

学位論文 博士 (工学)

**多様な情報源に対するランキング学習
に関する研究**

2014 年度

慶應義塾大学大学院理工学研究科

数原 良彦

主 論 文 要 旨

報告番号	甲 乙 第	号	氏 名	数原 良彦
主 論 文 題 目： 多様な情報源に対するランキング学習に関する研究				
(内容の要旨)				
ユーザが望む順序で結果を提示するランキングは様々なアプリケーションにおいて重要である。特に情報検索や推薦においては、ユーザが求める情報を優先的に上位に表示することにより、より少ないコストでユーザが目的を達成する。このように、情報処理の広い分野において、同時には提示不可能な大量の候補群から、ユーザの希望に合うアイテムに優先順位をつけて提示する方法の基盤技術としてランキング技術が用いられてきた。近年では、教師あり学習の枠組みを用いて最適なランキングモデルを構築するランキング学習と呼ばれる技術分野が情報検索コミュニティで盛んに研究されてきた。ランキング学習においては、あらかじめ評価ラベルが付与された訓練事例集合をもとに予測モデルを構築し、未知の事例集合に対して評価ラベルと整合する順位を予測する。本研究では、既存のランキング学習が抱える(1) 複数情報源に対する適用、(2) 大規模データに対する適用、(3) 学習データに含まれるノイズへの対応、(4) 少量の特徴のみ利用可能なデータへの適用という4つの課題を達成し、多様な情報源に対する実用的なランキング学習実現を目指す。 以下に本論文の構成を示す。はじめに第1 章において本研究の背景と目的について述べる。第2 章では、ランキング学習とその関連研究について述べる。第3 章では、人手評価データとクリックログデータという適合性分布が異なる複数情報源を用いる、ブースティングに基づくランキング学習アルゴリズムであるTRankBoostを提案し、実データを用いた評価実験の結果を述べる。第4 章では、検索評価指標のマージンへの導入、ノイズに頑健な損失関数の採用、最大損失を与える事例の選択といった特長を持つ選択的ペアワイズ手法によるオンラインランキング学習であるPARank-NDCG を提案し、従来のオンラインランキング学習に対して優れた性能を示すことについて述べる。第5 章では、ランキングに用いる多数の要因を抽出できない状況において高精度なランキングを実現するため、自己組織化マップを教師あり順序学習に拡張したOrderSOM を提案し、その評価について述べる。				

SUMMARY OF Ph.D. DISSERTATION

School School of Science for Open and Environmental Systems	Student Identification Number	SURNAME, First name SUHARA, Yoshihiko
Title A Study on Learning to Rank for Various Information Sources		
Abstract Ranking plays an important role in many applications. In search systems or recommender systems, a user satisfies his/her goal with lower cost by displaying appropriate contents in higher positions. Ranking methods have been used as a core technique to achieve a good search result, which are selected from a massive amount of search result candidates. Recently, learning to rank, which generates a ranking model based on a supervised learning framework, has been studied among information retrieval communities. A learning-to-rank algorithm generates a ranking model based on training data. The ranking model outputs a ranking list from unsorted documents. This paper tackles following four challenges to achieve a practical learning-to-rank algorithm: (1) Learning to rank from multiple sources. (2) Learning to rank from large-scale datasets. (3) Learning to rank from noisy training data. (4) Learning to rank based on small feature sets. This paper is organized as follows: Chapter 1 describes backgrounds and the purpose of this research. Chapter 2 introduces a learning-to-rank problem and related work. Chapter 3 describes a novel learning-to-rank algorithm called TRankBoost, which uses multiple sources that consist of different relevance distributions such as training data made from human assessors and that of click-through logs. Chapter 4 describes a novel online learning-to-rank algorithm called PARank-NDCG, which introduces information evaluation measure into margin size, ramp loss function for robust learning, and max-loss update strategy. Chapter 5 describes OrderSOM, which is an extension of self-organizing maps, to achieve high-accuracy rankings with a few features. Chapter 6 concludes the paper with a summary of future issues.		