Interweaving of Syntax and Semantics in Algorithms For Recognising Chinese Characters

David Al-Dabass [1] And Manling Ren[2]

[1] Dept of Computing, The Nottingham Trent University, Nottingham, NG1 4BU, UK
david.al-dabass@ntu.ac.uk
[2] Xerox Limited Technical Centre, Welwyn Garden City, Hertfordshire AL7 1HE, UK
manlingren@hotmail.com

Abstract: The structure of Chinese characters is reviewed and seen to be best represented as 3-layer hierarchy of character, radical and stroke. Fuzzy possibilistic reasoning is then put forward as an appropriate set of conceptual tools to investigate the automatic recognition of these characters. An associative memory artificial neural network algorithms form a suitable technique for realising these concepts. Implementing these techniques several issues are explored: vagueness of radicals, their situation, position invariance, extraction order and shape. Extensive results are obtained to demonstrate the quality of the algorithms in dealing with the range of difficulties inherent in the problem.

Keywords: Chinese character recognition, fuzzy possibilistic reasoning, associate memory neural network, topological structure.

1 Introduction

The complex structure of Chinese characters is formulated through a long history (about 5,500 years recorded in history). Early Chinese characters were mainly symbols and *pictographs* that could also represent some abstract concepts of daily life as shown in Figure 1 (a). In order to express more complex ideas and concepts, *pictographs* were developed and combined to form *ideographs* for multiple meanings. These *ideographs* form some 90% of the total Chinese characters in current usage [Scu91]. Most *ideographs* are made up of two components: (a) a radical, i.e. a *pictograph* before it becomes part of an *ideograph*, which indicates the classification of a character; and (b) a '*phonetic*' symbol for partially aid the pronunciation of a character. Figure 1 (b) shows several examples of the *ideographs*' development. Chinese characters possess three major features in their structures and quantities: a two-dimension (2-D) pictorial format, topological structure and a large vocabulary.

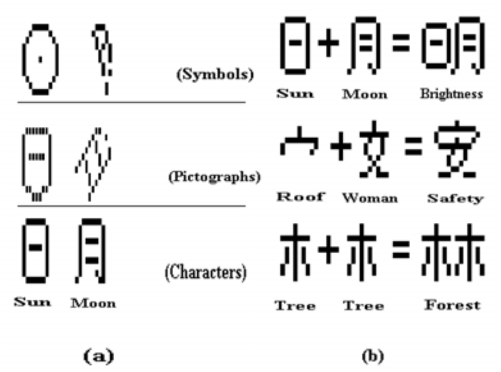


Figure 1. The historical development of Chinese characters

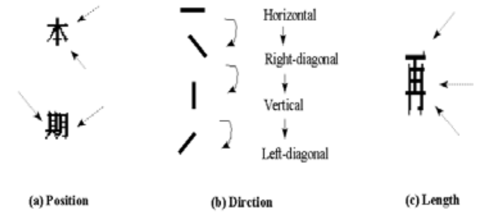


Figure 2. 2-D pictorial format of Chinese characters

Figure 3. Topological structure of Chinese characters

In the **2-D** pictorial format, basic components, strokes, can be situated at any position of a character. Figure 2 (a) shows that the stroke '*horizontal line*' can be located at several places in a character. In Figure 2 (b), the stroke '*horizontal line*' may change its identity once its direction is altered. Figure 2 (c) displays that the stroke '*horizontal line*' has three different lengths in a character.

The **topological structure** of a character means that the character is combined from or deconstructed into several components as shown in Figure 3 (a). The same component may appear in different characters as illustrated in Figure 3 (b). A component can be located at different positions in a character as shown in Figure 3 (c).

2 Structure of Chinese Characters

2.1 Three-Layer Hierarchy

In the three-layer hierarchy method, the structure of Chinese characters is represented in three layers: character, radical and stroke [Ren96]. Basic strokes in the bottom layer are treated as indexes to determine the shape of radicals in a character. Radicals in the second layer are used to deconstruct the internal topological structure of a character in order to reduce the number of characters to be learnt by computer. Characters in the top layer are recognised by restructuring radicals into a chain code, and verifying it by means of a code database. Based on this method, the process of recognising a character is carried out in sequence of character, radical and chain code. Figure 4 gives an example for illustration of the method.

