

医学知識発見におけるルールの興味深さ指標の評価

大崎 美穂^{†1} 阿部 秀尚^{†2} 津本周作^{†2}
横井 英人^{†3} 山口 高平^{†4}

本論文では、医療データセットを用いた2つの実証実験を通して、医学知識発見における興味深さ指標の有用性を調べ、興味深さ指標の活用方法を議論する。まず、髄膜炎の医療データセットから生成されたルールに対し、40種類の多様な興味深さ指標と医師による評価結果を得た。そして、F-Measureと相関係数を使って興味深さ指標を持つ医師の興味の推定性能を見積もり、興味深さ指標間で比較した。肝炎の医療データセットでも同様の実験を行った。その結果、Accuracy, Uncovered Negative, Peculiarity, Relative Risk, Chi-Square Measure for One Quadrant は、医学ドメインでの人間の興味を安定かつ妥当に推定する可能性が示された。また、推定性能は医師が立てる仮説の確かさに依存すること、興味深さ指標を組み合わせて、マイニングシステムとのインタラクションを通じた医師の仮説生成・検証を支援しうることが示唆された。

Evaluation of Rule Interestingness Measures in Medical Knowledge Discovery in Databases

MIHO OHSAKI,^{†1} HIDENAO ABE,^{†2} SHUSAKU TSUMOTO,^{†2}
HIDETO YOKOI^{†3} and TAKAHIRA YAMAGUCHI^{†4}

We discuss the usefulness of rule interestingness measures for medical KDD through experiments using clinical datasets and also consider how to utilize these measures in postprocessing based on the experimental outcomes. We first conducted an experiment to compare evaluation results by forty various interestingness measures with those by a medical expert for rules discovered in a clinical dataset on meningitis. We calculated and compared the performance of each interestingness measure to estimate a medical expert's interest with the F-Measure and correlation coefficient. We then conducted a similar experiment on hepatitis. The comprehensive results of experiments on meningitis and hepatitis showed that Accuracy, Uncovered Negative, Peculiarity, Relative Risk, and Chi-Square Measure for One Quadrant have a stable, reasonable performance in estimating real human interest in the medical domain. The results also indicate that the performance of interestingness measures is influenced by the certainty of a hypothesis made by a medical expert, and that the combinational use of interestingness measures will contribute to support medical experts to generate and confirm their hypotheses through human-system interaction.

1. はじめに

ルールの興味深さは、データマイニングやデータベースからの知識発見で重要な課題の1つである。こ

れまでに、ルールの興味深さの定式化や興味深さ指標による人間の代行評価の研究が行われた。しかし、従来の興味深さ指標の多くは個別に提案されており、理論的、あるいは実用的な有用性の検証や比較は十分ではなかった。近年、複数の興味深さ指標を理論的に分析する試みもあるが^{(15), (24), (32), (40), (44)}、実用面を考慮した試みは少ない。現実のドメインで、興味深さ指標が本当に興味深いルールの発見に役立つかを実証することが望まれる^{(7), (35)}。そこで本研究では、科学的・社会的に重要性が高い医療に的を絞って、興味深さ指標の有用性を検討する。実験では、医療データから得たルールに対する医師の評価と興味深さ指標による評価の一致度合いを調べ、興味深さ指標を持つ人間の興味の推

†1 同志社大学工学部
Faculty of Engineering, Doshisha University

†2 島根大学医学部医療情報学講座
Department of Medical Informatics, Shimane University, School of Medicine

†3 香川大学医学部附属病院医療情報部
Department of Medical Informatics, Kagawa University Hospital

†4 慶應義塾大学理工学部
Faculty of Science and Technology, Keio University

定性能、および、医学知識発見支援における興味深さ指標の活用方法を検討する。

人間を含む実験では、厳密さのために実験条件や手続きを統制し過ぎると、現実とかげ離れた状況を生む恐れがある。知識発見の場合、ドメイン専門家はマイニングシステムが出力するルールを見比べ、興味や着眼点を徐々に変化させながら新知識を導き出す。よって、統制で心理状態を静的に保ち、事前に決めた説明変数と心理尺度の関係を調べるアプローチは新知識の導出過程を扱うには適さない、と考えられる。また、価値ある知識を実際に見つける人的・時間的コストは非常に高いという問題もある。本研究では上記の問題を考慮し、マイニングシステムの調整、医師のルール評価と意見聴取を1~2年繰り返して医学知識発見に成功した2事例(1つは髄膜炎²³⁾, 1つは肝炎³⁶⁾に関する)を取り上げ、これらのルールと評価結果を用いる。事例数・医師数・知識発見過程の不統制さ等は科学的な厳密性には欠けるが、現実の成功事例を掘り下げて検討し、興味深さ指標の有用性を知ることは工学的に有意義であろうと我々は考えた。

本論文の2章では、興味深さ指標の調査と選定について述べる。3章では髄膜炎ドメインの実験I, 4章では肝炎ドメインの実験IIを示す。実験I, IIでは、興味深さ指標が持つ医師の興味の推定性能を見積もり、医学知識発見に至る医師の思考過程と興味深さ指標を活用した思考支援に関する知見を得る。最後に、5章で本研究のまとめと将来展望を示す。

2. 従来の興味深さ指標

興味深さ指標は個別に提案されたものが多く、同意異称のものもある。興味深さ指標を体系化すべく、指標間の等価性を数学的に証明した研究や挙動をシミュレーションで調べた研究^{15),24),32),40),44)}もあるが、用途や人間の主観との関係には言及していない。近年、これらの関係を検証する試みも出てきたが、検証方法が確立されていない、用いられる指標が限られている等の問題がある^{7),35)}。そこで我々は文献調査を行い、用途や人間の主観を加味して多様な興味深さ指標の分類を試みた。興味深さ指標は以下の要因で大別できる。主体(誰が評価するか。計算機か人間か)、対象(何を評価するか。相関ルールか分類ルールか)、単位(対象を個別に評価するか。1つ1つのルールかルール群か)、基準(どの基準に基づくか。絶対的な基準か相対的な基準か)、理論(どの理論に基づくか。事例数、確率量、統計量、情報量、ルールや属性の距離、ルールの複雑さのどれか)。

上記の要因のうち評価の主体が最も重要である。慣習的に、計算機が主体の興味深さ指標は客観的指標、人間が主体の興味深さ指標は主観的指標と呼ばれる。主流は客観的指標であり、少なくとも我々の文献調査では40を超えるものがある。客観的指標は、ルールを満たす事例やルールに付随する事例(前件部のみを満たす事例、すべての事例等)の分布構造を分析し、数理的観点からルールが有意義かを見積もる。客観的指標は基本的にドメイン知識を含まないので、現実に関心深いルールの抽出よりも、むしろ無意味なルールの除去に有効と考えられる。一方、主観的指標は10数程度提案されている。主観的指標は、事前にユーザが定式化したピリフ・バイアス・テンプレート等を参照し、これらへのルールの一一致度合いを距離計算で見積もる。主観的指標は事前に組み込んだドメイン知識を使うため、現実に関心深いルールの発見にある程度役立つが、その性能はユーザが興味をうまく定式化できるかに依存し、組み込んだドメイン知識とかげ離れた予期せぬルールを発見しにくい。

客観・主観にかかわらず、興味深さ指標は人間の興味を直接は反映していない。混同を防ぐため、客観的指標、主観的指標、人間の興味を以下に定義する。客観的指標: ルールの一般性、正確性、独自性等の性質を表す指標。データ分布構造の数理的な分析で算出され、人間の興味は含まない。主観的指標: 事前に人間が与えた興味深さに関する情報とルールから得られる情報の類似度を表す指標。事前情報としては人間の興味を含むが、類似度は数理的な距離計算で算出される。人間の興味: ルールを見て人間が実際に感じる興味。認知特性、ドメイン知識、個人的な経験、過去に評価したルールの影響等の総合で形成される。

我々は客観的指標に着目し、客観的指標と人間の興味の間関係を調べる。そこで一般的で多様な客観的指標を選定し、 $IF A THEN C$ 形式の分類ルール用に定義式を揃えた。分類ルールの前件部 A は、属性 A_1 (属性値 a_{11}, a_{12}, \dots), 属性 A_2 (属性値 a_{21}, a_{22}, \dots), \dots の連言、たとえば $A_1 = a_{14}$ and $A_2 = a_{21}$ and \dots の形式を持つ。後件部 C はクラス c_1, c_2, \dots の1つである。最も基本的な客観的指標 Coverage, Prevalence, Precision, Recall, Support, Uncovered Negative を以下に説明する。これらは本来、事例集合を True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN) に分ける混同行列(表1上側)に基づき、式(1)~(6)の第1項で定義される。分類ルールを対応付けると表1下側の混同行列と、前件部 A と否定 $\neg A$, 後件部 C と否定

-C で表現される式 (1) ~ (6) の最後の項が求まる．他の客観的指標の多くはこれらの組合せからなる．今回選定した客観的指標を表 2 に示す．

$$Cov = \frac{N_{TP} + N_{FP}}{N_{TP} + N_{FP} + N_{TN} + N_{FN}} = \frac{N_{A \wedge C} + N_{A \wedge \neg C}}{N_{A \wedge C} + N_{A \wedge \neg C} + N_{\neg A \wedge \neg C} + N_{\neg A \wedge C}} = \frac{N_A}{N} = P(A) \quad (1)$$

$$Prev = \frac{N_{TP} + N_{FN}}{N_{TP} + N_{FP} + N_{TN} + N_{FN}} = \frac{N_{A \wedge C} + N_{\neg A \wedge C}}{N_{A \wedge C} + N_{A \wedge \neg C} + N_{\neg A \wedge \neg C} + N_{\neg A \wedge C}} = \frac{N_C}{N} = P(C) \quad (2)$$

$$Prec = \frac{N_{TP}}{N_{TP} + N_{FP}} = \frac{N_{A \wedge C}}{N_{A \wedge C} + N_{A \wedge \neg C}} = \frac{N_{A \wedge C}}{N_A} = P(C|A) \quad (3)$$

$$Rec = \frac{N_{TP}}{N_{TP} + N_{FN}} = \frac{N_{A \wedge C}}{N_{A \wedge C} + N_{\neg A \wedge C}} = \frac{N_{A \wedge C}}{N_C} = P(A|C) \quad (4)$$

$$Sup = \frac{N_{TP}}{N_{TP} + N_{FP} + N_{TN} + N_{FN}} = \frac{N_{A \wedge C}}{N_{A \wedge C} + N_{A \wedge \neg C} + N_{\neg A \wedge \neg C} + N_{\neg A \wedge C}} = \frac{N_{A \wedge C}}{N} = P(A \wedge C) \quad (5)$$

$$UncNeg = \frac{N_{TN}}{N_{TP} + N_{FP} + N_{TN} + N_{FN}} = \frac{N_{\neg A \wedge \neg C}}{N_{A \wedge C} + N_{A \wedge \neg C} + N_{\neg A \wedge \neg C} + N_{\neg A \wedge C}}$$

表 1 上側は事例集合を TP, FP, TN, FN に分ける混同行列．下側は分類ルール IF A THEN C に対する混同行列．Pos は Positive, Neg は Negative の略．N_X は条件 X を満たす事例数

Table 1 The upper part represents a confusion matrix which divides a set of instances into TP, FP, TN, and FN. The lower part represents a confusion matrix for a classification rule of IF A THEN C. Pos and Neg are the abbreviation for Positive and Negative, respectively. N_X is the number of instances which satisfies the X.

