Research Journal of Applied Sciences, Engineering and Technology 6(11): 1993-1997, 2013
DOI:10.19026/rjaset.6.3814
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: November 24, 2012 Accepted: January 17, 2013 Published: July 25, 2013

Research Article

Data-Fusion Approach Based on Evidence Theory Combining with Fuzzy Rough Sets for Urban Traffic Flow

Ning Chen
School of Mechanical and Automotive Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China

Abstract: The traffic detecting result is always short of accuracy by different kinds of individual sensors in urban China. A new data fusion approach is raised in this paper to solve the issue, based on fuzzy rough set theory combining with evidence theory. The method is improved to concise attribute rules and to measure fuzzy likelihood. Furthermore, a new combination rule is given to dissolve the confliction among the traffic evidence data collected by different individual sensors. Finally, the experiment to fuse the traffic data from an intersection in Hangzhou City showed that the proposed approach could obtain a high accuracy.

Keywords: Combination rule, data confliction, data fusion, intelligent transportation system, urban traffic flow

INTRODUCTION

At present, there are several modes to collect traffic flow data in urban China, such as loops, video detector, and dynamic OD analyzer. Because of their individual limitations, they might give the contradictory outcome each other and it is difficult to determine the credibility of the collected traffic data. Therefore, the multi-source data fusion method gets a wide application. For example, there have been classical fusion algorithms such as Kalman filter (Xie *et al*., 2010), Bayes reasoning (Meng *et al*., 2012), Fuzzy set theory (Brooks and Kaupp, 2007). Unfortunately, these kinds of algorithm lack capacities to figure out the conflict among the traffic flow message to be fused.

It is a relief that D-S evidence theory can fuse the uncertain message with unknown conditions by means of both trusted function and likelihood function coming from the essential probability function, and the fused outcome becomes more accurate (Cao and Dai, 2008; Bogler, 1987). However, to fuse the traffic data using D-S evidence theory, there are also several problems to be solved such as the traffic data redundancy need be reduced, the essential probability formula of each evidence group should be extracted to avoid subjective effect like being assigned merely by the experts' experience. Meanwhile, the conflicts of evidence should be figured out also. There probably exists the deficiency or difference among the data due to the failure of one or more sensors in a multi-sensor fusing system (Carmine *et al*., 2006).

To solve these issues, the new measures are proposed in our research. For example, the raw data could be preprocessed by means of the attribute reduction of rough set, which is improved from the classical reduction principle considering the dependency between the existing attributes of the reduction set and the new ones. The fuzzy likelihood measure is used to obtain the essential probability formula of the traffic flow message. More over, on the basis of Yager improved combination rule of evidence theory (Yavuz, 2007), a new combination rule is raised to eliminate conflicts in the fusing process with a consideration of the conflict degree between evidence data.

EVIDENCE THEORY COMBINING WITH FUZZY ROUGH SET

The fusion parameters of traffic flow are defined as follows: vehicular flux, lane occupancy ratio (A_{flu}), average speed (B_{ocp}), queue length (C_{spe}), waiting time (D_{seq}), average traveling time (E_{tim}). These parameters constitute vector x = ($A_{flu,}$ B_{ocp}, C_{spe}, D_{seq}, E_{tim}, F_{dur}, K), where K denotes the different collecting method.

Data preprocessing based on rough set: The classical attribute reduction uses importance degree to describe the influence to decision attribute D after new attribute 'a' of condition attribute set C joined into reduction attribute set R. But it lacks consideration about the influence to set R. Using dependent degree to judge whether the addition of the new attribute makes the certain ones of set R become unimportant is proposed in this paper. The algorithm as follows:

