A Novel Classification Method and its application in Flow Forecast

Li Qianmu,† and Xu Manwu2††,

Nanjing University, State Key Laboratory for Novel Software Technology, Jiangsu, China
Nanjing University of Science and Technology, Computer Science Department, Jiangsu, China

Summary
In this paper, a new adaptive classification fusion method was proposed based on the Dempster-Shafer's theory of evidence and fuzzy Kohonen neural network. The new method integrated ideas from unsupervised neural network model and using neighborhood information in the framework of the Dempster-Shafer theory of evidence. This approach mainly consists in considering each neighbor of a pattern to be classified as an item of evidence supporting certain hypotheses concerning the class membership of that pattern. This evidence is represented by basic probability assignment (BPA's) and pooled using the Dempster's rule of combination. Experiments demonstrate the excellent performance of this method as compared with existing neural network techniques.

Key words:
Classification method, flow forecast, Dempster-Shafer, intelligence decision

Introduction

Since the early 1980's, the Dempster-Shafer (D-S) theory of evidence has generated considerable interest in the Artificial Intelligence community. In contrast, applications in statistical pattern recognition have been very limited until recently. In literature [1] and [2], D-S theory was shown to provide a suitable framework for combining the results of several independent classifiers, and thereby improve classification accuracy. D-S evidence theory has also been introduced to multi-information fusion of Remote Sensing [3], and proves that the method has fine application prospect in the automatic classification of Remote Sensing image based on the fusing analysis of ERS SAR data and TM image. Because not only the fine prospect has been shown in the application of evidence theory to process uncertainty information [4-9], but also the neural network has a series of unique characteristic, such as a large scale parallel processing, distributing information storage, excellent self adaptability and self organizing as well as strong function of learning, association and tolerance fault. Especially from the point of view of practical application, expert knowledge is better described by introducing fuzzy subset [5-6]. Within the framework of pattern recognition, many methods of classification were developed. More recently, techniques using the Dempster-Shafer's theory of evidence tried to deal with the problem related to classification of the uncertainty in Remote Sensing image. Application of the method has also been very limited. Therefore, this paper propose a new adaptive approach to classification fusion of multi-sources Remote Sensing images based on the unsupervised classification of fuzzy Kohonen neural network and using neighborhood information in the framework of the Dempster-Shafer theory of evidence.
The paper is organized as follows. In Section 2, the basics of the D-S theory of evidence and fuzzy Kohonen neural network are briefly recalled. In Section 3, the proposed classification fusion method is described. Some experimental results of multi-sources network information fusion are discussed in Section 4. Finally, the excellent performance of this method is concluded compared with existing neural network techniques.

2. Backgroud

2.1 Dempster-Shafer theory of evidence

In this section, only the main concepts of D-S theory of evidence will be recalled. For a more complete description can be found in Shafer's original work [4] and for more up-to-date sources can be also found in reports of recent developments [5,7].
In Dempster-Shafer theory of evidence, a problem is represented by a finite set Θ of mutually exclusive and exhaustive hypotheses. Each proposition of problem corresponds to a subset of Θ. Due to the selection of Θ depending on the level of knowledge and understanding of problem, the finite set Θ is called the frame of discernment. A mass function: $2^{\Theta} \rightarrow [0,1]$ verifying:

$$\begin{aligned} & m(\phi) = 0 \\ & \sum_{A \subseteq \Theta} m(A) = 1 \end{aligned} \qquad (1)$$

Called Basic Probability Assignment (BPA) or Basic Belief Assignment (BBA) in the frame of discernment Θ. For any $A \subset \Theta$, the quantity $m(A)$ can be interpreted as a measure of the belief that is willing to commit exactly to A,

Manuscript received September 5, 2006.
Manuscript revised September 25, 2006.