	Pos と予測	Neg と予測	合計
Pos が正解	N _{TP}	N _{FN}	N _{TP} + N _{FN}
Neg が正解	N _{FP}	N _{TN}	N _{FP} + N _{TN}
合計	N _{TP} + N _{FP}	N _{FN} + N _{TN}	N
	A と予測	¬A と予測	合計
C が正解	N _{A∧C}	N _{¬A∧C}	N _C
¬C が正解	N _{A∧¬C}	N _{¬A∧¬C}	N _{¬C}
合計	N _A	N _{¬A}	N

$$= \frac{N_{\neg A \wedge \neg C}}{N} = P(\neg A \wedge \neg C) \quad (6)$$

ここで、本研究の動機「なぜ客観的指標と人間の興味の間を調べるのか」を述べる．客観的指標は一般性や信頼性といった数式で定義可能なルールの性質を表し、ドメインにおけるルールの意味を考慮しない．よって、客観的指標と人間の興味の間で高性能の一致を求めても、有意義ではない．しかし、ある程度の性能で人間の興味を推定可能ならば、ユーザがルールを見て新知識を考え出す支援に客観的指標を活用できるかもしれない．よって、本研究を行うに至った．

3. 実験 I：髄膜炎ドメインでの実験

3.1 実験条件

実験 I では髄膜炎ドメインで客観的指標と医師によるルールの評価結果を比較し、客観的指標を持つ医師の興味を推定性能を調べる．具体的には、医師が評価したルールと同じものを客観的指標で評価し、両者の評価の一致度合いを定性的・定量的に求める．

過去の研究²³⁾では、髄膜炎の医療データ⁴²⁾から予後を表すルールの発見を試みた．マイニングシステム CAMLET¹⁾(分類学習基準は Accuracy)を適用してテキストベースの予後分類ルールを生成し、医師の評価結果を得た．クラスは二値、多値の 2 種類とした(たとえば二値では「原因はウイルスと細菌のどちらか」、多値では「原因はどの種類のウイルスか」等)．そして、新しい医学知識を提供しうるルールをいくつか発見できた．図 1 に、医師が興味を持ち医学的に価値があると判断したルールの一例を示す．

医師によるルール評価過程を以下に述べる．医師は個々のルールに対して医学的な解釈を行い、それをコメントとして記述した．その後、医師と本研究の実験者でコメントを見ながらルールの質ラベルを考案し、医師が各ルールに質ラベルの 1 つを与える作業を行った．質ラベルは、I：興味深い (Interesting), NU：理解不能 (Not Understandable), NI：興味がない

```
IF LOC ≤ 2.0 THEN C_COURSE = negative
IF LOC > 2.0 THEN C_COURSE = aphasia
```

図 1 医師が興味を持った髄膜炎予後を表すルールの一例．上側は、意識不明になってから 2 日以内に入院すれば言語障害を免れることを、下側はその裏命題を示唆している

Fig.1 An example of rules on meningitis prognosis to which a medical expert showed an interest. The upper part suggests that language disorder may be avoided if admitted to hospital within 2 days after losing consciousness. The lower part indicates the converse of contraposition.

表 2 本研究で使用した客観的指標。定義がルールの種類に依存する場合、分類ルールに対する定義を採用した。客観的指標ごとに、基礎となる理論、客観的指標の名称(略称)、[文献番号]、定義式を記載した。理論の列において、確：確率量、統：統計量、情：情報量、事：事例数、距：ルールや属性の距離

Table 2 The objective measures used in the present study. For the cases in which the definition of an objective measure depends on the type of rule, we adopted the definition of the objective measure for classification rules. Also shown are the base theories, the denomination (abbreviation) and [reference number] for each objective measure. The nomenclature in the theory column is as follows; P: probabilistic magnitude, S: statistic value, I: information amount, N: number of instances, D: distance among rules or among attributes.

確	Coverage (Coverage) ^{8),32)}	確	Prevalence (Prevalence) ²⁵⁾
	$P(A)$, $P(A)$: 前件部の生起確率.		$P(C)$, $P(C)$: 後件部の生起確率.
確	Precision (Precision) ^{8),32)}	確	Recall (Recall) ^{8),32)}
	$P(C A)$, $P(C A)$: 後件部の前件部条件付き確率.		$P(A C)$, $P(A C)$: 前件部の後件部条件付き確率.
確	Support (Support) ³²⁾	確	Uncovered Negative (UncNeg) ¹⁵⁾
	$P(A \wedge C)$, $P(A \wedge C)$: 前件部と後件部の共起確率.		$P(\neg A \wedge \neg C)$, $\neg A$, $\neg C$: 前件部, 後件部の否定.
確	Accuracy (Accuracy) ³²⁾	確	Specificity (Specificity) ³²⁾
	$P(A \wedge C) + P(\neg A \wedge \neg C)$		$P(\neg A \neg C)$
確	Lift (Lift) ⁴⁾	確	Added Value (AV) ⁴⁰⁾
	$\frac{P(C A)}{P(C)}$		$P(C A) - P(C)$
確	Leverage (Leverage) ³⁷⁾	確	Klösgen's Interestingness (KI) ³¹⁾
	$P(C A) - P(A) \times P(C)$		$\sqrt{P(A \wedge C) \times \{P(C A) - P(C)\}}$
確	Relative Risk (RR) ³⁾	確	Brin's Interest (BI) ⁶⁾
	$\frac{P(C A)}{P(C \neg A)}$		$\frac{P(A \wedge C)}{P(A) \times P(C)}$
確	Brin's Conviction (BC) ⁶⁾	確	Certainty Factor (CF) ⁴⁰⁾
	$\frac{P(A) \times P(\neg C)}{P(A \wedge \neg C)}$		$\frac{P(C A) - P(C)}{1 - P(C)}$
確	Jaccard Coefficient (Jaccard) ^{26),40)}	確	F-Measure (F-M) ³⁸⁾
	$\frac{P(A \wedge C)}{P(A) + P(C) - P(A \wedge C)}$		$\frac{2 \times P(C A) \times P(A C)}{P(C A) + P(A C)}$
確	Odds Ratio (OR) ⁴⁰⁾	確	Yule's Q (Yule's Q) ⁴⁰⁾
	$\frac{P(A \wedge C) \times P(\neg A \wedge \neg C)}{P(A \wedge \neg C) \times P(\neg A \wedge C)}$		$\frac{OR - 1}{OR + 1}$
確	Yule's Y (Yule's Y) ⁴⁰⁾	確	Kappa (Kappa) ^{13),40)}
	$\frac{\sqrt{OR} - 1}{\sqrt{OR} + 1}$		$\frac{P(A \wedge C) + P(\neg A \wedge \neg C) - P(A) \times P(C) - P(\neg A) \times P(\neg C)}{1 - P(A) \times P(C) - P(\neg A) \times P(\neg C)}$
確	Gray and Orłowska's Interestingness (GOI) ¹⁹⁾	確	Collective Strength (CST) ^{2),40)}
	$((\frac{P(C A)}{P(A) \times P(C)})^k - 1) \times ((P(A) \times P(C))^m)$ k, m : 独立性, 一般性の係数. $k = m$.		$\frac{P(A \wedge C) + P(\neg A \wedge \neg C)}{P(A) \times P(C) + P(\neg A) \times P(\neg C)} \times \frac{1 - P(A) \times P(C) - P(\neg C) \times P(\neg A)}{1 - P(A \wedge C) - P(\neg A \wedge \neg C)}$
確	Gini Index (Gini) ^{17),40)}		
	$P(A) \times \{P(C A)^2 + P(\neg C A)^2\} + P(\neg A) \times \{P(C \neg A)^2 + P(\neg C \neg A)^2\} - P(C)^2 - P(\neg C)^2$		
確	Goodman and Kruskal's Lambda-Asymmetric (GKL-A) ¹⁸⁾		
	$\sum_i \frac{MAX_j(P(A_i \wedge C_j) - MAX_j(P(C_j)))}{1 - MAX_j(P(C_j))}$, A_i : 属性 A の値. C_j : 属性 C の値.		
確事	Credibility (Credibility) ²²⁾	統	χ^2 Measure for One Quadrant (χ^2 - M1) ^{18),34)}
	$\beta_i \times P(C) \times P(R_i C) - P(R_i) \times T(R_i)$ $\beta_i = \frac{1}{2 \times P(R_i) \times (1 - P(R_i))}$ $P(R_i)$: ルール R_i の生起確率. $T(R_i)$: ルール R_i に含まれる事例数の閾値.		$\sum_{event} \frac{(T_{event} - O_{event})^2}{T_{event}}$ T_{event} : 事例 $event$ に含まれる事例数 (理論値). O_{event} : 事例 $event$ に含まれる事例数 (観察値). $event: A \rightarrow C$

