

# 要 旨

## 詳細識別を行う組合せ論理における 閾値記述法

傍士 智史

文字認識の重要な課題として類似文字識別があり、その一解決手法として組合せ論理を用いた識別がある。その中の一手法として組合せ論理のサブカテゴリー化がある。しかし、この手法には 1 特徴を用いて識別を行っているため、誤読が発生しやすいという問題があった。そこで 2 特徴を用いた組合せ論理の閾値記述法を提案し、誤読の発生を抑制できることを示す。

2 特徴を用いた組合せ論理の閾値記述法は次の通りである。級内分散級間分散分散比最大基準が最も大きい特徴と他の特徴との全ての組合せに対して、カテゴリー  $C_1$ ,  $C_2$  の分布の上限値, 下限値を用いて識別論理をそれぞれ作成する。この識別論理により 2 特徴におけるカテゴリー  $C_1$ ,  $C_2$  の分布を表すことができる。作成した全ての識別論理に対して学習パターンを入力し、2 カテゴリーの特徴分布が重複する領域に存在するパターン集合をそれぞれ作成する。前記パターン集合の中で最もパターン数が少なくなる識別論理を用いてサブカテゴリーを作成する。また、そのパターン集合を次サブカテゴリー作成時の学習パターンとする。級内分散級間分散分散比最大基準が上記でを使用した特徴の次に大きい特徴を用い、2 カテゴリーに属するパターン集合が無くなるか全特徴を使用するまで上記の処理を繰り返すことにより逐次的にサブカテゴリーを作成する。

類似文字対「鳥, 鳥」, 「東, 東」などを含む 10 組を対象とし、実験を行った。このとき、学習パターンにより作成されたサブカテゴリー数は平均 7.6 個であった。評価パターンに対する識別結果は正読率 77.9%, 誤読率 21.1%, リジェクト率 1.0%となった。これは、1 特徴

を用いた組合せ論理のときに比べて正読率は 2.9%の増加, 誤読率は 3.9%の減少となっている。また, 1 特徴を用いた組合せ論理のサブカテゴリー数は平均 31.8 個であるので, 2 特徴を用いた組合せ論理ではより少ないサブカテゴリー数で識別が可能であるという結果も得られた。これらの実験結果により, 2 特徴を用いた組合せ論理の有効性が確認できた。

キーワード 文字認識, 類似文字, 詳細識別, サブカテゴリー, 組合せ論理, 2 特徴を用いた  
組合せ論理

# Abstract

## Threshold description method of combinatorial logic on precisely classification

Satoshi Hoji

Similar shaped character recognition is an important problem in the field of character recognition. There is a method using combinatorial logic as a solution method of the problem. Moreover, there is combinatorial logic composed by subcategory as one of method of combinatorial logic. The method uses one feature, therefore, there is a problem that error is caused easily. To solve the problem, we propose combinatorial logic using two features.

The making procedure of threshold description method of combinatorial logic using two feature is as follows. First, the discriminant logic is created by upper bound value and lower bound value of feature distribution for the feature whose Fisher's criterion becomes maximum and all other features. This discriminant logic can represent feature distribution of category  $C_1$  and  $C_2$  in two features. Second, learning pattern is input to all created discriminant logic, and the pattern set is formed by learning pattern in area where feature distribution of two categories overlaps. The subcategory is created by discriminant logic minimizing the number of patterns in all formed pattern set. The pattern set is used as learning pattern for making next subcategory. Next, the above mentioned process is repeated until pattern set that belongs to two categories disappears or all features are used completely. As a result, a subcategory is created one after another.

On the experiment using similar shaped character pair including “ 鳥, 烏 ” and “ 東, 束 ”, 7.6 subcategories on the average per character pair were generated. Correct rate, error rate and reject rate became 77.9%, 21.1% and 1.0%, respectively. Correct rate increase 2.9%, and error rate decrease 3.9% compared with combinatorial logic using one feature, respectively. In addition, generated number of subcategories decrease, too. These results shown the effectiveness of combinatorial logic using two features.

**key words** Character recognition, Similar shaped character, Precisely classification, Subcategory, Combinatorial logic, Combinatorial logic using two feature