平成 14 年度　修士論文概要

述語表現された事例の確率的分類に関する研究	世木研究室	伊藤嘉信
	学籍番号	No. 13417012

1 はじめに

機械学習 (Machine Learning) は対象とする既存の事例から，知識や規則を自動的に抽出し，これを未知の対象に適用する．

分類アルゴリズムの実際的な適用，特にノイズを含むデータや，複雑なモデルを持つ問題に適用する場合，統計的アプローチは非常に重要である．一方，ルールに基づいたアプローチはデータマイニングの分野で興味深い．なぜなら，潜在的にある知識のはっきりした記述を与えるからである．帰納論理プログラミング (ILP : inductive logic programing) は，論理形式の知識を扱う学習法である．高コストな学習法であるが，近年比較的大きなデータセットにへの適用も研究されている．ILP の最近の研究課題の一つは，論理形式の枠組で，統計的情報を扱うことである．

2 対象とする問題

ILP の枠組では，対象とする世界の事例に関しての知識表現には，述語論理式を用いる．単一の関係表からしか学習出来ない命題論理学習器に比べ，複数の関係表にまたがった事例や，論理で記述された知識を扱う事が出来る．また，表現方法が自然であり人にとって理解し易い．

述語論理表現された事例から問題を解決するために学習される規則は，

$$P \leftarrow Q_1, Q_2, \ldots, Q_n \tag{1}$$

という形式の確定節 (definite clause) と呼ばれる形に制限する．これは，前件部を満たせば後件部が成り立つという意味である．

式 (1) で表されるような述語論理式を用いて分類を行ないたい．通常の ILP が扱う問題は，ルールによって真偽を決定する．しかし，3 つ以上のクラスからなる分類問題を扱う場合，これらのルール同士には矛盾が生じ得る．従って，そのルールのみで分類を行なうのでなく，ルール全体で最終的な分類を決定する．つまり，各ルールの重み付き投票のようなしくみを考える必要がある．

MACCENT[1] は，このしくみを確率的に実現するアルゴリズムである．このアルゴリズムは，確率的操作と論理表現を統合することに特徴を持つが，しかしながら，データの欠損への対応がされていない．データ欠損を扱うためには，論理表現されたルールの真偽を確率的に与える方法が必要である．本研究は，このための枠組を与える．

即ち，本論文では，欠損を含む部分観測された知識から学習を行なうため，ルールが適用できなかったものを偽とせず，確率的な真理値を推定する方法を提案する．確率的な真理値によって，データ欠損に対応した学習アルゴリズムを与える．

3 論理式の真理値の確率的推定

述語表現された事例に対して，ルールを適用する場合について考える．

$$^{\forall}\mathbf{X}[q \leftarrow r_1 \wedge r_2 \wedge \cdots \wedge r_{|R|}] \tag{2}$$

式 (2) において，前件部は $R = r_1 \wedge r_2 \wedge \cdots \wedge r_{|R|}$ であり，R は，$|R|$ 個のリテラル r_i を持つ．ここで，式 (2) の後件部に現れる変数を $\mathbf{Y} \subset \mathbf{X}$ とすると，後件部 q を基礎例 (ground instance) $q \cdot s$ とする置換 s を，式 (2) に適用した時の前件部を改めて次のように書く．

$$\exists \mathbf{X}' r'_1 \wedge r'_2 \wedge \cdots \wedge r_{|R'|} \tag{3}$$

これが真となる確率 $Pr(^{\exists}\mathbf{X}'R' = T)$ を考えたい．ここで $\mathbf{X}' = \mathbf{X} - \mathbf{Y}$，$R' = r'_1 \wedge r'_2 \wedge \cdots \wedge r'_{|R'|}$ とする．次に，全ての変数に対する置換を行ない，基礎例にすることを考える．すると，存在限定された命題の真となる事象は，各代入によって得られる命題が真となる事象の和事象となる．しかし，この和事象の確率を計算することは，実際上困難であるため，次の近似を行う．

- $Pr(^{\exists}X'\ R') \simeq \max_{\mathbf{a} \in Range(X')} Pr(R' \cdot \mathbf{a})$

ここで $Range(X)$ は，変数集合 X の定義域内の定数への全ての置換の集合を表す．リテラル R' が成り立つには，$R' \cdot \mathbf{a} = T$ となる $\mathbf{a}$ が 1 つ存在すれば良いので，真となる確率の最大を考えれば適当と言えるだろう．

基礎リテラルの積について考えるため，次の仮定を行なった．

- ルールの前件部を構成するリテラルに対して，その基礎例の真偽は独立である．

これによって，各リテラルが真となる確率の積として考えることが出来る．

$$Pr(^{\exists}\mathbf{X}'R' = T) \simeq \prod_{i=1}^{|R'|} Pr(r'_i \cdot \hat{\mathbf{a}} = T) \tag{4}$$

