

# メタ学習に基づくルール評価モデル構築支援手法の評価

## Evaluating Learning Algorithms from Constructive Meta-Learning Scheme for a Rule Evaluation Support Method

阿部 秀尚\*1      津本 周作\*1      大崎 美穂\*2      山口 高平\*3  
 Hidenao Abe      Shusaku Tsumoto      Miho Ohsaki      Takahira Yamaguchi

\*1 島根大学医学部医学科医療情報学講座

Department of Medical Informatics, Shimane University School of Medicine

\*2 同志社大学工学部情報システムデザイン科

Department of Information System Design, Faculty of Engineering, Doshisha University

\*3 慶應義塾大学理工学部管理工学科

Department of Administration Engineering, Faculty of Science and Technology, Keio University

This paper presents a method to support an evaluation procedure of a data mining process with a human-system interaction. Post-processing of mined results is one of the key issues to make a data mining process successful. However, it is difficult for human experts to evaluate many thousands of rules from a large dataset with noises completely. As an active support method, we have designed a prediction method based on rule evaluation models, which are learned from objective measure values and evaluations of a human expert for each rule. To evaluate performances of learning algorithms composed by constructive meta-learning system called CAMLET, we have done a case study with a meningitis data mining result, comparing with the five machine learning and statistical learning algorithms. Then we discuss about the availability of the constructive meta-learning system for our rule evaluation support method.

## 1. はじめに

近年、自然科学、社会科学、ビジネスの領域において、計算機を伴うシステムに蓄積されるデータは膨大となり、その活用が求められている。このようなデータを活用する手法として、データマイニング技術が注目され、多くのアルゴリズムの開発や研究が成されてきている。特に、IF-THEN形式のルールは領域専門家にとって可読性が高く、簡潔な表現形式としてその有用性が論じられている。しかしながら、大規模なしかもノイズの多いデータに対しては、既知の事実などを含む数十万ものルールが生成されることがあり、専門家が全てを評価することは困難が伴い、有益な知識の選別も困難となる。

以上の課題に対し、従来、マイニングアルゴリズムから生成されるルールなどの結果を選択する研究が行われてきた。中でも、評価用データに対するマイニング結果の数理構造から計算される数値によって、専門家の興味を推定するルール選択手法の多くは評価作業の自動化による専門家の負担軽減を目指している [Hilderman 01]。しかし、専門家の興味や関心は個人の知識量や時間経過に伴って変化するため、数理構造による単一の客観的指標の値によって専門家の評価基準を推定することは困難を伴う。これに対し、我々は専門家の興味や関心を推定するルール評価モデルを客観的指標の各値を基に学習し、適応的に専門家の評価を支援する手法を提案する。

本稿では、初めにルール評価作業の支援を行う手法について説明する。次に、専門家の評価基準を推定するルール評価モデルを獲得する学習アルゴリズムの性能評価実験を構成的メタ学習システムによって構築された学習アルゴリズムおよび代表的な5種類の学習アルゴリズムについて示す。性能評価実験においては、髄膜脳炎データベースとそれに対するデータ

連絡先: 阿部 秀尚, 島根大学医学部医学科医療情報学講座,  
 〒 693-8501 島根県出雲市塩冶町 89-1, 0853-20-2174,  
 0853-20-2170, abe@med.shimane-u.ac.jp

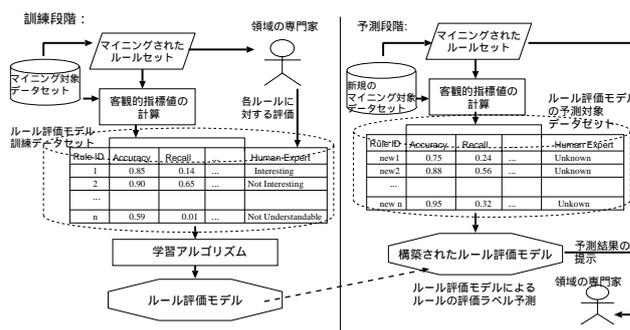


図 1: 客観的指標群に基づくルール評価支援手法の概観。

マイニング結果 [Hatazawa 00] を対象とする。このデータマイニング結果を評価する 39 種類の評価値 (以下、客観的指標と呼ぶ) と各ルールに対する専門医の評価をデータセットとして、各学習アルゴリズムによるルール評価モデルの性能、有効なモデル構築に必要なデータ数、モデルに用いられる客観的指標の内容について評価結果を示し、構成的メタ学習の有用性について考察する。

