

1. 基本技術とその応用上の課題

Learning from Examples: Application to Industry 1. Basic Techniques and Application Issues by Yasuhiro AKIBA (NTT Communication Science Laboratories), Hussein ALMUALLIM (The Department of Information and Computer Science, King Fahd University of Petroleum & Minerals) and Shigeo KANEDA (Graduate School of Policy and Management/Faculty of Engineering, Doshisha University).

秋葉 泰弘¹ フセイン・アルモアリム² 金田 重郎³

¹ NTTコミュニケーション科学研究所

² サウジアラビア国立石油鉱物大学

³ 同志社大学大学院総合政策科学研究科・同志社大学工学部

1. はじめに

機械学習技術は、計算機を賢くすることを目指し、人間による判断事例を基に、与えた事例以外の未知の事例に対しても、計算機による自動的な判断を実現する技術である。本連載では、非数値のシンボルで事例を表現する機械学習技術「例からの学習」に焦点を絞る。

「例からの学習」は、種々の機械学習技術の中で、特に、実応用への適用が着実に進んでいる。適用分野としては、1) 天体観測画像からの天体分類、2) テキスト読み上げシステムの音声合成ルールの獲得、3) オゾン濃度予報、などがある。実問題には、事例に誤りを含むなどの、実問題特有の課題があるが、「例からの学習」アルゴリズムは、種々の手法でこの課題に対処している。

本連載は、実問題に「例からの学習」を適用するための手法を紹介するためのものであり、2回からなる。初回の本稿は「例からの学習」を中心とする機械学習を初学者向けに紹介する。今回は、実問題特有の課題に対する機械学習アルゴリズム側の対処法を、掘り下げて紹介する。

以下、第2章では、例からの学習について具体例を交えて紹介し、第3章では応用例を述べる。第4章では、実問題特有の課題を示す。第5章はまとめである。

2. 例からの学習

2.1 位置付け

本連載で紹介する「例からの学習」は、経験的学習技術の範疇に入る。経験的学習では、学習アルゴリズムをベンチマーク用データ[☆]でテストして、性能（主に、未知の事例に対する正解率）を評価する。すなわ

[☆] ベンチマーク用の事例データとしては、実応用の中から独自に収集したり、あるいは、標準ベンチマークデータを利用する。U. C. Irvine校のIrvine Databaseは、よく知られた標準ベンチマークデータである。

ち、優れた性能の学習アルゴリズムの提案を目指した、実験科学的な学問分野である。

経験的学習は、問題解決に関する状態空間を図-1のように複数の部分領域に分割する手法である。たとえば、機械の故障診断や患者の病名診断などの診断問題であれば、「○」「△」などは、学習に用いた事例（以下、これを「訓練事例」と呼ぶ）であり、記号の使い分けは、事例が所属する「クラス」機械の故障部位や病名に対応している。分割された各領域は特定のクラスに対応付けられる。図-1中、「◆」は、クラスが未知の事例であり、どの領域に所属するかでクラスが決まる。概念の獲得や技能の獲得などは、この枠組みに入る。

ここで、領域分割にあたっては、何らかの評価関数（正解率、領域分割の記述量など）を最適化する必要がある。先の例では、最小化が望まれるのは、未知事例に対する診断誤りである。

上記の分割と最適化が、学習アルゴリズムの役割であり、実問題に適用するためには、以下の5つのステップを踏む必要がある。

- (StepI) 入手可能な入力情報の種類の明確化
- (StepII) 入力情報の表現形式の決定
- (StepIII) 分割領域の表現形式の選定と適切な学習アルゴリズムの選定

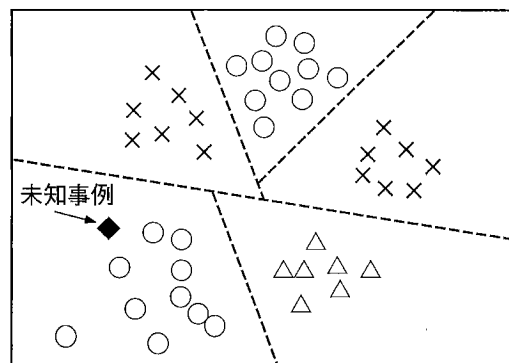


図-1 領域分割

	入力	入力情報の表現	分割の表現	学習手法
T1	事例	ベクトル	決定木	決定木学習
			決定リスト	決定リスト学習
T2	事例&領域知識	述語論理	命題理論	ルール学習
			Instance-Base	Nearest Neighbour
T3	アクション&その報酬	ベクトル	If-Thenルール	強化学習

