典型性に基づく概念学習アルゴリズム

上原邦昭† 谷澤正幸†† 前川禎男†

本稿では，予め分類された訓練例集合から，特徴の典型性と情報量の期待値という二つの尺度に基づいて概念を学習し，その概念を基にして新たな事例を分類するアルゴリズムを提案する．従来の帰納的学習では，ID3 のように，訓練例集合からカテゴリー間の違いを最小限のルールとして獲得するアルゴリズムの開発に焦点があてられていた．本稿で提案するアルゴリズムは，カテゴリーの典型的な特徴と弁別的な特徴に着目して，未学習の事例と各カテゴリーの距離によって分類するものである．このため，訓練例に誤った特徴が含まれていても，学習される概念は平均化されたものとなり，その影響を小さくすることができるという特徴がある．また，平均化された概念には多くの情報が残っているために，訓練例が十分に得られない場合でも，偶然的な影響を小さくすることができるという特徴がある．さらに，本アルゴリズムは複数のカテゴリーを分類候補として求めることもできるために，分類候補の絞り込みアルゴリズムとしても利用できるという特徴がある．

Prototype Based Concept Learning Algorithm

KUNIAKI UEHARA,† MASAYUKI TANIZAWA†† and SADAO MAEKAWA†

This paper describes a prototype based concept learning algorithm from cases. Most of the existing inductive classification algorithms, such as ID 3, mainly concentrate on the extraction of minimum discrimination rules to separate categories under consideration. On the other hand, our approach is to classify a new case into nearest categories by using prototype based distance metric, where prototype theory was proposed E. Rosch. Since prototype based classification approach highly uses statistical information extracted from training cases, compared with rule induction approaches, it achieves high accuracy to classify from even small training cases. Furthermore, this approach can derive some possible candidate categories for a new case.

1. はじめに

帰納的学習とは，外部から与えられた事例集合から帰納的推論を行い，それらの背後にある一般的な概念あるいは規則（知識）を獲得するプロセスである．帰納的学習は，さらに例からの学習と観察による学習に大きく分けることができる．例からの学習では，目標概念に対する正例を説明し，負例を除外するような一般的記述を決定することがタスクとなる．一般的記述としては，決定木[10]，ニューラルネットワーク[3]，ルール[7),11]などが用いられることが多い．

本稿では，例からの学習の新たな枠組として，典型性に基づく概念学習アルゴリズム（Prototype-Based Learning Algorithm，以降では PBL と略記する）を提案する[16),17]．PBL では一般的記述として，典型度と呼ぶ特徴の出現傾向を表す情報と，重要度と呼ぶ特徴のカテゴリー間での偏りを表す情報を用いている．分類すべき事例が与えられると，両者を用いて各カテゴリーとの類似性の度合を求め，類似性がもっとも高いカテゴリーに事例を分類するようになっている．

従来の例からの学習では，事例の弁別に有用な特徴に着目して学習を行うために，訓練例集合が小さい場合，訓練例にノイズ（分類誤り，計測誤り，計測誤差）が含まれていると，分類誤りの場合には一般的記述が得られないこと，計測誤りや計測誤差の場合には誤った複雑なルールを導くことなどの問題があった．これに対して，PBL では，弁別的な特徴に加えて，与えられた事例が，どの程度，カテゴリーと類似しているかという度合に基づいて分類しているために，訓練例集合が小さくてもある程度の分類精度を得ることができるという特長がある．また，事例集合にノイズが含まれていても，特徴の平均的な出現傾向を用いて分類しているために，ノイズの影響を受けにくいという特長がある．

† 神戸大学工学部情報知能工学科
Department of Computer and Systems Engineering, Faculty of Engineering, Kobe University
†† 三菱電機(株)制御製作所
Mitsubishi Electric Corporation

本稿では，まず2章で，訓練例集合に含まれる典型的な特徴と弁別的な特徴について考察する．さらに，事例を分類する際に，これらの特徴をどのように用いればよいかについて考察して，PBLの基本アルゴリズムPBL1を提案する．3章では，典型性に基づいて事例を分類する場合には，どのようにして各特徴に重要度を割り当てればよいかという問題を検討する．さらに，この考察にしたがって，情報量の期待値に基づいて特徴の重要度を変化させるアルゴリズム PBL2 を提案する．また，PBL2 を大分類と詳細分類の2段階に適用するアルゴリズム PBL3 を提案する．4章では，他の類似したアルゴリズムと比較しながら，PBLの有効性について検討する．最後に5章では，本アルゴリズムの問題点および今後の課題などについて議論する．

2. 特徴の典型性

2.1 基本的な考え方

ある単一のカテゴリーに属する事例に数多く出現しているが，そのカテゴリーのすべての事例に出現するとは限らない特徴のことを典型的な特徴[13]と呼ぶ．また，ある事例を分類する上で，他の事例（カテゴリー）と明確に識別するのに有効な特徴を弁別的な特徴と呼ぶ．この定義に従うと，「空を飛ぶ」という特徴は鳥の典型的な特徴になる．一方，弁別的な特徴については，訓練例集合が不十分な場合，他の特徴の影響を受けて発見が困難になることもある．しかしながら，このような場合でも「空を飛ぶ」という特徴は鳥の典型的な特徴であるために，分類に利用することができる．このように，訓練例集合が不十分な場合には，典型的な特徴を用いて分類する手法が有効になる．言い換えると，訓練例集合に含まれる頻度の高い特徴を分類に利用するという考え方である．