統	χ^2 Measure for Four Quadrants (χ^2 -M4) ^{18),34)}	
	定義式は $\chi^2 - M1$ と同じ . ただし , $event: A \rightarrow C, A \rightarrow \neg C, \neg A \rightarrow C, \neg A \rightarrow \neg C$	
情	J-Measure (J-M) ³⁹⁾	情 K-Measure (K-M) ³⁵⁾
	$P(A) \times (I_{diff}(C A; C) + I_{diff}(\neg(C A); \neg C))$ $I_{diff}(X; Y) = P(X) \times \log_2 \frac{P(X)}{P(Y)}$	$I_{diff}(C A; C) + I_{diff}(\neg(C A); \neg C) -$ $I_{diff}(C A; \neg C) - I_{diff}(\neg(C A); C)$
情	Yao and Liu's Interestingness based on one-way support (YLI1) ^{43),44)}	
	$P(C A) \times \log_2 \frac{P(A \wedge C)}{P(A) \times P(C)}$	
情	Yao and Liu's Interestingness based on two-way support (YLI2) ^{43),44)}	
	$P(A \wedge C) \times \log_2 \frac{P(A \wedge C)}{P(A) \times P(C)}$	
情	Mutual Information (MI) ⁴⁰⁾	情 Yao and Zhong's Interestingness (YZI) ⁴⁴⁾
	$\log_2 \frac{P(A \wedge C)}{P(A) \times P(C)}$	$\sum_{X, Y} P(X \wedge Y) \times \log_2 \frac{P(X \wedge Y)}{P(X) \times P(Y)}$ $X: A, \text{もしくは} \neg A, Y: C, \text{もしくは} \neg C$
事	Cosine Similarity (CSI) ⁴⁰⁾	
	$\frac{N(A \wedge C)}{\sqrt{N(A) \times N(C)}}$, $N(A \wedge C)$: ルールに含まれる事例の数 . $N(A), N(C)$: 前件部, 後件部に含まれる事例の数 .	
事	Piatetsky-Shapiro's Interestingness (PSI) ³⁷⁾	
	$N(A \wedge C) - \frac{N(A) \times N(C)}{N(U)}$, $N(U)$: 全事例集合における事例の数 .	
事	Laplace Correction (LC) ⁴⁰⁾	事 ϕ Coefficient (ϕ) ⁴⁰⁾
	$\frac{N(A \wedge C) + 1}{N(A) + 2}$	$\frac{N(A \wedge C) \times N(\neg A \wedge \neg C) - N(\neg A \wedge C) \times N(A \wedge \neg C)}{\sqrt{N(A) \times N(C)} \times \sqrt{N(\neg A) \times N(\neg C)}}$
距	Gago and Bento's Interestingness (GBI) ¹⁶⁾	
	$\sum_{j=1}^{N_R} D(R_i, R_j)$ $\frac{\sum_{j=1}^{N_R} D(R_i, R_j)}{N_R}$, $R_i: i$ 番目のルール . N_R : ルール数 . $D(R_i, R_j)$: 属性の重複度合いに基づいて求めた i 番目と j 番目のルール間の距離 .	
距	Peculiarity (Peculiarity) ⁴⁵⁾	
	$\sum_{a_{ij} \in Rule} \frac{PF(a_{ij})}{N}$, $PF(a_{ij}) = \sum_k a_{ij} - a_{ik} ^\alpha$, a_{ij} : データ表の i 番目の行, j 番目の列の値 . N : ルールに含まれる属性の数 . α : 定数 (デフォルト 0.5) .	

(Not Interesting) である . その結果 , 二値クラスルール 112 個のうち 23 個 , 多値クラスルール 132 個のうち 25 個が **I** を得た . 客観的指標によるルール評価過程は以下のように設計した . すべての客観的指標で同じルールを評価し , 客観指標ごとに評価値の降順にルールをソートする . 医師が与えた質ラベルの数に応じて , 上位から順に **I** , **NU** , **NI** を割り振る .

3.2 実験結果の分析と考察

二値クラスの実験結果を図 2 に , 多値クラスの実験結果を図 3 に各々示す . 紙面の都合上 , **NI** と **NU** のルールは省略した . 客観的指標によるルールの評価結果が医師によるルールの評価結果とどの程度一致するかは , 定性的・定量的の両方から検討する .

定性的な分析では , 全体傾向の把握のため一致の仕方を可視化した . 客観的指標と医師で評価結果が完全に一致では白 , 確率的に一致ではグレー , 不一致では

黒を各ルールに配色した . 図 2 , 図 3 では , 白が多いほど客観的指標の推定性能が高い . 各客観的指標が持つ白・グレー・黒のパターンはその推定の特徴を表す . なお , 確率的な一致とは以下のとおりである . 多くのルールに同じ評価値を与える客観的指標もあった . たとえば二値クラスのルール 112 個のうち 23 個が医師には興味深かったが , Specificity は 74 個に同じ最も高い評価値を与えた . 74 個中 , 本当に興味深いルールは 15 個含まれていたが , それらが上位 23 個に配置されて **I** を与えられるかは , ルールの元の配置やソート手順によって異なる . そこで , 15 個のうち k 個が上位 23 個に配置される (正解のルールと一致する) 事象の生起確率を求めた . そして , 生起確率と一致数 k の積を $0 \leq k \leq 15$ で加算して求まる一致数の期待値を , Specificity の一致数と見なした . このような状況を確率的な一致と呼ぶことにした .

Rule ID	4	5	8	13	16	20	22	30	40	41	51	60	65	72	80	84	85	91	92	95	106	108	111	MC(I)	MC(N)	MC(O)	CMC	
Human	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I					
UncNeg																								0.5424	+ 0.8364	+ 0.4986	+ 0.6049	+
Accuracy																								0.5000	+ 0.8636	+ 0.4833	+ 0.5867	+
RR																								0.5000	+ 0.8636	+ 0.4738	+ 0.5844	+
Lift																								0.5000	+ 0.8333	+ 0.4517	+ 0.5713	+
Peculiarity																								0.5217	+ 0.8764	+ 0.3338	+ 0.5634	+
χ^2 -M1																								0.4783	+ 0.8652	+ 0.4319	+ 0.5634	+
AV																								0.3750	+ 0.8295	+ 0.3703	+ 0.4875	+
YLI1																								0.3750	+ 0.8295	+ 0.3327	+ 0.4781	+
K-M																								0.3913	+ 0.8427	+ 0.1905	+ 0.4540	+
J-M																								0.3478	+ 0.8315	+ 0.1396	+ 0.4167	+
OR																								0.3478	+ 0.8315	+ 0.1396	+ 0.4167	+
Yule's Q																								0.3478	+ 0.8315	+ 0.1396	+ 0.4167	+
MI																								0.2174	+ 0.7978	+ 0.1405	+ 0.3433	+
Kappa																								0.2609	+ 0.8090	+ 0.0379	+ 0.3421	+
χ^2 -M4																								0.2609	+ 0.8090	+ 0.0282	+ 0.3397	+
GBI																								0.2174	+ 0.7978	+ 0.1021	+ 0.3337	+
YLI2																								0.2609	+ 0.8090	+ -0.0005	+ 0.3325	+
KI																								0.2128	+ 0.7910	+ 0.0779	+ 0.3236	+
Gini																								0.2174	+ 0.7978	+ 0.0253	+ 0.3145	+
CST																								0.2174	+ 0.7978	+ -0.0130	+ 0.3049	+
YZI																								0.2174	+ 0.7978	+ -0.0514	+ 0.2953	+
Credibility																								0.1304	+ 0.7753	+ -0.0006	+ 0.2589	+
BC																								0.2887	+ 0.4567	+ -0.0527	+ 0.2453	+
Specificity																								0.2887	+ 0.4567	+ -0.0527	+ 0.2453	+
ϕ																								0.1304	+ 0.7753	+ -0.1532	+ 0.2207	+
Leverage																								0.1277	+ 0.7684	+ -0.1551	+ 0.2171	+
CF																								0.2526	+ 0.4496	+ -0.1191	+ 0.2089	+
Precision																								0.2526	+ 0.4496	+ -0.1191	+ 0.2089	+
Yule's Y																								0.2526	+ 0.4496	+ -0.1191	+ 0.2089	+
PSI																								0.0435	+ 0.7528	+ -0.2423	+ 0.1494	+
Prevalence																								0.0968	+ 0.6543	+ -0.2609	+ 0.1467	+
CSI																								0.0435	+ 0.7528	+ -0.2550	+ 0.1462	+
LC																								0.0426	+ 0.7458	+ -0.2522	+ 0.1447	+
BI																								0.0435	+ 0.7528	+ -0.2924	+ 0.1368	+
Coverage																								0.0435	+ 0.7528	+ -0.2934	+ 0.1366	+
F-M																								0.0435	+ 0.7528	+ -0.2934	+ 0.1366	+
GOI																								0.0435	+ 0.7528	+ -0.2934	+ 0.1366	+
Jaccard																								0.0435	+ 0.7528	+ -0.2934	+ 0.1366	+
Recall																								0.0435	+ 0.7528	+ -0.2934	+ 0.1366	+
Support																								0.0435	+ 0.7528	+ -0.2934	+ 0.1366	+
Lower Limit																								0.2054	+ 0.7946	+ 0.0000	+ 0.3013	+

図 2 髄膜炎ドメインの実験 I で二値クラスを用いた結果。各列はルールを、各行は客観的指標を表す
 Fig. 2 The results of Experiment I for binary classes. Each row represents a rule, while each column represents an objective measure.