図-2 経験的学習の分類

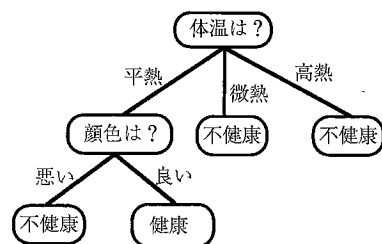
(StepIV) StepIIの表現形式への入力情報の変換, および, 学習アルゴリズムへの入力

(StepV) 所望の表現形式による分割の評価

経験的学習技術を上記の観点から区分して, 図-2に示す. 入力情報として事例のみを用いるタイプ (T1) が最もよく知られており, 実問題への適用も多い. 「例からの学習 (Learning from Examples)」と呼ばれている. ほかに, (T2) 事例と領域知識を用いる手法, (T3) アクションとその報酬を用いる手法がある. 前者の例は, 「説明に基づく学習」(EBL: Explanation Based Learning) であり, 領域知識および事例は, 述語論理で表現される. 直観的には, 訓練事例をガイドとして利用することにより, 領域知識から導出される定理を見出す手法である. また, 後者の例は, 強化学習である. ここでは, あるアクションとそれにより得られる報酬を事例として多数与え, 最適な戦略を学習する. 刻々変化する外部情勢に適応できるとされ, 研究が活発化している.

事例からの学習 (T1) では, 事例の表現法が, ベクトル形式と, 述語論理形式に大別できる. 後者は, 帰納的論理プログラム (ILP: Inductive Logic Programming) と呼ばれ, 事例の表現形式が自由であるため, 事物と事物の関係の学習に適する⁸⁾. また, その名前のおり, 事例を出力する (論理型の) プログラムを出力できる. ただし, 学習に必要な時間が, 事例数の指数オーダーのアルゴリズムが多く, 実用上のネックとなっている.

また, ベクトル形式の事例のアルゴリズム中でも, 事例ベース (Instance-Based) 学習では, 訓練事例をそのまま未知事例の判定に利用する. 図-1で説明すれば, 最も近い訓練事例のクラスで未知事例のクラスを決めたり (Nearest Neighbour法), あるいは, 最近傍のk個の事例のクラスの多数決により決める (k-



(a) 決定木

(体温=平熱) AND (顔色=良い) ⇒ 健康
真 ⇒ 不健康

(b) 決定リスト

((体温=平熱) AND (顔色=悪い)) OR
(体温=微熱) OR
(体温=高熱)

(c) ルール

図-3 例からの学習における分割表現 (概念「健康状態」の表現)

Nearest Neighbour法). 事例ベース学習は, 未知事例の判定に時間がかかるが, CPU/メモリのコストが大幅に下がっているため, 魅力ある手法である. ただし, 学習結果が人間に可読ではない.

以下, 例からの学習 (T1) の中で, ベクトルで事例を表現するアルゴリズムに焦点を絞る. ほかのアルゴリズムについては文献9) を参照されたい.

2.2 「例からの学習」アルゴリズム

本節では, 例からの学習が取り上げる問題設定を説明する⁹⁾.

入力情報の表現

入力情報は, シンボルまたは数値を成分とするベクトルで表現された, 複数の事例である. 事例のベクトル (以下, 属性ベクトルと呼ぶ) は, 一般には, 以下のように表される.

$$(V_0, V_1, V_2, \dots, V_{n-1}, V_n)$$

属性ベクトルの各成分の値を属性値という. 番号が最大の成分 (V_n) は特別で, その事例が属するクラス名である.

