

入学年度 平成 10 年度

学籍番号 10117910

氏名 江本 喜久男

論文題目 概念学習とベイズ学習を用いた多クラス分類手法の考察

犬塚研究室

1 はじめに

大量に蓄積されたデータから、価値のある情報を抜き出すデータマイニングにおいて、データが何に属するかを決定する問題を分類問題と呼ぶ。複数の分類候補から 1 つを決定する分類問題を特にマルチクラス分類問題と呼ぶ。過去の分類済みデータを計算機に与えて分類の規則を学習させる、機械学習の技術が利用される。

マルチクラス分類問題を二項関係とベイズ学習を用いて学習する手法が提案されている [1, 2]。本研究ではこの手法 (下図) において図中の 1, 2 の箇所の改良を検討する。

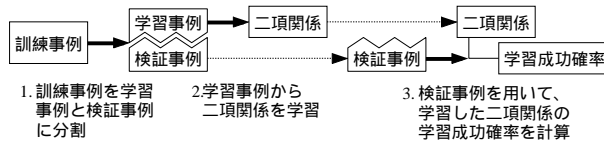


図 1: 本研究の枠組

2 二項関係による分類問題の学習

上図はマルチクラス分類問題を二項関係に分割して解決している。二項関係とは、2つのクラスを比べた時にどちらを選択するべきかという二者択一の関係である。 n クラスあるとき、 n^2 個の二項関係問題となる。 n^2 個の問題の解 R を統合してマルチクラス分類を決定する。

この決定の際、2つのクラス c, c' に対する二項関係問題が正しく決定される確率を $\psi_{cc'}$ としたとき、 n 個のクラスからの選択を次式によって行なう。

$$c_{MAP} = \operatorname{argmax}_{c \in C} \prod_{\substack{(c, c') \in R \\ \wedge c \neq c'}} \psi_{cc'} \prod_{\substack{(c, c') \in C^2 - R \\ \wedge c \neq c'}} (1 - \psi_{cc'}) \prod_{\substack{(c', c) \in C^2 - R \\ \wedge c \neq c'}} \psi_{c'c} \prod_{\substack{(c', c) \in R \\ \wedge c \neq c'}} (1 - \psi_{c'c}) \quad (1)$$

c_{MAP} が最終的に決定されるクラスであり、 $\psi_{cc'}$ を見積もるには学習事例とは別の検証事例が必要である。

3 検証事例と学習事例のトレードオフ

本手法では正確な二項関係の学習とその学習精度の正確な見積もりの両方が必要であり、検証事例と学習事例の割合にトレードオフが考えられる。高精度を得るには多くの学習事例が必要であるが、これによって検証事例が減り、精度の見積もりが困難となる。逆も同様である。

本研究では両事例の割合を変化させ、実験を行なった。

4 クラス対毎の学習

式 (1) では、二項関係の結果が正しいという事象の独立性を仮定している。しかし、従来手法 [1] ではすべての二項関係をまとめて学習し、1つの分類器で求めており、独立性が疑わしい。

そこで独立性を高めるために、クラス対毎に分類器を分割して学習する手法を提案し、分類器をまとめた手法との比較実験を行なった。

5 実験

分類器を分割しない手法と分割する手法で実験を行なった。両手法とも二項関係の学習には C4.5 を用いた。実験対象には UCI の機械学習用データベースを使用し、学習事例と検証事例の割合を 10% 毎に変化させた。

実験結果のグラフの一部を図 2 に示す。事例数の少ないデータセットでは、多くの場合グラフの形が山なり (図 2 左) になっており、トレードオフが発生したことが観察できる。事例数の多いデータセットでは、多くの場合グラフは右肩上がり (図 2 右) になった。これは、学習事例の割合を大きくしても検証事例が十分に確保されるためと思われる。全体として、分類精度が最大となる学習事例の割合は一定ではなかった。また、学習事例の割合を小さくした場合はかなり精度が低かった。

節 4 の分類器を分割する手法と分割しない手法の結果に有意な差は見られなかった。

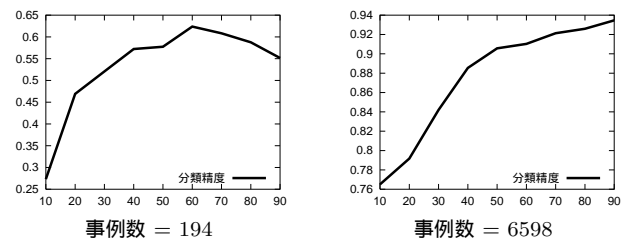


図 2: 割合の変化と分類精度の変化

横軸は学習事例の割合 (%) で残りは検証事例、縦軸は分類精度

6 まとめ

本研究では、二項関係とベイズ学習を用いた分類手法で問題となっていた検証事例と学習事例の割合におけるトレードオフを検討した。また、二項関係毎に分類器を分割する方法を提案した。事例の割合を一定にせず、条件などによって動的に設定することで高精度に分類を行なうことができる可能性がある。分類器を分割するか否かは、計算量などを考慮して決定すれば良い。

参考文献

- [1] 倉島敏史 中野智文 犬塚信博: “2 項関係の帰納とベイズ学習による多クラス分類,” 2001 年度人工知能学会全国大会 (第 15 回) 論文集 (CD).
- [2] Nakano, Inuzuka, and Itoh: “Solving selection problems using preference relation based on Bayesian learning,” ILP-2000, LNAI 1866, pp. 147-163 (2000).