SIMULATION OF FUZZY POSSIBILISTIC ALGORITHMS FOR RECOGNISING CHINESE CHARACTERS

MANLING REN[1] and DAVID AL-DABASS[2]

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

Abstract: The structure of Chinese characters is reviewed and seen to be best represented as a 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. Associative memory artificial neural network algorithms form a suitable technique for realising these concepts. To implement 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.

Keyword: *Chinese character recognition, fuzzy possibilistic reasoning, associative memory neural networks, topological structures.*

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.

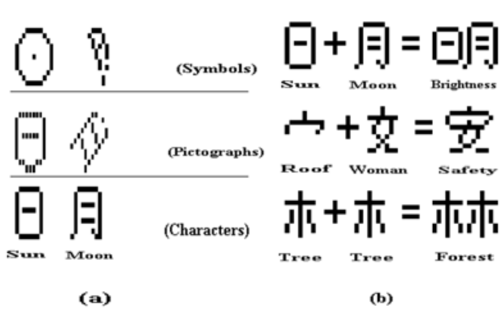

Figure 1: The historical development of Chinese characters

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.

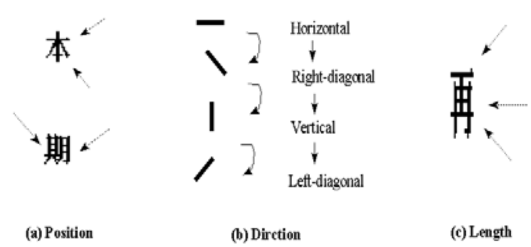

Figure 2: 2-D pictorial format of Chinese characters

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).

 ISSN 1473-804x online, 1473-8031 print

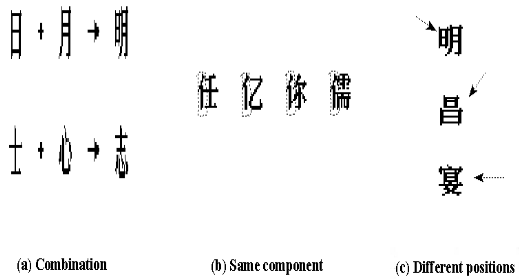

Figure 3: Topological structure of Chinese characters

The **vocabulary** of Chinese characters is defined as 3,500 characters for daily use, 7,000 characters necessary in writing, and 60,000 characters in total that include complex and simplified styles. Based on the feature of a 2-D picture, each Chinese character may be seen as a pattern different from others. Therefore, an adequate representation of a character requires a matrix of pixels about 10 times the number needed for a Roman letter. The vocabulary of Chinese characters is roughly equivalent to Western words in total [Gov90].

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.

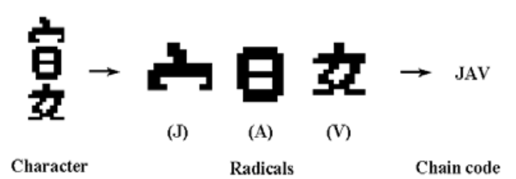

Figure 4: An example proceeded by the three-layer hierarchy

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.

i) Character

The analysis of characters is based on their topological structure embedded within the 2-D pictorial format. According to the investigation of characters in common use, 90% of characters are a combination of several components [Scu91]. The remnant is radicals existing as independent characters. These components come from original characters in the historical development of Chinese characters. Some of the original characters still keep as individual characters and some of them degenerate as only components of a character. In current Chinese dictionaries, most of these original characters are termed radicals.

A character can be seen as a pattern with the appearance of rectangle or square. The appearance is basically determined by the shape of basic strokes and their combination (It is a reason for calling Chinese characters '*square words*' a nickname, sometimes).

ii) Radical

The radical layer consists of two parts: radical and token-radical. A radical is an original pictograph in Chinese characters. It may inherit its individual meaning and occupy a completely independent position in a character. A token-radical is developed from such radicals that are further simplified to become ultimately simple combinations of basic strokes. Although some of these are overlaid or crossed in a character, token-radicals common to many characters might be classified into a certain category of radicals in terms of their abstract concepts and shape structures. In the three-layer hierarchy, radicals and token-radicals are classified into 26 different categories (24 for radicals and 2 for special cases) alongside the categories in the *Cang-Jie* method, shown in Figure 5.

