平成 12 年度卒業研究		
入学年度　平成 9 年度	学籍番号　09117907	氏名　伊藤 嘉信
論文題目　述語論理を用いた確率的クラス分類における属性値欠損への対応		犬塚 研究室

1　はじめに

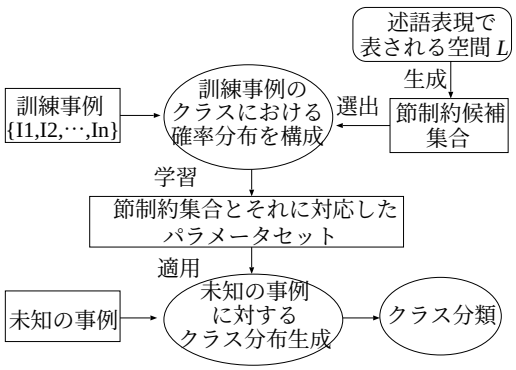
背景知識を用いて論理に基づいた制約を帰納し、これによってクラス上の確率分布を与える事でクラス分類を行なう手法 [1] が注目される。本研究ではこの方法がクラスの決定を確率的に行なうことに注目し、属性値が欠損した場合の補完について新しい方法を提案する。

2　述語表現を用いた確率的クラス分類

制約を与える事でそこから確率分布を構成する事が出来る。制約 Q_k に節形式を用い、クラス C_j と Q_k(節制約 $CC_{j,k}$) に関して次の $f_{j,k}$ を考える。$f_{j,k}$ の経験的確率分布の期待値 $\tilde{p}(f_{j,k})$ と、学習された確率分布の期待値 $p(f_{j,k})$ を等しくする節制約を見つける事で確率分布を構成する事が出来る。

$$f_{j,k}(I,C) = \begin{cases} 1 & : \text{クラス} C = C_j \text{かつ} Q_k \text{を満たす} \\ 0 & : \text{その他} \end{cases}$$

この時、事例 I に対して定義される各クラス C 上の確率



図の上段部で学習を行ない、下段部で学習結果を未知の事例に適用してクラス分類を行なう。学習されるものは節制約集合 $\mathcal{CC} = \{CC_{j_1,k_1}, \ldots, CC_{j_M,k_M}\}$ とそれに対応するパラメータセット $\Lambda = \{\lambda_1, \ldots, \lambda_M\}$ である。

図 1: クラス分類の流れ

分布は式 (1) で構成される。$\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_M\}$ の M 個の各パラメータ λ_m は、関数 f_{j_m,k_m} を基にした節制約 CC_{j_m,k_m} に対応している。λ_m は節制約の重み的役割をはたし、f_{j_m,k_m} と掛け合わせた重み和によって事例における各クラスの重みを計算、確率分布を構成する。$Z_\Lambda(I)$ は正規化関数である。

$$p_\Lambda(C|I) = \frac{1}{Z_\Lambda(I)} \exp\left(\sum_{m=1}^{M} \lambda_m f_{j_m,k_m}(I,C)\right) \qquad (1)$$

3　属性値欠損に対する提案手法

観測された事例の属性値が欠損していた場合、節制約を当てはめることができない。そこで、確率的に属性値を補完する方法を与える。

$$f_{j,k}(I,C) = \begin{cases} 1 & : \text{クラス} C = C_j \text{かつ} Q_k \text{を満たす} \\ p & : \text{クラス} C = C_j \text{かつ属性値が欠損している} \\ & \quad \text{属性を除いた} Q_k \text{を満たす} \\ 0 & : \text{その他} \end{cases}$$

属性値欠損によって Q_k を満たさなかった事例をそのまま 0 にするのではなく、訓練事例における属性ごとの取り得る属性値の各生起確率 p を用いる。これによって与えた確率から式 (1) における重み和の計算を属性値欠損に対して柔軟にする事が出来る。

4　実験

データベースとして属性値欠損がないものを用意し、属性値を欠損させる。属性値欠損率を 0% 5% 10% 20% 30%とし、実験を行なった。各% 毎に属性値欠損を 5 回作り変えて実験を行ない、各分類精度の平均をとった。確率的クラス分類の属性値欠損補完の有無、更に C4.5 で実験を行ない分類精度の比較を行なった。結果を図 2 に示す。

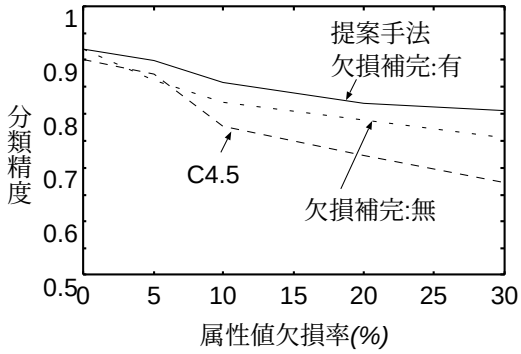


図 2：UCI のデータベース zoo に
属性値欠損を加えた場合の結果

実験のどのパーセントにおいても欠損補完手法有りの方が無しの時と比べて精度の向上が見られた。また C4.5 との分類精度の比較においても勝る事が出来、欠損確率が増すにしたがって精度が下がっていく中で C4.5 に比べ精度の減少率が緩やかであった。

5　まとめ

本研究では、統計的な量と帰納された述語論理を用いることで背景知識、帰納された述語ともに利用可能な、確率に基づいたクラス分類を行なった。そして、確率的クラス分類の点から属性値欠損に対する補完を行なった。属性値欠損に対する実験で欠損補完の有無、C4.5 との比較において本手法の有効性を示す事が出来た。

参考文献

[1] L. Dehaspe. *"Maximum entropy modeling with clausal constraints"*, In Proc. 7th Intl. Workshop on ILP, LNAI 1297, pp.109-125. Springer, 1997.