つぎに，十分な訓練例集合が与えられた場合について考える．このような場合に分類で重要となる特徴は，特定のカテゴリーに固有の弁別的な特徴であって，典型的な特徴が必ずしもカテゴリーの区別に役立つとは限らない．前述の例では，「空を飛ぶ」という特徴は典型的な特徴であるが，鳥と蝶を分類する上では有用ではない，両者を分類する上では，「口ばし」とか「羽毛」などの弁別的な特徴が有効である．言い換えると，事例の中には典型的な特徴であっても分類には全く関係ない不要な特徴が含まれていることもある．このように，十分な訓練例集合が与えられた場合には，典型的な特徴に加えて弁別的な特徴も同時に考慮することにより，さらに分類精度を向上させることができる．

また，人間は分類作業を行う場合，新たな事例を唯一のカテゴリーに分類することなく，複数のカテゴリー候補を提示する場合も多い．たとえば，医療の診断の場合，初期の診断から唯一の病気に断定した上で治療を進めてしまうと，誤った診断であった場合には取り返しがつかなくなるような事態が生じる可能性もある．このため，実用を考えた信頼性の高い分類を実現するためには，可能性のある分類候補とその可能性の度合を示すこと，さらにそれらの限られた候補の中で事例の再分類を行うことが必要となる．

2.2 PBL1

本節では，典型性に基づく概念学習アルゴリズムの最も基本的なアルゴリズムである PBL1 について説明する．まず，事例の記述形式について示す．事例は，属性とその値からなる特徴（以降では，属性とその値からなる対を特徴と呼ぶ）とその事例が属するカテゴリーからなる．n 個の特徴を持つ j 番目の事例 I_j は以下のように表される．

$$I_j=(c_j, a_{1j}, a_{2j}, \cdots, a_{nj})$$

ただし，c_j はカテゴリー，a_{ij} は i 番めの属性を表すものとする．

PBL1 では，まず与えられた訓練例集合から，各カテゴリーごとにすべての特徴の典型度を求める．カテゴリーにおける特徴の典型度とは，カテゴリー中でその特徴がどの程度出現しているかという頻度を示すものである．カテゴリー c において属性 a_i が値 k を持つときの典型度を以下のように定義する．

$$\text{典型度}\ (k, i)=\frac{\sum_{j=1}^{N} f(a_{ij}, k)}{N} \qquad (1)$$

ただし $f(a_{ij}, k)$ は，もし $a_{ij}=k$ ならば1，そうでなければ0となる関数である．k は i 番目の属性 a_i が取り得る値とする．また，N はカテゴリー c に含まれる事例数を表している．

分類すべき事例 $I_{new}=(b_1, b_2, \cdots, b_n)$ が与えられると，特徴の典型度を用いて連想度と呼ぶ類似性の度合を算出する．連想度は，典型度の高い特徴が分類事例に多く含まれていれば高い数値を示し，少なければ低い数値を示すように定義している．カテゴリー c における事例 I_{new} の連想度を以下のように定義する．

$$\text{連想度}\ (c, I_{new})=\sum_{i=1}^{n} \text{典型度}\ (b_i, i) \qquad (2)$$

ここで，n は事例が持つ属性数である．

連想度はすべてのカテゴリーに対して計算され，与えられた事例は連想度の最も大きなカテゴリーに分類される．以降では，(2)式で定めた連想度を用いて分類を行うアルゴリズムをPBL1と呼ぶ．なお，PBL1は，連想度の値によって一意にカテゴリーを決めず，連想度の大きな複数のカテゴリー候補を出力することもできるという特徴がある．複数のカテゴリー候補を利用したアルゴリズムについては，3.3節で再び検討する．

2.3 ID3とPBL1の比較

ID3とPBL1をsoy bean data（大豆の病気に関するデータ）に適用して分類精度の比較を行った．ID3[10]は，弁別的な特徴に注目して訓練例集合から未学習の事例を分類するための決定木を生成するアルゴリズムである．決定木の各ノードは属性に対応しており，情報量の期待値を用いて適切な属性を決定できるようになっている．soy bean dataは，文献7)においてルール帰納アルゴリズムAQと専門家の比較に用いられたデータである．なお，soy bean dataは，気温，発病時期，葉や茎の状態など，50の属性からなる大豆の症例を17種の病名に分類したデータである．実験で用いたsoy bean dataは289例からなり，アルゴリズムの比較には，訓練例として145例，分類のテストとして残りの144例を用いている．訓練例の増加による分類精度の変化を**図1**に示す．この分類精度は訓練例と未学習の例をランダムに20回選び直して平均をとった値である．

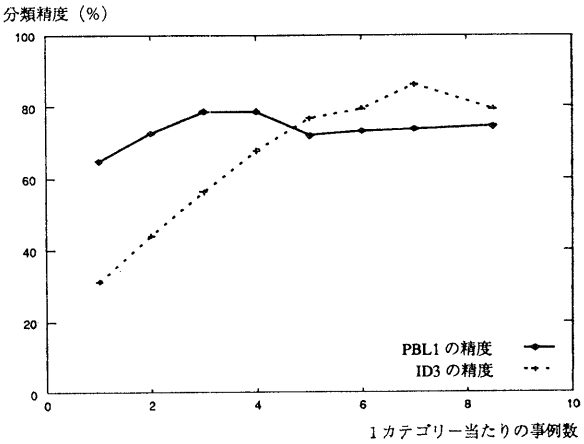


図1 PBL1とID3の分類精度の比較
Fig. 1 Performance accuracy of PBL1 and ID3.