Step 1: Select condition attribute set C = (A_{flu}, B_{ocp}, C_{spe}, D_{seq}, E_{tim}, F_{dur}), decision attribute D = K. The current collecting data, historical data and sensor characteristic constitute the decision attribute table:

$$P=\begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}=\begin{pmatrix} 1 & \tilde{A}_{flu}{}^1 & \tilde{B}_{ocp}{}^1 & \tilde{C}_{spe}{}^1 & \tilde{D}_{seq}{}^1 & \tilde{E}_{tim}{}^1 & \tilde{F}_{dur}{}^1 & \tilde{K}^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ i & \tilde{A}_{flu}{}^i & \tilde{B}_{ocp}{}^i & \tilde{C}_{spe}{}^i & \tilde{D}_{seq}{}^i & \tilde{E}_{tim}{}^i & \tilde{F}_{dur}{}^i & \tilde{K}^i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ n & \tilde{A}_{flu}{}^n & \tilde{B}_{ocp}{}^n & \tilde{C}_{spe}{}^n & \tilde{D}_{seq}{}^n & \tilde{E}_{tim}{}^n & \tilde{F}_{dur}{}^n & \tilde{K}^n \end{pmatrix}=F(X)$$

where,

$$X=\begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}=\begin{pmatrix} A_{flu}{}^1 & B_{ocp}{}^1 & C_{spe}{}^1 & D_{seq}{}^1 & E_{tim}{}^1 & F_{dur}{}^1 & K^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{flu}{}^i & B_{ocp}{}^i & C_{spe}{}^i & D_{seq}{}^i & E_{tim}{}^i & F_{dur}{}^i & K^i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{flu}{}^n & B_{ocp}{}^n & C_{spe}{}^n & D_{seq}{}^n & E_{tim}{}^n & F_{dur}{}^n & K^n \end{pmatrix}$$

is the original collected data set. In $F=(f_{A_{flu}}, f_{B_{ocp}}, f_{C_{spe}}, f_{D_{seq}}, f_{E_{tim}}, f_{F_{dur}}, E)$ each component expresses the mapping from x_i to y_i.

Step 2: In set C, if $\forall$ b $\in$ C, select it as the original elements of set R.

Step 3: If $\forall$ b $\in$ C ^ $\notin$ R, compute its importance degree

$$SGF(a,R,D)=\frac{Card(pos_{R\cup\{a\}}(D))}{Card(U)}-\frac{Card(pos_R(D))}{Card(U)}$$

If SGF ($\acute{\alpha}$, R,D) = max ({SGF(α, R, D)|α ∈ C}), and $\exists a^{'}\in C$: denote $\bar{R}=R\cup\{a^{'}\}$.

Step 4: Compute $k_{a^{'}b}=r(a^{'},b)=\frac{card(POS_{IND(a^{'})}IND(b))}{card(U)}$ for the new attribute α and $\forall$ b ∈ C. If ∈ b' ∈ R and $k_{\alpha' b}$ = max ({$k_{\alpha' b}$= r(α' , b) b ∈ R }), delete the attribute b' temporarily from reduction set $\bar{R}$. Then $\bar{R}$ is denoted as R' (if the dependent degree are all equal, b' is selected as the element which has the longest survival time in the set).

Compute SGF (b', R', D). If | SGF (b', R', D). SGF (α, R, D) < δ. δ is given in advance as threshold value), do not delete attribute b from set $\bar{R}$, and denoted as R = $\bar{R}$.

Step 5: Compute $\gamma_R(D)$. If $\gamma_R(D)$ = $\gamma_R(D)$ = $\gamma_C(D)$, R satisfies the condition, then the calculation is over. Otherwise calculation turns to step 2.

Calculation on essential probability function based on fuzzy likelihood measurement: To avoid the subjectivity while obtaining the essential probability function, this method is proposed base on fuzzy likelihood measure (case study of three collecting methods). The algorithm as follows:

