

平成 22 年 4 月 19 日現在

研究種目：若手研究(B)
 研究期間：2008 ～ 2009
 課題番号：20700139
 研究課題名(和文) 専門家の評価基準に適応するルール評価支援システムの開発
 研究課題名(英文) Development of a Rule Evaluation Support System with Adaptive Rule Evaluation Model Learning
 研究代表者
 阿部 秀尚 (ABE HIDENAO)
 島根大学・医学部・助教
 研究者番号：00397853

研究成果の概要(和文)：

本研究では、データマイニングにおける専門家によるルール評価作業について、繰り返し行われる評価作業において専門家の評価基準が時間経過に伴って変化する状況を想定した。このとき、ルール評価作業を支援するより正確な学習モデルを定常的に得るため、初期の分類学習モデルが有用となるまでの状況とルール評価モデルの評価からの評価基準の観測可能性について、実データおよび共通データセットを用いて実験を行った。この結果、1. 少数の訓練データで正解率が高い学習モデルを得るには、均一なクラス配分のデータセットが求められることから想定するルール評価作業の初期段階のコストは低く見積もることができること、2. ルールの正確さなどを測る評価指標について実用上の類似グループによる並び替え提示でも効果的な評価作業支援が可能であること、3. 状況の変化を計量的に観測するには、学習モデルの正解率だけでなくルール評価指標をさらに適用することによって捕らえることが可能であること、の3点が主に明らかとなった。

研究成果の概要(英文)：

In this study, I assumed the iterative situation about if-then rule evaluations by human experts in post-processing phase of data mining. From the results of my empirical studies, we can get the following lessons-of-learn: (1) the initial cost to build rule evaluation models can be reduced, because the method only needs small balanced subsets of labeled rules sampled by using the accuracy of trial rule evaluation models. (2) At the initial rule evaluation iteration, a human expert can use some functional groups of objective rule evaluation indices to measure rules. And, the system can also build rule evaluation models by using these groups of indices. (3) The transitions of criteria of human experts can be detected by measuring rule evaluation models consisting of if-then rules.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,200,000	360,000	1,560,000
2009年度	900,000	270,000	1,170,000
総計	2,100,000	630,000	2,730,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：データマイニング, 機械学習, ルール評価作業支援, ルール評価指標, メタ学習

1. 研究開始当初の背景

近年、情報技術の発展に伴い、医療分野においても病院情報システムなどが普及し、大量のデータが蓄積される状況にある。例えば、一日1000人以上の外来をもつ大学付属病院では、テキストデータだけでも一年に160GB以上のデータが蓄積される。これらの医療データを解析し、診療に役立てることは、病院における意思決定支援のために重要な課題となっている。一方、このように大量に蓄積されたデータを分析し、活用する手法として1990年代後半からデータマイニング手法が注目されている。データマイニングの実行にあたって、マイニング手法から出力されるif-thenルールや決定木などの学習モデルを解釈し、実際に行動を起こすための知識とするためには、問題領域の専門家による評価が必要である。このため、データマイニングプロジェクトの成功は、専門家が高度に評価する有益な知識を得ることが要件となる。しかしながら、可読性の高いif-thenルールを出力するマイニング手法として用いても、 n 個の属性値をもつデータセットから生成可能なルールは最悪の場合 $2^{(n+1)}$ であるため、大規模でしかもノイズを含むデータに対しては、数万個のルールが出力されることが容易に起こりうる。このようなルールを全て評価し、有益な知識を選別することは非常に困難を伴う作業である。

以上のことから、問題領域の専門家によるデータマイニングの出力の評価作業を効率よく支援するシステムが必要とされている。

2. 研究の目的

本研究では、問題領域の専門家によるデータマイニングの出力の評価作業を効率よく支援するシステムが必要とされている現状に対して、まず、専門家の評価基準の差異をルール評価モデルの客観的な評価指標の差異として定量化した、ルール評価モデル構築アルゴリズムの選定基準を明らかにする。次に、この選定基準に適したルール評価モデル構築アルゴリズムによって、問題領域の専門家の関心や興味深さといった評価基準に対し適応的なルール評価予測を提示するシステムを開発する。本研究では、専門家の評価基準に適応するルール評価支援システムを実現するため、以下の点を明らかにする。

