

客観的指標群に基づく ルール評価モデルの構築 ー髄膜脳炎データベースを対象としてー

阿部 秀尚¹⁾, 岡本 和也²⁾, 大崎 美穂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⁾

静岡大学情報学部情報科学科²⁾

同志社大学工学部情報システムデザイン科³⁾

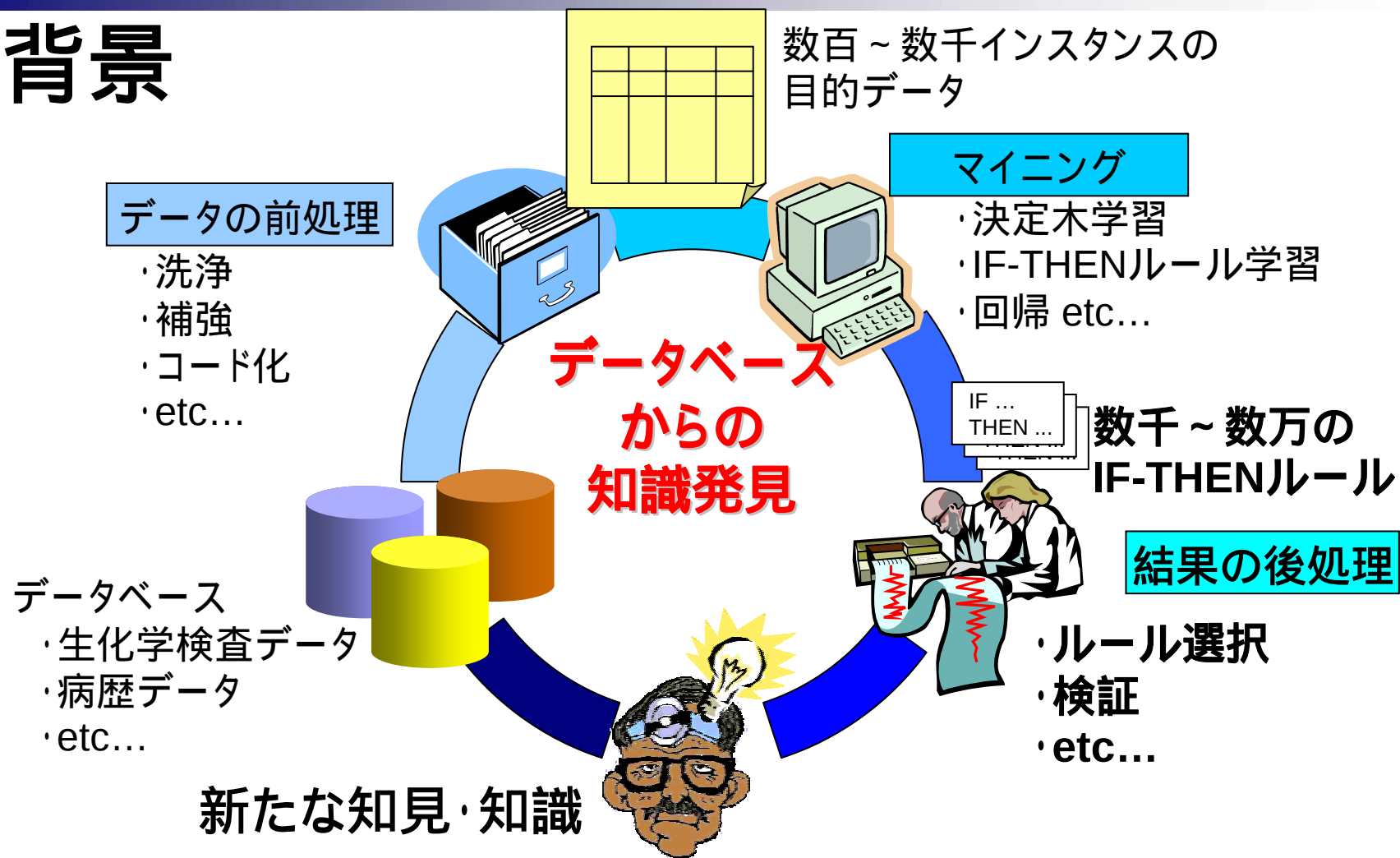
慶應義塾大学理工学部管理工学科⁴⁾



発表内容

- 背景
- 客観的指標によるルール選択支援
 - 関連研究
 - 客観的指標の定義
 - 単一指標によるルール選択支援
 - ルール評価モデル構築によるルール評価支援法
- ルール評価モデル構築実験
- まとめ・今後の課題

背景



マイニングによって得られた膨大なルールを人間が全てを評価することは困難

専門医(ユーザ)による評価を客観的指標を用いて支援

客観的指標とは

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

本稿における客観的指標(計39)