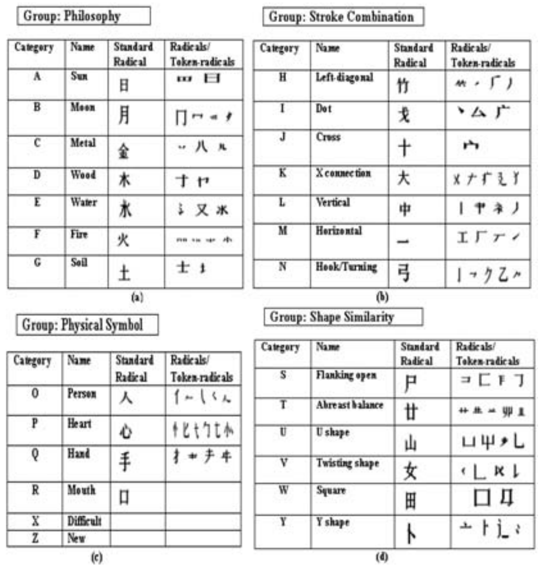

Figure 5: Classification of the *Cang-Jie* method

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.

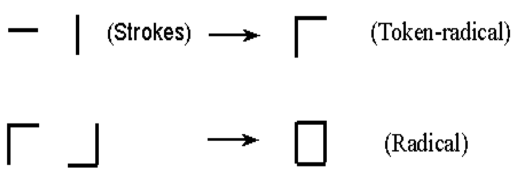

Figure 6: Combination of strokes

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 $Poss_\pi$.

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'.

The relation $\Re$ of all rules is

$$\Re = \bigcap_{j=1}^{r} \Re_j$$

iii) Possibility and Probability

There are similarities and differences between probability theory and possibility theory [Kli95]. The two theories are similar in the sense that both are subsumed not only under fuzzy measure theory, but also under the more restricted evidence theory. The differences in mathematical properties of the two theories make each theory suitable for modelling certain types of uncertainty and less suitable for modelling other types. As is well known, for instance, probability theory is an ideal tool for formalising uncertainty in situations where class frequencies are known or where evidence is based on outcomes of a sufficiently long series of independent random experiments. Possibility theory, on the other hand, is ideal for formalising incomplete information expressed in terms of fuzzy propositions.

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:

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

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:

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

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.

$$Y_j^{n+1}=\begin{cases}1 & \sum_i W_{ij}X_i^n-\varphi_j>0\\ Y_j^n & \sum_i W_{ij}X_i^n-\varphi_j=0\\ -1 & \sum_i W_{ij}X_i^n-\varphi_j<0\end{cases}$$
$$X_i^{n+1}=\begin{cases}1 & \sum_j W_{ij}Y_j^n-\theta_i>0\\ X_i^n & \sum_j W_{ij}Y_j^n-\theta_i=0\\ -1 & \sum_j W_{ij}Y_j^n-\theta_i<0\end{cases} \quad \cdots (3.3)$$

When a portion of the original pattern is used as a retrieval cue, the algorithm is denoted as auto-associative memory. When the desired output is different from the input, the algorithm is called hetero-associative. The static method requires that the network is trained in advance with specific patterns, and retrieves them from portions or distorted inputs. When the learning and retrieval are embedded in the training process randomly, it is defined as a 'dynamic' method.

ii) Evaluation

The network with the associative function has recurrence (closed loops) and non-linearity, i.e. positive weights are excitatory and will strengthen connectivity; and negative weights are inhibitory and weaken connectivity [Nel91].

It has the advantage of eliminating noise to recognise a pattern from its incomplete version. Compared with the back-propagation network, which has advantages in storing many more patterns than the number of dimensions and learning a large variety of pattern mapping relationship [Wan93] but with no guarantee of convergence [Nel91], the associative memory network can converge to a stable point even though it may be slow sometimes.