## 2. 客観的指標群に基づくルール評価支援

本ツールでは、可読性の高いマイニング結果である IF-THEN ルールの専門家による評価作業のインタラクティブな支援を行う。我々は、先行研究において、専門家の関心による興味深さの評価基準に適用可能な客観的指標の比較 [Ohsaki 04] と客観指標を組み合わせたルール評価モデルによる評価支援手法 [Abe 05] を提案してきた。

図 1 に本手法の概観を示す。

本手法では、始めに入力されたルール集合に対する各客観的指標の計算を行い、同時に専門家がルール集合の一部または全てを評価する。システムは、各ルールに与えられている客観的指標値と専門家による評価を訓練データセットとしてルール評価モデルを学習し、二回目以降の評価作業ではこのルール評価モデルによる評価ラベルの推定を行うことを可能とする。専門家は、二回目以降の評価において、単一の客観的指標による並び替えによる提示、過去のルール評価から推定された評価ラベルの提示の双方を利用することが可能である。

本手法では、より正確に専門家の評価ラベルを推定するため、予測精度の高いルール評価モデルが必要である。4. 節では、髄膜脳炎データマイニングの結果を用いて、ルール評価モデルを学習するアルゴリズムの評価を示す。

### 3. 構成的メタ学習による学習アルゴリズムの選定

構成型メタ学習を実現するツールである CAMLET[Abe 04]では、まず従来の学習アルゴリズムを分析し、機能単位であるメソッドを同定する。次に、これらのメソッドを体系化し、データセットにあわせて学習アルゴリズムを再構成することを可能にする。メソッドを体系化したものを、メソッドリポジトリと呼ぶ。CAMLET では、メソッドリポジトリを用いて、実際に学習アルゴリズムを所与のデータセットに対して実行し、ユーザから入力された要求を満たす学習アルゴリズム(マイニングアプリケーションと呼ぶ)を構築することにより、所与のデータセットに適した学習アルゴリズムを選定する。

#### 3.1 マイニングメソッドリポジトリ

マイニングメソッドリポジトリを構築するため、8 種類の帰納学習アルゴリズムの分析を行った [Abe 04]。分析対象となったのは、ヴァージョン空間法・AQ15・Classifier Systems・ID3・C4.5[Quinlan 93]・ニューラルネットワーク [Hinton 86]・Bagged/Boosted C4.5 である。これらのアルゴリズムから似たような機能部分(メソッド)を6種類同定し、アルゴリズムを構成する8種類の制御構造を整理した。6種類のメソッドは、「訓練・検証データセット生成」「分類器集合生成」「分類器集合の評価」「訓練・検証データセット更新」「分類器集合更新」「テストデータセットでの分類器集合評価」である。

以上の6種類のマイニングメソッドに対して、実装レベルと対応するよう各メソッドを階層構造を用いて体系化したものがメソッドリポジトリである。階層を生成するため、各ノードには「入力」「出力」「参照」のデータ構造と前後のメソッドを宣言的仕様として記述する。

また、上記の帰納メソッドで扱われるデータ構造の階層は、メソッドリポジトリの階層化に利用される。

#### 3.2 CAMLET の基本動作

CAMLET の基本動作を図2に示す。CAMLET の入力は、訓練データセットと必要がある場合はテストデータセット、要求(現在はテストデータに対する正解率)であり、出力は要求精度を満たすか、指定された更新回数の範囲内で最良のマイニングアプリケーションの仕様と所与のデータセットに対する実行結果である。

