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

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

b College of Information Science and Technology
Southwest Jiaotong University, Chengdu 610031, Sichuan, PR China
e-mail: alvis.cc@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.

Molodstov (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.,
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 $\mu_F$ 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 $\mathcal{F}(U)$.

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

$$\mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\}$$
where } A, B \in F(U) \text{ and } x \in U. \\ 
If } \mu_A(x) \leq \mu_B(x), A \text{ is called a subset of } B, \text{ and this relationship is denoted by } A \subseteq B. \text{ Obviously, } A = B \text{ if } A \subseteq B \text{ 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 \text{ be the initial universe of objects and } E \text{ be the set of parameters in relation to objects in } U. \text{ Both } U \text{ and } E \text{ are assumed to be nonempty finite sets. Let } P(U) \text{ be the power set of } U \text{ and } A \subseteq E.

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

For any parameter } e \in A, F(e) \subseteq U \text{ may be considered the set of } e\text{-approximate elements of the soft set } (F, A). \text{ 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) \text{ is called a fuzzy soft set over } U, \text{ where } A \subseteq E \text{ and } F \text{ 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, \ldots, x_m\} \text{ and the parameters set is } E = \{e_1, e_2, \ldots, e_n\}. \text{ For an object } x_i \in U \text{ and a parameter } e_l \in E, \text{ denote by } h_{il} \text{ the value of } x_i \text{ on } e_l. \text{ 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) \text{ shown in Table 1, all membership values of objects on parameters are known except those of } x_2, x_3 \text{ on } e_2. \text{ The unknown data are denoted by ‘∗’ in the information table, i.e., } h_{22} = * \text{ 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 n\} \text{ and set } 0/0 = 0.

Consider an incomplete fuzzy soft set } (F, E) \text{ over } U \text{ and let } h_{jl} \text{ be the unknown value which is to be predicted.

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

<table>
<thead>
<tr>
<th>U</th>
<th>e_1</th>
<th>e_2</th>
<th>e_3</th>
<th>e_4</th>
<th>e_5</th>
<th>e_6</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_1</td>
<td>0.9</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>x_2</td>
<td>0.8</td>
<td>*</td>
<td>0.5</td>
<td>0.7</td>
<td>0.6</td>
<td>0.3</td>
</tr>
<tr>
<td>x_3</td>
<td>0.4</td>
<td>*</td>
<td>0.9</td>
<td>0.9</td>
<td>0.5</td>
<td>0.9</td>
</tr>
<tr>
<td>x_4</td>
<td>0.9</td>
<td>0.8</td>
<td>0.9</td>
<td>0.4</td>
<td>0.7</td>
<td>0.5</td>
</tr>
</tbody>
</table>

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

\[ d_{jk} = \frac{h_{jk} - h_{ik}}{m_{ik} - m_{jk}}. \]

Moreover, the complete distance from } x_i \text{ to } x_j \text{ on all the parameters is defined by

\[ d_{ij} = \frac{\sum_{k=1}^{n} d_{jk}}{|\{k | (a \in U_k) \wedge (j \in U_k)\}|}. \]

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

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

\[ h_{jl} = \frac{\sum_{i \in U_l} (h_{il} - d_{ij})}{|U_l|}. \]

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

\[ r_{ik} = \frac{h_{ik} - h_{il}}{h_{ik} + h_{il}}. \]

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

\[ c_{kl} = \frac{\sum_{i \in U_k \cap U_l} r_{ik}}{|U_k \cap U_l|}. \]

Moreover, the degree of average dominance of } e_k \text{ to } e_l \text{ is given by

\[ v_{kl} = \frac{c_{kl}}{\sum_{q \in U_k \cap U_l \neq \emptyset} |c_{ql}|}. \]
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 $G_j = \{k | (h_{jk} \neq \star) \cap (U_k \cap U_l \neq \emptyset), 1 \leq k \leq n\}$.

