

客観的指標群に基づくルール評価支援ツールの開発 — 髄膜脳炎データマイニングを対象として —

A Visual Rule Evaluation Tool with Active Interaction based on Interestingness Measures
— A Case Study on a Meningitis Data Mining —

阿部 秀尚*1
Hidenao Abe

北口 真也*2
Shinya Kitaguchi

大崎 美穂*3
Miho Ohsaki

津本 周作*1
Shusaku Tsumoto

山口 高平*4
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 静岡大学大学院情報学研究科

Graduate School of Informatics, Shizuoka University

*3 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

*4 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. We have designed a tool to support rule evaluation procedure based on objective measures, which are calculated to each rule to evaluate mathematically. 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 of rule evaluation models, we have done a case study with a meningitis data mining result. Then we discuss about the availability of our rule evaluation support tool.

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、我々は単一の客観的指標による専門家へのルールの提示に加え、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から学習し、より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、まず、実際にルール評価作業の支援を行うツール

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、専門医による評価を予測するルール評価モデルの学習と予測精度の評価を行った。この結果を元にルール評価支援ツールにおける専門家の評価推定に用いるルール評価モデルとその構築について考察する。

2. 客観的指標群に基づくルール評価支援ツール

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 [阿部 04] を提案してきた。これらの研究の考察から、ツールを利用する専門家に適応するよう、単一の客観的指標による提示 (受動的支援) と客観的指標群から学習アルゴリズムによって得たルール評価モデルによる評価ラベルの推定 (能動的支援) を随時切替えられる機能が必要だと判断した。

図 1 にシステムの概観を示す。

本ツールでは、マイニングアルゴリズムから出力された IF-THEN 形式のルールを GUI を通じて専門家に提示し、評価の入力を受け付ける。また、要求に応じて客観的指標による並び

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

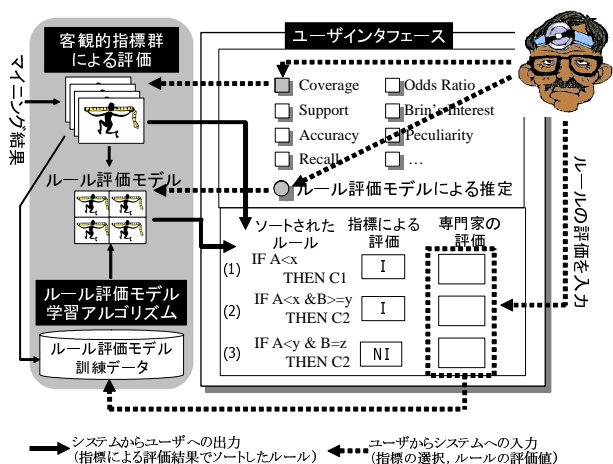


図 1: 客観的指標群に基づくルール評価支援ツールの概観

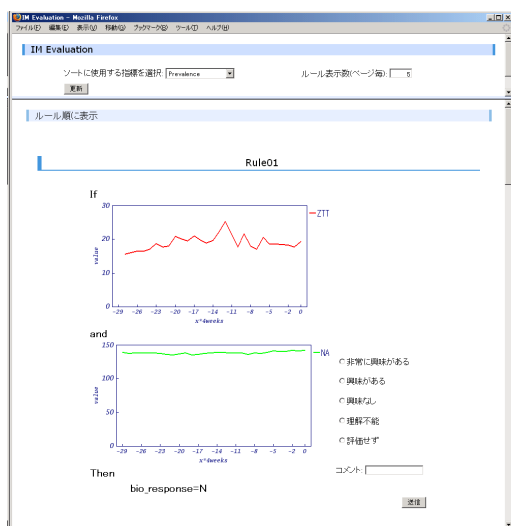


図 2: 客観的指標群に基づくルール評価支援ツールの動作画面。画面は時系列データから得られた IF-THEN ルールの例。

替えやルール評価ラベルの推定を行った結果を専門家に提示し、評価作業をインタラクティブに行うことを可能にする。図 2 に本ツールの動作画面を示す。

専門家がルールを評価する際は、始めに適当な客観的指標による並び替えを行い、ルールを評価する。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。

