

# 要 旨

## 距離に関する複数の情報を用いた大分類の候補削減法

久保 光弘

本論文では大分類の候補数削減を目的として、距離比と距離差の選択的使用法を提案する。距離比を用いる方法では、大分類における第 1 位候補の距離値  $d_f$  に距離比  $r_f$  を乗じた値  $d_f r_f$  の範囲内に存在するカテゴリを大分類結果として出力し、距離差を使用する場合は第 1 位候補の距離値  $d_f$  に距離差  $\Delta d_f$  を加算した値  $d_f + \Delta d_f$  の範囲内に存在するカテゴリを大分類結果として出力する。距離比、距離差は、(1) 学習パターンを大分類にかけて 1 位候補のとり距離に対応する区間に入力パターンを振り分ける、(2) 各区間に割り当てられたすべてのパターンの距離比と距離差を算出する、(3) 当該区間で最も大きい距離比を  $r_f$  として、最も大きい距離差を  $\Delta d_f$  として設定する、という手順で求められる。両者の選択的使用法として、2 種の方法を提案する。方法 1 は、各区間毎に距離比と距離差の選択がトップダウン的に与えられているものである。方法 2 は、各区間に 2 つの基準値を設定し、その基準値と第 1 位候補の距離との大小関係によって距離比と距離差の選択的に使用するものである。JIS 第 1 水準の漢字を含む 3036 字種に適用した結果、従来の大分類候補数 64 個を方法 1 では 83.0 %、方法 2 では 81.0 %削減することができた。また、分類率は、距離比を用いたときと比較して方法 1 では 0.13 %低下したが、方法 2 では距離比を用いたときと比較してほぼ同じ値を保っている。これにより、分類率に着目せずに候補を多く削減するためには方法 1 が、分類率を保ちつつ候補を削減するためには方法 2 が有効であると考えられる。

キーワード 文字認識，大分類，候補数削減，距離比，距離差

# Abstract

## Candidate reduction method using plural information concerning distance in preliminary classification

Mitsuhiro Kubo

This paper proposes candidate reduction method using plural information concerning distance in preliminary classification . The method using the ratio of distance  $r_f$  , output the candidates included within distance  $r_f d_f$  , where  $d_f$  is the first candidate's distance , as preliminary classification results. The method using the difference of distance  $\Delta d$  , output the candidates included within distance  $d_f + \Delta d$  as preliminary classification results. Concrete processing of setting  $r_f$  and  $\Delta d$  are as follows. (1)Allocate each learning sample to the division section decided corresponding to the  $d_f$  of a learning sample. (2)Calculate  $r_f$  and  $\Delta d$  for each learning sample in each division section. (3)Set the largest  $r_f$  and  $\Delta d$  in the learning sample as  $r_f$  and  $\Delta d$ , respectively, in the division section. This paper proposes two methods as a selection method. The method 1 specifies the method which should use in the top-down from  $r_f$  and  $\Delta d$  for each division section where  $d_f$  is located . Method 2 sets two standard values every each division section and chooses  $r_f$  or  $\Delta d$  according to the relationship between  $d_f$  and two standard values. On the experiments using ETL9B(3,036 characters  $\times$  200 samples) under the condition that the number of section division is 5, the method 1 and method 2 reduced the number of candidate by 83.0 % and 81.0 %, respectively, in comparison with the traditional method . Classification ratio of method 2 for test data keeps the same value as that of the method which uses only  $r_f$  , while classification ratio of method

1 for test data has decreased 0.13 % in comparison with the method which uses only  $r_f$ . Method 2 is effective to reduce candidates while keeping classification rate, though method 1 is effective to reduce candidates disregarding the decrease of classification rate.

**key words** Character recognition , Preliminary classification , Candidate reduction , Ratio of distance , Difference of distance