

審査の結果の要旨

氏名 鈴木 惇

大量のデータからの知識を獲得するための機械学習において、近年ではグラフなどの複雑データを扱うことが多くなってきている。このような複雑データの本質を抽出し、高精度な機械学習を行うためには低次元空間への埋め込みという方法が重要となる。そのような低次元空間は一定の曲率をもった Riemann 空間であることが多い。これを空間形とよぶ。空間形上で機械学習を行うためには、どのように埋め込みを行うかといった問題（モデル化の問題）、空間形上でいかに最適な表現を得るかといった問題（最適化の問題）、空間形上の学習の良さを評価する問題（評価の問題）といった3つの大きな問題がある。本論文では、上記3つの問題を解決するための新しい方法論を提示した。モデル化の問題に対しては、順序付き双曲空間埋め込みと多重関係グラフ埋め込みという新しい埋め込みモデルを提案した。最適化の問題に対しては、すでに存在する Riemann 確率勾配降下法と自然確率勾配降下法という最適化アルゴリズムの特性を新しい視点から明らかにした。評価の問題については、記述長最小原理に基づいてモデル選択するための新たな効率的計算方法を提案した。上記方法論につき、理論的な性能保証または実験的検証を行うことにより、空間形上の機械学習の体系を築いた。

本論文は「Machine Learning over Space Forms」（空間形上の機械学習）と題し、7章からなる。

第1章「Introduction」（序）では、機械学習の問題を独自の視点から整理し、モデル化、最適化、評価の問題を基礎的問題として抽出している。そして、空間形上においてこれらの問題を解決することの意義を説いている。

第2章「Preliminaries: Differential Geometry for Machine Learning」（準備）では、空間形上の機械学習の解析に必要な微分幾何学の基本的知識をまとめている。

第3章「Relational Data Embedding in Space Forms I: Ordinal Embedding」（空間形上の関係データ埋め込みI）では、通常のグラフよりもやや緩い関係性を有する順序付きデータを対象にして、これを埋め込むための方法として「順序付き埋め込み」を提案している。順序付き埋め込みのアイデアのポイントは、埋め込み先の双曲空間の距離関数と順序関係を整合させた損失関数を定義したことにある。この損失関数を最小化するようなデータの埋め込みが有効であることを実験的に検証している。具体的には、

与えられたグラフの一部から未知の順序関係を予測する問題に適用し、Euclid空間に埋め込む場合より双曲空間に埋め込む場合の方が高い予測精度を実現することを示した。

第4章「Relational Data Embedding in Space Forms II: Multi-relational Graph Embedding」（空間形上の関係データ埋め込みII）では、ノード同士が複数の階層的関係を有するという、通常のグラフより強い関係性がある場合を対象にして、「多重関係グラフ埋め込み(multi-relational graph embedding)」を提案している。多重関係の埋め込みに関してはTransEなどの先行研究が存在していた。そこでは曲率ゼロの埋め込み空間上でノード間の関係を直線的に移動して得られるものとして表現していた。しかしながら、多数のノードに1つのノードが対応し、かつ複数そのような関係を有する場合には、TransEではこれを適切に表現することはできなかった。本論文では、曲率がゼロでない多様体上にノードを埋め込み、その関係性を多様体上の測地線に沿って表現する方法を提案した。これにより、多重の多対1の関係性をもつデータについても、効果的に可視化でき、しかもTransEを含む従来の埋め込み法に比べて、ラベルの分類精度が優位に上回ることを実験的に検証した。

第5章「Optimization in Space Forms」（空間形上の最適化）では、双曲空間などの多様体上でデータを最適に配置する際の最適化アルゴリズムについて解析している。代表的な多様体上での一次オーダーの最適化アルゴリズムとして、自然確率勾配法と(NSG)リーマン確率勾配(RSG)法が存在する。本論文では、双曲空間上での与えられた点集合の重心を求めるという問題に対して、RSGでは最適解に収束するが、NSGでは必ずしもそうはならないことを理論的、実験的に示し、それらの本質的な違いを明らかにした。

第6章「Information Criterion in Space Forms」（空間形上の情報量規準）では、記述長最小原理に基づいて機械学習で得られたモデルを評価する問題を扱っている。MDL原理では、確率モデルのクラスを用いてデータを正規化最尤符号(Normalized Maximum Likelihood: NML)化する際の符号長をモデルの評価基準にしているが、そこに現れる正規化項の計算が一般的には難しいとされている。本論文では、正規化項をFourier変換の技法を用いて解析的に算出する新しい手法を提案した。それによると指数分布族に属するモデルに対してNML符号長が全て解析的に計算できることになり、空間形上の機械学習の評価の可能性を大きく切り開いた。

第7章「Conclusion」（結論）では全体を総括し、将来の展望を与えている。

以上を要するに、空間形上の機械学習において、埋め込みモデルの新しい方法論を提案し、最適化のアルゴリズムの基本的特性付けを行い、普遍的な評価手法を確立した。このことによって、機械学習分野に意義のある貢献を果たしている。空間形上の機械学習は今なお発展が著しい分野であるが、その基礎的な部分に対して本論文は有効な方法論を体系的に提示しており、数理情報学の発展に大きく寄与している。

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。