Finally, the unknown entry $h_{jl}$ can be predicted by linearly weighting $h_{jl}^\text{object}$ 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 $d_{jk}^k$ 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 $r_{ij}^k$, $c_{kl}$ 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 $d_{14} \approx -0.253$, $d_{23} \approx -0.221$, $d_{43} \approx -0.011$, $d_{63} \approx -0.292$, $d_{63} \approx -0.299$, so we have $h_{31}^\text{object} \approx 1.095$;

<table>
<thead>
<tr>
<th>$U$</th>
<th>$c_1$</th>
<th>$c_2$</th>
<th>$c_3$</th>
<th>$c_4$</th>
<th>$c_5$</th>
<th>$c_6$</th>
<th>$c_7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>0.9</td>
<td>0.4</td>
<td>0.1</td>
<td>0.9</td>
<td>0.6</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>$x_2$</td>
<td>0.8</td>
<td>0.6</td>
<td>0.5</td>
<td>$\star$</td>
<td>0.5</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>$x_3$</td>
<td>0.9</td>
<td>0.8</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>$x_4$</td>
<td>0.9</td>
<td>0.8</td>
<td>0.9</td>
<td>0.8</td>
<td>$\star$</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>$x_5$</td>
<td>0.9</td>
<td>0.2</td>
<td>0.2</td>
<td>0.6</td>
<td>0.3</td>
<td>0.4</td>
<td>$\star$</td>
</tr>
<tr>
<td>$x_6$</td>
<td>0.9</td>
<td>0.2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>

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 $d_{ij}$ is the complete distance from $x_i$ to $x_j$ should be subtracted from $h_{il}$. Through every $h_{jl}$ ($h_{jl} \neq \star, i \in \{1, 2, \ldots, m\}$), 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 $h_{il}$ ($h_{il} \neq \star, i \in \{1, 2, \ldots, m\}$), it is no surprise that the average of these estimates, i.e., $h_{jl}^\text{parameter}$, 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 $h_{jk}$ ($h_{jk} \neq \star, k \in \{1, 2, \ldots, n\}$), 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 $h_{jk}$ ($h_{jk} \neq \star, k \in \{1, 2, \ldots, n\}$), 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 \( h_{ij} \) be an unknown item. Write \( h_{ij}^1 = h_{il} - d_{ij} \) for the estimate of \( h_{jl} \) obtained by considering the distance \( d_{ij} \) from \( x_i \) to \( x_j \), and \( h_{ij}^2 = h_{il} + d_{ij} \) for the estimate of \( h_{jl} \) obtained by considering the distance \( d_{ij} \) from \( x_i' \) to \( x_j' \). If \( x_i \) and \( x_i' \) have different effects on \( x_j \), then \( h_{ij}^1 \) and \( h_{ij}^2 \) are expected to have different effects on \( h_{jl} \). That is, one should set different weights for \( h_{ij}^1 \) and \( h_{ij}^2 \) when calculating \( h_{jl} \), 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 \( U_l = \{i | h_{il} \neq *, 1 \leq i \leq n\} \) and \( E_l = \{l | h_{il} \neq *, 1 \leq l \leq n\} \). It is obvious that \( U_l \subseteq U \) and \( E_l \subseteq E \). 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

\[
\begin{align*}
  d_{jk}^{ih} &= h_{ik} - h_{jk}.
\end{align*}
\]

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

\[
\begin{align*}
  d_{ij} &= \frac{\sum_{k \in E \cap E_l} d_{jk}^{ih}}{|E \cap E_l|}.
\end{align*}
\]

Since \( h_{ik}, h_{jk} \in [0, 1] \), by Eqns. (9) and (10) it is easy to obtain that \( d_{jk}^{ih} \in [-1, 1] \) 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 \( h_{ij} \) predicted by an already known value \( h_{kl} \), which is denoted by \( h_{ij}^{kl} \), can be calculated in the following manner:

\[
\begin{align*}
  h_{ij}^{kl} &= h_{il} - d_{ij}.
\end{align*}
\]

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

\[
\begin{align*}
  h_{ij}^{kl} &= \begin{cases} 
    h_{il} & \text{if } d_{ij} = 0, \\
    h_{il} - d_{ij} & \text{if } -1 \leq d_{ij} < 0, \\
    h_{il} - d_{ij} + d_{ij} h_{il} & \text{if } 1 \geq d_{ij} > 0,
  \end{cases}
\end{align*}
\]

i.e.,

\[
\begin{align*}
  h_{ij}^{kl} &= \begin{cases} 
    h_{il} & \text{if } d_{ij} = 0, \\
    h_{il} - d_{ij} + d_{ij} h_{il} & \text{if } -1 \leq d_{ij} < 0, \\
    h_{il} - d_{ij} & \text{if } 1 \geq d_{ij} > 0.
  \end{cases}
\end{align*}
\]

Lemma 1. By Eqn. (13) \( h_{ij}^{kl} \) 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 \( 0 \leq h_{ij}^{kl} = h_{il} \leq 1 \).

