

審査の結果の要旨

氏名 肥村 洋輔

本論文は、「Unsupervised Structural Pattern Analyses for the Management of Network Traffic and Configurations（和訳：ネットワークのトラフィックおよび構成のマネジメントにおける教師なし構造パターン分析）」と題し、コンピュータネットワークの管理（マネジメント）における教師なし構造パターン分析手法の提案と実践を行っている。ネットワークの管理において、特にトラフィックや構成の管理は重要な活動である。このようなネットワーク管理に教師なしデータ分析手法を用いることで、傾向把握や異常検出などへの応用が期待できる一方で、分析結果に対する解釈が難しく適切な対処が取りにくいという問題がある。この問題に対して、トラフィックデータ・構成データそれぞれに対し、ドメイン知識を持つ者が視覚的に解釈可能な構造的なデータモデルに基づく教師なし分析（教師なし構造パターン分析と呼ぶ）の適用方法を提案し、実データを用いて評価している。

本論文は、5章から成り立っており、コンピュータネットワークにおける、トラフィックをおよびネットワーク構成の管理を対象とした、教師なし構造パターン分析に関する議論を中心に行っている。

第1章「Introduction」では、本論文がスコープとして扱う、コンピュータネットワーク管理の動機、統計的データ分析を用いた管理支援の技術的な契機、その中でも教師なし分析の活用の妥当性と課題について議論している。コンピュータネットワーク、特に、インターネットにおいて、その安定稼働を実現する管理業務の重要性は増大している一方で、接続ホスト数の増加・仮想化技術の普及により管理が複雑化している。これに対して、前述の接続ホスト数の増大・仮想化技術の普及から、相互比較可能な多数のオブジェクトが単一管理ドメイン内のネットワーク上に観測される（個々のホストのトラフィックや、個々の仮想ネットワークの構成が相互比較可能である）という着眼点に基づき、データ間の関連分析により典型パターンや統計的異常を検出する教師なし分析を活用することの有効性を議論している。他方で、教師なし分析のような統計分析は一般的にドメイン非依存の抽象的な数値データを対象とすることから、分析結果に対する適用先ドメインの文脈に合わせた解釈が難しく、分析結果に対して適切な対処が取りにくい点を考察している。これに対して、ドメイン知識を持つ者が視覚的に解釈可能な構造的なデータモデルに着目して、そのデータモデルに対して教師なし分析（教師なし構造パターン分析）を適用するアプローチの有望性を議論している。

第2章「Structural Traffic Pattern Analysis - Synoptic Graphlet : Bridging the Gap between Supervised and Unsupervised Profiling of Host-level Network Traffic」では、ネットワークトラフィックの教師なし構造パターン分析について議論している。ネットワーク上を流れるデータであるトラフィックを分析することは、利用傾向の把握や、不適切な通信の検出などの応用があり、重要な管理タスクである。トラフィック分析の研究領域において、通信特性を視覚的なグラフ構造により表現する graphlet と呼ばれるデータモデルが知られているが、このモデルに関する分類ルールを手で経験に基づき作成する従来方法では未知の挙動への対応などに課題があり、他方で事前定義された分類ルールを必要としない教師なし分析を行うにも前述の理由から分析結果に対する解釈が難しいという問題がある。この問題に対して、本研究では、graphlet におけるグラフ構造の特徴に関するものを経験的に選択した特徴量ベクトルを抽出し、階層型クラスタリングの教師なし分析結果として得られるクラスタ中心の特徴量ベクトルから、クラスタの特徴を表現する graphlet 構造を復元し可視化する手法を提案している。提案手法を、実トラフィックデータを用いて、複数のルールベースの分析手法による分析結果を用いたクロスバリデーション

を行うことで、その有効性を評価している。

第 3 章「Structural Configuration Pattern Analysis - Discovering Configuration Templates of Virtualized Tenant Networks in Multi-tenancy Datacenters via Graph-mining」では、ネットワーク構成の教師なし構造パターン分析について議論している。ネットワークの構成を分析することは、典型的な構成パターンを抽出して、再利用可能なテンプレートを作成して利用するなど、構成管理の効率を向上することに有効であると考えられる。本研究では、複数の仮想ネットワークが共有の物理インフラ上に集約されるマルチテナント型データセンターネットワークを対象とし、特に仮想ネットワークのトポロジ構成を分析している。様々な構成要素からなる仮想ネットワークトポロジを属性付きグラフ構造に抽象化し、グラフ編集距離により類似度を定義し、階層型クラスタリングの教師なし分析を適用することにより、典型構成パターンを抽出可能にしている。グラフ編集距離は、トポロジ間での構成要素やそれら接続関係の追加削除に関するオペレーションの差分の量としても理解できることから、この手法は解釈性の意味でも適していると考えられる。また、本手法に関して実践的な評価も行っている。

第 4 章「Discussion and Future Perspectives Towards Efficient Analysis for Management of Field Domains」では、コンピュータネットワーク領域を対象とした第 2 章・第 3 章の知見を発展させ、実フィールド領域における教師なし構造パターン分析に向けての発展的な議論を行っている。近年、実フィールドにおいても、センサデバイスやそれらを繋ぐネットワークの普及が予測され、データを活用したマネジメントの効率化が期待される。特に、実フィールドにおいても、例えばビルの中の個々の部屋や機器のように、類似したオブジェクトのデータが集まるようになることから、相互比較型の教師なし分析の潜在的な有効性を議論している。一方で、データの特性として、コンピュータネットワーク領域ではホスト間の一定の通信プロトコルに基づいた相互作用に比較的閉じる系であることから、直接的な観測データから前章・前々章で示したような解釈性は相対的に得やすいことに対して、実フィールド領域ではサブシステム間の相互作用に加えて利用者の意図や外部環境などの外的要因にも依存する開いた系であることから、観測データのみからは分析結果に対する解釈が難しくなることを考察している。これに対して、単一種類のデータに対する相互比較型の教師なしの分析から開始し、解釈向上に必要な追加情報の考察と補完データの導入を順次行うプロセスを検討している。その実例として、本研究では、ビル内の様々な情報を計測したビルセンサデータのデータセットの中から、個々の部屋の空調機器の温度データを分析することから開始して、その後、空調機器の設定情報・稼働状態や、さらに気象情報（外気温）のような追加データを導入しつつ教師なし分析を行うことで、部屋内の温度異常イベントの抽出およびその解釈性を向上させる例を示している。最後に、第 5 章では、「結論」として、本論文の総括と貢献についてまとめを行うとともに、今後の課題についての議論を行っている。

以上を要するに、本論文は、データ分析を用いたネットワークマネジメント性の向上に向けて、上述のように、教師なし構造パターン分析に基づいて解釈性を考慮した分析手法の提案や実データを用いた評価を行っており、また、将来的に重要性が増してくる実フィールドマネジメントに向けた考察と実データを用いた検討も行っており、情報理工学における実践的研究と今後の発展に寄与・貢献が少なくない。

よって、本論文は、博士（情報理工学）の学位請求論文として合格と認められる。