Rule Extraction from a Modified FMM Neural Network for Sign Language Recognition

Ho-Joon Kim, Seung-Kang Lee

Abstract—This paper presents a rule extraction method using a modified FMM neural network. The suggested method supplements the hyperbox definition with a frequency factor of feature values in the learning data set. We have defined a relevance factor between features and pattern classes. A modified hyperbox membership function is introduced and the learning algorithm with the model is presented in this paper. The excitatory features and the inhibitory features can be classified by the proposed method and they can be used for the rule generation process. From the experiments of sign language recognition, the proposed method is evaluated empirically.

Keywords—Rule extraction, FMM neural network, sign language recognition, pattern classification

I. INTRODUCTION

The limit of knowledge acquisition and representation is one of the major problems in the knowledge-based approaches for visual pattern recognition. A principal advantage of the neural network is its ability of acquiring the knowledge of the given problem from the instance-based learning process. However, a drawback of the conventional neural network approaches is that the system is essentially doing pattern matching. While the learning algorithm can find the association between the input and the output, the system does not know how and why they are related. As a result, the system lacks the explanation capability, and this may intolerable in some applications.

Many researchers have been working on the recognition of various sign languages and gestures, but this research poses major difficulties due to the complexity on hand and body movements in sign language expression[1-2]. Recently several approaches to represent the motion information for the human action recognition in video have been reported. Weinland et al. proposed a human action recognition model using 3D volume structures called Motion History Volume (MHV) as a free-viewpoint representation for human actions in the case of multiple calibrated video cameras[3]. Yilmaz et al. proposed a novel action representation method named action sketch which is generated from a view-invariant action volume[4].

Simpson introduced a fuzzy min-max (FMM) neural network based on fuzzy hyperbox sets representing the data clusters [5]. As a sequel to this, another FMM model for unsupervised learning was proposed by redefining the membership function and the hyperbox creation criteria. A few years later, by integrating the classification and clustering features of the original two FMM models and generalizing some features, Gabrys and Bargiela developed a general fuzzy min-max (GFMM) neural network [6]. However, the FMM model, the basis of the GFMM neural network, has shortcomings in that classification results can be significantly affected by both the order of training patterns and small amount of distorted information. Therefore, the operation of the GFMM network also shows these traits. For this reason, in [7], the classification performance of the FMM model was improved on the account of the frequency of training patterns. Quteishat et al. proposed a two-stage pattern classification and rule extraction system based on a modified FMM neural network and a GA rule extractor[8].

Fig. 1 The underlying sign language recognition system

In this paper, we employ the frequency factor used in [7] and propose a rule extraction method using the relevance factors between feature values and pattern classes. We have applied the proposed method to the sign language recognition problem. Six types of sign language patterns have been used for the experiments. We have extracted three types of features, motion history volume, motion energy data and hand-shape features. From the experimental results, we discuss the validity of the proposed model.
II. THE UNDERLYING SYSTEM MODEL

As shown in Fig. 1, our underlying system model consists of three modules: preprocessing module, feature extraction module and pattern classification module. In the preprocessing module, the feature regions are detected through the skin color analysis process. We have used the motion energy data and the motion history data and the hand-shape features for the feature extraction. An extended version of CNN model[9] is used for the feature map generation. The model generates a three dimensional feature map from the motion history data. For the pattern classification, we have adopted the weighted FMM model[7] which can provide a feature analysis facility using a feature relevance measure. A modified version of the WFMM model are introduced and it application to rule generation is presented in this paper.

III. THE ORIGINAL GFMM MODEL

The modified features of the FMM model in the GFMM neural network can be described by four main points. First, the input can be fuzzy hyperbox sets as well as crisp-point patterns. Second, the membership function is modified to a form that use fuzzy set operation, and the hyperbox expansion constraint has been changed. Third, the network is capable of being trained by both labeled data and unlabeled data, and can perform classification and clustering separately or in combination. Fourth, the maximum hyperbox constraint can adaptively be changed. The hyperbox defined by Simpson is defined in the GFMM model as the ordered set

\[ B_j = \{X_h, U, V_j, b_j(X_h, U, V_j)\} \]  

(1)