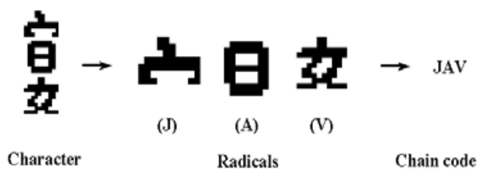


Figure 4. An example proceeded by the three-layer hierarchy

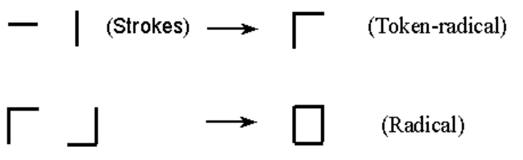


Figure 6. Combination of strokes

The structure representation in the three-layer hierarchy centres on the relationship of objects in these layers in two aspects: extraction and classification. Extraction takes into account the relationship of objects in their shape alteration between layers, which is also called vertical alteration. In the second aspect, classification generalises the relationship of objects in the same layer and forms different categories of these objects, which is termed horizontal classification.

iii) Stroke: A stroke is defined as a dot or a continuous line. Although radicals form the foundation of extraction and classification of the hierarchy, the structure of their shape has to be defined entirely by these primitive strokes. For instance, a combination of horizontal and vertical strokes can form a token-radical, or even a radical, as shown in Figure 6. Therefore, the structural representation in this layer will concentrate on these combinations as well as strokes themselves.

3 Algorithms

3.1 Fuzzy Possibilistic Reasoning

Fuzzy possibilistic reasoning in knowledge representation approaches is well suited to dealing with imperfect, uncertain and vague information [Kru94]. Reducing the complexity of imperfect information is achieved by information-compressed representations based on if-then rules. These rules are interpreted as logical implications and are termed as possibilistic inference rules defined by the notation $\Re$.

i) Concepts: Based on the approximate reasoning and probability theory, fuzzy possibilistic representation uses conjunctly combined rules to validate a possible resolution from various restrictions. In these if-then rules, antecedent in the *IF* clause and consequence in the *THEN* clause are constrained by their possibility distributions denoted by π. The possibility distributions are related with the interpretation of vague concepts as contour functions of random sets. Physical quantities of the distributions are defined by the possibility measures denoted by Poss_π.

ii) Expression: Generally, a possibilistic inference rule $\Re_j$ can be expressed by

$\Re_j$: **IF** ξ^{S}_{j} **is** μ_j **THEN** ξ^{T}_{j} **is** v_j, $j = 1, ..., r$,
or
$\Re_j$: **IF** $\xi^{S\,(1)}_{j}$ **is** $\mu_j^{(1)}$ **AND** $\xi^{S\,(2)}_{j}$ **is** $\mu_j^{(2)}$ **THEN** ξ^{T}_{j} **is** v_j, $j = 1, ..., r$,

where μ_j, $\mu_j^{(1)}$, $\mu_j^{(2)}$ and v_j are subsets of possibility distributions on the space sets S_j and T_j with regard to j. ξ is a variable whose values can be arbitrary possibility distributions on S_j or T_j. The symbol **is,** appearing in possibilistic inference rules, serves as a linguistic description of the operator $\subseteq$ and is therefore to be interpreted as 'is at least as specific as'.

3.2 Associative Memory Neural Networks

The associative memory neural networks, **Bi-directional associative memory** (BAM) introduced by B. Kosko in 1985 [Kos87] and **Hopfield memory** introduced by J. Hopfield in 1982 [Hop82], are able to recognise an incomplete pattern with their associative memory. The network architecture can be built up with neurones and connectivity on one layer or more layers.

i) Algorithms: The mathematical formula for the associative memory function is established on the construction of an energy equation E [Hop82] [Kos87] [Kos88], called the Steepest Gradient Descent algorithm:

$$E = -\sum_i \sum_j X_j W_{ij} Y_j + \sum_i \theta_i X_i + \sum_j \varphi_j Y_j \quad \cdots \quad (3.1)$$

Where, θ_i and φ_j are constants of the energy equation E.