given a certain piece of evidence. For any $A = \Theta$, the quantity $m(A)$ represents the belief that do not know how to assignment. The subsets A of Θ with $m(A) > 0$ are called the focal elements of mass function. All sets of focal elements of mass function are called the core of mass function, that is:

$$F = \{A \subseteq \Theta \mid m(A) > 0\} \qquad (2)$$

Associated with mass function are a belief or credibility function Bel and a plausibility function Pl, is defined, respectively, for all $A \subseteq \Theta$ as

$$Bel(A) = \sum_{B \subseteq A} m(B) \qquad (3)$$

$$Pl(A) = \sum_{A \cap B = \phi} m(B) \qquad (4)$$

The quantity Bel (A) can be interpreted as a total measure of one's belief that hypothesis A is true, while Pl (A) may be viewed as the amount of belief that could potentially be placed in A. In general, Bel (A) measures the minimal or necessary support whereas Pl (A) reflects the maximal or potential support for that hypothesis. These two measures span an uncertainty interval $[Bel(A), Pl(A)]$ for this hypothesis.

Given two BPA m_1, m_2 on Θ, induced by two independent sources of evidence A and B respectively, can be combined by the so-called Dempster's rule of combination to yield a new BPA $m = m_1 \oplus m_2$, called the orthogonal sum of m_1 and m_2, and defined as:

$$m(A) = \begin{cases} \dfrac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \phi} m_1(A_i) m_2(B_j)} & A \neq \phi \\ 0 & A = \phi \end{cases} \qquad (5)$$

Where $A_1, A_2, \cdots, A_k$ and $B_1, B_2, \cdots, B_l$ belong to the focal elements of set A and B respectively. According to the rule of combination, many entries independent sources of evidence are merged.

For given $A \subset \Theta$, the quantity $\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)$ is total belief that is willing to commit exactly to hypothesis A. If the total belief of empty set ϕ is not zero, then the total belief of non-empty set is normalized.

2.2 Fuzzy Kohonen Neural Network (FKNN)

FKNN is associated with fuzzy set theory and Kohonen Self-Organization Mapping Network [14, 15], which is one of Self-Organizing fuzzy neural network. Its main idea is introducing fuzzy membership function μ_{ij} to replace learning rate $\alpha(t)$ of basic Kohonen algorithm. The earlier studies have shown a close relationship between numerical results generated by basic Kohonen network and Fuzzy C-Mean (FCM). So the update rule of learning rate is defined as following:

$$a_{ijk}(t) = (\mu_{ijk}(t)^{m(t)}) \qquad (6)$$

$$u_{ijk}(t) = \left(\sum_{s=1}^{M} \left(\frac{\left\| x_{ij} - w_{jk}(t-1) \right\|}{\left\| x_{ij} - w_{js}(t-1) \right\|} \right)^{2/(m(t)-1)} \right)^{-1} \qquad (7)$$

$$m(t) = m_0 - t((m_0 - 1)/t_{\max}) \qquad (8)$$

Where $m(t)$ is the weight index, which is calculated via (8) for t=1, $\cdots$, $t_{\max}$. Weight index $m(t)$ reduces from m_0 (m_0 is any positive constant greater than one) to 1. In FKNN algorithm, with the increasing of the learning count, the degree of membership function is gradually changed, i.e., the winning node's membership degree is increased and others are decreased gradually until the algorithm ends. And the connection weight vector is modified according to the following (9) formula:

$$w_{jk}(t+1) = w_{jk}(t) + \left(\sum_{i=1}^{N} a_{ijk}(t)(x_{ij} - w_{jk}(t)) \right) \Big/ \sum_{i=1}^{N} a_{ijk}(t) \qquad (9)$$

Where t and t+1 are the last and present iterative time respectively. Each training vector x_i is assumed to possess a degree of membership μ_{ijk} to each class k, with the constraint $\sum_{k=1}^{M} \mu_{ijk} = 1$ and $0 \leq \sum_{i=1}^{N} \mu_{ijk} \leq N$. The weight vector of fuzzy Kohonen neural network and the class membership of each training samples are determined by learning of training set in each independent information sources.

3. New Adaptive Classification Method

The classification is an important application in network troubleshooting. But the basis of traditional classification is probability and statistics. So its classification accuracy is low for the sample of non-normal distribution. Moreover, each pixel is classified to one determinate class. In fact, there exist some mix-pixel, which is not classified to a certain class, in traditional Remote Sensing classification. That is to say, there exists some fuzzy in processing of classification. At the same time, the complexity of classification is increasing with data aggrandizing. For above-mentioned problems, a new adaptive network troubleshooting method based on the Dempster-Shafer theory of evidence is introduced.

