

OBJECT-PARAMETER APPROACHES TO PREDICTING UNKNOWN DATA IN AN INCOMPLETE FUZZY SOFT SET

YAYA LIU ^a, KEYUN QIN ^{a,*}, CHANG RAO ^b, MAHAMUDA ALHAJI MAHAMADU ^c

^aCollege of Mathematics
Southwest Jiaotong University, Chengdu 610031, Sichuan, PR China
e-mail: yayaliu@my.swjtu.edu.cn, qinkeyun@home.swjtu.edu.cn

^bCollege of Information Science and Technology
Southwest Jiaotong University, Chengdu 610031, Sichuan, PR China
e-mail: alvis.c@hotmail.com

^c Council for Scientific and Industrial Research
PO 132, Accra, Ghana
e-mail: mamahamadu@yahoo.com

The research on incomplete fuzzy soft sets is an integral part of the research on fuzzy soft sets and has been initiated recently. In this work, we first point out that an existing approach to predicting unknown data in an incomplete fuzzy soft set suffers from some limitations and then we propose an improved method. The hidden information between both objects and parameters revealed in our approach is more comprehensive. Furthermore, based on the similarity measures of fuzzy sets, a new adjustable object-parameter approach is proposed to predict unknown data in incomplete fuzzy soft sets. Data predicting converts an incomplete fuzzy soft set into a complete one, which makes the fuzzy soft set applicable not only to decision making but also to other areas. The compared results elaborated through rate exchange data sets illustrate that both our improved approach and the new adjustable object-parameter one outperform the existing method with respect to forecasting accuracy.

Keywords: fuzzy soft set, incomplete fuzzy soft set, object-parameter approach, prediction, similarity measures.

1. Introduction

A number of real life problems in engineering, social and medical sciences, economics, etc., involve uncertainties. Classical mathematical tools are not sufficient to handle those. Various theories such as fuzzy set theory (Zadeh, 1965), rough set theory (Pawlak, 1982; Zhong and Skowron, 2001), vague set theory (Gau and Buehrer, 1993) and intuitionistic fuzzy set theory (Atanassov, 1986) have been proposed and proven to be useful mathematical approaches to modeling uncertainties. However, all the above theories are associated with an inherent limitation—the inadequacy of the parametrization tool.

Molodtsov (1999) initiated soft set theory as a new mathematical tool for dealing with uncertainties. This

tool is free from the limitation mentioned above, which affected other commonly used approaches. The absence of any restrictions on the approximate description in soft set theory makes it easily applicable in practice. We can use any parametrization we prefer: with the help of words and sentences, functions, mappings, real numbers, and so on. Since then, the generalized models of soft sets have been developed rapidly to meet various demands in practical situations by combining soft sets with fuzzy sets (Maji *et al.*, 2001), rough sets (Feng *et al.*, 2011), vague sets (Xu *et al.*, 2010), interval-valued fuzzy sets (Yang *et al.*, 2009), interval-valued intuitionistic fuzzy soft sets (Jiang *et al.*, 2010), and with other theories. Currently, research on soft set theory and hybrid soft set theory has been very active and there has been some progress concerning practical applications (Roy and Maji, 2007; Jiang *et al.*, 2011; Herawan and Deris, 2011; Quin *et al.*,

*Corresponding author

2012b; Jun *et al.*, 2009; Xiao *et al.*, 2009; Kong *et al.*, 2011; Alcantud, 2016; Muthukumar and Krishnan, 2016; Zhao and Guan, 2015; Li *et al.*, 2015a; 2015b; Li and Xie, 2014; Xie *et al.*, 2015).

All of the data sets mentioned above are based on complete information. However, incomplete data sets widely exist in practice. A slight mistake in the process of measuring, a restriction of data collecting, an error of data understanding and many other factors may directly result in missing data. For example, if some attendees ignore or misunderstand some questions when filling out a questionnaire, missing data or incomplete information will appear (Nowicki, 2010); if the detection equipment for some atmospheric features is backward or inaccurate, missing data or incomplete data will appear when data mining for prediction of air pollution (Siwek and Osowski, 2016). Data sets under incomplete information are referred to as incomplete data sets. For example, soft sets under incomplete information are referred to as incomplete soft sets, fuzzy soft sets under incomplete information are referred to as incomplete fuzzy soft sets, etc. The simplest approach to transforming an incomplete data set to a complete one is to delete all objects related to missing information and transform an incomplete data set to a complete one. However, this method will of course result in a loss of information. On the other hand, predicting unknown data is a more effective method for dealing with incomplete information.

Zou and Xiao (2008) initiated the study on data analysis approaches under the incomplete soft sets environment. They presented a weighted-average method for incomplete soft sets and an average-probability method (called a fuzzy method) for incomplete fuzzy soft sets to predict unknown data in the corresponding information tables. For an incomplete crisp soft set, the weighed-average method can only predict the sum of values of every object on all parameters but cannot quantify each individual unknown entry in information tables, which makes the method only applicable to decision making problems. For an incomplete fuzzy soft set, although the average-probability method can predict each individual unknown entry of fuzzy soft sets, the predicted values of all unknown entries in a certain parameter column are equal, which makes this method of low accuracy. Qin *et al.* (2012a) propose a data filling approach for incomplete soft sets. In their approach, the missing data are filled in terms of the association degree between the parameters when a strong association exists between the parameters or in terms of the distribution of other available objects when no strong association exists between the parameters. However, the approach cannot be used for predicting unknown data in incomplete fuzzy soft sets.

Deng and Wang (2013) proposed an object-parameter approach (or the “Deng–Wang

approach”, to distinguish it from other methods in this work) to predict the unknown entries in fuzzy soft sets. They introduced the notions of “the complete distance” between two objects and “the average dominant degree” between two parameters to reveal hidden information in a fuzzy soft set. Compared with the two approaches mentioned earlier, the Deng–Wang method has three main advantages: (i) the predicted values of different objects on a certain parameter vary from their entries; (ii) it takes both the information between objects and the information between parameters into account; (iii) it can predict unknown data in incomplete soft sets as well as in incomplete fuzzy soft sets. However, it suffers from two main limitations: (i) the estimated value may be not in the interval $[0, 1]$; (ii) the information between the objects considered is not comprehensive, just like the information between parameters.

In order to overcome these limitations of the Deng–Wang approach, we redefine the notion of the dominant degree, distinguish the roles of the average distance and the dominant degree when predicting unknown data, and then give an improvement of the Deng–Wang method.

The remainder of this paper is organized as follows. Section 2 presents the notions of soft sets, fuzzy soft sets and incomplete fuzzy soft sets. In Section 3, the Deng–Wang approach for predicting unknown data in incomplete fuzzy soft sets is reviewed. In Section 4, a counterexample and some analysis are presented to illustrate the limitations of the Deng–Wang approach. An improvement of the Deng–Wang method is presented in Section 5. A new adjustable object-parameter approach based on the similarity measures of fuzzy sets is presented in Section 6. Experiments are implemented in Section 7, and conclusions are drawn in Section 8.

2. Preliminaries

In this section we briefly recall some concepts that are useful for subsequent discussions.