The algorithm contains two phases: learning and training. In the learning phase, the associative memory function is used to form the connectivity matrix W for training a set of input patterns X_i (u) and output patterns Y_j (u),

where u = 1, 2 ... M; i, j = 1, 2 ... N, the weight W (i, j) is determined by the Hebbian rules:

$$W(i, j) = \begin{cases} \sum_u X_i(u) Y_j(u) & if \quad i \neq j \\ 0 & if \quad i = j \end{cases} \quad \cdots \quad (3.2)$$

In the training phase, the algorithm aids convergence because its value in equation (3.1) is either reduced or to remain constant during the recall procedure [Wan94], providing the following conditions are satisfied.

$Y_j^{n+1} = \begin{cases} 1 & \sum_i W_{ij} X_i^n - \varphi_j > 0 \\ Y_j^n & \sum_i W_{ij} X_i^n - \varphi_j = 0 \\ -1 & \sum_i W_{ij} X_i^n - \varphi_j < 0 \end{cases}$ $\cdots \quad (3.3)$ $X_i^{n+1} = \begin{cases} 1 & \sum_j W_{ij} Y_j^n - \theta_i > 0 \\ X_i^n & \sum_j W_{ij} Y_j^n - \theta_i = 0 \\ -1 & \sum_j W_{ij} Y_j^n - \theta_i < 0 \end{cases}$	The relation $\Re$ of all rules is $\Re = \bigcap_{j=1}^{r} \Re_j$

4 Implementation

4.1 Vagueness of Radicals

According to the definition of the three-layer hierarchy method, radicals are seen as the basis of recognising a character. Essentially, a radical possesses pictographic features of uncertain position, shape and extracting order. The **position** means that a radical can be located at any place in a character, for instance, bottom, left or outside as shown in Figure 7 (a). A radical can be within a rectangle, square, u or y **shape** as illustrated in Figure 7 (b). The definition of shape is referred to Figure 3.3 in Section 3.2.2. The extracting **order** indicates the sequence of radicals extracted from a character. For instance, a radical located at the top of a character is extracted first, as examples shown in Figure 7 (c).

Figure 7. Vagueness of radicals

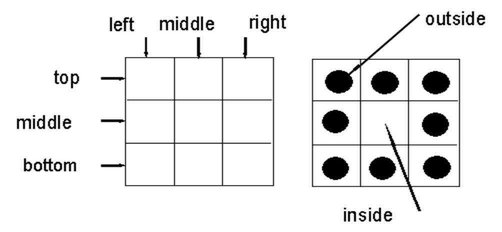


Figure 8. Possible position of a radical

4.2 Situation

The situation representation uses inference rules defined by the above interpretation for determining radicals in a character. The representation focuses descriptions on (a) the position of a radical in a character, and (b) the order of extracting a radical from a character.

i) Position Variance: The investigation of position variance of radicals in a character is based on their features of a two-dimensional picture and a rectangular appearance, one of the major characteristics in the structure of Chinese characters. The domain of position variable is defined by

P = {***width, length***}.

Because a radical may keep an independent position in a character, the possibility distribution of position variance of a radical on the domain P, shown in Figure 8, is defined by

π(P) = {***outside, inside, top, bottom, left, right, middle***}
Poss$_\pi$ (P) for π(P) is defined by, for instance,
Poss$_\pi$ (*left*) = {***width ≤ 2/3 width of P, length = length of P***}.

ii) Extraction Order: The extraction order indicates the sequence of radicals extracted from a character that might consist of two or more radicals. The domain of extraction order is expressed by

O = {***first, last***}

The possibility distribution of extraction order on the domain O is represented by

π(O) = {***outside → inside, inside → outside, top → (middle → bottom), top → bottom, left → (middle → right), left → right***}.

The notation '→' stands for the sequence from the first to the latter. Distinction of some distribution representations, such as, 'outside → inside' and 'inside → outside', will depend on inference rules between order, position and shape mentioned in the next section.

