Research on OWA Based Multi-Source Heterogeneous Data Fusion

Jianhua Jiang, Buyun Sheng and Mingzhong Yang
School of mechanical and electronic engineering, Wuhan University of Technology, Wuhan 430070, China

Abstract *For the problem of multi-source heterogeneous data fusion, an architecture model of multi-source heterogeneous data fusion was designed. In the process of data fusion, triangular fuzzy number (TFN) was used to give uniform express of multi-data description in quantity, the ordered weight average (OWA) was used to deal with the preference of decision-maker and design the algorithm of data fusion. At last, the feasibility was validated by an example.*

Keywords *data fusion; triangular fuzzy number; ordered weight average*

1. Introduction

Data fusion is a process of multi-source data cooperating as to reduce data redundancy and capture comprehensive information from data. It has become a research focus in data processing, object identification, situation assessment and intelligent decision-making domains. In general view, the data for data fusion come from multi-sensors[1], the data type is almost numerical and the methods are mainly from statistics and artificial intelligent, and some literatures have researched the multi-source data fusion[2,3,4]. In fact, except for numerical value, there are many other expression ways of data such as language and symbols, and the multi-ways of data expression will led to the ambiguity, difference and heterogeneity of data. On the other hand, in order to make an ultimate decision, a decision-maker must integrate a wide range of heterogeneous data and information. Therefore, in allusion to the features of heterogeneous data, this paper set focus on the method of multi-source heterogeneous data fusion and its application.

2. Multi-source Heterogeneous Data Fusion Methods and Its Architecture

2.1 multi-source heterogeneous data fusion methods

The methods for data fusion in decision level mainly include weight average[5], D-S evidence theory[6] and voting[7].

(1) Weight Average Method

The method uses formula $\sum w_i t_{ij}$ to compute decision support value, where w_i denotes weight of data source i and t_{ij} denotes support value of data source i to decision j. It evaluates every decision according the support value, it is easy to operate and gives full consideration to the importance of data sources. But it is difficult to eliminate the impact of subjective factors in determining the weights.

(2) D-S Evidence Theory

It defines the space including all possible results of the object to be identified as a framework set D, its subset marked 2^D and for $\forall A \subseteq D$:

$$m: 2^D \to [0,1]$$

Where $m(\phi)=0$, $\sum_{A\subseteq 2^D} m(A)=1$, ϕ is empty set, and m is called basic probability assignment function (BPAF) which actually refers to assign the trust value to subsets of D on basis of the available evidence. In practice, many BPAFs should be combined for different evidences of a problem may led to different m_i , it can be realized by the following formal:

$$m(A) = K^{-1} \times \sum_{\cap A_i = A} \prod m_i(A_i) \quad (1 \le i \le n)$$

$$\text{Where} \quad K = \sum_{\cap A_i \ne \phi} \prod m_i(A_i) \quad \text{and} \quad A_i \subseteq D$$

D-S evidence theory is based on BPAF and able to compute uncertainty caused by unknown factors. But it requests all the items in D mutually exclusive and is difficult to calculate when there are too many BPAFs.

(3) Voting

It treats every data source as a voter and chooses a decision by comparing the votes obtained, the votes is defined as follow:

$$Sup(a_i) = F(Sup_j(a_i))$$

Where a_i denotes decision i ($1 \le i \le n$), $Sup(a_i)$ denotes the total votes it obtained, $Sup_j(a_i)$ denotes the support value of data source j to decision i and its value is 0 or 1, and F may be defined as sum.

It is difficulty to find BPAF for D-S evidence theory and distinguish two decisions have same votes, this paper uses OWA to fusion multi-source data in full considering the preference of decision-makers.

2.2 Multi-source Heterogeneous Data Fusion Architecture

Literature [2] proposes an architecture for multi-source data fusion as shown in figure 1, it takes into account user requirements and source credibility, and uses proximity knowledge base, knowledge of reasonableness and voting method to resolve data conflicts.

