

Missing value invariant data mining

Although data mining algorithms have been improved extensively over the last two decades, incomplete data is still hard to process by traditional anomaly detection techniques. The most straightforward approach, to overcome missing values, is leaving out records listwise in case, a single value is missing. However, if the dimensionality of the data increases, the probability of a row to be full will decrease exponentially. Another approach is to impute missing values before applying the algorithm. This, however, can be costly and adds a bias.

The aim of this project is to develop novel or adapt data mining algorithms (Clustering, Outlier Detection, Classification, Regression, Feature Selection), which overcome the problem of missing data. This requires some knowledge of statistics, machine learning and programming, so a candidate should feel comfortable in these areas. For writing a thesis, a candidate can either bring his/her own idea, or we will provide you one.

Contact

Email: thomas.goerttler@hpi.de

Web: <https://hpi.de/mueller/personen/thomas-goerttler.html>