

ノイズを含む事例からのルール抽出: RF3 アルゴリズム†

齊藤 和巳^{††} 中野 良平^{††}

知識獲得はエキスパートシステム構築の最重要課題である。エキスパートからの知識(ルール)抽出は困難で、可能な限りの自動化が強く望まれている。本論文では、応用の広い分類問題を対象に、ノイズを含む事例から、高い信頼性で良い正答率を保証する簡潔な分類ルールの抽出を目標とする RF3 法を提案する。RF3 法の特長は、ルール集合選択のための新評価尺度 MEF の導入にある。MEF 尺度は、任意の汎化誤り率の許容限界に対し、抽出したルール集合の汎化誤り率が許容限界より悪くなり、抽出失敗となる確率の期待値の最小化を意図したものである。また、MEF 尺度は、抽出したルール集合の複雑さと例外事例の個数の和を最小化する尺度としても解釈可能である。RF3 法の能力を検証するための実験を行った。実験の範囲では、RF3 法が抽出したルール集合は、事例を生成するのに用いた元のルール集合とほぼ同等なものであった。また、訓練誤り率の許容限界を順次変化させ、MEF 尺度と別途推定した汎化誤り率を比較したところ、両曲線が極めて類似することが明らかとなった。したがって、MEF 尺度を用いることにより、高い信頼性で汎化誤り率を最小にするルール集合の選択が可能になると期待できる。

1. はじめに

知識獲得はエキスパートシステム構築の最重要課題である。エキスパートからの知識(ルール)抽出は困難で、可能な限りの自動化が強く望まれている。特に、事例からのルール抽出は広く研究され、その現実問題への応用が強く期待されている。ところが、現実問題では、事例がノイズを含むことを想定しなければならない。したがって、ノイズを許容しつつ、高い信頼性で良い正答率を保証する抽出法が重要になる。

ノイズを含む事例からのルール抽出(仮説選択)の理論的な枠組みとしては、少なくとも、以下の3種概念学習へのアプローチが知られている。

第1に、PAC(Probably Approximately Correct)学習に基づく方法がある²⁹⁾。PAC学習では、事例が一定の確率分布に従って毎回独立に出現すること、および、仮説の属する概念クラスが仮定される。そして、汎化誤り率とアルゴリズムの失敗確率の各許容限界を設定すれば、仮定した概念クラスの複雑さと2つの許容限界の値から、学習結果を保証するのに必要な訓練事例数を求めることができる。しかし、PAC学習の評価尺度は最悪ケースの解析に基づいているので、一般に、理論を適用するにはかなりの数の訓練事例が必要となる。ところが、一般の問題では、十分に訓練事例を得られない場合が多い。したがって、その解析を平均ケースにすることにより、必要とする訓練

事例数を減らすことができれば、より実用的な尺度が構築できると考える。さらに、ルール抽出では、与えられた訓練事例数に依存した評価尺度が求められるが、PAC学習では、学習に必要な訓練事例数が得られるだけである³⁾。

第2に、MDL(Minimum Description Length)原理に基づく方法がある²⁰⁾。MDL原理では、仮説の属する概念クラスが仮定され、仮説と例外事例をコード化する方法が与えられる。各コードに対しては、コード長が短くなるにつれ大きな確率を割り当てる。そして、仮説とその仮説の下での例外事例に割り当てた確率の和を求め、その値を最大にする仮説を探索し、最終結果とする。なお、仮説と例外事例に割り当てた確率は、Bayesianの立場から、事後確率を求めるための事前確率と条件付き確率と解釈できる。ところが、MDL原理では、各コードへ確率を割り当てるのに必要なパラメータの値を定めねばならない。一般に、その値の決定は困難な問題として残されている。

第3に、再サンプリング(resampling)法に基づくbootstrap, cross-validation等の方法がある⁶⁾。訓練事例を2つに分割(再サンプリング)して、一方を仮説の選択、他方を仮説の評価に用いれば、すべての訓練事例を用いて選択する仮説の汎化誤り率を推定できる。しかし、1回の実験では、高い信頼性で推定できないので、再サンプリングを繰り返して仮説選択を行うことにより、推定値の信頼性を高めていく。すなわち、1つの仮説の評価に、数十から数百回の再サンプリングが必要である⁶⁾。最終的にルール抽出するには、仮説の生成だけでなく、複数の仮説からの仮説選択という負荷の重い処理が必要になる。各仮説に対し

† Rule Extraction from Noisy Facts: RF3 Algorithm by KAZUMI SAITO and RYOHEI NAKANO (NTT Communication Science Laboratories).

