

客観的指標群に基づく ルール評価支援手法の評価 ー慢性肝炎データマイニングを対象としてー

阿部 秀尚¹⁾ , 大崎 美穂²⁾ , 横井英人³⁾ , 山口 高平⁴⁾ , 津本 周作¹⁾

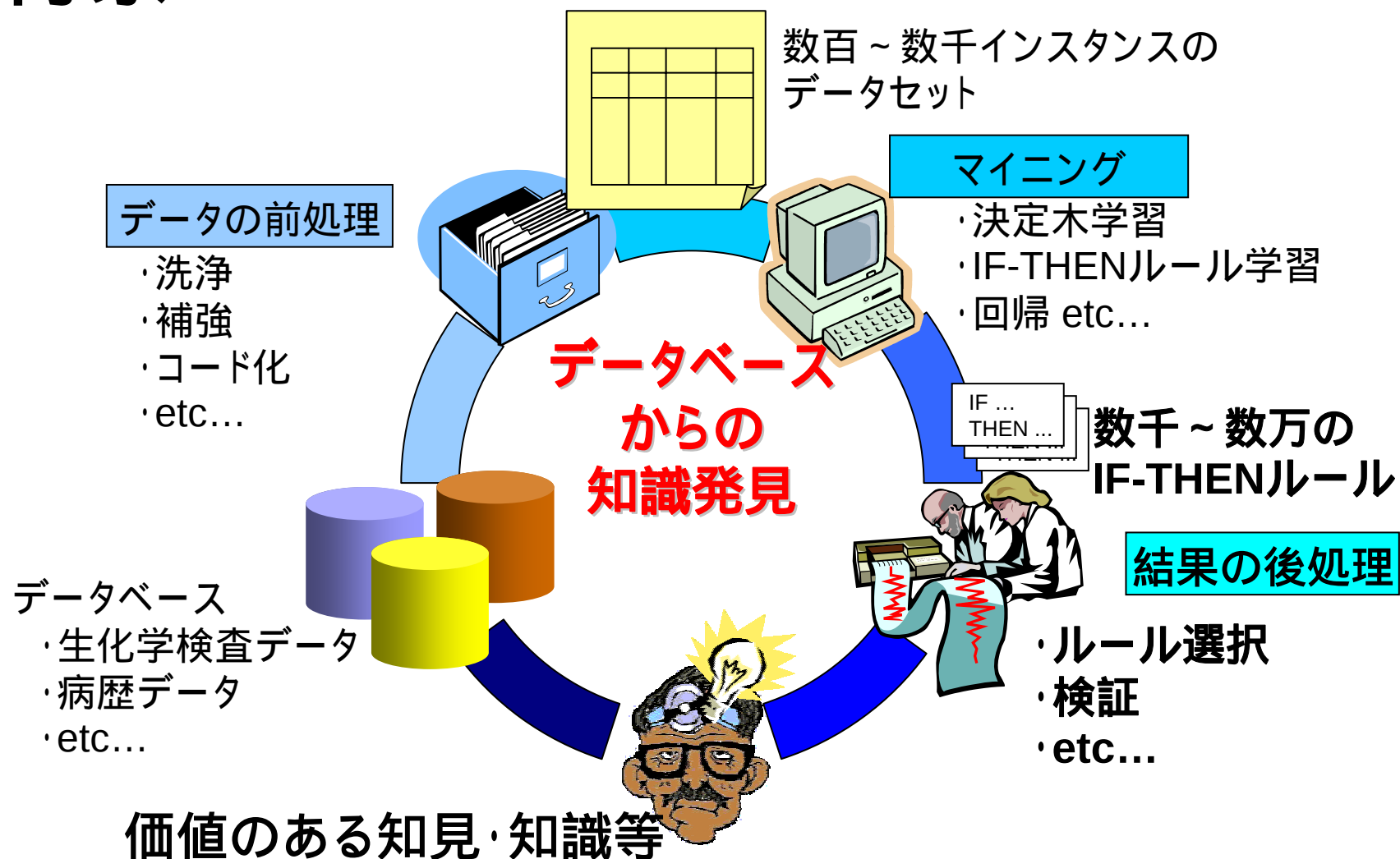
島根大学医学部医学科医療情報学講座¹⁾

同志社大学工学部情報システムデザイン学科²⁾

香川大学医学部附属病院医療情報部³⁾

慶應義塾大学理工学部管理工学科⁴⁾

背景



マイニングによって得られた膨大なルールから価値のあるルールを見出すことは困難



ルール評価支援と課題

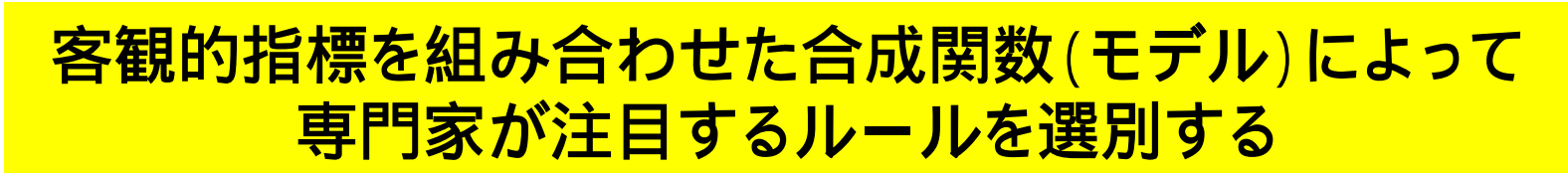
- ルールの正答率や再現率などの評価値を利用したルール選択
- 専門家の興味や関心を推定する評価値
 - ルールの評価値に基づく指標による興味の推定
 - ルールの類似度による興味の推定

- 
- ルールの評価値(客観的指標と呼ぶ)が専門家の評価基準をどのように反映するのか
 - 興味深さの指標が提案された領域以外で有効なのか



研究目的

- データマイニングの結果の後処理におけるルール評価支援手法の提案
 - 客観的指標群の組み合わせによるモデル
 - 専門家による評価を再利用
- 複数の学習アルゴリズムによるルール評価モデル構築実験
 - ルール評価モデルによる専門家の評価推定性能の評価
 - データマイニング過程の推移とルール評価モデルの変化の関係を考察





ルール評価モデル構築アルゴリズム の比較実験

- 慢性肝炎データマイニングに対する専門家による評価と各客観的指標値をデータ化
- ルール評価モデル構築のための学習アルゴリズムの性能評価
 - 訓練データセット全体での分類精度(正解率・各クラスの正答率と再現率)の評価
 - Leave-One-Outによる予測精度の評価
- データマイニング過程の推移とルール評価モデルの内容と性能変化の関係について考察

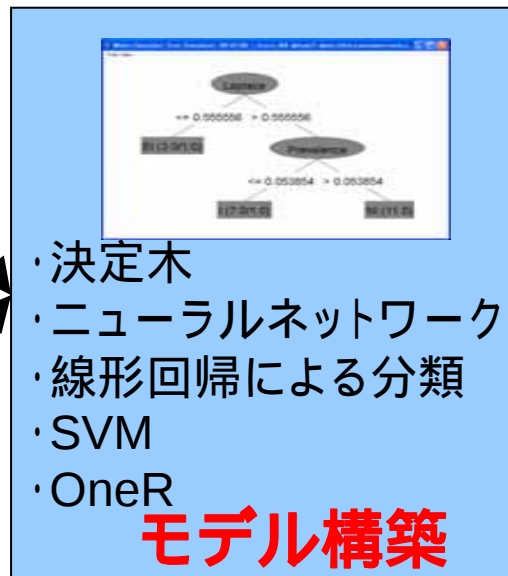
ルール評価モデル構築実験

慢性肝炎データ
 ・GPTパターン予測
 ・IFN効果判定予測
 (各2回)



39種類の客観的
 指標による評価
 属性:客観的指標
 属性値:客観的指標値

専門家による評価
 クラス: EI, I, NU, NI



Leave-one-out
 交差検定法で
 モデル構築
 アルゴリズムの
