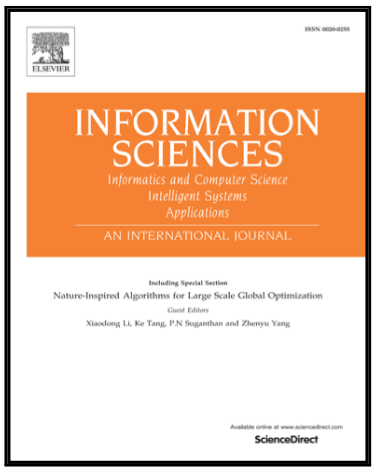
Accepted Manuscript



A Tree Ensemble-Based Two-Stage Model for Advanced-Stage Colorectal Cancer Survival Prediction

Yuyan Wang , Dujuan Wang , Xin Ye , Yanzhang Wang , Yunqiang Yin , Yaochu Jin

PII: S0020-0255(18)30759-X
DOI: https://doi.org/10.1016/j.ins.2018.09.046
Reference: INS 13957

To appear in: *Information Sciences*

Received date: 15 February 2018
Revised date: 13 September 2018
Accepted date: 22 September 2018

Please cite this article as: Yuyan Wang , Dujuan Wang , Xin Ye , Yanzhang Wang , Yunqiang Yin , Yaochu Jin , A Tree Ensemble-Based Two-Stage Model for Advanced-Stage Colorectal Cancer Survival Prediction, *Information Sciences* (2018), doi: https://doi.org/10.1016/j.ins.2018.09.046

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

A Tree Ensemble-Based Two-Stage Model for Advanced-Stage Colorectal Cancer Survival Prediction[1]

Yuyan Wang [a], Dujuan Wang [a,*], Xin Ye [a], Yanzhang Wang [a], Yunqiang Yin [b], Yaochu Jin [a,c,*]

[a] *School of Management Science and Engineering, Dalian University of Technology, Dalian, 116023, PR China*

[b] *Data Science Research Center, Kunming University of Science and Technology, Kunming, 650093, China*

[c] *Department of Computer Science, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom*

Abstract

Classification techniques have widely been applied to cancer survival prediction for predicting survival or death of patients. However, little attention has been paid to patients who are predicted to die. In this work, we consider survival prediction as a two-stage task, where the first stage is to predict whether the outcome is survival or not, and the second stage is to predict the remaining lifespan for patients whose predicted outcome is death. To this end, we propose a two-stage machine learning model to enhance cancer survival prediction. At the first stage, a tree-based imbalanced ensemble classification method is proposed for classification of the survivability of advanced-stage cancer patients. At the second stage, a selective ensemble regression method is proposed for survival time prediction, where *a priori* knowledge is adopted for feature selection and the mean proportion of error interval is proposed for selecting base learners. Extensive computational studies performed on colorectal cancer data from SEER database demonstrate that the proposed two-stage model can achieve a more accurate prediction compared to the one-stage regression model. The results show that the proposed classification approach can effectively handle the imbalanced survivability data, and the proposed regression method outperforms several state-of-the-art regression models.

Key words: Colorectal cancer; survival prediction; two-stage model; ensemble learning; semi-random regression tree

* Corresponding author.

E-mail addresses: wangdujuan@dlut.edu.cn (D. Wang); yaochu.jin@surrey.ac.uk (Y. Jin).

Download English Version:

https://daneshyari.com/en/article/11021168

Download Persian Version:

https://daneshyari.com/article/11021168

Daneshyari.com