Its disadvantages lie mainly in the limitations of associative memory and convergence to a local solution rather than a global minimum [Day90]. In addition, the memory, defined as M and used for evaluating the capability of associated with patterns, is limited as a constant value as shown in equation (3.4) while the number of learning patterns n increases to a big value. For instance, M is equal to 25 when n is 1000.

$$M=\frac{n}{4\log_2 n} \quad \cdots (3.4)$$

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).

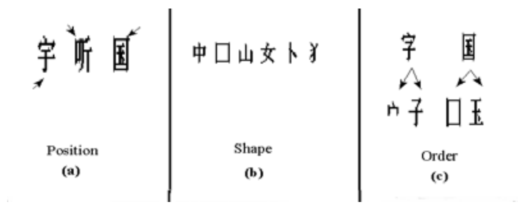

Figure 7: Vagueness of radicals

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.

According to the fuzzy possibilistic reasoning theory described in Section 3.3.5 in Chapter 3, let the notation P and O stand for two domains of the position of radicals in a character and the order of extracting a radical from a character respectively. Their possibility distributions can be defined as $\pi(P)$ and $\pi(O)$. The possibility measures are given by the notations $Poss_\pi$ (P) for $\pi(P)$, and $Poss_\pi$ (O) for $\pi(O)$. The possibilistic inference rules are represented by the notation $\Re^{(PO)}$. If p and o denote variables with the domains P and O respectively, the $Poss_\pi$ (p) is the possibility measure of p on $\pi(P)$; similarly, $Poss_\pi$ (o) for o on $\pi(O)$.

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

$\pi(P)$ = {***outside, inside, top, bottom, left, right, middle***}

$Poss_\pi$ (P) for $\pi(P)$ is defined by, for instance,

$Poss_\pi$ (*left*) = {***width ≤ 2/3 width of P, length = length of P***}.

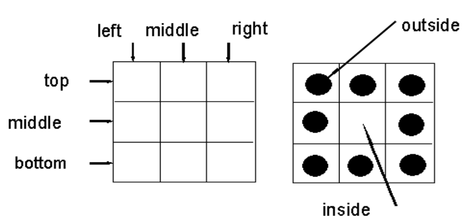

Figure 8: Possible position of a radical

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

$\pi(O)$ = {***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***}.

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

$Poss_\pi$ (*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.

4.3 Shape

The shape representation method centres on the shape domain of radicals, their possibility distributions and measure technique. Inference rules are established for the representation of radicals' relationships between their shape, position and order.

i) Possibility Distributions and Measures

The domain of radicals is defined as a rectangle in different sizes in terms of features of combined strokes. The shape domain of radicals is expressed by

S = {***rectangle***}.

Different combinations of basic strokes are assigned as the possibility distributions on the domain S, where the validity of the combinations is checked. The models of combinations are classified as connection and disconnection. The possibility distributions are represented by

π(S) = {***combination of basic strokes, basic strokes***}.

In order to determine the shape of a radical, possibility measures are based on evaluation of a continuous line, direction of a line connecting with other lines, priority of such direction and disconnecting distance. For instance, one of the possibility measures $Poss_\pi$ (S) for π(S) is defined as follows:

$Poss_\pi$ (*priority of up $\rightarrow$ down*) = {***up $\rightarrow$ down, up $\rightarrow$ down left, right $\rightarrow$ left, up $\rightarrow$ down right***}.

ii) 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.

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.

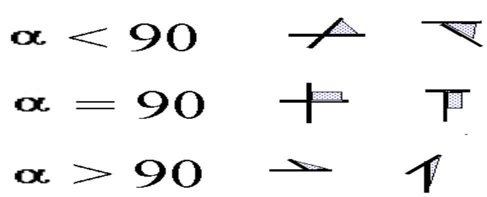

Figure 9: Angle of strokes connected

The location relation stands for the intersection point of two connected strokes. Figure 10 gives several examples to show the location relation.

Figure 10: Location of strokes connected

The continuous relation expresses the possibility of a contour as part of a radical. A continuous contour is defined if a contour is formed with an angle.

