Cognitive Pattern Recognition and Its Application on Chinese Character Intelligent Formation without Character Library

Youguo Pi, Wenzhi Liao, Mingyou Liu
School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, Guangdong, China

ABSTRACT

Cognitive pattern recognition regards computer as human brain and brings human beings' brain mechanism of information processing in computer to strengthen its pattern recognition intelligence, and is different from machine pattern recognition. The system structure of cognitive pattern recognition is proposed by studying the presentation of perceptive constancy, prototype matching and so on in this paper. As an application of the proposed theory and method, the system of Chinese Character Intelligent Formation without Character Library is studied and discussed, with the formation experiments of 70244 Chinese Characters.

Keywords: Cognitive Pattern Recognition, Topology, Prototype Matching, Chinese Character Intelligent Formation.

1. INTRODUCTION

Cognitive psychology compares human brain to computer, and regards human brain as the information processing system similar to computer. It aims to reveal the internal psychological mechanism of human in cognitive process, which contains how to acquire, store, process and use information. Pattern recognition, as a typical perceptual process, relies on existing knowledge and experience. Then the research on pattern recognition of cognitive psychology includes matching, knowledge, memory, etc [1,2].

So far, machine pattern recognition (pattern recognition for short) mainly contains two basic methods: statistics pattern recognition and structure (syntax) pattern recognition. Structure pattern recognition, basing on image features of structure, accomplishes pattern recognition by using dendriform information of the layered structure of pattern and subschema [3]. Statistics pattern recognition, basing on the type probability density function of samples in feature space, separates pattern statistics into types. The multiple statistics pattern recognition is achieved by searching the hyperplane of optimal decision under the decision rule of Bayesian decision theory which aims to minimize the expected risk of prior probability and lost function. Researches develop surrounding decision rules and probability density distribution [4-10]. Statistical pattern recognition is based on the statistical theory, requiring the number of samples be able to reflect the statistical law. Although some papers claim that the SVM can solve the small sample size problem, the decision rule of it has not been reported yet. The learning abilities of human beings are outstanding, pattern recognition, as an intelligent system, also needs to improve its system learning ability.

Pattern recognition is a hot topic both in the fields of cognitive psychology and information science, the former focusing on exploring the intelligent mechanism of human beings, while the latter applying its mechanism into computer and constituting intelligent system by extending it.

This paper proposes that cognitive psychology is information processing psychology, which borrows ideas from the information processing mechanism of computer and regards human beings as computer. Similarly, the information science can also borrow ideas from the achievements of psychology, regarding computer as human beings, and carry out the study of cognitive pattern recognition.

The contribution of this paper are: (1) researching on the expression of the related theories of prototype matching in information science, (2) proposing the theory and...
system architecture of cognitive pattern recognition, (3) applying the theory of cognitive pattern recognition on Chinese Character Intelligent Formation without Character Library.

The rest of this paper is organized as follows: perceptual constancy, object composition theory, prototype theory are discussed after introducing the theory of prototype matching in cognitive psychology in section 2. The system architecture of cognitive pattern recognition is proposed in section 3. In section 4, experiment of the theory of cognitive pattern recognition is presented and the system of Chinese Character intelligent formation without character library is introduced. Finally, conclusion and future work are drawn.

2. THEORY OF COGNITIVE PATTERN RECOGNITION BASED ON PROTOTYPE MATCHING

Current cognitive psychology has proposed such theoretical matching models or hypothesis about pattern recognition as the Theory of Template, the Theory of Prototype, and the Theory of Feature. The theory of prototype matching believes that the storage in the long-term memory is the prototype, the inside attribute of one kind of objects, referring to abstractive characteristics of all individuals in one certain type or category. Therefore, the sense to objects of human comes from the matching between input information and prototype [2]. Theory of cognitive pattern recognition based on prototype matching is discussed as following:

2.1. Perceptive Constancy

Size constancy and shape constancy are two main types of the perception constancy in cognitive psychology, making the perception of an object invariable from different distances and angles. Therefore, under the sufficient condition to realize patterns, the geometry changing in the size and shape does not affect people's perception for the object.

The constancy relies on a character of the geometric structure named topologic property, which means an object in elastic deformation never changes its inner property. Therefore, the size constancy and shape constancy of perception can be interpreted with topology invariance, that is, people can perceive the objects in topological transformation as the same object.

2.2. Composition Of Patterns

Biederman's theory holds that the limited components have almost infinite combination, thus compose almost countless objects. It is to say, that all the objects of the world can be decomposed into limited components, or that all the objects are composed by a few components according to certain combinations of them.

Definition 1. The objects which have the same characteristics in some aspects compose a set $O$.

Definition 2. The elements in one object set $O$ are called object or pattern $m$, $m \in O$.

Definition 3. Any $m \in O$ can be decomposed into one or several components $C_i$ ($i$ is the natural number) according to a certain regulations.

Definition 4. For any $m \in O$, the mutual relation of the components is called structure $S$ or $S(C_1, C_2, \ldots, C_n)$.

Therefore, every pattern can be denoted as:

$$M=S(C_1, C_2, \ldots, C_n)$$

where $S$ is the compound mode of the components $C_1$, $C_2$, $\ldots$, $C_n$, also called structure.

2.3. The Relationship Between Prototype And The Components Of Objects

The theory of prototype believes that the storage in the long-term memory is the prototype rather than the templates corresponding to the external pattern. Prototype, rather than inner copy of a specific model, is a class of inner characterization of the object and a class of all individual’s general characterization of the objects.

