

# A Weighted Attribute Decision Making Approach in Incomplete Soft Set

Lishi Zhang  
 School of Science  
 Dalian Ocean University  
 Dalian, China  
 e-mail: dwtag@sohu.com

**Abstract**—Kong. et al. (2014) [1] presented an efficient approach which gives weighted-average of all possible choice values of soft sets under incomplete information system, However, the weights of the attributes are not considered, although all the entries are “1”, the importance of all attributes are not all the same, in our paper, we first combined the weights of attributes and Kong’s approach together, the weights of attributes are derived from the data directly, much computation are involved, It is demonstrated that the proposed approach are reasonable in finding the choice values of soft sets. Second, we give a concise proof of the main theorem in Kong. et al. (2014) [1]

**Keywords**—component; soft set; attribute; incomplete information; choice value; weighted-average.

## I. INTRODUCTION

Information originates from real applications is useful for data mining which aims at finding the knowledge in it, however, these data are often inherently related with incompleteness, uncertainty, fuzziness, which is inherent in data, the phenomena comes from incompleteness of data, limitations of measuring device, and error of data collecting and even the inputting process. In data processing, the information must be whole in general, the missing data should be estimated in some way, before machine program is employed, a whole data is necessary.

There are many ways to get the whole data from the incomplete one: most ways are related to the statistic approach: replacing the empty entries with the sample mean; using a learning algorithm such as EM, max likelihood to estimate the missing entry.

As a usual technique for dealing with incomplete data, soft set theory has been successfully applied in many categories such as decision making[2-10], data analysis[11-12], rule mining[13].

Decision making analysis in the field of a soft set is one of the most important tasks. The complete data has been extensively explored, not many researches are carried out on data analysis approaches under incomplete information. Zou et al. gave a data analysis approaches of soft sets under incomplete information, the final decision value of an object is calculated by weighted-average of all

possible choice values of the object [11]. It is similar to the expectations of variables in probability. An object-parameter method is proposed to infer missing data in incomplete fuzzy soft sets by Deng [12]. The weighted-average of all possible choice values approach is complicated with huge volume of calculations. After adding parameters or deleting parameters, the all data must be reordered and calculated again, the process is time assuming, Kong [1] present a simplified probability approach, and illustrated the equivalence between the weighted-average of all possible choice values approach and the simplified probability approach.

In our paper, we proposed an attribute-weighted approach, we first calculated the weights of the attributes by the Kong’s [1] method, then the data was transformed into a new database with the weights, furthermore, we calculate the choice value with the Kong’s [1] method.

The rest of this paper is structured as follows. Section II reviews the basic definitions of soft set theory. Section III discusses the Weight-average of all possible choice value approach proposed in [1]. In Section IV, the new approach is proposed. In Section V, two approaches are compared and demonstrated the effectiveness of our proposed approach, Finally Section IV draws the conclusion.

## II. PRELIMINARIES

In the present section we will briefly recall some basic definitions and background knowledge of soft sets

**Definition 2.1.** A pair  $(F, E)$  is called a soft set (over  $U$ ) if and only if  $F$  is a mapping of  $E$  into the set of all subsets of the set  $U$ , i.e.,  $F: E \rightarrow P(U)$ , where  $P(U)$  is the power set of  $U$ .

The soft set is a parameterized family of subsets of the set  $U$ . Every set  $F(e)$ ,  $e \in E$ , from this family may be considered as the set of  $e$ -elements of the soft set  $(F, E)$ , or as the  $e$ -approximate elements of the soft set.

**Example 2.1.** Let universe  $U = \{h_1, h_2, h_3, h_4\}$  be a set of houses, a set of parameters  $E = \{e_1, e_2, e_3, e_4\}$  be a set of status of houses which stand for the parameters “beautiful”, “cheap”, “in green surroundings”, and “in good location” respectively. The mapping  $F$  be a mapping of  $E$  into the set of all subsets of the set  $U$ . Now consider a soft set  $(F, E)$  that describes the “attractiveness of houses for purchase”, where  $F(e_1) = \{h_1, h_3, h_4\}$ ,  $F(e_2) = \{h_1, h_2\}$ ,  $F(e_3) = \{h_1, h_3\}$ , and  $F(e_4) = \{h_2, h_3, h_4\}$ .

A two-dimensional table is used to represent the soft set  $\{F, E\}$ . Table I is the tabular form of the soft set  $\{F, E\}$ . If  $h_i \in F(e_j)$ , then  $h_{ij} = 1$ , otherwise  $h_{ij} = 0$ , where  $h_{ij}$  are the entries (see Table I).