- 確率に基づく指標
 - 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, KI, Brin's Interest, Brin's Conviction, GOI, Credibility, Laplace Correction, Collective Strength
- 統計検定ベースに基づく指標・・・ある条件(仮説)を設定して統計量を算出
 - カイ2乗統計量, Gini Gain, GKI
- 情報量に基づく指標・・・ルールが存在することによって得られる情報量に注目して算出
 - Mutual Information, GKI, J-Measure, YLI1, YLI2, YZI, K-Measure
- 事例数に基づく指標・・・ルールに被覆される事例数のカウントに基づいて算出
 - Φ coefficient, PSI, Cosine Similarity
- 属性値の差を見る指標・・・ルール間やデータに対する距離を仮定し値を算出
 - Peculiarity, GBI

先行研究(慢性肝炎データセットにおけるケーススタディ)

■ 慢性肝炎データセットから得られたルール

- 仮説生成フェーズ(フェーズ1)のルールセット
- 仮説検証フェーズ(フェーズ2)のルールセット

単一の客観的指標
による並び替え
(計39種類)

ruleID	Accuracy	予測
Rule13	0.94	I
Rule11	0.85	I
Rule12	0.84	NI
Rule1	0.81	NI
Rule10	0.81	NI
...	...	...

専門医による評価

ruleID	HumanExpert
Rule1	NI
Rule10	NI
Rule11	I
Rule12	I
Rule13	I
...	...