The theory of fuzzy sets, first introduced by Zadeh (1965), provides an appropriate framework for representing and processing vague concepts by allowing partial memberships.

Definition 1. (Zadeh, 1965) A fuzzy set F in the universe U is defined as

$$F = \{(x, \mu_F(x)) / x \in U, \mu_F(x) \in [0, 1]\}$$

where μ_F is called the membership function of F and $\mu_F(x)$ indicates the membership degree of x to F . We denote the set of all fuzzy sets on U by $F(U)$.

The *fuzzy intersection*, *union*, and *complement* are respectively defined as follows:

$$\mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\},$$

$$\mu_{A \cup B}(x) = \max\{\mu_A(x), \mu_B(x)\},$$

$$\mu_{\neg A}(x) = 1 - \mu_A(x),$$

where $A, B \in F(U)$ and $x \in U$.

If $\mu_A(x) \leq \mu_B(x)$, A is called a *subset* of B , and this relationship is denoted by $A \subseteq B$. Obviously, $A = B$ if $A \subseteq B$ and $B \subseteq A$.

To overcome the inadequacy of the parametrization tools of fuzzy set theory and other commonly used approaches to deal with uncertainty, Molodtsov (1999) introduced the concept of soft set theory.

Let U be the initial universe of objects and E be the set of parameters in relation to objects in U . Both U and E are assumed to be nonempty finite sets. Let $P(U)$ be the power set of U and $A \subseteq E$.

Definition 2. (Molodtsov, 1999) A pair (F, A) is called a *soft set* in the universe U , where F is a mapping given by $F : A \rightarrow P(U)$.

For any parameter $e \in A$, $F(e) \subseteq U$ may be considered the set of e -approximate elements of the soft set (F, A) . In other words, the soft set is not a kind of set in the ordinary sense, but a parameterized family of subsets of the set U .

Maji *et al.* (2001) initiated the study on hybrid structures involving both fuzzy sets and soft sets. They introduced the notion of fuzzy soft sets, which can be seen as a fuzzy generalization of crisp soft sets.

Definition 3. (Maji *et al.*, 2001) A pair (F, A) is called a *fuzzy soft set* over U , where $A \subseteq E$ and F is a mapping given by $F : A \rightarrow F(U)$.

A fuzzy soft sets can be viewed as an information table. In the information table of a fuzzy soft set, each entry belongs to the interval $[0, 1]$ and is decided by the membership degree of an object on a parameter.

In the foregoing discussion, assume that the domain of every fuzzy soft set is $U = \{x_1, x_2, \dots, x_m\}$ and the parameters set is $E = \{e_1, e_2, \dots, e_n\}$. For an object $x_i \in U$ and a parameter $e_l \in E$, denote by h_{il} the value of x_i on e_l . If incomplete data exist in the information table of a fuzzy soft set, then the fuzzy soft set is called an incomplete fuzzy soft set and the unknown data are denoted by the sign ‘*’. For example, in the following fuzzy soft set (F, A) shown in Table 1, all membership values of objects on parameters are known except those of x_2, x_3 on e_2 . The unknown data are denoted by ‘*’ in the information table, i.e., $h_{22} = *$ and $h_{32} = *$.

3. Deng–Wang approach

In this section, the object-parameter approach presented by Deng and Wang (2013) is reviewed.

Let $U_k = \{i | h_{ik} \neq *, 1 \leq i \leq m\}$ and set $0/0 = 0$. Consider an incomplete fuzzy soft set (F, E) over U and let h_{jl} be the unknown value which is to be predicted.

Table 1. Tabular representation of the incomplete fuzzy soft set (F, A) .

U	e_1	e_2	e_3	e_4	e_5	e_6
x_1	0.9	0.4	0.5	0.4	0.8	0.8
x_2	0.8	*	0.5	0.7	0.6	0.3
x_3	0.4	*	0.9	0.9	0.5	0.9
x_4	0.9	0.8	0.9	0.4	0.7	0.5

Definition 4. (Deng and Wang, 2013) Let (F, E) be a fuzzy soft set over U . For $x_i, x_j \in U$ and $e_k \in E$, if h_{ik} and h_{jk} are already known, the *relative distance* from x_i to x_j with respect to e_k is defined by

$$d_{jk}^{ik} = \frac{h_{ik} - h_{jk}}{\sum_{\{l \in U_k\}} |h_{lk} - h_{jk}|}. \tag{1}$$

Moreover, the *complete distance* from x_i to x_j on all the parameters is defined by

$$d_{ij} = \frac{\sum_{k=1}^n d_{jk}^{ik}}{|\{k | (i \in U_k) \wedge (j \in U_k)\}|}. \tag{2}$$

The quantity d_{jk}^{ik} is used to evaluate the difference between the values of objects x_i and x_j on the parameter e_k . The complete distance d_{ij} is used to measure the average distance from the values of x_i to the values of x_j (the average distance from x_i to x_j , for brevity).

Based on the complete distance, the unknown entry h_{jl} is evaluated according to the information from the relationship between the values of objects on a certain parameter e_l by

$$h_{jl}^{\text{object}} = \frac{\sum_{i \in U_l} (h_{il} - d_{ij})}{|U_l|}. \tag{3}$$

Definition 5. (Deng and Wang, 2013) Consider an incomplete fuzzy soft set (F, E) on U . Let $x_i \in U$ and let $e_k, e_l \in E$. If h_{ik} and h_{il} are already known, the *degree of e_k being relatively dominant* to e_l regarding x_i is defined by

$$r_{il}^{ik} = \frac{h_{ik} - h_{il}}{h_{ik} + h_{il}}. \tag{4}$$

Definition 6. (Deng and Wang, 2013) Suppose (F, E) is a fuzzy soft set over U . For $e_k, e_l \in E$, the *degree of e_k being definitely dominant* to e_l is defined by

$$c_{kl} = \frac{\sum_{i \in U_k \cap U_l} r_{il}^{ik}}{|U_k \cap U_l|}. \tag{5}$$

Moreover, the *degree of average dominance* of e_k to e_l is given by

$$v_{kl} = \frac{c_{kl}}{\sum_{\{q | U_q \cap U_l \neq \emptyset\}} |c_{ql}|}. \tag{6}$$

The average dominance degree is used to study the relationship between parameters in an incomplete fuzzy soft set. Based on the average dominance degree, the unknown entry h_{jl} is evaluated according to the information between parameters regarding the object x_j by

$$h_{jl}^{\text{parameter}} = \frac{\sum_{k \in G_j} (h_{jk} - v_{kl})}{|G_j|}, \quad (7)$$

where $G_j = \{k | (h_{jk} \neq *) \wedge (U_k \cap U_l \neq \emptyset), 1 \leq k \leq n\}$.

Finally, the unknown entry h_{jl} can be predicted by linearly weighting h_{jl}^{object} and $h_{jl}^{\text{parameter}}$ as follows:

$$h_{jl} = w_1 \times h_{jl}^{\text{object}} + w_2 \times h_{jl}^{\text{parameter}}, \quad (8)$$

where w_1 and w_2 stand for the weights of objects and parameters on the impacts on unknown data, respectively.