定量的な分析では、推定性能の評価基準(メタ基準と呼ぶ)を以下に決めた。情報検索や機械学習の性能評価では、ROC 曲線や PR 曲線の AUC が一般的である¹⁰⁾。本研究では、客観的指標が医師が望むルールを漏れなく、かつ、間違いなく推定できるかを知りたいので、PR 曲線の AUC を検討する。客観的指標は、I, NU, NI の各々や、NU ∨ NI 等の組合せに対して PR 曲線を持ち得る。PR 曲線を描くには客観的指標が与えた評価値の降順にルールを並べ替え、質ラベルの境界を決めるルール数の閾値を設定する。閾値を徐々に変化させて (Recall, Precision) の点を描画する。図 4 に実際に得た PR 曲線の一例を示す。

Precision と Recall は独立ではないため、PR 曲線の補間方法と AUC の算出方法はいまだ議論されている¹⁰⁾。また、本研究では推定結果をユーザ支援に利用したいので、ユーザが疲労なく見られるルール数を考慮すべきである。I の PR 曲線でルール数の閾値が 100 のとき、Precision と Recall の値が最大にな

ると仮定しよう。このとき、客観的指標は 100 個ものルールを興味深いと推定しユーザに提示してしまい、Precision と Recall が最大でも最良の推定結果とはいえない。PR 曲線の AUC は (Recall, Precision) の全点の積分値なので、このようなルール数の制約を反映しにくい。

以上の検討より、今回は PR 曲線の AUC を用いず、医師が与えた質ラベルの個数を閾値とした点 (Recall, Precision) を最良の推定性能を示す代表点と見なす。そして、Precision と Recall の調和平均である F-Measure³⁸⁾ をメタ基準とする。たとえば医師が 20 個のルールに I を与えるならば、客観的指標の評価点上位 20 個に I を与え、その状態の点 (Recall, Precision) で F-Measure の値を求める。

Precision-Recall ベースのメタ基準を以下に定式化する。医師が質ラベル label を与えたルール集合 R_{label} を式 (7)、客観的指標が質ラベル label を与えたルール集合 \hat{R}_{label} を式 (8) と表す。R は全ルール集合、r

Rule ID	2	3	4	5	7	10	17	23	31	35	40	45	48	47	85	87	90	95	89	105	108	126	131	MC(I)	MC(N)	MC(C)	GMC	
Human	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Accuracy																									0.4706 +	0.8732 +	0.3313 +	0.5364 +
Peculiarity																									0.4800 +	0.8785 +	0.2981 +	0.5342 +
χ^2 -M1																									0.4706 +	0.8732 +	0.3218 +	0.5340 +
Kappa																									0.4400 +	0.8692 +	0.3528 +	0.5255 +
UncNeg																									0.4314 +	0.8638 +	0.3574 +	0.5210 +
RR																									0.4000 +	0.8598 +	0.3327 +	0.4981 +
Lift																									0.4000 +	0.8598 +	0.3249 +	0.4962 +
χ^2 -M4																									0.4000 +	0.8598 +	0.2690 +	0.4822 +
CST																									0.3922 +	0.8545 +	0.2747 +	0.4784 +
K-M																									0.3600 +	0.8505 +	0.2225 +	0.4482 +
J-M																									0.3200 +	0.8411 +	0.1469 +	0.4070 +
OR																									0.3200 +	0.8411 +	0.1469 +	0.4070 +
Yule's Q																									0.3200 +	0.8411 +	0.1469 +	0.4070 +
YLI1																									0.3214 +	0.8173 +	0.1499 +	0.4025 +
YLI2																									0.2800 +	0.8318 +	0.1371 +	0.3822 +
AV																									0.3103 +	0.8058 +	0.0941 +	0.3802 +
YZI																									0.2400 +	0.8224 +	0.0899 +	0.3481 +
Gini																									0.2308 +	0.8113 +	0.0110 +	0.3210 +
KI																									0.2308 +	0.8113 +	0.0110 +	0.3210 +
ϕ																									0.2000 +	0.8131 +	0.0054 +	0.3046 +
BC																									0.2941 +	0.5556 +	-0.0372 +	0.2766 +
Specificity																									0.2941 +	0.5556 +	-0.0372 +	0.2766 +
Credibility																									0.1600 +	0.8037 +	-0.0776 +	0.2615 +
Leverage																									0.1600 +	0.8037 +	-0.0791 +	0.2611 +
CF																									0.2772 +	0.5521 +	-0.0674 +	0.2598 +
Precision																									0.2772 +	0.5521 +	-0.0674 +	0.2598 +
Yule's Y																									0.2772 +	0.5521 +	-0.0674 +	0.2598 +
LC																									0.1569 +	0.7981 +	-0.0876 +	0.2561 +
GBI																									0.1200 +	0.7944 +	-0.0889 +	0.2364 +
MI																									0.1200 +	0.7944 +	-0.0889 +	0.2364 +
Recall																									0.1200 +	0.7944 +	-0.0889 +	0.2364 +
CSI																									0.1200 +	0.7944 +	-0.1263 +	0.2270 +
F-M																									0.1200 +	0.7944 +	-0.1263 +	0.2270 +
Jaccard																									0.1200 +	0.7944 +	-0.1263 +	0.2270 +
BI																									0.0800 +	0.7850 +	-0.0987 +	0.2116 +
Support																									0.0800 +	0.7850 +	-0.1735 +	0.1929 +
PSI																									0.0784 +	0.7793 +	-0.1743 +	0.1905 +
Coverage																									0.0769 +	0.7736 +	-0.1752 +	0.1881 +
GOI																									0.0000 +	0.7664 +	-0.2678 +	0.1246 +
Prevalence																									0.0430 +	0.4795 +	-0.3978 +	0.0419 +
Lower Limit																									0.1894 +	0.8106 +	0.0000 +	0.2973 +

図 3 髄膜炎ドメインの実験 I で多値クラスを用いた結果 . 構成は図 2 と同じである
 Fig. 3 The results of Experiment I for multivalued classes. The configuration is the same as Fig. 2.

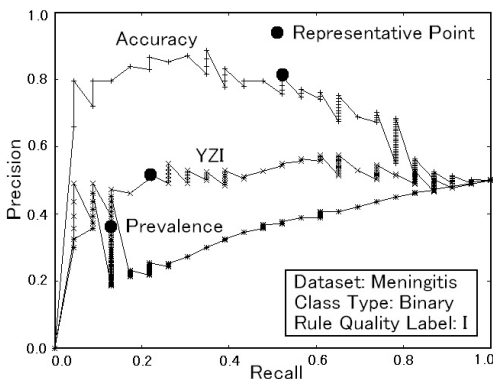


図 4 ルールの質ラベル I に対する PR 曲線の一つ例
 Fig. 4 Examples of PR curves for the quality label I.

は 1 つのルール, $L_H(r)$ はルール r に医師が与えた質ラベル, $L_M(r)$ は客観的指標が与えた質ラベルである . 客観的指標と医師で質ラベルが一致するルール集合 X_{label} を式 (9) と表す . よって, 質ラベル $label$ に

対するメタ基準は式 (10) となる . 興味深いルールの識別に比べ, 理解不能なルールと興味がないルールの識別はあまり重要ではない . そこで, 式 (10) の $label$ を I としたメタ指標 $MC(I)$, $NU \vee NI$ としたメタ指標 $MC(N)$ を導入する .

$$R_{label} = \{r \mid r \in R \wedge L_H(r) = label\} \quad (7)$$

$$\hat{R}_{label} = \{r \mid r \in R \wedge L_M(r) = label\} \quad (8)$$

$$X_{label} = \{r \mid r \in R_{label} \wedge r \in \hat{R}_{label}\} \quad (9)$$

$$MC(label) = \frac{2 \frac{|X_{label}|}{|R_{label}|} \frac{|X_{label}|}{|\hat{R}_{label}|}}{\frac{|X_{label}|}{|R_{label}|} + \frac{|X_{label}|}{|\hat{R}_{label}|}} \quad (10)$$

ルールの質ラベルの順序を考慮するため, 医師と客観的指標の間の質ラベルの相関係数もメタ基準に加える . 相関ベースのメタ基準 $MC(C)$ を, 同順位補正を含むスピアマンの順位相関係数⁵⁾ に基づき式 (11) に定義する . $RANK_H(r)$ は, 医師が与えた質ラベルを持つルールセット R_{label} におけるルール r の順位を

意味する． $RANK_M(r)$ は、客観的指標が与えた質ラベルを持つルールセット \hat{R}_{label} におけるルール r の順位を意味する． S, K は式 (11) の簡易化のために用いた． N_R は全ルール数、 T_H は R_{label} の同順位補正項、 T_M は \hat{R}_{label} の同順位補正項である．

$$MC(C) = \frac{K - S}{\sqrt{K - 2T_H} \sqrt{K - 2T_M}} \quad (11)$$

where

$$S = \sum_{I, NU, NI} (RANK_H(r) - RANK_M(r))^2$$

$$K = \frac{N_R(N_R^2 - 1)}{6},$$

$$T_H = \frac{\sum N_{T_H}(N_{T_H}^2 - 1)}{12}, \quad \text{and}$$

$$T_M = \frac{\sum N_{T_M}(N_{T_M}^2 - 1)}{12}$$

$MC(I), MC(N), MC(C)$ を集約する総合メタ基準を式 (12) に定義する．知識発見では、ルール集合でデータセットを正確にモデル化するよりも、1 つでもよいのでドメインにとって価値あるルールを見つけ出す方が重要な場合も多い．少数の価値あるルールを優先するため、モデルの正確さの犠牲を迫られることもある．よって、理解不能なルールや興味がないルールの推定性能や全ルール、質ラベル順序の推定性能よりも、興味深いルールの推定性能を重視し、 $MC(I), MC(N), MC(C)$ に 2 : 1 : 1 の比率で重み W_I, W_N, W_C を割り当て加算する．

$$CMC = \sum_{i=I, N, C} W_i \times MC(i) \quad (12)$$

where $W_I = 0.50$, and $W_N = W_C = 0.25$

客観的指標の絶対的な有用性はドメインに依存し、厳密に見積もることは難しい．また、髄膜炎に特化した有用性よりも、ある程度汎化的に有用性を見積もる方が役立つであろう．そこで、ルールの質ラベルのランダム割当てを想定してメタ基準の下限値を定義し、これに対する客観的指標の相対的な有用性を見積もる．Recision-Recall ベースのメタ基準の下限値は次の段落で述べる．相関ベースのメタ基準の下限値は無相関を意味する 0 とする．総合メタ基準の下限値は、各メタ基準の下限値を式 (12) に代入して求める．