†† NTT コミュニケーション科学研究所

て、再サンプリングによる評価を行えば、全体ではかなりの計算量が必要となる。

これら3つのアプローチに基づいて、ノイズを含む事例からルール抽出するいくつかのアルゴリズムが提案されている^{17), 18), 19)}。いずれも、理論的な課題の解明、または、決定木 (decision tree) の構築に関するものである。決定木の構築では、プルーニング (pruning) と呼ばれるいくつかの効率的なアルゴリズムが提案されている¹⁸⁾。しかし、決定木では、それ自身の表現能力が不十分である¹⁶⁾。例えば、 $x_1x_2 \vee x_3x_4x_5$ のような和結合した概念 (disjunctive concept) の表現が困難である。したがって、より実用的なルール抽出の枠組みを設定し、高い信頼性で良い正答率を保證するルール抽出法の提案が望まれている。

本論文では、応用の広い分類問題を対象に、ノイズを含む事例から、高い信頼性で良い正答率を保證する簡潔な分類ルールの抽出を目標とする RF3 法の提案を行う。なお、RF3 法は RF 法²¹⁾ (RF2 法はその改良版) をノイズを含む環境化でも適用できるように拡張したものである。まず、ルール抽出の枠組みについて述べる。次いで、複数のルール集合から適当なルール集合を選択するための新評価尺度 MEF (Minimum Expected Failure) を提案し、さらに、その位置付けについて述べる。次いで、RF3 法のアルゴリズムの概略について述べる。最後に、RF3 法の実験結果について述べる。

2. フレームワーク

分類ルールの抽出を行う枠組みについて説明する。

概念 (concept) は、その概念の基本構成要素となる属性 (attribute) によって表現される。例えば、病名という概念は {体温, 血圧, 性別, ...} という属性によって表現できる。一般に、属性を用いて概念を表現する形式は論理式のクラスとして定式化される。ここではその概念クラスとして選言標準形 (DNF: Disjunctive Normal Form) を仮定する。DNF は、人間にとっても直感的で理解しやすく、ルール抽出を考えるのに適当な概念クラスであると思われる。

DNF は項 (term) の論理和として表現され、項はリテラル (literal) の論理積として表現される。各リテラルは対応する1つの属性にのみ関係し、その属性の値によって、リテラルの値は真または偽となる。本論文では、項を1つのルール、DNF をルールの集合とみなす。

概念クラスを仮定すれば、そのクラスに属す任意の概念は、属性を用いた論理式として表現できる。したがって、それらをあらかじめ仮説の集合 $\{f_1, f_2, \dots, f_n\}$ として生成することにより、概念学習は仮説の選択問題として定式化できる。しかし、一般に、生成し得る仮説の個数は莫大であり、到底すべてを実際に数え上げることはできない。

事例は、分類したい概念に属す事例と属さない事例からなる。前者を正の事例、後者を負の事例と呼ぶ。例えば、正の事例はその病名の患者であり、負の事例は非患者である。そして、各事例は属性空間内の点として位置付けられる。本稿では、各事例は任意の分布 P_D に従って毎回独立に出現すると仮定する。事例 e_i は (x_i, y_i) で表せる。ここで、 x_i は属性値からなるベクトル、 y_i は事例の正負を示すラベルを表す。事例の集合 $\{e_1, e_2, \dots, e_E\}$ を訓練事例と呼ぶ^{*}。なお、属性値はノイズを含むとする。

3. 新評価尺度

以下の議論で用いるいくつかの表記法を導入する⁶⁾。まず、任意の事例 e_i と仮説 f_n に対して損失関数 L を以下のように定める:

$$L(e_i, f_n) = \begin{cases} 0 & \text{if } y_i = f_n(x_i) \\ 1 & \text{otherwise.} \end{cases}$$

訓練事例 $\{e_1, e_2, \dots, e_E\}$ に対する仮説 f_n の誤り率 err は以下のように表現できる:

$$\text{err}(\{e_i | 1 \leq i \leq E, f_n\}) = \sum_{i=1}^E \frac{L(e_i, f_n)}{E}.$$

今後、この値を訓練誤り率と呼ぶ。確率分布 P_D に従って新たに出現する事例 e に対する仮説 f_n の平均誤り率 Err を以下とする:

$$\text{Err}(f_n) = E(L(e, f_n))$$

今後、この値を汎化誤り率と呼ぶ。

複数のルール集合から適当なルール集合を選択するための MEF 尺度を導く。ルール抽出では、高い信頼性で汎化誤り率をできる限り小さくすることが望まれる。そこで、これらを定量的に表現するため、汎化誤り率の許容限界を ε 、実際の汎化誤り率が ε 以上となり、アルゴリズムが失敗となる確率を δ とする。望ましい評価尺度の1つは、任意の汎化誤り率の許容限界 ε に対して、アルゴリズムの失敗確率 δ の期待値を最

* ノイズを含む事例からのルールの抽出では、各事例の出現頻度が重要な意味を持ち、重複した事例の存在を許容しなければならない。したがって、厳密には、訓練事例の集合は、重複を許した多重集合 (multi-set) として定義される。

小にすることである。この期待値を直接計算することは困難であるが、期待値の良い近似としてその上限を次の命題より求めることができる。なお、この結果は Devroye が証明した定理⁹⁾の1つの系として示すことも可能である。