Given an incomplete fuzzy soft set (F, E) over U , h_{jl} stands for unknown data going to be predicted by the proposed object-parameter method. Based on the definitions above, Deng and Wang proposed an algorithm to predict the unknown data in an incomplete fuzzy soft set presented as Algorithm 1.

Algorithm 1. Object-parameter algorithm to predict the unknown data in an incomplete fuzzy soft set (Deng and Wang, 2013).

Step 1. For any object $x_i \in U$ and any parameter $e_k \in E$, according to Eqns. (1) and (2) we obtain d_{jk}^{ik} and d_{ij} . Thus the evaluation of h_{jl} regarding the relationship between objects is proceeds through Eqn. (3).

Step 2. According to Eqns. (4)–(6), the values of r_{il}^{ik} , c_{kl} and v_{kl} are obtained, respectively.

Step 3. By Eqn. (7) the evaluation of h_{jl} regarding the relationship between parameters is made.

Step 4. Given a pair of weights w_1 and w_2 , the unknown entry h_{jl} is predicted by Eqn. (8).

4. Counterexample and analysis

In this subsection, a counterexample is provided to show that the Deng–Wang approach may be unreasonable in some cases.

Example 1. In Table 2 an incomplete fuzzy soft set is given. There are 6 objects and 7 parameters, and 5 unknown entries required to be predicted.

We use Algorithm 1 to predict the unknown data in Table 2:

1. By Eqns.(1)–(3), we obtain $d_{13} \approx -0.253$, $d_{23} \approx -0.221$, $d_{43} \approx -0.011$, $d_{53} \approx -0.292$, $d_{63} \approx -0.299$, so we have $h_{31}^{\text{object}} \approx 1.095$;

Table 2. Tabular representation of an incomplete fuzzy soft set.

U	e_1	e_2	e_3	e_4	e_5	e_6	e_7
x_1	0.9	0.4	0.1	0.9	0.6	0.3	0.4
x_2	0.8	0.6	0.5	*	0.5	0.3	0.3
x_3	*	0.8	0.9	*	0.9	0.9	0.9
x_4	0.9	0.8	0.9	0.8	*	0.8	0.9
x_5	0.9	0.2	0.2	0.6	0.3	0.4	*
x_6	0.9	0.2	0.4	0.4	0.4	0.3	0.3

2. By Eqns.(4)–(7), we obtain $c_{21} \approx -0.448$, $c_{31} \approx -0.410$, $c_{51} \approx -0.329$, $c_{61} \approx -0.380$, $c_{71} \approx -0.335$, $v_{21} \approx -0.247$, $v_{31} \approx -0.212$, $v_{51} \approx -0.170$, $v_{61} \approx -0.197$, $v_{71} \approx -0.173$, so we have, $h_{31}^{\text{parameter}} \approx 1.080$.

In this case, no matter how we select the weights of objects and parameters, by Eqn. (8) we have that $h_{31} > 1$, which is obviously unreasonable.

To understand why this unreasonable situation would occur, we have to analyze the principle of the Deng–Wang approach.

Suppose that h_{jl} stands for missing data to be predicted in the fuzzy soft set (F, E) . To estimate h_{jl} based on the relationship between elements, Deng and Wang (2013) firstly calculated the so-called “complete distance” from each element to x_j by Eqn. (2). The “complete distance”, which can be positive or negative, is actually used to measure the average distance from the value of one object to another on all the parameters. In order to estimate the value of h_{jl} an already known data h_{il} , the complete distance d_{ij} (d_{ij} is the complete distance from x_i to x_j) should be subtracted from h_{il} . Through every h_{il} ($h_{il} \neq *$, $i \in \{1, 2, \dots, m\}$), Deng and Wang (2013) obtained an estimate value of h_{jl} . Finally, the average of these estimated values was calculated by Eqn. (3). Since $h_{il} - d_{ij}$ may be greater than 1 for each h_{il} ($h_{il} \neq *$, $i \in \{1, 2, \dots, m\}$), it is no surprise that the average of these estimates, i.e., h_{jl}^{object} , may be greater than one.

To estimate h_{jl} based on the relationship between parameters, Deng and Wang (2013) calculated the so-called “relate dominant degree”, “definitely dominant degree” and “average dominant degree” of each parameter over e_l . In order to estimate h_{jl} by already known data h_{jk} , the “average dominant degree” v_{kl} should be subtracted from h_{jk} . Through every h_{jk} ($h_{jk} \neq *$, $k \in \{1, 2, \dots, n\}$), Deng and Wang (2013) obtained an estimate value of h_{jl} . Finally, the average of these estimates was calculated by Eqn. (7). Since $h_{jk} - v_{kl}$ may be greater than 1 for each h_{jk} ($h_{jk} \neq *$, $k \in \{1, 2, \dots, n\}$), it is no surprise that the average of these estimates values, i.e., $h_{jl}^{\text{parameter}}$, may be greater than 1.

In fact, the “relatively dominant degree” calculated by Eqn. (4) is used to measure the distance between values

of an object on two parameters; the “definitely dominant degree” calculated by Eqn. (5) is used to measure the average distance between the values of all objects on two parameters; the “average dominant degree” (calculated by Eqn. (6)) is used as the average distance from the values on one parameter to the values on another (the average distance from one parameter to another, for brevity) in Eqn. (7). That is, although Deng and Wang (2013) introduced the notion of the “dominant degree”, they did not distinguish the roles of the “average distance” and the “dominant degree” when predicting unknown data in an incomplete fuzzy soft set.

Since Deng and Wang (2013) computed the average of the estimates by Eqns. (3) and (7), they ignored the varying effects of different elements on a fixed element, and of different parameters on a fixed parameter. We present an illustrative example. Let h_{jl} be an unknown item. Write $h_{jl}^1 = h_{il} - d_{ij}$ for the estimate of h_{jl} obtained by considering the distance d_{ij} from x_i to x_j , and $h_{jl}^2 = h_{i'l} - d_{i'j}$ for the estimate of h_{jl} obtained by considering the distance $d_{i'j}$ from x'_i to x_j . If x_i and x'_i have different effects on x_j , then h_{jl}^1 and h_{jl}^2 are expected to have different effects on h_{jl}^{object} . That is, one should set different weights for h_{jl}^1 and h_{jl}^2 when calculating h_{jl}^{object} , rather than computing the average by Eqn. (3).

Based on the discussion above, we should present an improvement of the Deng–Wang approach in this work. In our improvement, the “average distance” will be used to predict an unknown item from an already known item, and the “dominant degree” will be redefined and used to describe the possibility that the unknown one will be determined by this already known one. As an illustration, in Table 2, to predict h_{45} ($h_{45} = *$) by h_{46} ($h_{46} \neq *$) according to the relationship between parameters e_5 and e_6 , let c_{65} be the average distance from e_6 to e_5 and C_{65} be the dominant degree of e_6 over e_5 . Then the value of h_{45} predicted by h_{46} is $h_{46} - c_{65}$. C_{65} describes the possibility that h_{45} is determined by h_{46} , i.e., the possibility that h_{45} is equal to $h_{46} - c_{65}$. Besides, the estimated value of every unknown item will be within the interval $[0, 1]$ by using our improved approach.

