第 29 卷 第 4 期
2003 年 7 月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 29, No. 4
July, 2003

A Recognition Method for Static Words of Chinese Sign Language Based on Fuzzy-Neuro Network[1)]

ZOU Wei YUAN Kui DU Qing-Xiu XU Chun
(*Institute of Automation, Chinese Academy of Sciences, Beijing* 100080)
(E-mail: wei.zou@mail.ia.ac.cn)

Abstract In this paper, a novel recognition method of single-hand static words of Chinese Sign Language based on fuzzy-neuro network is introduced. First, the fuzzy reasoning rules and the network structure are established using empirical knowledge. Then the membership function parameters for each fuzzy subset are obtained by learning. For the learning process, a new kind of empirical risk function is proposed which is differentiable and can be minimized by gradient descent strategy. This method is compared with others through experiments and its validity and reliability are confirmed.

Key words Chinese sign language, fuzzy-neuro network, reasoning rules, empirical risk function, CAS-Glove

1 Introduction

As a kind of nature language, sign language is the primary mode of communication for most deaf people. The aim of the recognition of sign language is to provide computers with an efficient and accurate mechanism to translate human sign language into text or speech[1]. Words of Chinese sign language (CSL) are mainly expressed by two types of information: one is the hand pose information, which includes bending angles of hand joints and relative positions of fingers; the other is the gesture information which is the trajectory formed when the hand moves in the 3D-space. So, from the viewpoint of recognition, the CSL words can be divided into single-hand words and dual-hand words and the single-hand words can be further divided into two categories: static words and dynamic words. In this paper we only consider static single-hand words.

In this paper, a novel recognition method of single-hand static words of CSL based on fuzzy-neuro network (FNN) is introduced, which uses the CAS-Glove[2] as the input device. There are 18 bending sensors fixed on the CAS-Glove to measure joints' angles of the user's fingers and wrist. The fuzzy reasoning rules and the structure of the radical basis function (RBF) networks are established using the empirical knowledge. Then the membership function parameters of each fuzzy subset are obtained by learning. For the learning process, a new kind of empirical risk function is proposed which is differentiable and can be minimized by gradient descent strategy. In experiments this method is compared with the method of fuzzy reasoning based on clustering and the method of feature matching, and the validity and reliability of this method are confirmed.

2 Structure of fuzzy-neuro network[3~6]

Suppose $X=[x_1, x_2, \cdots, x_N]$ is a vector in the N dimensional space R^N, and the classifier D is the mapping: $R^N \rightarrow W$, where $W=\{\omega_1, \omega_2, \cdots, \omega_M\}$ is the set of M class labels.

The fuzzy inference classifier is mainly based on the M fuzzy rules such as:

R_m: IF (x_1 is A_{1m}) and (x_2 is A_{2m}) and…and (x_N is A_{Nm}) THEN (class is ω_m)

---

1) Supported by National Natural Science Foundation of P. R. China (60273028) and "863" Project of P. R. China (2001AAN114200)

Received May 17, 2002; in revised form November 27, 2002
收稿日期 2002-05-17; 收修改稿日期 2002-11-27

where $A_{im}(1<i<N, 1<m<M)$ are the input fuzzy term sets defined in the domain of the input variable x_i. All the rules can be expressed by the neural network just shown as Fig. 1, which can be divided into 5 layers:

