論文審査の結果の要旨

氏名 Stefan Jan Skudlarek

本論文は「Unsupervised Anomaly Detection within Non-Numerical Sequence Data (非数値系列データにおける教師無し異常検出)」と題し、4章から構成されている。さまざまな複雑なシステムにおいて、異常が発生した場合に、そのモニタリングデータから異常を自動的に検出できることが望まれている。しかし、異常は滅多に生起しないため、異常時のデータを事前に全て知ることは困難である。そのため、多くの異常検出アルゴリズムでは、正常時のデータを事前に得てその特性を解析し、その特性と比較することにより異常時データを検出することがなされている。そのような場合を教師有り異常検出というが、複雑なシステムにおいては、異常が全く含まれていないことが保証された正常データを事前に得ることが困難な場合も多い。そのような場合は、モニタリングデータのみを用いて異常の有無を検出しなければならない。このような場合を、教師無し異常検出というが、モニタリングデータが非数値データからなる場合は、数値データの場合に比べて、異常検出が特に難しくなる。本論文は、この最も難しい場合を取り扱っており、新しい異常検出アルゴリズムを提案すると共に、その性能を理論解析およびシミュレーションにより、評価している。また、従来法との比較を行い、提案手法の優れている点を明らかにしている。

第1章「Introduction」では、異常検出に関する研究の背景、目的および従来手法を紹介し、本研究の位置付けを与えている。さらに、本論文で取り扱う異常検出のフォーマルな問題設定を与えている。

第2章「Unsupervised Anomaly Detection based on Average Index Difference」では，同じシンボルが出現するインデックスの差の平均値を用いる新しい教師無し異常検出法を提案している。まず、2.1節で平均インデックス差関数を定義し、その関数の特性を理論的に明らかにしている。次に 2.2 節で、異常データ系列が1カ所に固まっている場合に対して、平均インデックス差を用いた教師無し異常検出アルゴリズムを提案している。さらに、その場合の平均インデックス差関数の特性を解析し、理論的に異常検出が可能であることを明らかにしている。2.3節では、異常データ系列が全系列の中に分散している場合を考え、そのような場合でも異常検出ができるアルゴリズムを提案している。理論解析により、そのアルゴリズムで異常検出が可能であることを示すと共に、適切なパラメータ値を理論的に導出している。2.4節では，数値実験により提案アルゴリズムの性能を評価している。人工的に作成したデータおよび実問題としてコンピュータネットワークの不正使用データを用いて異常検出を行った場合の性能を、従来知られている手法と比較し、提案手法の長所を明らかにしている。なお、本章において示された定理の証

明は、「Appendix」にまとめて記載されている。

第 3 章「Unsupervised Anomaly Detection based on Representative Sequence Selection」では、データ系列を距離空間に写像し、その空間内の距離を利用して、データの異常を検出する手法を提案している。従来の手法と異なる点は、正常データの特性を求めるのではなく、正常データの代表点を求め、その代表点からの距離に基づき分類することにより、計算量が少なくてすむように工夫されている。3.1 節で上記のような新しい異常検出アルゴリズムを提案し、パラメータが満たすべき条件を与えている．3.3 節で上記アルゴリズムで使用する距離行列を与えるカーネル関数の定義を与えている．3.4 節では提案アルゴリズムの計算コストを評価している。さらに、3.5 節で、人工的なデータおよび実データとしてプロテインデータを使用して、性能評価を行なっている。一般に、正常データの分散が異常データの分散よりかなり大きい場合は異常検出が難しく、従来の手法ではうまく検出できていなかった。しかし，提案した異常検出アルゴリズムは、そのような場合でも異常検出の成功率が高いという優れた特長があることを明らかにしている。

第 4 章「Conclusion」では、本論文の成果をまとめると共に、今後の研究課題を明らかにしている。

なお、本論文の成果は、山本博資との共同研究であるが、論文提出者が主体となって新しいアルゴリズムの提案、解析、数値実験を行なったものであり、論文提出者の寄与が十分であると判断する。

したがって、博士（科学）の学位を授与できると認める。