一致度を比較

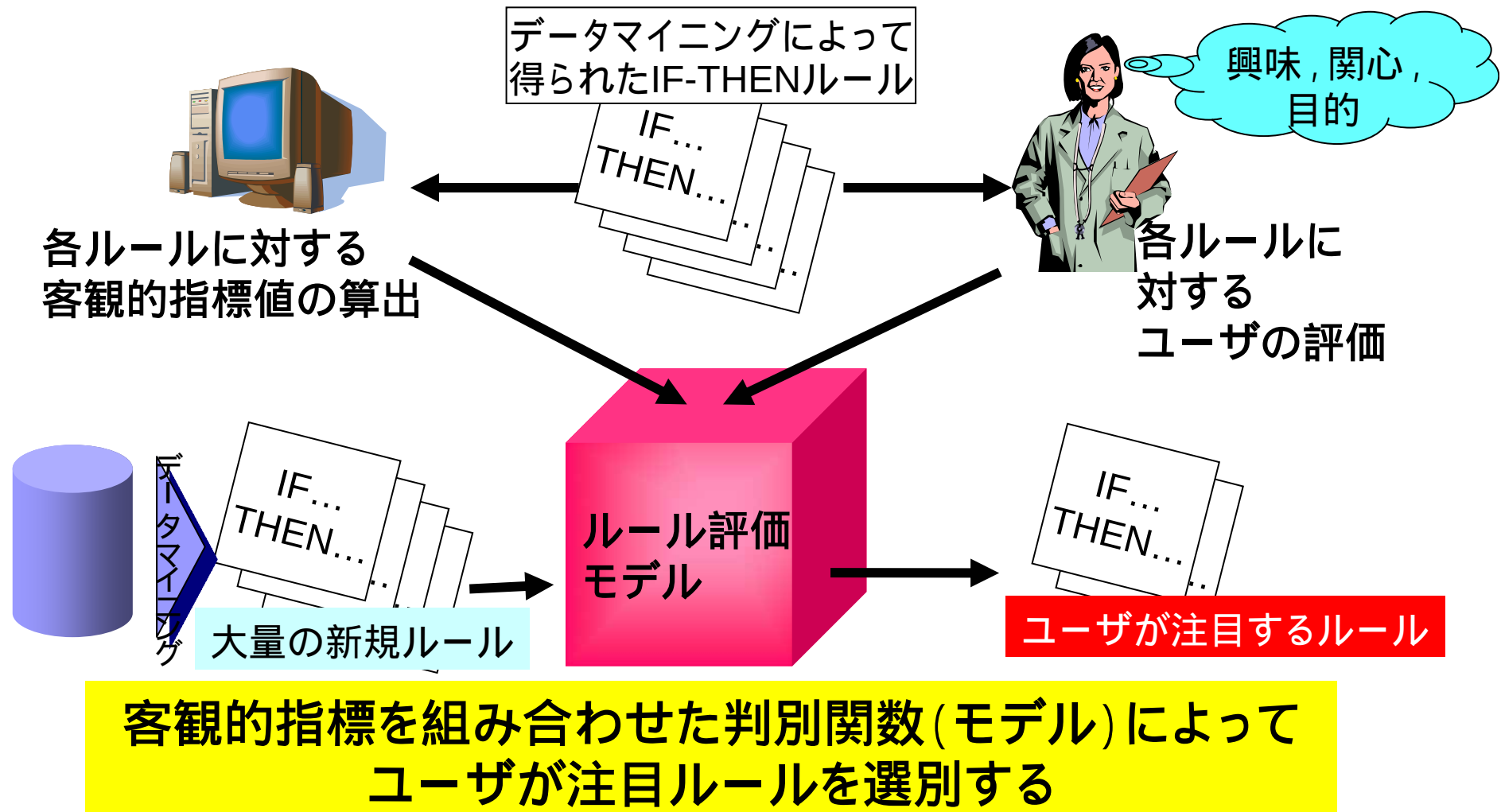
課題

- 専門医によるルール評価と完全に一致する単一の客観的指標は無い
- 一致度が高い客観的指標の選択, 閾値の設定を一般化することが困難

研究目的

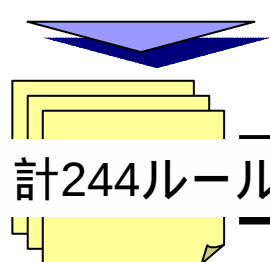
- 客観的指標を組み合わせ, 専門医の評価ラベルと予測が一致するモデルを構築

ルール評価モデル構築による ルール評価支援



ルール評価モデル構築実験

140事例の
髄膜脳炎に関する
6(2×3)種類の
診断鑑別データ



計244ルール

39種類の客観的
指標による評価

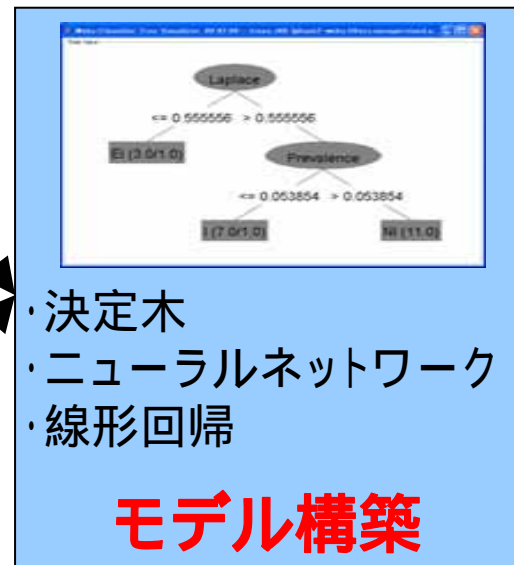
属性:客観的指標

属性値:客観的指標値

専門家による評価

クラス:専門家の評価

クラス値: 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

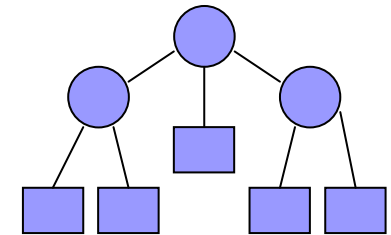
ruleID	Accuracy	Added Value	...	YulesQ	YulesY	HumanExpert
Rule1	0.81	0.41	...	0.73	0.44	NI
Rule10	0.81	0.43	...	0.75	0.45	NI
Rule11	0.85	0.46	...	0.79	0.49	I
Rule12	0.84	0.56	...	0.87	0.58	I
Rule13	0.94	0.44	...	0.88	0.59	I
Rule14	0.81	0.43	...	0.75	0.45	NI
Rule15	0.84	0.55	...	0.86	0.57	NI
Rule16	0.91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

使用したマイニングアルゴリズムの概要 (データマイニングツールWekaを利用)

■ 決定木 (C4.5)

- 情報利得比を用いて分割条件を決定

$$\text{entropy}(p_1, p_2, \dots, p_n) = -p_1 \log p_1 - p_2 \log p_2 \dots - p_n \log p_n$$



■ ニューラルネットワーク (バックプロパゲーション学習)