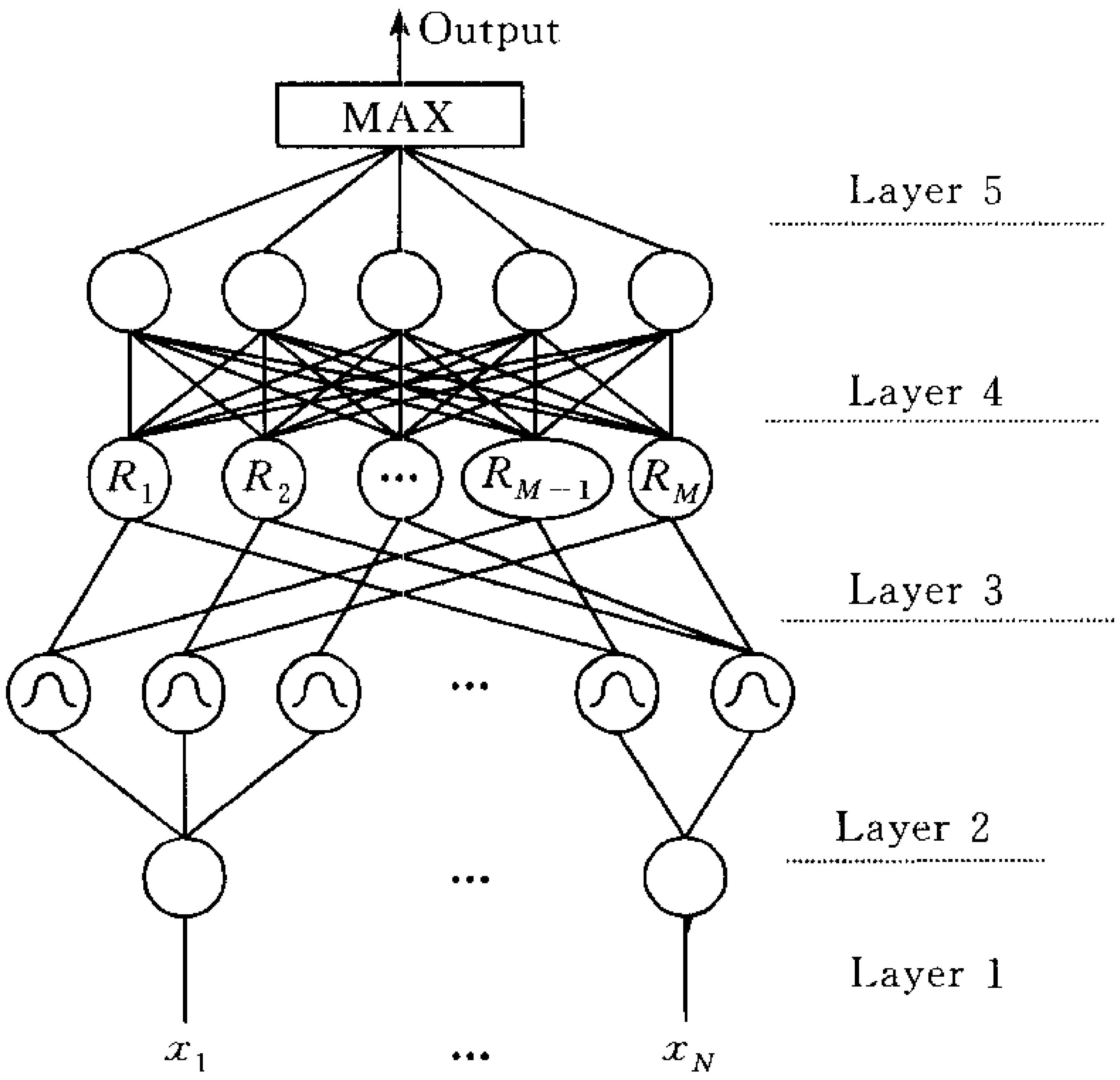

Fig. 1 A fuzzy-neuro network used as classifier

Layer 1. Input layer. In this layer, the number of units equals the dimensions of the input vector.

Layer 2. Fuzzification layer. In this layer, $\mu_{im}(1\leqslant i\leqslant N, 1\leqslant m\leqslant M)$, the degree of membership to fuzzy term set A_{im} of input variable x_i, can be obtained by membership function (MF). Here, MF is a radial basis function defined as:

$$\mu_{im} = \exp(-(x_i - a_{im})^2/b_{im}) \tag{1}$$

Layer 3. Rule layer. In this layer, the firing level of the m-th rule, μ_m, is obtained by using the product operator:

$$\mu_m = \prod_{i=1}^{N} \mu_{im}, \quad 1 \leqslant m \leqslant M \tag{2}$$

Layer 4. Normalization layer. Nodes in this layer perfom the normalization of firing level coming from Layer 3 as:

$$s_m = \mu_m / \sum_{j=1}^{M} \mu_j, \quad 1 \leqslant m \leqslant M \tag{3}$$

Layer 5. Output layer. The strategy of Winner-Take-All is adopted in this layer to select the best output from all the rules as the recognition result. Namely, if $k=\arg\max_m\{S_m\}$, the output class label is ω_k and $X\in$ class ω_k.

