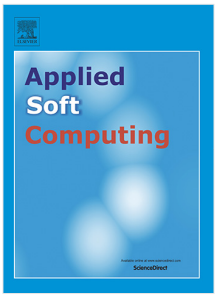
Accepted Manuscript



Improving performance of classification on incomplete data using feature selection and clustering

Cao Truong Tran, Mengjie Zhang, Peter Andreae, Bing Xue, Lam Thu Bui

PII: S1568-4946(18)30543-X
DOI: https://doi.org/10.1016/j.asoc.2018.09.026
Reference: ASOC 5105

To appear in: *Applied Soft Computing Journal*

Received date : 19 June 2017
Revised date : 20 September 2018
Accepted date : 22 September 2018

Please cite this article as: C.T. Tran, et al., Improving performance of classification on incomplete data using feature selection and clustering, *Applied Soft Computing Journal* (2018), https://doi.org/10.1016/j.asoc.2018.09.026

***Highlights (for review)**

Highlights

- We propose the combination of imputation and clustering which can crucially reduce the computation time, and can achieve comparable accuracy to using only imputation. The key reason is that clustering can produce a smaller set of representative training data to perform imputation in the application process.
- We propose the combination of imputation and feature selection which also can remarkably reduce the computation time, and can achieve better accuracy than using only imputation. The underlying reason is that feature selection can generate better and smaller training data, and reduce the number of incomplete instances in the testing process
- We investigate the combination of imputation, feature selection and clustering which not only can further speed up imputation, but also can improve classification accuracy, simultaneously.