- 専門家の評価基準の変化と学習モデル評価指標の関心の定量化方法
- 定量化された差異に基づく学習戦略および学習アルゴリズムの選定の有用性

学習モデルの評価指標は、教師あり学習、半教師あり学習、教師なし学習の各学習戦略で共通のもの、異なるものが複数存在する。例えば、正答率は各学習戦略で共通して用いられる。これらの評価指標群を入力として、学習戦略および学習アルゴリズムの選定基準を出力とする問題は、またメタ学習の問題として捉えることができる。

本研究は、これらの課題について、実データと共通データを用いて実証的に明らかにする。

3. 研究の方法

本研究では、専門家の評価基準の変化と学習モデル評価指標の関心の定量化を行うため、既存の評価済みルール集合を用いて、有用な分類学習モデルを得るための最小被評価ルール数の見積もりとルール評価モデルの変化を捉える指標の検討を行った。初年度は、共通データセットから得たルールに評価ラベルにあたるクラスを付し、39種類のルール評価指標とこのクラスから成るデータセットと、すでに専門家から評価を得た髄膜炎と慢性肝炎に関する分類ルールに同様にルール評価指標を算出して作成したデータセットを対象とした。これら両データセット群に対して、ランダムに均一・不均一の評価ラベルの分布であるクラス分布を設定して、分類学習アルゴリズムを適用した。それぞれのクラス分布について、分類学習モデルの交差検証による正解率をもって、分類学習モデルがより有用に機能するために必要な訓練データのクラス分布を求めた。同時に、クラス分布に合わせて、訓練データをより少数になるように設定し、正解率を求めた。初年度においては、また、初期の分類学習による有用なルール評価モデルを得るための被評価ルール数は、数十程度であることから、ルール評価指標による提示の改善が必要と判断した。このため、迅速に最初の分類学習によるルール評価モデルを得るため、分類ルール評価指標の値の類似性について、相関係数とそれに基づく主成分分析を行った。また、主成分分析によって得た合成指標を基にルール評価モデルを作成した場合の精度についても検討を行った。次年度は、主成分分析で求めた合成ルール

評価指標を用いたルール評価モデルを求めることが可能であるか、について個々のルール評価指標による分類学習モデルとの比較を交差検証の正解率を基に行った。また、ルール評価モデルを用いて専門家の評価基準の変化を計量化するため、ルール評価指標をルール評価モデルに再び適用することを想定した手法を提案し、実装した。本手法は、共通データセットに対してルール集合を求め、各ルール集合のルール評価指標の代表値に基づく分類をクラスタリングによって求めた。この分類結果を従来のデータセット特徴量による分類と比較した。さらに、ルール評価指標に基づく分類が個々のルールのルール評価指標の値によって分類予測可能であるかどうかを検討した。

4. 研究成果

本研究では、ルール評価作業の初期段階で分類学習モデルが適用可能となる被評価ルール数の見積もりと評価基準の変化を捉えるための計量化指標として分類ルール評価指標が適用できることを知見として得ることができた。

本研究において想定するルール評価作業支援の枠組みでは、はじめに専門家によるルール評価作業が必要であった。このため、少数の事例からどのように有用な学習モデルを得るかが重要である。そこで、実際に専門家から評価を得たルール集合を用い、評価ラベルにあたるクラスの分布を均一なものと同様に不均一なものに設定して学習モデルの正解率の評価を行った。分類学習アルゴリズムの評価に用いられる共通データセットでも同様にクラス分布とデータ数を設定して比較したところ、少数の訓練データ集合であっても、クラス分布が均一であれば、有用な分類学習によるルール評価モデルが得られることが示された。以上の成果から、ルール評価モデルによる専門家の評価作業支援を行うため、教師付学習と分類学習モデルに基づく半教師付学習を適用するための初期条件を明らかにすることができた。

