



Title	大量データのためのコミッティ・マイニング手法に関する研究
Author(s)	寺邊, 正大
Citation	大阪大学, 2001, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/43459
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、 〈a href="https://www.library.osaka-u.ac.jp/thesis/#closed"〉 大阪大学の博士論文について 〈/a〉 をご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

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

The University of Osaka

氏名	寺邊正大
博士の専攻分野の名称	博士(工学)
学位記番号	第 16567 号
学位授与年月日	平成13年11月28日
学位授与の要件	学位規則第4条第1項該当 工学研究科通信工学専攻
学位論文名	大量データのためのコミッティ・マイニング手法に関する研究
論文審査委員	(主査) 教授 元田 浩 (副査) 教授 小牧 省三 教授 森永 規彦 教授 塩沢 俊之 教授 北山 研一 教授 河崎善一郎

論文内容の要旨

本論文は、大量データのためのコミッティ・マイニング手法について、筆者が株式会社三菱総合研究所、ならびに大阪大学大学院工学研究科通信工学専攻在学中に行った研究成果をまとめたものであり、以下の6章により構成した。

第1章は序論であり、本研究の背景と概要について述べ、本研究の位置づけ、およびその意義を明らかにした。

第2章では、データマイニングについて概説した。まず、データベースからの知識発見過程の部分過程であるデータマイニング過程とこれまでに報告されている実問題への適用事例について紹介した。次に、データマイニング過程のうち、本研究で研究対象とするデータからの分類器の導出過程について、これまでに提案されている手法と導出された分類器を利用したクラス判別方法について述べた。さらに、本研究で用いる分類器導出手法である決定木アルゴリズムについて、決定木の記述表現とともに概説した。

第3章では、コミッティ・マイニング手法の概念と研究課題について概説した。まず、コミッティ・マイニング手法の基本概念について述べたあと、関連する従来研究分野として、分散並列データマイニング、マルチエージェント学習、および組織学習をとりあげ、コミッティ・マイニング手法との関連について述べた。さらに、コミッティ・マイニング手法の研究課題である、分類器の性能向上と分類器導出の高速化について、その課題内容を具体的に述べた。

第4章では、分類器の性能を向上する手法として相関ルールによる属性生成手法を提案し、実験を通じて手法の有効性を評価した。まず、提案手法に関連する従来研究をあげ、これらの特徴と提案手法との関連について述べた。次に、相関ルールによるデータ間共起の記述方法と相関ルールを高速に抽出するアルゴリズムである Apriori について説明した。そして、提案手法である相関ルールにもとづく属性生成手法のアルゴリズムについて処理手順にそって詳述した。さらに、提案手法について実験用データセットを用いた評価実験を行い、その結果をもとに提案手法の評価と考察を行なった。

第5章では、分類器導出の高速化を実現する手法としてサブサンプリングとコミッティ学習の1つである Bagging を併用する手法である S^3 Bagging を提案し、実験を通じて手法の有効性を評価した。まず、提案手法の関連研究として、コミッティ学習手法とサンプリング手法についてとりあげ、これらの手法について分類器導出の高速化への有効性という面から概説し、残された研究課題について明らかにした。そして、提案手法である S^3 Bagging について手法の概要を概説したあと、処理手順について詳述した。さらに、提案手法について実験用データセットを用いた評価実験を行い、その結果をもとに提案手法の評価と考察を行なった。

第6章は結論であり、本研究で得られた大量データのためのコミッティ・マイニング手法に関する成果を総括した。

論文審査の結果の要旨

本論文では、分散並列情報処理環境の特徴を有効に活用したデータマイニング手法であるコミッティ・マイニング手法の概念について提案しており、概念を具体化するための課題としてデータマイニングにより抽出される分類器の性能向上と導出の高速化をとりあげ、これらの課題を克服するための新たな手法を提案している。その主な成果は、以下のように要約される。

- (1)分散並列情報処理環境を利用して、複数の自律的な情報処理主体がコミッティを構成し、これら情報処理主体が協調することにより各々の主体が独立に知識を抽出するよりも高速に、かつ精度の高い知識を抽出することが可能となることを述べている。そして、このような技術を実現するための手法としてコミッティ・マイニング手法を提案している。
- (2)コミッティ・マイニング手法を実現するための具体的な課題として、大量データから導出される分類器性能の向上と大量データからの分類器導出の高速化をとりあげ、課題の内容について述べている。
- (3)従来の属性生成手法では、大量データを扱う場合の処理速度やデータマイニングの初期過程において先験的な知識を必要とする点で、コミッティ・マイニング手法としては問題があることを指摘し、これを克服することが可能な手法として相関ルールにもとづく属性生成手法を提案している。
- (4)提案手法について実験用データセットを用いた評価実験を行い、提案手法が大量データからでも高速に新規属性を生成することが可能であり、かつ分類器の性能向上に効果があることを確認している。
- (5)提案手法の性能に影響を与えるパラメータである最小支持度と最小確信度の設定方法について実験データをもとに検討を行い、1属性あたりの属性値とクラス数の積から最小支持度を決定するための計算式を導出している。
- (6)大量データからの分類器導出の高速化を実現する手法としてサブサンプリングとコミッティ学習の1つであるBaggingを併用する手法であるS³ Baggingを提案し、実験を通じて手法の有効性を評価している。
- (7)提案手法について実験用データセットを用いた評価実験を行うことにより、提案手法が大量データからの分類器導出の高速化に有効であることを示している。
- (8)サンプリングの打ちきりを決定するための指標として、サンプリングしたデータの多様性を表す記述長を提案し、属性間の相関が小さいデータでは有効な指標であることを示している。

以上のように、本論文は分散並列情報処理環境におけるデータマイニング手法であるコミッティ・マイニング手法について多くの知見を含んでおり、当該分野の技術的課題の解決と発展に寄与するところが大きい。よって本論文は博士論文として価値あるものと認める。