

要 旨

組合わせ論理のサブカテゴリー化による

詳細識別

田中 宏之

文字認識の重要な課題として、類似文字識別がある。本論文では、組み合わせ論理のサブカテゴリー化を用いて類似文字の詳細識別を行う手法を提案する。組合わせ論理とは、自カテゴリーの特徴分布を記述する特徴を多数集めることにより形成している。しかし、この方法ではリジェクトが多数発生してしまう。そこで、1 個の組合わせ論理では識別が困難であると判断し、組合わせ論理のサブカテゴリー化を検討した。

サブカテゴリーの作成手順は、次の通りである。まず、学習パターンを用いて級内分散・級間分散比最大基準の最も大きい特徴の分布を作成し、対抗カテゴリーの排他的分布領域の境界を閾値として設定する。これを用いて 1 個の識別論理を作成し、1 段目のサブカテゴリーとする。次に、前記識別論理を用いた識別を行い、リジェクトとなる学習パターン集合を作成する。当該集合において、級内分散・級間分散比最大基準の最も大きい特徴を選択し、前記の手順で閾値設定を行い、2 段目の組合わせ論理のサブカテゴリーを作成する。この操作を繰り返すことにより、サブカテゴリーを多段に渡って作成していく。

類似文字対 (東, 束), (問, 間) を含む 10 組に対する実験を行った。このとき、作成されたサブカテゴリー数は平均 21 個であった。評価パターンの正読率は 83.5%, 誤読率は 16.5%であり、従来の組合わせ論理で発生していたリジェクトパターンを全て識別できた。これにより、組合わせ論理のサブカテゴリー化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, 組合わせ論理, サブカテゴリー

Abstract

Precisely classification using combinatorial logic composed by subcategory

Hiroyuki Tanaka

Similar shaped character pattern recognition is an important problem in the field of character recognition. We propose precisely classification using combinatorial logic composed by subcategory. The combinatorial logic consists of a pile of the discrimination logic. However, the combinatorial logic yields a lot of reject. To solve the problem, combinatorial logic should be composed by plural subcategories because discrimination using only one combinatorial logic is very difficult.

The making procedure of the subcategory is as follows. First, the feature distribution whose Fisher's criterion becomes maximum is selected using the learning pattern. Second, the boundary which decides the exclusive distribution area of the rival category is set as a threshold of combinatorial logic. As a result, a combinatorial logic of first stage is made. Third, reject patterns are created by discrimination using the first stage combinatorial logic. Feature is selected and threshold is decided by the method described above, then the second subcategory of the combinatorial logic is created. In this way, the subcategory breeds itself into multi stage structure by repeating this operation.

On the experiment using similar shaped character pair including (東, 束) and (問, 問), 21 subcategories on the average per character pair were generated. The combinatorial logic with subcategory achieved the recognition rates of 83.5 % for test pattern and discriminated all the reject patterns generated by the traditional combinatorial logic.

These results draw the effectiveness of combinatorial logic with subcategory.

key words Character recognition, Similar shaped character, Precisely classification,
Combinatorial logic, Subcategory