医療診断で例えると, 各属性は患者の症状 (顔色, 体温, 痛みの所在など) や種々の検査項目 (血液検査, 尿検査, レントゲン検査) となる. 属性値は, 顔色の場合, 赤や黄色や紫などのシンボルであり, 体温の場合, 37.5度などの数値となる. 一方, クラス名は病名である. ベクトルを具体例で示すと, (顔面蒼白, 37.5度, 下腹部の激痛, ..., 白血球激増, 尿異常なし, 右下腹部に異常陰影あり, ..., 急性虫垂炎) となる.

分割の表現

学習アルゴリズムが対象とするタスクは, 属性ベク

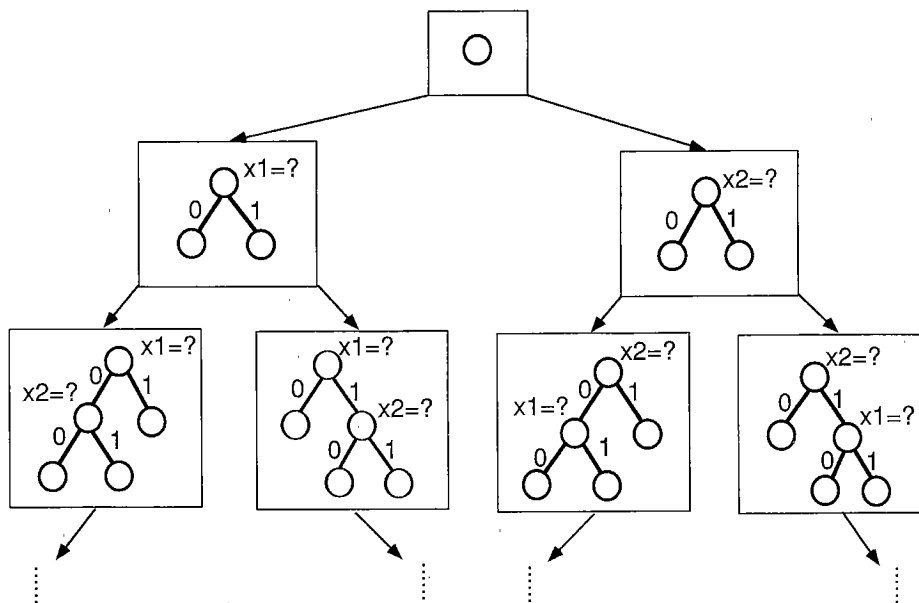


図-4 決定木の探索空間（属性が3つで、属性値が2値の場合）

トルで表現された訓練事例をうまくカバーする分割を分割の表現方法の中から探し出すことであり、基本的には、モデル空間上の探索となる（次節で解説する）。分割表現としては、決定木、決定リスト、命題論理などがある。

決定木（Decision Tree）は、図-3（a）に示すような、木構造をしている。木の葉を除く各ノードには質問が対応付けられ、各ノードから伸びる各エッジは、質問に対する回答に対応する。各葉はクラスに対応付けられている。決定木を利用したクラス判定では、ルートから順に各ノードに対応する質問に答え、その質問の答えに対応するエッジを辿り、葉に至ったらその葉のクラスをその事例のクラスとする。木のルートから各葉に至る経路は、その決定木で表現される分割領域に対応し、幾何学的イメージとしては互いに重なり合わない超立方体群である。

決定リスト（Decision List）は、図-3（b）に示すような、順序を持つIf-Thenルール群であり、各If-Thenルールの条件部は、1属性に関する条件を論理演算子ANDで結合した論理式、帰結部はクラス名である。最下位のルールは、デフォルトルール、すなわち条件部が恒真のルールである。決定リストでは、上位のルールが優先的に照合され、最初に条件を満足したルールの帰結を、その事例のクラスとする。決定リストにより表現された領域分割は、各If-Thenルールがカバーする領域から、そのルールより上位のルールがカバーする領域を差し引いた領域である。幾何学的イメージとしては、虫食い超立方体群である。