コンストラクションでは、マイニングアプリケーションの基本となる仕様が決まる。最初に8種類の制御構造から1つを選択する。次に、各メソッドについての具体的なメソッドをメソッドリポジトリの各ノードに与えられた宣言的仕様に基づいて選択する。インスタンスエーションでは、マイニングアプ

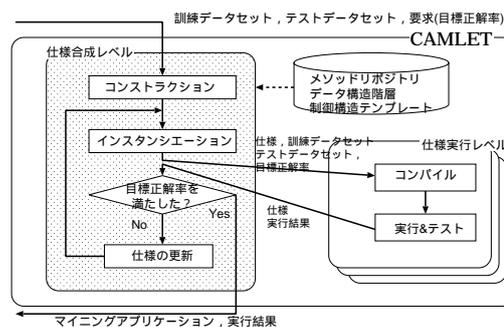


図2: CAMLET の基本動作。

リケーションの仕様を実体化する。そのため、メソッド間で出力と入力のデータ構造が違う場合、変換可能なデータ構造については変換メソッドを挿入するなど仕様の調整を行う。コンパイルでは、実体化された仕様から、実行コードを生成する。実行&テストでは、所与のデータセットに対してマイニングアプリケーションを実行し、評価に必要な正解率などを得る<sup>\*1</sup>。実行されたマイニングアプリケーションが要求された正解率を満たさない場合、仕様の更新によって、新たな仕様が生じられ、インスタンスエーションへと渡される。仕様の更新において、評価が改善するように仕様の更新を行うことは、すなわち仕様の洗練化となる。

### 4. ルール評価モデル学習アルゴリズムの性能評価実験

本実験では、先行研究 [Hatazawa 00] において、140 レコードの髄膜脳炎データセット得られた6種類の鑑別診断に関するルール計 244 ルールに対し、専門医のコメントから3種類の評価ラベル (I:興味あり・NI:興味なし・NU:理解不能) を与えた評価結果を用いる。これらの各ルールに対し、確率、統計検定、情報量、事例数、属性値の差異の各観点に基づく39種類の客観的指標 [Ohsaki 04] を算出し、客観的指標毎に1~0に正規化を行い、データセットを得た。このデータセットに対し、構成的メタ学習システム CAMLET によって選定された学習アルゴリズム(以下、この学習アルゴリズムをCAMLETと表記する)、決定木 C4.5(J4.8)[Quinlan 93]・バックプロパゲーション学習型ニューラルネットワーク (BPNN) [Hinton 86]・SVM(Support Vector Machine)[Platt 99]<sup>\*2</sup>・線形回帰式による分類 (CLR)[Frank 98]<sup>\*3</sup>・OneR[Holte 93] の各学習アルゴリズムによって得られる、ルール評価モデルを評価する。CAMLET 以外の学習アルゴリズムでは、Weka[Witten 00] による実装を用いた<sup>\*4</sup>。

ルール評価モデルの評価観点は、以下の3つである。

- 訓練データセットおよび Leave-One-Out(LOO) での性能
- 各ルール評価モデルの正解率がデフォルトクラス割合を越えるのに必要な訓練データ数
- ルール評価モデルに用いられる指標

\*1 以降の実験では、10 回交差検定の平均正解率を仕様の評価値とした。

\*2 カーネル関数には多項式カーネルを用いた。

\*3 多重共線性のある属性の削除と貪欲探索による属性選択を設定した。

\*4 weka-3-4-7 を用いた。

#### 4.1 CAMLET による学習アルゴリズムの選定

本実験では、約 6000 種類の学習アルゴリズムを生成可能なメソッドリポジトリを用いて、CAMLET により 100 回の仕様更新を上限として、髄膜炎ルール集合からのデータセットに対する学習アルゴリズムの選定を行った。この結果、得られた学習アルゴリズムは図 3 に示すように、情報利得比を規準とした決定木生成を基本に、Boosting をランダム分割した訓練データセットに対して繰り返し行い、最終的な分類は Boosting で用いられる重みつき投票によって行われる。