 信頼性を評価

ruleID	Accuracy	Added Value	...	YulesQ	YulesY	HumanExpert
Rule1	0.81	0.41	...	0.73	0.44	EI
Rule10	0.81	0.43	...	0.75	0.45	EI
Rule11	0.85	0.46	...	0.79	0.49	I
Rule12	0.84	0.56	...	0.87	0.58	I
Rule13	0.94	0.44	...	0.88	0.59	I
Rule14	0.81	0.43	...	0.75	0.45	NI
Rule15	0.84	0.55	...	0.86	0.57	NI
Rule16	0.91	0.58	...	0.91	0.65	NI
Rule17	0.81	0.61	...	0.89	0.61	NU

客観的指標

IF-THENルール個々に対して、所与のデータとの関係で算出される値

対象とする客観的指標(計39種)

- 確率に基づく指標(26種)
 - Coverage, Prevalence, Precision, Recall, Support, Accuracy, Specificity, Lift, Leverage, Added Value, Relative Risk, Jaccard, Certainty Factor, Odds ratio, Yule's Q, Yule's Y, Kappa, Koelesgen's Interestingness, Brin's Interestingness, Brin's Conviction, GOI(一般性重視), Credibility, KSI, Laplace Correction, Collective Strength
- 統計検定量に基づく指標(3種)
 - Chi-Square (True/Positive だけの理論値との差, 4分表全ての理論値との差), Gini Gain
- 情報量に基づく指標(5種)・・・事象の生起確率が離散値である場合を対象
 - Mutual Information, J-Measure, YLI1, YLI2, YZI, K-Measure
- 事例数に基づく指標(3種)・・・事例数のカウント
 - Φ coefficient, PSI, Cosine Similarity
- 属性値の差異を見る指標(2種)・・・ルール間の距離を仮定
 - GBI, Peculiarity