After developing such basic possibility distribution of order $\pi(O)$ above, a complex distribution could be derived, for instance,

$\pi(O)^{(1)}$ (*top → bottom (left → right)*) = {***top → bottom left → bottom right***}.

Poss_{π} (O) for $\pi(O)$ is defined by, for instance,

Poss_{π} (*top → bottom*) = {***there are two rectangles***}.

Now, possibilistic inference rules $\Re^{(PO)}$ might be established for representing relations between the position and order of a radical. As examples, several rules are shown as follows

$\Re^{(PO)}_{(1)}$: **IF** position **is** top **THEN** order **is** first,
$\Re^{(PO)}_{(2)}$: **IF** position **is** bottom **THEN** order **is** last.

iii) Shape Vagueness and Possibilistic Inference Rules: To produce a general concept of forming a radical, the shape vagueness of radicals is investigated for expressing the relation of combining two strokes. The relations can be classified as angle, location, continuous, distance and discontinuous.

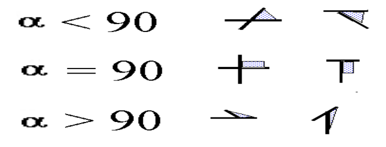


Figure 9. Angle of strokes connected

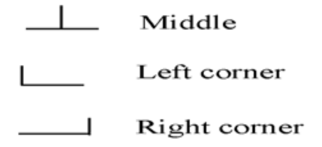


Figure 10. Location of strokes connected

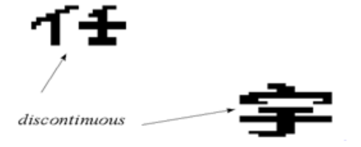


Figure 11. Discontinuous contour

The angle relation indicates a contour expression of two connected strokes. For example, it is defined as a contour if two connected strokes form an angle. Figure 9 shows three different types of angles from two connected strokes.

The location relation stands for the intersection point of two connected strokes. Figure 10 gives several examples to show the location relation. The discontinuous relation implies the possibility of a contour that may be broken down into two radicals. The distance of two disconnected strokes decides a discontinuous contour. Figure 11 gives two examples for showing the discontinuous contour.

Possibilistic inference rules are established by representations of relations between shape vagueness denoted by $\square^{(S)}$; between shape and position by $\square^{(SP)}$; and between shape, position and order by $\square^{(SPO)}$. For example, the inference rules shown below are defined to divide a character into two parts: c1 and c2 from the inside to outside.

$\Re^{(S)}_{(1)}$: **IF** contour of c1 **is** square **AND** c2 **is** continuous contour of c1
AND angle of c1 connecting with c2 **is** 90
AND location of c2 **is** on the top middle of c1
THEN shape **is** combination of c1 and c2 (c1+c2).
$\Re^{(SP)}_{(2)}$: **IF** shape **is** c1+c2 **THEN** c1 position **is**
outside.
$\Re^{(SP)}_{(3)}$: **IF** shape **is** c1+c2 **THEN** c2 position **is**
inside.
$\Re^{(SPO)}_{(4)}$: **IF** shape **is** c1+c2 **AND** position **is** outside
THEN order **is** last.
$\Re^{(SPO)}_{(5)}$: **IF** shape **is** c1+c2 **AND** position **is** inside
THEN order **is** first.