Suppose  $U = \{h_1, h_2, \dots, h_n\}$ ,  $E = \{e_1, e_2, \dots, e_m\}$ ,  $(F, E)$  is a soft set with tabular representation.

Define,

$$f_E(h_i) = \sum_j \{h_{ij}\}$$

where  $h_{ij}$  are the entries in soft set table.

TABLE I. THE TABULAR REPRESENTATION OF  $\{F, E\}$

$U$	$e_1$	$e_2$	$e_3$	$e_4$
$h_1$	1	1	1	0
$h_2$	0	1	0	1
$h_3$	1	0	1	1
$h_4$	1	0	0	1

Zou’s algorithm based on weight-average of all possible choice value of the object is shown as follows [11]:

1. Calculate all possible choice values for each object.
2. Calculate probabilities  $p_{e_i}$  and  $q_{e_i}$  that an object belongs to and does not belong to  $F(e_i)$  if for parameter  $e_i$  there are incomplete information, that is the entry  $h_{ki}$  is unknown,  $k = 1, \dots, n$ .
3. Calculate the weight of possible choice value for each object according to the following equations. Where  $d$  is the number of incomplete information in each object.

$$k = \begin{cases} \prod_{e \in E_0^*} q_e & x = 0 \\ \sum_{C_j^*} \left( \left( \prod_{e_j \in E_0^*} p_{e_j} \right) \left( \prod_{e_i \in E_0^*} q_{e_i} \right) \right), & 0 < x < d \\ \prod_{e \in E_1^*} p_e & x = d \end{cases}$$

4. Calculate the decision values  $d_i = \sum_{i=1}^m k_i c_i$ . The optimal decision is  $\max\{d_i\}$ . Where  $k_i$  is the weight of the choice value  $c_i$ ,  $c_i$  is the choice value of the object  $h_i$ ,  $c_i = \sum_{j=1}^y h_{ij}$ , Where  $y$  is the number of parameters in a soft set.

TABLE II. THE TABULAR REPRESENTATION OF  $\{F, E\}$

$U$	$e_1$	$e_2$	$e_3$	$e_4$
$h_1$	1	1	1	0
$h_2$	0	1	*	1
$h_3$	1	*	1	1
$h_4$	*	0	0	*

TABLE III. THE TABULAR REPRESENTATION OF  $\{F, E\}$

$U$	$c_i$	$d_i$
$h_1$	{2}	3
$h_2$	{2,3}	3
$h_3$	{2,3}	2.67
$h_4$	{1,2,3}	2

TABLE IV. THE TABULAR REPRESENTATION OF  $\{F, E\}$

$U$	$e_1$	$e_2$	$e_3$	$e_4$	$d_i$
$h_1$	1	1	1	0	3
$h_2$	0	1	0	1	3
$h_3$	0	2/3	1	1	2.67
$h_4$	1/3	0	0	2/3	2

**Example 3.1.** Let us consider the following information, the tabular representation of  $(F, E)$  is shown in Table II. By Zou’s approach, the calculation process is as follows:

1. Calculate all possible choice values for each object,  $C_1 = \{3\}$ ;  $C_2 = \{2, 3\}$ ;  $C_3 = \{2, 3\}$ ;  $C_4 = \{1, 2, 3\}$ . The decision values of all objects in the soft set  $(F, E)$  are listed in Table III.

2. Calculate probabilities  $p_{e_i}$  and  $q_{e_i}$

$$q_{e_1} = 2/3, p_{e_1} = 1/3; q_{e_2} = 1/3, p_{e_2} = 2/3;$$

$$q_{e_3} = 0, p_{e_3} = 1; q_{e_4} = 1/3, p_{e_4} = 2/3;$$

3. Calculate the weight of possible choice value for each object:

For object  $h_1$ , the choice value is complete, so the weight is not necessary.

For object  $h_2$ ,  $k_{21} = q_{e_3} = 0, k_{22} = p_{e_3} = 1$ .

For object  $h_3$ ,  $k_{31} = q_{e_2} = 1/3, k_{32} = p_{e_2} = 2/3$ .

