論 文 の 内 容 の 要 旨

論文題目　　ネットワークログの因果解析による障害の
原因究明支援技術に関する研究

氏　　名　　小林　諭

現代の情報社会を支えるネットワークシステムはその重要性を増す一方であり、システムの継続的な運用が必要不可欠となっている。特にシステムの復旧のためのトラブルシューティングにおいては、情報の具体性、普及による一般性の面で優れるシステムログの活用が有効である。しかしシステムの規模の拡大および複雑化に伴いシステムログをはじめとした運用データの規模は増大しており、効率的な活用のため自動解析技術への需要が高まっている。

システムログを用いた自動解析の分野では異常検知、異常箇所の特定、障害の原因究明など広い視点からアプローチがなされてきた。特に障害の原因究明を目的とした研究においては、他のデータよりも文脈的な情報を示すシステムログの活用は大きく注目されている。しかし、既存のシステムログを用いた障害の原因究明支援技術は運用者にとって必ずしも実用的なものとはなっていない。これは既存技術の多くが過剰な情報抽出により運用上真に有用な情報を埋もれさせてしまっていることに起因する。

過剰な情報抽出を防ぐためのアプローチの1つとして、因果解析が挙げられる。障害の原因究明においては、擬似相関などの不適切な情報を抑制できる点で運用情報活用の効率化に貢献できる。一方でログを対象とする既存の因果解析技術は障害の周辺イベントを対象とした診断の形を取るものが主流であり、解析の効率化のため扱うデータの範囲に制限が発生するものとなっている。オペレータによるシステムログを用いたトラブルシューティングは本来システムの定常状態や過去の事例との対比を伴うものであり、より広範囲のデータを用いた探索的なアプローチに基づく自動解析が実現できれば従来よりも実オペレーションの手順に即したより高度な原因究明支援が可能となるだろう。そのためには、システムログの因果解析をより高速に、効率的に行うための手段が必要である。

システムログの因果解析の効率化は、ログ中の情報の取捨選択によってなされる。この実現には、システムログのデータ的な、あるいは時系列的な構造に関する知見が必要となる。過去の研究において行われてきたシステムログの構造についての解析は、その多くがシステムごとの汎用性を考慮して統計的な性質を主とする手法であり、システムログを出力するシステムの性質に踏み込んだものではなかった。これに対し本研究では大規模ネットワークシステムという環境に焦点を当て、その環境およびログの構造的性質を活用した解析を行うことで運用者にとってより信頼に足る解析とすることを重視する。

本論文では、大規模ネットワークシステムにおいて運用者のトラブルシューティングに対するより実用的な情報提供を行うという形での支援を目指す。この情報提供を、システムログの自動解析によりログ中のイベント間の因果関係を推定する技術を提案することにより実現する。この技術の軸は因果推論に基づくグラフ解析手法であるPCアルゴリズムの利用にある。さらに解析の効率化のため、ネットワークシステムのログの構造に関する知見に基づく手法選択及び前処理・後処理を行なう。このうち、ログのフォーマットの半言語的構造に基づくLog templateの生成手法、ログ出力の周期性・恒常性に基づく時系列の前処理手法、ログデータの時系列的なスパース性を考慮した条件付き独立検定手法、得られる因果関係の恒常性に着目した因果DAGの後処理手法、の4つがログの構造の知見に由来する特に重要な要素技術となっている。これらにより、広範囲のシステムログを対象とする探索的な因果解析を可能としている。

この技術の実際のネットワークシステムにおける有用性を評価するため、本論文では実運用されている大規模ネットワークの運用データを用いた解析を行なっている。複数のケーススタディの検証により、実際に発生した障害において有用な情報の提供に成功していることを確認する。またこのネットワークシステムの運用チームにより記録されたトラブルチケットとの対応付けにより、発生した大規模な障害の74%において運用者にとって有用な因果情報の検出に成功していることを示している。さらに提案手法がトラブルチケットに記録されていない障害についても因果情報を検出しており、そのような障害の再発を未然に防ぐ上でも重要な貢献があることを示唆している。

本研究はネットワークシステムのトラブルシューティングの支援においてより重要性の高い情報の提供を可能とした点、そしてその情報を運用者が効率的に利用できる範囲まで精選・圧縮をおこなった点に大きな貢献がある。得られた因果関係情報は探索的解析により知識ベースとしての価値があり、他のデータや手法と連携した更なる解析を行う助けとなるだろう。またシステムログの構造についての知見は今後のシステム運用データ解析分野の更なる発展に寄与することが期待される。