

報告番号	※甲	第	号
------	----	---	---

主論文の要旨

論文題目	A Study on Knowledge Discovery among Multiple Evolving Data Streams (データストリームにおける知識発見に関する研究)
氏名	范薇

論文内容の要旨

提出した論文「データストリームにおける知識発見に関する研究」では、実世界の社会に行われる事象間の複雑の相関の解明、イベント発生の解釈を目指し、データストリームにおける離散的相関・連続的相関を発見する手法の提案から、複数の情報源からストリームデータ間の相関を解析する一連の処理方法を提案するまでの研究成果をまとめた。

第1章は序論で、研究における前提としてデータストリーム処理の必要な要素を整理、事象間の相関の捉えために、基礎となる相関関係を分類している。そして、研究の目的として、分類された二つの基礎な相関関係（離散的相関と連続的相関）を効率的に解析する上で、事象間に行われた複雑な相関関係を分析できるとともに、既存研究と比較により本研究に提案した相関知識の発見する手法を示した。

第2章では、本研究におけるストリームデータの処理モデルを紹介する上で、離散的相関、連続的相関、さらに複雑な相関知識を発見する既存研究の提案手法の問題設定と欠点を議論している。既存研究と比較により本研究の新規性と有効性を述べた。

第3章では、データストリーム間の離散的相関知識を発見する手法を提案している。本手法では連続的に流れてくるトランスザクションデータから属性間の相関性かつ数値的比率関係を抽出している。相関している属性の発見により、既存研究の頻出パターンをマイニング手法では希有なかつ重要なイベントを発見できる。また、データストリームから比率関係の抽出もはじめての研究であり、既存の論理的属性しか扱わない制限から抜け出す、量的関係を明らかにした。

第4章では、データストリーム間の連続的相関知識を発見する手法を提案している。連続的相関では、時間上に連続的であるデータ間の依存性を考慮し、第4章と第5章では二つの手法を提案している。この章では、複数の時系列データから相関分析により、隠れ変数を求めている。提案手法のポイントとして、ストリームデータの変動の検出と逐次処理の実現であり、この手法により、次元の削減やデータ圧縮や異常検出などの応用に繋がっている。

第5章では、クラスタリング技術に基づき複数ストリームデータ間の連続的相関

係を抽出する手法を提案している。本手法では研究対象が部分系列と捉え、柔軟性を持つ任意の部分系列間のクラスタリングを実現している。一つは観測されたデータストリームが時間上で最近のデータと強い相関性を持つことを考慮した、recent-biased 近似手法により、高速に類似計算できる点である。もう一つは、データストリームの近似データが階層構造に保存される点である。そこで、データストリームの変動性により類似関係の変動の検出や、任意の期間内の部分ストリームデータのクラスタリングを実現している。

第 6 章では、研究目的として社会に行われる事象間の複雑の相関の解明にあたり、一つの例における、提案した二つ基礎的相関関係（離散的相関関係と連続相関関係）を組み合わせることで解明する手法を提案している。この例では、オンライン新聞記事により株価データの変動を予測することを目的し、新聞記事に現れる語句と株価データ間の相関関係を解明した。一つの新聞記事から株価データを求めるのが離散的相関を解明することともに、ある期間中に連続的に公開された新聞記事間の関係と株価時系列データのトレンドの変化を考慮した。提案手法はこれらが連続的相関関係の範囲と見なしている。そして、提案された解析フレームワークには離散的相関関係の抽出と連続的相関関係の抽出する処理を組み合わせることで、高い予測精度を得ている。

第 7 章は結論であり、本論文の総括と今後の研究課題について述べている。