For object  $h_4$

$$k_{41} = q_{e_1} q_{e_4} = 2/9,$$

$$k_{42} = q_{e_1} p_{e_4} + p_{e_1} q_{e_4} = 4/9 + 1/9 = 5/9,$$

$$k_{43} = p_{e_1} p_{e_4} = 2/9.$$

4. Calculate the decision value.

$$d_1 = 3$$

$$d_2 = 2 \times k_{21} + 3 \times k_{22} = 3$$

$$d_3 = 2 \times k_{31} + 3 \times k_{32} = 2.67$$

$$d_4 = 1 \times k_{41} + 2 \times k_{42} + 3 \times k_{43} = 2$$

The above approach is rendered by Zhou [11], while a simplified approach is given by Kong [1] as follows:

Let  $(F, E)$  be a soft set with incomplete information over  $U = \{h_1, h_2, \dots, h_n\}$ ,  $E = \{e_1, e_2, \dots, e_m\}$ . For an

object  $h_i$ , there are  $d$  cells with “\*” in the corresponding row in the tabular representation, let

$$h_{ij} = *, E^* = \{e_j \mid h_{ij} = *\};$$

Let  $s$  be the number of cells with the value of 1 in this row, let

$$h_{ij} = 1, E^1 = \{e_j \mid h_{ij} = 1\};$$

While  $m-s-d$  be the number of cells with the value of 0 in this row, let

$$h_{ij} = 0, E^0 = \{e_j \mid h_{ij} = 0\};$$

For an object  $h_i$ , if  $h_{ij} = *$ , probabilities

$$p_{e_j} = \frac{n_1}{n_1 + n_0}, q_{e_j} = \frac{n_0}{n_1 + n_0}$$

stand for an object belonging to or not to  $F(e_j)$ , where  $n_1$  and  $n_0$  stand for the number of objects that belong to and does not belong to  $F(e_j)$ . According to the simplified probability approach, the decision value is calculated by

$$d_i' = s + \sum_{e_j \in E^*} p_{e_j}'$$

### III. THE SHOTR PROOF OF KONG’S THEOREM

In this section, a concise proof is given to verify the main theorem.

**Proof:** Let the  $k$  row has  $t$  position with “\*” and  $p$  position with “1” in the corresponding column, the probabilities of the column are

$$r_1^0, r_1^1, r_2^0, r_2^1, \dots, r_t^0, r_t^1,$$

Where  $r_s^0, r_s^1$  denote the percentage of “0” and “1” of the  $s$  Column respectively, and  $r_s^0 + r_s^1 = 1$ .  $d_1$  denotes the choice value of Zhou’s approach, then

$$\begin{aligned} d_1 &= \sum_{s=0}^t (p+s) \sum_{k_1+k_2+\dots+k_t=s} r_1^{k_1} r_2^{k_2} \dots r_t^{k_t} \\ &= p r_1^0 r_2^0 \dots r_t^0 + (p+1)(r_1^1 r_2^0 \dots r_t^0 + r_1^0 r_2^1 \dots r_t^0 + \dots + r_1^0 r_2^0 \dots r_t^1) \\ &\quad + (p+2)(r_1^1 r_2^1 r_3^0 \dots r_t^0 + r_1^0 r_2^1 r_3^1 \dots r_t^0 + \dots + r_1^0 r_2^0 \dots r_{t-1}^1 r_t^1) \\ &\quad + (p+t) r_1^1 r_2^1 r_3^1 \dots r_t^1 \\ &= p(r_1^0 r_2^0 \dots r_t^0 + r_1^1 r_2^0 \dots r_t^0 + r_1^0 r_2^1 \dots r_t^0 + \dots + r_1^0 r_2^0 \dots r_t^1) \\ &\quad + L + r_1^1 r_2^1 r_3^0 \dots r_t^0 + r_1^0 r_2^1 r_3^1 \dots r_t^0 + \dots + r_1^0 r_2^0 \dots r_{t-1}^1 r_t^1 \\ &\quad + L + r_1^1 r_2^1 r_3^1 \dots r_t^1 + \dots \dots \dots (0) \end{aligned}$$

$$1(r_1^1 r_2^0 \dots r_t^0 + r_1^0 r_2^1 \dots r_t^0 + \dots + r_1^0 r_2^0 \dots r_t^1) + \dots \dots \dots (1)$$

$$2(r_1^1 r_2^1 r_3^0 \dots r_t^0 + r_1^0 r_2^1 r_3^1 \dots r_t^0 + \dots + r_1^0 r_2^0 \dots r_{t-1}^1 r_t^1) + \dots \dots \dots (2)$$

$$\dots \dots \dots$$

$$t(r_1^1 r_2^1 r_3^1 \dots r_t^1)$$

he item  $s \sum_{k_1+k_2+\dots+k_t=s} r_1^{k_1} r_2^{k_2} \dots r_t^{k_t}$ , the addition component number is  $kC_n^k$ , as  $kC_n^k = nC_{n-1}^{k-1}$ , reordering the addition components, we get that there are  $C_{n-1}^{k-1}$  items containing  $r_s^1$ , the sum of all the components is

$$\begin{aligned} r_s^1 (r_1^0 + r_1^1) \dots (r_{s-1}^0 + r_{s-1}^1) (r_{s+1}^0 + r_{s+1}^1) \dots (r_t^0 + r_t^1) &= r_s^1 \\ p(r_1^0 + r_1^1) (r_2^0 + r_2^1) \dots (r_t^0 + r_t^1) &= p \end{aligned}$$