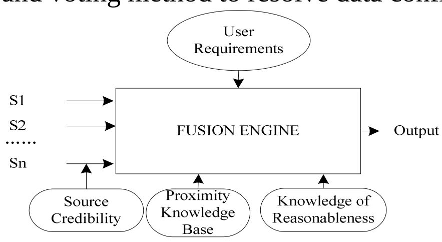

Figure 1 Schematic of data fusion

Under the guidance of the above model, an architecture model of multi-source heterogeneous data fusion in decision level was designed as shown in figure 2. The fusion engine is made up of four parts of data warehouse, decision support value computing, OWA operator weight vector computing and data conversion and sorting.

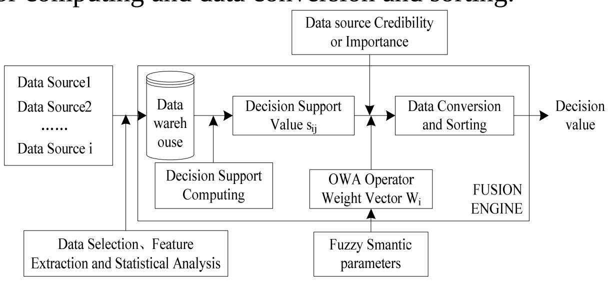

Figure 2 multi-source data fusion architecture

(1) Data warehouse is used to integrate multi-source data and eliminate data heterogeneity and differences by operation of data selection, feature extraction and statistical analysis.

(2) Decision support computing module obtains related data from data warehouse according to decision-making attributes and computes the support value s_{ij} of data source i to decision j.

(3) OWA operator weight vector computing module computes the weight of data source according to the fuzzy semantic parameters provided by decision-maker, and these parameters reflect the preferences of decision-makers to the data source.

(4) Data conversion and sorting calculates $s_{ij}^{'}$ in combining with data source credibility (or importance degree) and OWA weight, and sorts $s_{ij}^{'}$ according its value. At last, the final decision value can be calculated by $s_{ij}^{'}$ and w_i.

3. Algorithm of Data Fusion

3.1. Data Type and Its Features

Data can be expressed in quantity and quality, and numbers for quantity while language variables for quality[8]. According to difference of data description, data type can be divided into quantitative and qualitative two types, this paper focus on 4 types of data description, as shown in table 1.

Table 1 data type

type	Data Description	notes
quantitative	Random variable	Random variable submits to certain distribution
	Two value type	Data value either 1 or 0, or true or false
qualitative	Degree type	The degree general described by 7 or 9 standard
	Terminology based	Description method depends on vocabulary space

In case of large samples, random variable submits to normal distribution marked $X \sim (\mu, \sigma^2)$, where μ denotes expectation and σ denotes standard variance, and $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9974$. Two-value type data is used to give answer for affirming or denying, if agreed the value is 1, otherwise 0, it can also be described by true or false. Degree type data is generally described by degree adverbs such as good, a little good and so on, the grading standard widely used is 7 or 9. Terminology based data is described by terms from a pre-defined vocabulary space and the space depends on specific circumstance.

3.2 Computing of TFN based Support Value

In allusion to the ambiguity in data description, TFN is used to compute the decision support value.

(1) Transform Method for Random Data

Supposed that:

$$x_0 = u - 3\sigma \qquad x^{'} = \frac{x - x_0}{6\sigma}$$

If support value increases with the increased of random variable x and $[\mu - 3\sigma, \mu + 3\sigma]$ is divided into n intervals, support value can be defined as:

$$s(x) = \begin{cases} (0,0,0), x \le \mu - 3\sigma \\ (\frac{i}{n}, x^{'}, \frac{i+1}{n}), \frac{6\sigma i}{n} + x_0 < x \le \frac{6\sigma(i+1)}{n} + x_0 \\ (1,1,1), x > u + 3\sigma \end{cases} \tag{1}$$

If support value decreases with the increased of random variable x, support value defined as:

$$s'_{(x)} = (1,1,1) - s(x)$$

(2) Transform Method for Two Value Type Data

Supposed that two-value type data described by set {1, 0}, the number of answer 1 and 0 is *n* and *m* respectively. If support value is defined by number of value 1, support value can be defined as:

$$s(x) = (n/n+m, n/n+m, n/n+m) \tag{2}$$

(3) Transform Method for Degree Type Data

If data described by the 7 grading standard, the support value can be defined according to table 2.