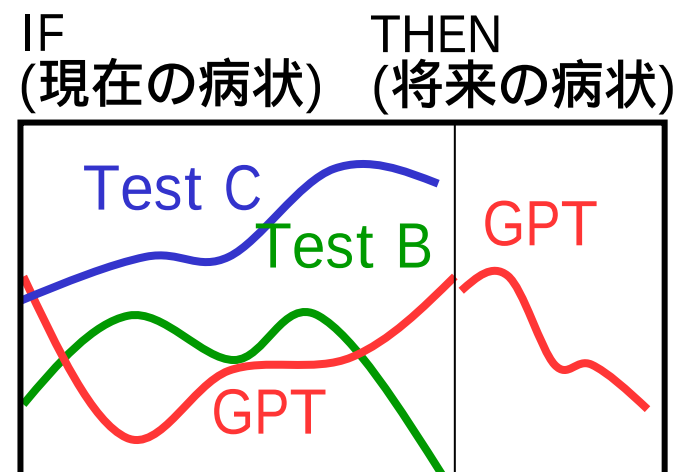


実験対象の学習アルゴリズム

- 決定木: J4.8(C4.5のJava実装版)
- ニューラルネットワーク: 3層完全結合
 - BP学習のパラメータ: 学習率 = 0.3, 慣性項 = 0.2
 - 出力層: 各クラスに対応
- 線形回帰による分類: ClassificationViaLinearRegression
 - 線形回帰式: 各クラスに対して“1 vs. その他”の式を得る
 - 属性選択: 貪欲探索法 (+ 線形結合を排除)
- SVM: Sequential Minimal Optimization [Platt98]
 - SVMによる3クラスの分類: 各クラスに対して“1 vs. その他”のモデルを得る
 - カーネル関数: polynomial kernel
- OneR
 - 各属性値によって並び替え後, クラスの変更点を閾値とした条件節を持つルール集合を作成し, 最も正解率の高いルール集合を選択

GPTパターンによる病状経過予測

- 入力: 慢性肝炎患者の血液・尿検査データ
(千葉大学医学部附属病院提供)
- 出力: 観察期間の検査項目のパターンを条件部, 予後のGPTのパターンを結論部とする
IF-THENルール

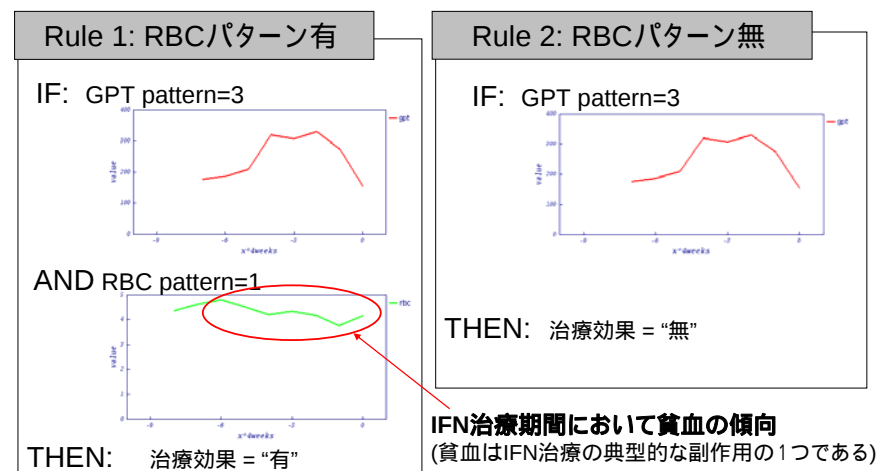


インターフェロン(IFN)効果判定予測

- 入力: 慢性肝炎患者の血液・尿検査データ, IFN治療効果判定データ(千葉大学医学部附属病院提供)
- 出力: IFN投与終了までの任意期間の検査項目パターンを条件部,

専門家による効果判定
を結論部とするIF-THEN
ルール

異なる治療効果判定となる代表的なルール



各ルール集合のサイズ・クラス分布

		得られた ルール数	評価ラベルの分布			
			特に興味あり (EI)	興味あり(I)	妥当 (NI)	理解 不能 (NU)
GPT パターン 予測	1回目	30	3	8	16	3
	2回目	21	2	6	12	1
IFN 効果判定 予測	1回目	26	4	7	11	4
	2回目	32	15	5	11	1

