IAPR Workshop on CV - Special Hardware and Industrial Applications OCT.12-14, 1988, Tokyo

ENTROPY-REDUCED TRANSFORMATION APPROACH TO PATTERN RECOGNITION OF COMPLEX DATA SET

Y. Y. Tang, Y. Z. Qu and C. Y. Suen

Department of Computer Science Concordia University
1455 De Maisonneuve Blvd. West
Montreal, Quebec H3G 1M8, Canada

ABSTRACT

The term of complex data set indicates that the target data set to be recognized consists of multiple categories of pattern samples. Recognition of complex data set is a challenging research topic in computer vision and pattern recognition. The entropy reduction approach has been widely used to solve the problem of recognition of single category data set. In this paper we generalize this concept in terms of Entropy-Reduced Transformation (ERT) which contains several important properties which enable us to produce the concrete solution for the practical applications. Validation of the generalized approach is demonstrated by an example.

Index Terms : Pattern recognition, complex data set, large set, entropy-reduced, transformation, multiple categories text.

I. INTRODUCTION

The term of complex data set indicates that the target data set to be recognized consists of multiple categories of pattern samples. Generally a complex data set is also a very large data set because it is not only a combination of multiple categories of data but also each category of data may have many kinds of distortions such as size distortion, rotation distortion etc. which will result in an increasing the number of pattern samples. Some typical examples are as shown in Figs. 1 and 2. In Fig. 1 three categories of data exist in the same text: Chinese characters, Roman letters and numerals. Also for each category of data there are several kinds of fonts with size distortion and rotation distortion. Fig. 2 gives some examples of rotation distortion which is often found in blueprints and typewritten documents. Recognition of such complex data sets is a very challenging research topic because it can not be treated as a simple combination of the existing methods appropriate to each single category subset of the complex data set. If a subset of data inherits the characteristics of multiple categories of data simultaneously, it will make all the existing methods which are single category-oriented out of functions. The more severe problem is that the total number of pattern samples of a complex data set is so large that the costs of memory and time for recognition are unacceptable for practical applications.

In this paper, we propose an efficient approach to handle the problem of recognizing a complex data set: Entropy-Reduced Transformation (ERT). Although the concept of entropy reduction has been widely used in pattern recognition as a suitable criterion for the feature selection and the design of tree-like classifiers, the existing methods are often restricted to single category data set [7,8]. Section II is devoted to generalize the concept of entropy reduction as a generic transformation which contains two important properties of such a transformation. These properties enable us to find appropriate functions to reduce the entropy (i.e. the uncertainty) of a complex data set step by step (i.e. find the algorithm to simplify the problem). A concrete simulation example is presented in Section III to illustrate how to set up the entropy-reduced transformation to solve the problem efficiently.

II. ENTROPY-REDUCED TRANSFORMATION

First let us define the notations used in this paper. For a given data set $W = \left\{ w_i \mid i=1,2,\ldots,n \right\}$, we can define an entropy space $\Omega = (W, P_W, H_W)$, where P_W is the priori probability defined on W such that

$$(i) \quad P_W(w_i) = p_i,$$
$$(ii) \quad 0 < p_i < 1, \tag{1}$$
$$(iii) \quad \sum_{i=1}^{n} p_i = 1.$$

H_W is the uncertainty measure defined on W according to Shannon's entropy theory [4]

$$H_W = H(p_1, p_2, \cdots p_n) = -\sum_{i=1}^{n} p_i \log p_i . \tag{2}$$

A complex data set can be represented as

$$U = \bigcup_{i=1}^{m} W^i. \qquad (3)$$

In its entropy space P_U is determined by the relationships among the subsets W^i 's (i=1,2,...,m), and the uncertainty of U will have an upperbound $H_U \le \sum_{i=1}^{m} H_{W^i}$.

The concept of entropy-reduced transformation is defined as follows.

Definition:

Let $\Omega_i = (W^i, P_{W^i}, H_{W^i})$, $\Omega_j = (W^j, P_{W^j}, H_{W^j})$, $i \neq j$. $|W^i|$ and $|W^j|$ stand for the numbers of elements (i.e. pattern samples) contained in W^i and W^j respectively. Assume that $|W^i| < |W^j|$. If a function F_{ji} can be defined such that

$$(i) \quad |F_{ji}(W^j)| = |W^i|,$$
$$(ii) \quad H_{F_{ji}(W^j)} = H_{W^i}, \qquad (4)$$

then F_{ji} is called an entropy-reduced transformation. We say that Ω_j is normalized to Ω_i by F_{ji}.