- 三層, 各層の間は完全結合

■ 線形回帰式による分類

$$x = \omega_0 + \omega_1 a_1 + \omega_2 a_2 + \dots + \omega_k a_k$$

x : クラス, $a_1, a_2, \dots, a_k$: 属性値, $\omega_1, \omega_2, \dots, \omega_k$: 重み

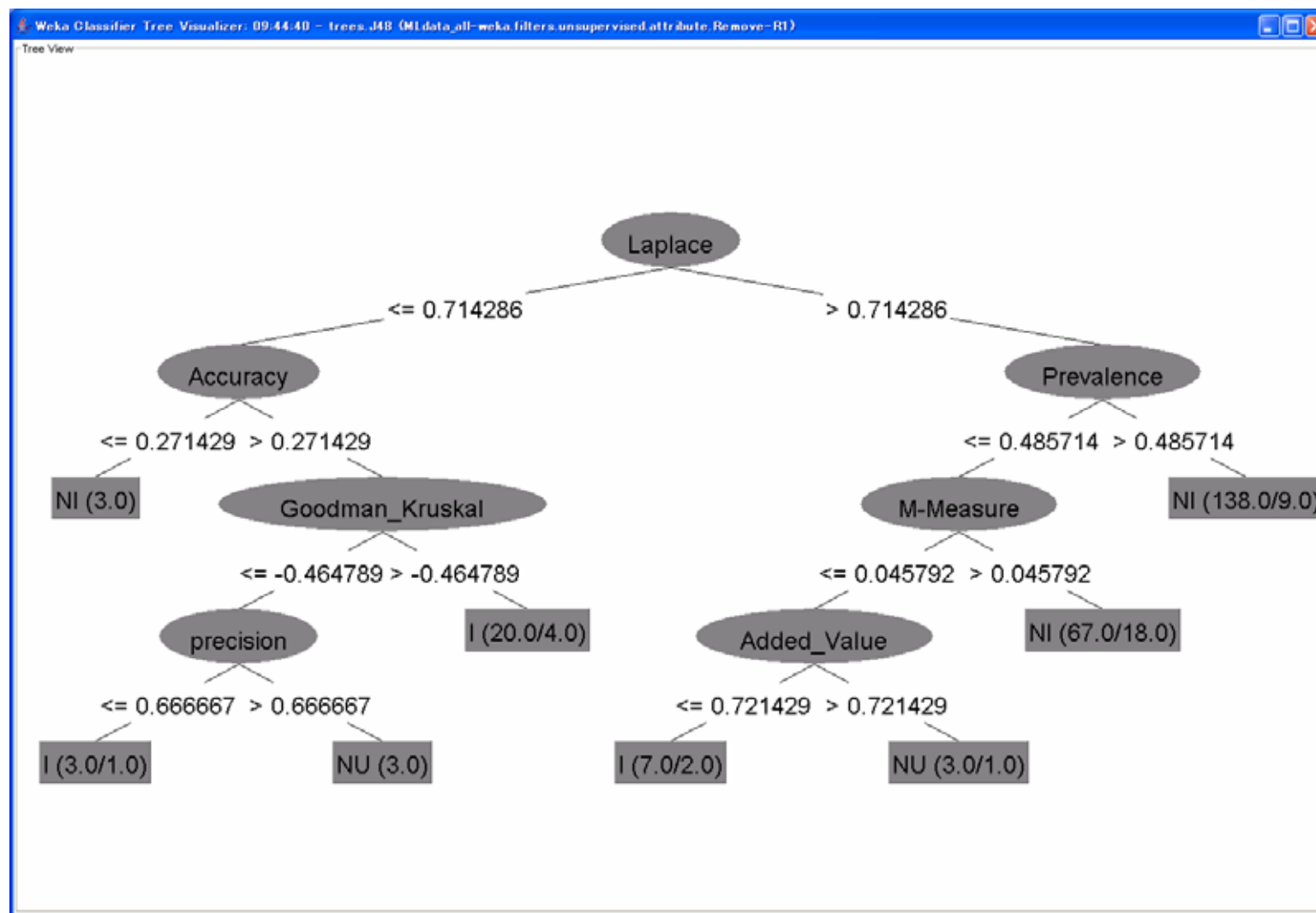
ルール評価モデルと単一の客観的指標による一致度の比較(ルール数244)

評価方法 アルゴリズム	訓練データでの評価		Leave-one-out
	# 1	# 2	正解率
決定木	23/48* (47.92%)	209/244* (85.66%)	191/244* (78.29%)
ニューラルネットワーク	43/48* (89.58%)	220/244* (90.16%)	190/244* (77.87%)
線形回帰	19/48* (39.58%)	202/244* (83.20%)	192/244* (78.69%)
最良の単一指標	23/48* (47.92%)	195/244* (79.92%)	
Lift			
最低の単一指標	1/48 (2.08%)	150/244 (61.48%)	
GOI-D			

#1:興味深いルール的一致カウント数 / 専門家が興味深いと判断したルール数

#2:全ルール的一致カウント数 / 全ルール数

決定木として得たルール評価モデル





まとめ・今後の課題

■ まとめ

- データマイニングの結果の後処理におけるルール評価モデル構築手法の提案
- 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
 - 全体の一致度で2.8%以上向上

