

審査の結果の要旨

氏 名 藤原 靖宏

本論文は「Efficient Sequence Data Analysis with Hidden Markov Models (隠れマルコフモデルによる高速な系列データの解析手法)」と題し、英文7章から構成されている。自然言語処理、バイオインフォマティクス、音声認識等の系列データの解析に広く用いられている隠れマルコフモデル(HMM)の復号には、最適解を与えるビタビアルゴリズムや、様々な近似アルゴリズムが適用されているが、ビタビアルゴリズムは状態数やデータ量の大きい問題に対して非効率であり、近似アルゴリズムには誤り限界を保証できないという問題点が存在する。本論文では、HMMの復号、モデル選択、ストリームモニタリングという3つの問題設定において、複数の状態をまとめ上げた粗粒度の状態遷移における近似尤度を算出し、最適の尤度を取り得ないパスやモデルを効果的に枝刈りする手法を提案し、ビタビアルゴリズムより高速に最適解を求められることを示している。

第1章は、「Introduction (序章)」であり、本論文の背景および目的について概観し、本論文の構成を述べている。

第2章は、「Preliminary and Related Work (基本事項及び関連研究)」と題し、隠れマルコフモデル及びビタビアルゴリズムの基本的な解説を行い、その高速化及び応用に関する関連研究をまとめている。

第3章は、「Efficient Likelihood Computation for HMM (HMMのための効率的な尤度計算手法)」と題し、単一のHMMを高速かつ最適に復号するアルゴリズムを提案している。本アルゴリズムは縮退ラティスと呼ばれるデータ構造を用いて効果的な枝刈りを行うものである。自然言語処理における系列ラベリングタスクによる評価実験を行い、ビタビアルゴリズムと比較して最大で約400倍の高速化が可能であることを示している。

第4章は、「Efficient Search Method for HMM Data Set (HMMデータセットの効率的な探索手法)」と題し、与えられたデータ系列に対し、多数のHMMから最大の尤度を持つモデルを効率的に探索するアルゴリズムを提案している。ゲノムデータ、DNAデータ、交通量データ等を用いた評価実験を行い、提案アルゴリズムが、状態数、モデル数の大きな問題に対し、ビタビアルゴリズムと比較して数十から数百倍高速であるこ

とを示した。

第5章は、「Fast Data Stream Monitoring with HMM Data Set (HMMデータセットを用いた高速なデータストリームモニタリング手法)」と題し、与えられたデータストリームをモニタリングし、多数のHMMから最近のウィンドウ幅のデータ系列に関して最も尤度の高いモデルを効率的に探索する手法を提案している。DNAデータ、交通量データ等を用いた評価実験を行い、状態数、モデル数の大きな問題に対し静的な解法及びビタビアルゴリズムと比較して、数十から数百倍高速であることを示した。

第6章は、「Efficient Centrality Monitoring for Time-evolving Graphs (時系列グラフにおける効率的な中心性モニタリング手法)」と題し、これまでに提案してきた高速なHMMアルゴリズムが、時間によって変化する時系列グラフのモニタリング問題に対しても有効であることを示している。変化するグラフをモニタリングし、中心性と呼ばれる尺度の計算を更新し続ける問題に対して、これまでと同様の縮退、枝刈りの手法を適用したアルゴリズムを提案している。P2Pネットワーク、ソーシャルネットワーク、Webのリンク構造を用いた評価実験を行い、10万から50万ノード規模のグラフにおいて既存手法と比較して最大で約100倍の高速化が可能であることを示している。

第7章「Conclusions (結論)」では、本論文の成果と今後の課題について総括している。

以上これを要するに、本論文は、隠れマルコフモデルを用いた復号、モデル選択、及びストリームモニタリング問題に対して、最適解を保証しながらビタビアルゴリズムよりも高速なアルゴリズムを構築する手法を提案するものであり、様々な実データを用いた評価実験によりその有効性を明らかにすると共に、同様の手法が時系列グラフの中心性モニタリングにも適用可能であることを示しており、情報理工学上貢献するところが少なくない。

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。