この比較実験から，ID3では各カテゴリーに属する訓練例が少ない場合，分類には不適切な属性が決定木のノードとして選択されるために，分類精度が極端に悪くなっていることがわかる．また，ID3は1カテゴリーあたりの訓練例が増加するにつれて分類精度が向上しているのに対して，PBL1は訓練例が少ない場合でも，ある程度，正確に分類しているが，訓練例が増加しても余り精度の改善がみられないことがわかる．

3. 特徴の重要度

3.1 PBL2

前章で提案したPBL1において，訓練例が増加しても分類精度の向上が得られない原因として，分類に関係のない不適切な特徴の影響を受けていることが挙げられる．言い換えると，PBL1では典型的な特徴のみに着目して分類しているために，ある特徴が各カテゴリーに均等に含まれている場合は，それが典型的な特徴であっても分類精度に与える影響はほとんどない．逆に，一部のカテゴリーに偏在している特徴は，典型的な特徴でなくても分類には重要な特徴となる．たとえば，乗り物の分類において，「タイヤ」という特徴は「乗用車」の典型的な特徴であるが，「自転車」や「オートバイ」においても典型的な特徴であるために，これらを分類する上で「タイヤ」は役立たない．しかしながら，「荷台」という特徴は「トラック」のみに存在する特徴であるために，「荷台」のみで事例を「トラック」に分類することができる．以上の考察から，分類に有効な特徴の典型度をより強調し，分類に有効ではない特徴の典型度を弱めるように重み付けを変えることができれば，PBL1の分類精度をさらに改善することができると考えられる．

特徴の重み付けとして，記憶に基づく推論などでさまざまなものが提案されているが，重みは与えられる事例集合にも依存しているために，試行錯誤的に決定されている場合も多い．たとえば，事例に基づく学習システムBloom[1]では，逐次的に重みを少しずつ更新していく手法，専門家から知識獲得するシステムProtos[9]では，専門家の説明から発見的に特徴の重み付けを獲得する手法などが提案されている．

本稿では，特徴の分布情報を示す概念として，情報量の期待値（エントロピー）[4]を用いて特徴の重み付けを定義する．エントロピーはID3でも採用されているが，弁別的な特徴を明示的に求めるための判断基準

として用いられている．これに対して，以降で提案する PBL 2 では，特徴の典型度を用いて分類する際に，特徴が分類にどの程度重要であるかどうかを表すための数値として用いている．また，ID 3 では決定木の生成過程で各属性のエントロピーを求めても，ノードとなるべき属性が決定された段階でその値は捨て去られるのに対し，PBL 2 では特徴の重みとして最後まで保存されるという違いがある．

特徴の重み付けとエントロピーには非常に深い関係がある．たとえば，ある特徴を持つ事例が各カテゴリー中に均等に存在しているような場合，その特徴を持つ事例がどのカテゴリーに属するかは全く等確率であり，分類には不要な特徴になる．この場合，エントロピーはカテゴリー数 n に対して最大値 $\log_2 n$ をとる．逆に，ある特徴を持つ事例が一つのカテゴリーのみに存在するような，分類において重要な特徴である場合，エントロピーは小さくなる．つまり，重要な特徴ほどエントロピーは小さく，分類に役立たない特徴ほどエントロピーは大きくなる．

したがって，エントロピーの小さな特徴ほど重み付けを大きくすればよいということになる．しかしながら，エントロピーの取り得る範囲は状況によって変化すること，エントロピーと特徴の重み付けは負の相関関係を持っていることから，エントロピーをそのまま特徴を重みとして利用することは不適切である．このため，エントロピーに対して式(3)に示す変換を施し，取りうる値の範囲を0から1の間に正規化する．この値を特徴の重要度と呼ぶ．属性 a_i が値 k を持つときの重要度を以下のように定義する．

$$\text{重要度}\ (k,i)=2^{-(-\sum_{c_l\in C}P(c_l|a_i=k)\log_2 P(c_l|a_i=k))}=\prod_{c_l\in C}P(c_l|a_i=k)^{P(c_l|a_i=k)} \quad (3)$$

ただし，C はカテゴリー集合，$P(c_l|a_i=k)$ は，属性 a_i が値 k を持つことが観測されたときに，その事例がカテゴリー c_l に属する確率を示している．式(3)によって，重要度が1のときにその特徴が最も分類に重要であり（単一のカテゴリーのみに存在する特徴），重要度が0に近づくほど分類に無関係な特徴となる．

特徴の重要度によって拡張された連想度を式(4)に示す．式(4)は，各属性ごとに典型度と重要度の積をとったものの総和で，カテゴリー c における事例 I_{new} の拡張連想度と呼ぶ．なお，以降では，重要度を用いて拡張されたアルゴリズムを PBL 2 と呼ぶ．

$$\text{拡張連想度}\ (c, I_{new})=\sum_{i=1}^{n}\text{典型度}\ (b_i,i)\times\text{重要度}\ (b_i,i) \quad (4)$$

PBL 2 を用いて soy bean data を分類した結果を**図 2** に示す．この実験結果から，特徴の重要度の導入によって，PBL 2 は ID3 や PBL 1 と比較して，かなり高い分類精度を示すようになったことがわかる．

3.2 重要度の感度分析

本節では，前節で導入した重要度を式(5)のように指数乗した拡張重要度について検討する．なお，式(5)の乗数 x を重要度乗数と呼ぶ．

$$\text{拡張重要度}\ (k,i)=\text{重要度}\ (k,i)^x \quad (5)$$

拡張重要度は，重要度乗数 x の値によってエントロピーによる重み付けの度合を変化させるようにしたものである．重要度は0から1の間の値を取るために，重要度乗数が大きくなれば拡張重要度はエントロピーに敏感に反応する．また，重要度乗数が小さくなれば，エントロピーにあまり反応しなくなるという特性を持っている．言い換えると，重要度乗数を大きくすれば，重要な特徴のみが強調されてエントロピーの大きな特徴は分類に影響しなくなり，重要度乗数を小さくすれば，あまり重要でない特徴も考慮した分類となる．重要度乗数による soy bean data の分類精度の変化を**図 3** に示す．図3の結果から，重要度乗数が 1.7 前後で分類精度が最大になっていることがわかる．なお，当然のことながら，重要度乗数が1の場合には，前節で導入した重要度と同一の結果を示すことになる．