Precision-Recall ベースのメタ基準の下限値は、ランダムラベル割当ての推定成功率の期待値である式 (13) で求める． ${}_a C_b$ は a 個のルールから b 個を取り出す組合せ数 $\frac{a!}{b!(a-b)!}$ である． N_R は全ルール数、 N_{label} は医師が質ラベル $label$ を与えたルール数、かつ、客観的指標が質ラベル $label$ を与えたルール数で

もある． ${}_{N_R} C_{N_{label}}$ はランダムラベル割当てが生成する全事象数である． ${}_{N_{label}} C_i$ と $({}_{N_R - N_{label}} C_{(N_{label} - i)})$ の積は、 N_{label} 個のルール中 i 個でランダムラベル割当てが推定成功する事象数である．式 (13) の第 1 項は N_{label} 個のルール中 i 個で推定成功する確率、第 2 項はそのときのメタ基準値である．

Lower Limit(label)

$$= \sum_{i=0}^{N_{label}} \frac{{}_{N_{label}} C_i ({}_{N_R - N_{label}} C_{(N_{label} - i)})}{{}_{N_R} C_{N_{label}}} \times \frac{2 \frac{i}{N_{label}} \frac{i}{N_{label}}}{\frac{i}{N_{label}} + \frac{i}{N_{label}}} \quad (13)$$

定量的な分析では、客観的指標ごとにメタ基準値、総合メタ基準値、これらの下限値を求め、図 2 と図 3 の右側に掲載した．客観的指標は総合メタ基準値の降順にソートした．メタ基準値や総合メタ基準値が下限値よりも大きい場合は、右横に記号+を付加した．

図 2 の CMC を見ると、全 40 個の客観的指標のうち 20 個 (50%) が記号+を得ており、ランダムラベル割当てよりも推定性能が高い．上位 5 位は UncNeg, Accuracy, RR, Lift, Peculiarity, $\chi^2 - M1$ である．これらの CMC は下限値の 1.8 倍を超えており (ランダムラベル割当てよりも 1.8 倍正解しやすく)、医師の興味推定に有用である可能性が示唆された．下位 5 位は Support, Recall, Jaccard, GOI, F-M, Coverage である． CMC は下限値の 0.5 倍未満であり (ランダムラベル割当ての 0.5 倍しか正解できず)、今回の限られた実験条件では断言できないが、有用でない可能性がある．

図 3 の CMC を見ると、20 個 (50%) が記号+を得ており、ランダムラベル割当てよりも推定性能が高い．上位 5 位は Accuracy, Peculiarity, $\chi^2 - M1$, Kappa, UncNeg である．これらの CMC は下限値の 1.7 倍を超え、医師の興味推定に有用である可能性が示唆された．下位 5 位は Prevalence, GOI, Coverage, PSI, Support である．これらの CMC は下限値の 0.7 倍未満であり、有用でない可能性がある．

次に図 2, 図 3, 論文後半に掲載した表 3 から、二値クラスと多値クラスの包括的な傾向を検討する．UncNeg, Accuracy, Peculiarity, $\chi^2 - M1$, RR は二値と多値のクラスに共通して高い推定性能を持ち、有用である可能性が高まった．一方、Prevalence, GOI, Coverage, Support, PSI は共通して低い推定性能を持ち、有用性でない可能性が高まった．図 2, 図 3 とともに、順位に近い客観的指標の白・グレー・黒のパターンは類似しており、医師によるルール評価が安定して

いたと示唆される．これは「確実な仮説を立て、それに基づいてルールを評価した」という医師のコメントに合致する．共通して上位にあった *UncNeg*, *Accuracy*, *Peculiarity*, $\chi^2 - M1$, *RR* は、医師が確実な仮説を持ちつつ医学知識発見を試みる際、興味深いであろうルールの推奨に役立つと考えられる．

我々は医師のルール評価過程を観察し、これが2つの段階に分かれると気づいた．第1段階は仮説生成であり、医師は様々な仮説の元を思い付き抽象的なレベルで仮説を選択する．第2段階は仮説検証であり、具象的なレベルで仮説の信頼性を高める．この段階分けは、文献 12), 20) で提唱され、その後、創造的思考に関する様々な研究^{11), 28) ~ 30)} で妥当性が確認された発散思考と収束思考に合致する．仮説生成と仮説検証を時系列に配置すれば、ルール評価を円滑化し医学知識発見を促進できる可能性がある．

4. 実験 II : 肝炎ドメインでの実験

4.1 実験条件

実験 I と同様の手続きで、肝炎ドメインにおける客観的指標の推定性能を調べる．過去の研究³⁶⁾ では、肝炎の医療データ⁴¹⁾ から病状変化を表すルールの発見を試みた．マイニングシステムには典型的な時系列データマイニング手法の1つ⁹⁾ を用いた．これは、時間窓による検査履歴からの時系列パターン切り出し、クラスタリングによる代表パターン抽出、分類学習による予後分類ルール (過去の代表パターンの組合せから現在の代表パターンの組合せを予測するグラフベースルール) の生成からなる．クラスタリングには EM アルゴリズムと K-Means アルゴリズム (距離基準はユークリッド距離) を、分類学習には C4.5 の商用版 C5.0 (分類学習基準は情報利得比) を用いた．

マイニング過程は仮説生成・検証の2段階に分けられ、予後分類ルールは医師 (実験 I とは別人) が評価した．仮説生成に対応する1回目のマイニングでは、医師が肝炎の病状に関する仮説を立てることに貢献したルールを得た．仮説検証に対応する2回目のマイニングでは、医師の仮説検証に貢献したルールを得た．図5に、医師の仮説生成・検証に貢献したルールの一例を示す．医師はこれらに興味を持ち、医学的に価値があると判断した．

医師によるルール評価過程は実験 I とほぼ等しいが、医師より、仮説に至るルールと仮説までは至らないが関心を持ったルールの2種類があると指摘を受け、EI: 特に興味深い (Especially Interesting) というルールの質ラベルを加えた．EI は仮説の生成・検証

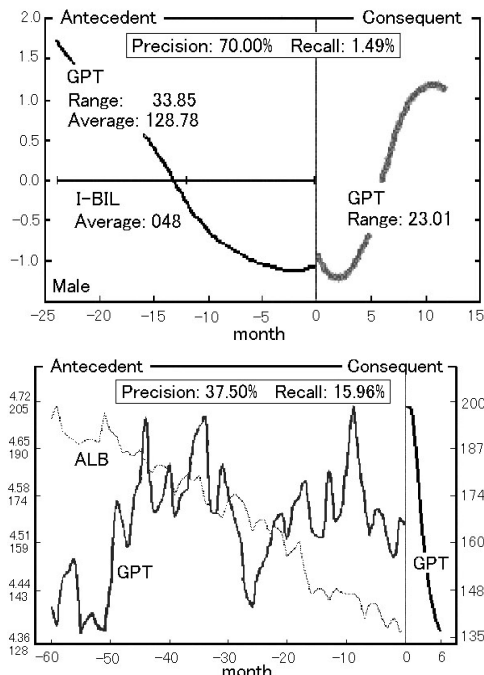


図5 医師が興味を持った肝炎病状を表すルールの一例．上側は1回目のマイニングで得られたルール．病状把握に重要な検査項目 GPT の値が約3年の周期で時間変動することを示唆している．下側は2回目のマイニングで得られたルール．GPTの3年周期時間変動をより明確に示唆している

Fig. 5 An example of rules on hepatitis prognosis to which a medical expert showed an interest. The upper part of the figure corresponds to the first mining. This shows that the value of GPT, which is a necessary medical test to understand the symptoms of hepatitis, changes over a period of around three years. The lower part of the figure corresponds to the second mining, and more clearly shows the three-year-period described above.

に貢献したルールを意味する．仮説の生成・検証には貢献しなかったが医師が興味を持ったルールには、EIではなくIを与えた．結果として、1回目のマイニングでは、ルール30個のうち3個がEIを、9個がIを得た．2回目のマイニングでは、ルール21個のうち2個がEIを、6個がIを得た．客観的指標によるルール評価過程は実験 I と同様である．

4.2 実験結果の分析と考察

1回目のマイニングの実験結果を図6に、2回目のマイニングの実験結果を図7に各々示す．紙面の都合上、NIとNUのルールは省略した．客観的指標と医師の間でルールの評価結果がどの程度一致するかは、定性的・定量的の両方から検討する．