The distance relation is to measure a scope of two disconnected strokes.

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.

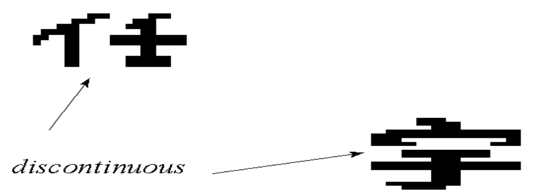

Figure 11: Discontinuous contour

Possibilistic inference rules are established by representations of relations between shape vagueness denoted by $\Re^{(S)}$; between shape and position by $\Re^{(SP)}$; and between shape, position and order by $\Re^{(SPO)}$. 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 Classification of Radicals

Since radicals have been determined as major objects for recognition, the policy of classifying radicals has to be considered carefully. Three basic principles of determining categories are developed: (a) a member in a category should have the physical properties of the category and major features of the group to which the category belongs; (b) each member in a category may be a radical or a token-radical or some combinations of basic strokes; and (c) combinations are allowed between a token-radical and basic strokes to form a new integrated radical. The policy has some benefits in transforming knowledge of a radical identified abstractly by human analysis into its shape recognised by computer.

4.5 Architecture of the Associative Memory Neural Network

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.

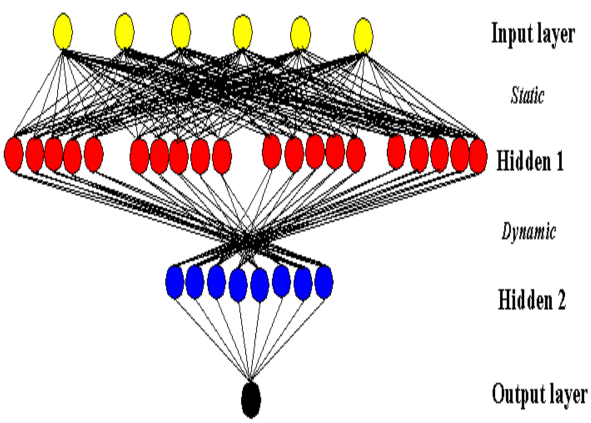

Figure12: Architecture of the associative memory neural network

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.

When the network is connected as a whole, its inter-connectivity is low but intra-connectivity is high [Ben93], while the inter-connectivity is connectivity of neurones between layers, and the intra-connectivity is ones at the same layer. The intra-connectivity can be reduced while the architecture of sub-nets is being used.

Weights in the network can be shared for saving space and time, i.e. weights that are associated with different (e.g. translated) input features may be shared, and weights that are associated with different times may be shared. In the current network, weights shared in feature space are considered, but not in time.

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:

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

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.

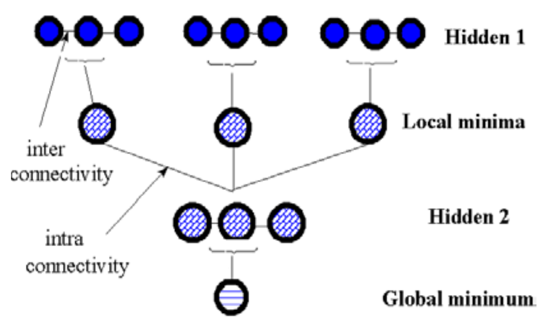

Figure 13: Function of the hidden-2 layer

The converged quality, indicated by the parameter ‘reliability rate’, is also taken into account for a global comparison of different results when referring to learning patterns. There are three possibilities of results: recognition, mis-recognition and failure. The reliability rate produced depends on the matching quality and training quantity of a pattern [Ren97].

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

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.

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.

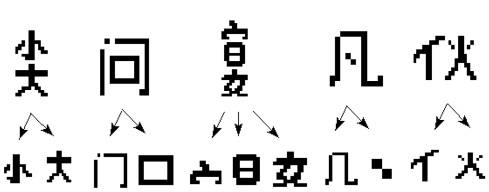

Figure 15: Different shapes of radicals

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.