Two important properties of the entropy-reduced transformation are given by the following theorems:

Theorem 1 :

Given $\Omega_i = (W^i, P_{W^i}, H_{W^i})$, $\Omega_k = (W^k, P_{W^k}, H_{W^k})$, and $\Omega_j = (W^j, P_{W^j}, H_{W^j})$. If F_{ik} normalizes Ω_i to Ω_k, and F_{kj} normalizes Ω_k to Ω_j, such that

$$(i) \quad \sum_{w_l \in F^t_{ik}(W^i)} p_l = p_t, \quad t = 1,2,\dots, |W^k|,$$
$$(ii) \quad \sum_{w_t \in F^r_{kj}(W^k)} p_t = p_r, \quad r = 1,2,\dots, |W^j| \qquad (5)$$

where $F^t_{ik}(W^i)$ stands for the t-th cluster of $F_{ik}(W^i)$, $F^r_{kj}(W^k)$ stands for the r-th cluster of $F_{kj}(W^k)$, then we have

$$F_{ij}(W^i) = F_{kj}(F_{ik}(W^i)). \qquad (6)$$

Theorem 2 :

Given $U = \bigcup_{i=1}^{m} W^i$ and $\left\{ \Omega_i = (W^i, P_{W^i}, H_{W^i}) \mid i=1,2,\dots,m \right\}$ is known , and $|W^j| = \min(|W^i|)$, where i,j = 1, 2, ..., m. If for all Ω_i we can find an entropy-reduced transformation which normalizes Ω_i to Ω_j , then we can find an entropy-reduced transformation F which normalizes Ω_U to Ω_j .

We can call Theorem 1 the cascade principle and Theorem 2 the parallel principle. They bring to light two basic ways to reduce the uncertainty of a complex data set. Obviously, if one of them is used together with the other one, more complicated problems can be solved. Although these theorems impose some crucial principle to design an efficient discriminator to recognize a complex data set, details are described in [1,2,3], here we would like to stress another important consequence implied by these theorems as follows.

Theorem 3 :

Let $U = \bigcup_{i=1}^{m} W^i$. D is a well-designed discriminator for $\Omega_j = (W^j, P_{W^j}, H_{W^j})$, where $W^j \subset U$ and $|W^j| = \min(|W^i|)$, i, j = 1, 2, ..., m. If Ω_U can be normalized to Ω_j by entropy-reduced transformation according to either Theorem 1 or Theorem 2 or both of them, then Ω_U can be discriminated by D without any quality loss.

Details of the proofs of the theorems above are presented in [1,2].

III. AN EXAMPLE OF APPLICATION

To demonstrate the application of the entropy-reduced transformation approach to recognize complex data sets, we present an illustrative example. Our target data set to be recognized includes the two categories of subsets W^1 and W^2 . W^1 is a set of 3200 Chinese characters, W^2 is a set of 94 keyboard symbols (including 26 uppercase and 26 lowercase English letters, and 10 numerals). Due to the necessity to process real life samples like those shown in Figs. 1 and 2, we allow all Chinese characters, English letters, and numerals to have 10 different sizes and plus a rotation of x degrees, where x = 1°, 2°, ..., 360°. If we treat each kind of distortion as a pattern sample, this complex data set will be a huge data set totalling more than ten million pattern samples ((3200+94)×10×360 = 11,858,400). The uncertainty of this complex data set comes up to about 24 bits if all the pattern samples with even priori probability are counted. This will make almost all the existing methods suitable for single category data set either out of function or very unefficient. To deal with this situation the entropy-reduced transformation approach is proposed. Our purpose is to reduce the uncertainty of this complex data set to that of a much smaller data set which contains only 3200+94 samples with a standard size (one of 10 kinds of sizes is selected as


ABD 技术及其在最优部分匹配检索算法设计中的应用

ABD TECHNIQUE AND ITS APPLICATION TO THE OPTIMAL PARTIAL-MATCH RETRIEVAL ALGORITHM DESIGN

一、引 言

一个直接执行向量 ALGOL60 语言的计算机

ARCHITECTURE OF A DIRECT-EXECUTING VECTOR-ALGOL 60 COMPUTER

Fig. 1

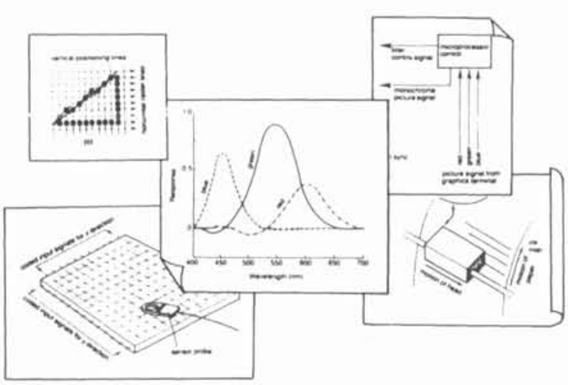

Fig. 2

the standard size) and without any rotation. First we think the complex data set consisting of 10 distinct subsets. Each subset contains all the samples with the same size (i.e. cluster all the samples into 10 groups according their sizes). Based on the parallel principle of Theorem 2, we define a linear transformation for each subset except the one with a standard size as follows