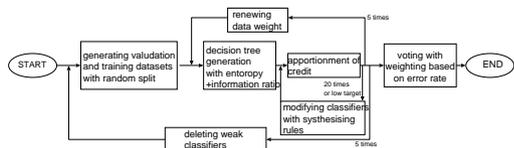


図 3: CAMLET によって選定された学習アルゴリズム。

#### 4.2 ルール評価モデルの性能評価

ここで、評価ラベル (クラス)  $i$  についての再現率  $Recall(D_i)$  は正しく分類されたインスタンス数を  $Correct(i)$  として、 $Recall(D_i) = (Correct(i)/|D_i|) \times 100$  と計算される。同様に、予測正答率  $Precision(D_i)$  はクラス  $i$  と予測されたデータ数  $Predicted(i)$  を用いて、 $Precision(D_i) = (Correct(i)/Predicted(i)) \times 100$  と計算される。データセット  $D$  に対する正解率  $Acc(D)$  は、 $Acc(D) = (Correct(D)/|D|) \times 100$  と計算される。

各学習アルゴリズムによる結果を表 1 に示す。この結果、ルール評価モデルによる評価ラベルの分類精度は、最も多数を占めるクラスである ‘NI’ を予測した割合を全てにおいて上回っている。

表 1: 各学習アルゴリズムのルール評価モデルによる再代入分類精度 (%) と LOO における予測精度 (%)。

Learning Algorithms	訓練データセットでの評価								
	Acc.	Recall			Precision				
		I	NI	NU	I	NI	NU		
CAMLET	89.4	70.8	97.9	11.1	85.0	90.2	100.0		
J4.8	85.7	41.7	97.9	66.7	80.0	86.3	85.7		
BPNN	86.9	81.3	89.8	55.6	65.0	94.9	71.4		
SVM	81.6	35.4	97.3	0.0	68.0	83.5	0.0		
CLR	82.8	41.7	97.3	0.0	71.4	84.3	0.0		
OneR	82.0	56.3	92.5	0.0	57.4	87.8	0.0		

Learning Algorithms	Leave-One-Out(LOO)での評価								
	Acc.	Recall			Precision				
		I	NI	NU	I	NI	NU		
CAMLET	80.3	7.4	73.0	0.0	7.4	73.0	0.0		
J4.8	79.1	29.2	95.7	0.0	63.6	82.5	0.0		
BPNN	77.5	39.6	90.9	0.0	50.0	85.9	0.0		
SVM	81.6	35.4	97.3	0.0	68.0	83.5	0.0		
CLR	80.3	35.4	95.7	0.0	60.7	82.9	0.0		
OneR	75.8	27.1	92.0	0.0	37.1	82.3	0.0		

表 1 より、CAMLET により構築された学習アルゴリズムは、その他の学習アルゴリズムと比較して、訓練データセットでの評価では最も高い正解率となっている。さらに、Recall および Precision を見ると、正解率で本学習アルゴリズムに次ぐニューラルネットワークよりも、正確に ‘興味あり (I)’ と ‘妥当 (NI)’ を分類している。また、LOO による予測精度の評価でも、SVM や CLR と同等の 80% 台の信頼性を持つことが示された。

以上の結果は、分類精度・信頼性共に高い情報利得比による決定木 (J4.8 に相当) を複数統合することにより、より高い分類精度を信頼性を損なわずに達成できたものと考えられる。

#### 4.3 有効なルール評価モデルを構築するために必要な訓練データ数の評価

本手法では、専門家によるルール集合の評価が少なくとも一度以上必要である。このため、有効なルール評価モデル構築に必要な最低限のルール評価回数を見積もることは、専門家のルール評価作業の負担を軽減するために重要な評価観点である。本節では、髄膜炎ルール集合の評価によるデータセットを訓練データセットとして、この訓練データセットの部分集合に対して訓練されたルール評価モデルを構築し、各学習アルゴリズムの訓練データセット全体に対する正解率を算出した。この結果を表 2 に示す。