Step 1: Select identification frame Θ = {X, Y, Z}, where X, Y, Z, respectively denote the traffic flow parameters of the three collecting methods. The collected data set is denoted as:

$$Q=\begin{pmatrix} t_{11} & & \cdots & & t_{16} \\ t_{21} & \cdots & t_{ij} & \cdots & t_{26} \\ t_{31} & & \cdots & & t_{36} \end{pmatrix}$$

after attribute reduction. (i ∈[1, 3], i ∈ Z; j ∈[1, 6], j ∈ Z). Where t_{ij} denotes the fuzzy membership function of the j-th traffic parameters which are collected by the j-th collecting methods respectively. Historical fusion data in the same condition are selected to be the basic traffic flow data, which are expressed as the fuzzy membership function of relevant attribute, and denoted as S =($\bar{t}_1$, $\bar{t}_2$, … , $\bar{t}_j$) ,(j ∈[1, 6], j ∈ Z) where $\bar{t}_i$ is the fuzzy membership function of relevant attribute.

Step 2: Matrix multiplication is defined as the fuzzy likelihood calculation between two fuzzy membership functions:

So

$$\bar{M}=Q\bullet S=\begin{pmatrix} \rho(t_{11},\bar{t}_1) & & \cdots & & \rho(t_{16},\bar{t}_6) \\ \rho(t_{21},\bar{t}_1) & \cdots & \rho(t_{ij},\bar{t}_j) & \cdots & \rho(t_{26},\bar{t}_6) \\ \rho(t_{31},\bar{t}_1) & & \cdots & & \rho(t_{36},\bar{t}_6) \end{pmatrix}_{3\times 6}$$

(j∈[1, 3], j ∈ Z; (j ∈[1, 6], j ∈ Z)) $\rho(t_{ij},\bar{t}_j)=p(t_{ij}\cap\bar{t}_j\neq\varnothing)=p(A\leq(M\wedge N)(x))=$ supx min {M(x); N(x)}. where M(x) and N(x) are the relevant membership functions of t_{ij} and $\bar{t}_i$.

Step 3: Compute the essential probability function of $\bar{M}$ divided by column to have normalization processing. The outcome is M= {(m_{i1}, m_{i2}, m_{i3}, m_{i4})}i ∈[1, 6], i ∈ Z} and each group is as follows:

$$m_{i1}=\frac{\rho(t_{i1},\bar{t}_i)}{(\rho(t_{i1},\bar{t}_i)+\rho(t_{i2},\bar{t}_i)+\rho(t_{i3},\bar{t}_i)+\rho_i(\Theta))}$$

$$m_{i2}=\frac{\rho(t_{i2},\bar{t}_i)}{(\rho(t_{i1},\bar{t}_i)+\rho(t_{i2},\bar{t}_i)+\rho(t_{i3},\bar{t}_i)+\rho_i(\Theta))}$$

$$m_{i3}=\frac{\rho(t_{i3},\bar{t}_i)}{(\rho(t_{i1},\bar{t}_i)+\rho(t_{i2},\bar{t}_i)+\rho(t_{i3},\bar{t}_i)+\rho_i(\Theta))}$$

$$m_{i4}=\frac{\rho_i(\Theta)}{(\rho(t_{i1},\bar{t}_i)+\rho(t_{i2},\bar{t}_i)+\rho(t_{i3},\bar{t}_i)+\rho_i(\Theta))}$$

Table 1: Comparison of the combination outcome

	K	ε	$\prod_{i=1, j\leq i}^{n}(1-\varepsilon_{ij})$	M(A)	M(B)	M(C)	M(X)
D-S combination formula	0.99901	0.368	-	0	0	1	0
Yager combination formula	0.99901	0.368	-	0	0	0.00099	0.99901
The combination formula in literature [9]	0.99901	0.368	-	0.321	0.003	0.188	0.488
The new combination formula in this paper	0.99901	-	0.2525	0.3929	0.0066	0.3663	0.2523

($\rho_i(\Theta) = 1 - \max(\rho(t_{i1}, \bar{t_i}), \rho(t_{i2}, \bar{t_i}), \rho(t_{i3}, \bar{t_i}))$,
(j ∈[1, 6], j ∈ Z)

m_{i1}, m_{i2}, m_{i3}, and m_{i4} denote respectively the essential probability functions of the message collected by three collecting methods and the uncertain message.