Table 2 Fuzzy quantification of degree type data

Inverse proportion	Direct proportion	$s(x)$
very high	very low	(0.00,0.00,0.17)
high	low	(0.00,0.17,0.34)
a little high	a little low	(0.17,0.34,0.50)
common	common	(0.34,0.50,0.67)
a little low	a little high	(0.50,0.67,0.84)
low	high	(0.67,0.84,1.00)
very low	very high	(0.84,1.00,1.00)

(4) Transform Method for Terminology based Data

If the vocabulary space w includes n terminologies, and they are ordered by their contribution value for decision from low to high as $w = \{w_0, w_1, ..., w_{n-1}\}$, the support value defined as:

$$s(w_i) = (i/(n-1), i/(n-1), i/(n-1)) \tag{3}$$

3.3 Computing of OWA Operator[9] Weight Vector

Supposed that $F: R^n \to R$ and a n dimension vector $w = (w_1, w_2, ..., w_n)$ related to F makes:

$$F(a_1, a_2, ..., a_n) = \sum_{i=1}^{n} w_i b_i \tag{4}$$

Where $w_i \in [0,1]$, $1 \le i \le n$ and $\sum_{i=1}^{n} w_i = 1$. b_i is the *i*th largest element of a_i. Then F is called n dimension OWA Operator and $w = (w_1, w_2, ..., w_n)$ can be represented by a function as:

$$w_i = f(i/n) - f((i-1)/n) \tag{5}$$

Where $i = 1,2,...,n$, f is named fuzzy semantic operator (FSO) and defined as:

$$f(x) = \begin{cases} 0 & , x < a \\ (x-a)/(b-a) & , a \le x \le b \\ 1 & , b < x \end{cases} \tag{6}$$

Where $x, a, b \in [0,1]$.

In addition, OWA also defines a measure operator for reflecting preference of decision-makers named $orness(w)$ and it is defined as:

$$c = orness(w) = \frac{1}{n}\sum_{i=1}^{n}(n-i)w_i \tag{7}$$

3.4 Algorithm of Data Fusion

Supposing there are n decisions $A=(A_1, A_2,..., A_n)$, m data sources $S=(S_1, S_2,..., S_m)$ and p_i is the credibility or importance of S_i, the algorithm of data fusion is described as:

Step1: Compute support value of data sources to decisions

Access data from data warehouse and transform it to support value of every decision according to the method of section 3.1. The support value marked as:

$$S_{ij} = (a_{ij}, b_{ij}, c_{ij})$$

Where S_{ij} is support value of data source i to decision j, (a_{ij}, b_{ij}, c_{ji}) is TFN expression of S_{ij} and $0 \le a_{ij} \le b_{ij} \le c_{ij} \le 1$.

Step2: Compute OWA operator weight vector

Considering the preference of decision-makers, choose fuzzy semantic parameters for formula 6. In most cases, fuzzy semantic parameters is defined as "the majority", "at least half" and "as much as possible", their parameters are $(0.3,0.8)$, $(0,0.5)$ and $(0.5,1)$ respectively. According to the parameters, the FSO $f(x)$ can be determined. With FSO, formula 5 and 7, OWA weight vector $w=(w_1, w_2,..., w_n)$ and $orness(w)$ can be calculated.