補題 F を仮説 $\{f_i\}_{1 \leq i \leq F}$ の個数、 E を訓練事例 $\{e_i\}_{1 \leq i \leq E}$ の個数、 ε を選択した仮説の汎化誤り率の許容限界、 ε' を選択した仮説の訓練誤り率の許容限界とする。各事例が任意の分布 P_D に従って毎回独立に出現する場合、もし、選択した仮説 $f \in \{f_i\}$ が、少なくとも、 $(1-\varepsilon')E$ 個の訓練事例を正答すれば、 f の汎化誤り率が ε より大となるアルゴリズムの失敗確率 δ の上限は、

$$\delta \leq Fe^{-2(\varepsilon' - \varepsilon)^2 E},$$

となる。また、任意の汎化誤り率の許容限界 ε に対するアルゴリズムの失敗確率 δ の期待値の上限は、

$$E(\delta) \leq \varepsilon' + \sqrt{\frac{\log F}{2E}} + \sqrt{\frac{1}{8E \log F}}, \quad (1)$$

となる。

証明 汎化誤り率が ε より大である仮説 f_i が、少なくとも、 $(1-\varepsilon')E$ 個の訓練事例を正答する確率は、二項展開式の $\varepsilon'E$ までの和でおさえられる：

$$P(\text{Err}(f_i) > \varepsilon) \leq \sum_{k=0}^{\varepsilon'E} \binom{E}{k} \varepsilon^k (1-\varepsilon)^{E-k}.$$

したがって、選択した任意の仮説 f に対する失敗確率 δ の上限は、Hoeffding の不等式¹¹⁾ を用いて求められる：

$$\begin{aligned} \delta &= P(\text{Err}(f) > \varepsilon) \\ &\leq F \sum_{k=0}^{\varepsilon'E} \binom{E}{k} \varepsilon^k (1-\varepsilon)^{E-k} \\ &\leq Fe^{-2(\varepsilon' - \varepsilon)^2 E}, \end{aligned}$$

失敗確率 δ の期待値は、 ε で定積分することにより求められる：

$$E(\delta) = \int_0^1 P(\text{Err}(f) > t) dt$$

したがって、上記の結果と Gordon の不等式⁹⁾を用いれば以下の結果を得る：

$$\begin{aligned} E(\delta) &\leq \int_0^u dt + \int_u^\infty Fe^{-(\varepsilon' - t)^2 E} dt \\ &\leq u + \sqrt{\frac{1}{8E \log F}} \end{aligned}$$

ここで、

$$u = \varepsilon' + \sqrt{\frac{\log F}{2E}}.$$

□

(1)式で、第2項に対する第3項の比を計算すれば $1/2 \log F$ となる。ここで、 $F=100$ としても、 $1/2 \log F < 0.11$ なので、現実的なルール抽出では、安全に第3項を無視できる。したがって、本稿で扱う MEF 尺度としては以下を用いる：

$$\varepsilon' + \sqrt{\frac{\log F}{2E}}. \quad (2)$$

MEF 尺度が過剰評価となる可能性について考察する。例えば、

$$\varepsilon' + \sqrt{\frac{\log F}{2E}} \ll \frac{1}{4}$$

の場合、Hoeffding の不等式は $\varepsilon < 1/4$ の範囲で、Chernoff の不等式に置き換えることがより厳密な評価となる⁷⁾。また、もし、

$$\varepsilon' \gg \sqrt{\frac{\log F}{2E}}$$

である場合には、0から

$$\varepsilon' - \sqrt{\frac{\log F}{2E}}$$

までの積分値を考慮すべきである。しかし、現実には、仮説の数に比して十分に訓練事例が得られないことが一般的であり、実用的にはこれらのケースを安全に無視できる。

4. MEF 尺度の位置づけ

MEF 尺度の仮定の1つは、機械学習の分野では標準と見なされてきた PAC 学習と同じものである。すなわち、各事例が任意の一定の分布に従って毎回独立に出現することを仮定している (distribution-free learning)。ここで、既に指摘したように、PAC 学習の問題点の1つは、その評価が最悪ケースの解析に基づくものであり、十分な数の訓練事例を得られず適用困難な場合が起こることであった。一方、MEF 尺度では、任意の汎化誤り率の許容限界に対して、抽出アルゴリズムが失敗する確率の期待値を最少にすることを試みている。したがって、より現実の問題に適した評価尺度を構築していると考えられる。

MEF 尺度は、例外事例の個数とルール集合の複雑さの和を最小化する MDL 原理からみると、次のように解釈できる。訓練事例数 E が決まれば、その値を (2)式に乗ずることにより、

$$\varepsilon'E + \sqrt{\frac{E \log F}{2}}, \quad (3)$$

を得る。ここで、第1項は例外事例の個数の上限にはかならず、第2項はルール集合の複雑さに関する項と

なる。なぜなら、第2項の変数は F だけであり、仮説の個数が多いということは、必然的に複雑なルール集合を仮定していることになる。

計算量の観点から MEF 尺度を考えると、1つの仮説の評価には、(2)式の計算だけで十分である。したがって、再サンプリング法と比較すれば、各仮説の評価に必要な計算はかなり効率的なものである。

5. RF3 アルゴリズム

5.1 訓練誤り

ノイズを含む事例からルール抽出を行う場合、一般に、すべての訓練事例を正答させることにはあまり意味がない。抽出アルゴリズムが例外事例をも説明することを試みて、必要以上に複雑なルール集合を生成してしまうからである。そこで、抽出アルゴリズムに対して、訓練誤りを許容するため、訓練誤り率の許容限界 (MPE: Maximum Permissible Error) を導入する必要がある。