5. Improvement of the Deng–Wang approach

Consider an incomplete fuzzy soft set (F, E) on U and let h_{jl} be an unknown item. Let $U_l = \{i | h_{il} \neq *, 1 \leq i \leq m\}$ and $E_i = \{l | h_{il} \neq *, 1 \leq l \leq n\}$. It is obvious that $U_l \subseteq U$ and $E_i \subseteq E$. By convention, we set $0/0 = 0$.

Definition 7. Let (F, E) be a fuzzy soft set over U . For $x_i, x_j \in U$ and $e_k \in E$, if h_{ik} and h_{jk} are already known, the distance from x_i to x_j with respect to e_k is defined by

$$d_{jk}^{ik} = h_{ik} - h_{jk}. \tag{9}$$

Furthermore, the average distance from the values of x_i to the values of x_j on all the parameters (the average distance from x_i to x_j , for brevity) is defined by

$$d_{ij} = \frac{\sum_{k \in E_i \cap E_j} d_{jk}^{ik}}{|E_i \cap E_j|}. \tag{10}$$

Since $h_{ik}, h_{jk} \in [0, 1]$, by Eqns. (9) and (10) it is easy to obtain that $d_{jk}^{ik} \in [-1, 1]$ and $d_{ij} \in [-1, 1]$.

Remark 1. It should be noted that one can use different formulas to calculate the distance from a certain value to another one. Both Eqns. (1) and (9) can be used to compute the distance from h_{ik} to h_{jk} . Here, we select the latter, rather than the former because it is simpler to compute.

The value of h_{jl} predicted by an already known value h_{il} , which is denoted by $h_{jl}^{il'}$, can be calculated in the following manner:

$$h_{jl}^{il'} = h_{il} - d_{ij}. \tag{11}$$

Since $d_{ij} \in [-1, 1]$, by Eqn. (11) it is easy to obtain that $h_{jl}^{il'} \in [h_{il} - 1, h_{il} + 1]$. To assure that the predicted value can be within the interval $[0, 1]$, we formalize Eqn.(11) as follows:

$$h_{jl}^{il} = \begin{cases} h_{il} & \text{if } d_{ij} = 0, \\ h_{il} - d_{ij} \frac{1-h_{il}}{(h_{il}+1)-h_{il}} & \text{if } -1 \leq d_{ij} < 0, \\ h_{il} - d_{ij} \frac{h_{il}-0}{h_{il}-(h_{il}-1)} & \text{if } 1 \geq d_{ij} > 0, \end{cases} \tag{12}$$

i.e.,

$$h_{jl}^{il} = \begin{cases} h_{il} & \text{if } d_{ij} = 0, \\ h_{il} - d_{ij} + d_{ij}h_{il} & \text{if } -1 \leq d_{ij} < 0, \\ h_{il} - d_{ij}h_{il} & \text{if } 1 \geq d_{ij} > 0. \end{cases} \tag{13}$$

Lemma 1. By Eqn. (13) h_{jl}^{il} is restricted to the interval $[0, 1]$.

Proof.

(i) Suppose that $d_{ij} = 0$. Since $0 \leq h_{il} \leq 1$, it is easy to get $0 \leq h_{jl}^{il} = h_{il} \leq 1$.

(ii) Suppose that $-1 \leq d_{ij} < 0$. Since $-1 \leq h_{il} - 1 \leq 0$ and $0 \leq 1 + d_{ij} < 1$, we have $h_{il} - d_{ij} + d_{ij}h_{il} = h_{il} + d_{ij}(h_{il} - 1) \geq h_{il}$ and $h_{il} - d_{ij} + d_{ij}h_{il} - 1 = h_{il}(1 + d_{ij}) - (1 + d_{ij}) = (h_{il} - 1)(1 + d_{ij}) \leq 0$, i.e., $h_{il} - d_{ij} + d_{ij}h_{il} \leq 1$. Thus we obtain $h_{il} \leq h_{jl}^{il} \leq 1$. Also, since $h_{jl}^{il} = h_{il} - d_{ij} + d_{ij}h_{il} = (h_{il} - 1)d_{ij} + h_{il}$ and $-1 \leq h_{il} - 1 \leq 0$, we know that the value h_{jl}^{il} increases monotonically as d_{ij} decreases monotonically. If d_{ij} reaches the minimal value -1 , it is easy to obtain that

$$h_{jl}^{il} = h_{il} - d_{ij} + d_{ij}h_{il} = 1.$$

(iii) Suppose that $1 \geq d_{ij} > 0$. Since $0 \leq h_{il} \leq 1$ and $0 \leq 1 - d_{ij} < 1$, we obtain $h_{il} \geq h_{il} - d_{ij}h_{il} = h_{il}(1 - d_{ij}) \geq 0$, i.e., $h_{il} \geq h_{jl}^{il} \geq 0$. Also, it is obvious that the value $h_{jl}^{il} = h_{il} - d_{ij}h_{il}$ decreases monotonically as d_{ij} increases monotonically. If d_{ij} reaches the maximal value 1, it is easy to get that $h_{jl}^{il} = h_{il} - d_{ij}h_{il} = 0$. ■

Definition 8. Let (F, E) be a fuzzy soft set over U . For any $x_i, x_j \in U$, denote the average distance from x_i to x_j by d_{ij} ; then the *dominant degree* of x_i over x_j ($i \neq j$) is defined by

$$D_{ij} = \frac{1 - |d_{ij}|}{\sum_{\{p|E_p \cap E_j \neq \emptyset, 1 \leq p \leq m, p \neq j\}} (1 - |d_{pj}|)}. \quad (14)$$

Write $M_l = \{i | (h_{il} \neq *) \wedge (E_i \cap E_j \neq \emptyset), 1 \leq i \leq m\}$. The *relative dominant degree* of x_i over x_j with respect to e_l is defined by

$$D_{jl}^{il} = \frac{1 - |d_{ij}|}{\sum_{p \in M_l} (1 - |d_{pj}|)}. \quad (15)$$

In Eqn. (14), the larger the difference $|d_{ij}|$ between x_i and x_j , the smaller the dominance degree of x_i over x_j . There may be many methods to construct the dominant degree of one object over another—we just propose one of these by Eqn. (14).

The relative dominant degree D_{jl}^{il} calculated by Eqn. (15) is used to describe the possibility that the value of h_{jl} ($h_{jl} = *$) will be determined by h_{il} ($h_{il} \neq *$). For $e_l \in E$, if there is one and only one object x_j such that the membership value of x_j on e_l (h_{jl}) is unknown, then the relative dominant degree degenerates to the dominant degree of x_i over x_j , i.e., $D_{jl}^{il} = D_{ij}$.