Consequently,

$$d_1 = p + r_1^1 + r_2^1 L + r_t^1$$

### IV. PROPOSED APPROACH

The entries of the objects are all “1” in the table, but all the importance of attributes are not equal, for a given data set, the weights of attributes are not always given in advance, it is reasonable to derive them from the concrete dataset, in the present paper, we first derive the weights from the data by Kong’s method, then we transform the dataset with the weights, then again with Kong’s method, we get the final choice value.

The proposed algorithms is as the follows

Let  $(F, E)$  be a soft set with incomplete information over  $U = \{h_1, h_2, \dots, h_n\}$ ,  $E = \{e_1, e_2, \dots, e_m\}$ .

1. For the attribute  $e_k$ , sum all the entries with “1” to  $t_k$ , find all the missing data with the marks “\*”, find the corresponding rows with “\*”, Calculate probabilities  $r_{h_i}$  and  $s_{h_i}$  that the attributes which object shares and does not share.

2. Calculate the weights

$w_k = t_k / \sum r_{h_i}$ ,  $h_i$  are all the objects that have “\*” at attributes  $e_k$ , let

$$v_k = w_k / \sum_{k=1}^m w_k$$

3. The  $k$ th column of data are multiplied by  $v_k$ , from the new dataset, with the Kong’s method, calculate the choice value.

TABLE V. THE TABULAR REPRESENTATION OF  $\{H, A\}$

$U$	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$	$e_9$	$e_{10}$
$h_1$	1	0	1	0	1	1	1	*	1	0
$h_2$	0	1	1	*	0	1	0	1	1	1
$h_3$	1	1	0	1	1	1	1	0	1	1
$h_4$	1	*	0	1	0	0	0	0	1	1
$h_5$	1	0	1	0	1	1	*	0	*	0
$h_6$	1	0	1	*	0	1	1	*	1	1
$h_7$	1	1	0	1	1	1	0	0	1	0
$h_8$	*	1	1	1	0	0	0	1	1	0
$h_9$	0	0	0	1	1	0	1	1	0	0
$h_{10}$	0	1	0	1	1	1	*	0	1	1
$h_{11}$	1	0	1	1	1	1	0	0	*	1
$h_{12}$	0	0	0	1	1	0	0	0	1	0

**Example 3.1.** We use example in [11], 10 factors are used to evaluate the quality of information systems, a questionnaire with 50 attendees was given in the staff of a company. But in the process of collecting and coordinating data, because of missing or unclear data, the result of investigation is incomplete. So the data analysis approach under incomplete information must be used. The universe  $U$  of  $(H, A)$  only comprises of 12 objects which represent different user preferences which are denoted by  $h_1$  to  $h_{12}$ , parameter set  $A$  represents 10 factors used to appraise the quality of the information system. These factors are “authenticity”, “share”, “correctness”, “in time”, “economy”, “validity”, “integrality”, “security”, “consistency”, “dependability” which are denoted by  $e_1$  to  $e_{10}$ . By computing, we get for  $e_4$

$$t_4 = 5, r_{h_2} = 6/9, r_{h_4} = 6/8, w_4 = 5 + 6/9 + 6/8 = 6.4167$$

All the weights are listed as follows

$$w_1 = 7.3333, w_2 = 5.5556, w_3 = 9, w_4 = 6.4167, w_5 = 8$$

$$w_6 = 8, w_7 = 5.4028, w_8 = 6.4167, w_9 = 10.2417, w_{10} = 6$$

Normalizing them, we get

$$v = [0.1041, 0.0789, 0.1278, 0.0911, 0.1136,$$

$$0.1136, 0.0767, 0.0627, 0.1462, 0.0852]$$

Then attribute  $e_4$  corresponds to the 4<sup>th</sup> column dataset

$$[0, *, 0, 0.0911, 0.0911, *, 0, 0, 0.0911, 0.0911, 0.0911, 0]$$

By comparing two approaches, we have that Kong's simplified approach is simpler, but they do not consider the weights of attributes, the proposed approach is somewhat complex which involves more computation, but the weights of attributes are used to transformed into the data, it is more applicable to mine the knowledge.