訓練誤りは2種の誤りに分類できる²²⁾。すなわち、FP (false-positive) 誤りと FN (false-negative) 誤りである。ここで、FP 誤り率とは、訓練に用いたすべての負の事例のうち、少なくとも1つ以上のルールにカバーされてしまう負の事例の割合である。一方、FN 誤り率とは、訓練に用いたすべての正の事例のうち、どのルールにもカバーされない正の事例の割合である。一般に、ルール集合の複雑さを一定にして、訓練誤り率を最少にすることを試みれば、この2種の誤りはトレードオフの関係にある。

この2種の誤りに対しても、それぞれ、FP 許容限界と FN 許容限界が導入できる。ところが、FP 許容限界では、最終的に抽出されるルールがすべて確定するまで、実際の FP 誤りを計算できないので、RF3 法にとっては扱いにくいものとなる。そこで、各ルールごとにカバーする負の事例の割合を RFP (rule false-positive) 誤り率として定義し、これに対して RFP 許容限界を導入する。なお、RFP 誤り率の逆数はルールの確信度と考えることもできる。

5.2 ルール集合の複雑さ

本稿では、ルールの集合として DNF を仮定した。しかし、DNF で表現可能なルール集合の個数は莫大であり、その数を仮説の個数として採用することは現実的でない。そこで、抽出を試みる任意の仮説の複雑さは、抽出したルール集合のものと同程度以下であると仮定し、アルゴリズムにより抽出したルール集合の

複雑さから仮説の個数を推定することを考える。まず、 N を属性数とし、 K をルール集合に現れる条件 (リテラル) の総数とする。すると、そのルール集合を表現するのに必要なビット数の近似は：

$$K \times \log_2 N,$$

となる¹⁶⁾、したがって、新評価尺度を求めるのに使われる仮説の個数 F は：

$$\log F = K \times \log N,$$

として求めることができる。

5.3 RF3 アルゴリズムの詳細

RF3 法は、訓練誤り率の許容限界 MPE を順次変化させ、ルール集合を抽出し、その中で MEF 尺度を最小にするルール集合を最終結果とする。ルール集合の抽出は、ルール候補の生成、および、精練という2つのフェーズからなる。すなわち、正の事例の記述を逐次一般化することによりルール候補の集合を生成し、IDA*¹⁴⁾ と呼ばれる探索手法を用いてそれらを最適なルール集合に精練する。図1には、RF3 法のフローチャートを示す。

5.3.1 訓練事例からのルール候補の生成

第1フェーズでは、すべての訓練事例と RFP 許容限界が入力され、ルール候補の集合が出力される。まず、正の事例を順番に seed として選択する。次に、ヒューリスティクスを用いて正の事例を順次選択する

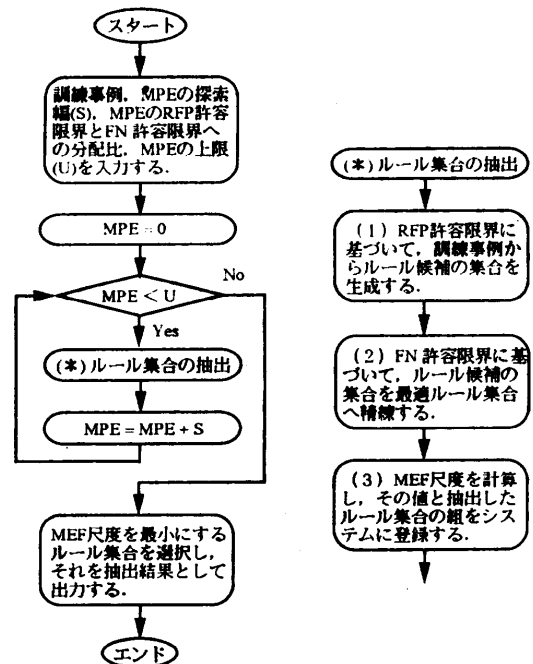


図1 RF3 法のフローチャート
Fig. 1 Flow chart of RF3 algorithm.

ことにより, seed の記述を最小限に一般化する. すなわち, 選択した正の事例をカバーし, カバーしてしまう負の事例の割合が RFP 許容限界以下となるよう一般化する. ただし, 条件を満たす一般化が存在しない場合には, seed の記述はそのままとなる.

一方, 従来方法^{9), 15)-17)}では, ヒューリスティック評価関数等を利用して属性を選択することにより, ルールを生成する. しかし, この属性選択方式を用いては, 例えば, 不要な属性を持つパリティ問題では, ルール抽出が困難となる.

5.3.2 ルール候補の最適ルールへの精練

第2フェーズでは, 第1フェーズで生成したルール候補の集合と FN 許容限界が入力され, 最適に精練されたルール集合が出力される. 第1フェーズ完了後, 一般に, ルール候補の集合には冗長なタームやリテラルが存在する. それらを除くことは, カバーしない正の事例の割合を FN 許容限界以下にする SC (set covering) 問題として定式化できる. しかし, SC 問題は NP 困難なので⁸⁾, 一般に, 厳密解を求めることは大変難しく, 多くの場合, 欲張り法 (greedy algorithm)¹⁸⁾ に代表される近似解法が採用される¹⁰⁾.