In this definition, the \( h \)-th input pattern set \( X_h = [X_h^1, X_h^u] \), where \( X_h^l \) and \( X_h^u \) are the vectors of the min point and the max point of the input hyperbox, respectively. For hyperbox \( B_j \) in a \( n \)-dimensional space, \( U_j = (u_{j1}, u_{j2}, \ldots, u_{jn}) \) is the min point vector, \( V_j = (v_{j1}, v_{j2}, \ldots, v_{jn}) \) is the max point vector. The characteristic function \( b_j \) determines the membership value as follows:

\[ b_j(X_h) = \min(\min([1 - f(v_h^u - v_{hi}, \gamma_j)], [1 - f(u_h^l - x_{hi}, \gamma_j)])) \]  

(2)

where

\[ f(r, \gamma) = \begin{cases} 1 & \text{if } r\gamma > 1 \\ r\gamma & \text{if } 0 \leq r\gamma \leq 1 \\ 0 & \text{if } r\gamma < 0. \end{cases} \]  

(3)

The ramp threshold function \( f \) determines the extent of the distance between the input pattern and the hyperbox boundary. The sensitivity parameter \( \gamma \) indicates the steepness of the slope for the fuzzy boundary region. The GFMM algorithm allows hyperbox to expand when the following criterion is met:

\[ \forall (\max(v_{hi}, x_{hi}) - \min(u_{hi}, x_{hi})) \leq \theta \]  

(4)

The growth parameter \( \theta \) determines the hyperbox maximum range for each dimension. If the input pattern \( x_h \) is within the minimum distance \( \theta \), the hyperbox in consideration is expanded; otherwise a new hyperbox is created.

Among these new features, the feature that affects the classifiers performance most would be the second feature where the form of the decision space depends on. However, the property of this membership function is that the contribution of individual input training patterns is ignored. While dealing with practical problems, this property can be problematic in two aspects that are discussed in the next section.

IV. FEATURE EXTRACTION FROM SIGN LANGUAGE PATTERNS

Convolutional neural networks(CNN) incorporate constraints and achieve some degree of shift and deformation invariance using spatial subsampling and local receptive fields[9]. When an image pattern is input, spatially-localized subset of units (receptive fields) are passed through the two-dimensional processing element in the subsequent layers. The convolution layers have orientation-selective filter banks where elementary visual features are extracted from the spatial template. The filtered image is then subsampled by the subsampling layer. Spatial resolution is reduced in this process and certain amount of translation is ignored. Therefore, each sub-layer generates a feature map which reflects successively larger ranges of the preceding unit. In this paper, we introduce an extended version of the CNN for temporal feature extraction as shown in Fig. 2.

![Fig. 2 The extended CNN model for feature extraction](image)

The input data for the feature extractor are represented as a spatiotemporal volume which is described in the previous section. The spatial structure of the receptive field in the model is extended along the time axis. The center of the three-dimensional processing element shifts through the spatial and temporal domain of the cube by two positions. Thus, the proposed model is not only robust to spatial variance but also to temporal variance.

The size of the initial feature map used in this research is \( 23 \times 23 \times 23 \). After the input data is processed through the feature extractor, a final feature map of size \( 3 \times 3 \times 3 \) is generated. This feature map becomes the input of the pattern classifier.

Motion history information is used for the feature extraction module in our model. We have CNN model to extract feature maps from the motion history volume[3]. Fig. 3 shows an
example of the feature map generated from the sign pattern in video. In the figure, the direction of time sequence is from the left column to the right column. For each frame in the image sequence, the object region is cut out by a background subtraction and contour detection method. We refer to motion as the occurrence of object region pixels between contiguous images, i.e. if the object region did not exist at an image point (x, y) at time t and appeared at the same location at time t+1, it indicates that the point is a region of motion. By stacking the motion information along the time dimension, we obtain a spatiotemporal volume data. Since motion is to occur near the boundary of the object region, the template provides a certain degree of shape information as well as the direction of the object movement.