4.4 Architecture of the Associative Memory Neural Network

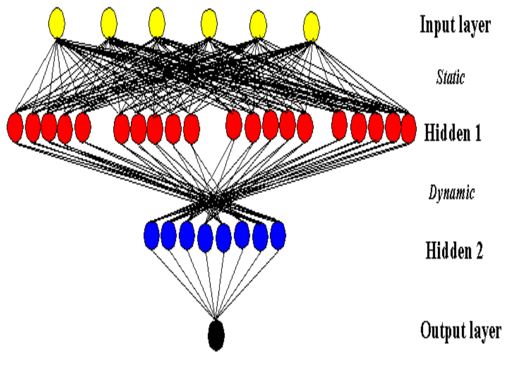


Figure 12. Architecture of the associative memory neural network

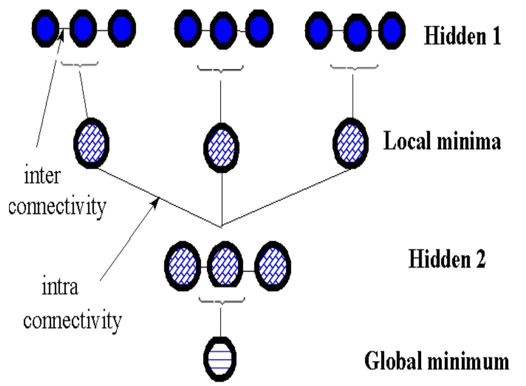


Figure 13. Function of the hidden-2 layer

The associative memory neural network in the subsystem consists of four layers: input, hidden-1, hidden-2 and output. The hidden-1 layer consists of multi sub-nets where each sub-net deals with radicals in a category. The number of neurones in each sub-net is decided by the learning patterns in the category. The connectivity from the input to the hidden-1 layer is static. Neurones in the hidden-2 layer are created by the results gained from the hidden-1 layer. The connectivity between the two hidden layers is dynamic. The design of the hidden-2 layer with a dynamic structure is used to further enhance convergence on global minimum of the associative algorithms. Figure 12 shows the architecture of the network.

In the learning phase, radicals are classified into categories. Each is represented by a sub-network that is used for reducing the connectivity of the whole network and for using shared weights. The 26 different sub-nets are composed of a whole neural network with the associative memory function. The major task in the learning phase is to learn formal radicals and to form inter-connectivity for training a pattern.

In the training phase, sub-nets in the hidden-1 layer are trained to converge to local minima. The hidden-2 layer is generalised by re-learning these patterns of local minima. Eventually, the global minimum will be converged to the output layer.

4.5 A Modified Network

The modification of the associative memory function in the network aims to enhance convergence to a global minimum [Ren95]. This modification has been made in both of the learning and training phases. In the learning phase, the modification centres on changing reasonable parameters for Hebbian rules shown in equation (3.2), so that the convergence is ideally forced to search for all patterns. Two assumptions are made:

$$\theta_i = \varphi_j = 0 \quad \textit{or} \quad \theta_i = \varphi_j = \frac{1}{2}\sum W_{ij} \qquad \cdots \quad (4.1)$$

In the training phase, the modification is concentrated on how to enhance local minima to converge to a global one. The enhancement is dynamically formed in the hidden-2 layer, as shown in Figure 13.

5 Results

5.1 Extraction of Radicals

In order to test the fuzzy possibilistic inference rules run by the Preprocessing subsystem for extracting radicals from a character effectively, some experiments have been conducted. In the extraction processing, the experimentation focused on different structures of radicals in a character and the extraction technique.

50 standard writing characters with different structures, which can be extended to more characters with the same structure, were used to examine the extraction rules, shown in Figure 14. These characters were chosen with representation of radicals in different positions, shapes and orders. The correct radicals have been extracted from 48 out of 50 test characters, i.e. a 96% success rate. In the implementation of the preprocessing, a dynamic scheme was employed for updating possibilistic inference rules while they were being evolved around a sample set of special cases.

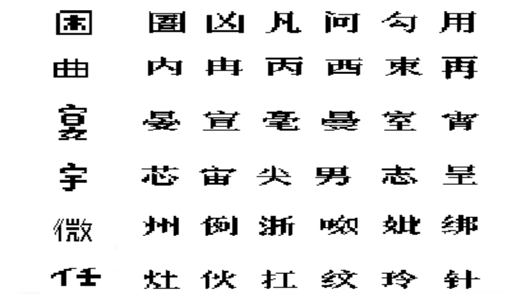


Figure 14. Test characters

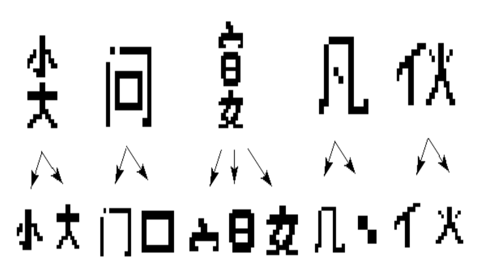


Figure 15. Different shapes of radicals

Different Positions: The implementation of extracting radicals that are located at different places in characters has examined the capability of rules to cope with various structures of characters.