3 Parameter learning

3.1 Empirical risk function

If the output of the classifier is ω_k, we can define the M dimensional output vector of network as $Y=\{0,\cdots,\overset{k-1}{0},\overset{k}{1},\overset{k+1}{0},\cdots,0\}$. Let the group of learning data be $\{X^{(k)}, D^{(k)}\}$, where $1\leqslant k\leqslant K$ and K is the number of data, $X^{(k)}$ is the input vector, $D^{(k)}$ is the teacher vector of the classifier corresponding to the k-th input vector and $D^{(k)}=\{d_m^{(k)}\}, 1\leqslant m\leqslant M$. Then for the k-th input vector, if the actual output vector of the network is $Y^{(k)}=\{y_m^{(k)}\}$, $1\leqslant m\leqslant M$, the empirical risk of this vector and the total empirical risk of all learning data can be defined as:

$$E_w^{(k)} = \frac{1}{2}\sum_{m=1}^{M} | y_m^{(k)} - d_m^{(k)} | \tag{4}$$

$$E_w = \frac{1}{2K}\sum_{k=1}^{K} E_w^{(k)} = \frac{1}{2K}\sum_{k=1}^{K}\sum_{m=1}^{M} | y_m^{(k)} - d_m^{(k)} | \tag{5}$$

The vector $Y^{(k)}$ can be determined according to the output vector $S^{(k)}$ of the 4-th layer in Fig. 1, where $S^{(k)}=\{s_m^{(k)}\}, 1\leqslant m\leqslant M$. If $s_{\mathrm{Max}}^{(k)}=\max\limits_{m}\{s_m^{(k)}\}$, then

$$y_m^{(k)} = \lim_{w_1\to\infty} (s_m^{(k)}/s_{\mathrm{Max}}^{(k)})^{w_1}, \quad 1\leqslant m\leqslant M$$

and $s_{\mathrm{Max}}^{(k)}$ can be expressed as:

$$s_{\mathrm{Max}}^{(k)} = \lim_{w_2\to\infty} \| s^{(k)} \|_{w_2} = \lim_{w_2\to\infty} \Big(\sum_{j=1}^{M} s_j^{(k)^{w_2}} \Big)^{1/w_2}$$

If w_1 and w_2 are both big enough and if $w=\max\{w_1, w_2\}$, then we can get

$$y_m^{(k)} \approx \Big[s_m^{(k)} / \Big(\sum_{j=1}^{M} s_j^{(k)^w} \Big)^{1/w} \Big]^w = s_m^{(k)^w} / \sum_{j=1}^{M} s_j^{(k)^w}$$

Substitute this equation into Equation (4), and refer to Equation (3). The empirical risk function can be deduced as:

$$E_w^{(k)} \approx \frac{1}{2}\sum_{m=1}^{M} \Big| s_m^{(k)^w} / \sum_{j=1}^{M} s_j^{(k)^w} - d_m^{(k)} \Big| = \frac{1}{2}\sum_{m=1}^{M} \Big| \mu_m^{(k)^w} / \sum_{j=1}^{M} \mu_j^{(k)^w} - d_m^{(k)} \Big| = \frac{\sum\limits_{\substack{m=1\\ m\neq m_k}}^{M} \mu_m^{(k)^w}}{\sum\limits_{j=1}^{M} \mu_j^{(k)^w}} \tag{6}$$

where m_k is the index of the element whose value is 1 in vector $D^{(k)}$.

