



Title	APPROACHES TO THE UTILIZATION OF RULES INDUCED FROM DATASETS
Author(s)	大木, 基至
Citation	大阪大学, 2018, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/69624
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

論文内容の要旨

氏 名 (大 木 基 至)

論文題名

APPROACHES TO THE UTILIZATION OF RULES INDUCED FROM DATASETS
(データセットから抽出されたルールの利用に向けての取り組み)

論文内容の要旨

近年、計算機やネットワークシステムなどの発展により多くのデータが蓄積されるようになった。本研究では、データからラフ集合アプローチにより抽出されたルールの利用に関して三つのアプローチを提案する。

はじめに、ルールを設計知識として利用する上で役立つルールの評価指標について考察する。ルールの条件部の命題を満足させるように対象を設計すれば、ルールの結論部が実現できる。条件部の命題の解釈を誤る可能性を考慮すれば、条件部の命題の一部の解釈を誤っても結論部を概ね満たすルールが望ましい。残念ながら、従来の評価指標を用いてこのようなルールを選べない。そこで、このようなルールの選択に適した頑健性指標を提案する。数値実験により提案した頑健性指標の有用性を明らかにする。

次に、知識提供のためデータから抽出されたルールを公開する際に生じるプライバシー保護の問題を取り上げる。公開されたルールに適合する対象数が少ないとき、ルールと他の知識を連結することにより、対象のプライバシー情報が漏洩することがある。適合する対象数が k 以上のルールのみを公開することが考えられるが、 k 未満のルールの削除により、公開されたルールを用いた分類器の性能が劣化し利便性が損なわれるおそれがある。そこで、ルール間の類似度と一致度を用いてルールをクラスタリングし、適合対象数が k 以上となるルールを生成する方法を提案する。数値実験により、分類性能とプライバシー保護性能の観点から提案法の有用性を明らかにする。

最後に、公開されたルールの有効な利用法を考察する。公開されたルールは利用者が所有するデータと異なるバイアスがかかっている可能性がある。それらを安易に利用し分類器を拡張すると、分類精度が低下することがある。そこで、所有するルールによる分類器と公開されたルールによる分類器とを混合させる方法を提案する。それぞれの分類器から得られる妥当性スコアを混合するモデルがEMアルゴリズムにより推定される。データのバイアスを想定した数値実験により提案法の有用性を明らかにする。

論文審査の結果の要旨及び担当者

氏 名 (大 木 基 至)			
論文審査担当者	(職)		氏 名
	主 査	教 授	乾 口 雅 弘
	副 査	教 授	潮 俊 光
	副 査	教 授	飯 國 洋 二

論文審査の結果の要旨

本論文は、データ集合から抽出されたルールを想定して、ルールの利用に関する三つの手法に関する研究成果を纏めたものである。

第1章では、計算機やネットワークシステムなどの発展に伴った、蓄積されたデータからの機械学習や知識発見の有効利用の必要性が述べられ、本研究で提案する三つの手法とそれらの関係が説明されている。第2章では、本論文に必要な基礎事項と既存手法が紹介されている。第3章では、設計知識としてのルールの有用性を評価する指標が提案されている。設計者による条件部の解釈ミスを想定し、条件部に完全に適合しなくとも結論部が成立するようなルールが望ましいと考え、それに応じた評価基準として頑健性指標が提案されている。ルール選択に他の指標よりもこの指標を用いた方が有効であることが数値実験により示されている。第4章では、ルールを公開する際に必要となるプライバシー保護の問題が取り上げられている。適合する対象が少ないルールを連結することにより、ルールの匿名性を高める方法を提案し、ルール間の類似度と一致度に基づいたルールクラスターリング法が提案されている。ルール群の分類性能とプライバシー保護性能の観点から、提案法の有用性が数値実験により確認されている。第5章では、公開されているルールの有効利用法が考察されている。公開されているルールに地域の特色や年齢分布などによるバイアスが存在すると仮定したもとで、所有するルール群と公開されているルール群とを適切に混合するモデルが提案されている。数値実験により、提案モデルの有用性が明らかにされている。第6章では、本論文で得られた成果が纏められ、本論文に関連する今後の研究課題が述べられている。本論文では、高度情報化社会における蓄積されたデータや知識の活用に関連する三つの課題に対する新しいアプローチを与えており、その貢献は学術的に高く評価できる。これより、博士（工学）の学位論文として価値のあるものと認める。