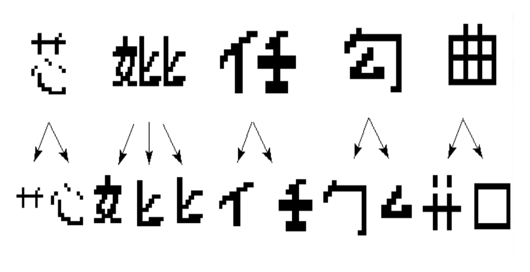


Figure 16. Results of different orders

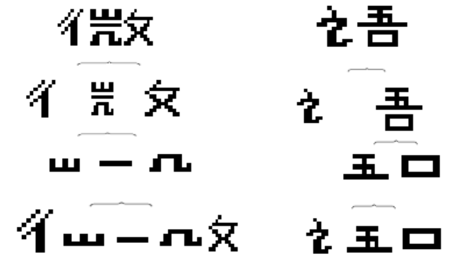


Figure 17. Examples for expansibility of rules

Different Shapes: Examining different shapes of radicals centred on dealing with basic shapes and complex shapes. The basic shapes include the shapes of basic strokes, and complex shapes are combination of basic strokes, for instance, shapes of a rectangle, square, cross, y or u shape. Figure 15 shows some results of the implementation.

Different Orders: The extraction order is able to decide which radical has to be extracted first in the case of a character being composed of more than one radical. Examples shown in Figure 16 are some results of testing order rules.

Expansibility of Rules: In addition to dividing a character as a picture into several radicals, i.e. sub-pictures, some sub-pictures might be further segmented into smaller sub-pictures, i.e. sub-sub-pictures, where the sub-pictures are formed from two or more radicals. The segmentation is termed expansibility of possibilistic rules. Examples in Figure 17 show the implementation results of the expansibility of rules.

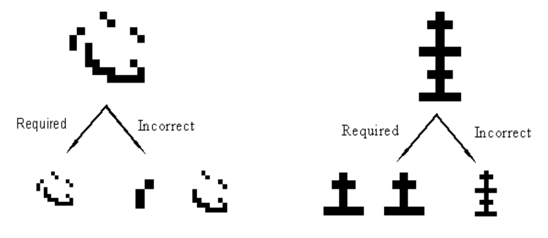


Figure 18. Two special cases

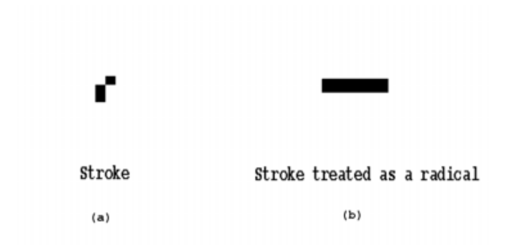


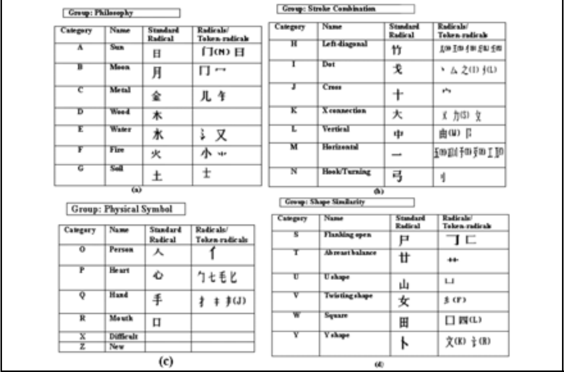
Figure 19. A case of strokes treated as a radical

For the first case, a complete radical is divided into two parts by the current inference rules. There are two methods to deal with this problem according to human analysis. One method is to have special rules, probably against existing rules, to deal with these special radicals. The other is to attach new rules that could examine a rectangular area occupied by a radical. If the area is small enough only for a stroke rather than a radical, the extraction in this case will be invalid or it will be treated as a single radical without extraction. However, it needs a statistic value to decide a minimum area for tolerating a radical because some of the strokes can be treated as radicals as shown in Figure 19 (b), but some cannot be as shown in Figure 19 (a). In the second case, two radicals are connected, or overlaid together, and form a continuous shape. This is very difficult to deal with by only applying inference rules. Other methods should be further investigated for exploring these radicals. Currently, such radicals are treated as difficult ones.