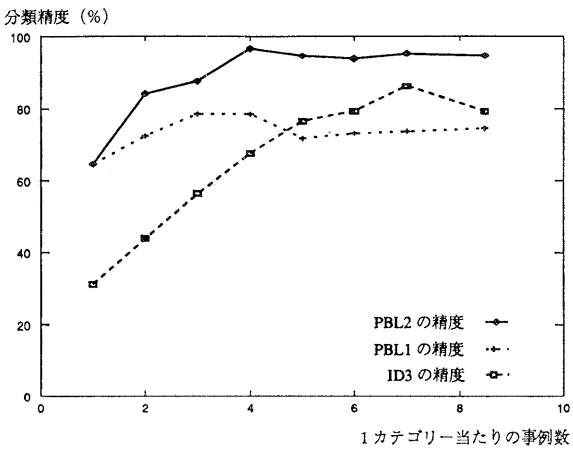


図 2 PBL1 と PBL2 および ID3 との分類精度の比較
Fig. 2 Performance accuracy of PBL1, PBL2 and ID3.

3.3 重要度の状況依存性

PBL 2では，特徴の重要度が訓練例集合から一意に決定されると仮定している．しかしながら，重要度のなかには状況によって変化するものが存在する．たとえば，「乗り物の分類」において，「タイヤが二本」という特徴は「自転車」と「オートバイ」を区別する際には重要でないが，「オートバイ」と「乗用車」を区別する場合には重要な特徴となる．このような状況依存した重要度に対応するために，本節では，PBL 2を2段階に適用するアルゴリズム PBL 3を提案する．

PBL 3では，まず第1段階で連想度によるカテゴリー候補の絞り込みを行う．つぎに第2段階では，絞られた複数のカテゴリー候補の中で重要度を再計算し，新たに得られた重要度を用いて訓練例の再分類を行い，最も高い連想度を持つカテゴリーに分類するようにしている．なお，第2段階の分類作業に移行するのは，上位のカテゴリー候補の連想度が近接している場合のみに限定している☆．この制限は，第1段階でカテゴリーが一意に絞り込める場合には，第2段階での重要度の再計算が無駄になること，第2段階ではカテゴリー候補が少なくなるために，一部のカテゴリーにのみ存在する特徴の重要度が相対的に弱められる可能性があり，逆に誤分類につながることなどの理由による．

PBL 3 と PBL 2 の分類精度の変化を**図4**に示す．なお，比較に用いた重要度乗数は1段階目が1.7，2段階目が4.0である．結果としては，わずかながら(3.2%程度) PBL 3の分類精度が向上している．比較実験では，データをランダムに入れ換えて100回のシミュレーションを行い，分類精度の平均値を求めている．また，PBL 3の分類精度の向上については，信頼率1%の統計的検定によって確認している（**表1**参照）．

PBL 3において，第1段階と第2段階で重要度乗数を変えているのは，互いに分類の性質が異なるためである．この違いについて説明するために，図3にPBL 3の結果を併せたものを**図5**に示す．図5から，PBL 3（1段階目の重要度乗数は1である）は PBL 2 と比べて，2段階目の重要度乗数が2以下では分類精度が落ちていること，重要度乗数が大きくなっても分類精度はほとんど低下していないことなどがわかる．これ

☆ 連想度が最大のカテゴリーと比較して，値が0.8倍までのものをカテゴリー候補としている．

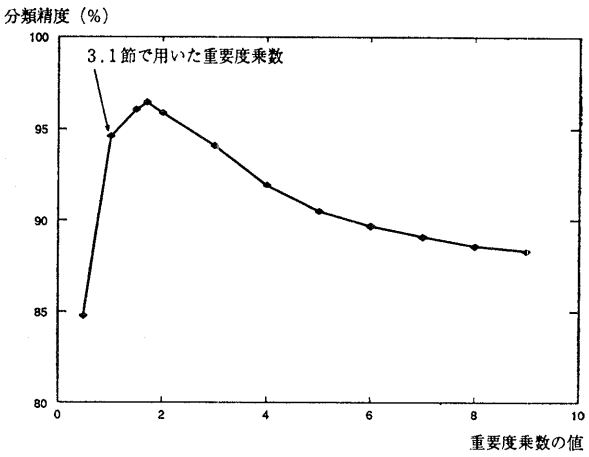


図3 重要度乗数による分類精度の変化
Fig. 3 Sensitivity to feature weight settings.

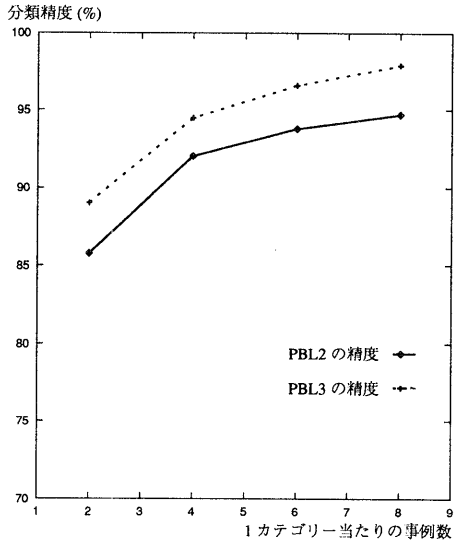


図4 PBL2 と PBL3 の分類精度の比較
Fig. 4 Performance accuracy of PBL2 and PBL3.