命題論理は、図-3（c）に示すようなルールで、単

一属性に関する条件を論理演算子ORや論理演算子ANDで結合して得られる論理命題である。どの命題も、一属性に関する条件を論理演算子ANDで結合し（これをタームという）、それらを論理演算子ORで結んだ形式（DNFという）に展開できる。したがって、命題論理で記述される領域分割は、各DNFでカバーされる領域群とカバーされない領域からなる。幾何学的イメージとしては、海に浮かぶ島である。結果的に、クラスの個数は2個限定であり、応用上は注意が必要である^{※2}

事例のみから学習を行う学習アルゴリズムには、ほかにニューラルネット学習やベイズ推定がある^{3), 9), 12)}。ニューラルネット学習では、領域分割が生物の神経結合を模倣したモデルで記述される。その幾何学的イメージは、複数の曲面による領域分割である。ベイズ推定では、領域分割が確率分布により記述される。その幾何学的イメージは、山と山の境目を境界とした領域分割である。許容される属性ベクトルの成分は、シンボルまたは数値である。ニューラルネット学習とベイズ推定も、候補モデル空間を探索することにより学習を進める。ただし、モデルが数式で定義され、解析的手法により探索を行う。結果として、領域分割が数式表現されるため、応用の立場からすると、その解釈が難しい。

^{※2} 決定木/決定リストは、3個以上のクラスに対応でき、未知事例のクラスも一意に決まる。これに対して、2個のクラスしか持たない学習アルゴリズムで、3個以上のクラスに対応するには、各クラスごとに、「そのクラス」と「それ以外のクラス」に訓練事例を分けて、学習させ、クラスごとに知識を作成すればよい。ただし、このように対処しても、未知事例のクラスが複数出力されたり、逆に、どのクラスの知識にも該当しない現象が生じ得る。

2.3 決定木の学習

本節では、決定木学習を取り上げ、例からの学習プロセスのアウトラインを示す。決定木の学習は、1) 学習結果が人間に可読であり、業務知識と照合できる、2) ニューラルネットや多変量解析では直接には扱えない、シンボルに基づく知識表現が可能である、3) 3個以上のクラスを許容できる（これを、マルチクラス学習アルゴリズムと呼ぶ）、4) 未知事例に対して必ず1つのクラスが決まり曖昧性がない、などの理由から、広く実問題に適用されている。

一般に、例からの学習におけるモデル探索空間は、可能なすべてのモデル（具体的には、たとえば、生成し得る決定木全体の集合）からなり、そのモデルの構造的複雑さに基づいた、半順序構造が入っている。決定木の場合のモデル空間は、図-4に示すとおりで、ある決定木は、すぐ上の決定木上のある葉を分岐させた構造である。

モデル探索空間は巨大であり、探索空間全体を調べるのは不可能である。そのため、探索空間を部分的に探索することになる。例からの学習で最もよく利用される探索方法は山登り探索で、探索空間上のあるポイントを起点にし、その起点のモデルより入力事例を旨くカバーしているモデルが半順序構造の直下であれば、そのモデルを起点に同様の操作を繰り返し、改善が望めなくなったら最終起点が探索結果となる。事例のカバーの良さを測るための評価関数は、学習アルゴリズムによりさまざまである。

最もよく知られたQuinlanの決定木学習アルゴリズムID3を用いて、上述の山登り探索をシミュレートする¹⁵⁾。ほかの代表的な決定木学習には、CARTがある¹⁾。CARTも、ID3と同様によく知られており、応用によっては、ID3を上回る性能を示している。

今、事例（訓練事例）として、次の属性ベクトルの集合Sが入力されたとしよう。

- $$S = \{(\text{低い}, \text{ブロンド}, \text{青}, +),$$
- $$(\text{低い}, \text{黒}, \text{青}, +),$$
- $$(\text{高い}, \text{黒}, \text{茶}, -),$$
- $$(\text{高い}, \text{ブロンド}, \text{茶}, -),$$
- $$(\text{高い}, \text{黒}, \text{青}, -),$$
- $$(\text{高い}, \text{ブロンド}, \text{茶}, -),$$
- $$(\text{高い}, \text{赤}, \text{青}, +),$$
- $$(\text{高い}, \text{ブロンド}, \text{青}, +)\}$$