学習アルゴリズムによるルール評価モデルの性能比較(訓練データ集合)

	Acc	Precision				Recall			
		EI	I	NI	NU	EI	I	NI	NU
GPT 1回目									
J4.8	96.7	100.0	88.9	100.0	100.0	66.7	100.0	100.0	100.0
BPNN	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SVM	56.7	0.0	100.0	68.2	14.3	0.0	12.5	93.8	33.3
CLR	63.3	0.0	66.7	62.5	0.0	0.0	50.0	93.8	0.0
OneR	60.0	0.0	66.7	59.3	0.0	0.0	25.0	100.0	0.0
GPT 2回目									
J4.8	90.5	66.7	85.7	100.0	0.0	100.0	100.0	91.7	0.0
BPNN	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SVM	95.2	100.0	100.0	92.3	100.0	50.0	100.0	100.0	100.0
CLR	85.7	50.0	100.0	85.7	0.0	50.0	83.3	100.0	0.0
OneR	85.7	0.0	75.0	92.3	0.0	0.0	100.0	100.0	0.0
INF 1回目									
J4.8	88.5	80.0	100.0	83.3	100.0	100.0	71.4	90.9	100.0
BPNN	100.0								
SVM	46.2								
CLR	53.8								
OneR	50.0								
INF 2回目									
J4.8	90.6	88.2	100.0	90.9	0.0	100.0	80.0	90.9	0.0
BPNN	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SVM	56.3	72.7	0.0	45.0	100.0	53.3	0.0	81.8	100.0
CLR	65.6	63.2	100.0	60.0	0.0	80.0	60.0	54.5	0.0
OneR	68.8	62.5	0.0	87.5	0.0	100.0	0.0	63.6	0.0

学習アルゴリズムによっては少数クラスの分類が困難な
全体の正解率・そのほかのクラスの分類精度は単純な
精度を上回る

学習アルゴリズムによっては少数クラスの分類が困難だが、全体の正解率・そのほかのクラスの分類精度は単純な分類精度を上回る

学習アルゴリズムによるルール評価モデルの性能比較 (Leave-One-Out検定)

		Acc	Precision				Recall			
			EI	I	NI	NU	EI	I	NI	NU
GPT 1回目										
	J4.8	50.0	0.0	60.0	60.0	0.0	0.0	75.0	56.3	0.0
	BPNN	30.0	0.0	12.5	50.0	0.0	0.0	12.5	50.0	0.0
	SVM	46.7	0.0	0.0	65.0	11.1	0.0	0.0	81.3	33.3
	CLR	40.0	0.0	14.3	50.0	0.0	0.0	12.5	68.8	0.0
	OneR	43.3	0.0	25.0	55.6	0.0	0.0	37.5	62.5	0.0
GPT 2回目										
	J4.8	76.2	0.0	66.7	90.9	0.0	0.0	100.0	83.3	0.0
	BPNN	66.7	0.0	33.3	81.8	0.0	0.0	83.3	75.0	0.0
	SVM	81.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0
	CLR	76.2	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0
	OneR	81.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0
INF 1回目										
	J4.8	19.2	37.5	0.0	20.0	0.0	75.0	0.0	18.2	0.0
	BPNN	26.9	40.0	22.2	25.0	25.0	50.0	28.6	18.2	25.0
	SVM	34.6	21.4	0.0	54.5	0.0	75.0	0.0	54.5	0.0
	CLR	19.2	33.3	0.0	28.6	0.0	25.0	0.0	36.4	0.0
	OneR	19.2	0.0	11.1	23.5	0.0	0.0	14.3	36.4	0.0
INF 2回目										
	J4.8	75.0	76.5	66.7	75.0	0.0	86.7	40.0	81.8	0.0
	BPNN	37.5	50.0	28.6	22.2	0.0	53.3	40.0	18.2	0.0
	SVM	31.3	36.4	0.0	28.6	0.0	26.7	0.0	54.5	0.0
	CLR	34.4	41.2	20.0	30.0	0.0	46.7	20.0	27.3	0.0
	OneR	68.8	60.0	0.0	100.0	0.0	100.0	0.0	63.6	0.0

GPTパターン予測，IFN効果判定予測とも，
1回目よりも2回目で，ほぼ全ての学習アルゴリズム
において予測精度が向上

GPTパターン予測, IFN効果判定予測とも,
1回目よりも2回目で, ほぼ全ての学習アルゴリズム
において予測精度が向上



データマイニング過程の推移と ルール評価モデルの変化

- データマイニング過程の推移
 - 仮説生成: 専門家が仮説に気がつく
(GPT 1回目)
 - 仮説検証: 専門家の仮説を検証する
(GPT 2回目, IFN 1回目)
 - 仮説洗練化: 専門家の仮説の因果関係を明らかにする
(IFN 2回目)

