

称号及び氏名 博士(理学) 蔣 錦賽

学位授与の日付 平成**28**年**3**月**31**日

論文名 分散データからのプライバシーを保護したファジィルールの抽出に関する研究

論文審査委員 主査 馬野 元秀
副査 瀬田 和久
副査 藤本 典幸

分散データからのプライバシーを保護したファジィルールの抽出に関する研究

蒋 錦賽

PCの普及により、さまざまな場所でさまざまな種類の電子データが大量につくられるようになった。このような大量のデータから有用な知識を取り出して有効に利用するのがデータマイニングである。このような知識は人間が利用するので、人間にとって分かりやすい形で表現することが重要である。知識をファジィ集合を用いて表現すると、自然言語に近い分かりやすい知識が得られるだけでなく、曖昧さを含むデータに対してもデータマイニングが可能になる。これがファジィデータマイニングである。

また、データマイニングが普及してくると、個人情報を含むデータからデータマイニングを行なうことが多くなり、個人情報を保護することを考えたプライバシー保護データマイニングが研究されている。プライバシー保護データマイニングには、主に2つのアプローチがある。1つは、個人を特定できないようにデータを変更する方法で摂動(perturbation)技術と呼ばれる。もう1つのアプローチは安全なマルチパーティ計算方式で、複数のパーティが協調して有効な知識を得る方法である。これは、あえて複数のパーティで処理することにより、それぞれが全体を見ないようにすることによりプライバシー保護を行なうものである。

本研究では、複数のパーティに同じ属性を持つ分散したデータ(カテゴリデータと数値データを含んでもよい)に対して、データ自身を他のパーティに転送せずに、計算過程に必要な値だけを転送して、分散データを全体としてのファジィルールの抽出する方法を提案する。提案する手法は、安全なマルチパーティ計算方式によりプライバシーを保護しながら、分散データに対してデータマイニングを行なう手法である。ルール抽出法は従来の1ヶ所にあるデータ集合からの量限定子を持つファジィルールの抽出手法に基づいている。そして、各分散データから得られたルールとデータ全体のルールを比べることにより、各分散データにどのような特徴があるかを調べることができ、データについてより多くの知識が得られる。そのために、各分散データからのルールとデータ全体からのルールを比べるためのルールの類似度を提案する。

また、同じ属性を持つ分散データからのルール抽出手法を基にして、分散データが異なる属性部分を持つ場合に対して、全体としてのファジィルールの抽出する手法を提案する。全体としては、データが欠損していると考え、欠損データを持つパーティにファジィ推論に用いるファジィルールの抽出する。このとき、同じ属性を持つ分散データからのルール抽出手法を拡張することにより、プライバシー保護を行いながら推論に用いるルールを抽出する手法を提案する。そして、抽出されたルールを用いて欠損データを推論する。最後に、同じ属性を持つ分散データからのルール抽出手法を用いて、プライバシーを保護しながら全体としてのファジィルールの抽出する。

実際のデータとして、**R.A.Fisher** のアヤメのデータと **UCI** の機械学習リポジトリのワインデータを用いてファジィルールを抽出する。

本論文は **6** 章で構成されており、第 **1** 章では、本研究の背景と目的について述べる。

第 **2** 章では、主なデータマイニング手法について述べ、プライバシー保護データマイニングの主な二つのアプローチである摂動技術と安全なマルチパーティ計算方式について述べる。

第 **3** 章では、ファジィ集合とその演算、ファジィ推論、主なファジィデータマイニング手法およびファジィ量限定子について述べる。また、ファジィデータマイニング手法であるファジィ量限定子を持つファジィルールの抽出手法について述べる。

第 **4** 章では、ファジィ量限定子を持つファジィルールの抽出手法を基にして、同じ属性を持つ分散データから、計算過程に必要な値だけを転送して、分散データ全体としてのファジィルールを抽出する手法について述べる。また、各分散データにおけるデータの特徴を表現するために、**2** つルールの比較方法を提案し、それを用いて各分散データから得られたルールとデータ全体のルールを比べる。そして、各分散データのルールとデータ全体のルールの中の似たものを削除すると、残ったものを各分散データにおけるデータの特徴を見なすことができる。実験では、アヤメのデータとワインデータを用いた結果を述べる。

第 **5** 章では、同じ属性を持つ分散データからのルール抽出手法を基にして、分散データが異なる属性部分を持つ場合について考え、全体としてのファジィルールを抽出する手法について述べる。実験では、アヤメのデータを **4** つの分散データに分け、適当に属性のデータを欠損させた。この手法の有効性を示すために、**2** つファジィルール集合の類似度の計算方法を提案し、実験の結果と元のアヤメデータの結果を比較する。

第 **6** 章では、本研究で得られた結果をまとめ、今後の課題を示す。

