LEUENBERGER Michael (2018) : Environmental data mining using machine learning algorithms: methodological developments and case studies

Abstract

Due to the large amount and complexity of data available nowadays in geo- and environmental sciences, we face the need to develop and incorporate more robust and efficient methods for their analysis, modelling and visualization. An important part of these developments deals with an elaboration and application of a contemporary and coherent methodology following the process from data collection to the justification and communication of the results. Recent fundamental progress in machine learning can considerably contribute to the development of this emerging field - environmental data science.

The main purpose of this Thesis is to develop coherent and self-consistent methodologies for the analysis of environmental phenomenon using machine learning algorithms. In particular, this Thesis gives an overview of machine learning algorithms for environmental data mining. It highlights and investigates the different issues that can occur when dealing with complex and high dimensional environmental data using cutting-edge machine learning algorithms. In addition, several important topics of data driven modelling, including data splitting, complexity analysis, residuals assessment, feature selection and uncertainties are discussed.

Moreover, a special attention is paid to the Extreme Learning Machine algorithm (ELM). Being an efficient artificial neural network, it gained recently a great popularity in the domain of machine learning. By taking advantage of its quickness, another objective of this Thesis is to extract the potential of ELM for the tasks of feature selection and uncertainty quantification. Both of these approaches can give valuable information about the hidden relationship between input and output variables, which indirectly reflects the behaviour of the studied phenomenon.

In this regard, the general leitmotif of this Thesis is focused on the development of coherent methodologies for the analysis of environmental phenomenon using machine learning algorithms. The applied part of the research deals with an application of the methodology and the developed methods for simulated, modelled and real environmental data, such as forest fires, pollution and wind fields.