- ルール評価モデルの変化との対応についての考察
 - ルール評価モデル予測精度の変化
 - ルール評価モデルの内容の変化

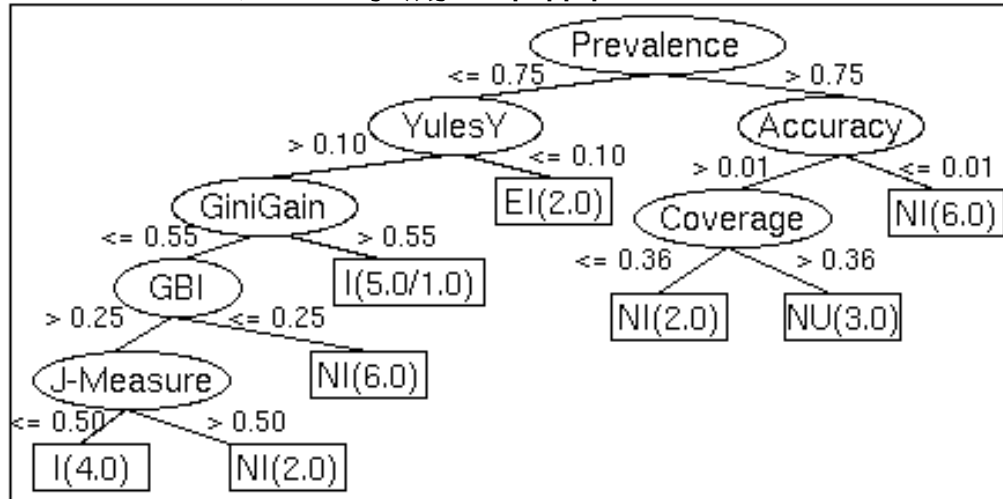
データマイニング過程の推移とルール評価モデルの予測精度の変化

		Acc	Precision				Recall			
			EI	I	NI	NU	EI	I	NI	NU
GPT 1回目										
J4.8		50.0	0.0	60.0	60.0	0.0	0.0	75.0	56.3	0.0
BPNN		30.0	0.0	12.5	50.0	0.0	0.0	12.5	50.0	0.0
SVM		46.7	0.0	0.0	65.0	11.1	0.0	0.0	81.3	33.3
CLR		40.0	0.0	14.3	50.0	0.0	0.0	12.5	68.8	0.0
OneR		43.3	0.0	25.0	55.6	0.0	0.0	37.5	62.5	0.0
GPT 2回目										
J4.8		76.2	0.0	66.7	90.9	0.0	0.0	100.0	83.3	0.0
BPNN		66.7	0.0	83.3	81.8	0.0	0.0	83.3	75.0	0.0
SVM		81.0	0.0							100.0
CLR		76.2	0.0							0.0
OneR		81.0	0.0							0.0
GPTパターン予測, IFN効果判定予測とも, 1回目よりも2回目で, 「特に興味あり」 「興味あり」と「妥当」「理解不能」の間での予測精度が向上										
INF 1回目										
J4.8		19.2	37.5							0.0
BPNN		26.9	40.0							25.0
SVM		34.6	21.4							0.0
CLR		19.2	33.3	0.0	28.6	0.0	25.0	0.0	36.4	0.0
OneR		19.2	0.0	11.1	23.5	0.0	0.0	14.3	36.4	0.0
INF 2回目										
J4.8		75.0	76.5	66.7	75.0	0.0	86.7	40.0	81.8	0.0
BPNN		37.5	50.0	28.6	22.2	0.0	53.3	40.0	18.2	0.0
SVM		31.3	36.4	0.0	28.6	0.0	26.7	0.0	54.5	0.0
CLR		34.4	41.2	20.0	30.0	0.0	46.7	20.0	27.3	0.0
OneR		68.8	60.0	0.0	100.0	0.0	100.0	0.0	63.6	0.0

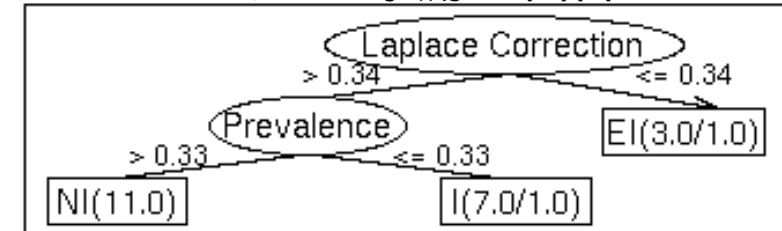
GPTパターン予測, IFN効果判定予測とも,
1回目よりも2回目で, 「特に興味あり」
「興味あり」と「妥当」「理解不能」の間での
予測精度が向上

