0325	0.0291	0.1533
bank32NH	0.0318	0.0350	0.0333	0.0300	0.0332	0.0314	0.0316	0.0317	0.0317
elevators	0.0646	0.0407	0.0385	0.0334	0.0722	0.0620	0.0589	0.0613	0.0597
housing	0.1044	0.1098	0.1100	0.1031	0.1392	0.1088	0.1027	0.0915	0.1141
machinecpu	0.0601	0.0992	0.0614	0.0490	0.0878	0.0646	0.0548	0.0758	0.0791
puma32H	0.0698	0.1542	0.1515	0.3091	0.1504	0.1590	0.0752	0.0595	0.0844
pyrim	0.1455	0.1415	0.1613	0.1365	0.1560	0.1260	0.1409	0.1715	0.1240
stock	0.0384	0.1252	0.0837	0.0439	0.1604	0.0774	0.0427	0.0446	0.0296
triazines	0.1771	0.1784	0.1910	0.2412	0.1976	0.1867	0.2022	0.1881	0.1659
AVE	0.0876	0.1038	0.0914	0.1032	0.1087	0.0918	0.0838	0.0830	0.0905

Table 9
Results of statistical test.

	GP	LR	MLP	SLR	AR	RBD	RT	FRNN
CorrCoef	2.1998	0.6895	0.5582	2.4624	1.1491	0.4925	0.5582	1.1491
RootMSErr	−1.0835	0.0328	0.6238	−0.9521	−0.0985	0.6238	0.8208	0.0328

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, $y(t_i - s\Delta t)$ is the wind-speed before $s\Delta t$ minutes, and $y(t_i + k\Delta t)$ is the wind-speed in $k\Delta t$ minutes. In this work, $\Delta t = 10$ 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. $y(t_i + \Delta t)$ with $y(t_i - s\Delta t), \ldots, y(t_i - \Delta t), y(t_i)$. 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.

$t_1 - s\Delta t$	$\cdots$	$t_1 - \Delta t$	t_1	$t_1 + \Delta t$	$\cdots$	$t_1 + k\Delta t$
$y(t_1 - s\Delta t)$	$\cdots$	$y(t_1 - \Delta t)$	$y(t_1)$	$y(t_1 + \Delta t)$	$\cdots$	$y(t_1 + k\Delta t)$
$y(t_2 - s\Delta t)$	$\cdots$	$y(t_2 - \Delta t)$	$y(t_2)$	$y(t_2 + \Delta t)$	$\cdots$	$y(t_2 + k\Delta t)$
$y(t_3 - s\Delta t)$	$\cdots$	$y(t_3 - \Delta t)$	$y(t_3)$	$y(t_3 + \Delta t)$	$\cdots$	$y(t_3 + k\Delta t)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$y(t_{n-1} - s\Delta t)$	$\cdots$	$y(t_{n-1} - \Delta t)$	$y(t_{n-1})$	$y(t_{n-1} + \Delta t)$	$\cdots$	$y(t_{n-1} + k\Delta t)$
$y(t_n - s\Delta t)$	$\cdots$	$y(t_n - \Delta t)$	$y(t_n)$	$y(t_n + \Delta t)$	$\cdots$	$y(t_n + k\Delta t)$

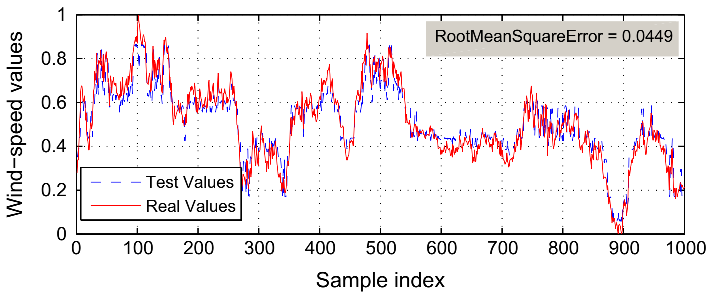

Fig. 7. Wind-speed prediction.

