

ベクトル量子化に基づく判別分析

1 はじめに

n 個の項目 $\{q_1, q_2, \dots, q_n\}$ のデータが与えられたとき、これが m 個のクラス $\{C_1, C_2, \dots, C_m\}$ のいずれに属するデータか判定する問題は、判別問題と呼ばれる。これまではニューラルネットワークの学習機能や、ファジシステムによるルールベースなどにより判別を行う研究がされてきた [1]。本研究では判別問題に対して与えられたデータをデジタル画像と考え、ベクトル量子化法の基本操作である LBG アルゴリズムを用いて判別を行う手法を提案する。

2 ベクトル量子化

アナログ量をデジタル値に変換することを一般に量子化 (quantization) という。特に対象量をベクトルで量子化するものをベクトル量子化と呼ぶ。ベクトル量子化は通信信号に対して用いることができるが、特にデジタル画像に適用したとき高い圧縮率が得られる。

3 ベクトル量子化に基づく判別方法

ベクトル量子化法の基本操作である LBG アルゴリズムを使って判別を行う手順を次に示す。

1. 項目 q_1, q_2, \dots, q_n に基づいて各カテゴリーを表現するパターンを作成する。 \sqrt{n} が整数になるときパターンのサイズは縦と横のサイズが $\sqrt{n} \times \sqrt{n}$ の正方形の配置を取るよう決定する。次にこのパターンのピクセルにおおのこの項目を割り当てる。ある項目の最小値を q_{min} 、最大値を q_{max} とするとき、画像中の任意の値 q_i を、式 (1) によって計算した結果である g_i に変換する処理により階調値の範囲 $[0, 255]$ に変換する。その値を先ほどのパターンに割り当てる。

$$g_i = 255 \times \frac{q_i - q_{min}}{q_{max} - q_{min}} \quad (1)$$

各カテゴリーを表現するパターンを作って並べていくことによって 1 つの大きな画像を作成する。

2. 1 で作成された画像 (原画像) をベクトル量子化法を用いて処理する。まず原画像を 1 で処理したパターンのサイズで正方形上のブロックに分割した後、各ブロックをそれぞれ多次元空間中のベクトルであるとみなす。
3. 原画像を構成する $\sqrt{n} \times \sqrt{n}$ 画素のブロックをすべて集めた後判別したいクラス数である m 個のパターンで K -平均アルゴリズムを用いてクラスタリングする。クラスタリングが終了すると全てのパターンについて m 個の代表ベクトルのパターンで置換されているので、どのパターンで置換されたかにより判別が可能となる。

4 分類問題と検証結果

ベクトル量子化を用いた判別法が有効であるか、あやめの分類問題と肝臓病の診断問題に適用して実験を行った。

あやめの分類問題は、3 種類のあやめ (setosa, versicolor, virginica) があり、花びらの幅と長さおよび、がくの幅と長さという 4 つの属性値からいずれに属するのか分類する問題である。

また肝臓病の診断問題とは血液検査のデータから、健康者 (Healthy Person) 慢性肝炎 (Chronic Hepatitis) 肝硬変 (Liver Cirrhosis) となっている病気の状態を判別する問題である。本研究では 9 つの血液検査のデータから健康者、慢性肝炎、肝硬変の 3 つのクラスを分類した。

結果を表 1, 2 に示す。

表 1: 提案手法によるあやめの認識結果

種類	問題数	認識	認識率 (%)
setosa	22	22	100
versicolor	21	21	100
virginica	21	21	100

表 2: 提案手法による肝臓病の認識結果

病名	問題数	認識数	診断率 (%)
健康者	22	22	100
慢性肝炎	21	13	61.9
肝硬変	21	20	95.2

表 1 に示すようにあやめの認識結果としては認識率が 100 % という結果が得られた、これは同様の過去の研究例と比較しても良好な結果といえる。また表 2 の肝臓病の診断結果であるが、従来法であるアソシアトロン診断率が、健康体...100 %、慢性肝炎...60 %、肝硬変...60 % であるのでこれと比較すれば良好な結果が出ているといえる。

5 まとめ

本論文ではベクトル量子化に基づく判別分析について提案した。検証実験によりその有効性を確認することができた。

参考文献

- [1] S.Abe and M.Lan, "A method for fuzzy extraction directly from numerical data and its application to pattern classification," *IEEE Trans.on Fuzzy Systems*, Vol.3, No.1 (1995)