ルール評価モデル(決定木)の変化

(a) GPTパターン予測 1回目

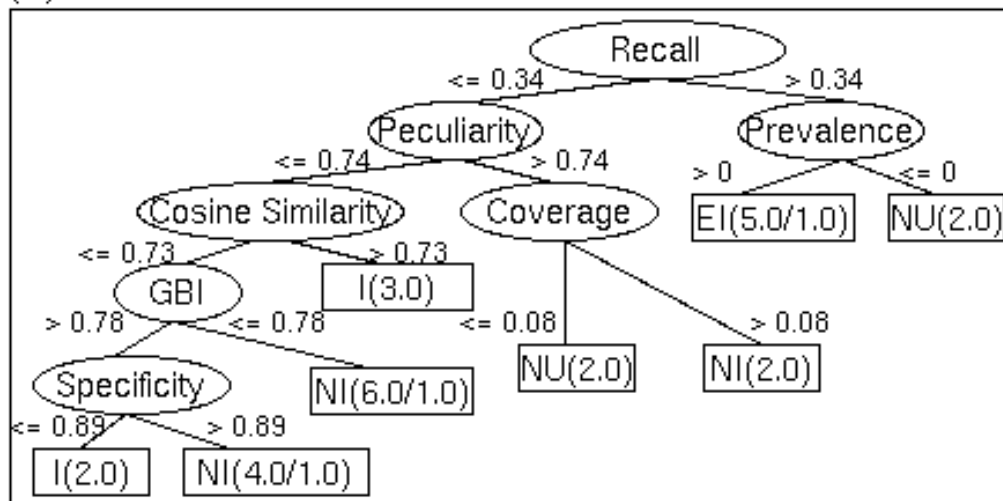


(b) GPTパターン予測 2回目

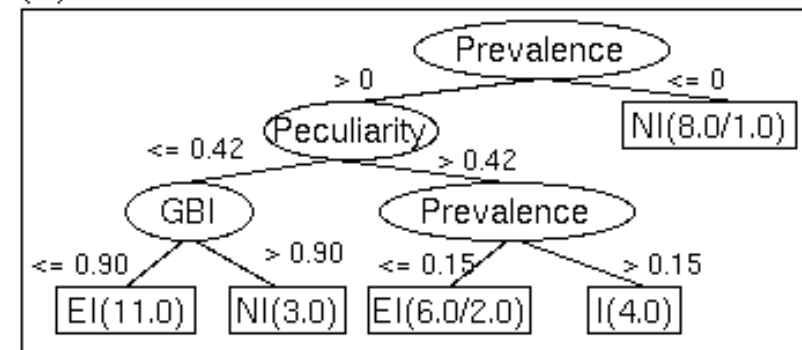


- ・1回目より2回目の決定木が簡略化する傾向
- ・データマイニング過程の推移がルール評価モデルに反映