The empirical risk function proposed is differentiable and can be minimized by gradient descent strategy, so it is very convenient to be applied to the parameter learning process. From analysis of deduction of Equation (6), it can be achieved that if w is big enough, the minimization of $E_w^{(k)}$ only guarantees that the maximum element in $S^{(k)}$ and the element of 1 in D^k have the same index. However, a more strict condition is to ensure the sum of all the non-maximum elements in $S^{(k)}$ is minimized so that the recognition result from Winner-Take-All strategy is more believable. If $w=1$, the risk function becomes:

$$E^{(k)} = E_1^{(k)} = \frac{\sum\limits_{\substack{m=1\\ m\neq m_k}}^{M} \mu_m^{(k)}}{\sum\limits_{j=1}^{M} \mu_j^{(k)}} = \sum_{\substack{m=1\\ m\neq m_k}}^{M} S_m^{(k)} \tag{7}$$

This risk function can meet both constrains above. Moreover, it is so simple that when applied to learning, the complexity of computation becomes less and it is easier to deduce the learning algorithm.

3.2 Parameter learning

The parameters that should be adjusted in the learning process are a_{im} and b_{im} ($1\leqslant i\leqslant N$, $1\leqslant m\leqslant M$), which lie in the MF of each fuzzy term set. According to the gradient descent strategy, the parameters can be learned by:

$$a_{im}(u+1) = a_{im}(u) - \eta\frac{\partial E^{(k)}}{\partial a_{im}} = a_{im}(u) - \eta\sum_{t=1}^{M}\frac{\partial E^{(k)}}{\partial \mu_t^{(k)}}\frac{\partial \mu_t^{(k)}}{\partial a_{im}}$$

$$b_{im}(u+1) = b_{im}(u) - \eta\frac{\partial E^{(k)}}{\partial b_{im}} = b_{im}(u) - \eta\sum_{t=1}^{M}\frac{\partial E^{(k)}}{\partial \mu_t^{(k)}}\frac{\partial \mu_t^{(k)}}{\partial b_{im}}$$

where $0\leqslant\eta\leqslant 1$ is the learning step, u is the number of iterations. A fuzzy term set can be contained in antecedents of different rules, so suppose F is the set of the rules whose antecedents are all interrelated with parameters a_{im} and b_{im}. Refer to Equations (1), (2) and

(7), the iteration of all the parameters can be deduced as:

$$a_{im}(u+1)=a_{im}(u)-\eta\sum_{t\in F}\left\{\left[(1-\delta_{tm_k})\frac{s_{m_k}^{(k)}}{\sum_{j=1}^{M}\mu_j^{(k)}}+\delta_{tm_k}\frac{s_{m_k}^{(k)}-1}{\sum_{j=1}^{M}\mu_j^{(k)}}\right]2\mu_t^{(k)}\frac{(x_i^{(k)}-a_{im}(u))}{b_{im}(u)}\right\}$$

$$=a_{im}(u)-2\eta\sum_{t\in F}\left\{[(1-\delta_{tm_k})s_{m_k}^{(k)}+\delta_{tm_k}(s_{m_k}^{(k)}-1)]s_t^{(k)}\frac{(x_i^{(k)}-a_{im}(u))}{b_{im}(u)}\right\} \quad (8)$$

and

$$b_{im}(u+1)=b_{im}(u)-2\eta\sum_{t\in F}\left\{[(1-\delta_{tm_k})s_{m_k}^{(k)}+\delta_{tm_k}(s_{m_k}^{(k)}-1)]s_t^{(k)}\frac{(x_i^{(k)}-a_{im}(u))^2}{b_{im}^2(u)}\right\} \quad (9)$$

where $\delta_{tm_k}=\begin{cases}0, & t\neq m_k\\ 1, & t=m_k\end{cases}$

4 Experiments and result analysis

Recognition of 32 static sign words from [7] is experimented with the network described above. The fuzzy inference rule for each word is established based on the empirical knowledge and the parameters are learned according to Equations (8) and (9).

In the experiments, the CAS-Glove is used as the input device, on which 18 bending sensors are fixed.

