

論文の内容の要旨

論文題目 Acquisition of Probabilistic Knowledge from
Real-World Data and Its Application

(和訳 実世界データからの確率的知識の獲得とその利用)

氏名 鶴岡 慶雅

計算機が実用化されてから数十年が経過し、今まで人間にしかできなかった数多くの知的作業が計算機によって可能となりつつある。計算機による知的処理に対する近年の主要なアプローチの一つは、事例から様々な知識を学習し、その知識を利用して最適な意思決定を行なおうとする方法である。事例から知識を獲得しようとするとき、完全に誤りのない知識だけを学習しようとする、ノイズを含む事例からの学習が不可能なため、学習できる内容は非常に限られたものになってしまう。実世界データは一般に多くのノイズを含んでいるため、そのようなデータから学習を行なうためには、確率的な知識という形で学習を行なうことが不可欠である。

そのような、実世界データからの確率的知識の獲得と利用に関して、本論文では大きく分けて以下に述べる3つの研究を行なった。一つには、事例からの知識の獲得という問題に対し、確率的な知識の信頼性をベイズ学習の枠組を利用して正確に推定する手法を、自然言語処理における語義曖昧性解消の問題に適用しその効果を示した。また、教師情報が付与されていない事例も利用して学習する手法の一つとして、決定リストと確率的ラベルを用いる手法を提案し、疑似単語判定の問題を通してその効果を示した。さらに、意思決定のための確率的知識の利用法として、状態遷移に関する確率的知識を利用する効率的ゲーム木探索手法を提案し、コンピュータ将棋での実装によりその効果を示した。

事例から仮説を学習する場合、仮説の信頼度、すなわちその仮説が正しい確率を正確に推定することは非常に重要である。しかし事例の数が少ない場合、通常の最尤推定法による確率値の推定では誤差が非常に大きくなってしまふ。それに対して、ベイズ学習の枠組を利用すれば、適切な事前分布を設定することで、最尤推定法よりもはるかに正確な推定を行なうことができる。本論文では、ベイズ学習による確率値の推定手法を決定リスト分類器に適用する。決定リスト分類器は、様々な自然言語処理においてその有効性を示しており、特に語義曖昧性解消問題に対しては、語義

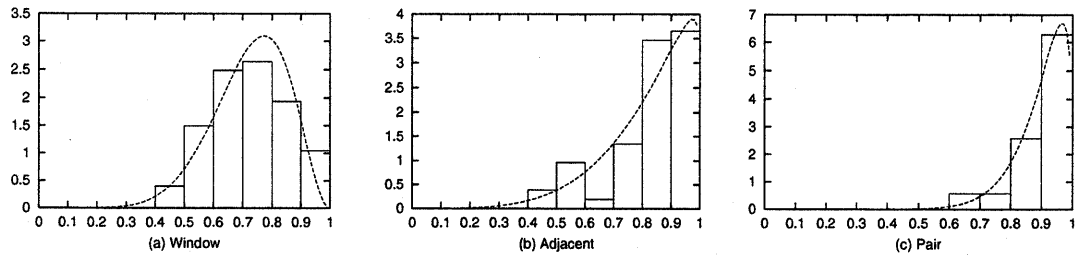


図 1: 仮説の確率値の分布と事前分布

判別の精度を競う競技会である Senseval-1 において，決定リストを階層的に拡張した手法が最も良い成績をあげている。

ベイズ学習において最も重要なことは事前分布のどのように設定するかである．本論文では詳細さの異なる複数の種類の文脈上の証拠を用い，各々の種類について事前分布を設定する．すなわち，ある種類に属する仮説の確率は，その種類に属する他の多くの仮説の確率値と同じような性質を持っていると考える．そして，事例の数が多く確率値の推定が容易な仮説の確率値を利用して，事前分布のハイパーパラメータを設定する．図 1 に事前分布の例を示す．グラフ中のヒストグラムは，事例が多い仮説の確率値の分布を示し，曲線はそれらをもとに設定された事前分布を示す．証拠の種類によって，仮説の確率値の事前分布に明瞭な違いがあることがわかる．決定リスト中の仮説の確率値は，これらの事前分布と訓練データから計算される事後分布の期待値として得られる。

本手法の有効性を確かめるために，決定リストによる語義曖昧性解消の実験を行なった．その結果，提案手法によって事前分布を利用して確率値を推定することで，仮説の確率値の誤差が事前分布を一様分布とした場合と比較して半減することが確認された．また，決定リストのルールの信頼度の計算法として従来用いられてきた最尤推定法や，対数尤度比による手法と分類性能の比較を行なった結果，提案手法が分類性能の点でも優れていることを確認した．

事例から知識を学習しようとする場合，それぞれの事例にラベルがついている，いわゆる教師付き学習を考えることが多い．しかし教師付き学習を行なうためには，事例に対するラベル付けという非常にコストの大きな作業を人間が行なわなければならない．そこで最近注目を集めているのが，ラベルのついていない事例も利用して学習を行なおうとする手法である．そのような手法としては，EM アルゴリズムと Naive Bayes 分類器を組み合わせる手法，Co-training と呼ばれる手法，Transductive Support Vector Machine などが提案され，テキスト分類や語義曖昧性解消などの実験を通して，ラベルなしデータを利用することで分類器の性能向上が可能なが報告されている．