(c) IFN効果判定予測 1回目



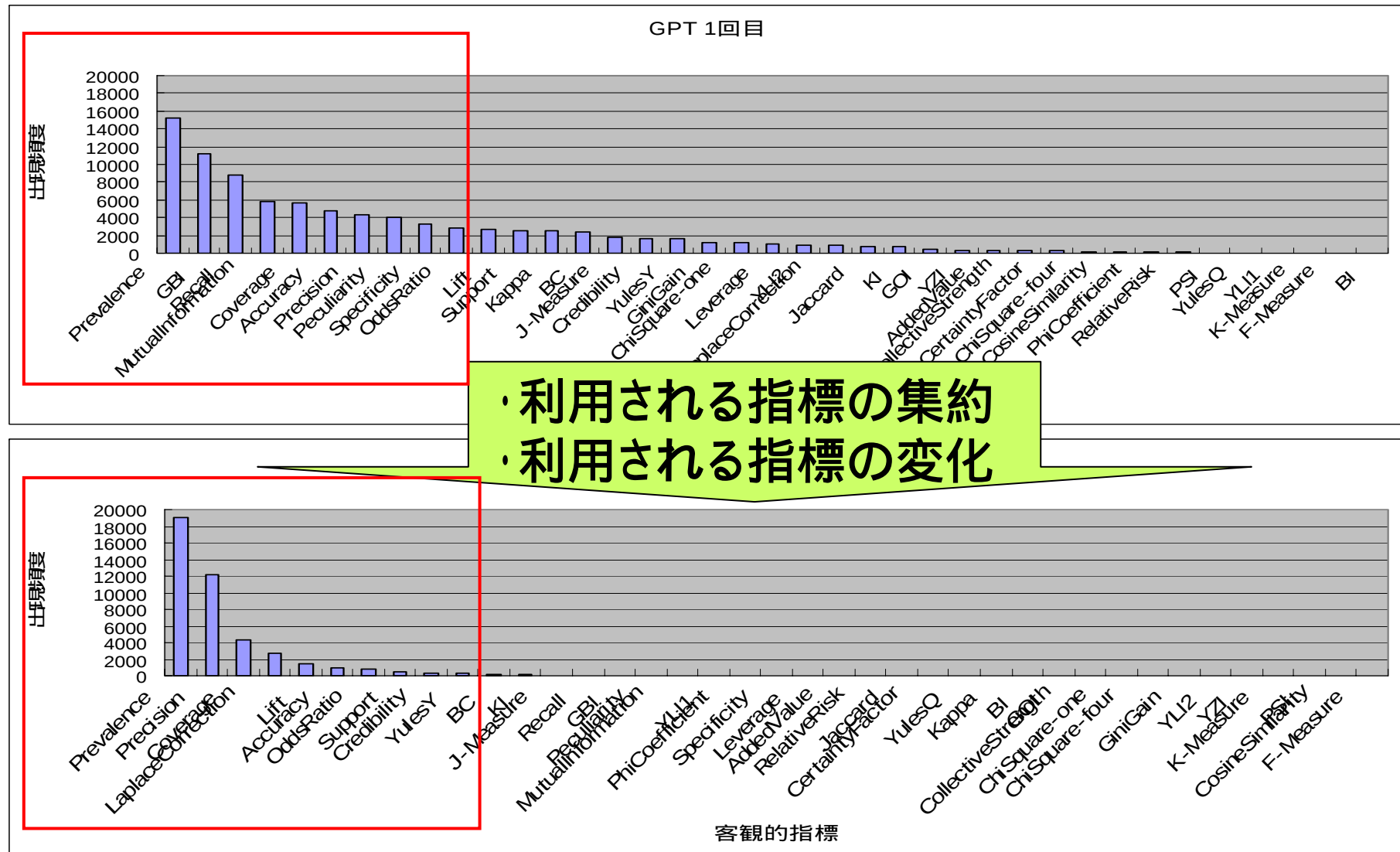
(d) IFN効果判定予測 2回目



図中の各枝に示された値は正規化された指標値

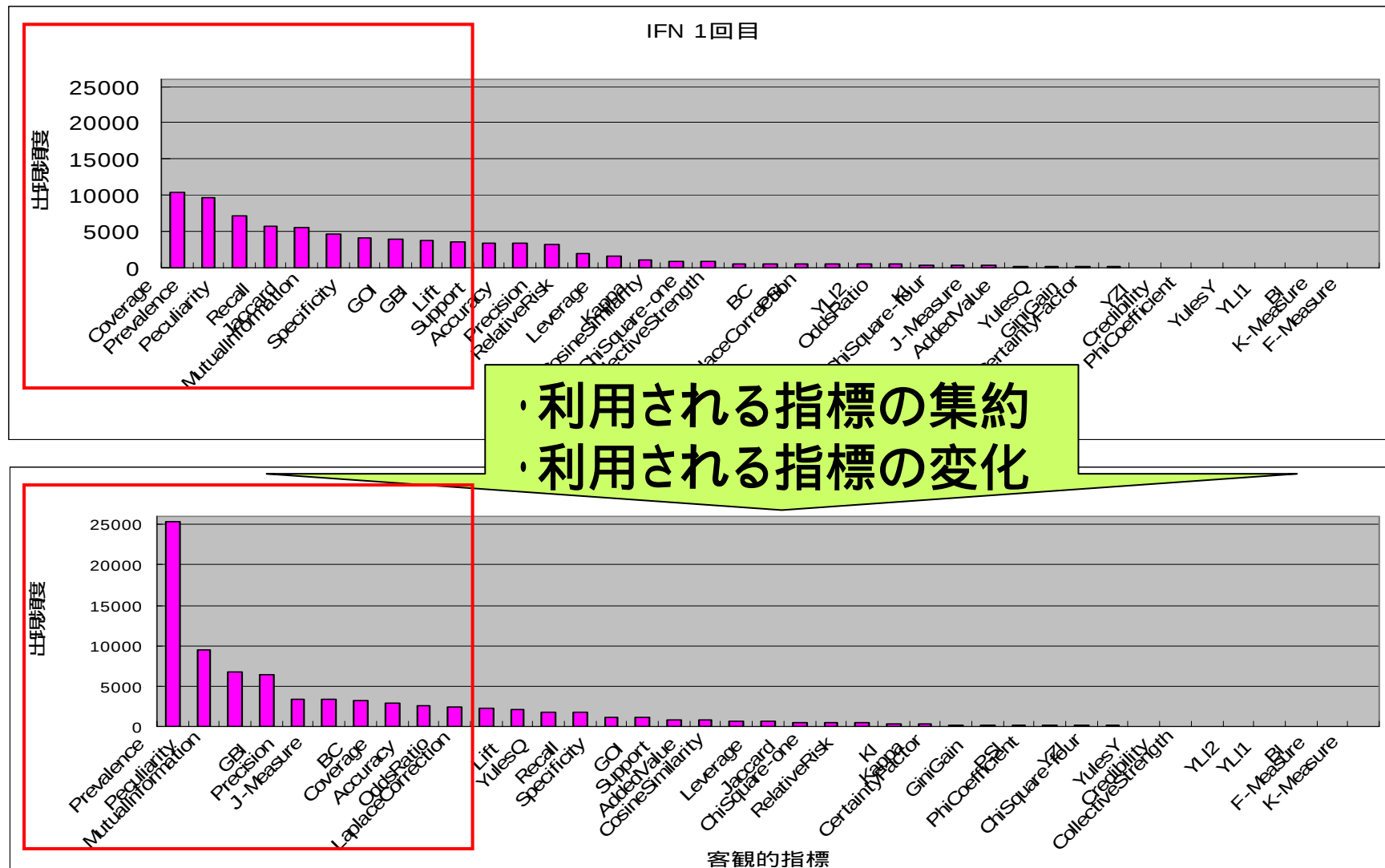
決定木に利用される指標の変化 (GPTパターン予測)

