

要 旨

遺伝的アルゴリズムを用いた類似文字識別

中津留 真一

本論文では、遺伝的アルゴリズムを用いた類似文字識別を提案する。類似文字「王」と「玉」の識別を行うには人間は両者の差分となる「丶」の部分のストロークの有無を用いていると考えられる。差分となるストローク部分を使用して識別を行うには、類似文字のセットごとに偏やつくりを直接取り出す論理を設計者が組む必要があった。本論文では類似文字識別に必要な特徴を GA によって自動的に発見できるであろうかという問題提起を行う。GA を類似文字識別に適用するにあたり、GA の各世代における特徴次元数の削減率を適応度とすることで個体間の優劣の差を明確にする工夫を行った。類似文字「王」、「玉」に適用した結果、特徴数の平均が 92.7 個で認識率の平均が 98.3% という結果を得た。一方、分散比による特徴選択を用いる手法では特徴数 686 個で認識率 97.5% であった。これより、GA は少ない特徴数で高い認識率を上げていることがわかる。また、GA が識別に使用している特徴の可視化を行った結果、人間が見分けているのと同様に差分のストロークを用いていることがわかった。

キーワード 文字認識，類似文字識別，遺伝的アルゴリズム，特徴選択

Abstract

Similar-shaped Character Recognition using Genetic Algorithm

Shinichi Nakatsuru

This paper proposes a method of similar-shaped character recognition using Genetic Algorithm(GA). It seems that the human uses presence of “丷” in character pattern to distinguish “王” and “玉” which are similar-shaped characters. A designer had to develop algorithm which extract directly radicals each set of similar-shaped character to discriminate between similar-shaped characters using difference stroke. This paper addresses a question whether GA can find effective features for similar-shaped character recognition automatically. In application of genetic algorithm for similar-shaped character recognition, the proposed method enable rank of fitness value to scale largely because the proposed method uses fitness value as reduction rate of the number of features for each generation. As a result of experiment which were made on two similar Kanji patterns “王” and “玉”, the number of features in the proposed method are 92.7 and recognition rate is 98.3%. On the other hand the traditional method using Fisher’s criterion achieved 97.5% correct recognition rate with 686 features. GA achieved higher recognition rate than Fisher’s criterion though the number of features extracted by GA is reduced. As a result of feature visualization, GA uses difference stroke to discriminate similar Kanji patterns like human.

key words Character recognition , Similar-shaped character recognition, Genetic algorithm , Feature selection