3.1 Classification fusion based on neighborhood information

Let us consider the problem where some pattern $x \in R^p$ has to be classified in one of M categories or classes $\{\omega_1, \omega_2, \cdots, \omega_M\}$ using a training set T of N P-dimensional patterns with known degree of membership function of each classes. The set of these patterns to be classified X is called a testing set. As a result of the capability of Dempster-Shafer theory of evidence in processing the problem of uncertainty, the paper proposed new classification that deals with all kinds of uncertainty in the processing of Remote Sensing classification using the Dempster-Shafer theory of evidence. The proposed classification method is implemented from four stages. First, according to prior knowledge of classification problem, the number of M classes was determined. And the set of classes is denoted by $\Theta = \{\omega_1, \omega_2, \cdots, \omega_M\}$, which is called the frame of discernment. Secondly, the available information is assumed to consist in a training set T consisting of many pairs $(x_{ij}, (\mu_{ij1}, \cdots, \mu_{ijM}))$ that is obtained using fuzzy Kohonen neural network, $\forall i = 1, \cdots, N;\ j = 1, \cdots, P; k = 1, \cdots, M$. The form of the training set T is defined by:

$$T = \{((x_{11}, (\mu_{111}, \cdots, \mu_{11M})), \cdots, (x_{1P}, (\mu_{1P1}, \cdots, \mu_{1PM}))), \cdots, \\ ((x_{N1}, (\mu_{N11}, \cdots, \mu_{N1M})), \cdots, (x_{NP}, (\mu_{NP1}, \cdots, \mu_{NPM})))\}$$

Each pair consisting of a training sample and the degree of membership function committed to each class of the frame of discernment Θ. These pairs constitute a distinct item of evidence regarding the class membership of some pattern $x \in R^p$ to be classified. Thirdly, compared pattern to be classified with each sample in training set under some measure. According to the relevant metric d, if x is close to some sample in training set, then one will be inclined to believe that the sample in training set is an item of evidence that may affect x belonging to the degree of some class. Consequently, based on these distances and on the degree of membership function of training samples to each class, this item of evidence may be postulated to induce a basic probability assignment (BPA) over Θ. Finally, testing sample is classified and fused. Due to the problem of classification taking into account multi-sources information fusion, a testing sample has to be classified in all classes of each information source. In the same information source, the item of evidence affecting the categories of x can be combined using the Dempster's rule of combination to form a resulting BPA. After that, the BPA of multi-sources information can be integrated using the Dempster's rule of combination to form a final BPA synthesizing one's final belief regarding the class of x. And the final classification will be obtained. The whole processing of classification fusion is shown in Fig. 1.

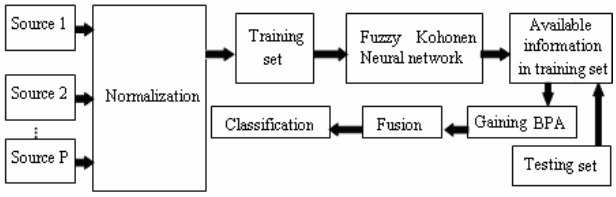

Fig.1. The processing chart of classification fusion in network troubleshooting

3.2 Obtaining Basic Probability Assignment (BPA)

According to fuzzy Kohonen neural network that is introduced in the above second subsection, unsupervised learning is implemented by each element in training set of multi-sources information that is normalized. Consequently, the degree of membership function that training sample belong to certain class is obtained in multi-information sources. When the learning of fuzzy Kohonen neural network finished, the available information is assumed to consist in a training set T of N P-dimensional patterns and their corresponding each class membership functions taking values in $[0,1]$ are obtained.