Step3: Transform s_{ij} according to p_i and s_{ij}

In order to use OWA weight vector, s_{ij} must be transformed and sorted, the transform method as follow:

Supposed that:

$$\begin{aligned} s_{ij\_\min} &= p_i \times s_{ij} \\ s_{ij\_\max} &= p_i + s_{ij} - p_i \times s_{ij} \\ s_{ij\_average} &= \frac{n}{\sum_{i=1}^{n} p_i} p_i s_{ij} \end{aligned} \tag{8}$$

Define:

If $c \le 0.5$

$h_{(c)} = 0$, $m_{(c)} = 2c$, $l_{(c)} = 1-2c$

If $c \ge 0.5$

$h_{(c)} = 2c-1$, $m_{(c)} = 2-2c$, $l_{(c)} = 0$

s_{ij} can be transformed into $s_{ij}^{'}$ by following formula:

$$s_{ij}^{'} = h_{(c)} s_{ij\_\max} + m_{(c)} s_{ij\_average} + l_{(c)} s_{ij\_\min} \tag{9}$$

Step 4: Compute final support value for every decision

The final support value of decisions can be computed by the following formula:

$$s_j = \sum_{i=1}^{m} w_i b_{ij} \qquad j = 1,2,..., n \tag{10}$$

Where b_{ij} denotes the *i*th largesse element of $(s_{1j}^{'}, s_{2j}^{'},..., s_{nj}^{'})$.

Step5: Make a decision according to the final support value of decisions.

4. An Example

This paper takes steam turbine product development decision-making as example. Supposing there are 5 kinds of product named A_1, A_2, A_3, A_4 and A_5, the data can be collected include market demand, product feedback, product parameters, historical data, failure statistics, expert advice and so on. Based on these data sources, products can be compared from 6 aspects, they are assessment of market demand a_1, average failures per year a_2 ($\mu = 3.5, \sigma = 0.8$) , longest average no failure time a_3 (month as unit and $\mu = 12.28, \sigma = 2.53$), economy evaluation a_4, customer evaluation a_5 and expert advice a_6, and the preliminary results are shown in table 3.

Table 3 support value of product and data source credibility

a_i	p_i	A1	A2	A3	A4	A5
a_1	0.5	large	common	Very large	A little small	Very small
a_2	0.9	5.25	8.64	3.28	1.81	4.76
a_3	0.9	8.36	13.28	15.14	7.08	17.31
a_4	0.3	common	Very good	A little bad	A little good	good
a_5	0.6	0.65	0.73	0.48	0.54	0.67
a_6	0.4	Improve	Fully recommend	As usual	Out of market	Improve

(1) In table 3, a_1 and a_4 are degree-type data and transformed by table 2, a_2 and a_3 are random data and transformed by formula 1 (supposing n=15), a_5 is two-value type data(the value in table 3 is the ratio of positive evaluation) and transformed by formula 2, a_6 is terminology based data and transformed by formula 3. The transformed results are shown in table 4.

(2) If "the majority" is selected as FSO, the parameters a and b in formula 6 are 0.3 and 0.8 respectively. According to formula 5 and 6, the OWA weight vector can be calculated as:

$w = (0, 0.067, 0.33, 0.33, 0.27, 0)$

Put every w_i into formula 7, the value of c can be calculated and $c = 0.37$.

(3) According to formula 8 and 9, the data in table in table 4 can be transformed as shown in table 5.

(4) For every column data in table 5, sorted them from high to low according the second value and computed final support value by formula 10, the result is shown in table 6.

(5) From table 6, product A_3 gets the highest support value, so it should be developed full

Table 4 disposed attribute values

a_i	p_i	A1	A2	A3	A4	A5
a_1	0.5	(0.67,0.84,1.00)	(0.34,0.5,0.67）	(0.84,1.00,1.00)	(0.17,0.34,0.50)	(0.00,0.17,0.34)
a_2	0.9	(0.13,0.14,0.20）	(0.00,0.00,0.00)	(0.53,0.55,0.60)	(0.80,0.85,0.86)	(0.02,0.23,0.27)
a_3	0.9	(0.20,0.24,0.26)	(0.53,0.56,0.6)	(0.66,0.68,0.73)	(0.13,0.16,0.20)	(0.80,0.83,0.86)
a_4	0.3	(0.34,0.50,0.67）	(0.84,1.00,1.00)	(0.17,0.34,0.50)	(0.50,0.67,0.84)	(0.67,0.84,1.00)
a_5	0.6	(0.65,0.65,0.65)	(0.73,0.73,0.73)	(0.48,0.48,0.48)	(0.54,0.54,0.54)	(0.67,0.67,0.67)
a_6	0.4	(0.33,0.33,0.33)	(1.00,1.00,1.00)	(0.66,0.66,0.66)	(0.00,0.00,0.00)	(0.33,0.33,0.33)

