

要 旨

大分類における距離による 候補数削減の閾値設定

山元 祐太

本論文では大分類の候補数削減を目的として区間分割型距離比設定法を提案する．区間分割型距離比を用いた大分類では，大分類における 1 位候補の距離に対応する区間の距離比 r_f をカテゴリ - 距離比テーブルから読み出し，1 位候補の距離 d_f に距離比 r_f を乗じた距離 $d_f r_f$ の範囲内に存在するカテゴリを大分類結果として出力する．距離比 r_f は正解カテゴリを含むよう大量の学習パターンを統計処理して設定されている．区間を細分化して距離比を設定すれば候補を十分削減することが期待できる．閾値設定法は，(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各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を設定する，(3) 当該区間で最も大きい距離比を r_f として設定する，という手順をとる．JIS 第 1 水準の漢字を含む 3,036 字種に適用した結果，従来の大分類法の候補数 64 個を 84.1 %削減することができた．また，分類率の低下も 0.26 %と軽微であった．これより，本手法の有効性が確認された．

キーワード 文字認識，大分類，候補数削減，距離比，区間分割

Abstract

Threshold setting-up method for reduction of the number of candidates by distance in preliminary classification

Yuta Yamamoto

This paper propose setting-up method of ratio of distance using section division to reduce number of candidates in preliminary classification. Preliminary classification module using reads the ratio of distance r_f corresponding to the first candidate from the predetermined table where r_f has been experimentally decided so as to contain the correct candidate, and calculate the distance $r_f d_f$ by multiplying ratio of distance r_f by the first candidate's distance d_f . The candidates included within distance $r_f d_f$ are output as preliminary classification results. Ratio of distance is set up so that a correct category may be included by statistical processing using numerable learning patterns. If the section is subdivided and ratio of distance is set up, it is expectable to reduce candidates enough. The setting-up method takes the procedure (1)Each pattern is distributed to the section corresponding to the distance in which the 1st place of a candidate takes, covering it over preliminary classification, (2)The ratio of distance of all the patterns assigned to each section is set up, and (3)The largest ratio in the section concerned is set up as r_f . As a result, in the experiment using 3,036 different characters including JIS.1 Kanji level, the proposed method reduced the number of candidate by 84.1% in comparison with the traditional method. The decrease of classification rate is 0.26%, and is negligible. These results draw the effectiveness of the proposed method.

key words Character recognition, Preliminary classification, Reduction of number

of candidates, Ratio of distance , Section division