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

Data	FRRPA	GP	LR	MLP	SLR	AR	RBD	RT	FRNN
10min(k=1)	0.0698	0.2251	0.0365	0.0400	0.0366	0.0972	0.1096	0.0795	0.0932
30min(k=3)	0.0850	0.2295	0.0619	0.0721	0.0614	0.1344	0.1248	0.1006	0.1129
60min(k=6)	0.1067	0.2355	0.0875	0.0894	0.0867	0.1593	0.2117	0.1245	0.1506
2hours(k=12)	0.1519	0.2486	0.1264	0.1180	0.1264	0.2250	0.2410	0.2301	0.1934
3hours(k=18)	0.1840	0.2613	0.1651	0.1534	0.1651	0.2316	0.2371	0.2037	0.2021
4hours(k=24)	0.1939	0.2681	0.1952	0.1937	0.1952	0.2311	0.2369	0.2183	0.2247
5hours(k=30)	0.2483	0.2862	0.2247	0.1822	0.2247	0.3049	0.3006	0.2703	0.2418
6hours(k=36)	0.2528	0.2982	0.2422	0.1637	0.2422	0.2887	0.3190	0.2893	0.2542
7hours(k=42)	0.2588	0.3077	0.2567	0.1885	0.2567	0.3040	0.3331	0.3024	0.2670
8hours(k=48)	0.2646	0.3162	0.2704	0.2038	0.2704	0.3293	0.3474	0.2748	0.2734
9hours(k=54)	0.2617	0.3253	0.2836	0.2438	0.2836	0.3541	0.3242	0.3138	0.2793
10hours(k=60)	0.2589	0.3329	0.3007	0.2442	0.2986	0.3558	0.3406	0.3317	0.2954
AVE.	0.1947	0.2779	0.1876	0.1577	0.1873	0.2513	0.2605	0.2282	0.2157
Stat. test	0	−2.7135	0.0577	1.5011	0.0577	−1.5588	−1.7321	−1.3856	−0.8660

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 $k*10$ $(k = 1, 2, \ldots)$ minutes. For example, if we want to predict the wind-speed value in $k*10$ minutes, the format of training data set can be denoted as

$$Tr = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i = \{y(t_i - s\Delta t), \ldots, y(t_i)\}, y_i = y(t_i + k\Delta t), i = 1, \ldots, n\}. \tag{21}$$

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, $t_{22}0.05 = 1.717$. 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

[1] S. An, Q.H. Hu, Fuzzy rough decision trees, Rough Sets Curr. Trends Comput. Lect. Notes Comput. Sci. 7413 (2012) 397–404.
[2] R.B. Bhatt, M. Gopal, On fuzzy-rough sets approach to feature selection, Pattern Recogn. Lett. 26 (2005) 965–975.
[3] C.L. Blake, C.J. Merz, UCI Repository of Machine Learning Databases, 1998. <http://www.ics.uci.edu/mlearn/MLRepository.html>.
[4] P. Bonissone, J. Cadenas, M.C. Garrido, R.A.D. Valladares, A fuzzy random forest, Int. J. Approx. Reason. 51 (2010) 729–747.
[5] J. Buckley, T. Feuring, Linear and non-linear fuzzy regression: evolutionary algorithm solutions, Fuzzy Sets Syst. 112 (2000) 381–394.
[6] D.G. Chen, S.Y. Zhao, Local reduction of decision system with fuzzy rough sets, Fuzzy Sets Syst. 161 (2010) 1871–1883.
[7] K.Y. Chana, C.K. Kwongb, T.C. Fogartyc, Modeling manufacturing processes using a genetic programming-based fuzzy regression with detection of outliers, Inf. Sci. 180 (2010) 506–518.
[8] C. Cornelis, R. Jensen, G. hurtado martin, D. Slezak, Attribute selection with fuzzy decision reducts, Inf. Sci. 180 (2) (2010) 209–224.
[9] L.-F. Deng, K.-Y. Cai, A. So, Fuzzy nonlinear regression with fuzzified radial basis function network, IEEE Trans. Fuzzy Syst. 13 (2005) 742–760.
[10] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Gen. Syst. 17 (1990) 191–209.
[11] E. Frank, R.R. Bouckaert, Conditional density estimation with class probability estimators, in: First Asian Conference on Machine Learning, Berlin, 2009, pp. 65–81.
[12] J.H. Friedman, Stochastic gradient boosting, Comput. Statist. Data Anal. 38 (4) (2002) 367–378.
[13] R.M. Heiberger, E. Neuwirth, Simple linear regression, A Spreadsheet Interface for Statistics, Data Analysis, and Graphics, Springer, New York, 2009. pp. 193–212.
[14] P.-Y. Hao, J.-H. Chiang, Fuzzy regression analysis by support vector learning approach, IEEE Trans. Fuzzy Syst. 16 (2008) 428–441.
[15] B. Heshmaty, A. Kandel, Fuzzy linear regression and its applications to forecasting in uncertain environment, Fuzzy Sets Syst. 15 (1985) 159–191.
[16] Q.H. Hu, S. An, D.R. Yu, Soft fuzzy rough sets for robust feature evaluation and selection, Inf. Sci. 180 (2010) 4384–4400.
[17] Q.H. Hu, S. An, X. Yu, D.R. Yu, Robust fuzzy rough classifiers, Fuzzy Sets Syst. 183 (1) (2011) 26–43.
[18] Q.H. Hu, D.R. Yu, Z.X. Xie, Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern Recogn. Lett. 27 (2006) 414–423.
[19] Q.H. Hu, D. Yu, W. Pedrycz, D. Chen, Kernelized fuzzy rough sets and their applications, IEEE Trans. Knowl. Data Eng. 23 (2011) 1649–1667.