Fig. 3 An example of the spatio-temporal volume

V. RULE EXTRACTION METHOD

In this paper, we propose a method for rule extraction for pattern classification. We have defined a modified membership function of hyperbox in FMM neural networks as follows:

$$B_j(\mathbf{h}_i) = \frac{1}{Z} \sum_{i=1}^{n} w_{ji} f(a_{hi}, I_{ji})$$  \hspace{1cm} (5)

In the equation, $A_h = (a_{h1}, a_{h2}, \ldots, a_{hn}) \in I^n$ is the h-th input pattern which consists of n features. $I_{ji}$ is the feature value interval of the i-th dimension in the j-th hyperbox. $w_{ji}$ is the connection weight between the i-th feature and the j-th hyperbox.

The function $f$ is defined as

$$f(a_{hi}, I_{ji}) = \begin{cases} 
1.0 & \text{if } I_{ji}^L < a_{hi} < I_{ji}^U \\
1.0 + \gamma (a_{hi} - I_{ji}^L) & \text{if } a_{hi} < I_{ji}^L \\
1.0 - \gamma (a_{hi} - I_{ji}^U) & \text{if } a_{hi} > I_{ji}^U 
\end{cases}$$  \hspace{1cm} (6)

where the parameter $\gamma$ controls the slope of the fuzzy membership function at the boundaries of the feature range.

The learning process of the model consists of two processes: hyperbox creation and hyperbox expansion. In the learning process the weight updating rule is defined as

$$w_{ji}^{new} = \begin{cases} 
w_{ji}^{old} + \lambda & \text{if } |I_{ji}^{old}| = |I_{ji}^{new}| \\
w_{ji}^{old} \cdot \frac{|I_{ji}^{old}|}{|I_{ji}^{new}|} & \text{otherwise} 
\end{cases}$$  \hspace{1cm} (7)

where $|I_{ji}|$ means the size of the i-th feature range of the j-th hyperbox. In other words, each weight value increases in proportional to the frequency factor of the feature value and in inverse proportional to the size of the feature range. The weights between the features and the hyperboxes reflect the relevance factors between them as different values. Therefore, we can resolve the hyperbox overlapping problem without using the hyperbox contraction process. From the trained FMM neural network, the relevance factor between features and pattern classes can be calculated as

$$FR(I_{ji}, k) = \left( \frac{1}{N_k} \sum_{i=1}^{n} w_{ji}^k - \frac{1}{N_{B_k}} \sum_{i=1}^{n} w_{ji}^k \right) / \sum_{i=1}^{n} w_{ji}^k$$  \hspace{1cm} (8)

where the function $FR(I_{ji}, k)$ means the relevance factor between the feature range $I_{ji}$ and the pattern class k. In the equation, $C_k$ is the set of hyperboxes which belong to the class k, $N_B$ is the total number of hyperboxes and $N_k$ is the number of the hyperboxes which belong to the class k.

We have classified the feature ranges of each dimension as 5 fuzzy partitions as shown in Fig. 4.

Fig. 4 Fuzzy partitions for feature ranges

If the relevance factor $FR(I_{ji}, k)$ has positive value, it means that the feature $I_{ji}$ is an excitatory signal for the pattern class k. On the other hand, an negative value of relevance factor means the inhibitory relationship between the feature and the pattern class.

From the analysis of the relevance factors we can generate a set of rules for the pattern classification as follows. The first step is to generate the feature list in which the absolute value of the relevance factors are larger than a threshold value. The second step is to find the IF-THEN rules by analyzing the relevance factors. If the relevance factor $FR(I_{ji}, k)$ has a positive value m, and $I_{ji} = 0.15$, the following rule is generated.

IF (xi is small) THEN pattern k with (cf = m)
where \(cf\) means ‘confidence factor’ which is defined as the absolute value of the relevance value.

We have considered 27 motion history features which are named as \(M = \{m1, m2, \ldots, m27\}\). Each motion history feature has the meaning related with its spatio-temporal location in the motion history volume. For example, the feature \(m4\) means ‘left-center-early motion’. Therefore the symbolic rules like the following example are generated by the rule extraction process.

\[
\text{IF (lower-center-early motion is medium large)} \\
\quad \text{THEN the sign pattern is ‘depart’ with (}\ cf = 0.48)\]

VI. EXPERIMENTAL RESULTS

The first experiment is to evaluate the classification capability of the proposed model for the hyperbox overlapping area. Fig. 5 shows the comparison of the two different decision makings inside the hyperbox region.