From table VI, we see that the proposed method keeps the first two choices, that is the optimal value and suboptimal value, while the last four ones are also the same.

TABLE VI. THE COMPARISON OF THE CHOICE VALUE OF  $\{H, A\}$

$U$	$d_i$	Order(from big to small)	$d_p$	Order(from big to small)
$h_1$	6.3	10	0.4598	10
$h_2$	6.5	3	0.6211	3
$h_3$	7	11	0.9720	6
$h_4$	5.54	6	1.1100	1
$h_5$	6.3	2	1.1753	11
$h_6$	6.8	1	1.2811	7
$h_7$	6	5	1.2842	2
$h_8$	3.64	7	1.3022	5
$h_9$	5	4	1.3487	4
$h_{10}$	7.4	9	1.4036	9
$h_{11}$	6.9	8	1.4183	8
$h_{12}$	2	12	1.5342	12

## V. CONCLUSION

In this paper, the decision making problem in incomplete soft set is discussed. In Zhou's paper, the weight-average of all possible choice values are used, approach, it coincides with the expectation of variables in probability, while its the weighted-average of all possible choice values approach refers to great amount of calculations. In Kong's paper, a simplified probability approach is used which is straightforward in computation, it is relatively simple and easy to compute. An equivalence between the weighted-average of all possible

choice values approach and the simplified probability approach is established, the present paper use all the above method, furthermore, the weights of attributes are introduced from the raw data, it reflects the uncovered favored tendency the attributes have, the comparison results illustrate that the new proposed approach has a almost the same choice order with the former approaches. The future effort ought to make in incomplete data in other objects with uncertainty.

## REFERENCES

- [1] Z. Kong, G.D. Zhang and L.Wang. "An efficient decision making approach in incomplete soft set," Appl. Math. Model, Vol 38, 2014, pp.2141–2150.
- [2] P.K. Maji, A.R. Roy and R. Biswas. "An application of soft sets in a decision making problem," Comput. Math. Appl, Vol 44, No 8-9, 2002, pp.1077–1083.
- [3] A.R. Roy, P.K. Maji. "A fuzzy soft set theoretic approach to decision making problems," J. Comput. Appl. Math, vol 203, No 2, 2007, pp.412–418.
- [4] C. Naim, E. Serdai. "Soft matrix theory and its decision making," Comput. Math. Appl, vol 59, No10, 2010, pp.3308–3314.
- [5] Z. Kong, L.Q. Gao and L.F. Wang. Comment on "A fuzzy soft set theoretic approach to decision making problems," J. Comput. Appl. Math, Vol 223, No 2, 2009, pp.540–542.
- [6] Y.C. Jiang, H. Liu and Y. Tang. "Semantic decision making using ontology-based soft sets," Math. Comput. Model. Vol 53, No 5–6, 2011, pp.1140–1149.
- [7] F. Feng, Y.M. Li and V.L. Fotea. "Application of level soft sets in decision making based on interval-valued fuzzy soft sets," Comput. Math. Appl. Vol 60, No 6, 2010, pp.1756–1767.
- [8] F. Feng, Y.B. Jun and X.Y. Liu. "An adjustable approach to fuzzy soft set based decision making," J. Comput. Appl. Math, Vol 234, No 1, 2010, pp.10–20.
- [9] X.B. Yang, T.Y. Lin and J.Y. Yang. "Combination of interval-valued fuzzy set and soft set," Comput. Math. Appl, Vol 58, No 3, 2009, pp. 521–527.
- [10] Y. Yang, X. Tan and C.C. Meng. "The multi-fuzzy soft set and its application in decision making," Appl. Math. Model, Vol 37, No 7, 2013, pp.4915–4923.
- [11] Y. Zou, Z. Xiao. "Data analysis approaches of soft sets under incomplete information," Knowl.-Based Syst, Vol 21, No 8, 2008, pp.941–945.
- [12] T.Q. Deng, X.F. Wang. "An object-parameter approach to predicting unknown data in incomplete fuzzy soft sets," Appl. Math. Model, Vol 37, No 6, 2013, pp.4139–4146.
- [13] T. Herawan, M.M. Deris. "A soft set approach for association rules mining," Knowl.-Based Syst, Vol 24, No 1, 2011, pp.186–195