5.2 Classification and Recognition of Radicals

The experiments for classification and recognition of radicals followed on from results of the Normalisation subsystem. In the process of using the associative memory neural network with sub-net structure, the experimentation was focused on different classification, recognition and modification of the network.

120 radicals covering 24 categories have been used to examine functions of classification, recognition and translation of the system. Some ambiguous cases will be discussed in Section 5.5.5.



Group: Philosophy

Category	Name	Standard Radical	Radicals/ Token-radicals
A	Sun	日	冂(M) 日
B	Moon	月	冂 [illegible]
C	Metal	金	儿 [illegible]
D	Wood	木	
E	Water	水	氵 又
F	Fire	火	小 [illegible]
G	Soil	土	士

(a)

Group: Stroke Combination

Category	Name	Standard Radical	Radicals/ Token-radicals
H	Left-diagonal	竹	[illegible]
I	Dot	戈	丶 厶 [illegible]
J	Cross	十	[illegible]
K	X connection	大	[illegible]
L	Vertical	中	由(M) [illegible]
M	Horizontal	一	[illegible]
N	Hook/Turning	弓	[illegible]

(b)

Group: Physical Symbol

Category	Name	Standard Radical	Radicals/ Token-radicals
O	Person	人	亻
P	Heart	心	[illegible]
Q	Hand	手	扌 [illegible]
R	Mouth	口	
X	Difficult		
Z	New		

(c)

Group: Shape Similarity

Category	Name	Standard Radical	Radicals/ Token-radicals
S	Flanking open	尸	[illegible] 匚
T	Abreast balance	廿	艹
U	U shape	山	凵
V	Twisting shape	女	[illegible] (F)
W	Square	田	口 [illegible](L)
Y	Y shape	卜	文(K) [illegible](R)

(d)

Figure 20. Classification of test radicals

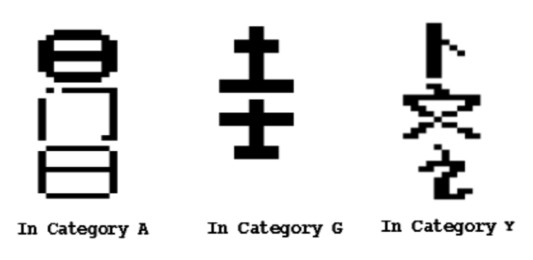


Figure 21. Some learning patterns

Classification: According to standards of classification, test radicals were divided into the categories shown in Figure 20.

Within these categories, a radical with a tag represents a combination of radicals or token-radicals that are independent in different categories in the *Cang-Jie* method. The tag is used for referring to a database of Chinese characters in the *Cang-Jie* method, instead of building up a new one.

Learning Phase: Radicals in each category were learnt by the learning phase of the network to form connectivity schemes among its sub-nets in hidden-1 layer. Some learning patterns are shown in Figure 21.

The Modified Network: Modification of the network was centred on the structure of neurons and improvement of global convergence. Results in Figure 28 show the convergence of local minima to a global minimum of radicals from the hidden-1 layer when parameters θ_i and φ_j in Equation (3.3) are set to 0. After modifying the parameters θ_i and φ_j to $\frac{1}{2} \sum W_{ij}$, the mis-recognition rates of these tests were reduced. Figure 29 shows the results of the enhancement. Compared to Figure 28, the convergence to a global minimum in Figure 29 has been improved and the number of iterations is reduced as well.