Prototype, the general characterization of all individuals in a category or area, reflects a class of the objects’ basic characteristics, as the high degree of general and abstract of the object’s characteristics. The object is constituted by some elements that are concrete and determined under its structure. Therefore, there exists a mapping,
from general to determine and from abstract to concrete, between the prototypes and the components of the objects.

**Definition 5.** The set of components C which abstracted from all components C of all objects M in one kind O is named the prototype of O, denoted as P.

The relation between the prototype and component can be described as:

Suppose \( M=\{m_1, m_2, \ldots, m_n\} \) is the set of all the objects in O, each element in it denotes an object; \( C=\{c_{i1}, c_{i2}, \ldots, c_{in}\} \) denotes the components of the \( i \)th object in M. \( C=\{C_1, C_2, \ldots, C_n\} \) denotes the collection of all components sets; \( P=\{p_1, p_2, \ldots, p_n\} \) denotes the set of all the prototypes which is abstracted from the component collection \( C \) of all the objects \( M \) in \( O \), that is \( P=\bigcap_{i=1}^{n} C_i \) represents the set of all prototypes in \( M \).

Suppose \( \Gamma_P \) is the power set of the prototypes \( P \),

\( \Gamma_{C_i} \) is the power set of \( C_i \), \( \Gamma_M \) is the power set of \( M \).

Then:

**Proposition 1.** \( \Gamma_P \) is a topology of the prototype set \( P \),

\( (P, \Gamma_P) \) constitutes a discrete topological space.

**Proposition 2.** The collection \( C=\{C_1, C_2, \ldots, C_n\} \) is a basis of the prototypes’ topological space \( (P, \Gamma_P) \).

So any prototypes in the prototype set \( P \) are abstracted from the component collection \( C \) of all the objects \( M \) in \( O \).

**Proposition 3.** If \( f \) is the mapping from the topological space of the components \( C \) to the topological space of the prototypes \( P \), then \( f \) is surjective.

\[
M=\mathcal{S}(C)=\mathcal{S}(f^{-1}: P \rightarrow C)
\] (2)

Therefore the mathematical model, which expresses the course that the components compose the objects, is shown in Fig.2.

### 3. THE SYSTEMIC CONSTRUCTION OF COGNITIVE PATTERN RECOGNITION

The framework of cognitive pattern recognition is shown in Fig.3 [11]. The process of pattern collecting is the same as that in traditional pattern recognition. The functions of the pattern analysis processing are stronger than that of traditional pattern recognition, analyzing the simulating signal collecting from the real world, including feature extraction, prototype analysis, topological judgment, prototypes combination description, background description and so on, preparing for further works including knowledge searching, character/prototype searching, and matching decision-making.

The database of knowledge, together with the database of prototype, corresponds to the long-term memory of human beings. The database of prototype stores the prototypes of the external objects, which has extended the ability and is able to append new prototype into them. The database of the knowledge stores transcendent knowledge, such as rules of prototype combining, feature relation knowledge and so on. The section of
matching decision-making combines the result of pattern analysis, firstly searching the prototype and the knowledge from the corresponding database, then matching them and evaluating the result with some rules, and finally outputting the results of the recognition. If the matching is fail, the system will add those new items to their databases correspondingly. The system therefore can learn and memorize new things.

4. APPLICATION OF COGNITIVE PATTERN RECOGNITION TO CHINESE CHARACTER INTELLIGENT FORMATION

So far, Chinese and international organizations have developed Chinese character information processing system using the technology named Chinese character library, which has some drawbacks such as unable to establish stable standards, not passing down the civilization of Chinese character well, not conforming to the mechanism of human cognition of Chinese character, out of touch with education and unable to meet the demand for application in various fields. As the only ideograph language existing in the world so far, the Chinese character is a combination of either single hieroglyphic or self-explanatory symbol, or several of them basing on meaning and pictophnetic rules. The hieroglyphic and self-explanatory symbols are the Chinese character prototype, and are basic elements of Chinese character (basic units). The components of Chinese characters are the topological mapping of basic units in the Chinese character structure. The position, size and shape of the basic units may be different at different Chinese characters, but the basic units are topo-

Fig.4 The input and display process of Chinese character “历”

4.1. Experiments of Chinese Character Intelligent Formation Without Character Library

Based on the above analysis, a platform of Chinese character intelligent formation is developed as shown in Fig.4.

The Chinese Character formation experiment includes inputting the structures and basic units of Chinese characters in the platform, then the system constructs the Chinese characters and displays them, such as inputting the Chinese character “历”, its procedure of inputting and displaying is shown in Fig.5. According to all 70244 Chinese characters formation experiments in GB18030-2005, it proves that all Chinese characters can be constructed correctly, and the theory of Chinese character intelligent formation is validity.

5. CONCLUSION AND FUTURE WORK

The theories of information processing in cognitive
psychology and pattern recognition are combined in this paper. Taking computer as the human brain, the basic theory and the system structure of the cognitive pattern recognition are proposed. Limited in space, some corresponding theorems were not given the proof. We propose that the procedure of writing and reading Chinese character is just a process of cognition and recognition, basing on which the system of Chinese Character Intelligent Formation without Character Library is proposed. 70244 Chinese characters in GB18030-2005 are successfully constructed according to the proposed theory.

Realizing the psychological process with computer may be a new research topic in pattern recognition, though we just make a courageous and initial try.

6. REFERENCE