

述語論理を用いた確率的分類システムの学習精度改善

川田 啓史(指導教員: 世木 博久)
名古屋工業大学工学部情報工学科

研究背景・目的

分類問題

機械学習やデータマイニングの分野で研究

応用

診断、予測、パターン認識、制御、自然言語処理など多数

アプローチ

- 決定木学習、最近傍法など
- 帰納論理プログラミング(ILP)**
 - 述語論理を用いた帰納推論
 - 背景知識が使える
 - 複数の表にまたがった事例を扱える
 - 人間にとって理解しやすい表現

ILPの最近の傾向

「統計的関係学習」
論理の枠組みと確率統計の概念を統合



述語論理を用いた確率的分類システムMACCENT[Dehaspe 97]

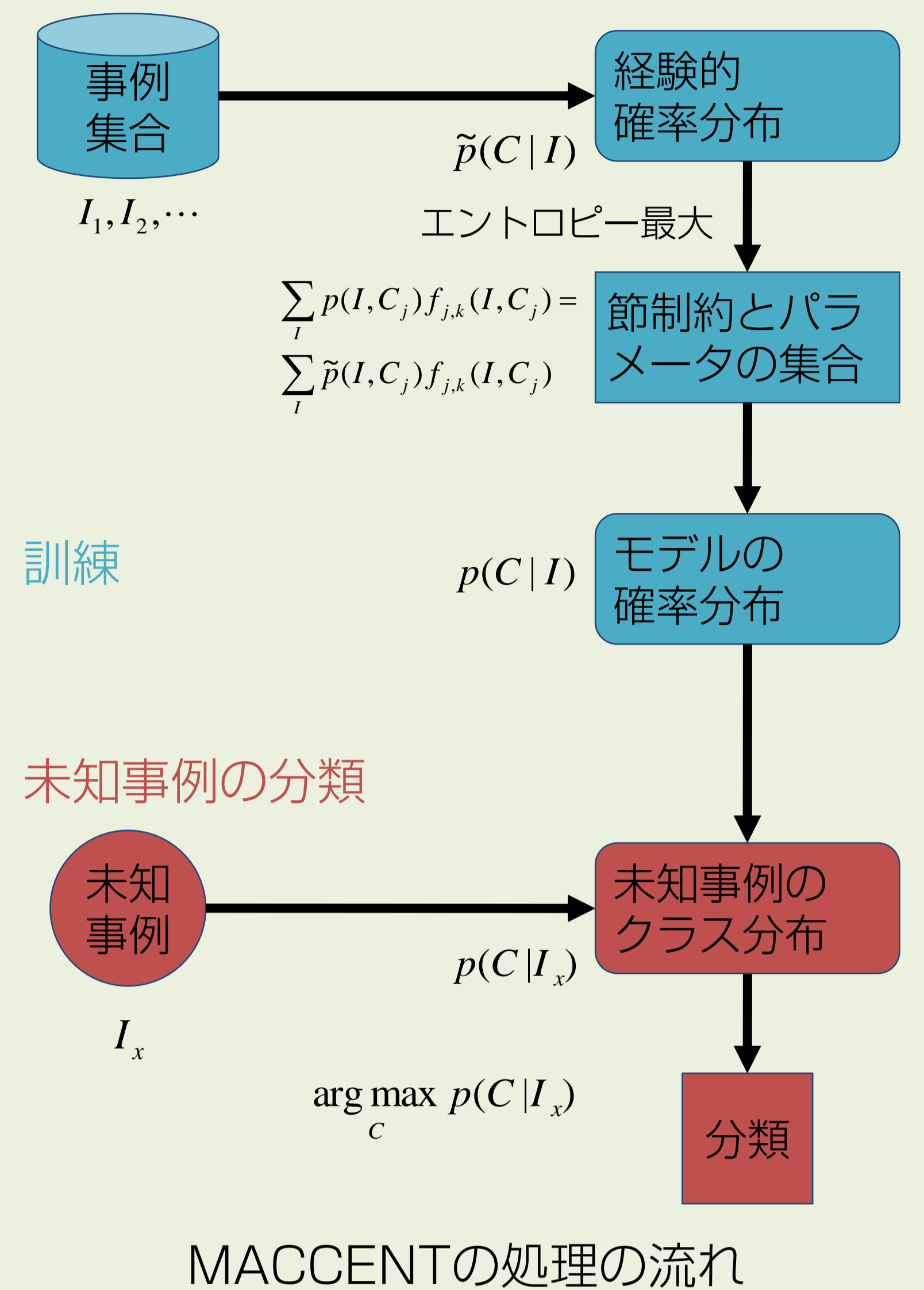
- ILP + 最大エントロピーモデル
- 最大エントロピーモデルの素性関数に述語表現を用いて制約を節制約にする