ここで、新たに加えた EI に対するメタ基準 $MC(EI)$ を定義し、メタ基準 $MC(I)$ 、総合メタ基

Rule ID	2	3	11	4	5	8	12	13	22	23	24	27	MC(EI)	MC(I)	MC(N)	MC(C)	CMC
Human	EI	EI	I	I	I	I	I	I	I	I	I	I					
F-M													0.3333 +	0.6667 +	0.7778 +	0.4704 +	0.5830 +
Jaccard													0.3333 +	0.6667 +	0.7778 +	0.4704 +	0.5830 +
Recall													0.3333 +	0.6667 +	0.7778 +	0.4403 +	0.5770 +
CST													0.3333 +	0.6667 +	0.7778 +	0.4263 +	0.5742 +
Kappa													0.3333 +	0.6667 +	0.7778 +	0.3381 +	0.5565 +
YL12													0.3333 +	0.5833 +	0.7222 +	0.3333 +	0.5111 +
Gini													0.0000	0.5833 +	0.7222 +	0.3144 +	0.4407 +
χ^2 -M4													0.0000	0.5833 +	0.7222 +	0.1947 +	0.4167 +
χ^2 -M1													0.0000	0.5833 +	0.7222 +	0.1947 +	0.4167 +
Leverage													0.3333 +	0.4167 +	0.6111 +	0.1773 +	0.3910 +
PSI													0.3333 +	0.4167 +	0.6111 +	0.1773 +	0.3910 +
YZI													0.0000	0.5000 +	0.6667 +	0.2764 +	0.3886 +
Support													0.3333 +	0.4000 +	0.5714 +	0.1818 +	0.3773 +
Peculiarity													0.3333 +	0.4167 +	0.6111 +	0.0875 +	0.3730 +
J-M													0.0000	0.5000 +	0.6667 +	0.1962 +	0.3726 +
CSI													0.0000	0.5000 +	0.6667 +	0.1515 +	0.3636 +
Accuracy													0.0000	0.5000 +	0.6667 +	0.0875 +	0.3508 +
Coverage													0.3333 +	0.3846 +	0.5294 +	0.0854 +	0.3435 +
RR													0.0000	0.5000 +	0.6667 +	0.0355 +	0.3404 +
Lift													0.0000	0.5000 +	0.6667 +	0.0070 +	0.3347 +
Yule's Y													0.0000	0.5000 +	0.6667 +	0.0000	0.3333 +
GBI													0.0000	0.5000 +	0.6667 +	-0.0165	0.3300 +
YL11													0.0000	0.5000 +	0.6667 +	-0.0431	0.3247 +
KI													0.0000	0.4167 +	0.6111 +	0.0686 +	0.3026 +
UncNeg													0.0000	0.4167 +	0.6111 +	0.0213 +	0.2931
K-M													0.0000	0.4167 +	0.6111 +	0.0118 +	0.2913
ϕ													0.0000	0.4167 +	0.6111 +	-0.0307	0.2827
AV													0.0000	0.4167 +	0.6111 +	-0.0493	0.2790
BC													0.0000	0.4167 +	0.6111 +	-0.0493	0.2790
MI													0.0000	0.3333 +	0.5556 +	0.0449 +	0.2534
Specificity													0.0000	0.4167 +	0.6111 +	-0.2055	0.2478
GOI													0.0000	0.3333 +	0.5556 +	-0.0071	0.2430
OR													0.0000	0.3333 +	0.5556 +	-0.1064	0.2232
Yule's Q													0.0000	0.3333 +	0.5556 +	-0.1064	0.2232
LC													0.0000	0.3333 +	0.5556 +	-0.1097	0.2225
CF													0.0000	0.3333 +	0.5556 +	-0.1127	0.2219
Precision													0.0000	0.3333 +	0.5556 +	-0.1408	0.2163
Credibility													0.0000	0.3333 +	0.5556 +	-0.2435	0.1957
BI													0.0000	0.2500 +	0.5000 +	-0.0827	0.1835
Prevalence													0.0000	0.3077 +	0.4706 +	-0.2666	0.1639
Lower Limit													0.1000	0.4000	0.6000	0.0000	0.3000

図 6 肝炎ドメインの実験 II における第 1 回目のマイニングの結果。構成は図 2 と同じである

Fig. 6 The results of Experiment II in the first mining. The configuration is the same as Fig. 2.

準 *CMC* を定義し直す。特に興味深いルールと興味深いルールの漏れない検出に比べ、両者の識別はあまり重要ではない。そこで、 $MC(EI)$ は式 (10) の *label* を *EI* としたもので、 $MC(I)$ は $EI \vee I$ としたもので、 CMC は式 (12) の総和を $i = EI, I, N, C$ で行ったものとする。仮説の生成・検証に貢献した/しないにかかわらず、医師が興味を持ったルールの推定性能を重視するため、 $MC(EI)$ 、 $MC(I)$ 、 $MC(N)$ 、 $MC(C)$ には、1:2:1:1 の比率で重み W_{EI} 、 W_I 、 W_N 、 W_C を割り当てる。他の式や手続きは実験 I と等しい。

図 6 の *CMC* を見ると、全 40 個の客観的指標のうち 24 個 (60%) が記号+を得ており、ランダムラベル割当てよりも推定性能が高い。上位 5 位は F-M, Jaccard, Recall, CST, Kappa である。これらの *CMC* は下限値の 1.8 倍を超えており (ランダムラベル割当てよりも 1.8 倍正解しやすく)、医師の興味推定に有用である可能性が示唆された。下位 5 位は Prevalence, BI, Credibility, Precision, CF である。これ

らの *CMC* は下限値の 0.8 倍未満であり (ランダムラベル割当ての 0.8 倍しか正解できず)、今回の限られた実験条件では断言できないが、これらは肝炎ドメインで有用でない可能性がある。

図 7 の *CMC* を見ると、15 個 (38%) が記号+を得ており、ランダムラベル割当てよりも推定性能が高い。上位 5 位は UncNeg, Accuracy, RR, Peculiarity, Lift である。これらの *CMC* は下限値の 1.8 倍を超え、医師の興味推定に有用である可能性が示唆された。下位 5 位は Prevalence, GOI, LC, ϕ , KI である。これらの *CMC* は下限値の 0.3 倍未満であり、医師の興味推定に有用でない可能性がある。

次に図 6, 図 7, および表 3 から、1 回目と 2 回目のマイニングの包括的な傾向を検討する。客観的指標の推定性能にばらつきは見られたが、F-M, Jaccard, Recall, CST, Kappa は共通して比較的高い推定性能を持ち、有用性である可能性が高まった。一方、Prevalence, LC, GOI, ϕ , BI は共通して比較的低い推定

Rule ID	13	21	14	15	17	18	19	MC(E)	MC(D)	MC(N)	MC(C)	CMC
Human	EI	EI	I	I	I	I	I					
UncNeg								0.0000	0.7500 +	0.8462 +	0.6531 +	0.5999 +
Accuracy								0.0000	0.7500 +	0.8462 +	0.6088 +	0.5910 +
RR								0.0000	0.7500 +	0.8462 +	0.6088 +	0.5910 +
Peculiarity								0.0000	0.7500 +	0.8462 +	0.3832 +	0.5459 +
Lift								0.0000	0.7500 +	0.8462 +	0.3197 +	0.5332 +
χ^2 -M1								0.0000	0.6667 +	0.7500 +	0.3412 +	0.4849 +
CST								0.0000	0.5000 +	0.6923 +	0.2623 +	0.3909 +
F-M								0.0000	0.5000 +	0.6923 +	0.2623 +	0.3909 +
Jaccard								0.0000	0.5000 +	0.6923 +	0.2623 +	0.3909 +
Kappa								0.0000	0.5000 +	0.6923 +	0.2623 +	0.3909 +
Recall								0.0000	0.5000 +	0.6923 +	0.2590 +	0.3903 +
PSI								0.4000 +	0.3750	0.6154	0.0454 +	0.3622 +
YL1								0.0000	0.5217 +	0.4211	0.1813 +	0.3292 +
Support								0.4000 +	0.3333	0.5000	-0.0032	0.3127 +
Coverage								0.5000 +	0.3158	0.4348	-0.0324	0.3068 +
AV								0.0000	0.4545 +	0.4000	0.0985 +	0.2815
YL12								0.0000	0.3750	0.6154	0.0263 +	0.2783
GBI								0.0000	0.3750	0.6154	0.0201 +	0.2771
J-M								0.0000	0.5517 +	0.0000	0.2259 +	0.2659
OR								0.0000	0.5517 +	0.0000	0.2259 +	0.2659
Yule's Q								0.0000	0.5517 +	0.0000	0.2259 +	0.2659
CSI								0.0000	0.3529	0.5600	0.0075 +	0.2547
Credibility								0.0000	0.4211 +	0.5217	-0.2262	0.2275
χ^2 -M4								0.0000	0.2500	0.5385	0.0584 +	0.2194
Leverage								0.0000	0.3333	0.5000	-0.0811	0.2171
BC								0.0000	0.4545 +	0.4000	-0.2439	0.2130
OF								0.0000	0.4545 +	0.4000	-0.2439	0.2130
Precision								0.0000	0.4545 +	0.4000	-0.2439	0.2130
Specificity								0.0000	0.4545 +	0.4000	-0.2439	0.2130
Yule's Y								0.0000	0.4545 +	0.4000	-0.2439	0.2130
K-M								0.0000	0.5517 +	0.0000	-0.0984	0.2030
Gini								0.0000	0.2500	0.5385	-0.2631	0.1551
YZI								0.0000	0.2500	0.5385	-0.2631	0.1551
MI								0.0000	0.2500	0.5385	-0.3042	0.1469
BI								0.5000 +	0.1905	0.1905	-0.4195	0.1304
KI								0.0000	0.2222	0.4167	-0.4607	0.0801
ϕ								0.0000	0.1176	0.4000	-0.5470	0.0177
LC								0.0000	0.1176	0.4000	-0.5573	0.0156
GOI								0.0000	0.1111	0.3333	-0.5536	0.0004
Prevalence								0.1429 +	0.1000	0.1818	-0.7147	-0.0380
Lower Limit								0.0952	0.3810	0.6190	0.0000	0.2952

図 7 肝炎ドメインの実験 II における第 2 回目のマイニングの結果。構成は図 2 と同じである
 Fig. 7 The results of Experiment II in the second mining. The configuration is the same as Fig. 2.

性能を持ち、有用でない可能性が高まった。

図 7 では順位に近い客観的指標の白・グレー・黒のパターンが類似しているが、図 6 ではパターンが異なり、全体的に細かいモザイク状を呈している。これより、1 回目のマイニングでは医師によるルール評価が不安定であったこと、2 回目では安定したことが見て取れる。これらは「評価基準を様々に変えながらルールを評価し、曖昧な仮説から確実な仮説を導き出そうとした」という医師のコメントに合致する。

以上の議論から、医師が思考を発散・収束できるように多様な視点を与える支援、医師が持つ非明示的な興味を明示的な形で示し気づきを促す支援が、医学知識発見を促進すると推測される。また、客観的指標の推定特徴が多様な場合、相補関係にある客観的指標を組み合わせれば推定性能の向上を期待できる。

表 3 における実験 I, II の総合順位を見ると、上位 5 位は Accuracy, UncNeg, Peculiarity, RR, χ^2 -M1 である。これらの CMC は下限値の 1.6 倍を超えてお

り(ランダムラベル割当てよりも 1.6 倍正解しやすく)、実験 I, II でも比較的上位にあった。これらは医師の興味推定に有用と考えられる。下位 5 位は Prevalence, GOI, LC, BI, ϕ である。これらの CMC は下限値の 0.7 倍未満であり(ランダムラベル割当ての 0.7 倍しか正解できず)、実験 I, II でも比較的下位にあった。これらは今回のドメインでは有用でない可能性が高い。

1 位は Support と UncNeg の和である Accuracy、2 位は UncNeg であった。表 2 に示すように、Support は前件部と後件部がともに成り立つ事例が多いほど、UncNeg は前件部と後件部がともに成り立たない事例が多いほどルールに高い評価値を与える。表 3 から、UncNeg の推定性能は高いが、Support の推定性能はあまり高くないと分かる。以上より、今回は「ルールがあたる以上に、はずれないことを重視した」と推測される。上位 5 位のうち、Peculiarity のみがルールの独自性を表し、他はルールの正確性を表

表 3 実験 I, II の結果をまとめたもの (上位と下位の抜粋). 記号+は下限値より大きいことを意味する
 Table 3 The summary of the results of Experiment I and II. A plus sign + to the right of a CMC value means that the value is larger than the lower boundary of CMC.