Supposed x is a new vector to be classified on the basis of the information contained in T. If x is close to some sample in training set T according to the relevant metric, then one will be inclined to believe that both x and sample vectors belong to the same class. And the sample in training set is an item of evidence may affect the degree of x belonging to some class. If the training sample x_{ij} belongs to ω_k, then the whole set will be classified into two subsets in the frame of discernment Θ: singleton $\{\omega_k\}$ and Θ itself. Consequently, based on these distances between testing sample and training sample and on the degree of membership function of training samples to each class, this item of evidence may be postulated to induce a basic probability assignment (BPA) over Θ. A BPA that provides a description of the uncertainty pertaining to the class of the current pattern is computed, given the available evidence. This information may be used to implement various decision rules allowing for ambiguous pattern rejection and novelty classification. The outputs of several information sources may also be combined in a data fusion context, yielding decision procedures, which are very robust to object recognition.

For each item of evidence x_{ij} and each class ω_k, a simple basic probability assignment function $m_{ij}(\omega_k)$ represents the item of evidence x_{ij} supporting hypothesis ω_k, and all

$m_{ij}(w_p)$, with $p \neq k$, represents the item of evidence supporting the supplementary set of ω_k . Therefore we can use $m_{ij}(\omega_k)$ as a degree of support for a simple basic probability assignment function with focal element ω_k . In a similar manner, if $p \neq k$, then $m_{ij}(w_p)$ represents the degrees of support for simple basic probability assignment function with a common focal element ω_k . The combination of these simple basic probability assignment functions of the supplementary set with focal element ω_k is a separable basic probability function with the degree of support $1-\prod_{p\neq k}(1-m_{ij}(\omega_p))$. This yields the corresponding basic probability assignment function as following:

$$M^i_{jk}(\omega_k)=m_{ij}(\omega_k)=\mu_{ij}(\omega_k)\phi_{ij}(d_{ii'j}) \qquad M^i_{jk}(\Theta)=1-m_{ij}(\omega_k)$$

$$M^i_{j\neg k}(\neg\omega_k)=1-\prod_{p\neq k}(1-m_{ij}(\omega_p)) \qquad M^i_{j\neg k}(\Theta)=1-M^i_{j\neg k}(\neg\omega_k)=\prod_{p\neq k}(1-m_{ij}(\omega_p))$$

Where the value of $\phi_{ij}(d_{ii'j})$ is verifying between 1 and 0. In this paper, an exponential form was postulated for $\phi_{ij}(d_{ii'j})$. $\phi_{ij}(d_{ii'j})$ is a decreasing function such as $\lim_{d\to 0}\phi(d)=1$ and $\lim_{d\to\infty}\phi(d)=0$. $d_{ii'j}$ is a Euclidean distance via formula (13). Detailed formula is defined as follow:

$$\phi_{ij}(d_{ii'j})=\exp(-d_{ii'j})\ ;\ d_{ii'j}=(x_{i'j}-x_{ij})^2$$

Combining knowledge about focal element ω_k , the total basic probability assignment functions $m'_{ij}(\omega_k)=M^i_{jk}(\omega_k)\oplus M^i_{j\neg k}(\neg\omega_k)$ are obtained for class ω_k and item of evidence x_{ij} .

$$m'_{ij}(\omega_k)=\frac{m_{ij}(\omega_k)\prod_{p\neq k}(1-m_{ij}(\omega_p))}{1-m_{ij}(\omega_k)[1-\prod_{p\neq k}(1-m_{ij}(\omega_p))]}$$

So basic probability assignment function (BPA) that represents the item of evidence x_{ij} belonging to each singleton set is finally obtained in the frame of discernment, and the rest BPA is assigned to the frame of discernment Θ itself.

3.3Classification fusion and decision rule

In order to classify a new vector x , these BPA of N evidence item need to integrated in P information sources respectively. First, the BPA of N evidence item is combined in the same information source. The resulting BPA represents the degree of belief that a new vector x belongs to each class in the frame of discernment Θ . Secondly fusing P information sources, the degree of belief that a new vector x belongs to each class in the frame of discernment Θ is finally obtained. The magnitude of contribution that N evidence item support the belief degree of pattern to be classified is different. Those training patterns situated far from x actually provide very little information, even build down the belief degree that pattern to be classified belong to some class, i.e. it is conflictive between some evidence item of pattern in training set T and testing set X . As a result of considering training pattern in turn, if N BPAs that can be combined using the Dempster's rule of combination, the complexity of computation is very high. Therefore, in order to cut down the complexity of computation and the negative influence of belief degree that pattern to be classified belongs to some classes in Θ ; it is sufficient to consider the K nearest neighbors of T in this sum. The BPA representing all evidence of training set with K P-dimension is defined as the orthogonal sum therefore:

$$m_j(\omega_k)=\bigoplus_{i=1}^{K}m_{ij}'(\omega_k)\ ;\ m(\omega_k)=\bigoplus_{j=1}^{P}m_j(\omega_k)$$