属性は、背丈、髪の色、目の色、クラスの4つで、各属性の取り得る値は、順に {高い, 低い}, {黒, 赤, ブロンド}, {青, 茶}, {+, -} である。決定木の学習での探索の起点は、ルートノードだけからなる決定木である。このポイントに隣接する探索ポイントは、

表-1 決定木学習の実問題への適用

パターン認識問題	ハッブル宇宙望遠鏡画像における銀河の弁別 (ジョンホプキンス大)
	ハッブル宇宙望遠鏡画像における宇宙線衝突の弁別 (ジョンホプキンス大学)
	パロマ天文台による天体写真から、星などのカタログを作成 (カリフォルニア工科大)
	航空写真から道路を検出 (マサチューセッツ大)
	テキスト読み上げにおける区切りルールの自動学習 (AT & Tベル研)
診断問題	天秤の釣り合い過程の学習メカニズムの解明 (カナダ・デルハウジー大)
	イオンクロマトグラフ検出器の選択 (オーストラリア・ニューサウスウェールズ大)
	腫瘍の弁別知識の学習 (ベルギー・エラスムス病院)
戦略学習	オクタン値推定と分子部分構造検出 (オーストラリア・ヨハンスケプラー大)
	大気中のオゾン濃度予測 (カナダ環境庁)
	艦船用対空防衛兵器の利用モデルの獲得 (米海軍大学院)
	DSSの戦略獲得支援 (モントリオール大)
その他	複合型電力システムの安全管理 (ギリシャ工科大)
	ソフトウェアの開発工数見積り (米国のコンサルタント会社)
	テキスト分類およびキーワード抽出 (富士通)
	図書館における本の利用予測 (ハーバード大)
	自律型エージェントの知識獲得 (オーストラリア・ニューカッスル大)

このルートノードから、枝分れを起こした3つの決定木である。ID3の場合、カバーの良さを測る評価関数として、情報量の期待値を利用する。各枝分れを起こした時に、どれだけ情報量が増えるかをみるわけである。各々の木について評価値を計算すると、「髪の色は何色か?」という質問で、分岐した決定木が一番良いことが分かる。以上の一連の操作を、期待情報量が改善されなくなるまで続けることで、決定木が完成する。

3. 決定木学習アルゴリズムの応用

本章では、最近の海外ジャーナル論文から、決定木学習アルゴリズムの応用例を紹介する。表-1に応用例をまとめた。大きく分けて、パターン認識、診断問題、戦略の学習がある。なお、より以前の応用例は、文献7)を参照されたい。

機械学習がこのように期待をもって迎えられているのは、1つには、判断対象が複雑化・大規模化して、人手による判断知識作成では対応できなくなってきているためと思われる。また、例からの学習アルゴリズムの中でも、ID3やCARTが多用されているのは、ID3/CARTの学習時間が訓練事例数に比例しており、

高速である点も大きな要因であると思われる。

3.1 パタン認識

画像認識手法の中に、画像から交点の個数や、曲線タイプなどを抽出して認識に利用するモデルベース認識がある。画像が特徴パラメータ化されるモデルベース認識では、決定木は効果的であり、いくつかの適用例が報告されている。

この中で、最も活発な研究は、ハッブル宇宙望遠鏡、および、パロマ山天文台のシュミットカメラにより撮影された天文画像の自動分析である。この分野では、あまりに膨大なデータ量ゆえに、従来の人手による解析手法が無効化しつつあり、機械学習への期待が大きい。カリフォルニア工科大のWeirらは、パロマ山天文台が撮影した天文画像から、星/星雲/人工物を弁別するために、学習アルゴリズムを利用した¹⁶⁾。識別に利用された属性数は8個、識別精度は90%を超える。学習アルゴリズムは、ID3を改良したものである。ただし、天文画像がきわめて薄く、人間ですら認識が困難なため、そのままでは訓練事例として利用できない。そこで、高感度CCDによる撮影データを併用して、訓練事例を獲得している。