は，PBL 3では重要な特徴を第1段階と比べてより強調しなければならないこと，また，第2段階ではあまり重要でない特徴は分類に影響していないことによる．すなわち，第1段階の大分類では，多くのカテゴリーを考慮するために重要度の小さな特徴も無視できないのに対し，第2段階の詳細分類では，絞られたカテゴリー候補間でのみ分類が行われるために，重要度の低い特徴は逆に誤分類につながることを示している．したがって，PBL 3では両段階とも同一の手法が用いられているが，分類の性質は互いに異なったものとなっている．最後に，本節までに述べた，PBL の全

体アルゴリズムを**図6**に示す．

4. 性能評価および考察

4.1 類似したアルゴリズムとの比較

前章までは，soy bean data のみを用いて PBL の有効性を検証してきた．本節では，PBL を種々のデータベース[8]に適用して分類精度の検証を行う．適用したデータベースの各種条件を**表2**に，分類結果を**表3**に示す．ただし，PROTO-TO と C4（ID3 の改良アルゴリズム）の分類結果は文献 6）から引用したものである．なお，表の空欄の部分は未計測または計測不能のデータであることを示している．また，分類精度については，一部を除いて，訓練例集合とテスト例集合をランダムに 50 回選び直して適用した平均を示している．さらに，PBL では実数値データを直接扱えないために，実数値の取り得る範囲を適当な数に分割している．このため，分割の仕方によって分類精度が大きく変わるデータに関しては，範囲を持たせて示している．

PBL 2 については，発見的に求めた最適な重要度乗数による分類精度を示し，括弧内にはデフォルトの重要度乗数 1.0 の場合の精度を示している．なお，紙面の都合で詳しく説明できないが，発見的な重要度乗数の求め方は以下の通りである．まず，試行錯誤的に重要度乗数を設定して連想度を求め，訓練例集合の再分類を行う．この結果から分類誤差を計算し，最も分類誤差が小さくなるまで上記の操作を繰り返し，最適な重要度乗数を求めるようにしている．PBL 3 については，重要度乗数の調整が複雑であるために，デフォルトの重要度乗数として 1 段階目は 1.0，2 段階目は 4.0 としている．

表 3 の結果から，PBL 2 は glass, hepatitis, tic-tac-toe を除いて高い分類精度を示していることがわかる．このうち glass については，すべての属性が実数値で与えられているために，実数値の分割が分類精度に大きく影響していることがわかる．したがって，実数値の分割点または最適な分割数を決定するアルゴリズムの開発，あるいは PBL に実数値を取り扱う能力を持たせることが必要であると考えられる．

hepatitis は 2 カテゴリーしか持たないにもかかわ

表 1 PBL2 と PBL3 の分類精度の比較
Table 1 Comparison of performance accuracy between PBL2 and PBL3.

訓練事例数		$n=2$	$n=4$	$n=6$	$n=8$
PBL2	平均 (%)	85.8	92.0	93.8	94.7
	分散 (10^{-4})	9.18	4.59	4.30	2.76
PBL3	平均 (%)	89.0	94.5	96.6	97.9
	分散 (10^{-4})	6.39	3.16	2.78	1.65
検定結果		PBL3>PBL2	PBL3>PBL2	PBL3>PBL2	PBL3>PBL2

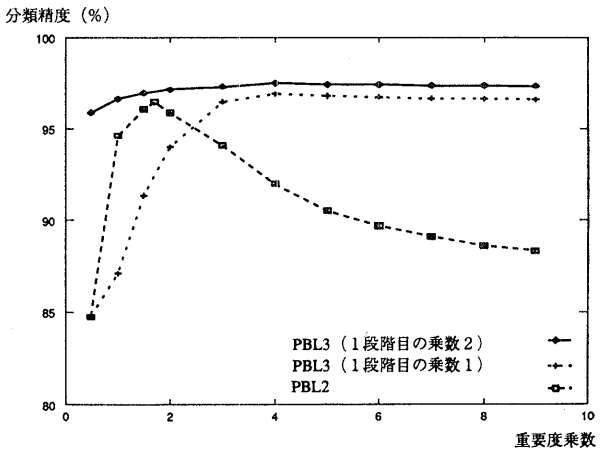


図 5 重要度乗数による分類精度の変化
Fig. 5 Sensitivity to context-sensitive feature weight settings.

学習部
1 訓練例集合を入力する．
2 カテゴリーごとに特徴の典型度を算出する．（重要度乗数の調整）
3 特徴ごとに重要度を算出する．

分類部
分類を行う事例を入力する．
1 すべてのカテゴリーに対して連想度を算出する．
2 連想度の値に応じてカテゴリーをソートする．
3 If PBL3 を用いる（カテゴリー数 3 以上）
4 then 連想度の値が最も大きいカテゴリーおよび値が近接しているカテゴリー候補を求める．
6 If カテゴリー候補が 2 以上ある
7 then カテゴリー候補の中で重要度を再計算する．
8 各カテゴリー候補に対して連想度を算出する．
9 連想度の値に応じてカテゴリーをソートする．
10 連想度の値が最も高いカテゴリーに事例を分類する．

図 6 PBL の全体アルゴリズム
Fig. 6 PBL algorithm.