Conflict solution of the evidence combination: The disaccord to the real traffic scene may be occurred by the fused outcome if there is high conflict evidence, namely the conflict coefficient *k* → *1* .Yager has improved the D-S composite formula. And the new formula is as follows (two evidence sources):

- $m(\emptyset) = 0$;
- $m(A) = \sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j), A \neq \emptyset, X$
- $m(X) = \sum_{A_i \cap B_j = X} m_1(A_i) m_2(B_j) + k$ ($k = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)$)

Yager's formula shows that if the conflict evidence can't be resolved reasonably, it should be thrown into unknown field, but it will induce another issue. Although most of evidences have proved the conclusion is right, the combination outcome would be negative. Based on Yager's formula, Sun Quan proposed an evidence combination formula which transforms the conflict by the credibility in the literature (Sun *et al*., 2011). But this method ignores the evidence contribution to the combination outcome when computing the credibility of each group of conflict evidence (Selzer and Gutfinger, 1988).

In this paper we have improved the Yager evidence combination formula based on the literature (Selzer and Gutfinger, 1988), with the consideration about the credibility of group conflict evidence. The evidence credibility is used as proportional coefficient of the probability of the conflict evidence in the combination formula. The new evidence combination formula is as follows:

- $m(\emptyset) = 0$
- $m(A) = p(A) + kq(A)$, $A \neq \emptyset$, X
- $m(X) = p(X) + k q(X) + k \prod_{i=1, j\leq i}^{n}(1-\varepsilon_{ij})$

$$p(A) = \sum_{\substack{A_i \in F_i \\ \bigcap_{i=1}^{n} A_i = A}} m_1(A_1) m_2(A_2) \cdots m_n(A_n)$$

$$q(A) = \sum_{i=1}^{n} \beta_i m_i(A)$$

The credibility between two evidences, m_i and m_j, is denoted as:

$$\varepsilon_{ij} = e^{-k_{ij}}$$

That is decreasing function. The conflict magnitude between the two evidences is denoted as:

$$k_{ij} = \sum_{A_i \cap A_j = \emptyset} m_i(A_i) m_j(A_j)$$

The average credibility between evidence m_i and other evidences is denoted as:

$$\alpha_i = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \varepsilon_{ij}$$

The weight value of normalization is denoted as:

$$\beta_i = \frac{\alpha_i}{\alpha_1 + \alpha_2 + \cdots + \alpha_n}$$

It could be proved that m(A) could be essential probability function, as there exists the below conditions.

- $m(\emptyset) = 0$
- $0 \leq m(A) \leq 1$
- $\sum_{A \subset X} m(A) = 1$

It is shown that the normalization credibility β_i of the evidence in each group is used as the weight. It embodies fully the contribution degree of the evidence in each group to the combination outcome.

Table 1 shows the effect of the new combination formula of evidence theory:

$$m_1 : m_1(A) = 0.98, m_1(B) = 0.01, m_1(C) = 0.01$$

$$m_2 : m_2(A) = 0, m_2(B) = 0.01, m_2(C) = 0.99$$

$$m_3 : m_3(A) = 0.9, m_3(B) = 0, m_3(C) = 0.1$$

Table 2: Collected data of Qingcun-Yan'an intersection

Parameter /Method	Flux (vehicle/5min)	Traveling time (sec)	Queue length (m)	Waiting time (sec)	Average speed (km/h)	Lane occupancy ratio
Loops	16(1)	-	-	-	-	0.8(1)
Video Detector	25(2)	-	56(2)	90(2)	60(2)	-
OD Analyzer	18(3)	196(3)	-	-	53(3)	-

Table 3: Historical fusion data of Qingcun-Yan'an intersection at (t-T), (t-2T) and (t-3T)

Parameter /Method	Flux (vehicle/5min)	Traveling time (sec)	Queue length (m)	Waiting time (sec)	Average speed (km/h)	Lane occupancy ratio
t-T	20(1)	150(3)	60(2)	80(2)	56(3)	0.6(1)
t-2T	18(1)	160(3)	70(2)	150(2)	43(2)	0.7(1)
t-3T	16(3)	240(3)	40(2)	130(2)	30(3)	0.5(1)