しかし, 欲張り法は, 最も多くの要素を含む集合から選択するので, 単純なケースでも最適解を得られないことがある. そこで, RF3 法では, 第1フェーズでルール候補と各リテラルがかなり絞り込める点に着目し, 第2フェーズで, IDA* アルゴリズム¹⁴⁾を用いて, 厳密解を探索する. IDA* は探索の閾値を逐次変えながら深さ優先探索を実行するので, 計算機のメモリを使い尽くすことなく, 大きな問題に対しても厳密解を求めることが可能である.

5.3.3 パラメータ

RF3 法を用いてルール抽出するには, ユーザに入力させる3種のパラメータが必要である. まず, 訓練誤りの許容限界 MPE は, 0を初期値とし, 一定の値を加えることにより変化させる. なぜなら, MPE を入力, MEF を出力と考えれば, RF3 法は, 一次元区間を定義域とするシステムの最小値探索となる. したがって, システムの特別な性質が明らかでないので, ユーザの指定した一定の探索幅で入力サンプル点を選り, システムの出力を調べればよいと考える. また, 一般に, この探索幅がシステムの性能と効率のトレードオフになると考える.

次に, MPE の RFP 許容限界と FN 許容限界への分配比である. 一般に, RFP 誤りと FN 誤りの深刻

度は適用問題に依存して異なっている. 例えば, 医療診断問題では, 初期診断で真の患者を病気でないと誤る FN 誤りは, RFP 誤りよりも深刻な問題になると思われる. したがって, この分配比の決定もユーザに任せることが望ましい.

最後に, MPE を変化させる区間の上限である. 一般に, 上限の設定は正負の事例の割合と関係がある. すなわち, この割合が偏っている場合, 新たに出てくる事例に対して, 事例の多いほうのクラスを常にその分類結果とすれば, 一般に, ある程度の正答率を得ることができる. ここで, 抽出したルールの正答率がそれ以下では意味がない. したがって, 正負の事例の割合から, 訓練誤りの許容限界 MPE を変化させる区間の上限の目安を得ることができる. 例えば, 正負の事例の割合が 1:2 の場合には, 上限の目安は 66% となる.

6. 実験

6.1 実験の設定

MEF 尺度を導入した RF3 法の能力を検証するため, 以下の実験を行った. 表1には, 以下で説明する実験設定の要約を示す.

実験には, mx6, par4, dnf4 と呼ばれる3つの概念を用いた¹⁶⁾. これら概念の性質は, mx6 が Multiplexor, par4 が Parity, dnf4 が Random DNF であり, 属性数に比較して簡潔な DNF 記述で表現でき, したがって, それぞれにはかなりの不要な属性がある. 図2には, 3つの概念の属性数と完全な DNF 記述を示す.

事例は, 一様分布に基づき毎回独立に生成した. すなわち, 各事例の属性値には, ランダムに0または1を与え, そして, その属性値ベクトルがそれぞれの

表1 実験設定の要約
Table 1 Summary of experimental design.

事例を発生させる分布 各属性へのノイズの与え方	一様分布		
	5%の割合で属性値を反転		
目標概念	mx6	par4	dnf4
訓練事例数	360	640	1,320
テスト事例数	5,000		
MPE を変化させる区間	0%~49%		
MPE を変化させる定数	1%		
RFP 許容限界と FN 許容限界	1:1 の分配比		

$$\begin{aligned}
 & \text{mx6: (16-attrs)} \\
 & \overline{x_1}x_2x_3 \vee \overline{x_1}x_2x_4 \vee \overline{x_1}x_2x_5 \vee \overline{x_1}x_2x_6 \\
 \\
 & \text{par4: (16-attrs)} \\
 & \overline{x_1}x_2x_3x_4 \vee \overline{x_1}x_2x_3x_5 \vee \overline{x_1}x_2x_3x_6 \vee \\
 & \overline{x_1}x_2x_3x_4 \vee \overline{x_1}x_2x_3x_5 \vee \overline{x_1}x_2x_3x_6 \vee \\
 & \overline{x_1}x_2x_3x_4 \vee \overline{x_1}x_2x_3x_5 \\
 \\
 & \text{dnf4: (64-attrs)} \\
 & x_1x_2x_3x_4x_5x_6 \vee x_1x_2x_3x_4 \vee x_2x_3x_4x_5x_6 \vee \\
 & \overline{x_1}x_2x_3x_4x_5 \vee \overline{x_1}x_2x_3x_4x_6 \vee \overline{x_1}x_2x_3x_4x_5x_6 \vee \\
 & x_1x_2x_3x_4x_5 \vee x_1x_2x_3x_4x_6 \vee x_1x_2x_3x_4x_5x_6 \vee \\
 & \overline{x_1}x_2x_3x_4x_5x_6
 \end{aligned}$$

図 2 目標概念の記述
Fig. 2 Target DNF descriptions.