Table 1 Recognition result of static words in CSL

Methods	FNN		FCI		Methods	FNN		FCI		Average Results	
Indexes	RR(%)	ARB	RR(%)	ARB	Indexes	RR(%)	ARB	RR(%)	ARB		
thick	86	0.73	67	0.14	add	75	0.22	39	0.06	FNN	
thin	72	0.15	48	0.07	subtract	50	0.08	52	0.13		
good	92	0.35	57	0.26	meter	77	0.59	54	0.33		
bad	95	0.83	89	0.77	divide	99	0.78	87	0.84	RR	ARB
small	84	0.68	58	0.58	zero	69	0.42	34	0.12		
party	96	0.91	87	0.93	one	49	0.12	53	0.17	81%	0.48
ministry	87	0.43	62	0.12	two	91	0.33	76	0.19		
department	67	0.25	26	0.16	three	94	0.54	67	0.13	FIC	
bureau	52	0.23	32	0.09	four	99	0.61	86	0.65		
office	77	0.21	45	0.17	five	100	0.92	94	0.91		
section	86	0.58	51	0.53	six	92	0.61	74	0.42	RR	ARB
gun	67	0.24	54	0.04	seven	72	0.47	29	0.16		
yuan	74	0.39	43	0.39	eight	89	0.56	53	0.23	56%	0.33
jiao	78	0.40	46	0.32	nine	63	0.09	31	0.11		
toilet	97	0.87	65	0.37	ten	78	0.37	42	0.07		
ginger	92	0.64	26	0.18	mountain	93	0.75	68	0.78		

Because the objects to be recognized are static sign words, only 15 sensors shown in Fig. 2 are taken into consideration. According to the characteristics of Chinese sign language, the output angle spaces of all PIP sensors except the thumb PIP are divided into such fuzzy term sets as: Small (S), Smaller (SE), Bigger (BE) and Big (B), while the output space of all the other sensors (including the thumb PIP Sensor) are divided into Small (S), Middle (M) and Big (B).

The curves of initial MF and learned MF are drawn in real line and dotted line respectively in Fig. 3 (only 4

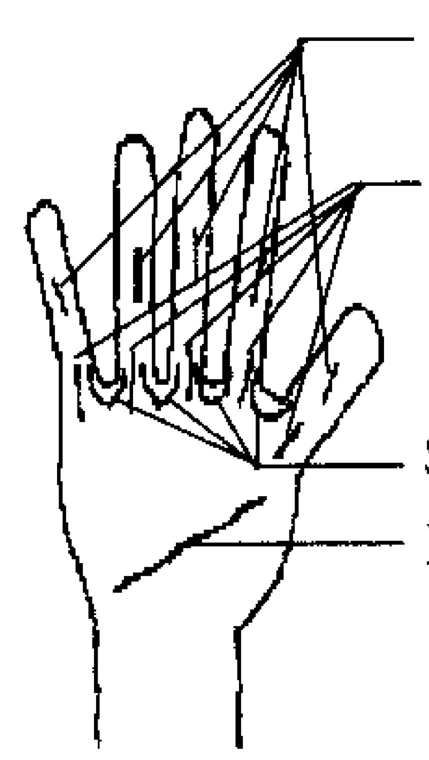

Fig. 2 Sensors on the CAS-Glove

sensors are given, others are similar). The learning data were sampled from 6 persons, who expressed each word normatively according to [7]. Averagely, the number of samples for each word was more than 70. It is obvious that there are differences between the initial MF and the learned MF of each fuzzy term set, which mainly result from the nonlinear components of the sensor characteristics, the disparity of stress and the uncertainty of initialization knowledge.