Table 4: The attribute reduction of decision table

Process	Candidate set	SGF(α, R.D)	a	k	b	α	R	$\gamma_R(D)$	$\gamma_C(D)$
1	$\{A_{flu}, B_{ocp}, C_{SPe,} D_{seq}\ E_{tim}\}$	{4/6,3/6, 3/6,3/6,4/6}	A_{flu}	{2/6}	--BEN	-	$\{A_{flu,} F_{dur}\}$	4/6	1
2	$\{B_{OCP}, C_{spe}, D_{seq}, E_{tim}\}$	{1/6,1/6, 1/6,1/6,2/6}	E_{tim}	{1/6, 1/6}	F_{dur}	- 2/6 > δ	$\{A_{flu}, E_{tim}\}$	5/6	1
3	$\{B_{ocp}, C_{spe}, D_{seq}, F_{dur}\}$	{0,1/6,1/6, 1/6}	C_{spe}	{0,0}	- -	- -	$\{A_{flu}, C_{spe}, E_{tim}\}$	1	1

Table 5: The traffic data after treatment with attribute reduction

Parameter /Method	Flux (vehicle/5min)	Queue length (m)	Average speed (km/h)
Loops	11	-	-
Video detector	36	30	36
OD Analyzer	15	-	48

The new combination formula could reduce uncertain probability caused by the conflict of different evidences. With fusing the essential probability functions obtained by fuzzy likelihood measure according the new combination formula, it can acquire the better fusion outcome based on data filtering by maximum value.

APPLICATION

In the urban area of Hangzhou, the principle traffic data come from the sensors of loops, video detector, and dynamic OD analyzer. Here set Qingcun-Yan'an intersection, one intersection in the transportation grid, as an example to testify the above algorithms. The time slice (t) is 12:00:00 to 12:05:00 on Jan 1st, 2007 and the traffic data are derive from one lane. The historical fused data of three periods before the time (the period length is T) as the Table 2 and 3 shows. The number in the bracket is the data collecting mode. ('-'denotes no data. In the bracket '1' denotes loops, '2' denotes video detector, '3' denotes OD analyzer)

Attribute reduction: The decision table is built based on the number of the collecting modes in the bracket. Condition attribute ($A_{flu,}$ B_{ocp}, C_{spe}, D_{seq}, E_{tim}, F_{dur}), decision attribute D = K. The reduction process is according to the algorithm of attribute reduction in this paper. Select R = $\{F_{dur}\}$ and δ =1/6. Table 4 shows the calculation steps of the algorithm. The final condition attribute is $\{A_{flu}, E_{tim}\}$.

Table 6: The impoetant parameters in the process

Key parameter	Value
Fuzzy membership matrix	$Q=\begin{pmatrix} e^{-(x-16)^2/18} & 0 & 0 \\ e^{-(x-25)^2/18} & e^{-(x-56)^2/18} & e^{-(x-60)^2/18} \\ e^{-(x-18)^2/18} & 0 & e^{-(x-53)^2/18} \end{pmatrix}$
Benchmark conversion matrix	$S=(e^{-(x-20)^2/18}, e^{-(x-60)^2/18}, e^{-(x-56)^2/18})$
Fuzzy likelihood matrix	$M = Q\square S = \begin{pmatrix} 0.8 & 0 & 0 \\ 0.9 & 0.8 & 0.8 \\ 0.7 & 0 & 0.9 \end{pmatrix}$
Vector of essential probability function	$\overline{M}$ = {(0.320,0.360,0.280,0.040),(0,0.800,0,0 .200),(0,0.444,0.500,0.056)}

