World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:15, No:12, 2021

Self-Organizing Maps for Exploration of Partially Observed Data and Imputation of Missing Values in the Context of the Manufacture of Aircraft Engines

Authors : Sara Rejeb, Catherine Duveau, Tabea Rebafka

Abstract : To monitor the production process of turbofan aircraft engines, multiple measurements of various geometrical parameters are systematically recorded on manufactured parts. Engine parts are subject to extremely high standards as they can impact the performance of the engine. Therefore, it is essential to analyze these databases to better understand the influence of the different parameters on the engine's performance. Self-organizing maps are unsupervised neural networks which achieve two tasks simultaneously: they visualize high-dimensional data by projection onto a 2-dimensional map and provide clustering of the data. This technique has become very popular for data exploration since it provides easily interpretable results and a meaningful global view of the data. As such, self-organizing maps are usually applied to aircraft engine condition monitoring. As databases in this field are huge and complex, they naturally contain multiple missing entries for various reasons. The classical Kohonen algorithm to compute self-organizing maps is conceived for complete data only. A naive approach to deal with partially observed data consists in deleting items or variables with missing entries. However, this requires a sufficient number of complete individuals to be fairly representative of the population; otherwise, deletion leads to a considerable loss of information. Moreover, deletion can also induce bias in the analysis results. Alternatively, one can first apply a common imputation method to create a complete dataset and then apply the Kohonen algorithm. However, the choice of the imputation method may have a strong impact on the resulting self-organizing map. Our approach is to address simultaneously the two problems of computing a self-organizing map and imputing missing values, as these tasks are not independent. In this work, we propose an extension of self-organizing maps for partially observed data, referred to as missSOM. First, we introduce a criterion to be optimized, that aims at defining simultaneously the best self-organizing map and the best imputations for the missing entries. As such, missSOM is also an imputation method for missing values. To minimize the criterion, we propose an iterative algorithm that alternates the learning of a self-organizing map and the imputation of missing values. Moreover, we develop an accelerated version of the algorithm by entwining the iterations of the Kohonen algorithm with the updates of the imputed values. This method is efficiently implemented in R and will soon be released on CRAN. Compared to the standard Kohonen algorithm, it does not come with any additional cost in terms of computing time. Numerical experiments illustrate that missSOM performs well in terms of both clustering and imputation compared to the state of the art. In particular, it turns out that missSOM is robust to the missingness mechanism, which is in contrast to many imputation methods that are appropriate for only a single mechanism. This is an important property of missSOM as, in practice, the missingness mechanism is often unknown. An application to measurements on one type of part is also provided and shows the practical interest of missSOM.

Keywords : imputation method of missing data, partially observed data, robustness to missingness mechanism, self-organizing maps
Conference Title : ICAMAS 2021 : International Conference on Applied Mathematics, Analysis and Statistics
Conference Location : Amsterdam, Netherlands
Conference Dates : December 02-03, 2021