ルール (DNF 記述) を満足するか否かにより, 正または負のラベルを与えた。

ラベルの付与された各事例に対しては, 各属性ごとに γ % の確率でその値を反転させることにより, ノイズを与えた。したがって, 矛盾する事例が出現する可能性もある。ノイズがあまりに大きいとアルゴリズムの優位性が分からなくなるので, ベンチマークの LED (LED display)²⁾ 等と比較して, γ を半分の 5% に設定した。

各概念における訓練事例数は, PAC 学習の計算式:

$$\frac{K \times \log N}{\eta} \quad (4)$$

に基づいて求めた。PAC 学習に関しては, 現実の問題では, 前もって知識の複雑さが分からず, 適用できないことが多いが, 人工的な問題では, アルゴリズム評価のための重要な指標を得ることができる。ここでは, PAC 学習は最悪ケースの解析がベースであることを考慮し, 同じ概念に対して行われた実験¹⁶⁾と比較して, 半分の訓練事例からでも有効なルール抽出ができることを確認するため, η を 2 倍の 20% に設定した。

汎化誤り率の推定には, 訓練事例とは独立にランダムに生成した 5,000 のテスト事例を用いた。ただし, テスト事例に対してもノイズを与えた。

RF3 法のパラメータに関しては, RFP 許容限界と FN 許容限界の分配比を等しくし, 訓練誤りの許容限界 MPE を 0% から 49% まで 1% 刻みで変化させた。なお, 実験に用いた概念では, 正負の事例の割合がほぼ等しいので, MPE の上限を 50% とした。

表 2 ルール集合の比較
Table 2 Comparison of rules.

概念	元のルール集合		抽出したルール集合	
	汎化誤り (%)	リテラル数	汎化誤り (%)	リテラル数
mx 6	9.9	12	9.9	12
par 4	16.6	32	18.5	33
dnf 4	11.6	41	15.9	30

6.2 実験結果

表 2 には, 3つの概念に対して, 元のルール集合と RF3 法が抽出したルール集合における汎化誤り率とルール集合のサイズの比較を示す。ここで, 元のルール集合とは, 事例を生成するために用いた元々の DNF 記述のことである。RF3 法が抽出したルール集合は, MEF 尺度を最少にするものであり, 表からも分かるように, 最適なルール集合とほぼ同等なものとする。

図 3 から図 5 には, 訓練誤り率の許容限界 MPE (x 軸) を順次変化させ, MEF 尺度 (MEF), 汎化誤り率 (Err) がそれぞれどのように変化したかを示す。なお, y 軸には, アルゴリズムの失敗確率の期待値 (MEF 尺度) と汎化誤り率の 2つの意味が当てられている。

6.2.1 mx 6 概念

図 3 には, 360 事例から mx 6 概念のルールを抽出したときの曲線を示す。この概念では, RF3 法は元のルール集合を完全に復元した。ノイズを含む事例からの抽出であることを考えると, このことは注目できる結果である。さらに, 汎化誤り率が最少になる部分は, MEF 尺度が最少になる場所と幅広く一致し, その範囲では, すべて元のルール集合を完全に復元していた。ここで, 同程度の複雑さを持つ他の概念 (ルール集合) と比較して, mx 6 概念には, 抽出が容易にできる特別な理由はないと考える。むしろ, 多くの属性選択方式のアルゴリズムを用いた場合には, 有効にヒューリスティック評価関数が働かず, 抽出が困難となる¹⁶⁾。したがって, RF3 法を用いれば, mx 6 概念と同程度の複雑さを持つシンプルなルールの場合には, ノイズを含む事例からでもかなり良いルールの抽出が期待できる。

6.2.2 par 4 概念

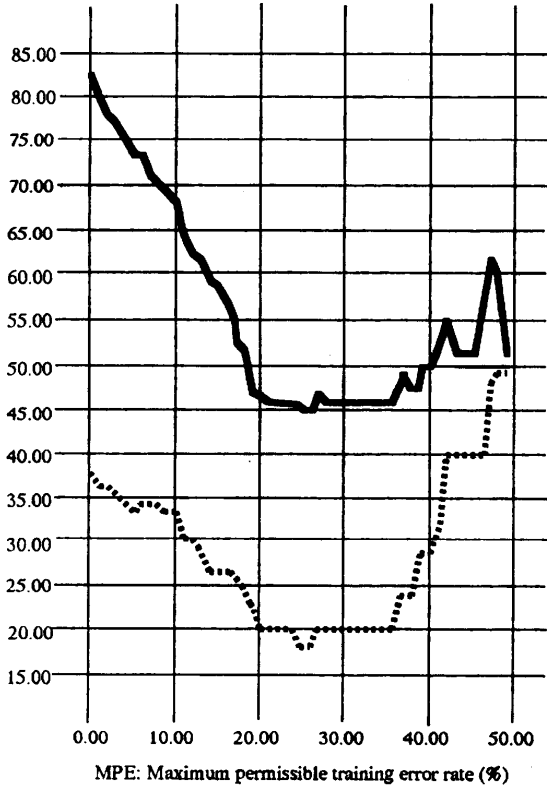
図 4 には, 640 事例から par 4 概念のルールを抽出したときの曲線を示す。この概念では, 1つのタームに 1つの余計なりテラルが現れたことを除き, 元の

MEF: MEF criterion (%) & Err: generalization error rate (%)



MPE: Maximum permissible training error rate (%)
 図 3 mx 6 概念に対する MEF 曲線と Err 曲線
 Fig. 3 MEF curve and Err curve for mx 6 concept.

