

# 不確実データベースからの負の相関ルールの発見

藤田 岳行

## 要旨

近年、多くの企業では、顧客データなどの仕事に関する情報を集め、独自の経営戦略に生かそうとする試みがある。また、クラウドシステムなど登場によってよりデータの蓄積が容易なものになり、データベース化をし利用しようと言う気運が高まっている。企業だけではなく、twitterに代表されるソーシャルネットワークの普及によって個人のレベルでも多くの人がそうした大規模なデータを手にすることが容易になっており、一般的な情報源の1つとなっている。多くの人が大量のデータに触れ、そこから情報を得る行為が日常となった今日、効率的に有益な情報を得るための技術に高い要求が集まっていくのは自然な流れであり、要求に応えるべく様々な研究が行われている。大規模なデータの分析を行うための手法として小売業などのバスケット分析で用いられる相関ルール分析が知られており、また相関ルールの拡張として、負の相関ルールが提案されている。一方、対象となるデータに関しても多様化している。例えば、センサーネットワークを代表とするノイズを含むデータを持つデータベースを不確実データベースと呼び盛んに研究が行われている。また近年では、個人情報保護の目的で人為的に通常のデータベースにダミーデータなどのノイズを加えることも行われており、そうしたものも不確実データベースとして扱う。不確実データベースを対象として通常のデータベースと同様に相関ルールの発見を行う手法が提案されている。今後データマイニングの社会的普及がより進み、それに伴い、こうした不確実データベースに対する分析を行う機会が増加していくことが予想される。不確実データベースの分析の手助けを行うべく、本研究では既存研究の拡張として、新たに不確実データベースを対象とした負の相関ルールの抽出を行う手法の提案する。また評価実験を行い、現実的な時間で答えを得ることができることを確認する。