実験 I: 髄膜炎			実験 II: 肝炎			実験 I, II の総合		
順位	客観的指標	CMC 平均	順位	客観的指標	CMC 平均	順位	客観的指標	CMC 平均
1	UncNeg	0.5630 +	1	F-M	0.4870 +	1	Accuracy	0.5162 +
2	Accuracy	0.5616 +	1	Jaccard	0.4870 +	2	UncNeg	0.5047 +
3	Peculiarity	0.5488 +	3	Recall	0.4836 +	3	Peculiarity	0.5041 +
4	χ^2 -M1	0.5487 +	4	CST	0.4825 +	4	RR	0.5035 +
5	RR	0.5412 +	5	Kappa	0.4737 +	5	χ^2 -M1	0.4998 +
6	Lift	0.5337 +	6	Accuracy	0.4709 +	6	Lift	0.4838 +
7	K-M	0.4511 +	7	RR	0.4657 +	7	Kappa	0.4538 +
8	YLI1	0.4403 +	8	Peculiarity	0.4595 +	8	CST	0.4371 +
9	Kappa	0.4338 +	9	χ^2 -M1	0.4508 +	9	YLI1	0.3836 +
10	AV	0.4338 +	10	UncNeg	0.4465 +	10	YLI2	0.3761 +
31	CSI	0.1866	31	CF	0.2175	31	MI	0.2450
32	Recall	0.1865	32	Precision	0.2147	32	Coverage	0.2437
33	F-M	0.1818	33	Credibility	0.2116	33	Credibility	0.2359
33	Jaccard	0.1818	34	MI	0.2001	34	CF	0.2259
35	BI	0.1742	35	KI	0.1913	35	Precision	0.2245
36	PSI	0.1699	36	BI	0.1569	36	ϕ	0.2064
37	Support	0.1647	37	ϕ	0.1502	37	BI	0.1656
38	Coverage	0.1623	38	GOI	0.1217	38	LC	0.1597
39	GOI	0.1306	39	LC	0.1190	39	GOI	0.1262
40	Prevalence	0.0943	40	Prevalence	0.0629	40	Prevalence	0.0786
	Lower Limit	0.2993		Lower Limit	0.2976		Lower Limit	0.2985

すものである。本実験では、独自性の客観的指標として GBI と Peculiarity を用いた。表 2 に示すように、GBI は他のルールと属性が異なるほど、Peculiarity は他のルールと属性値が異なるほどルールに高い評価値を与える。表 3 から、Peculiarity の推定性能は高いが、GBI の推定性能はあまり高くないと分かる。以上より、医師は「ルールの正確さに加えて珍しさも重視した」「他のルールと異なる属性の有無以上に、属性値の差異をルールの珍しさの判断基準とした」と推測される。

ここで、客観的指標と人間の興味の間関係を調べた関連研究^{7),35)}を参照し、本実験の結果の一般性について検討する。文献 35) では 1 種類のデータセットのみを用いたため、実験結果の一般性を議論できなかった。そこで、文献 7) では異なる 8 種類のデータセットを用いたが、客観的指標の推定性能はデータセットに依存し、結果的に一貫した傾向を得られなかった。一方、本実験では医療的を絞ることで、実験 I, II においてある程度一貫した傾向を見出せた。なお、文献 35) では Recall を、文献 7) では相関係数をメタ基準に用いており、このことも関連研究と本研究の実験結果の差異に影響したと考えられる。実験 I, II で一貫した傾向が現れたことから、以下の 3 点が示唆される。医師間で評価のずれが大きくない可能性。髄膜炎

と肝炎のドメイン間でルールの良し悪しに共通点がある可能性。今回のメタ基準が妥当である可能性。本実験のみでは、結果を複数の医師や医療ドメイン全体に一般化はできないが、少なくとも、共通の傾向を見出しうるユーザ・ドメインの切り分け、および、メタ基準の定義に関する指針を示せたといえる。

上記の議論から、今後の研究の 2 つの方向性が示唆される。1 つ目は、メタ基準を確立して一般的な傾向を得るべく、共通の傾向を見出しうるユーザ内、ドメイン内で複数のデータセットを用いた実験を積み重ねていく方向である。2 つ目は、人間の興味の可変性や個人性を許容して知識発見過程を支援すべく、適応的に客観的指標を活用する方向である。以下に後者の方向を検討する。実験 I, II では、知識発見支援の枠組みを考えるうえで重要な 2 つの情報を得られた。1 つ目は「医学的な意味を含まなくても、数個の客観的指標、および客観的指標の組合せは医師の興味をある程度推定しうる」という情報である。2 つ目は「医師は仮説生成でいろいろな考えを巡らせ、仮説検証で考えを絞り込む思考過程をたどる」という情報である。これらの情報より、作り込んだ客観的指標をマイニングアルゴリズムに固定し、ルールを自動生成する枠組みには限界があると示唆される。少数の単純な客観的指標でルールを生成し、後処理で複数の客観的指標を柔

軟に組み合わせるとルールのソート提示や推奨すれば、医学知識発見の支援効果が高いであろう。この支援には、時々刻々と移り変わる医師の思考に沿うように、医師の能動的な介入を許す半自動的な枠組みが望まれる。我々は現在、この枠組みに基づく知識発見支援のユーザインタフェースを開発しつつある。

5. おわりに

本研究では、髄膜炎と肝炎の検査履歴データから得たルールと医師によるルールの評価結果を用いた実験で、客観的な興味深さ指標が本当に興味深いルールの発見に役立つかを検証した。具体的には、興味深さ指標と医師のルール評価の一致度合いから、興味深さ指標が持つ医師の興味の推定性能を見積もった。その結果、Accuracy, UncNeg, Peculiarity, RR, χ^2 -M1が医師の興味の推定に役立つ可能性、および、複数の興味深さ指標をライブラリ化し組み合わせた医学知識発見支援の可能性が示唆された。今後は、興味深さ指標を組み合わせるとルール評価時の医師の思考を支援するユーザインタフェースを開発し、有効性の検証と他の医療データへの適用を試みたい。

謝辞 本研究は、科研費特定領域(2)13131205, 科研費若手(B)17700162, 文科省学術フロンティア「医学研究の新展開(同志社大)」の支援を受けた。

参考文献

- 1) Abe, H. and Yamaguchi, T.: Constructive Meta-Learning with Machine Learning Method Repository, *Proc. Int'l Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE-2004)*, pp.502-511 (2004).
- 2) Aggrawal, C. and Yu, P.: A New Framework for Itemset Generation, *Proc. ACM Symposium on Principles of Database Systems (PODS-1998)*, pp.18-24 (1998).
- 3) Ali, K., Manganaris, S. and Srikant, R.: Partial Classification Using Association Rules, *Proc. Int'l Conf. on Knowledge Discovery and Data Mining (KDD-1997)*, pp.115-118 (1997).
- 4) Bayardo, R., Agrawal, R. and Gunopulos, D.: Constraint-Based Rule Mining in Large, Dense Databases, *Data Mining and Knowledge Discovery Journal*, pp.217-240 (2000).
- 5) Brazdil, B.P. and Soares, C.: A Comparison of Ranking Methods for Classification Algorithm Selection, *Proc. Euro. Conf. on Machine Learning (ECML-2000)*, pp.63-74 (2000).
- 6) Brin, S., Motwani, R. and Silverstein, C.: Be-

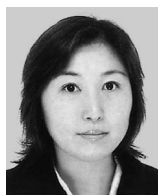
- yond Market Baskets: Generalizing Association Rules to Correlations, *Proc. ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD-1997)*, pp.265-276 (1997).
- 7) Carvalho, D.R., Freitas, A.A. and Ebecken, N.: Evaluating the Correlation between Objective Rule Interestingness Measures and Real Human Interest, *Proc. Euro. Conf. on Principles of Data Mining and Knowledge Discovery (PKDD-2005)*, pp.453-461 (2005).
- 8) Chen, M., Zheng, A., Lloyd, J., Jordan, M. and Brewer, E.: Failure Diagnosis Using Decision Trees, *Proc. Int'l Conf. on Autonomic Computing (ICAC-2004)*, pp.36-43 (2004).
- 9) Das, G., King-Ip, L., Heikki, M., Renganathan, G. and Smyth, P.: Rule Discovery from Time Series, *Proc. Int'l Conf. on Knowledge Discovery and Data Mining (KDD-1998)*, pp.16-22 (1998).
- 10) Davis, J. and Goadrich, M.: The Relationship Between Precision-Recall and ROC Curves, UW-CS Technical Report, TR1551 (2006).
- 11) Duch, W.: Computational Creativity, *Proc. IEEE World Congress on Computational Intelligence (WCCI-2006)*, pp.435-442 (2006).
- 12) Finke, R.A., Ward, T.B. and Smith, S.M.: *Creative Cognition—Theory, Research and Applications*, MIT Press, Cambridge (1992).
- 13) Fleiss, J.: *Statistical Methods for Rates and Proportions*, John Wiley and Sons (1981).
- 14) Freitas, A.A.: On Rule Interestingness Measures, *Knowledge-Based Systems*, Vol.12, No.5-6, pp.309-315 (1999).
- 15) Fürnkranz, J. and Flach, P.: ROC 'n' rule learning — Towards a better understanding of covering algorithms, *Machine Learning*, Vol.58, pp.39-77 (2005).
- 16) Gago, P. and Bento, C.: A Metric for Selection of the Most Promising Rules, *Proc. Euro. Conf. on the Principles of Data Mining and Knowledge Discovery (PKDD-1998)*, pp.19-27 (1998).
- 17) Gini, C.: *Variability and Mutability*, Tipografia di Paolo Cuppin, Bologna (1912) (in Italian).
- 18) Goodman, L.A. and Kruskal, W.H.: *Measures of association for cross classifications*, Springer Series in Statistics, 1, Springer-Verlag (1979).
- 19) Gray, B. and Orlowska, M.E.: CCAIIA: Clustering Categorical Attributes into Interesting Association Rules, *Proc. Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD-1998)*, pp.132-143 (1998).
- 20) Guilford, J.P.: The Structure of Intellect, *Psy-*