MEF: MEF criterion (%) & Err: generalization error rate (%)



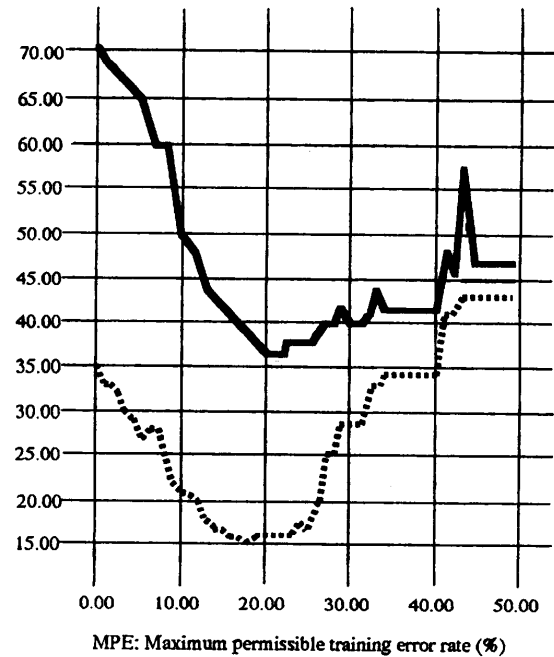
MPE: Maximum permissible training error rate (%)
 図 4 par 4 概念に対する MEF 曲線と Err 曲線
 Fig. 4 MEF curve and Err curve for par 4 concept.

ルール集合を復元できた。そして、汎化誤り率が最少になる点は、まさに、MEF 尺度が最少になる点と一致した。なお、パリティ概念に関しても、多くの属性選択法のアルゴリズムによる抽出が困難であることを考えると、RF 3 法の能力は注目し値する。

6.2.3 dnf 4 概念

図 5 には、1,320 事例から dnf 4 概念のルールを抽出したときの曲線を示す。この概念では、元のルール集合より、かなりシンプルなルール集合を抽出した。図 6 には、実際に抽出したルール集合を示す。元のルール集合と比較すれば、5つのリテラルからなるターム（ルール）を抽出できなかった。すなわち、2つのタームは完全に無視され、もう1つのタームでは、1つのリテラルが欠落してしまっただ。ここで、実験では、事例を一様分布に基づいて生成しているの、一般に、リテラルが1つ増えればカバーする事例は半分となる。したがって、カバーする事例数が少な

MEF: MEF criterion (%) & Err: generalization error rate (%)



MPE: Maximum permissible training error rate (%)
 図 5 dnf 4 概念に対する MEF 曲線と Err 曲線
 Fig. 5 MEF curve and Err curve for dnf 4 concept.

- $\bar{x}_{18}\bar{x}_{22}\bar{x}_{24} \vee \bar{x}_{30}\bar{x}_{46}\bar{x}_{48}\bar{x}_{58} \vee \bar{x}_9\bar{x}_{12}\bar{x}_{38}\bar{x}_{55} \vee$
- $\bar{x}_2\bar{x}_{29}\bar{x}_{48} \vee \bar{x}_{23}\bar{x}_{33}\bar{x}_{40}\bar{x}_{52} \vee \bar{x}_4\bar{x}_6\bar{x}_{39}\bar{x}_{52} \vee$
- $\bar{x}_6\bar{x}_{11}\bar{x}_{36}\bar{x}_{55} \vee \bar{x}_6\bar{x}_9\bar{x}_{39}\bar{x}_{46}$

図 6 dnf 4 概念に対して抽出したルール集合
 Fig. 6 Extracted rules for dnf 4.

いターム (small disjuncts) を抽出できなかつたことになる。ところが, small disjuncts は, 一般に汎化誤りを大きくする問題の1つとして指摘されている¹²⁾。したがって, 元のルールを完全に抽出できない場合には, small disjuncts を無視することは有効な1つの戦略なので, MEF 尺度は望ましい性質を備えていると考える。また, dnf4 概念でも, 汎化誤り率と MEF 尺度が最少になる点は, ほぼ一致した。

7. おわりに

ノイズを含む事例から, 高い信頼性で良い正答率を保証する簡潔な分類ルールの抽出を目標とする RF3 法の提案を行った。

RF3 法の特長は, ルール集合選択のための新評価尺度 MEF の導入にある。MEF 尺度では, 任意の汎化誤り率の許容限界に対し, アルゴリズムの失敗確率の期待値が近似的に最小化される。また, MEF 尺度は, 抽出したルール集合の複雑さと例外事例の個数の和を最小化する尺度としても解釈できた。

実験の範囲では, RF3 法を用いて, 元のルール集合とほぼ同等なルール集合を抽出できた。また, 訓練誤り許容限界を変化させ, MEF 尺度と汎化誤り率を比較したところ, 両者の傾向が酷似することが分かった。したがって, MEF 尺度を用いることにより, 高い信頼性で汎化誤り率を最小にするルール集合の選択が可能になると期待できる。