らず，カテゴリー間の事例数に大きな偏りがあること，データに数値データが含まれていること，事例が持つ特徴として弁別的な特徴がほとんど無いことなどが，分類精度の低い原因として考えられる．しかしながら，典型的な特徴はいくつか含まれているために，他のアルゴリズムと比較しても同程度，あるいはそれ以上の分類精度を示している．

tic-tac-toe は，特徴間の関係のみが分類に影響を与えるデータであるが，PBL などの確率的アルゴリズムでは特徴間の関係を学習に反映させることができないために，適切な概念の獲得が行われていない．このようなデータの分類には constructive induction[2] や記憶に基づく学習[15]などの手法を導入することが必要であると考えられる．

表 2 適用したデータベースと適用条件
Table 2 Summary of databases used in experiments.

データベース名	訓練例	テスト例	属性	カテゴリー	特徴の欠落	実数値の属性
breast cancer	350	349	9	2	あ り	な し
glass	107	107	9	6	な し	すべての属性
hepatitis	78	77	19	2	あ り	一部あり
house-vote	218	217	16	2	あ り	な し
iris	75	75	4	3	な し	すべての属性
tic-tac-toe	479	479	9	2	な し	な し
zoo	51	50	17	7	な し	な し

表 3 複数のデータベースに適用した分類結果
Table 3 Summary of experimental results.

データベース名	PBL3	PBL2	PROTO-TO	C4(ID3)	Bayesian
breast cancer	—	95.2(91.5)%	—	—	97.2%
glass	45-55%	43-50%	48.0%	65.5%	—
hepatitis	—	84.2(82.5)%	79.9%	79.8%	84.8%
house-vote	—	92.1(90.3)%	90.4%	95.3%	90.5%
iris	—	95.4(95.4)%	96.0%	94.2%	95.3%
tic-tac-toe	—	72.9(65.6)%	—	—	67.7%
zoo	92.8%	93.0(92.7)%	—	—	93.8%

4.2 PROTO-TO との比較

PBL2 と独立して開発された概念学習アルゴリズムとして PROTO-TO[6] がある．PROTO-TO と PBL2 は，特徴の存在確率と重要度の積を用いて分類を行っている点が類似している☆．また，両者の相違点としては，PROTO-TO は属性ごとに，PBL 2 は属性値ごとに特徴の重要度を求めていることが挙げられる．このため，PROTO-TO は，属性値によって重要度が異なる特徴や比較するカテゴリーの組によって特徴の重要度が変化するデータに対しては，十分な分類精度を得ることができないと考えられる．逆に，PBL 2 では，属性値ごとに重要度を求めているために，直接的に実数値を扱うことができず，予めデータ集合の前処理を行って適切に分割しておく必要があるという問題がある．

PBL 2 と PROTO-TO を直接的に比較することはできないために，表 3 には文献 6)で PROTO-TO が用いたデータを用いて，PBL 2 をほぼ同一の条件で実行した結果を示している．表 3 の結果から，hepatitis と house-vote では PBL 2 が若干上回っているが，数パーセントの差であるため有意であるとは断定できない．glass では，PROTO-TO は実数値を直接扱えるようにしているにもかかわらず，PBL 3 よりも分類精度が低くなっていることから，PROTO-TO は実数値の取り扱いが十分ではないと考えられる．

4.3 Bayesian Classifier との比較

PBL と同様に確率的な情報を用いたアルゴリズムとして Bayesian Classifier[5] がある．Bayesian Classifier はすべての特徴が Boolean であるデータを対象としており☆，式(6)の score が高いカテゴリーに事例を分類するアルゴリズムである．

$$score(F,C)=\frac{k}{n}\prod_{i=1}^{r}\begin{cases}\dfrac{u_i}{k} & \text{if } F_j=1\\ \dfrac{k-u_i}{k} & \text{otherwise}\end{cases} \qquad (6)$$

ただし，F は分類する特徴集合，C は対象とするカテゴリー，n はすべての訓練例数，k はカテゴリー C の訓練例数，u_i はカテゴリー C における特徴 i を持つ訓練例数である．

PBL と Bayesian Classifier の分類精度の比較を表 3 に示す（試行回数 50 回）．表 3 の結果から，両者はほぼ同等の分類精度であることがわかる．一方，アルゴリズムの違いからみると，PBL と Bayesian

☆ 分類を行う場合に，PBL2 は得られた連想度の値の和が最も大きくなるカテゴリーを，PROTO-TO では典型的なカテゴリーとの差の最も小さいカテゴリーを選択するなど，分類に用いる式の細部は異なっている．

☆ PBL と同様に，直接的に実数値属性を取り扱うことはできない．

Classifier は，カテゴリーとの近接尺度を表す関数として，特徴の存在確率の和をとるか積をとるかという点が異なっている．このため，Bayesian Classifier では，特徴の存在確率として0が代入されると score は0となり，未学習の事例が適切なカテゴリーに分類されなくなるという問題がある．すなわち，ある特徴が訓練例集合には出現し，未学習の事例には出現しない場合，その特徴の存在確率は0となるために，誤分類の可能性が高くなるという問題がある．したがって，Bayesian Classifier では，特徴の存在確率が0になる場合には，値を $1/2n$ に置き換えるという操作を追加している．**図7**からもわかるように，この置換えを用いなければ，soy bean data の分類精度は50%程度にまで落ちている．また，置換えを行った場合でも，soy bean data は特徴数が多いために，訓練例が少ない場合は偶然性の影響を受けやすいということがわかる．これに対してPBLでは，一部の特徴の存在確率が偶然0となっても，存在確率の和をとっているために，影響を受けることは少なくなっている．

