

Tutorial

“Variable selection for efficient design of neural networks and other machine learning-based models: efficient approaches for industrial applications”

Valentina Colla, Pd.D., Silvia Cateni M.Sc.,
Scuola Superiore Sant'Anna, Pisa, Italy
colla@sssup.it, mvannucci@sssup.it

In many real word applications of neural networks and other machine learning approaches, large experimental datasets are available, containing a huge number of variables whose effect on the considered system or phenomenon to analyse and/or reproduce, such as some numerical variables to forecast or a classification outcome, in not completely known or not deeply understood.

Variable selection procedures identify a small subset from original feature space in order to point out the input variables, which mainly affect the considered target. The identification of such variables subset leads to very important advantages, such as lower complexity of the model and of the learning algorithm, savings of computational time and improved performance. Moreover, variable selection procedures can help to acquire a deeper knowledge of the considered problem, system or phenomenon can be achieved by identifying the factors which mostly affect it. This concept is strictly linked to the crucial aspect of the *stability* of the variable selection, defined as the sensitivity of a machine learning model with respect to variations in the dataset that is exploited in its training phase.

In the tutorial, different categories of variable section procedures will be presented and discussed, by also proposing exemplar case studies coming mainly from the industrial field, in order to highlight strengths and weaknesses of each method in relation to the different tasks and to the variables of the considered dataset. Finally some open challenges and future directions in the research on this field will be outlined.

Valentina Colla got her Master Degree in Telecommunication Engineer at the University of Pisa in 1994 and a PhD at Scuola Superiore San'Anna (SSSA) in Robotics in 1998. She is currently Technical Research Manager at SSSA and she is the responsible of the Center of Information and Communication Technologies for Complex Industrial Systems and Processes (ICT-COISP) of the

TeCIP Institute of SSSA. Her research fields include standard and Artificial Intelligence-based data processing, data mining and machine learning tools and techniques. She is deeply involved in research activities related to modeling, simulation, optimization and control of industrial processes, with a particular focus on manufacturing industry and process industry. She is also active in the field of simulation of complex industrial processes and application of multi-objective optimization techniques aimed at improving resource efficiency and reducing production costs and environmental impact of process industries. She is co-author of about 240 papers in International Journals and Conferences.

She has a considerable experience in the process industry and in manufacturing fields. She has been involved in more than 45 EU funded projects and in many projects supported by industries. She is coordinator of 3 projects supported by the EU through the Research Fund for Coal & Steel.

Silvia Catenigot her master degree in telecommunication engineering, in 2005, from Pisa University. Her research activity includes mathematical modelling and data analysis through statistical and artificial intelligence-based techniques. She participated to several research projects particularly dealing with steelmaking industry. She is presently a research assistant at TeCIP Institute of SSSA.