表 2: 各学習アルゴリズムの訓練部分集合による訓練データセット全体に対する正解率 (%)。

Learning sample	10	20	30	40	50	60	70	80	90	100
CAMLET	76.7	78.4	80.8	81.6	81.7	82.6	82.8	84.8	84.6	89.3
J4.8	73.4	74.7	79.8	78.6	72.8	83.2	83.7	84.5	85.7	85.7
BPNN	74.8	78.1	80.6	81.1	82.7	83.7	85.3	86.1	87.2	86.9
SMD	78.1	78.6	79.5	79.8	79.8	80.0	79.9	80.2	80.4	81.6
CLR	76.6	76.5	80.2	80.2	80.9	80.7	80.8	81.4	81.0	82.9
OneR	76.2	73.4	77.5	78.0	77.7	77.5	79.0	77.8	78.9	82.4

表 2 から、CAMLET によって選定された学習アルゴリズムは、訓練データセットの 10% で全体を訓練データに用いたときの 85% 以上の正解率をもつルール評価モデルを構築することが可能であることが分かる。また、この学習アルゴリズムでは、少数の訓練部分集合においては SVM や CLR のように高い正解率を達成し、後半においては J4.8 や BPNN と同等以上の高い正解率を得ることが可能である。このことから、複数の分類器を統合する学習アルゴリズムを適切に選定することにより、単一の分類器をルール評価モデルとする場合より適応的なルール評価モデルの構築が可能であることが明らかとなった。

#### 4.4 ルール評価モデルに用いられる指標の評価

本節では、髄膜炎ルール集合の評価から得られたデータセットからブートストラップサンプリングに対する、CAMLET により選定された学習アルゴリズム・J4.8・CLR・OneR によるルール評価モデル生成をそれぞれ 10000 回繰り返し行い、ルール評価モデルに用いられる客観的指標の出現頻度を計測した。本実験の結果を図 4 に示す。

以上の結果から、それぞれの学習アルゴリズムで用いられる客観的指標は異なっていることが明らかとなった。しかし、ルール評価モデルに用いられる客観的指標の種類は、ルールの正確さを表す指標 (Accuracy, Precision, Coverage) などが大半をしめている。これに加え、異なった種類の指標である Peculiarity (ルールの特異性) [Zhong 03] や GBI [Gago 98] なども出現頻度が上位となる結果となった。このことは、専門家が実際の評価において、ルールが示す事実の正確性に加え、主観的な基準による別の観点から関心のあるルールを ‘興味あり’ と評価していたことと合致する。

### 5. おわりに

本稿では、客観的指標によりインタラクティブに専門家の IF-THEN ルールの評価作業を支援する手法について、より正確な支援を可能とするため、ルール評価モデルの構築に構成的メタ学習を適用し、ルール評価モデルの構築性能について評価実験を行った。専門家の評価基準を推定するためのルール評価モデルは、各ルールに対する客観的指標群の値と専門家の評価から学習アルゴリズムによって獲得するため、高い予測精度が求められる。

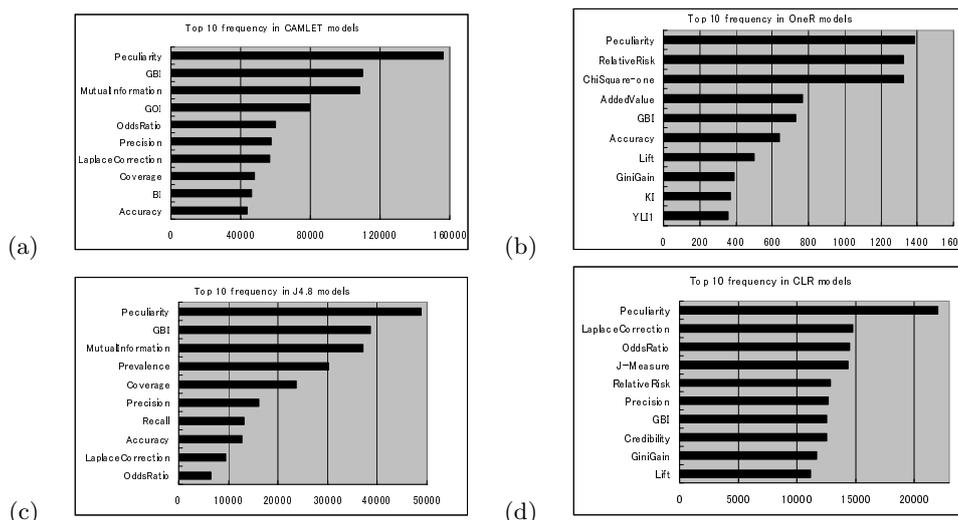


図 4: 10000 回のブートストラップサンプルに対して, CAMLET により選定された学習アルゴリズム (a), OneR (b), J4.8 (c), CLR (d) により構築されたルール評価モデルに出現する客観的指標の頻度.