ここで、より正確に専門家の評価ラベルを推定するためには、予測精度の高いルール評価モデルが必要である。次節では、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. 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、データセットを得た。それぞれの属性である客観的指標毎に 1~0 に正規化を行った。このデータセットに対し、決定木 C4.5 [Quinlan 93]・ニューラルネットワーク [Hinton 86]・SVM(Support Vector Machine)[Platt 99]^{*1}・線形回帰式による分類 [Frank 98]^{*2}・OneR[Holte 93] の各学習アルゴリズムによって得られる、ルール評価モデルの予測精度を評価する。学習アルゴリズムは、Weka[Witten 00] による実装を用いた^{*3}。

ルール評価モデルの評価基準は、以下の 3 つである。

- 訓練データセットでの「興味あり」の再現率
- 訓練データセットでの正解率
- Leave-One-Out(LOO) による正解率

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。データセット D に対する正解率 $Acc(D)$ は、正しく分類されたインスタンス数を $Correct(D)$ としたときに $Acc(D) = (Correct(D)/|D|) \times 100$ 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予測精度は、無作為に評価ラベルを予測した場合を全てにおいて上回っている。表中、上段の数値は '/' の左に正しく分類されたインスタンス数、右に対象となるデータのインスタンス数が示される。

3.1 訓練データセットでの評価

訓練データセットでの評価では、「興味あり」の再現率においてニューラルネットワークが最も良い結果となった。他の学習アルゴリズムは、単純な並び替えと閾値設定によりルール集合を学習する OneR とほぼ同程度の再現率にとどまった。これは、「興味あり」の評価ラベルを持つインスタンスが 39 種類の客観的指標による空間中で線形分離困難な分布であることを示唆している。

全体の正解率では、OneR に比べてその他の学習アルゴリズムが優れており、単一の客観的指標と比較して複数の客観的指標によるルール評価モデルのほうが分類能力が高い結果となった。

3.2 Leave-One-Out による頑強性の評価

LOO による評価結果は、新規データに対して本手法が頑強であるかを示す。各学習アルゴリズムによる結果より、75.82%~81.58%の正解率が得られた。ニューラルネットワークは訓練データセットでの評価では最も高い性能を示したが、LOO による評価は低く過学習の傾向であった。OneR では、244 種類

*1 カーネル関数には多項式カーネルを用いた。
 *2 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。
 *3 weka-3-4-3 を用いた。

表 1: 各学習アルゴリズムのルール評価モデルによる予測精度. (NN:バックプロパゲーション学習ニューラルネットワーク, SMO:Weka に実装された SVM, CLR:線形回帰による分類を表す)

| | 訓練データ | | Leave-One-Out 正解率 |
|------|-----------|---------|----------------------|
| | 「興味あり」再現率 | 正解率 | |
| C4.5 | 20/48 | 209/244 | 193/244 |
| | 41.67% | 85.66% | 79.10% |
| NN | 39/48 | 212/244 | 189/244 |
| | 81.25% | 86.89% | 77.46% |
| SMO | 17/48 | 199/244 | 199/244 |
| | 53.42% | 81.56% | 81.56% |
| CLR | 20/48 | 202/244 | 196/244 |
| | 41.67% | 83.20% | 80.33% |
| OneR | 27/48 | 202/244 | 185/244 |
| | 56.25% | 81.97% | 75.82% |

の訓練データセットに対し適宜単一の指標を選択したが、新規データに対する予測精度には限界があることが示された。

その他の学習アルゴリズムでは、79.10%～81.56%の正解率が得られた。これらの学習アルゴリズムは、複数の客観的指標の組合せによるルール評価モデル、つまり合成関数としての新たな指標を学習することができた。

