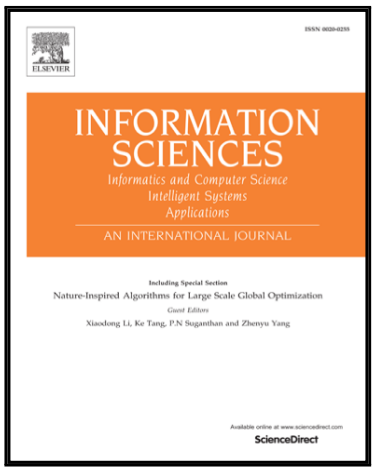

Semi-supervised trees for multi-target regression

Jurica Levatić, Dragi Kocev, Michelangelo Ceci, Sašo Džeroski

PII: S0020-0255(18)30210-X
DOI: 10.1016/j.ins.2018.03.033
Reference: INS 13511

To appear in: *Information Sciences*

Received date: 13 May 2016
Revised date: 4 January 2018
Accepted date: 11 March 2018

Please cite this article as: Jurica Levatić, Dragi Kocev, Michelangelo Ceci, Sašo Džeroski, Semi-supervised trees for multi-target regression, *Information Sciences* (2018), doi: 10.1016/j.ins.2018.03.033

Semi-supervised trees for multi-target regression

Jurica Levatić[a,b], Dragi Kocev[a,b,c], Michelangelo Ceci[c,d], Sašo Džeroski[a,b]

[a] *Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia*
[b] *Jožef Stefan International Postgraduate School, Ljubljana, Slovenia*
[c] *Department of Computer Science, University of Bari Aldo Moro, Bari, Italy*
[d] *CINI, Consorzio Interuniversitario Nazionale per l'Informatica, Italy*

Abstract

The predictive performance of traditional supervised methods heavily depends on the amount of labeled data. However, obtaining labels is a difficult process in many real-life tasks, and only a small amount of labeled data is typically available for model learning. As an answer to this problem, the concept of semi-supervised learning has emerged. Semi-supervised methods use unlabeled data in addition to labeled data to improve the performance of supervised methods.

It is even more difficult to get labeled data for data mining problems with structured outputs since several labels need to be determined for each example. Multi-target regression (MTR) is one type of a structured output prediction problem, where we need to simultaneously predict multiple continuous variables. Despite the apparent need for semi-supervised methods able to deal with MTR, only a few such methods are available and even those are difficult to use in practice and/or their advantages over supervised methods for MTR are not clear.

This paper presents an extension of predictive clustering trees for MTR and ensembles thereof towards semi-supervised learning. The proposed method preserves the appealing characteristic of decision trees while enabling the use of unlabeled examples. In particular, the proposed semi-supervised trees for MTR are interpretable, easy to understand, fast to learn, and can handle both numeric and nominal descriptive features. We perform an extensive empirical evaluation in both an inductive and a transductive semi-supervised setting. The results show that the proposed method improves the performance of su-

*Corresponding author
Email address: jurica.levatic@ijs.si (Jurica Levatić)