(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

\[
\begin{align*}
  h_{il} - d_{ij} + d_{ij} h_{il} &= h_{il} + d_{ij} (h_{il} - 1) \geq h_{il} \\
  h_{il} - d_{ij} + d_{ij} h_{il} - 1 &= h_{il} (1 + d_{ij}) - (1 + d_{ij}) \\
  &= (h_{il} - 1) (1 + d_{ij}) \leq 0, \quad \text{i.e.,} \quad h_{il} - d_{ij} + d_{ij} h_{il} \leq 1.
\end{align*}
\]

Thus we obtain \( h_{il} \leq h_{ij}^{kl} \leq 1 \). Also, since

\[
\begin{align*}
  h_{ij}^{kl} &= h_{il} - d_{ij} + d_{ij} h_{il} = (h_{il} - 1) d_{ij} + h_{il} \\
  &\leq -d_{ij} + h_{il} \leq -1 \leq h_{il} - 1 \leq 0,
\end{align*}
\]

we know that the value \( h_{ij}^{kl} \) increases monotonically as \( d_{ij} \) decreases monotonically. If \( d_{ij} \) reaches the minimal value \( -1 \), it is easy to obtain that
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

\[
D_{ij} = \frac{1 - |d_{ij}|}{\sum_{p \in M_i} (1 - |d_{pj}|)}. \tag{14}
\]

Write \(M_i = \{i|\{h_i \neq \ast\} \cap \{E_i \cap E_j \neq \emptyset\}, 1 \leq i \leq m\}\). The relative dominant degree of \(x_i\) over \(x_j\) with respect to \(e_l\) is defined by

\[
D_{ij}^{il} = \frac{1 - |d_{ij}|}{\sum_{p \in M_i} (1 - |d_{pj}|)}. \tag{15}
\]

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

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

\[
h_{jl}^{\text{object}} = \sum_{i \in M_l} h_{il}^{il} \times D_{ij}^{il}. \tag{16}
\]

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

\[
e_{kl}^{il} = h_{ik} - h_{il}. \tag{17}
\]

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

\[
c_{kl} = \frac{\sum_{i \in U_k \cap U_l} e_{kl}^{il}}{|U_k \cap U_l|}. \tag{18}
\]

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

Remark 2. Both Eqns. (4) and (17) can be used to measure the distance from \(h_{ik}\) to \(h_{lj}\). 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 \(h_{kl}\) can be predicted by \(h_{jk}\) in the following manner:

\[
h_{jl} = h_{jk} - c_{kl}. \tag{19}
\]

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

\[
h_{jl}^{k'} = \begin{cases} 
    h_{jk} & \text{if } c_{kl} = 0, \\
    h_{jk} - c_{kl} & \text{if } -1 \leq c_{kl} < 0, \\
    h_{jk} - c_{kl} & \text{if } 1 \geq c_{kl} > 0,
\end{cases} \tag{20}
\]

Lemma 2. By Eqn. (21) \(h_{jl}^{k'}\) 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

\[
C_{kl} = \frac{1 - |c_{kl}|}{\sum_{q \leq l} |U_q \cap U_l| (1 - |c_{ql}|)}. \tag{22}
\]

Write \(G_{il} = \{k|(h_{ik} \neq \ast) \cap (U_k \cap U_l \neq \emptyset), 1 \leq k \leq n\}\). The relative dominant degree of \(e_k\) over \(e_l\) regarding \(x_j\) is characterized by

\[
C_{jl}^{ik} = \frac{1 - |c_{kl}|}{\sum_{q \in G_{il}} (1 - |c_{ql}|)}. \tag{23}
\]

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 \(C_{jl}^{ik}\) in Eqn. (23) is used to describe the possibility that the value of \(h_{jl}\) \((h_{jl} = \ast)\) will be determined by \(h_{ik}\) \((h_{ik} \neq \ast)\). For \(x_j \in U\), if there is one and only one parameter \(e_l\) s.t. the membership value of \(x_j\) on \(e_l\) \((h_{jl} = \ast)\) is unknown, then the relative dominant degree degenerates to the dominant degree of \(e_k\) over \(e_l\), i.e., \(C_{jl}^{ik} = C_{kl}\).
Object-parameter approaches to predicting unknown data in an incomplete fuzzy soft set

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

\[
h_{ji}^{\text{parameter}} = \sum_{k \in G_j} h_{jk}^{il} \times C_{jk}^{il}.
\]

(24)

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

\[
h_{ji} = w_1 \times h_{ji}^{\text{object}} + w_2 \times h_{ji}^{\text{parameter}},
\]

(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 \( U \), \( h_{ji} \) 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 \( i \in M_i \), calculate \( h_{ji}^{il} \) by Eqns. (9), (10) and (13), and then calculate \( D_{ji}^{il} \) by Eqn. (15).