Table 7: The important parameters in the process

Key parameter	Value
Conflict coefficient $\overline{K}$	$\overline{K}$=(0.7392, 0.7900, 0.9282)
Credibility ε	ε =(0.4775, 0.4538, 0.3953)
Conflict coefficient k	0.3436
Average credibility α	α =(0.4657, 0.4364, 0.4245)
Value of normalization β	β =(0.3510, 0.3290, 0.3200)
Combination outcomes of evidence	M=(0.4209, 0.2833, 0.1831, 0.1127)

Fusion of the traffic data:

- The identification frame is. Θ = {X, Y, Z} (X, Y and Z denote the collected data of loops, video detector and OD analyzer). Compute the essential probability function. Table 5 shows the important parameters in this process.
- Based on the essential probability function and the new combination formula of evidence proposed in this paper, to get more credible fusion outcome. Table 6 shows the important parameters in the process.

In Table 7 the combination outcome of evidence shows that the traffic flow data which are collected by the loops have the maximal credibility. Table 8 shows the fusion outcomes, the practical data of the traffic flow which are collected by manual work in the same

Table 8: The fusion outcome of the traffic flow message

Detection parameter	Vehicular flux (vehicle/5min)	Traveling time (sec)	Queue length (m)	Waiting time (sec)	Average speed (km/h)	Lane occupancy ratio	Average value of relative error
Fusion outcome	25	196	56	90	60	0.8	-
Practical collecting value	22.0	160	45.0	78	45	0.60	-
Relative error	0.136	0.225	0.244	0.154	0.333	0.333	0.232

condition, the relative error and the average value of relative error of each traffic flow parameter.

Table 8 shows that the approach can obtain the fusion outcome effectively. The errors occurred due to the below factors. The threshold value δ that affects the final outcome is obtained by human experience in Table 4. Another one is that the variance of normal distribution is determined with the principle of 3 σ.

CONCLUSION

The method of attribute reduction has been improved based on rough set and it could consider the relation of attributes during reduction process. The essential probability function is obtained by the maximum fuzzy likelihood function that helps to diminish the effect of subjective factor. Finally, a new combination formula has been raised based on the Yager's formulas. It can reduce the negative effect on fusion accuracy caused by the conflict of different evidences. The experiment demonstrates that the proposed method is effective and practical to cope with issues such as urban traffic data fusion in urban Hangzhou.

ACKNOWLEDGMENT

This study is supported by National Natural Science Foundation of China(Grant No.61273240), and Key Project of Chinese Ministry of Education (Grant No.211070), and Zhejiang Provincial Natural Science Foundation of China (Grant No. Y12F030091).

REFERENCES

Bogler, P.L., 1987. Shafer-Dempster reasoning with applications to multisensor target identification system. IEEE T. Syst. Man Cyb., 17: 968-977.

Brooks, R. and M. Kaupp, 2007. ORCA: A Componment Model and Repository. Software Engineering for Experiment Robotics, pp: 231-251.

Cao, L. and R. Dai, 2008. Fundamentals, Concepts, Analysis, Design and Implementation. Beijing: Post and Telecom Press, China.

Carmine, G., L. Lucio and L. Vincenzo, 2006. Agent-based architecture for designing hybrid control system. Inform. Sci., 176: 1103-1130.

Meng, X., G. Bai, B. San and X. Han, 2012. Application of bayes date fusion on intelligent fault diagnosis in engine room. J. Dalian Maritime Univ., 38(3): 179-191.

Selzer, F. and D. Gutfinger, 1988. LADAR and FLIR based sensor fusion for automatic target classification. SPIE, 1003: 236-241.

Sun, Q., X. Ye and W. Gu, 2011. A new combination rules of evidence theory. Acta Electron. Sinica, 39(8): 611-628.

Xie, W., J. Ma and M. Yang, 2010. Agent-oriented architecture for intelligent service robot. Proceedings of the 8th World Congress on Intelligent Control and Automation, pp: 5964-5967.

Yavuz, H., 2007. An integrated approach to the conceptual design and development of an intelligent autonomous mobile robot. Robot. Auton. Syst., 55: 498-512.