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} D/d_j & 0 \\ 0 & D/d_j \end{bmatrix} \begin{bmatrix} x_j \\ y_j \end{bmatrix} \qquad (7)$$

where (X, Y) are the new coordinates of a point for a pattern sample with the standard size, (x_j, y_j) are the coordinates of a point for a pattern sample with the j-th kind of size, D is the standard size, d_j stands for the j-th kind of size. If we denote the standard size the 1-st kind of size out of ten, then j = 2,3,...,10. After doing this the uncertainty of the original complex data set has been reduced to the level of that of a set having (3200+94)×360 possible pattern samples. Then a rotation-invariant transformation defined in [3, 5] is applied to cluster all 360 kinds of rotations of a pattern sample into a unique reference pattern sample which belongs to a set W consisting of W^1 and W^2 with the standard size. Then according to the cascade principle of Theorem 1, we know that the uncertainty of the complex data set now has been reduced to the level of that of W, which is about 11 bits. After all these operations, we can now apply the tree-like discriminator [6, 8] which is only suitable for the recognition of a data set in which all the pattern samples have the standard size and without rotation like the set of W. Based on Theorem 3, we can use this restricted discriminator to recognize the complex data set described above. This approach has been tried on a CYBER-835 computer and the experimental results, shown in the Table 1, have completely supported our theoretical predictions (more experimental results are provided in [1,2]).

Table 1 Recognition Results

No. classes		3294*	3294**
I	Error Rate	1.5%	1.4%
	Recognition Rate	98.5%	98.6%
	Speed	945/sec.	958/sec.
II	Error rate in step 1	0.09375%	0.09421%
	Error Rate in step 4	0.01884%	0.01875%
	Error Rate	0.11305%	0.1125%
	Rejection Rate	0.00625%	0.00813%
	Recognition Rate	99.87%	99.88%
	Speed	868/sec.	873/sec.

I. Conventional Search

II. PLS - Search

* For complex text

** For single category

Table 1

IV. CONCLUSIONS

Based on a detailed analysis of Section II and the demonstration of Section III, we can conclude that the entropy-reduced transformation approach can provide an efficient solution for the recognition of complex data set. It contributes by finding and organizing a set of entropy-reduced methods to reduce the entropy of the entire complex data set. This approach also expresses that similar as other information processing system, entropy-reduction is a main principle for every stage of pattern recognition. Based on this principle we may develop many new efficient methods to tackle more complicated problems.

V. REFERENCES

[1] Y. Y. Tang, " Multiple-level Information Source and Entropy-Reduced Transformation Models and Their Application in Pattern Recognition," Ph.D Thesis, Dept. of Computer Science, Concordia University, 1988.

[2] Y. Z. Qu, Y. Y. Tang, and C. Y. Suen, "Entropy-Reduced Transformation (I) : Theoretical Analysis and Application in Pattern Recognition," to appear.

[3] Y. Z. Qu, Y. Y. Tang, and C. Y. Suen, "A New Method for the Design of Integrated Discriminator for Document Recognition," to appear.

[4] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.,* Vol.27, pp. 623-656, 1948.

[5] Y. Y. Tang, Y. Z. Qu, and C. Y. Suen, "Analysis and Design of an Integreted Discriminator Based on Entropy Reduction and its Application to Recognition of Complex Data Set," to appear.

[6] Y. Y. Tang, C. Y. Suen and Q. R. Wang, "Chinese character Classification by Glodarry Trained Tree Classifier and Fourier Descriptors of Condensed Patterns," *Proceedings The IEEE 1-st Int. Conf. on Computers and Applications,* pp. 215-220, Beijing, 1984.

[7] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles,* Addison-Wesley Publishing Company, 1974.

[8] Q. R. Wang and C. Y. Suen, "Analysis and Design of a Decision Tree Based on Entropy Reduction and its Application to Large Character Set Recognition," *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. PAMI-6, No. 4, pp. 406-417, 1984.