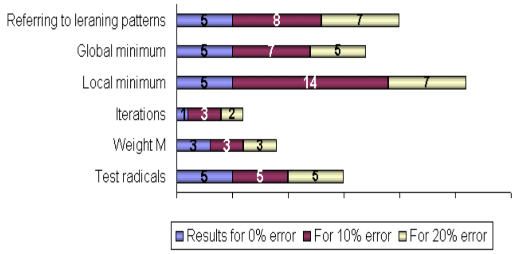


Figure 28. Test results with setting the parameters θ_i and φ_j to 0

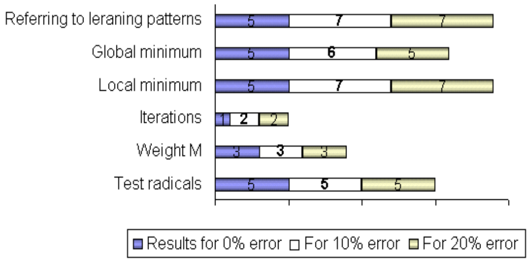


Figure 29. Test results with modifying the parameters θ_i and φ_j to $\frac{1}{2} \sum W_{ij}$

6 Conclusions

The work described in this article represents a significant advance towards using the method of three-layer hierarchy character-radical-stroke for the representation of the structure of Chinese characters, and the process of character-radical-chain code to translate a character from a 2-D pictorial format to a chain code for verification.

Although work in the preprocessing stage has classified positions of radicals in a character, a case that allows omitted and difficult radicals in a character has not been considered yet. Basically, a character in such a case has a very complex structure and it is written in a complex style. Applying fuzzy possibilistic rules to such characters and more complex characters can be investigated in future development.

References

[1] [Ben93] Bengio, Y.A., *Connectionist approach to speech recognition*, Advances in Pattern Recognition Systems Using Neural Network Technologies, 1993, World Scientific, USA, pp.3-24.

[2] [Day90] Dayhoff, J. E., Neural Network Architectures: An Introduction, 1990, Van Nostrand Reinhold, USA, pp.37-57.

[3] [Gov90] Govindan, V. K. and Shivaprasad, A. P., *Character recognition: a review*, Pattern Recognition, 1990, Vol.23, No.7, pp.671-683.

[4] [Hop82] Hopfield, J., *Neural networks and physical systems with emergent collective computational abilities*, Proc. Ntl. Acad. Sci. USA, 1982, Vol.79, pp.2554-2558.

[5] [Kli95] Klir, G. J. and Yuan, B., Fuzzy Sets And Fuzzy Logic: Theory And Applications, 1995, Prentice Hall PTR, USA, pp.200-208, pp.369-374.

[6] [Kos87] Kosko, B., *Adaptive bidirectional associative memories*, Applied Optics, 1987, 26(23), pp.4947-4960.

[7] [Kos88] Kosko, B., *Bidirectional associative memories*, IEEE Transactions on Systems, Man, and Cybernetics, 1988, 18(1), pp.49-60.

[8] [Kru94] Kruse, R., Gebhardt, J. and Klawonn, F. Foundations of Fuzzy Systems, 1994, John Wiley & Sons Ltd, UK, pp.81-155.

[9] [Nel91] Nelson, M. M. and Illingworth, W. T., A Practical Guide to Neural Nets, Addison Wesley Publishing Company, Inc., USA, 1991, pp.67-71.

[10] [Ren95] Ren, M., Su, D. and Al-Dabass, D., *An associative memory artificial neural network system*, ECAC'95-London: Proceedings of European Chinese Automation Conference, London, UK, 1995, pp.91-96.

[11] [Ren96] Ren, M., Al-Dabass, D. and Su, D., *A three-layer hierarchy for representing Chinese characters*, Research and Development in Expert Systems XIII: Proceedings of Expert Systems 96, the Sixteenth Annual Technical Conference of the British Computer Society Specialist Group on Expert Systems, Cambridge, UK, 1996, pp.137-146.

[12] [Scu91] Scurfield, E., Chinese, 1991, Hodder and Stoughton, UK.

[13] [Wan93] Wang, Y. F., Cruz, J. B. Jr. and Mulligan, J. H. Jr., *Multiple training concept for back-propagation neural networks for use in associative memories*, Neural Networks, 1993, Vol.6, pp.1169-1175.

[14] [Wan94] Wang, T., *Improving recall in associative memories by dynamic threshold*, Neural Networks, 1994, Vol.7, No.9, pp.1379-1385.