The total belief degree of that pattern to be classified belongs to each class is obtained based on multi-sources information fusion. According to fuzzy rule, final decision of pattern to be classified is made. The rule of making-decision is defined as: $L=\arg\max_{k=1,\cdots,M}(m(\omega_k))$, Where L takes values in $\{1,\cdots,M\}$. So the classification fusion of pattern to be classified finished in the testing set X .

4. Application In Network Troubleshooting

According to Gabarit approximation [6], data flow could be represented by continuum condition discrete time AR Markov model. If we let $\lambda(n)$ express the bit rate of No.n packets, then one rank AR Markov equation is shown as follows by the using of recursion relation: $\lambda(n-1)=l\cdot\lambda(n)+m\cdot\varpi(n)$, l and m are influence genes. Following experience, we can evaluate l ＝0.8781 . m ＝0.1108. $\varpi(n)$ is the independence gauss white noise sequence, and its mean is 0.572, its variance is 1[5]. Every node loading could be expressed by the formula: $L=\sqrt{\sum_{i=1}^{n}(k_i a_i^2)}$. In the formula, L represents the loading value of local node; $a_1,a_2,\dots,a_n$, are loading targets; $k_1,k_2,\dots,k_n$, are weights.

The simulation environment in this paper is NS2. Figure 2 presents the network topology of simulation experiment. Design a share bottleneck connection A and B in the topology, and the bandwidth of link is 100Mbit/s.

Other link bandwidths are all 5Mbit/s. The link delay between node A and B is a variable, whose value is between 10ms and 300ms. Let packets length be 1024bytes. The buffer length in router A and B are all 40M. m=n=40. The send-velocity minimum of every node is 300kbits/s， and the maximum is 1500kbits/s. We choose data link traffic prediction as a example.Set sampling cycle be 40ms，and get 400s sampled data to train. Get eight sets at random once more. In each sets, 320s data is to test. Based upon these hypothesis and parameter, the result of training is shown as figure 3. The result that predicts the queue length at the next time is shown as figure 4. The results of comparing BPN with our new method are shown in table 1.

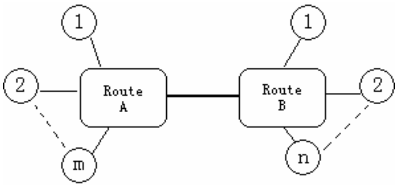

Fig.2. Simulation network topology

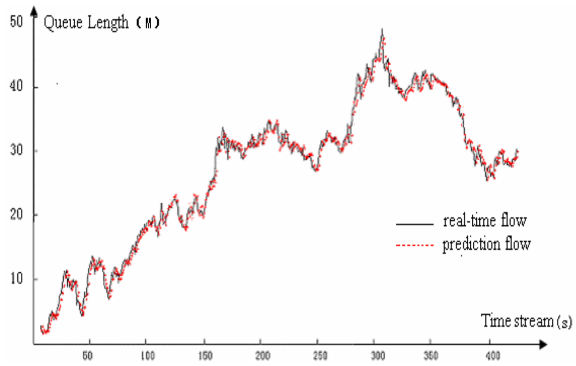

Fig.3.The result of training

The first row represents TSE (total squared error)， and the second and the third row represent iteration degree, when BPN or our new method reach corresponding TSE. We also can hold MSE (mean square error, MSE＝TSE / 400). From table 1, we can find that convergence rate of our new method is more quickly than BPN in evidence, and our new method could gain wee TSE and MSE in tolerable iteration degree.