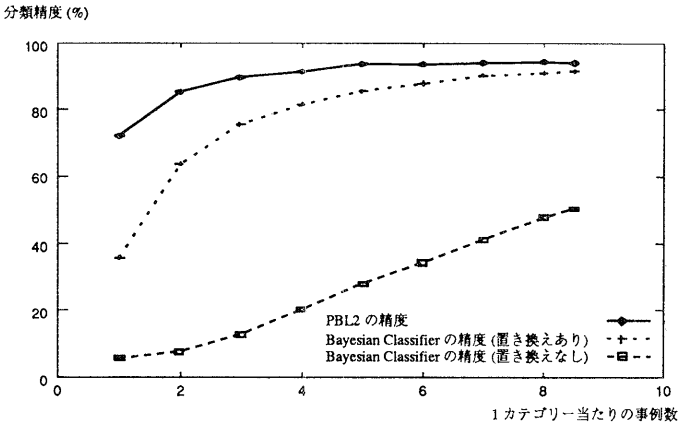


図7 PBL2 と Bayesian Classifier の分類精度の比較
Fig. 7 Comparison of performance accuracy between PBL2 and Bayesian Classifier.

以上のことから，訓練例が十分にあり，かつノイズがほとんどないデータに対しては，Bayesian Classifier のように，特徴の存在確率の積を用いて分類する方が曖昧性なく正確に分類できるということがわかる．しかしながら，このような理想的なデータはあまり現実的なものとは考えられない．したがって，PBL は Bayesian Classifier と比べて，ノイズの影響を受けにくいアルゴリズムであり，特に訓練例が十分でない場合には有効なアルゴリズムであると考えられる．

4.4 ニューラルネットワークとの比較

近年，人間の脳の情報処理に習ったニューラルネットワーク情報処理の研究が盛んに行われている．ニューラルネットワーク情報処理は，特にパターン認識などで成功を収めている手法であり，ノイズにも強くロバストな分類が可能である．PBL の入出力関係は階層型ニューラルネットワーク[3]の入出力関係と類似しているために，本節では，PBL と階層型ニューラルネットワークとの比較を行う．

両者の大きな違いとしては，ニューラルネットワークが訓練例集合を収束するまでパラメータを調整しつつ学習していくのに対して，PBL は，重要度乗数を調整する場合を除いて，連想度の値を1回計算すると終了する．このため，計算コストの点で大きな違いがある．同一の計算機で両者をC言語でコーディングを行い時間を計測した結果，PBL 2 の計算時間が約 0.1 秒であるのに対して，ニューラルネットワークでは収束するまでに約30分要している☆．

図8に，PBL とニューラルネットワークとの soy bean data に対する分類精度の比較を示す．なお，ニューラルネットワークには3層の階層型ニューラルネットワークを用いている．図8の結果から，PBL とニューラルネットワークの分類精度は非常に類似していることがわかる．一方，訓練例集合に対する分類精度はニューラルネットワークの方が高く，soy bean data に限定すると，収束した時点で完全に分類できるようになる．このように，ニューラルネットワークは訓練例の分類精度を上げるように内部のパラメータを調整していくため，訓練例に対する分類能力は高いが，過学習になったり，学習結果がローカルミニマムに陥るなど，未学習の例の分類精度が悪くなる場合がある．しかしながら，3層以上のニューラルネットワークでは，PBL では学習できない，特徴間の関係が分類に影響するようなデータに対する学習が可能であるという利点がある．

☆ 利用した計算機は，SPARC Station 2 (24.2 SPECmarks, 28.5 Mips, 4.2 MFlops) 相当のワークステーションである．

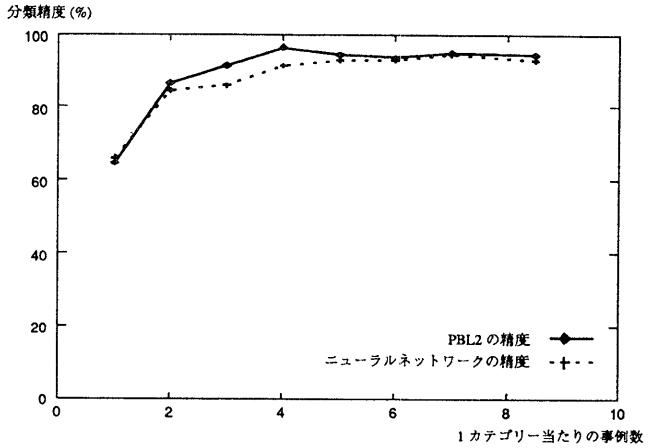


図 8 PBL2 とニューラルネットワークの分類精度の比較
Fig. 8 Comparison of performance accuracy between PBL2 and Neural Network.

5. おわりに

本稿では，予め分類された訓練例集合から，特徴の典型度と重要度という二つの尺度に基づいて概念の一般的記述を獲得し，その一般的記述を基にして新たな事例を分類するアルゴリズム PBL を提案した．本研究と類似な研究として，パターン認識におけるテンプレートの学習[14]がある．PBL における特徴の典型度はテンプレートに，重要度はテンプレートの強調に相当している．また，PBL 3 における2段階のカテゴリー候補の絞り込みは，パターン認識における階層的テンプレートマッチング技術に相当している．このように，本研究とパターン認識における学習は類似している点も多いが，パターン認識で開発された各要素技術がどのように機械学習と関連しているかについては未検討の課題である．

PBL は様々なデータに適用可能であると考えられるが，現状では，検討した例題が少ないために，適用できるデータの特性やその条件などが完全に把握されているとは言い難い．たとえば，PBL の大規模なデータへの適用についても未検討である．従来の例からの学習においても，ID3 から C 4.5[12] への改良のように，ノイズへの対処方法が提案されており，大規模なデータでは両者の分類精度の差異がほとんどなくなると思われる．これは本研究で行った比較実験からも予想されることである．また，tic-tac-toe のように特徴間の関係のみが分類精度に影響を与えるようなデータでは，C 4.5 のような手法の方が上回るものと考えられる．