The unknown entry h_{jl} can be evaluated according to the information between the values of objects on a certain parameter by

$$h_{jl}^{\text{object}} = \sum_{i \in M_l} h_{jl}^{il} \times D_{jl}^{il}. \quad (16)$$

Definition 9. Consider an incomplete fuzzy soft set (F, E) on U . Let $x_i \in U$, $e_k, e_l \in E$, h_{ik} and h_{il} are already known. The *distance* from e_k to e_l regarding x_i is defined by

$$c_{il}^{ik} = h_{ik} - h_{il}. \quad (17)$$

Furthermore, the *average distance* from the values of objects on e_k to the values of objects on e_l (the average distance from e_k to e_l , for brevity) is defined by

$$c_{kl} = \frac{\sum_{i \in U_k \cap U_l} c_{il}^{ik}}{|U_k \cap U_l|}. \quad (18)$$

Since $h_{ik}, h_{il} \in [0, 1]$, by Eqns. (17) and (18) it is easy to get that $c_{il}^{ik} \in [-1, 1]$ and $c_{kl} \in [-1, 1]$.

Remark 2. Both Eqns. (4) and (17) can be used to measure the distance from h_{ik} to h_{il} . Here, we select the latter, rather than the former because it is simpler to compute.

Analysing the average distance from e_k to e_l , the unknown value h_{jl} can be predicted by h_{jk} in the following manner:

$$h_{jl}^{jk'} = h_{jk} - c_{kl}. \quad (19)$$

Since $c_{kl} \in [-1, 1]$, by Eqn. (19) it is easy to obtain that $h_{jl}^{jk'} \in [h_{jk} - 1, h_{jk} + 1]$. To assure that the predicted value can be within the interval $[0, 1]$, we formalize Eqn. (19) as follows:

$$h_{jl}^{jk} = \begin{cases} h_{jk} & \text{if } c_{kl} = 0, \\ h_{jk} - c_{kl} \frac{1-h_{jk}}{(h_{jk}+1)-h_{jk}} & \text{if } -1 \leq c_{kl} < 0, \\ h_{jk} - c_{kl} \frac{h_{jk}-0}{h_{jk}-(h_{jk}-1)} & \text{if } 1 \geq c_{kl} > 0, \end{cases} \quad (20)$$

i.e.,

$$h_{jl}^{jk} = \begin{cases} h_{jk} & \text{if } c_{kl} = 0, \\ h_{jk} - c_{kl} + c_{kl}h_{jk} & \text{if } -1 \leq c_{kl} < 0, \\ h_{jk} - c_{kl}h_{jk} & \text{if } 1 \geq c_{kl} > 0. \end{cases} \quad (21)$$

Lemma 2. By Eqn. (21) h_{jl}^{jk} is restricted to the interval $[0, 1]$.

Proof. The proof is similar to that of Lemma 1 and, therefore, omitted here. ■

Definition 10. The *dominant degree* of e_k over e_l ($k \neq l$) is given by

$$C_{kl} = \frac{1 - |c_{kl}|}{\sum_{\{q|U_q \cap U_l \neq \emptyset, 1 \leq q \leq n, q \neq l\}} (1 - |c_{ql}|)}. \quad (22)$$

Write $G_j = \{k | (h_{jk} \neq *) \wedge (U_k \cap U_l \neq \emptyset), 1 \leq k \leq n\}$. The *relative dominant degree* of e_k over e_l regarding x_j is characterized by

$$C_{jl}^{jk} = \frac{1 - |c_{kl}|}{\sum_{q \in G_j} (1 - |c_{ql}|)}. \quad (23)$$

In Eqn. (22), the larger the difference $|c_{kl}|$ between e_k and e_l , the smaller the dominance degree of e_k over e_l . There are various methods to construct the dominant degree of one parameter over another—we just propose one of these by Eqn. (22).

The relative dominant degree C_{jl}^{jk} in Eqn. (23) is used to describe the possibility that the value of h_{jl} ($h_{jl} = *$) will be determined by h_{jk} ($h_{jk} \neq *$). For $x_j \in U$, if there is one and only one parameter e_l s.t. the membership value of x_j on e_l (h_{jl}) is unknown, then the relative dominant degree degenerates to the dominant degree of e_k over e_l , i.e., $C_{jl}^{jk} = C_{kl}$.

The unknown entry h_{jl} can be evaluated according to the information between the values of parameters pertaining to object x_j by

$$h_{jl}^{\text{parameter}} = \sum_{k \in G_j} h_{jl}^{jk} \times C_{jl}^{jk}. \quad (24)$$

Finally, the unknown entry h_{jl} can be predicted by linearly weighing h_{jl}^{object} and $h_{jl}^{\text{parameter}}$ as follows:

$$h_{jl} = w_1 \times h_{jl}^{\text{object}} + w_2 \times h_{jl}^{\text{parameter}}, \quad (25)$$

where w_1 and w_2 respectively denote the weight of objects and the weight of parameters quantifying the impact on unknown data. They satisfy the condition $w_1 + w_2 = 1$.

Given an incomplete fuzzy soft set (F, E) over U , h_{jl} is going to be predicted by the proposed object-parameter method. It can be viewed as an improvement of the Deng–Wang approach and can be implemented through the procedure presented in Algorithm 2.

Algorithm 2. Improved object-parameter algorithm to predict the unknown data in an incomplete fuzzy soft set.

Step 1. For any $i \in M_l$, calculate h_{jl}^{il} by Eqns. (9), (10) and (13), and then calculate D_{jl}^{il} by Eqn. (15).

Step 2. According to Eqn.(16), we estimate the value of h_{jl} based on the information between objects; the estimated value is denoted by h_{jl}^{object} .

Step 3. For any $k \in G_j$, calculate h_{jl}^{jk} by Eqns. (17), (18) and (21), and then calculate C_{jl}^{jk} by Eqn. (23).

Step 4. According to Eqn.(24), we estimate the value of h_{jl} based on the information between parameters; the estimated value is denoted by $h_{jl}^{\text{parameter}}$.

Step 5. Given a pair of weights w_1 and w_2 , estimate the value of h_{jl} by Eqn. (25).

Theorem 1. The value h_{jl} estimated by Algorithm 2 is in the interval $[0, 1]$.

Proof. By Eqn. (13) and Lemma 1, we deduce that $0 \leq h_{jl}^{il} \leq 1$. By Eqn. (15), we obtain that $\sum_{i \in M_l} D_{jl}^{il} = 1$. It is easy to obtain that $0 \leq \sum_{i \in M_l} h_{jl}^{il} \times D_{jl}^{il} \leq \sum_{i \in M_l} D_{jl}^{il} = 1$, i.e., $0 \leq h_{jl}^{\text{object}} \leq 1$. In a similar way, we conclude that $0 \leq h_{jl}^{\text{parameter}} \leq 1$. The weights w_1 and w_2 satisfy the condition $w_1 + w_2 = 1$, so we obtain $0 \leq h_{jl} \leq 1$ by Eqn. (25). ■

Remark 3. Comparing Algorithms 1 and 2, three main differences are noticeable:

- (i) Hidden information between objects revealed. In Algorithm 1, only the distance from one object

to another is considered. The effect of different elements on a fixed element is not taken into consideration. In Algorithm 2, we define the “dominant degree” of one object over another, and use the “dominant degree” to measure the effect of different objects on a fixed element. Both the average distance between objects and the dominant degree of one object over another are taken into consideration.

