

# Supervised clustering of label ranking data using label preference information

Mihajlo Grbovic · Nemanja Djuric · Shengbo Guo ·  
Slobodan Vucetic

Received: 10 January 2012 / Accepted: 29 April 2013 / Published online: 7 June 2013  
© The Author(s) 2013

**Abstract** This paper studies supervised clustering in the context of label ranking data. The goal is to partition the feature space into  $K$  clusters, such that they are compact in both the feature and label ranking space. This type of clustering has many potential applications. For example, in target marketing we might want to come up with  $K$  different offers or marketing strategies for our target audience. Thus, we aim at clustering the customers' feature space into  $K$  clusters by leveraging the revealed or stated, potentially incomplete customer preferences over products, such that the preferences of customers within one cluster are more similar to each other than to those of customers in other clusters. We establish several baseline algorithms and propose two principled algorithms for supervised clustering. In the first baseline, the clusters are created in an unsupervised manner, followed by assigning a representative label ranking to each cluster. In the second baseline, the label ranking space is clustered first, followed by partitioning the feature space based on the central rankings. In the third baseline, clustering is applied on a new feature space consisting of both features and label rankings, followed by mapping back to the original feature and ranking space. The RankTree principled approach is based on a Ranking Tree algorithm previously proposed for label ranking prediction. Our modification starts with  $K$  random label rankings and iteratively splits the feature space to minimize the ranking loss, followed by re-calculation of the  $K$  rankings based on cluster assignments. The MM-PL approach is a

---

Editors: Eyke Hüllermeier and Johannes Fürnkranz.

M. Grbovic (✉) · N. Djuric · S. Vucetic  
Department of Computer and Information Sciences, Center for Data Analytics and Biomedical Informatics, Temple University, Philadelphia, PA 19122, USA  
e-mail: [mihajlo.grbovic@temple.edu](mailto:mihajlo.grbovic@temple.edu)

N. Djuric  
e-mail: [nemanja.djuric@temple.edu](mailto:nemanja.djuric@temple.edu)

S. Vucetic  
e-mail: [slobodan.vucetic@temple.edu](mailto:slobodan.vucetic@temple.edu)

S. Guo  
Xerox Research Centre Europe, 6 chemin de Maupertuis, 38240 Meylan, France  
e-mail: [shengbo.guo@xrce.xerox.com](mailto:shengbo.guo@xrce.xerox.com)