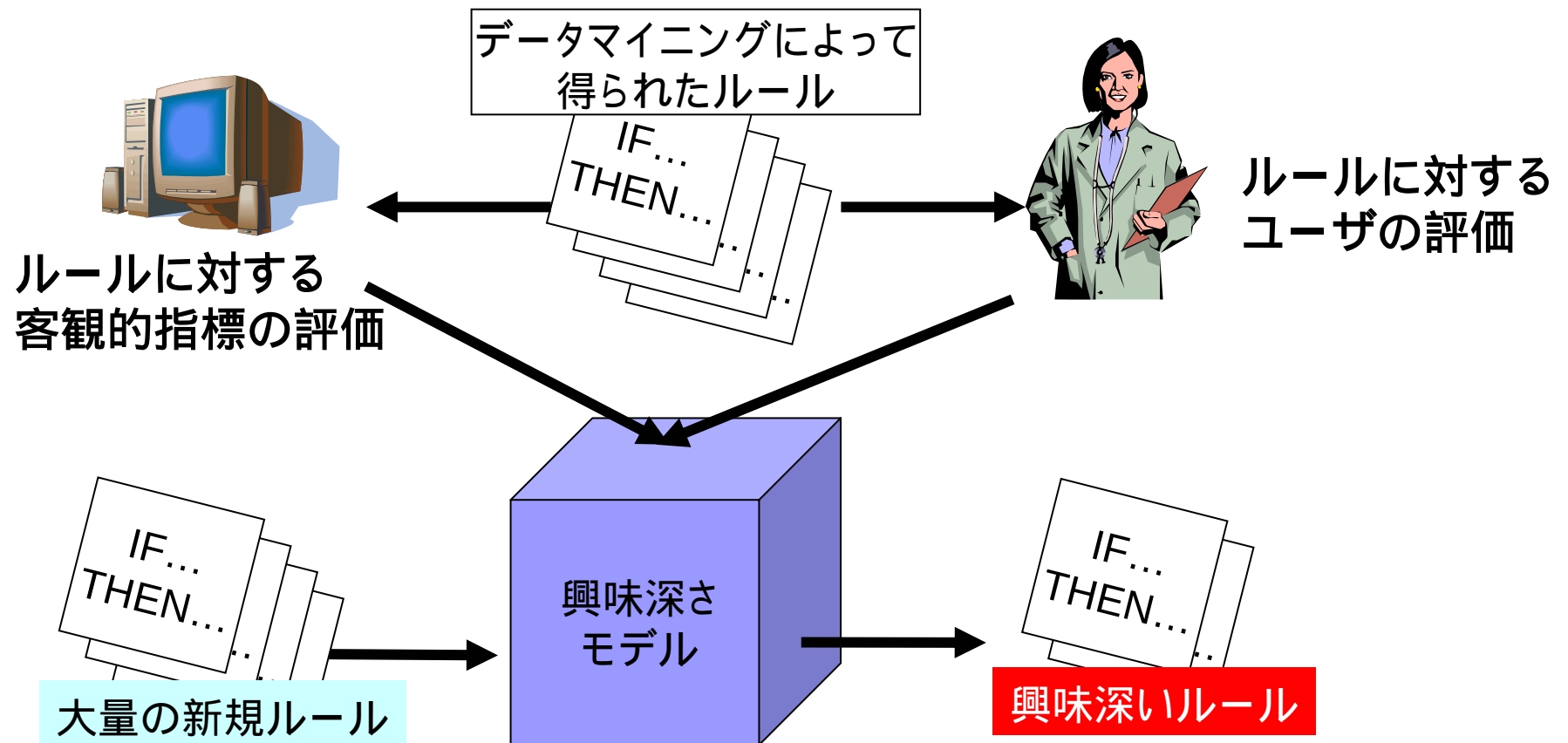
■ 今後の課題

- ルール評価モデル構築のための学習アルゴリズム選定
- ルール評価モデル構築に基づく結果の後処理支援システムの開発・評価



興味深さモデルとは

- ユーザによる評価を、客観的指標を組み合わせて記述



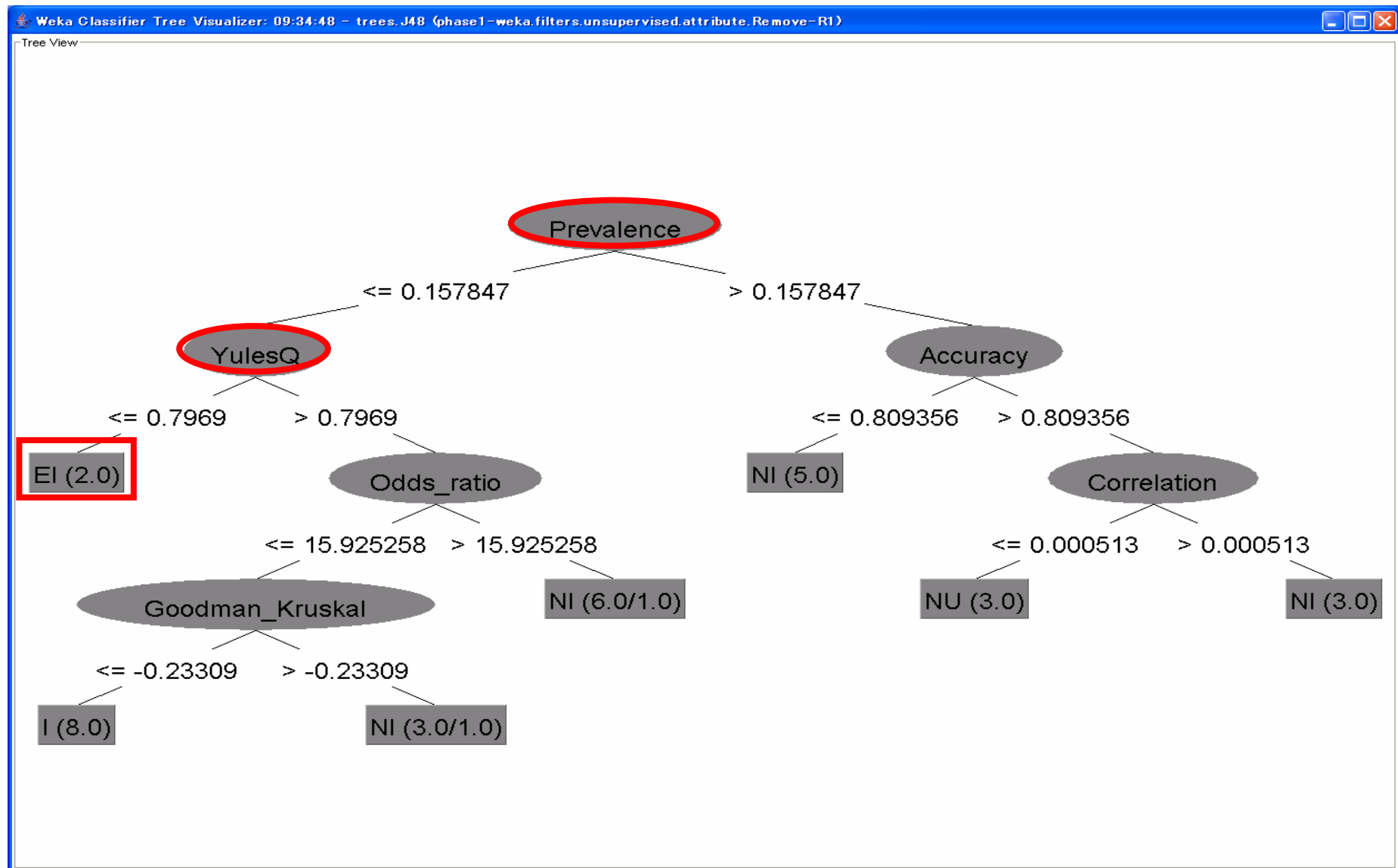
新規ルールからユーザが興味深いと思うだろうルールをフィルタリング可能

肝炎フェーズ1の結果(ルール数30)

評価方法 アルゴリズム	訓練データでの評価			Leave-one-out
	# 1	# 2	# 3	正解率(#3の平均)
決定木(J48)	10/12* (83.33%)	2/3* (66.67%)	28/30* (93.33%)	15/30 (50%)
ニューラルネットワーク	12/12* (100%)	3/3* (100%)	30/30* (100%)	11/30 (36.67%)
線形回帰	12/12* (100%)	3/3* (100%)	30/30* (100%)	7/30 (23.33%)
最良の単一指標	8/12* (66.67%)	2/3* (66.67%)	22/30* (73.33%)	
Recall				
最低の単一指標	0/12 (0%)	0/3 (0%)	18/30* (60%)	
KSI				

#1:興味深いルール的一致カウント数/専門家が興味深いと判断したルール数、
 #2:重要ルール的一致カウント数/重要ルール数、#3:全ルール的一致カウント数/全ルール数

肝炎フェーズ1 モデル (決定木)