- (ii) Hidden information between parameters is revealed. In Algorithm 1, the roles of the “average distance” and the “dominant degree” are not distinguished. In Algorithm 2, we redefine the “dominant degree” of one parameter over another, and use the “dominant degree” to measure the effect of different parameters on a fixed parameter. Both the average distance between parameters and the dominant degree of one parameter over another are taken into consideration in Algorithm 2.
- (iii) By Algorithm 1, the final estimated value may be not in the unit interval $[0, 1]$. By Algorithm 2, the final estimated value is restricted to $[0, 1]$.

Example 2. Let us come back to the example shown in Table 2, and let the weights of objects and parameters be equal, i.e., $w_1 = w_2 = 1/2$. By Algorithm 2, we obtain $h_{31} \approx 0.626$, $h_{24} \approx 0.987$, $h_{34} \approx 0.959$, $h_{45} \approx 0.841$, $h_{57} \approx 0.308$. ♦

6. New adjustable object-parameter method

In this section, we use the similarity measure of fuzzy sets to study the information between two objects, and between two parameters in the information table of a fuzzy soft set. The similarity measure between two fuzzy sets is defined by Wang (1983) as follows.

Definition 11. (Wang, 1983) A function $S : F(U) \times F(U) \rightarrow [0, 1]$ is called a similarity measure of fuzzy sets if it satisfies the following properties:

- (S1) $S(U, \emptyset) = 0$ and $S(A, A) = 1$ whenever $A \in F(U)$;
- (S2) $S(A, B) = S(B, A)$ whenever $A, B \in F(U)$;
- (S3) for all $A, B, C \in F(U)$, we have $S(A, C) \leq \min(S(A, B), S(B, C))$ whenever $A \subseteq B \subseteq C$.

Now we propose an adjustable method to predict the unknown data in an incomplete fuzzy soft set. This new method takes account of both the relationship between objects and between parameters, so we call it an adjustable object-parameter method.

Definition 12. Suppose (F, E) is a fuzzy soft set on U and $x_i, x_j \in U$. Let S' be a similarity measure of fuzzy sets. The *similarity* between x_i and x_j is defined by

$$S'_{ij} = S'(G_i, G_j), \tag{26}$$

where G_i and G_j are two fuzzy sets on $E_i \cap E_j$ such that $G_i(e) = F(e)(x_i)$ and $G_j(e) = F(e)(x_j)$ for $\forall e \in E_i \cap E_j$. If $E_i \cap E_j = \emptyset$, write $S'_{ij} = 0$.

S'_{ij} is actually used to measure the similarity between values of x_i and x_j on parameters $E_i \cap E_j$. When the amount of data in the information table is huge, and the unknown items are rare, the similarity between values of x_i and x_j on all parameters can be approximately estimated by S'_{ij} .

The unknown entry h_{jl} is estimated according to the information from the relationship between the values of objects on parameter e_l by

$$h_{jl}^{\text{object}} = \frac{\sum_{i \in \{i|h_{il} \neq *\}} h_{il} \times S'_{ij}}{\sum_{i \in \{i|h_{il} \neq *\}} S'_{ij}}. \tag{27}$$

Definition 13. Suppose (F, E) is a fuzzy soft set on U and $e_k, e_l \in E$. Let S be a similarity measure of fuzzy sets. The *similarity* between e_k and e_l is defined by

$$S_{kl} = S(F_k, F_l), \tag{28}$$

where F_k and F_l are two fuzzy sets on $U_k \cap U_l$ such that $F_k(x) = F(e_k)(x)$ and $F_l(x) = F(e_l)(x)$ for $\forall x \in U_k \cap U_l$. If $U_k \cap U_l = \emptyset$, set $S_{kl} = 0$.

S_{kl} is actually used to measure the similarity between the values of objects in $U_k \cap U_l$ with regard to e_k and e_l . When the amount of data in the information table is huge and the unknown items are rare, the similarity between values of all objects on E_k and E_j can be approximately estimated by S_{kl} .

The unknown value h_{jl} is estimated according to the information from the relationship between the values of parameters regarding to object x_j by

$$h_{jl}^{\text{parameter}} = \frac{\sum_{k \in \{k|h_{jk} \neq *\}} h_{jk} \times S_{kl}}{\sum_{k \in \{k|h_{jk} \neq *\}} S_{kl}}. \tag{29}$$

Finally, the unknown entry h_{jl} can be predicted by linearly weighing h_{jl}^{object} and $h_{jl}^{\text{parameter}}$ as follows:

$$h_{jl} = w_1 \times h_{jl}^{\text{object}} + w_2 \times h_{jl}^{\text{parameter}}, \tag{30}$$

where w_1 and w_2 respectively denote the weight of objects and the weight of parameters quantifying the impact on unknown data, and satisfy the condition $w_1 + w_2 = 1$.

Given an incomplete fuzzy soft set (F, E) over U , h_{jl} is going to be predicted by the proposed object-parameter method. Our new approach can be implemented through Algorithm 3.

Algorithm 3. New adjustable object-parameter algorithm to predict the unknown data in an incomplete fuzzy soft set.

Step 1. Select a similarity measure S' of fuzzy sets. For all $i \in \{i|h_{il} \neq *\}$, calculate S'_{ij} .

Step 2. Calculate h_{jl}^{object} by Eqn. (27).

Step 3. Select a similarity measure S of fuzzy sets. For all $k \in \{k|h_{jk} \neq *\}$, calculate S_{kl} .

Step 4. Calculate $h_{jl}^{\text{parameter}}$ by Eqn. (29).

Step 5. Given a pair of weights w_1 and w_2 , obtain the final estimate value of h_{jl} by Eqn. (30).

Theorem 2. The value h_{jl} estimated by Algorithm 3 is in the interval $[0, 1]$.

Proof. Since $0 \leq h_{il} \leq 1$, we have that $h_{il} \times S'_{ij} \leq S'_{ij}$ for $\forall i \in \{i|h_{il} \neq *\}$. It is easy to deduce that $0 \leq \sum_{i \in \{i|h_{il} \neq *\}} h_{il} \times S'_{ij} \leq \sum_{i \in \{i|h_{il} \neq *\}} S'_{ij}$, which indicates that

$$0 \leq \frac{\sum_{i \in \{i|h_{il} \neq *\}} h_{il} \times S'_{ij}}{\sum_{i \in \{i|h_{il} \neq *\}} S'_{ij}} \leq 1,$$

i.e., $0 \leq h_{jl}^{\text{object}} \leq 1$.

In a similar way, we can prove that

$$0 \leq h_{jl}^{\text{parameter}} \leq 1.$$

At the same time, w_1 and w_2 satisfy the condition $w_1 + w_2 = 1$, so we have that $0 \leq w_1 \times h_{jl}^{\text{object}} + w_2 \times h_{jl}^{\text{parameter}} \leq 1$, i.e., $0 \leq h_{jl} \leq 1$. ■

Example 3. Let us return to the example shown in Table 2 and let the weights of objects and parameters be equal, i.e., $w_1 = w_2 = 1/2$.

