

Efficient Learning Algorithms for Rankings and Other Combinatorial Concept Classes

末廣, 大貴

<https://doi.org/10.15017/1441266>

出版情報 : 九州大学, 2013, 博士 (情報科学), 課程博士
バージョン :
権利関係 : 全文ファイル公表済

(別紙様式2)

氏 名 : 末廣 大貴

論文題名 :

Efficient Learning Algorithms for Rankings and Other Combinatorial Concept Classes
(ランキング及び組み合わせ論的概念クラスに対する効率的な学習アルゴリズム)

区 分 : 甲

論 文 内 容 の 要 旨

近年、情報検索、商品推薦、金融リスクの予測、スケジューリングなど様々な分野において、アイテム間の順位を推定するランキング学習の技術が注目されている。これらの分野では訓練データが膨大であるため、得られるランキングの精度が高いだけでなく、ランキングが得られるまでの計算効率が良いことが、学習アルゴリズムに求められる重要な要件となっている。本論文は、統計学習およびオンライン予測と呼ばれる2つの理論的な学習モデルのもとで、ランキングに関する効率の良い学習アルゴリズムの設計法について論じたものである。

統計学習モデルにおいては、ランキング関数を学習する問題として、最も基本的な二部ランキング問題を取り上げる。二部ランキング問題とは、正のラベルが付いた事例（正例）および負のラベルが付いた事例（負例）からなる訓練サンプルが与えられたとき、負例よりも正例に大きい値を与えるような事例空間上の実数値関数（ランキング関数）を求める問題である。この問題は、正例と負例の直積空間上の二値分類問題に帰着できることが知られており、この帰着関数とサポートベクターマシン（SVM）などの適当な二値分類学習アルゴリズムを組み合わせることにより、性能が理論的に保証された二部ランキングアルゴリズムが得られる。この手法はランキングSVMと呼ばれ、広く用いられている。しかし、ランキングSVMには、帰着の過程でサンプルの大きさが2乗のオーダーで増大してしまうという問題がある。すなわち、学習アルゴリズムとして1ノルム正則化SVMや2ノルム正則化SVMを用いるとすると、ランキング関数を求めるためには、元のサンプルのサイズを m としたとき、それぞれ $O(m^4)$ の大きさの線形計画問題や2次計画問題を解く必要があり、 m が大きいとき極めて効率が悪い。本論文では、SVMによる定式化を見直すことにより、2部ランキング問題を $O(m^2)$ の大きさの最適化問題として定式化し、大幅に計算量を削減することに成功した。さらに、得られたランキング関数の汎化性能に関して理論的な保証を与えた。特に、1ノルム正則化ランキングSVMを再定式化することによって得られたアルゴリズムは、元の1ノルム正則化ランキングSVMに匹敵する性能を持つ初めての実用的なアルゴリズムである。

ランキング学習の応用研究として、本論文では、プロ棋士の対戦記録である膨大な棋譜データから将棋の評価関数を構築する問題に取り組んだ。棋譜データに現れる局面を正例、現れない局面を負例と考えることにより、この問題は、局面を事例空間とする二部ランキング学習の問題とみなすことができる。このように、コンピュータ将棋の分野では、機械学習の技術を用いて局面の評価

関数を自動獲得する手法が発達し、その強さがプロ棋士に匹敵するほどに至っている。しかし、局面の特徴ベクトルを構成する評価項目をどのように設計するか、すなわち、特徴抽出の問題は依然として大きな問題である。従来、各評価項目の設計は手動で行われているため、得られる評価関数は、結果として作成者の棋力や感覚に強く依存するものとなっている。そこで本研究では、カーネル法と呼ばれる技術を用いて、 $10^{13} \sim 10^{14}$ に及ぶ複雑な評価項目を網羅的に自動生成する手法を提案した。さらに本論文で開発した効率の良い2部ランキング学習の手法と組み合わせることで、実時間で評価関数の学習を行うアルゴリズムを開発した。いくつかの予備実験では、本提案手法が有望であることを示唆する結果を得ている。

オンライン予測問題とは、株式投資のように、予測に基づく意思決定と、環境からのフィードバックの開示が逐次的に繰り返される意思決定問題である。各試行において、アルゴリズムは自らの決定と環境からのフィードバックによって定まる損失を被る。アルゴリズムの目標は、累積損失を最小化することである。情報検索や推薦システムなどの多くの問題は、ランキング、すなわち、あるアイテム集合上の順列のクラスを決定空間とするオンライン予測問題としてモデル化することができる。同様に、多くの自然なオンライン組み合わせ最適化問題は、対応する組み合わせ集合を決定空間とするオンライン予測問題としてモデル化できる。例えば、ある固定されたグラフ上の道や全域木のクラスを決定空間とすることにより、それぞれオンライン最短路問題やオンライン最小全域木問題が得られる。このようなオンライン予測の問題に対し、射影および分解と呼ばれる幾何的な操作を用いる一般的なアルゴリズム設計スキームが知られている。しかし、順列、道、全域木などの組み合わせ構造を予測の対象とする場合、概念クラスの大きさが指数的となるため、射影と分解を多項式時間で実現するアルゴリズムが存在するか否かは自明ではなく、クラスごとに個別にアルゴリズムを設計する必要があった。そこで本論文では、組み合わせ構造の広いクラスに対し、劣モジュラ最適化の手法に基づき、射影と分解を多項式時間で実現するアルゴリズムを与えた。このクラスは、順列、全域木、 k 部分集合族などを部分クラスとして含むため、これら多くの組み合わせ構造のクラスに適用可能な、世界初の統一的なオンライン予測アルゴリズムが得られたことになる。

〔作成要領〕

1. 用紙はA4判上質紙を使用すること。
2. 本文の文字サイズは10.5ポイント（「論文内容の要旨」の文字は12ポイント）
1行の字数44字，行数42行、余白（左右20mm，上下25mm程度）をあげ，頁数は記入しない。
3. 要旨は1頁に2,000字程度（最大2頁以内を目安）にまとめる。
4. 図表・図式等は随意に使用のこと。
5. 氏名は外国人の場合，カタカナ表記（漢字圏の学生は漢字）で記入する。
6. 論文題名は論文目録と合わせる。（外国語の場合は和訳をカッコ書きで付記する。）
7. 区分には甲または乙を明示すること。

この原稿は，「九州大学博士学位論文内容の要旨及び審査結果の要旨」の原稿としてオフセット印刷するので，鮮明な原稿をクリップ止めで提出すること。