また、ジョンホプキンス大のOwenらは、ハッブル宇宙望遠鏡が撮影した画像において、従来は自動化が困難といわれていた、銀河のクラス分け（銀河は、その形態により、楕円型、凸レンズ型、渦巻き型などに分けられる）を行った¹⁰⁾。識別には、画像から測定された13個の属性が利用され、精度は約65%であった。学習アルゴリズムは、ID3の改良版を利用している。また、同じくジョンホプキンス大のSalzbergは、宇宙線が撮像素子に衝突することにより生じた、撮影乾板上のキズを、機械学習により抽出する研究を行っている¹¹⁾。学習アルゴリズムは、上記のOwenと同一である。属性20個を利用し、精度は95%である。

一方、AT&Tベル研究所のHirschbergらは、テキスト読み上げシステムにおける「フレーズ」ルールの自動獲得にCARTを利用している⁵⁾。フレーズは、読み上げの区切りやイントネーションを変化させ、聴き手が受け取る文の意味を変更させる。

学習には、音声的な特徴をテキスト上にあらかじめ人手で付加したテキスト（コーパス）を必要とする。しかし、このようなコーパスの入手は困難である。そこで、Hirschbergらは、テキスト側のみにフレーズ情報を人手で作成し、そのほかの属性は、テキスト分析により自動生成するアプローチを用いた。これにより、約9万単語のコーパスを数日で作成でき、95.4%の精度を確保できたという。

3.2 診断問題

診断問題への適用は、従来から試みられている応用分野であるが、最近も、引き続き、いくつかの研究が報告されている。

従来とはやや異なった応用としては、カナダ環境庁が進めている、オゾン濃度予報への決定木学習の適用がある²⁾。大気中のオゾン濃度予測には、従来、重回帰分析が利用されていたが、線形モデルでは近似が不十分なこと、また、気象上の測定パラメータ数が膨大となって、もはや、人手では予報に必要なパラメータの判定が困難になったことによる。このため、非線形の統計的推定ツールとしてCARTを採用した。属性数は57である。

CARTは良い性能を示し、学習結果は、物理的意味としても納得のいくものであった。決定木が、警報レベル（オゾン濃度80ppb以上）と予想したケースのうち、実際に80ppb以上となるか、または、警報に近い60ppb~79ppbとなったのは、92%に達した。CARTの性能は、実用可能なポテンシャルと結論付けている。

3.3 戦略学習

決定木は、リスクマネジメントなどの戦略を記述する表現法として広く利用されてきた。その意味では、戦略を表す決定木を自動生成するのは、自然なアプローチであり、従来から、ローンの設定可否の判定などに利用されてきた⁷⁾。最近の応用としては、以下のものがある。

スタンフォード大のSilversteinおよびハーバード大学のShieberは、図書館の蔵書の一部を遠隔の倉庫に保管する場合について、倉庫に保存する本を選ぶ戦略をID3で作成した¹³⁾。属性数6と少ないが、図書分野コードなどを属性としているために、1つの属性がとりうる値は数百種ある。ハーバード大学Widener図書館では、480万冊の叢書を持ち、1985年に20%を倉庫に移したが、その後年間34000回の取り出しを余儀なくされている。これを、ID3を利用した判定により、年間6200回に抑えられると結論付けている。

また、別の応用として、ソフトウェアの工数を見積もる研究がある¹⁴⁾。CARTの改良版を利用して、63個のソフト開発プロジェクトのデータを利用して、15個の別の開発データで評価した。属性数は33個であり、得られた推定ソフト工数と実際の工数の相関は、0.56となった。この値は、タスクそのものの困難性を考慮すれば、良好と思われる。

4. 実問題への適用における課題

実問題に機械学習技術を適用しようとする、実問題固有の種々の課題に遭遇する。本章では、その課題である、(1) 事例の量、(2) 事例の表現と質、(3) アルゴリズムの選択、(4) アルゴリズムのインプリメント、について述べる。

【事例の量】

未知事例に対する正解率を上げるには、分割境界を注意深く選ぶ必要があり、そのためには、大量の事例を必要とする。しかし、現実にはそれほど多くの事例を準備できないことが多い。たとえば、優れた故障診断システムの作成には、多数の故障事例を必要とする。しかし、多数の故障事例が集まるまで、故障診断システムを出荷しないのは自己矛盾である。また、医療診断では、真の病名を知るには、解剖などを必要とすることがある。しかし、これは、常に可能な行為とは限らない。