Using Algorithm 3, let

$$S'(A, B) = \frac{1}{n} \sum_{i=1}^n \frac{\min(A(x_i), B(x_i))}{\max(A(x_i), B(x_i))}$$

(cf. Fan, 2002),

$$S(A, B) = \frac{1}{n} \sum_{i=1}^n \frac{2A(x_i)B(x_i)}{A(x_i)^2 + B(x_i)^2}$$

(cf. Li et al., 2014). Then we have $h_{31} \approx 0.580$, $h_{24} \approx 0.880$, $h_{34} \approx 0.800$, $h_{45} \approx 0.725$, $h_{57} \approx 0.461$. ♦

Now we pay attention to the principle of this new object-parameter method.

Algorithm 3 is proposed based on the opinion that, if h_{jl} is an unknown item to be predicted, the larger S'_{ij} , the more possible is h_{jl} is equal to h_{il} ; the larger S_{kl} , the more possible is that h_{jl} is equal to h_{jk} .

To illustrate this, let us return to the example shown in Table 2. It is already known that $h_{11} = 0.9$; when

Table 3. Comparative results of performance of the object-parameter predicting methods.

	USD-RMB	GBP-RMB	CAD-RMB	JPY-RMB	KRW-RMB
A_1	0.128022855	0.132633390	0.133498108	0.147776700	0.1065110061
A_2	0.000950457	0.004730404	0.002032659	0.000029923	0.0000058517
A_3	0.001749141	0.008605328	0.004158835	0.000049918	0.0000088618

we estimate the unknown item h_{31} according to the other already known items in column 1, the more similar ray 1 and ray 3 are, the more possible is that $h_{31} = h_{11} = 0.9$. It is already known that $h_{32} = 0.8$. When we estimate the unknown item h_{31} according to the other already known items in ray 3, the more similar columns 1 and 2 are, the more possible is that $h_{31} = h_{32} = 0.8$.

In Eqn. (27), we have

$$h_{jl}^{object} = \frac{\sum_{i \in \{i|h_{il} \neq *\}} h_{il} \times S'_{ij}}{\sum_{i \in \{i|h_{il} \neq *\}} S'_{ij}}$$

$$= \sum_{i \in \{i|h_{il} \neq *\}} (h_{il} \times \frac{S'_{ij}}{\sum_{i \in \{i|h_{il} \neq *\}} S'_{ij}}).$$

Here, $S'_{ij}/\sum_{i \in \{i|h_{il} \neq *\}} S'_{ij}$ can be viewed as the weight of h_{il} when calculating h_{jl}^{object} . For $e_l \in E$, the value of $\sum_{i \in \{i|h_{il} \neq *\}} S'_{ij}$ is determined. Hence, we make sure that the larger S'_{ij} is and the larger the weight of h_{il} is, the more possible is that $h_{jl}^{object} = h_{il}$.

In a similar way, by Eqn. (29), we make sure that the larger S_{kl} is, the more possible is to get $h_{jl}^{parameter} = h_{jk}$, i.e., the more similar h_{jk} and the estimated value of h_{jl} (based on the relationship between parameters) are.

Remark 4. Algorithm 3 has two main advantages compared with Algorithms 1 and 2:

- (i) If we select different similarity measures, the final estimated values of the unknown items will be different, which makes this method adjustable.
- (ii) Compared with Algorithms 1 and 2, the computational complexity of Algorithm 3 is obviously reduced.

7. Experiment

This section presents an experiment to compare the efficiency of the modified method to that of the Deng–Wang approach. The experimental database, which regards the exchange rate on 2016/01/05 (quoted by the Bank of China), can be found on the official website of the Bank of China: <http://srh.bankofchina.com/search/whpj/search.jsp>. The experimental results are analyzed and comparisons are made based on five data sets, i.e., the USD-RMB exchange rate data set, the GBP-RMB exchange rate data set, the CAD-RMB exchange rate data set, the JPY-RMB exchange rate

data set and the KRW-RMB exchange rate data set on 2016/01/05. The Bank of China provided statistics about 6 attributes at 95 different periods on 2016/01/05 in each database. In order to ensure that the original data sets are complete, in our experiment we choose four attributes: “Buying Rate”, “Cash Buying Rate”, “Selling Rate” and “Cash Selling Rate”. We then obtain five complete data sets, each of which has 95 objects and four attributes. These data sets are information tables, rather than fuzzy soft sets. We transfer these information tables into fuzzy soft sets by dividing each entry by 1000. After the division, every entry in the five information tables is in the interval $[0, 1]$, and every information table is transformed into a complete fuzzy soft set with 95 objects and four attributes.

We compare the predicting approaches in terms of the predictive accuracy, which is described by the mean error of the missing entries in an incomplete fuzzy soft set. The mean error is defined as

$$M = \frac{1}{n} \sum_{t=1}^n |A_t - F_t|, \tag{31}$$

where n is the number of missing data in an incomplete fuzzy soft set, A_t is the actual value and F_t is the forecast value.

On each complete fuzzy soft set, we randomly delete 1% entries to obtain an incomplete fuzzy soft set, and then compute the “mean error” by different algorithms. In order to obtain greater credibility of the results, we repeat the experiment program 100 times to compute the average “mean error” as the final one. All the algorithms are implemented as MATLAB programs. The experimental results are shown in Table 3.

In Table 3, every figure is accurate to 9 decimal places. A_1 denotes the mean error when predicting the unknown data by using the Deng–Wang approach (Algorithm 1), A_2 denotes the mean error when predicting the unknown data by using the improved Deng–Wang approach (Algorithm 2), whereas A_3 denotes the mean error when predicting the unknown data by using the adjustable object-parameter approach (Algorithm 3, and the similarity measures used here are the same as in Example 3). The experimental results in Table 3 indicate that both Algorithms 2 and 3 outperform Algorithm 1. The prediction is always the best when using Algorithm 2.

8. Conclusion

In this paper, we proposed an improvement of the Deng–Wang approach for unknown data predicting in an incomplete fuzzy soft set. Furthermore, using the similarity measure of fuzzy sets, a new adjustable object-parameter approach was given to predict the unknown data in an incomplete fuzzy soft set. All the three approaches: the Deng–Wang one, its improvement and the new adjustable object-parameter, can be used to predict unknown data in incomplete soft sets as well as in incomplete fuzzy soft sets. The elaborated results included a comparison between the three approaches through five rate exchange data sets. They indicate that the improved Deng–Wang approach performs best with respect to the forecasting accuracy.

Acknowledgment

This work has been supported by the National Natural Science Foundation of China (grants no. 61473239, 61175044, 61175055, 61603307).

