

Ranking data with ordinal labels: optimality and pairwise aggregation

Stéphan Cléménçon · Sylvain Robbiano ·
Nicolas Vayatis

Received: 8 March 2011 / Accepted: 24 September 2012 / Published online: 12 December 2012
© The Author(s) 2012

Abstract The paper describes key insights in order to grasp the nature of K -partite ranking. From the theoretical side, the various characterizations of optimal elements are fully described, as well as the *likelihood ratio monotonicity* condition on the underlying distribution which guarantees that such elements do exist. Then, a pairwise aggregation procedure based on Kendall tau is introduced to relate learning rules dedicated to bipartite ranking and solutions of the K -partite ranking problem. Criteria reflecting ranking performance under these conditions such as the ROC surface and its natural summary, the volume under the ROC surface (VUS), are then considered as targets for empirical optimization. The consistency of pairwise aggregation strategies are studied under these criteria and shown to be efficient under reasonable assumptions. Eventually, numerical results illustrate the relevance of the methodology proposed.

Keywords K -partite ranking · Ordinal data · ROC surface · Volume under the ROC surface · Empirical risk minimization · Median ranking

1 Introduction

In many situations, a natural ordering can be considered over a set of observations. When observations are documents in information retrieval applications, the ordering reflects degree of relevance for a specific query. In order to predict future ordering on new data, the learning process uses past data for which some relevance feedback is some provided,

Editor: Johannes Fürnkranz.

S. Cléménçon · S. Robbiano (✉)
LTCI UMR Telecom ParisTech/CNRS No. 5141, Telecom ParisTech, Paris cedex 13, 75634, France
e-mail: robbiano@telecom-paristech.fr

S. Cléménçon
e-mail: stephan.clemencon@telecom-paristech.fr

N. Vayatis
CMLA UMR CNRS No. 8536, ENS Cachan & UniverSud, Cahan cedex, 94235, France