以上の実験に使用した学習アルゴリズムは、比較的単純な学習アルゴリズムである。ルール評価モデルの予測精度を向上させるには、従来から正解率の向上を目指して研究が行われているメタ学習アルゴリズムによってより高い正解率が得られることが期待される。

3.3 訓練データでのルール評価モデル

本節では、学習アルゴリズムから得られたルール評価モデルのうち、特に可読性が高い OneR と C4.5 による決定木について考察する。以下のモデルでの各指標値は、訓練データにおいて正規化した後の値である。

OneR によるルール集合を図 3 に示す。

| | | |
|----|-----------------------------------|-----------|
| IF | YLI1 < 0.02 THEN | "I" |
| IF | YLI1 >= 0.02 and YLI1 < 0.29 THEN | "NI" |
| IF | YLI1 >= 0.29 and YLI1 < 0.43 THEN | "I" |
| IF | YLI1 >= 0.43 and YLI1 < 0.44 THEN | "NI" |
| IF | YLI1 >= 0.44 and YLI1 < 0.55 THEN | "I" |
| IF | YLI1 >= 0.55 and YLI1 < 0.63 THEN | "NI" |
| IF | YLI1 >= 0.63 and YLI1 < 0.83 THEN | "I" |
| IF | YLI1 >= 0.83 | THEN "NI" |

図 3: OneR によって得られたルール評価モデル (ルール集合)

ここで用いられている Yao & Liu's Interestingness based on one-way support(YLI1)[Yao 97] は、ルールの分類結果の正確性を表す指標に補正を加えて、評価者の興味深さを予測する目的で提案された指標である。

一方、図 4 に示す C4.5 による決定木では、ルールの正確性に補正を加えた Laplace Correction[Tan 02] をはじめ、ルールの正確性を評価する指標 (Accuracy, Precision, Recall, Coverage) とルールが予測するクラス的一般性を評

価する指標 (Prevalence) が用いられている。GOI(Gray & Orłowska Interestingness)[Gray98] は、前述の指標群と同様にルールによる分類結果から計算される評価値である。Peculiarity[Zhong 03] は、ルールの特異性について、ルール同士の条件節に含まれる値の差異を積算して求められる。

これらのルール評価モデルに用いられた客観的指標の定義式を表 2 にまとめる。

表 2: 本実験における OneR と C4.5 によるルール評価モデルに現れる客観的指標の定義。

| 指標名 | 定義式 |
|--------------------|--|
| YLI1 | $P(C A) * \log_2 \frac{P(A \cap C)}{P(A) * P(C)}$ |
| Laplace Correction | $\frac{N(A \cap C) + 1}{N(A) + 2}$ |
| Coverage | $P(A)$: ルールの条件節が被覆する確率 |
| Prevalence | $P(C)$: ルールの結論節が被覆する確率 |
| Precision | $P(C A)$: ルールの条件節があるときに結論を満たす条件付き確率 |
| Recall | $P(A C)$: ルールの結論節のうち条件節によって被覆される条件つき確率 |
| Accuracy | $P(C A) * P(A) + P(\neg C \neg A) * P(\neg A)$ |
| GOI | $((\frac{P(C A)}{P(A) * P(C)})^k - 1) * ((P(A) * P(C))^m)$ |
| Peculiarity | $\sum_{a_{ij} \in Rule} \frac{PF(a_{ij})}{N}$ $PF(a_{ij}) = \sum_k a_{ij} - a_{ik} ^\alpha$ a_{ij} : i -th and j -th value in a relation. N : Number of attributes in the rule. α : Constant. Default = 0.5. |

以上の結果から、学習アルゴリズムによるルール評価モデルには各ルールの正確性を評価する指標が多く用いられるが、それぞれ異なる指標に基づいて、ルール評価モデルが構成されていることが示された。これは、一般のデータマイニングと同様、ルール評価モデルの学習についても、適切なデータの前処理とアルゴリズムの選定が必要であることを示唆している。