ここで，$\hat{\mathbf{a}}$ は，近似によって考えた確率が最大となる時の置換である．

各リテラルに関して $Pr(r'_i \cdot \hat{a} = T)$ となる確率を考える．ここで， $r'_i \cdot \hat{\mathbf{a}} = \ell_i(b_{i1}, b_{i2}, \ldots, b_{in})$ と書く．すると，

$$Pr(r'_i \cdot \hat{a} = T) = \frac{Pr(\ell_i = T)}{Pr(b_{i1}, \ldots, b_{in})} Pr(b_{i1}, \ldots, b_{in} | \ell_i = T). \tag{5}$$

ここで $Pr(b_{i1}, \ldots, b_{in} | \ell_i = T)$ は，既知のデータから推定できる．一方，$\frac{Pr(\ell_i=T)}{Pr(b_{i1},\ldots,b_{in})}$ は直接計算することが困難であるが，これが既知のリテラル（即ち，真となることが既知のリテラル）から推定する方法を与えた．その推定値を B_i とする．そして最後に次の仮定を置くことにより，述語 ℓ_i が真となった時の，各引数の定数項の起こる確率の積で表せる．

- 基礎例の解釈の下で，各引数は独立である．

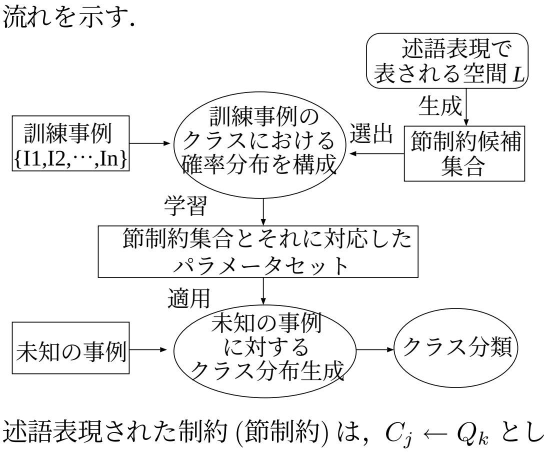
これで，リテラル r'_i ごとの 確率 $Pr(r'_i \cdot \hat{\mathbf{a}} = T)$ を求めることが出来た．よって，式 (5) のリテラル r'_i の連言を満たす確率は，次式 (6) で求まる．

$$\text{式 (5)} = \prod_{i=1}^{|R'|} B_i \prod_{j=1}^{n} Pr(b_{ij} | \ell_i = T) \tag{6}$$

以上より，真偽の分からないリテラルを含むルールを．式 (6) によって評価可能となった．

4　節制約を用いた確率的分類への応用

上記の推定法を用いて分類問題を扱う方法を述べる．学習器でまずルールを生成することが必要になってくる．そこで本研究では，制約として述語論理式を用いて確率的分類を行なっている手法 (MACCENT) を応用する．次の図は MACCENT の流れを示す．



述語表現された制約 (節制約) は，$C_j \leftarrow Q_k$ として得られる．式 (7) は，前件部 Q_k を満たし，クラスが C_j である場合に，値 1 となる．式 (8) の事後確率は，各節制約に対応した λ 値が，式 (7) によって足し合わされることによって求められる．(Z_Λ は正規化定数)

$$f_{j,k}(I, C) = \begin{cases} 1 & : \quad C_j \leftarrow Q_k \\ 0 & : \quad \text{otherwise} \end{cases} \tag{7}$$

$$p_\Lambda(C|I) = \frac{1}{Z_\Lambda(I)} \exp\left(\sum_{m=1}^{M} \lambda_m f_{j_m, k_m}(I, C) \right) \tag{8}$$

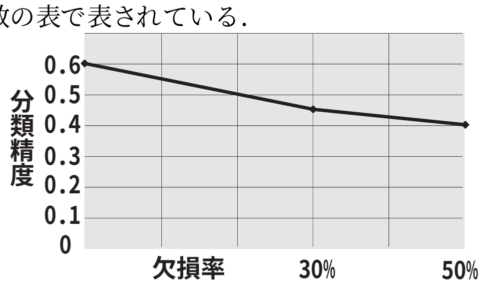
式 (7) では，節制約を用いて，0,1 の判定を行なっていた．そこで，式 (6) によって確率的に推定された真理値を用いる．

$$g_{j,k}(I, C) = \begin{cases} Pr(^{\exists}\mathbf{X} Q_k) & : \quad C = C_j \\ 0 & : \quad C \neq C_j \end{cases} \tag{9}$$

これは，クラス C_j である時に，前件部 Q_k を満たす確率を返す．この新たな式 (9) を用いて，式 (8) の事後確率を求める．これによって，欠損を含むデータに対するクラス分類法を，述語のルールによって行なうことが可能となった．

5　列車方向判定問題

列車方向判定問題 (4 クラス：東西南北) を用いて，本手法の有効性を見た．この問題は，事例が複数の表で表されている．



参考までに，MACCENT での欠損率 0%における分類精度は同程度である．

6　まとめ

述語論理によって，与えられたルールに対し真偽の判定できない真理化を，確率的に推定する方法を提案し，これを分類法に応用する手法を与えた．また，実験によって本手法の働きを確認した．確率の導出において仮定と近似をしており，これらの影響，改良について今後検討する必要がある．

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.