今後は, もっと多くの問題に適用して, 新評価尺度 MEF と RF3 法の有効性を検証する予定である。

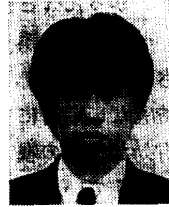
参 考 文 献

- 1) Angluin, D. and Laird, P.: Learning from Noisy Examples, *Machine Learning*, Vol. 2, No. 4, pp. 343-370 (1988).
- 2) Breiman, L., Friedman, J., Olshen, R. and Stone, C.: *Classification and Regression Trees*, Wadsworth International Group (1984).
- 3) Buntine, W.: A Critique of the Valiant Model, *IJCAI-89*, pp. 837-842, Detroit, MI (1989).
- 4) Clark, P. and Niblett, T.: The CN2 Induction Algorithm, *Machine Learning*, Vol. 3, No. 4, pp. 261-283 (1989).
- 5) Devroye, L.: Automatic Pattern Recognition: A Study of the Probability of Error, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 10, No. 4, pp. 530-543 (1988).
- 6) Efron, B.: Estimating the Error Rate of a Prediction Rule: Improvement on Cross-validation, *J. Am. Stat. Assoc.*, Vol. 78, No. 382, pp. 316-331 (1983).
- 7) Gallant, S.: A Connectionist Learning Algorithm with Provable Generalization and Scaling Bounds, *Neural Networks*, Vol. 3, No. 2, pp. 191-201 (1990).
- 8) Garey, M. and Johnson, D.: *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman, New York (1979).
- 9) Gordon, R.: Values of Mills' Ratio of Area to Bounding Ordinate and of the Normal Probability Integral for Large Values of the Argument, *Ann. Math. Statist.*, Vol. 12, pp. 364-366 (1941).
- 10) Haussler, D.: Quantifying Inductive Bias: AI Learning Algorithms and Valiant's Learning Framework, *Artif. Intell.*, Vol. 36, pp. 177-221 (1988).
- 11) Hoeffding, W.: Probability Inequalities for Sums of Bounded Random Variables, *J. Am. Stat. Assoc.*, Vol. 58, pp. 13-30 (1963).
- 12) Holte, R., Acker, L. and Porter, B.: Concept Learning and Accuracy of Small Disjuncts, *IJCAL-89*, pp. 813-818, Detroit, MI (1989).
- 13) Johnson, D.: Approximation Algorithms for Combinatorial Problems, *J. Comput. Syst. Sci.*, Vol. 9, pp. 256-278 (1974).
- 14) Korf, R.: Depth-first Iterative Deepening: An Optimal Admissible Tree Search, *Artif. Intell.*, Vol. 27, pp. 97-109 (1985).
- 15) Michalski, R.: Synthesis of Optimal and Quasi-optimal Variable-valued Logic Formula, *International Symposium on Multi-valued Logic*, pp. 76-87 (1975).
- 16) Pagallo, G. and Haussler, D.: Boolean Feature Discovery in Empirical Learning, *Machine Learning*, Vol. 5, No. 1, pp. 71-99 (1990).
- 17) Quinlan, J.: Induction of Decision Trees, *Machine Learning*, Vol. 1, No. 1, pp. 463-482 (1986).
- 18) Quinlan, J.: Simplifying Decision Trees, *Int. J. Man-Mach. Stud.*, Vol. 27, pp. 221-234 (1987).
- 19) Quinlan, J. and Rivest, R.: Inferring Decision Trees Using the Minimum Description Length, *Information and Computation*, Vol. 80, pp. 227-248 (1989).
- 20) Rissanen, J.: A Universal Prior for Integers and Estimation by Minimum Description Length, *Ann. of Statist.*, Vol. 11, No. 2, pp. 416-431 (1983).
- 21) Saito, K. and Nakano, R.: Rule Extraction from Facts and Neural Networks, *INNC-90-PARIS*, pp. 379-382, Paris, France (1990).

- 22) Swets, J. A.: Measuring the Accuracy of Diagnostic Systems, *Science*, Vol. 240, No. 3, pp. 1285-1293 (1988).
- 23) Valiant, L.: A Theory of the Learnable, *Comm. ACM*, Vol. 27, No. 11, pp. 1134-1142 (1984).

(平成3年8月27日受付)

(平成4年2月14日採録)



齊藤 和巳 (正会員)

1963年生. 1985年慶応義塾大学理工学部数理科学科卒業. 同年, NTT入社. 以来, 神経回路網, 機械学習などの研究に従事. NTTコミュニケーション科学研究所研究主任. 1991年9月より, オタワ大学客員研究員. 人工知能学会, 神経回路学会, 日本認知科学会各会員.



中野 良平 (正会員)

1947年生. 1971年東京大学工学部計数工学科卒業. 工学博士. 同年, 日本電信電話公社(現NTT)入社. 以来, 統計解析, 分散処理, データベース, 人工知能の研究に従事. 現在, NTTコミュニケーション科学研究所主幹研究員. 人工知能学会, 神経回路学会, 日本応用数理学会各会員.