[20] Q.H. Hu, L. Zhang, S. An, D. Zhang, D. Yu, On robust fuzzy rough set models, IEEE Trans. Fuzzy Syst. 20 (4) (2012) 636–651.
[21] R. Jensen, C. Cornelis, Fuzzy rough nearest neighbour classification and prediction, Theor. Comput. Sci. 412 (42) (2011) 5871–5884.
[22] R. Jensen, Q. Shen, Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches, IEEE Trans. Knowl. Data Eng. 16 (2004) 1457–1471.
[23] R. Jensen, Q. Shen, Fuzzy-rough attribute reduction with application to web categorization, Fuzzy Sets Syst. 141 (2004) 469–485.
[24] R. Jensen, Q. Shen, Fuzzy-rough sets assisted attribute selection, IEEE Trans. Fuzzy Syst. 15 (2007) 73–89.
[25] R. Jensen, Q. Shen, New approaches to fuzzy-rough feature selection, IEEE Trans. Fuzzy Syst. 17 (2009) 824–838.
[26] J. Kacprzyk, M. Fedrizzi, Fuzzy Regression Analysis, Physica-Verlag, Heidelberg, Germany, 1992.
[27] J.L. Lu, R.L. Wang, An enhanced fuzzy linear regression model with more flexible spreads, Fuzzy Sets Syst. 160 (2009) 2505–2523.
[28] J. Mielniczuk, J. Tyrcha, Consistency of multilayer perceptron regression estimators, Neural Networks 6 (7) (1993) 1019–1022.
[29] M.H. Mashinchi, M.A. Orgun, M. Mashinchi, W. Pedrycz, A TabuCHarmony search-based approach to fuzzy linear regression, IEEE Trans. Fuzzy Syst. 19 (2011) 432–448.
[30] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982) 341–356.
[31] H. Peng, F. Liu, X. Yang, A hybrid strategy of short term wind power prediction, Renew. Energy 50 (2013) 590–595.
[32] G. Peters, Fuzzy linear regression with fuzzy intervals, Fuzzy Sets Syst. 63 (1994) 45–55.
[33] H. Shakouri, R. Nadimi, A novel fuzzy linear regression model based on a non-equality possibility index and optimum uncertainty, Appl. Soft Comput. 9 (2009) 590–598.
[34] H. Tankaka, S. Uejima, K. Asai, Fuzzy linear regression model, in: Proceeding International Conference on Application Systems Research Cybernetics, Acapulco, Mexico, 1980, pp. 12–15.
[35] S. Trabelsi, Z. Elouedi, P. Lingras, Classification systems based on rough sets under the belief function framework, Int. J. Approx. Reason. 52 (2011) 1409–1432.
[36] E.C.C. Tsang, D.G. Chen, D.S. Yeung, X.Z. Wang, J.W.T. Lee, Attributes reduction using fuzzy rough sets, IEEE Trans. Fuzzy Syst. 16 (2008) 1130–1141.
[37] X.Z. Wang, J.H. Zhai, S.X. Lu, Induction of multiple fuzzy decision trees based on rough set technique, Inf. Sci. 178 (2008) 3188–3202.
[38] X.Z. Wang, E.C.C. Tsang, S.Y. Zhao, D.G. Chen, D.Y.S. Yeung, Learning fuzzy rules from fuzzy samples based on rough set techniques, Inf. Sci. 177 (2007) 4493–4514.
[39] C.K.I. Williams, C.E. Rasmussen, Gaussian processes for regression, in: Advances in Neural Information Processing Systems, vol. 8, 1996.
[40] I.H. Witten, E. Frank, Data Mining-Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufman Publishers, 2000.
[41] H.Y. Wu, Y.Y. Wu, J.P. Luo, An interval type2 fuzzy rough set model for attribute reduction, IEEE Trans. Fuzzy Syst. 17 (2009) 301–315.
[42] K.K. Yen, S. Ghoshray, G. Roig, A linear regression model using triangular fuzzy number coefficients, Fuzzy Sets Syst. 106 (1999) 167–177.
[43] M. Yesilbudaka, S. Sagiroglub, I. Colak, A new approach to very short term wind speed prediction using k-nearest neighbor classification, Energy Convers. Manage. 69 (2013) 77–86.
[44] L.A. Zadeh, Fuzzy sets, Inform Contr 8 (1965) 338–353.
[45] J.H. Zhai, Fuzzy decision tree based on fuzzy-rough technique, Soft Comput. 15 (2011) 1087–1096.
[46] S.Y. Zhao, E.C.C. Tsang, D.G. Chen, X.Z. Wang, Building a rule-based classifier by using fuzzy rough set technique, IEEE Trans. Knowl. Data Eng. 22 (2010) 624–638.
[47] A. Zollanvaria, U.M. Braga-Netoa, E.R. Doughertya, On the sampling distribution of resubstitution and leave-one-out error estimators for linear classifiers, Pattern Recogn. 42 (11) (2009) 2705–2723.