→ 事例の特徴を述語で表現

研究目的

述語論理を用いた確率的分類システムの学習精度を改善

- 方法1**: 欠損データへの対応
- 方法2**: 節制約の探索方法の改良



欠損値への対応

欠損値

データに欠損があることがある
→ 分類精度低下

対応方法

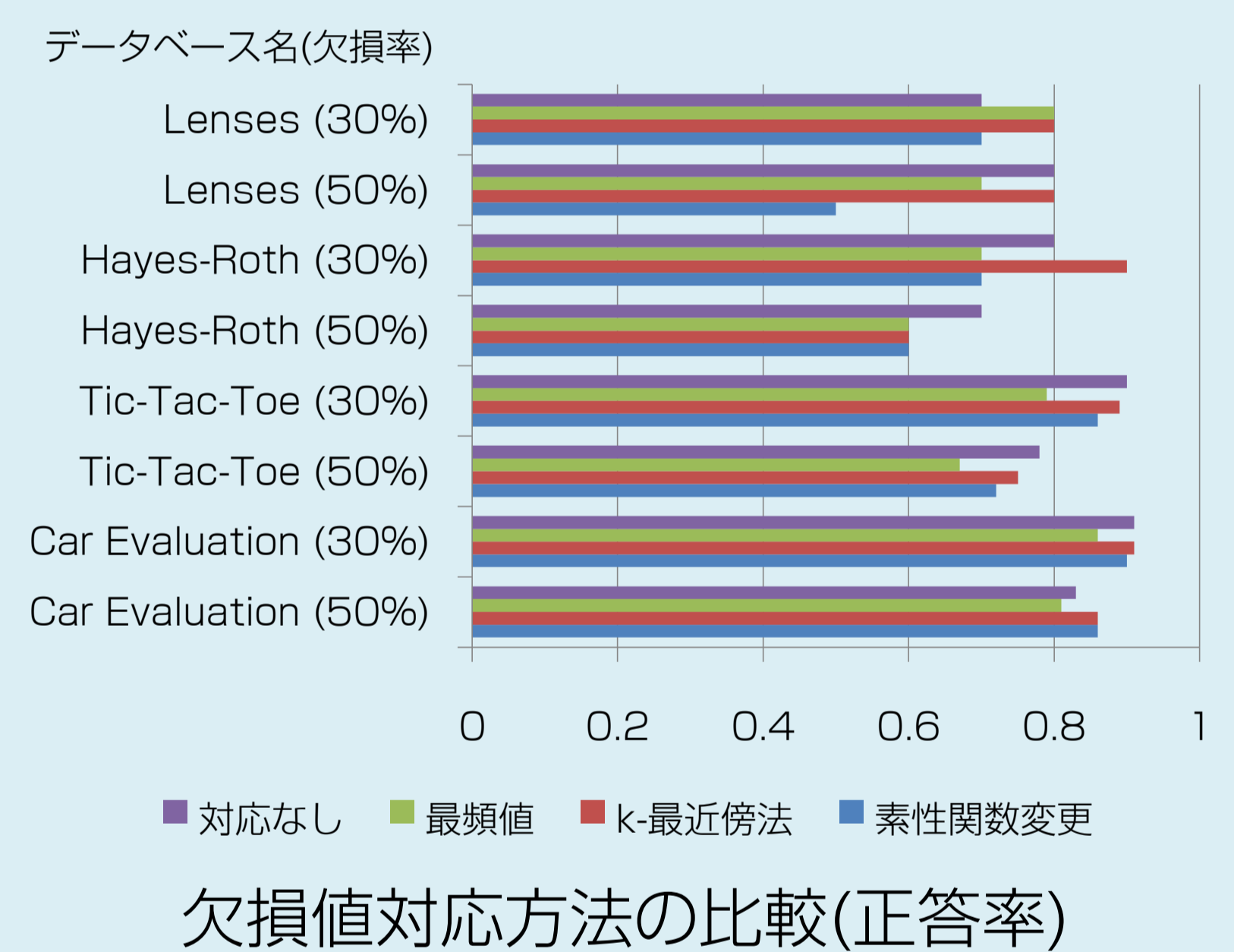
- 前処理として欠損値を補完
 - 最頻値
 - k-最近傍法
- 素性関数の値を0/1ではなく節の本体を真とする近似的確率に変更した方法[Ito et al. 04]と比較

処理時間

- 最頻値・k-最近傍法
→ 約1%の前処理時間
- 素性関数を変更した方法
→ 実行時間3~40倍

まとめ

- わずかな実行時間の増加で正答率が上がるデータがあるという結果
- データに合わせた適切な欠損値補完方法を選ぶ必要



探索方法の改善

節制約の探索

- 最適な節制約集合の求め方
対数尤度を最大にする節制約の集合を求める



対数尤度の増分(利得値)が最大の節制約を1つずつ集合に加える

- 従来法はビーム探索を行う
 - 利得値の“大きな”節
→ 節の本体にリテラルを加えた節を探索
 - 利得値の“小さな”節
→ それ以上探索しない
→ 探索されなかった節の中に重要な節があると、分類精度が低下

最短優先探索(SFS)

- 節の長さ l が小さい順に探索
- $l \leq l_{max}$ となる節をすべて探索

探索の冗長性除去

リテラルの順番が違っただけで論理的に等価な節の探索を回避



リテラル間に辞書式順序を導入し、リテラルが決められた順序になるようにする

まとめ

- SFSにより実行時間の増加を抑えつつ節制約の探索空間が拡大
- 従来法(ビーム探索)では見つけられない節制約の探索を可能にしている