References

- Alcantud, J.C.R. (2016). A novel algorithm for fuzzy soft set based decision making from multiobserver input parameter data set, *Information Fusion* **29**: 142–148.
- Atanassov, K.T. (1986). Intuitionistic fuzzy sets, *Fuzzy Sets and Systems* **20**(1): 87–96.
- Deng, T. and Wang, X. (2013). An object-parameter approach to predicting unknown data in incomplete fuzzy soft sets, *Applied Mathematical Modelling* **37**(6): 4139–4146.
- Fan, J. (2002). Some new similarity measures, *Journal of Xi'an Institute of Posts and Telecommunications* **3**(7): 69–71.
- Feng, F., Liu, X., Leoreanu-Fotea, V. and Jun, Y.B. (2011). Soft sets and soft rough sets, *Information Sciences* **181**(6): 1125–1137.
- Gau, W.L. and Buehrer, D.J. (1993). Vague sets, *IEEE Transactions on Systems, Man, and Cybernetics* **23**(2): 610–614.
- Herawan, T. and Deris, M.M. (2011). A soft set approach for association rules mining, *Knowledge-Based Systems* **24**(1): 186–195.
- Jiang, Y., Liu, H., Tang, Y. and Chen, Q. (2011). Semantic decision making using ontology-based soft sets, *Mathematical and Computer Modelling* **53**(5): 1140–1149.
- Jiang, Y., Tang, Y., Chen, Q., Liu, H. and Tang, J. (2010). Interval-valued intuitionistic fuzzy soft sets and their properties, *Computers & Mathematics with Applications* **60**(3): 906–918.
- Jun, Y.B., Lee, K.J. and Park, C.H. (2009). Soft set theory applied to ideals in d-algebras, *Computers & Mathematics with Applications* **57**(3): 367–378.
- Kong, Z., Wang, L. and Wu, Z. (2011). Application of fuzzy soft set in decision making problems based on grey theory, *Journal of Computational and Applied Mathematics* **236**(6): 1521–1530.
- Li, Y., Qin, K. and He, X. (2014). Some new approaches to constructing similarity measures, *Fuzzy Sets and Systems* **234**(1): 46–60.
- Li, Z., Wen, G. and Xie, N. (2015a). An approach to fuzzy soft sets in decision making based on grey relational analysis and Dempster–Shafer theory of evidence: An application in medical diagnosis, *Artificial Intelligence in Medicine* **64**: 161–171.
- Li, Z., Xie, N. and Wen, G. (2015b). Soft coverings and their parameter reductions, *Applied Soft Computing* **31**: 48–60.
- Li, Z. and Xie, T. (2014). The relationship among soft sets, soft rough sets and topologies, *Soft Computing* **18**(4): 717–728.
- Maji, P.K., Biswas, R. and Roy, A.R. (2001). Fuzzy soft sets, *Journal of Fuzzy Mathematics* **9**(3): 589–602.
- Molodtsov, D. (1999). Soft set theory—first results, *Computers & Mathematics with Applications* **37**(4): 19–31.
- Muthukumar, P. and Krishnan, G.S.S. (2016). A similarity measure of intuitionistic fuzzy soft sets and its application in medical diagnosis, *Applied Soft Computing* **41**: 148–156.
- Nowicki, R. (2010). On classification with missing data using rough-neuro-fuzzy systems, *International Journal of Applied Mathematics and Computer Science* **20**(1): 55–67, DOI: 10.2478/v10006-010-0004-8.
- Pawlak, Z. (1982). Rough sets, *International Journal of Computer & Information Sciences* **11**(5): 341–356.
- Qin, H., Ma, X., Herawan, T. and Zain, J.M. (2012a). DFIS: A novel data filling approach for an incomplete soft set, *International Journal of Applied Mathematics and Computer Science* **22**(4): 817–828, DOI: 10.2478/v10006-012-0060-3.
- Qin, H., Ma, X., Zain, J.M. and Herawan, T. (2012b). A novel soft set approach in selecting clustering attribute, *Knowledge-Based Systems* **36**: 139–145.
- Roy, A.R. and Maji, P. (2007). A fuzzy soft set theoretic approach to decision making problems, *Journal of Computational and Applied Mathematics* **203**(2): 412–418.
- Siwek, K. and Osowski, S. (2016). Data mining methods for prediction of air pollution, *International Journal of Applied Mathematics and Computer Science* **26**(2): 467–478, DOI: 10.1515/amcs-2016-0033.
- Wang, P. (1983). *Fuzzy Sets and Its Applications*, Shanghai Science and Technology Press, Shanghai.
- Xiao, Z., Gong, K. and Zou, Y. (2009). A combined forecasting approach based on fuzzy soft sets, *Journal of Computational and Applied Mathematics* **228**(1): 326–333.
- Xie, N., Han, Y. and Li, Z. (2015). A novel approach to fuzzy soft sets in decision making based on grey relational analysis and mycin certainty factor, *International Journal of Computational Intelligence Systems* **8**(5): 959–976.
- Xu, W., Ma, J., Wang, S. and Hao, G. (2010). Vague soft sets and their properties, *Computers & Mathematics with Applications* **59**(2): 787–794.

Yang, X., Lin, T.Y., Yang, J., Li, Y. and Yu, D. (2009). Combination of interval-valued fuzzy set and soft set, *Computers & Mathematics with Applications* **58**(3): 521–527.

Zadeh, L.A. (1965). Fuzzy sets, *Information and Control* **8**(3): 338–353.

Zhao, A. and Guan, H. (2015). Fuzzy-valued linguistic soft set theory and multi-attribute decision-making application, *Chaos, Solitons & Fractals* **89**: 2–7.

Zhong, N. and Skowron, A. (2001). A rough set-based knowledge discovery process, *International Journal of Applied Mathematics and Computer Science* **11**(3): 603–619.

Zou, Y. and Xiao, Z. (2008). Data analysis approaches of soft sets under incomplete information, *Knowledge-Based Systems* **21**(8): 941–945.

Yaya Liu received her MSc degree in mathematics from Southwest Jiaotong University, China. She is currently a PhD candidate at the College of Mathematics, Southwest Jiaotong University. Her research interests include soft sets, fuzzy sets, rough sets, and data mining.

Keyun Qin, PhD, Prof., received his BEd and MSc degrees in mathematics from Henan Normal University and Southwest Jiaotong University. He obtained his PhD from Southwest Jiaotong University in 1996. He is on the advisory committee of the CRSSC (Chinese Association for Artificial Intelligence, Rough Set and Soft Computing Society) and CCF TCMVFL (China Computer Federation, Multiple-Valued and Fuzzy Logic). As the main researcher, he has participated in 6 projects of the National Natural Science Foundation, 1 initiative of the Doctor Program of Higher Education, and 3 projects of Sichuan Provinces regarding applied basic research for major construction. He has published 60 research papers in domestic and foreign academic journals or conferences. His research interests include soft sets, multi-valued logic, uncertainty reasoning, rough set theory and methods.

Chang Rao received his Bachelor's degree in transportation engineering and control from Southwest Jiaotong University, China. He is now studying at Southwest Jiaotong University as a PhD candidate. His research interests include soft sets, data mining, railway signaling systems and safety-critical software testing.

Mahamuda Alhaji Mahamadu received his BSc degree in computer science and statistics from the University of Ghana in 1987 and his MSc degree in electrical engineering from the Kwame Nkrumah University of Science and Technology (KNUST) in 2007. He is currently pursuing a PhD program in communication engineering at Southwest Jiaotong University in China. His present research interests are in rough sets, data mining, wireless communications and multicarrier systems.

Received: 22 April 2016

Revised: 5 September 2016

Accepted: 15 October 2016