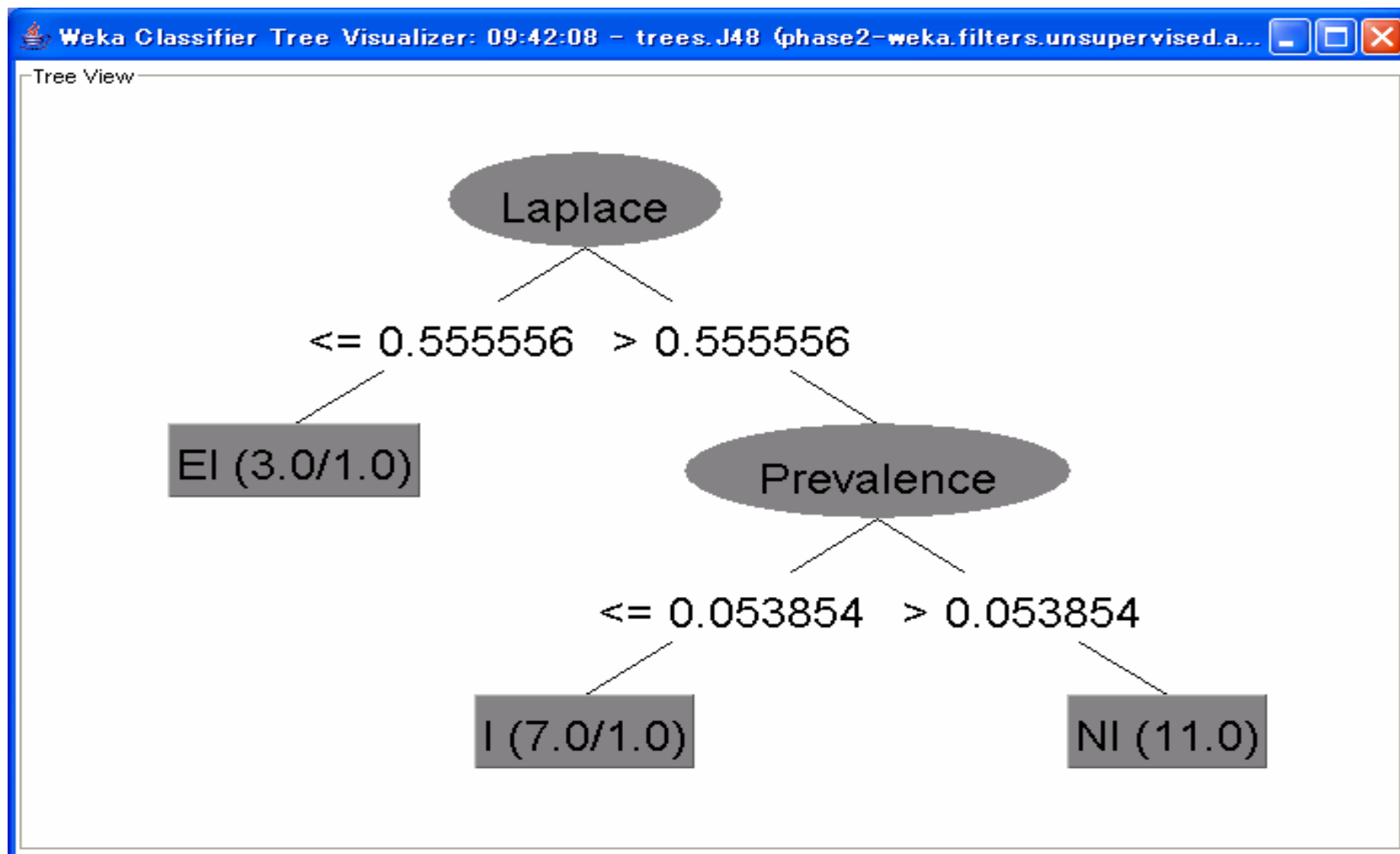


肝炎フェーズ2の結果(ルール数21)

評価方法 アルゴリズム	訓練データでの評価			Leave-one-out
	# 1	# 2	# 3	正解率
決定木(J48)	8/8* (100%)	2/2* (100%)	19/21* (90.48%)	16/21* (76.19%)
ニューラルネット ワーク	8/8* (100%)	2/2* (100%)	21/21* (100%)	14/21* (66.67%)
線形回帰	8/8* (100%)	2/2* (100%)	21/21* (100%)	16/21* (76.19%)
最良の単一指標	6/8* (75%)	1/2 (50%)	17/21* (80.95%)	
Credibility				
最低の単一指標	0/8 (0%)	0/2 (0%)	4/21 (19.05%)	
Prevalence				

#1:興味深いルール的一致カウント数/専門家が興味深いと判断したルール数、
#2:重要ルール的一致カウント数/重要ルール数、#3:全ルール的一致カウント数/全ルール数

肝炎フェーズ2 モデル (決定木)





先行研究で調査・提案された客観的指標

実験に使用した客観的指標(計40種)

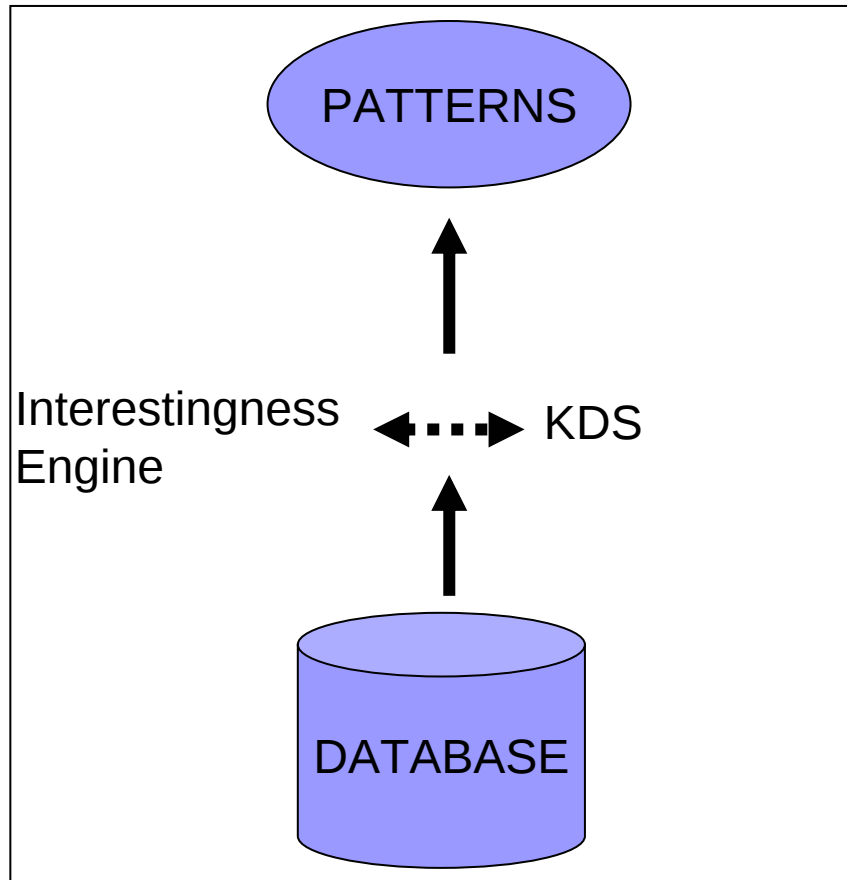
- 確率に基づく指標(25種)
 - Coverage, Prevalence, Precision, Recall, Support, Accuracy, Specificity, Lift, Leverage, Added Value, Relative Risk, Jaccard, Certainty Factor, Odds ratio, Yule's Q, Yule's Y, Kappa, KI, BI, Conviction, GOI, Credibility, KSI, Laplace Correction, Collective Strength
- 統計検定ベースに基づく指標(3種)・・・ある条件(仮説)を設定して使用
 - Measure, Gini Index, GKI
- 情報量に基づく指標(7種)・・・事象の生起確率が離散値である場合を対象
 - NMI, J-Measure, GKI, YLI1, YLI2, YLI3, K-Measure
- 事例数に基づく指標(3種)・・・事例数のカウントをベース
 - ϕ coefficient, PSI, CS
- 属性値の差を見る指標(2種)・・・ルール間の距離を仮定し使用
 - GBI, Peculiarity