4. 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ツールについて述べた。本ツールは、単一の客観的指標によるルール選択とルール評価モデルによる専門家の評価基準推定による支援の両方を実現する。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。5種類の学習アルゴリズムの予測精度の評価実験では、39種類の客観的指標の値と専門家の評価ラベルから成る訓練データに対する正解率が全く推定を行わない場合に比べて優れた結果となった。また、Leave-One-Out による新規データに対する頑強性の評価では、75.82%～81.58%の正解率を得た。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による専門家のルール評価作業支援が有効であることを示した。

今後は、新たな客観的指標の追加と学習アルゴリズム選定の機構を追加し、専門家のルール評価作業支援の効率化を目指す。さらに、他の領域におけるデータマイニング結果を用いた評価を行っていく。

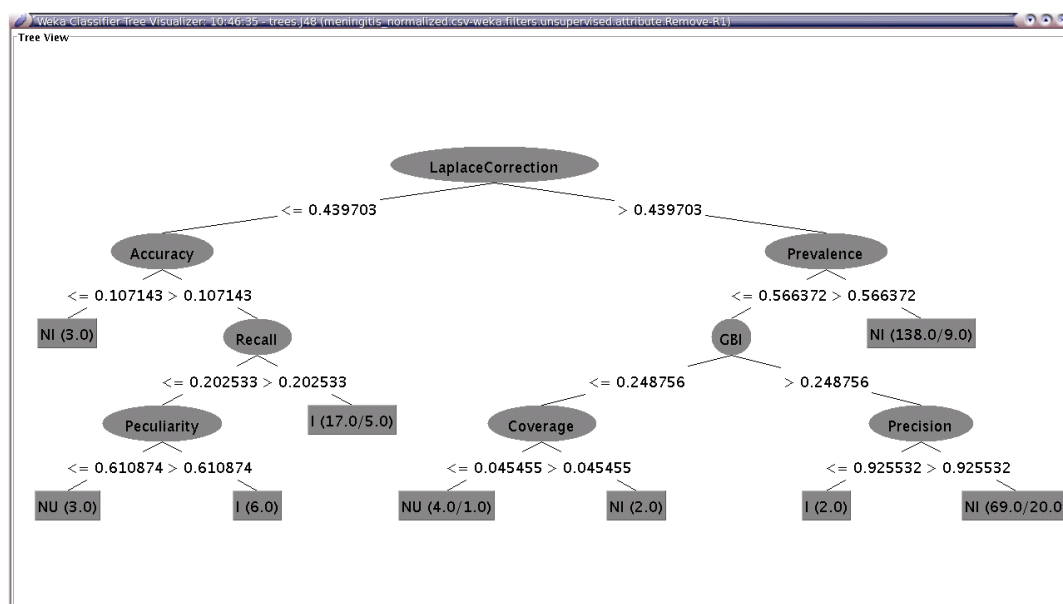


図 4: C4.5 によって得られたルール評価モデル (決定木)

参考文献

- [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).
- [Gray98] Gray, B., Orlowska, M. E.: CCAIA: Clustering Categorical Attributes into Interesting Association Rules. *Proc. of Pacific-Asia Conf. on Knowledge Discovery and Data Mining PAKDD-1998*, pp. 132-143 (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).
- [Tan 02] Tan, P. N., Kumar V., Srivastava, J.: Selecting the Right Interestingness Measure for Association Patterns. *Proc. of Int. Conf. on Knowledge Discovery and Data Mining KDD-2002*, pp. 32-41 (2002).
- [Witten 00] Witten, I. H and Frank, E.: *DataMining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, (2000).
- [Yao 97] Yao, J., Liu, H.: Searching Multiple Databases for Interesting Complexes. in Liu, H., Motoda, H., Liu H. (Eds.): *KDD: Techniques and Applications*, World Scientific (1997).
- [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).
- [阿部 04] 阿部 秀尚, 岡本 和也, 大崎美穂, 山口高平, 津本周作: 客観指標群に基づくルール評価モデルの構築 - 髄膜脳炎データベースを対象として -, 第 24 回医療情報学連合大会, 2-J-3-2, (2004).