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

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

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

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

**Theorem 1.** The value \( h_{ji} \) estimated by Algorithm 2 is in the interval \([0,1]\).

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

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

(i) Hidden information between objects revealed.

(ii) Hidden information between parameters is revealed.

(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 \( h_{31} \approx 0.626 \), \( h_{24} \approx 0.987 \), \( h_{34} \approx 0.959 \), \( h_{45} \approx 0.841 \), \( h_{57} \approx 0.308 \).
Definition 12. Suppose \((F, E)\) is a fuzzy soft set on \(U\) and \(x_i, x_j \in U\). Let \(S^{'i}\) be a similarity measure of fuzzy sets. The similarity between \(x_i\) and \(x_j\) is defined by

\[
S^{'i}_{ij} = S'(G_i, G_j),
\]

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 \(E_i \cap E_j = \emptyset\), define \(S^{'i}_{ij} = 0\).

\(S^{'i}_{ij}\) 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 \(S^{'i}_{ij}\).

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

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

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 \(E_l\) 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

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

Finally, the unknown entry \(h_{jl}\) can be predicted by linearly weighing \(h_{jl}^{\text{object}}\) 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 impacting the confidence on unknown data, and satisfy the condition \(w_1 + w_2 = 1\).

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 \(S^{'i}\) of fuzzy sets. For all \(i \in \{i|h_{il} \neq *\}\), calculate \(S^{'i}_{ij}\).

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

**Step 3.** Select a similarity measure \(S\) of fuzzy sets. For all \(k \in \{k|h_{ik} \neq *\}\), 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 \(h_{il} \times S^{'i}_{ij} \leq S^{'i}_{ij}\) for \(\forall i \in \{i|h_{il} \neq *\}\). It is easy to deduce that

\[
0 \leq \sum_{i \in \{i|h_{il} \neq *\}} h_{il} \times S^{'i}_{ij} \leq \sum_{i \in \{i|h_{il} \neq *\}} S^{'i}_{ij},
\]

which indicates that

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

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., \(0 \leq h_{jl} \leq 1\). \(\blacksquare\)

**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

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

(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 \(h_{31} \approx 0.580, h_{24} \approx 0.880, h_{34} \approx 0.800, h_{45} \approx 0.725, h_{52} \approx 0.461\). 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 \(S^{'i}_{ij}\), the more possible is \(h_{jl}\) is equal to \(h_{ij}\); the larger \(S_{kl}\), the more possible is that \(h_{ik}\) 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
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

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

Here, $S'_{ij}/\sum_{i \in \{i | h_{il} \neq *\}} S'_{ij}$ can be viewed as the weight of $h_{il}$ when calculating $h_{jl}^{\text{object}}$. For $e \in E$, 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 $h_{jl}^{\text{object}} = h_{il}$.

In a similar way, by Eqn. (29), we make sure that the larger $S_{jk}$ is, the more possible is to get $h_{jl}^{\text{parameter}} = h_{jk}$, 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://xhr.bankofchina.com/search/whpj/search.jsp](http://xhr.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

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

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.

<table>
<thead>
<tr>
<th></th>
<th>USD-RMB</th>
<th>GBP-RMB</th>
<th>CAD-RMB</th>
<th>JPY-RMB</th>
<th>KRW-RMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>0.12802285</td>
<td>0.13263390</td>
<td>0.13349811</td>
<td>0.14777670</td>
<td>0.10651101</td>
</tr>
<tr>
<td>$A_2$</td>
<td>0.000950457</td>
<td>0.004730404</td>
<td>0.002032659</td>
<td>0.000029923</td>
<td>0.0000088618</td>
</tr>
<tr>
<td>$A_3$</td>
<td>0.001749141</td>
<td>0.008605328</td>
<td>0.004158835</td>
<td>0.000049918</td>
<td>0.0000088618</td>
</tr>
</tbody>
</table>
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

Object-parameter approaches to predicting unknown data in an incomplete fuzzy soft set


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