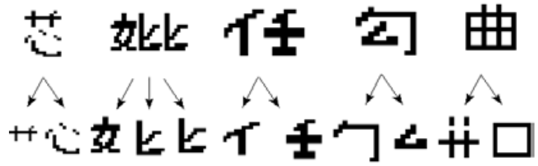

Figure 16: Results of different orders

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.

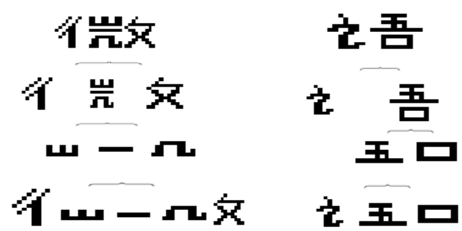

Figure 17: Examples for expansibility of rules

Special Cases: In analysing the above results, two types of incorrect results that appeared in the extraction process need to be carefully investigated. Incorrect results can arise in two special cases where (a) an individual radical has a discontinuous shape, and (b) the shape of two radicals connected together is without a discontinuous part, shown in Figure 18, for example.

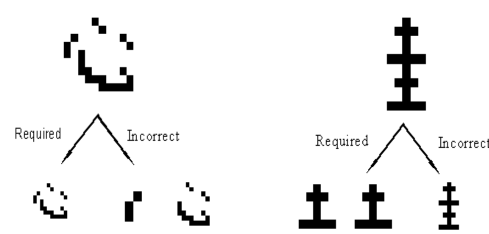

Figure 18: Two special cases

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).

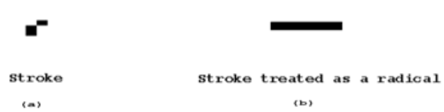

Figure 19: A case of strokes treated as a radical

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.

The implementation of recognising radicals included two phases: learning and training. The learning phase was concentrated on the effect of different weights and error tolerance of the sub-nets. The training phase examined optimal architectures and convergence 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.

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

Group: Philosophy

Category	Name	Standard Radical	Radicals/ Token-radicals
A	Sun	日	冂(M) 日
B	Moon	月	冂 [illegible]
C	Metal	金	儿 钅
D	Wood	木	
E	Water	水	氵 又
F	Fire	火	小 ⺌
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 之(I) [illegible](L)
J	Cross	十	[illegible]
K	X connection	大	乂 力(S) [illegible]
L	Vertical	中	由(W) [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] 七 毛 匕
Q	Hand	手	扌 [illegible] [illegible](J)
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	纟(F)
W	Square	田	口 四(L)
Y	Y shape	卜	文(K) 讠(R)

(d)

Figure 20: Classification of test radicals

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.

Figure 21: Some learning patterns

Different parameters for u =1, … M in Equation (3.2) were chosen for the structure of neurones in the hidden-1 layer, so that weights of the network can achieve better results. Partial **weights**, W (i, j), are shown in the horizontal axis in Figure 22 when M is equal to 2, 3, or 4 shown in the vertical axis. Figure23 shows how weights can affect the recognition of test radicals.

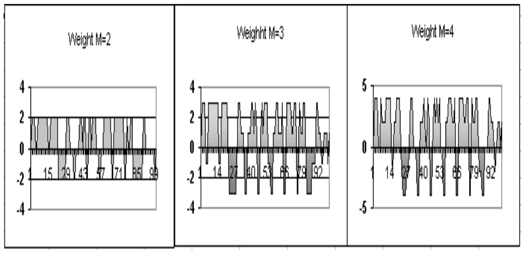

Figure 22: Different weights

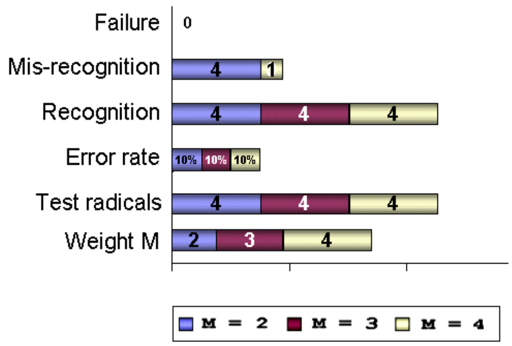

Figure 23: Effects of different weights

Three groups of test data with different error rates were used for examining the **error tolerance**. The error rates were 10%, 20% and 60%. The percentage of error rate indicates the scale of noise occurring in test data. Figure 24 gives the recognition results for this trial.