客観的指標の分類に関して

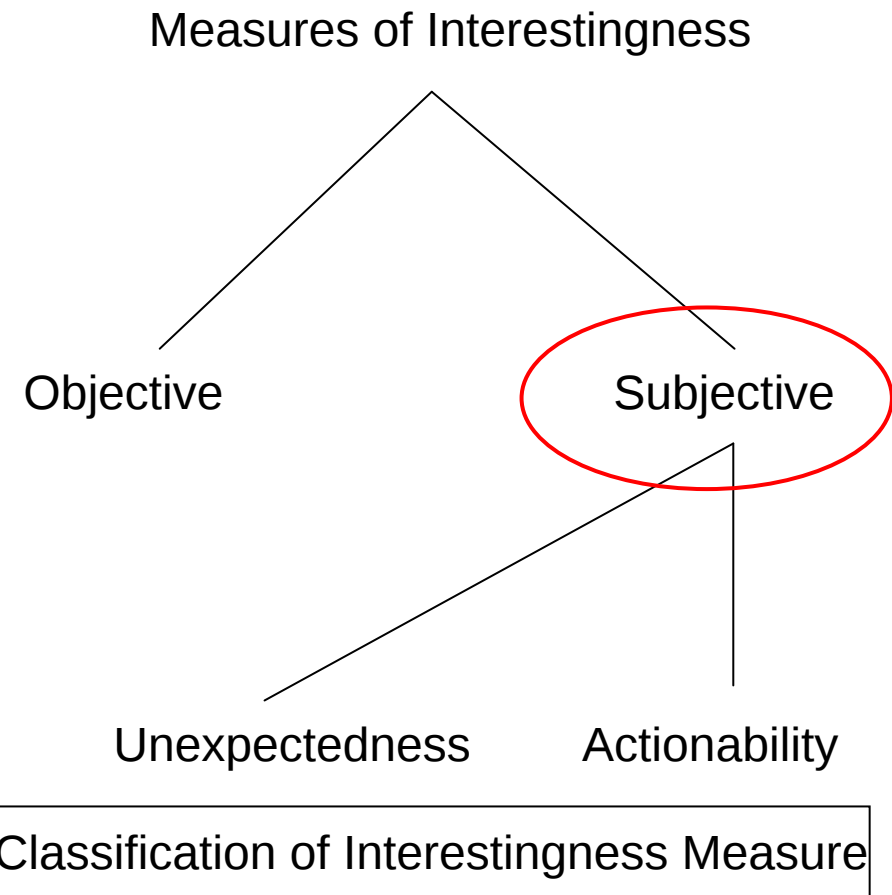
- 評価の高い指標は一般的な指標がほとんどで、特定の領域に特化した客観的指標の性能は低い
 - 特定の領域に特化した客観的指標は使用用途を限定して提案されているため

一般的な指標	特定の領域に特化した指標
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, K-Measure, Laplace Correction, Collective Strength, χ^2 Measure, Gini Index, Φ coefficient	KI, BI, Conviction, GOI, Credibility, GKI, NMI, J-Measure, GKI, YLI1, YLI2, YLI3, KSI, PSI, CS, GBI, Peculiarity, GBI

興味深さ指標



A Framework of Knowledge Discovery



(Avi Silberschatz and Alexander Tuzhilin, "On Subjective Measures of Interestingness in Knowledge Discovery")

2004/11/27



実験結果の考察

- 決定木, ニューラルネットワーク, 線形回帰が単一の客観的指標よりもユーザの評価と一致度が高い
- 決定木のノードで出現した客観的指標は、一般的な指標が多い
 - 一般的な指標を使って生成されたモデルは、他の新規データに適用できる可能性がある



ルール例

Rule 1:

STIFF > 1.0
FOCAL = -
EEG_FOCUS = +
Cell_Poly <= 220.0
CSF_PRO <= 75.0
-> class VIRUS

妥当

Rule 2:

AGE > 35.0
SEX = F
HEADACHE <= 4.0
EEG_FOCUS = -
Cell_Poly <= 220.0
-> class VIRUS

興味深い