この問題は、機械学習アルゴリズムの実用化に際して、意外に大きな問題であり、学習アルゴリズム性能向上を目指す原動力となっている。この問題を解決するため、学習アルゴリズム側でも、人間の作成した知識と事例を融合する手法などの対策が提案されている。また、訓練事例の収集を支援する仕組みを考へることもきわめて重要である。

【事例の表現と質】

例からの学習では、基本的には属性値に誤り（ノイズ）がないことを前提とし、分類に必要な属性は、過不足なく与えられるとしている。しかし、現実のデータには、(1) 属性値およびクラス値へのノイズの混入、(2) 数値データの存在、(3) 最終的な獲得知識に含まれるべき必須属性の欠如、(4) 過剰な非必須属性の存在、などの課題がある。

このような事例は学習アルゴリズムに入力できないか、入力できても未知事例に対する正解率が低い。このため、実用学習アルゴリズムでは、ノイズ、数値データ、同時に複数の値を持つ属性、構造を持つ属性値、などに対処する工夫を行っている。なお、非必須属性は、学習過程で自動的に排除されるが、不要な属性が多いほど、ある一定性能を確保するために必要な訓練事例数は増加する。また、必須属性が抜けていれば、どんな学習アルゴリズムも対処できない。精度の高い知識を獲得するには、属性の選定がきわめて重要である。

【アルゴリズムの選択】

図-1に示した領域の分割では、ある評価関数を最小化するが、モデル空間が巨大なため、最適な解を求め

ることは不可能である。したがって、ID3、CARTがある程度オールラウンドに優れた性能を示すとしても、全応用に最高の性能を持つアルゴリズムは存在せず、ある応用に最適なアルゴリズムをあらかじめ予想すること（これは「モデル選択」と呼ばれる）もできない。

したがって、個別の応用に最適な学習アルゴリズムは、実験的に選択するのが現実的であり、表-1のように種々の学習アルゴリズムが開発されている理由の1つはそこにある。しかし、より複雑な問題を扱うためには、分割の表現（決定木など）や学習戦略を動的に選択する必要も出てくる。今後、研究の必要な分野であり、自動モデル選択の試みも始まっている。

【アルゴリズムのインプリメント】

実際にインプリメントするには学習アルゴリズムの詳細について熟知する必要があるが、実問題への適用を試みる技術者/研究者にはネックとなる。これに対しては、QuinlanはID3の優れた汎用パッケージC4.5を提供しており、Quinlanの本を購入すれば利用できる⁴⁾。また、ほかの学習アルゴリズムについても、次回紹介するように、主要なアルゴリズムは、FTPで入手できる。その意味では、応用に学習アルゴリズムを適用する環境は、十分に整ってきたといえよう。

5. おわりに

機械学習技術、特に「例からの学習」技術について、例を交えて解説した。また、どのような応用分野に適用されているかを最近の応用例を通して紹介した。現実の問題に機械学習を適用する際には、(1) 事例の量、(2) 事例の表現と質、(3) アルゴリズムの選択、(4) アルゴリズムのインプリメントなどの課題がある。今回は、これら課題への対処法について紹介する。

謝辞 日頃ご指導・ご鞭撻をいただき、NTTコミュニケーション科学研究所各位、ならびに、King Fahd University of Petroleum and Minerals各位に深謝いたします。

参考文献

- 1) Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. : Classification and Regression Trees. The Wadsworth & Brooks/Cole Advanced Books & Software, California (1984) .
- 2) Burrows, W., Benjamin, M., Beauchamp, S., Load, R. E., McCollor, D. and Thompson, B. : CART Decision-Tree Statistical Analysis and Prediction of Summer Season Maximum Surface Ozone for the Vancouver, Montreal, and Atlantic Regions of Canada, J. of Applied Meteorology, Vol.34., pp.1848-1862 (1995) .
- 3) Duda, R. O. and Hart, P. E. : Pattern Classification and Scene Analysis, John Wiley & Sons Inc., New York (1973) .
- 4) 吉川康一監訳 (Quinlan著) : AIによるデータ解析, トッパン (1995) .