髄膜炎データマイニング結果を用いた実験では, 代表的な 5 種類の学習アルゴリズムと構成的メタ学習システムにより選定された学習アルゴリズムの性能比較を行った. この結果, 以上の評価より, 客観的指標群に基づくルール評価モデルの構築に構成的メタ学習を適用することにより, より正確な専門家のルール評価作業支援が可能であることを示した.

今後は, データの前処理を含めたルール評価モデル構築アルゴリズムの選定により, より適応的に専門家のルール評価支援を行うことを目指し, 構成的メタ学習システムの拡張と他のデータマイニング結果への適用を行う予定である.

## 参考文献

- [Abe 04] Abe, H. and Yamaguchi, T.: Constructive Meta-Learning with Machine Learning Method Repositories, in Proc. of the 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA/AIE 2004, pp. 502-511 (2004).
- [Abe 05] Abe, H., Tsumoto, S., Ohsaki, M., and Yamaguchi, T.: A Rule Evaluation Support Method with Learning Models Based on Objective Rule Evaluation Indexes, in Proc. of the Fifth International Conference on Data Mining (ICDM2005), pp. 549-552 (2005).
- [Frank 98] Frank, E., Wang, Y., Inglis, S., Holmes, G., and Witten, I. H.: Using model trees for classification, Machine Learning, Vol.32, No.1, pp. 63-76 (1998).
- [Gago 98] Gago, P., Bento, C.: A Metric for Selection of the Most Promising Rules, in Proc. of Euro. Conf. on the Principles of Data Mining and Knowledge Discovery PKDD-1998, pp. 19-27 (1998).
- [Hatazawa 00] Hatazawa, H., Negishi, N., Suyama, A., Tsumoto, S., and Yamaguchi, T.: Knowledge Discovery Support from a Meningoencephalitis Database Using an Automatic Composition Tool for Inductive Applications, in Proc. of KDD Challenge 2000 in conjunction with PAKDD2000, pp. 28-33, (2000).
- [Hilderman 01] Hilderman, R. J. and Hamilton, H. J.: Knowledge Discovery and Measure of Interest, Kluwer Academic Publishers (2001).
- [Hinton 86] Hinton, G. E.: "Learning distributed representations of concepts", *Proceedings of 8th Annual Conference of the Cognitive Science Society*, Amherst, MA. Reprinted in R.G.M.Morris (ed.) (1986).
- [Holte 93] Holte, R. C.: Very simple classification rules perform well on most commonly used datasets, Machine Learning, Vol. 11, pp. 63-91 (1993).
- [Ohsaki 04] Ohsaki, M., Kitaguchi, S., Kume, S., Yokoi, H., and Yamaguchi, T.: Evaluation of Rule Interestingness Measures with a Clinical Dataset on Hepatitis, in Proc. of ECML/PKDD 2004, LNAI3202, pp. 362-373 (2004).
- [Platt 99] Platt, J.: Fast Training of Support Vector Machines using Sequential Minimal Optimization, Advances in Kernel Methods - Support Vector Learning, B. Schölkopf, C. Burges, and A. Smola, eds., MIT Press, pp. 185-208 (1999).
- [Quinlan 93] Quinlan, R.: C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, (1993).
- [Witten 00] Witten, I. H and Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, (2000).
- [Zhong 03] Zhong, N., Yao, Y. Y., Ohshima, M.: Peculiarity Oriented Multi-Database Mining. IEEE Trans. on Knowledge and Data Engineering, 15, 4, pp. 952-960 (2003).