本論文では，自然言語処理において様々なクラス分類問題に適用されている決定リスト分類器でラベル無し事例を利用する手法を提案する．提案する学習アルゴリズムでは，信頼度が確率で得られる分類器を用い，ラベル無し事例につけた確率的なラベル（以後，確率ラベル）を繰返し更新していくことにより学習を行なう．繰返しのたびに，確率ラベルが付与された事例から分類器の学習を行い，その分類器によって確率ラベルの更新を行う．このような繰返しによって学習を行なうことにより，ラベル有り事例だけからでは学習することのできなかったルールを学習することができる．さらに，確率ラベルからその信頼度を考慮して学習を行なうことにより，間違っラベル付けが間違っ学習を引き起こすという悪循環を軽減している．

本学習アルゴリズムを疑似単語の判定問題に適用して実験を行なった結果，ラベル無し事例を

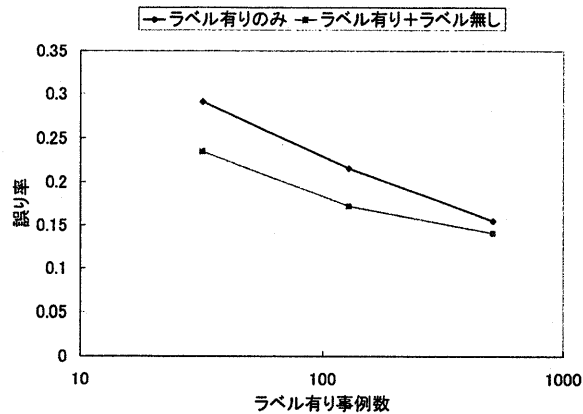


図 2: ラベル有り事例の数と分類誤り率

表 1: 指し手による遷移確率

カテゴリ	遷移確率
直前に動いた敵の駒を駒得しながら取る手	44%~89%
放置すると大きな駒損になる駒が逃げる手	21%~56%
駒損をしないで歩で王を叩く手	25%
駒損をしないで両取りを掛ける手	18%~40%
:	:

利用することで分類誤り率が大きく改善することが確認された。その効果は、ラベル有り事例の数を数倍に増やすことに相当する効果であった（図 2）。本手法では、分類性能の点だけでいえば、Naive Bayes 分類器 と EM アルゴリズムを組み合わせた手法に匹敵する性能は得られなかった。しかし決定リストによる学習では、学習内容が人間が理解可能な if-then 形式のルールの並びで得られるという利点がある。そのため、学習途中で人間が介入して学習効率を高めることも可能だと考えられる。

計算法による知的処理の大きな要素として、合理的な意思決定の問題がある。多種多様に変化する現実世界において、現時点での最も合理的な行動を決定するという問題は、人工知能の究極の目標の一つだと言ってもよい。その問題に対する一つのアプローチは、先読みした未来の状態を何らかの評価基準によって評価し、最も評価が高くなるような行動を選択するという方法である。その研究の題材として、古くから計算法によるゲームプレイヤが研究され、コンピュータチェスでは、全幅探索といういわば力任せの探索手法によって人間のチャンピオンを凌駕するにいたった。しかしチェスよりも分岐数の大きいゲーム、例えば将棋や囲碁などに関してはそのような力任せの手法は無力であり、コンピュータプレイヤの実力もアマチュアレベルの段階に留まっている。

本研究では、コンピュータ将棋プレイヤにおいて、プロ棋士の棋譜から抽出した確率的知識を探索に利用することで、より効果的な探索が可能になることを示す。従来の探索アルゴリズムの多くが深さを打ち切り条件として用いるのに対して、本アルゴリズムでは、局面の実現確率を打ち切り

条件として用いる。その理由は、将棋のエキスパートプレイヤーは、深さに基づいて指し手を読んでいるわけではなく「ありそうな展開」、すなわち実現確率の高い局面を中心に読んでいると考えられるからである。そして実現確率を計算するために、プロ棋士の棋譜から抽出した確率的知識を遷移確率として利用する。表1に、遷移確率の高い指し手のカテゴリのいくつかを、その確率値とともに示す。

本アルゴリズムの評価として深さ打ち切りプログラムとの対戦による評価を行なった結果、同一思考時間における勝率は9割近くにも達し、本アルゴリズムによる探索が極めて効果的であることが確認された。また、思考時間を深さ打ち切りの1/5としても勝率が5割を上回ることから、その効果は、探索を5倍以上高速化することに相当するといえる。さらに、次の一手問題によって、現在のトップレベルの将棋プログラムとの比較を行なった結果、同等以上の正解率を達成することができ、実用的な観点からも優れた探索アルゴリズムであることを確認した。