10,000個のブートストラップサンプルに対して決定木を構築



決定木に利用される指標の変化 (IFN効果判定予測)

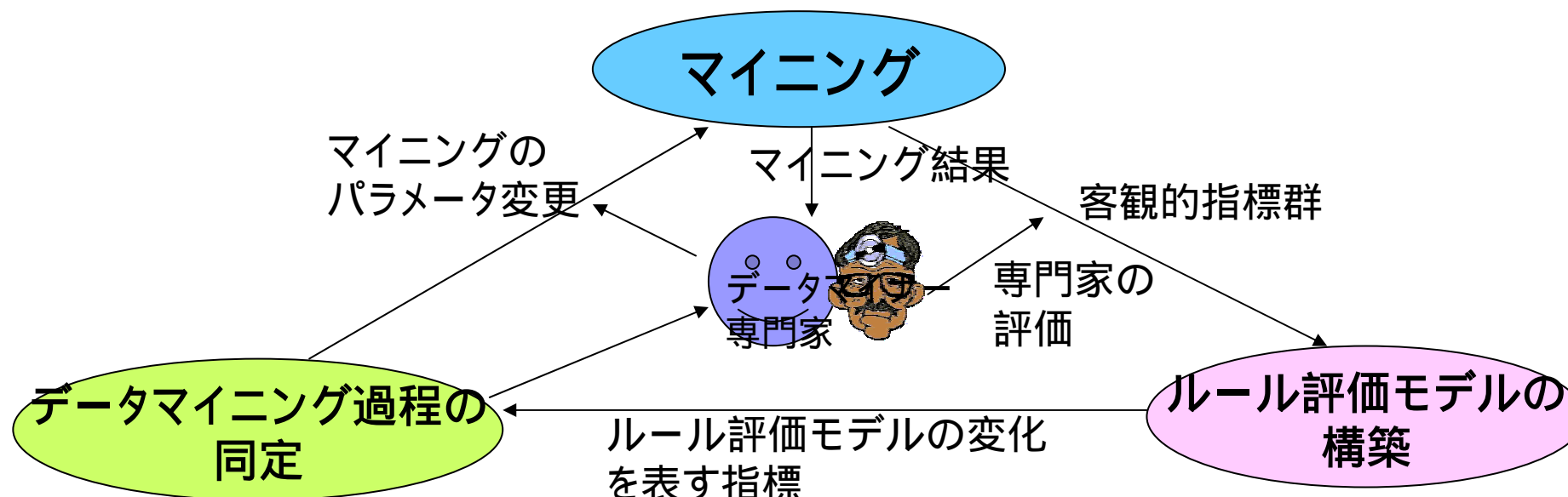
10,000個のブートストラップサンプルに対して決定木を構築



データマイニング過程の推移とルール評価モデルの変化についての考察

■ データマイニング過程の推移を定義

- ルール評価モデルが各段階に応じて変化
- 専門家・データマイナーがどのような状態でデータマイニングを行っているのか把握可能





まとめ・今後の課題

■ まとめ

- ルール評価モデル構築によるルール評価支援手法の提案
- ルール評価モデルによるユーザの評価の推定性能の比較・評価 → 有効なルール評価モデル構築が可能
- データマイニング過程の推移とルール評価モデルの変化の関係を考察

■ 今後の課題

- ルール評価モデル構築のための属性(客観的指標)集合の検討
- ルール評価モデル構築のための学習アルゴリズム選定
- データマイニング過程の状態を定量的に同定