- chological Bulletin*, Vol.53, pp.267–293 (1956).
- 21) Hamilton, H.J. and Fudger, D.F.: Estimating DBLearn's Potential for Knowledge Discovery in Databases, *Computational Intelligence*, Vol.11, No.2, pp.280–296 (1995).
 - 22) Hamilton, H.J., Shan, N., and Ziarko, W.: Machine Learning of Credible Classifications, *Proc. Australian Conf. on Artificial Intelligence (AI-1997)*, pp.330–339 (1997).
 - 23) Hatazawa, H., Abe, H., Komori, M., Tachibana, Y. and Yamaguchi, T.: Knowledge Discovery Support from a Meningoencephalitis Dataset Using an Automatic Composition Tool for Inductive Applications, *Post-Proc. of the Joint JSAI 2001 Workshop on New Frontiers in Artificial Intelligence*, pp.500–507 (2001).
 - 24) Hilderman, R.J. and Hamilton, H.J.: *Knowledge Discovery and Measure of Interest*, Kluwer Academic Publishers (2001).
 - 25) Huang, Y., Shekhar, S. and Xiong, H.: Discovering Colocation Patterns from Spatial Data Sets: A General Approach, *IEEE Trans. Knowledge and Data Engineering*, Vol.16, No.12, pp.1472–1485 (2004).
 - 26) Jaccard, P.: Nouvelles Recherches sur la Distribution Florale, *Bulletin de la Societe Vaudoise des Sciences Naturelles*, Vol.44, pp.223–270 (1908) (in French).
 - 27) Jaroszewicz, S. and Simovici, D.A.: A General Measure of Rule Interestingness, *Proc. Euro. Conf. on Principles of Data Mining and Knowledge Discovery (PKDD-2001)*, pp.253–265 (2001).
 - 28) Kaufman, J.C. and Sternberg, R.J. (Eds.): *The International Handbook of Creativity*, Cambridge University Press, Cambridge (2006).
 - 29) Kawakita, J.: *KJ Method — A Scientific Approach to Problem Solving*, Kawakita Research Institution, Tokyo (1975, 2000).
 - 30) Kerr, B. and Gagliardi, C.: Measuring Creativity in Research and Practice, *Positive Psychological Assessment — A Handbook of Models and Measures*, Lopez, S.J. and Synder, C.R. (Eds.), American Psychological Association, Washington (2003).
 - 31) Klösgen, W.: Explora: A Multipattern and Multistrategy Discovery Assistant, *Advances in Knowledge Discovery and Data Mining*, Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (Eds.), California, pp.249–271, AAAI/MIT Press (1996).
 - 32) Lavrac, N., Flach, P. and Zupan, B.: Rule Evaluation Measures: A Unifying View, *Proc. Int'l Workshop on Inductive Logic Programming (ILP-1999)*, pp.174–185 (1999).
 - 33) Liu, H., Lu, H., Feng, L. and Hussain, F.: Efficient Search of Reliable Exceptions, *Proc. Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD-1999)*, pp.194–203 (1999).
 - 34) Morimoto, Y., Fukuda, T., Matsuzawa, H., Tokuyama, T. and Yoda, K.: Algorithms for Mining Association Rules for Binary Segmentations of Huge Categorical Databases, *Proc. Int'l Conf. on Very Large Databases (VLDB-1998)*, pp.380–391 (1998).
 - 35) Ohsaki, M., Kitaguchi, S., Okamoto, K., Yokoi, H. and Yamaguchi, T.: Evaluation of Rule Interestingness Measures with a Clinical Dataset on Hepatitis, *Proc. Euro. Conf. on Principles of Data Mining and Knowledge Discovery (PKDD-2004)*, pp.362–373 (2004).
 - 36) Ohsaki, M., Sato, Y., Yokoi, H. and Yamaguchi, T.: A Rule Discovery Support System for Sequential Medical Data — In the Case Study of a Chronic Hepatitis Dataset, *Proc. Int'l Workshop on Active Mining AM-2002 in IEEE Int'l Conf on Data Mining (ICDM-2002)*, pp.97–102 (2002).
 - 37) Piatetsky-Shapiro, G.: Discovery, Analysis and Presentation of Strong Rules, *Knowledge Discovery in Databases*, Piatetsky-Shapiro, G. and Frawley, W.J. (Eds.), pp.229–248, AAAI/MIT Press (1991).
 - 38) Rijsbergen, C.: *Information Retrieval*, Chapter 7, Butterworths, London (1979). <http://www.dcs.gla.ac.uk/Keith/Chapter.7/Ch.7.html>
 - 39) Smyth, P. and Goodman, R.M.: Rule Induction using Information Theory, *Knowledge Discovery in Databases*, Piatetsky-Shapiro, G. and Frawley, W.J. (Eds.), pp.159–176, AAAI/MIT Press (1991).
 - 40) Tan, P.N., Kumar, V. and Srivastava, J.: Selecting the Right Interestingness Measure for Association Patterns, *Proc. Int'l Conf. on Knowledge Discovery and Data Mining (KDD-2002)*, pp.32–41 (2002).
 - 41) Tsumoto, S.: Discovery Challenge, *Euro. Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD-2002)* (2002). <http://lisp.vse.cz/challenge/ecmlpkdd2002/index.html>
 - 42) Tsumoto, S.: Guide to the Meningoencephalitis Diagnosis Data Set, Int'l Workshop of KDD Challenge on Real-world Data KDD-2000 (2000). <http://www.slab.dnj.ynu.ac.jp/challenge2000/menin.htm>

- 43) Yao, J. and Liu, H.: Searching Multiple Databases for Interesting Complexes, *KDD: Techniques and Applications*, Liu, H., Motoda, H. and Liu, H. (Eds.), World Scientific (1997).
- 44) Yao, Y.Y. and Zhong, N.: An Analysis of Quantitative Measures Associated with Rules, *Proc. Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD-1999)*, pp.479–488 (1999).
- 45) Zhong, N., Yao, Y.Y. and Ohshima, M.: Peculiarity Oriented Multi-Database Mining, *IEEE Trans. on Knowledge and Data Engineering*, Vol.15, No.4, pp.952–960 (2003).

(平成 18 年 8 月 11 日受付)

(平成 19 年 1 月 9 日採録)



大崎 美穂 (正会員)

1994 年九州芸術工科大学芸術工学部音響設計学科卒業。1996 年同大学院芸術工学研究科博士前期課程修了。1999 年博士後期課程修了。同年静岡大学情報学部助手。2004 年同志社大学工学部専任講師。2007 年より准教授。博士(工学)。人間と知的システムの能力を相補し、人間の知識・認知・感性活動の支援を試みる。医療データマイニング、聴覚障害補償の研究に従事。1998 年 Int'l Conf. on Soft Comp. and Inf./Intel. Sys., Best Paper Award。2003 年情報処理学会山下記念研究賞。2006 年 FAN シンポジウムベストプレゼンテーション賞。IEEE, 人工知能学会, 知能情報ファジィ学会, 音響学会各会員。



阿部 秀尚

2000 年静岡大学情報学部情報科学科卒業。2002 年同大学院情報学研究科情報学専攻修士課程修了。2004 年同大学院理工学研究科博士後期課程修了。現在、島根大学医学部助手。博士(工学)。医療データマイニング, データマイニング結果の後処理支援, データマイニングプロセスにおけるアプリケーション開発支援環境に関する研究に従事。人工知能学会, 日本医療情報学会各会員。



津本 周作

1963 年生。1989 年大阪大学医学部卒業。同年千葉大学医学部附属病院医員(神経内科)。1990 年松戸市立病院救急部医員。1991 年千葉大学医学部附属病院医員(医療情報部)。1993 年東京医科歯科大学助手。1999 年島根医科大学助教授。2000 年島根医科大学医療情報学教授。2003 年島根大学医学部医療情報学教授。現在に至る。博士(工学)。1993 年医療情報学連合大会優秀論文賞受賞。1995 年医療情報学連合大会研究奨励賞受賞。IEEE, ACM, AAAI, AMIA 各会員。



横井 英人

1996 年香川医科大学卒業。同年千葉大学第一内科にて臨床研修。2000 年千葉大学医学部附属病院医療情報部医員。厚生労働省等勤務を経て 2005 年香川大学医学部附属病院医療情報部講師。医学博士。研究分野: 医療情報システムの標準化, 医療情報のデータマイニング。日本医療情報学会, 日本内科学会, 日本消化器内視鏡学会, 日本超音波医学会, 日本遠隔医療学会各会員。



山口 高平 (正会員)

1979 年大阪大学工学部通信工科学科卒業, 1984 年同大学院工学研究科博士後期課程修了。同年大阪大学産業科学研究所助手。1989 年静岡大学工学部助教授。1997 年同大学工学部教授。2004 年より慶應義塾大学理工学部教授。工学博士。定理証明の研究を経て, 知識システム, 知識発見, セマンティック Web, オントロジ, 知能ソフトウェア工学に関する研究に従事。1991, 1998 年度人工知能学会研究会奨励賞。1992 年度人工知能学会全国大会優秀論文賞。2002 年度人工知能学会研究会優秀賞。人工知能学会, 電子情報通信学会, 日本認知科学会, 知能情報ファジィ学会各会員。