さらに，実数値を含む事例の取り扱い，より適切な重要度乗数の獲得，PBL によって得られた概念からのルール化などについても問題点が残されている．今後は，これらの問題点を解決していくとともに，PBL を多くの例題に適用して適用条件や有効性を検証し改良していくことが必要である．

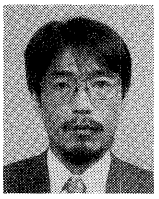
謝辞 本研究の一部は文部省科学研究費重点領域研究（知識科学における概念形成と知識獲得）および平成3年度大川情報通信基金の援助による．慶応義塾大学（当時）の開 一夫氏，King Fahd University of Petroleum and Minerals の Hussein Almuallim 氏にはいくつかの参考文献を教えて頂きました．また，本学大学院生衣川裕史君，福田慶郎君には PBL の評価について検討して頂きました．あわせて感謝いたします．

参考文献

1) Aha, D. W.: Incremental, Instance-based Learning of Independent and Graded Concept Descriptions, *Proc. of the 6th International Workshop on Machine Learning*, pp. 387-391 (1989).
2) Aha, D. W.: Incremental Constructive Induction: An Instance-Based Approach, *Proc. of the 8th International Workshop on Machine Learning*, pp. 117-121 (1991).
3) 麻生英樹：ニューラルネットワーク情報処理，産業図書，pp. 3-54 (1988).
4) 北川敏男：推測統計学 II，岩波全書，pp. 70-81 (1958).
5) Langley, P., Iba, W. and Thompson, K.: An Analysis of Bayesian Classifiers, *Proc. of the 10th AAAI*, pp. 223-228 (1992).
6) Maza, M.: A Prototype Based Symbolic Concept Learning System, *Proc. of the 8th International Workshop on Machine Learning*, pp. 41-45 (1991).
7) Michalski, R. S. and Chilausky, R. L.: Learning by Being Told and Learning from Examples: An Experimental Comparison of the Two Methods of Knowledge Acquisition in the Context of Developing an Expert System for Soybean Disease Diagnosis, *International Journal of Policy Analysis and Information Systems*, Vol. 4, No. 2, pp. 125-161 (1980).
8) Murphy, P. M. and Aha, D. W.: *UCI Reposi-*

tory of Machine Learning Databases [Machine-Readable Data Repository]: University of California, Department of Information and Computer Science, Irvine, CA (1992).
9) Porter, B. W., Bareiss, R. and Holte, R. C.: Knowledge Acquisition and Heuristic Classification in Weak-theory Domains, Technical Report AI-TR-88-96, Department of Computer Sciences, University of Texas at Austin (1989).
10) Quinlan, J. R.: Induction of Decision Trees, *Machine Learning*, Vol. 1, pp. 81-106 (1986).
11) Quinlan, J. R.: Generating Production Rules from Decision Trees, *Proc. of the 7th JICAI*, pp. 304-307 (1987).
12) Quinlan, J. R.: *C 4.5: Programs for Machine Learning*, Morgan Kaufmann (1993).
13) Rosch, E.: Principles of Categorization, in Rosch, E. and Lloyd, B. B. (eds.) *Cognition and Categorization*, Erlbaum (1978).
14) Shapiro, S. C. (ed.): *Encyclopedia of Artificial Intelligence*, John Wiley & Sons (1987).
15) Stanfill, C. and Waltz, D.: Toward Memory-Based Reasoning, *Comm. ACM*, Vol. 29, No. 12, pp. 1213-1228 (1986).
16) 谷澤正幸, 上原邦昭, 前川禎男: CBL と EBL を用いた体系的知識獲得, 第 43 回情報処理学会全国大会論文集, 3D-8 (1991).
17) 谷澤正幸, 上原邦昭, 前川禎男: 典型性に基づく概念学習アルゴリズム, 情報処理学会人工知能研究会資料, 86-5, pp. 33-40 (1993).

(平成5年10月19日受付)
(平成6年6月20日採録)



上原 邦昭（正会員）

昭和29年生. 昭和53年大阪大学基礎工学部情報工学科卒業. 昭和58年同大学院博士後期課程単位取得退学. 大阪大学産業科学研究所助手, 講師を経て, 平成2年神戸大学工学部情報知能工学科助教授. 神戸大学総合情報処理センター副センター長兼任. 平成元年より2年までOregon State University, Visiting Assistant Professor. 工学博士. 人工知能, 特に機械学習, 自然言語によるヒューマンインターフェイスの研究に従事. 1990年度人工知能学会研究奨励賞受賞. 人工知能学会, 電子情報通信学会, 計量国語学会, 日本ソフトウェア科学会, システム制御情報学会各会員.

谷澤 正幸

昭和 43 年生. 平成3年神戸大学工学部システム工学科卒業. 平成5年同大学大学院工学研究科システム工学専攻修士課程修了. 同年三菱電機(株)入社. 現在, 制御製作所にてソフトウェア生産支援ツールの開発業務に従事.

前川 禎男（正会員）

昭和6年生. 昭和 29 年大阪大学工学部通信工学科卒業. 昭和 34 年同大学院工学研究科博士課程単位取得退学. 大阪大学助手を経て, 昭和 36 年神戸大学工学部電気工学科助教授. 現在, 同大学工学部情報知能工学科教授. 工学博士. この間, システム理論, 高級言語マシン, 人工知能などの研究に従事. 人工知能学会, 電子情報通信学会, 電気学会, システム制御情報学会などの会員.