次に、専門家による評価作業を支援するため、39種類の客観的ルール評価指標について、主成分分析によって評価の方向性の似通った7つのグループを得た。これにより、専門家がルール評価指標個々の定義を把握することなく、似たような評価の方向性となる新たな指標を利用可能となることが明らかとなった。さらに、これらのグループ化された指標がルール評価モデルの構成要素としても元のルール評価指標によるものと同等の正解率が得られることを示した。

また、ルール評価作業の進展に伴って、専門家の評価基準が変移することを想定し、データ集合内での属性とクラスの変化を

計量的に観測する指標について考察した。この結果、これまでルール評価モデルの構成要素としていた分類学習モデルのルール評価指標の変化を計測する手法を提案し、共通データセットにおいて本手法が従来のデータセット特徴量に関する指標よりも容易にデータ集合の類似性を判別可能であることを示した。これら類似性に基づく分類は、各データセットに対応するルール集合の代表値から求めた。この分類が、個々のルールのルール評価指標値から分類予測を行うことを想定し、決定木学習を適用したところ交差検証でも99.3%の正解率を得ることができた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 2 件)

- ① Hidenao Abe, Shusaku Tsumoto, Miho Ohsaki, Takahira Yamaguchi: Improving a rule evaluation support method based on objective indices. Intl. J. of Advanced Intelligence Paradigms (IJAIP), 2(2/3), 180-197 (2010) 査読有
- ② Hidenao Abe and Shusaku Tsumoto: Investigating Accuracies of Classifications for Randomized Imbalanced Class Distributions. Fundam. Inform. 90(4): 369-378 (2009) 査読有

[学会発表] (計 6 件)

- ① 阿部秀尚, 津本周作: 分類ルール評価指標を用いたデータセット類似度分析. 人工知能学会 第12回 データマイニングと数理統計研究会: 26-29 (2010.3.29) 立川市 (東京都)
- ② Hidenao Abe, Shusaku Tsumoto: Detecting Similarity of Transferring Datasets Based on Features of Classification Rules. ICDM Workshops 2009 (Transfer Mining 2009): 412-415 (2009.12.6) マイアミ (アメリカ)
- ③ Hidenao Abe, Shusaku Tsumoto: Detecting Temporal Trends of Technical Phrases by Using Importance Indices and Linear Regression. The 18th International Symposium on Methodologies for Intelligent Systems (ISMIS 2009): 251-260 (2009.9.15) プラハ (チェコ)
- ④ Hidenao Abe, Shusaku Tsumoto: A Comparison of Composed Objective Rule Evaluation Indices Using PCA and Single Indices. The 4th International

Conference on Knowledge Technology (RSKT 2009): 160-167 (2009.7.15) ゴールドコースト (オーストラリア)

⑤ Hidenao Abe, Shusaku Tsumoto: Comparing Accuracies of Rule Evaluation Models to Determine Human Criteria on Evaluated Rule Sets. ICDM Workshops 2008 (The Second International Workshop on Reliability Issues in Knowledge Discovery (RIKD 2008)): 1-7 (2008.12.15) ピサ (イタリア)

⑥ Hidenao Abe and Shusaku Tsumoto: Analyzing Correlation Coefficients of Objective Rule Evaluation Indices on Classification Rules, 2008 IEEE International Conference on Granular Computing (GrC2008): 467-474 (2008.8.27) 杭州 (中国)

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

名称:

発明者:

権利者:

種類:

番号:

出願年月日:

国内外の別:

○取得状況 (計 0 件)

名称:

発明者:

権利者:

種類:

番号:

取得年月日:

国内外の別:

[その他]

ホームページ等

なし

6. 研究組織

(1) 研究代表者

阿部 秀尚 (ABE HIDENAO)

島根大学・医学部・助教

研究者番号: 00397853