Figure 24: Results of test data with different error rates

The results in Figures 23 and 24 show that a better recognition is achieved when the weight parameter M is 3.

Using the better weight parameter, i.e. M is equal to 3, 74 radicals in 24 categories were learnt and formed the 24 sub-nets in the hidden-1 layer of the associative memory neural network in the system. The partial weights of the sub-nets are shown in Figure 25.

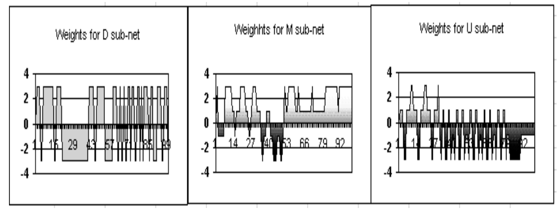

Figure 25: Weights for some sub-nets

Training Phase: Training the network was centered on the structure of neurons in the hidden-2 layer. Figure 26 shows three test groups of patterns in **different orders**. Figure 27 gives results of recognition when different structures of neurons in Figure 26 were used separately. It is clear that there is an optimal scheme, such as the order of patterns in (c) in this case, for the neuron structure, even if the difference between the three groups of learning patterns is quite small.

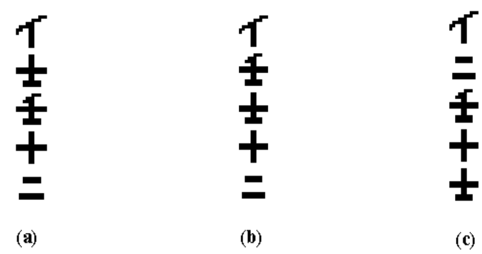

Figure 26: Patterns in different order

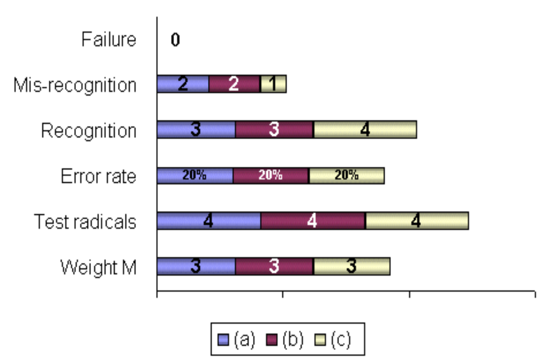

Figure 27: Results of different neuron structures

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.

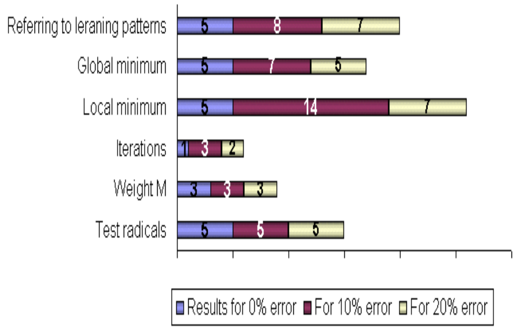

Figure 28: Test results with setting the parameters θ_i and φ_j to 0

After modifying the parameters θ_i and φ_j to ½ ΣW_{ij}, the mis-recognition rates of these tests were reduced. Figure 29 shows the results of the enhancement.

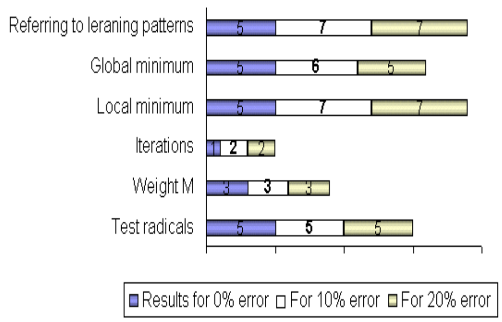

Figure 29: Test results with modifying the parameters θ_i and φ_j to ½ ΣW_{ij}

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.