Table 5 results of data transformed

a_i	A_1	A_2	A_3	A_4	A_5
a_1	(0.50,0.62,0.75)	(0.25,0.37,0.50)	(0.63,0.75,0.75)	(0.13,0.25,0.37)	(0.00,0.13,0.25)
a_2	(0.18,0.19,0.27）	(0.00,0.00,0.00)	(0.71,0.74,0.81)	(1.07,1.14,1.16)	(0.27,0.31,0.36)
a_3	(0.27,0.32,0.35)	(0.71,0.75,0.81)	(0.89,0.92,0.98)	(0.18,0.22,0.27)	(1.07.1.12,1.16)
a_4	(0.15,0.22,0.30）	(0.38,0.45,0.45)	(0.07,0.15,0.22)	(0.22,0.30,0.38)	(0.30,0.38,0.45)
a_5	(0.58,0.58,0.58)	(0.65,0.65,0.65)	(0.43,0.43,0.43)	(0.48,0.48,0.48)	(0.60,0.60,0.60)
a_6	(0.20,0.20，0.20)	(0.60,0.60,0.60)	(0.39,0.39,0.39)	(0.00,0.00,0.00)	(0.20,0.20,0.20)

Table 6 final support values

	A_1	A_2	A_3	A_4	A_5
value	(0.23,0.27,0.31)	(0.43,0.49,0.53)	(0.52,0.54,0.56)	(0.20,0.27,0.35)	(0.28,0.32,0.36)

5. Conclusion

In this paper, an architecture model for multi-source heterogeneous data fusion was constructed, TFN based data processing for multi-data description in quantity was researched and OWA operator based data fusion algorithm was designed, the practical application indicates that the algorithm designed is effective. The study of this paper presents a feasible option for constructing intelligent decision support system and has certain reference value for similar data processing and fusion.

Acknowledgements

This paper is supported by National Natural Science Fund Project of China (contract no.50620130441), Scientific and Technological Project of Wuhanc City (contact no. 200810321153) and Youth Science and Technology Chenguang Project of Wuhan City (contact no. 200750731289).

References

[1] D. L. Hall, J. Llinas. Hand Book of Multi-sensor Data Fusion. Beijing: Electronic Industry Press, 2008.
[2] R. R.Yager. A framework for multi-source data fusion. Journal of Information Sciences Vol.163 (2004) 175~200.
[3] Zhou Haiyin, Li Donghui, Jiang Yueping. High Precision Method for Determining the Position of Aerocraft Based on Mult-Sensors Information Fusion. Journal of Hunan University(natural Sciences), 34(4) (2007) 37~40.
[4] R.R.Yager. Modeling Intelligence information: Multi-source fusion. IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety,Orlando, FL, MAR 31-APR 01, 2005.
[5] Wang Guangyun, Li Weihua, Hua Wenjian, etal. A method for heterogeneous uncertain information fusion and its application. International Conference on Signal Processing Proceedings, vol.3 (2004) 2253~2256.
[6] Ma Linru, Yang Lin, Wang Jianxin. Research on Security Information Fusion from Multiple Heterogeneous Sensors. Journal of System Simulation, 20(4) (2008) 981~985.
[7] S. Kang. Two-Phase Identification Algorithm Based on Fuzzy Set and Voting for Intelligent Multi-sensor Data Fusion, Lecture Notes in Computer Science, vol. 4252 (2006) 769~776.
[8] F.Herrera, L.Martinez. An approach for combing linguistic and numerical information based on the 2-tuple fuzzy linguistic representation model in decision-making. International

Journal of Uncertainty, Fuzziness and knowledge-based systems, 8(5) (2000)539~562.
[9] Z.Zhang, C.Zhang. Result fusion in multi-expert systems based on OWA operator. Proceedings of the 23rd Computer Science Conference, (2000) 234~240.