

## データマイニングにおけるルールの興味深さ指標の評価: 髄膜炎の検査データを用いた実験

### Evaluation of Rule Interestingness Measures in Data Mining: Experiment Using a Clinical Dataset on Meningoencephalitis

三木 大司\* 大崎 美穂\* 北口 真也† 岡本 和也† 山口 高平‡  
Daishi Miki Miho Ohsaki Shinya Kitaguchi Kazuya Okamoto Takahira Yamaguchi

#### 1. まえがき

興味深さ指標に関する研究はデータマイニングの一分野として盛んである。近年、データマイニング時の基準、あるいはルールフィルタリング時の基準として、非常に多くの興味深さ指標が提案されている。しかし、実際には興味深さ指標の研究として数理的性質のみを議論することが多く、その指標が専門家にとって興味深いかはあまり議論されていない。このため、ユーザはどの指標をどういう用途で使えば良いかが分かりにくい。

以上の背景より、本研究では現実のデータセットを使い、興味深さ指標の実用性を検討する。今回は髄膜炎の検査データを対象にして、過去の研究で生成されたルールを使う。評価実験では、興味深さ指標による評価結果を行い、専門家によるルールの評価結果と比較することで、興味深さ指標が持つ真の興味の推定性能を調べる。本論文の第2.節では関連研究として、過去の我々の研究と従来の興味深さ指標を解説する。第3.節では評価実験の結果を示し、興味深さ指標の実用性を考察する。第4.節で本論文をまとめ、今後の課題を示す。

#### 2. 関連研究

##### 2.1 髄膜炎データを用いたケーススタディ

我々は過去の研究において、38種類の検査項目と3つの診断項目、140患者、140レコードの髄膜炎の診断データを用い、ルールと専門家による評価結果を得た。様々な検査項目から診断結果を推定するルールを得ようと考え、マイニングには帰納アプリケーション自動構築ツールCAMLETを用いた。属性には全検査項目を、クラスには各診断項目の診断結果を用いた。診断結果は二値と多値の2種類を持つため(例えば原因診断では、ウィルス性/細菌性(二値)、ウィルス性ならばウィルスの種類(多値))、各々の場合でルールを得た。ルールの評価では、生成されたルールをプロダクションルールの形式で専門家に提示した。専門家には、興味あり(I: Interesting)、理解不能(NU: Not Understandable)、興味なし(NI: Not Interesting)のどれかを各ルールに与えるとともに、その判定を下した理由をコメントしてもらった。

二値クラスでは112個のルールが得られ、全ルールを評価した結果、Iが23個、NU、NIが89個となった。二値クラスの場合、ある程度興味深いルールは得られたが、新しい知識に結び付くほど評価の高いものは得られなかった。一方、多値クラスでは、高評価のルールが複数あり、特に図1に示すルールは医学的に新しい発見と

して高い評価を得た。これより、専門家は粒度の高い詳細なルールを求めていたと分かった。また、この専門家はあらかじめ知識の種となる仮説をいくつかの心の中に描き、仮説と対応するルールに興味を示したということも推定できた。

LOC ≤ 2.0 → C_COURSE = negative
LOC > 2.0 → C_COURSE = aphasia

図1: 多値クラスにおいて専門家が高く評価したルールの一例。

##### 2.2 従来のルールの興味深さ指標

興味深さ指標は、評価対象(各ルール/ルール集合)、基礎となる理論(事例数/確率/統計量/情報量/距離/構造の複雑さなど)、評価基準の性質(絶対的/相対的)、評価基準の種類(一般性/信頼性/意外性など)といった要因で分類できる。それ以外にも様々な指標がある。指標の詳細については文献[4]を参照して欲しい。本研究では、各ルールを評価対象とする興味深さ指標に着目し、できる限り多様になるよう全39種類の指標を選定した。

ここで、本研究において、なぜ興味深さ指標と真の興味の関係を調べるかを詳細に述べる。興味深さ指標はデータの数理的な性質の解析結果であり、当然、医学的知識に基づく真の興味とは単純に一致しない。しかし、もし興味深さ指標が専門家を持つ真の興味をある程度の精度で推定できれば、次のような活用が可能になるだろう。興味深さ指標をルールでソートして専門家に提示し、新たな視点を専門家に与える。興味深さ指標(単独、あるいは複数指標の合成)と真の興味を対応付けし、真の興味の要素還元的に解析する。興味深さ指標を利用し、ルールを推奨する。以上より我々は、興味深さ指標と真の興味の比較を試みることにした。

#### 3. 興味深さ指標と真の興味の評価実験

実験では、第2.1節で得た髄膜炎のルールと専門家による評価結果を用い、第2.2節で選定した興味深さ指標が真の興味をどの程度推定できるかを調べた。興味深さ指標による評価では、算出した評価値でルールを降順に並べ、専門家が実験で与えた正解のルールと同数のルールを上位から選び出した。

ここで、クラスが二値、多値の各々で、性能が上位、下位だった興味深さ指標を考える。表1の左側において、二値クラスでは $\chi^2$ -Measureの性能が最も高く、Coefficientの性能が最も低い。表1の右側では、多値クラスではLiftの性能が最も高く、Gray and Orłowska's(D)

\*同志社大学工学部

†静岡大学情報学部

‡慶應義塾大学理工学部

表 1: 実験結果の上位, 下位をまとめたもの. 左は二値クラスの結果, 右は多値クラスの結果を示す.