- 5) Hirschberg, J. and Prieto, P. : Training Intonational Phrasings Rules Automatically for English and Spanish Text-to-speech; *Speech Communication*, Vol.18, pp.281-290 (1996) .
- 6) Langley, P.: *Elements of Machine Learning*, Morgan Kaufmann, California (1996) .
- 7) Langley, P. and Simon, H. A.: *Application of Machine Learning and Rule Induction*, *Communications of The ACM*, Vol.38, No.11, pp.55-64 (1995) .
- 8) Lavrac, N. and Dzeroski, S.: *Inductive Logic Programming*, Ellis Horwood, New York (1994) .
- 9) Mitchell, T. M.: *Machine Learning*, The McGraw-Hill Companies Inc., New York (1997) .
- 10) Owen, E. A., Griffiths, R. E. and Ratnatunga, K. U. : Using Oblique Decision Trees for the Morphological Classification of Galaxies, *Mon. Not. R. Astron. Soc.*, Vol. 281, pp.153-157 (1996) .
- 11) Salzberg, S., Chandar, R., Ford, H., Murthy, K. S. and White, R. : Decision Trees for Automated Identification of Cosmic-Ray Hits in Hubble Space Telescope Images, *Publications of the Astronomical Society of the Pacific*, Vol. 107, pp.279-288 (1995) .
- 12) Shavlik, J. W. and Dietterich, T. G.: *Readings in Machine Learning*, Morgan Kaufmann, California (1990) .
- 13) Silverstein, C. and Shieber, S. M. : Predicting Individual Book Use for Off-site Storage using Decision Trees, *Library Quarterly*, Vol.66., No.3., pp.266-293 (1996) .
- 14) Srinivasan, K. and Fisher, D. : Machine Learning Approaches to Estimating Software Development Effort, *IEEE Trans. on SE.*, Vol.21, No.2, pp.126-137 (1995) .
- 15) Quinlan, J. R.: *Induction of Decision Trees*, *Machine Learning*, Vol.1, pp.81-106 (1986) .
- 16) Weir, N., Fayyad, M. U., Djorgovski, S. G. and Roden, J. : The SKICAT System for Processing and Analyzing Digital Imaging Sky Surveys, *Publications of the Astronomical Society of the Pacific*, Vol. 107, pp.1243-1254 (1995) .

(平成9年9月8日受付)



秋葉 泰弘 (正会員)

1964年生。1988年早稲田大学教育学部数学科卒業。1990年同大学院理工学研究科数学専攻修士課程修了。同年、日本電信電話(株)入社。以来、機械学習、知識獲得、ニューラルネットの研究に従事。1995年人工知能学会全国大会優秀論文賞受賞。現在、NTTコミュニケーション科学研究所研究主任。
e-mail: akiba@cslab.kecl.ntt.co.jp



フセイン・アルモアリム

1961年生。1984年東京工業大学工学部電子物理学科卒業。1986年同大学院工学研究科修士課程修了。1992年オレゴン州立大学計算機科学科博士課程修了。同年、NTT情報通信網研究所勤務。現在、サウジアラビア国立石油鉱物大学・計算機学科助教授。Ph.D. in Computer Science。機械学習、文字認識などの研究に従事。1991年AAAI Honorable Mention Award 受賞。1995年人工知能学会全国大会優秀論文賞受賞。AAAI会員。e-mail: hussein@ccse.kfupm.edu.sa



金田 重郎 (正会員)

1951年生。1974年京都大学工学部電気第二学科卒業。1976年同大学院電子工学専攻修士課程修了。同年、日本電信電話公社・武蔵野電気通信研究所入所。以来、誤り訂正符号、フォールトトレラント技術、知識獲得、エキスパートシステムなどの研究に従事。1997年より同志社大学大学院総合政策科学研究科・同志社大学工学部教授。工学博士、技術士(情報処理部門)。電子情報通信学会、IEEE、人工知能学会各会員。e-mail: skaneda@mail.doshisha.ac.jp