Fig. 5 Comparison of the two different decision makings inside the hyperbox region

In this case, if an input pattern as shown in Fig. 5(a) (marked \(\times\)) is near data points of class B, the original classifier will determine this pattern as of class A, while it would seem more appropriate to classify it as class B. Moreover, half of the input training patterns that were subject to the hyperbox of class B are being excluded after the hyperbox contraction occurred. Within the hyperbox region, as in Fig. 5(c), the membership values are equal and thus the hyperbox contraction is required. The membership values are equal and thus the hyperbox contraction is required. Fig. 5(b) shows the hyperbox boundaries by using the proposed method. Although an overlap exists between the two hyperboxes, by comparing the hyperbox gain as shown in Fig. 5(d), the disambiguation can be achieved.

The second experiment is to evaluate the rule extraction method for the sign language recognition system. Six types of sign pattern classes (\(P1: \text{greeting}, P2: \text{meet}, P3: \text{depart}, P4: \text{glad}, P5: \text{thank you}, \) and \(P6: \text{very}\)) have been considered for the experiments. Fig.6 shows an example of the user interface of the system.

The examples of the motion history volume extracted from the sign pattern data are shown in Fig.7.

![Fig. 6 An example of the user interface of the sign language recognition system](image)

The CNN-based feature extractor generates \((3 \times 3 \times 3)\) three dimensional feature maps in which each cell has the meaning of the spatio-temporal location in the volume.

![Fig. 7 Examples of the motion history volumes: (a) greeting, (g) depart](image)

Examples of the motion history volume extracted from the sign pattern data are shown in Fig.7. The CNN-based feature extractor generates the three dimensional feature maps from these volumes. TABLE I shows the relevance factors between the features and the pattern classes. As shown in the table, we can classify the excitatory features and the inhibitory features from them.

<table>
<thead>
<tr>
<th>feature</th>
<th>meaning</th>
<th>Pattern</th>
<th>Relevance factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m1)</td>
<td>Left-upper-early</td>
<td>(P1)</td>
<td>0.250 (excitatory)</td>
</tr>
<tr>
<td>(m5)</td>
<td>Center-middle-early</td>
<td>(P3)</td>
<td>-0.493 (inhibitory)</td>
</tr>
<tr>
<td>(m12)</td>
<td>Right-upper-middle</td>
<td>(P5)</td>
<td>0.000 (don’t care)</td>
</tr>
<tr>
<td>(m14)</td>
<td>Right-middle-middle</td>
<td>(P6)</td>
<td>0.454 (excitatory)</td>
</tr>
<tr>
<td>(m16)</td>
<td>Left-lower-middle</td>
<td>(P6)</td>
<td>-0.137 (inhibitory)</td>
</tr>
<tr>
<td>(m22)</td>
<td>Left-middle-latter</td>
<td>(P6)</td>
<td>0.008 (inhibitory)</td>
</tr>
</tbody>
</table>

The following rules are the examples of the rule extraction results. From the relevance factors which have relatively large values, a set of IF-THEN rules are generated as follows:
If ('left_upper_early motion' is medium large)  
THEN 'greeting' with cf=0.250

If ('right-middle-middle motion' is medium)  
THEN 'regret' with cf=0.454

If (center_middle_early motion is large)  
THEN Not('depart') with cf=0.493

If (left_lower_middle motion is large)  
THEN Not('very') with cf=0.137

VII. CONCLUSION

In the conventional FMM neural networks, the hyperbox contraction process is required to resolve the ambiguity in the overlapping area. However, the proposed model can solve the ambiguity problem without using the overlapping test process and the contraction process. We have conducted the rule extraction experiments using the motion history data. For the performance improvement, the additional features such as the motion energy data and hand-shape data can be considered in the learning process. The modified CNN model achieves some degree of shift and deformation invariance using spatio-temporal 3D receptive fields. The modified FMM neural network model is capable of utilizing the feature distribution and the frequency factor in the learning process as well as the classification process. Since the weight factor effectively reflects the relationship between feature range and its distribution, the system can prevent undesirable performance degradation which may be caused by noisy patterns.

ACKNOWLEDGMENT

This research was financially supported by the Ministry of Education, Science Technology (MEST) and National Research Foundation of Korea (NRF) through the Human Resource Training Project for Regional Innovation.

REFERENCES