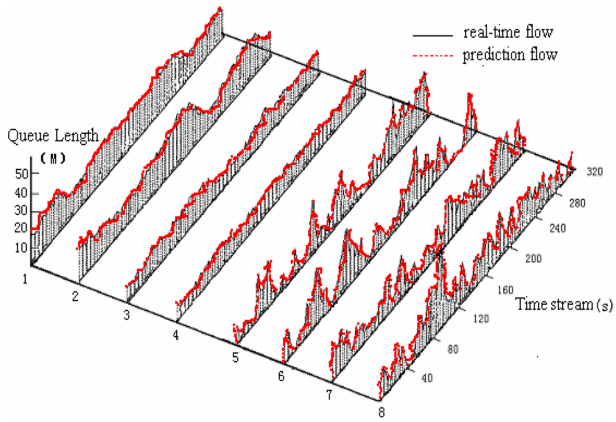

Fig.4. Prediction result of node data flow

Tab. 1. Iteration degrees of BPN and our new method

TSE	***2.43***	***2.23***	***2.03***	***1.83***	***1.63***
BPN	238	633	13232		
our method	58	69	81	95	118

5. Conclusion

This approach can be regarded as combining ideas from unsupervised neural network models and using neighborhood information in the framework of the Dempster-Shafer theory of evidence. This method consists in considering each neighbor of a pattern to be classified as an item of evidence supporting certain hypotheses concerning the class membership of that pattern. And this evidence is represented by basic probability assignments (BPA's). The final belief degree of a pattern to be classified is pooled using the Dempster's rule of combination in the frame of multi-sources information fusion. Experiments demonstrate that this classification scheme has more excellent performance compared with existing fuzzy Kohonen neural network techniques.

References

[1] G. Rogova, "Combining the results of several neural network classifiers," Neural Networks, 1944,7(1): 777~781.
[2] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," IEEE Trans. Syst., Man., Cybern., 2002, 22(3): 418~435.
[3] FANG Yong, "The Research of Data Fusion from Multiple Sources by Evidential Reasoning," Journal of Remote Sensing, 2004,4(2):106~111. (In Chinese)

[4] R. R. Murphy, "Dempster-Shafer Theory for Sensor Fusion in Autonomous Mobile Robots", IEEE Transactions of Robotics and Automation, 2004,14(2):121-127.
[5] M. Abdulghafour, A. Fellah, M. A. Abidi, "Fuzzy Logic-Based Data Integration: Theory and Applications," IEEE Conference on MFI94, Las Vegas, Nevada, 2004:241-248
[6] Christoph L, Schuba, Ivan V, Krsul. "Analysis of denial of service attack on TCP,"Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy, 2000: 342-348.
[7] A. Cheaito, M. Lecours, "Study of a modified Dempster-Shager approach using an expected utility interval decision rule," SPIE, Vol.3719,2005: 34~42.
[8] Caserri C, Meo M. "A new approach to model the stationary behavior of TCP connections," In: Proc IEEE INFOCOM2000, Tel Aviv, Israel, CA: IEEE Computer Society, 2000:245-251.
[9] Veres A, Boda M. "The chaotic nature of TCP network troubleshooting," In: Proc INFOCOM2000, Tel Aviv, Israel, CA:IEEE Computer Society,2000: 283-190.
[10] Floyd S, Fall K. "Promoting the use of End-to-End network troubleshooting in the Internet," IEEE/ACM Transaction Networking, 2002,7(4): 458-472.
[11] Harris B, Hunt R. "TCP/IP security threats and attack methods," Computer Communications, 2002,22(10): 885-897.
[12] Hamann T, Walrand J. "A new fair window algorithm for ECN capable TCP," In: Proc IEEE INFOCOM2000, Tel Aviv, Israel, CA: IEEE Computer Society, 2000: 435-441.
[13] He Hua-Can et al. Generalized logic in experience thinking. Science in China(E), 1996, 39(3): 225-234
[14] Buckley J J, Siler W. A new t-norm. Fuzzy Sets and Systems,2004,16(1):283-290

Li Qianmu received the B.S. and Ph.D. degrees in Computer Science and Technology from Nanjing University of SCI. and TECH. in 1999 and 2005, respectively. His research interests include wireless LAN, ad hoc networks, network troubleshooting and quality of service.

Xu manwu received the Ph.D. degree in Computer Science and Technology from Nanjing University in 1980. His research interests include wireless networks, quality of service and network performance.