Binary Class						Multi-Value Class					
Rank	Measure	C1	C2	C3	Meta	Rank	Measure	C1	C2	C3	Meta
1	$\chi^2$ -Measure	12.33/23+	90.66/112+	0.28+	0.57	1	Lift	11.89/25+	105.78/132+	0.31+	0.54
2	Lift	12.00/23+	90.00/112+	0.33+	0.56	2	Accuracy	11.50/25+	105.00/132+	0.29+	0.53
3	Relative Risk	11.50/23+	89.00/112+	0.30+	0.55	3	$\chi^2$ -Measure	11.00/25+	104.00/132+	0.28+	0.52
	⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮	⋮
22	Prevalence	1.77/23	69.54/112	-0.32	0.25	22	Mutual Information	3.00/25	88.00/132	0.00	0.31
22	Recall	1.00/23	68.00/112	-0.09	0.25	22	Pecularity	3.00/25	88.00/132	-0.04	0.31
22	Jaccard	1.00/23	68.00/112	-0.12	0.25	22	Laplace Correction	3.60/25	89.20/132	-0.23	0.30
23	Coverage	1.00/23	68.00/112	-0.24	0.24	23	Coverage	2.00/25	86.00/132	-0.19	0.27
23	Goodman and Kruskal's	1.00/23	68.00/112	-0.19	0.24	23	Support	2.00/25	86.00/132	-0.25	0.27
24	Support	1.00/23	68.00/112	-0.28	0.23	24	$\phi$ Coefficient	2.00/25	86.00/132	-0.22	0.27
24	Gray and Orłowska's (D)	1.00/23	68.00/112	-0.34	0.23	24	Prevalence	1.12/25	84.24/132	-0.32	0.24
24	Laplace Correction	1.00/23	68.00/112	-0.31	0.23	24	Credibility	0.00/25	82.00/132	-0.18	0.23
24	$\phi$ Coefficient	1.00/23	68.00/112	-0.28	0.23	24	Gray and Orłowska's (D)	0.00/25	82.00/132	-0.33	0.22

の性能が最も低い。また、今回の実験条件では二値・多値に関わらず、Lift,  $\chi^2$ -Measure, Accuracy が興味深いルールへの推定に有効であることが分かった。また、Gray and Orłowska's(D),  $\phi$  Coefficient, Support が性能が低いことが分かった。

ここで我々が過去に行った肝炎データに対する同様の実験 [4] の結果と髄膜炎の実験結果を比較する。これらの結果から、対象とするデータや評価を行った専門家に関わらず、 $\chi^2$ -Measure と Lift は常に最上位にあり、医療データにおいて広く有効である可能性が示された。逆に、Gray and Orłowska's(D) は常に最下位にあり、専門家の興味に反すると考えられる。

我々は興味深さ指標を一つ一つの正解・不正解の傾向についても検討した(詳細データは紙面の都合上掲載していない)。その結果、指標によって異なり、相補関係を持つことが分かった。よって、特定の指標を組み合わせることにより、より高い推定性能が見られることもわかる。肝炎データにおいても、髄膜炎データと傾向が類似しており、指標を組み合わせることにより、同様に高い推定性能が見られる [4]。このことから、医療データマイニングにおいて指標を組み合わせることでそのデータに基づいた新たな評価基準を作り、推定性能を上げ得ると分かった。

#### 4. むすび

データマイニングでは様々なルールの興味深さ指標が提案されているが、これらと専門家がルールに対して持つ真の興味との関係はほとんど検討されていない。そこで本研究では、髄膜炎に関するルールと専門家によるルールの評価結果をデータとし、興味深さ指標が持つ真の興味への推定性能を実験によって調べた。その結果、 $\chi^2$ -Measure と Lift が安定した高い性能を持ち、専門家の興味をある程度推定できると示された。また、興味深さ指標間で評価の傾向は異なり、お互いに相補的な関係にあることが分かった。このことから、興味深さ指標の合成によって推定性能の向上が期待される。今後、我々の研

究は次の2つの方向に展開する予定である。1つは興味深さ指標の数理的性質と挙動の解析である。これによって興味深さ指標の客観的な性質を知るとともに、今回調べた主観的な性質と比較したい。もう1つは興味深さ指標の合成による推定性能の向上である。我々は現在、合成方法として、興味深さ指標を属性・専門家の評価をクラスとした決定木学習、および、興味深さ指標を変数・専門家の評価を出力とした最小二乗法による関数近似を検討している。

#### 参考文献

- [1] R. J. Hilderman and H. J. Hamilton: Knowledge Discovery and Measure of Interest, Kluwer Academic Publishers (2001).
- [2] Meningoencephalitis Diagnosis Dataset, KDD Challenge in Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'00), <http://www.slab.dnj.ynu.ac.jp/challenge2000/> (2000).
- [3] H. Hatazawa, N. Negishi, A. Suyama, S. Tsumoto, and T. Yamaguchi: Knowledge Discovery Support from a Meningoencephalitis Database Using an Automatic Composition Tool for Inductive Applications, Proc. of KDD Challenge in Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD-2000), pp.28-33 (2000).
- [4] M. Ohsaki, S. Kitaguchi, K. Okamoto, H. Yokoi, and T. Yamaguchi: Evaluation of Rule Interestingness Measures with a Clinical Dataset on Hepatitis, European Conference on Machine Learning and European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD-2004), will appear (2004).