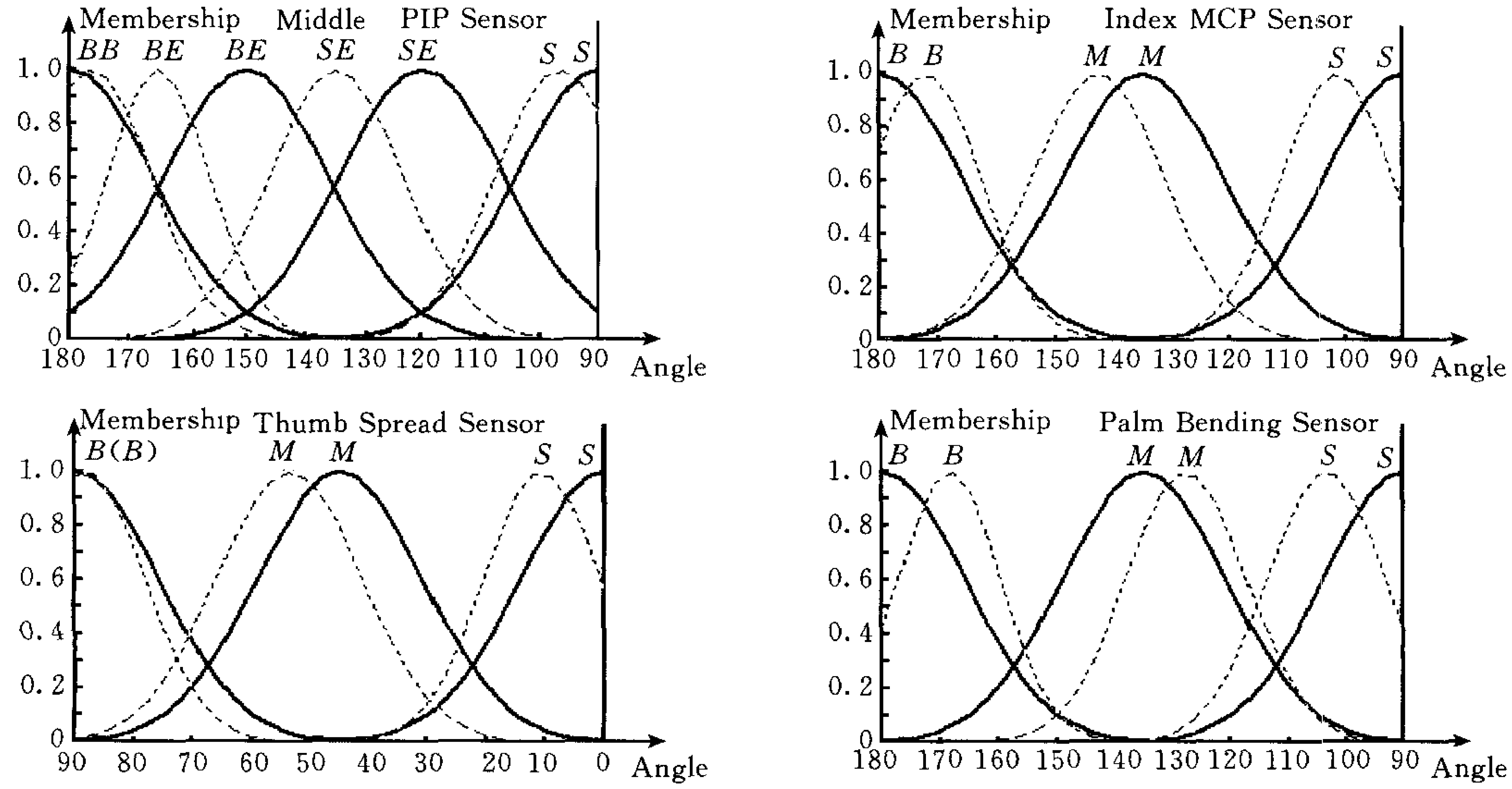

Fig. 3 MF of fuzzy term set

This FNN method is compared with the simple fuzzy inference method based on clustering (FIC) for CSL recognition. In the FIC, parameter a_{im} of each fuzzy term set is defined as the mean of all the data belonging to this set and b_{im} is the central second moment. Besides the recognition rate (RR), we define another performance index, i. e., the average recognition degree of belief (ARB). When a word is recognized correctly and the normalized output is s_m, $1 \leqslant m \leqslant M$, we denote $m_k = \arg\max\limits_{m}\{s_m\}$ and call $s_{m_k} - \max\limits_{m \neq m_k}\{s_m\}$ as the degree of belief (DB) of this word. The mean of DB on all times of correct recognition is called the average recognition degree of belief (ARB).