112 out of 120 radicals have been recognised by using the structure of the network, where 74 radicals were different from each other. Figure 30 shows the **recognition rate** of these **radicals** (both numbers indicated by the y-axis (vertical)) in different **categories** indicated by the x-axis (horizontal)). This figure was used for examining in which category radicals have been recognised successfully. Figure 30 shows the recognition frequency of the radicals in the categories (the number for frequency is indicated by y-axis (vertical), each radical in a category indicated by the x-axis (horizontal). The results in Figure 31 were used to test which radical is the most common radical appeared in characters.

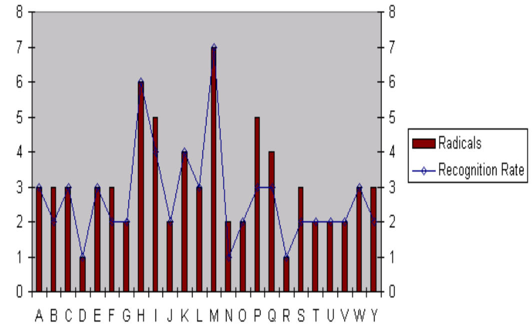

Figure 30: Recognition rate of radicals in different categories

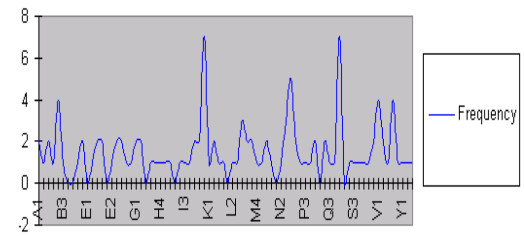

Figure 31: Recognition frequency of the radicals in the categories

5.3 Discussion

The associative memory algorithm has advantages of quicker convergence speed and the capability of recognising error patterns. According to the statistics of outcomes from the above trials, the number of iterations is less than 4 for reaching recognition. The capability of recognising the error patterns can be high, up to as much as 60%.

The ambiguity of recognising patterns is basically caused by converging to a local minimum, especially when two patterns had the same reliability rate. There is also another case of ambiguity as shown in Figure 32. The difference between patterns (a) and (b) might be treated as noise causing ambiguity of patterns to occur.

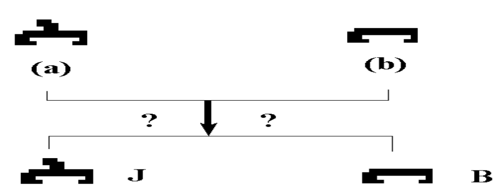

Figure 32: A case of ambiguity

In the system, dividing the whole network into sub-nets has solved the limitation of associative memory. However, convergence to local minima can still occur in the hidden-1 layer.

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.

Compared to the existing methods of Chinese character recognition, the three-layer hierarchy offers the advantages: (a) processing Chinese characters with a similar structure; (b) a more systematic representation of the internal topological structure of a character; (c) reducing the vocabulary of characters learnt by machine; and (d) using a chain code instead of a character to simplify the recognition process.

Having investigated different knowledge representation techniques, two methods of fuzzy syntactic and fuzzy possibilistic reasoning were applied. Recognising radicals used a neural network with associative memory function. The network learned different features of radicals and then recognised them. The enhancement of the network at several stages has improved its recognition rate to 96%. Outcomes of the execution time and recognition rate have shown that the network was successful.

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

[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.

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

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

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

[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.

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

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

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

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

[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.

[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.

[Ren97] Ren, M., Al-Dabass, D. and Su, D., *A fuzzy/possibilistic-reasoning AM-architecture for recognising radicals in Chinese characters*, ISFL'97-Zürich: Proceedings of Second International ICSC Symposium on Fuzzy Logic and Application, Zürich, Switzerland, 1997, pp.265-271.

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

[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.

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