Ensemble Learning for Multi-source Information Fusion

Jörg Beyer, Kai Heesche, Werner Hauptmann, Clemens Otte, and Rudolf Kruse

Abstract. In this chapter, we propose a new ensemble learning method. The main objective of this approach is to jointly use data-driven and knowledge-based submodels, like mathematical equations or rules, in the modeling process. The integration of knowledge-based submodels is of particular interest, since they are able to provide with information not contained in the data. On the other hand, data-driven models can complement the knowledge-based models with respect to input space coverage. For the task of appropriately integrating the different models, a method for partitioning the input space for the given models is introduced. Using that kind of ensembles, the advantages of both models are combined, i.e., robustness and physical transparency of the knowledge-based models and approximation abilities of the data-driven learning. The benefits of this approach are demonstrated for a real-world application.

Jörg Beyer
Siemens AG - CT IC 4, Otto-Hahn-Ring 6, 80200 Munich, Germany
Otto-von-Guericke-University Magdeburg - School of Computer Science,
Universitätsplatz 2, 39106 Magdeburg, Germany
e-mail: joerg.beyer.ext@siemens.com

Kai Heesche
Siemens AG - CT IC 4, Otto-Hahn-Ring 6, 80200 Munich, Germany

Werner Hauptmann
Siemens AG - CT IC 4, Otto-Hahn-Ring 6, 80200 Munich, Germany

Clemens Otte
Siemens AG - CT IC 4, Otto-Hahn-Ring 6, 80200 Munich, Germany

Rudolf Kruse
Otto-von-Guericke-University Magdeburg - School of Computer Science,
Universitätsplatz 2, 39106 Magdeburg, Germany
1 Introduction

Modern technical systems are characterized by an increasing degree of sophisticated behavior. The traditional way of modeling has been by mathematical equations representing the physical behavior. However, the identification of parameters is time-consuming and expensive. Data-driven models, like artificial neural networks, can be used to approximate physical phenomena. But, the data-driven modeling approach usually suffers from the lack of physical understanding of the model parameters. The resulting model only relies on the training data and does not use any other information source available. Thus, it is desirable to combine available information in terms of knowledge-based models, i.e., models which are based on domain or process knowledge, designed without training data and to complement this information with the data-driven approach.

The integration of knowledge-based submodels has several advantages:

- enhancing the interpretability, i.e., a domain expert can easily comprehend the decisions,
- providing information not contained in the training data and
- reducing the amount of required training data.

For these reasons, an important factor for the generation of adequate models of a technical system is the use of available information in terms of knowledge-based models and the supplementation of this information by data-driven models learnt on the training data. Since a knowledge-based model represents a particular subsystem, information with respect to its validity has to be included in the overall model.

The objective of the proposed approach is to generate an ensemble that is able to integrate the available knowledge-based submodels and to complement these submodels by data-driven ones. Using that kind of ensembles the advantages of both models are combined, i.e., the robustness and physical transparency of the knowledge-based models and the approximation abilities of data-driven learning.

The use of multiple models is also motivated by the paradigm that different partial models can complement each other by appropriate compensation of weaknesses and strengths of the individual models. Much of the work on ensemble techniques has strong parallels with the research on information fusion (IF) systems. In common with the research on IF, several architectures exist and different combination schemes have been developed. Later in this chapter, we give a review of IF.

The chapter is organized as follows: In Section 2, an introduction of IF is given and Section 3 describes different methods for creating ensembles. In Section 4, two ensemble models for combining data-driven and knowledge-based models are proposed. In Section 5, some experiments on a real-world application are outlined. Section 6 concludes the study.