The experimental results are shown in Table 1 (For the reason of space, the recognition times for each word is not listed. In experiments, every word was recognized more than 182 times and expressed by several persons. The lower the RR of a word, the more times this word was expressed). From the table, it can be concluded that the method proposed in this paper has a higher RR and is more credible. However, some words have a lower RR, such as "subtract" and "one". A main reason is that the hand poses of these words are similar and the difference is only the orientation of fingers. Because we only apply the information of hand pose, these words can not be distinguished from each other.

5 Conclusion

In this paper, a novel recognition method of single-hand static words of CSL based on fuzzy-neuro network is introduced, which uses the CAS-Glove as the input device. The fuzzy rules are set up and the network structure is built based on the empirical knowledge. The parameter learning algorithm is obtained by minimizing the empirical risk function proposed in this paper. This method combines both knowledge information and learning data adequately and effectively. Compared with other methods, the proposed method has a higher recognition rate and is more credible.

References

1 M Waleed Kadous. Machine recognition of auslan signs using Power Gloves: Towards large-lexicon recognition of sign languages. In:Proceedings of the Workshop on the Integration of Gesture in Language and Speech, Wilmington, DE, 1996. 165～174

2 Ma Xiao-Jun. Research on sign language information processing system and data fusion methods[Ph. D. Thesis]. Beijing: Institute of Automation, Chinese Academy of Sciences(in Chinese)

3 Ovidio Griore, Inge Gavet. Neuro-fuzzy models for speech pattern recognition. European Symposium on Intelligent Techniques. Orthodox Academy of Crete, Greec, 1999

4 Ludmila I Kuncheva, Dimiter V Lakov. RBF networks versus fuzzy if-then rules for classification. *International Journal Knowledge-Based Intelligent Eng Systems*, 1998, (2): 203～210

5 Klawonn F, Kruse R. Derivation of fuzzy classification rules from multidimensional data. Advances in Intelligent Data Analysis. The International Institute for Advanced Studies in Systems Research and Cybernetics, Windsor, Ontario, 1995. 90～94

6 Wang Shi-Tong. Fuzzy System, Fuzzy Neural Network and Application Programming. Shanghai: Shanghai Sciences and Technology Document Press, 1998(in Chinese)

7 Chinese Deaf-Mute Association. Chinese Sign Language. Beijing: Huaxia Publishing Company, 1991(in Chinese)

ZOU Wei Received his master degree from Shandong Universiy of Technology in 2000 and now is a Ph. D. candidate in Institute of Automation, Chinese Academy of Sciences. His research interests include pattern recognition, virtual reality,etc.

YUAN Kui Received his master and Ph. D. degrees from Kyushu University in 1985 and 1988, respectively. Now he is a professor and the director of Hi-tech Innovation Center, Institute of Automation Chinese Academy of Sciences. His research interests include autonomous robotics, virtual reality human-computer interaction, etc.

DU Qing-Xiu Received her master degree from Harbin Institute of Technology. Her research interests include virtual reality, human-computer interaction, etc.

XU Chun Received her bachelor degree from Northeastern University in 2000 and is pursuing her master degree in Institute of Automation,Chinese Academy of Sciences. Her research interests includes virtual reality,etc.

基于模糊神经网络的静态手语词汇识别

邹 伟 原 魁 杜清秀 徐 春

(中国科学院自动化研究所 北京 100080)

(E-mail: wei.zou@mail.ia.ac.cn)

摘 要 利用数据手套CAS-Glove作为输入设备，提出了一种基于模糊神经网络的中国手语单手静态词汇的识别方法：首先利用经验知识为每个词汇创建模糊规则，然后通过学习确定各模糊子集隶属函数中的参数．对于参数的学习，提出了一种适用于分类器的可微经验风险函数，该函数能够有效地利用梯度下降法进行最小化．在实验中通过比较证实了该方法的有效性和可靠性．

关键词 中国手语，模糊神经网络，推理规则，经验风险函数，CAS-Glove

中图分类号 TP391.4