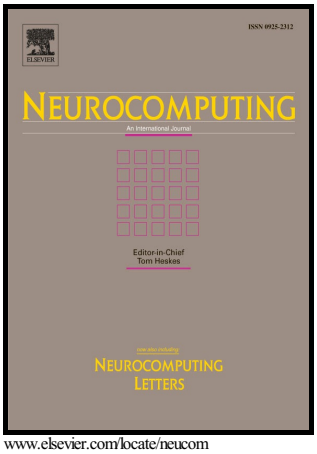

www.elsevier.com/locate/neucom

Semi-supervised Selective Ensemble Learning Based On Distance to Model for Nonlinear Soft Sensor Development

Weiming Shao, Xuemin. Tian

PII: S0925-2312(16)31167-5
DOI: http://dx.doi.org/10.1016/j.neucom.2016.10.005
Reference: NEUCOM17617

To appear in: *Neurocomputing*

Received date: 28 July 2015
Revised date: 26 September 2016
Accepted date: 7 October 2016

Cite this article as: Weiming Shao and Xuemin. Tian, Semi-supervised Selective Ensemble Learning Based On Distance to Model for Nonlinear Soft Senso Development, *Neurocomputing,* http://dx.doi.org/10.1016/j.neucom.2016.10.005

Semi-supervised Selective Ensemble Learning Based On Distance to Model for Nonlinear Soft Sensor Development

Weiming Shao, Xuemin Tian*,

College of Information and Control Engineering, China University of Petroleum, 66#, Changjiang West Road, Huangdao District, Qingdao, 266580, China

swmupc@163.com (W.Shao),

tianxm@upc.edu.cn (X. Tian)

*Corresponding author. Tel.: +86 532 86983467.

Abstract:

Data-driven soft sensors have been widely used in process systems for delivering online estimations of hard-to-measure yet important quality-related variables. However, in many data-driven soft sensor applications, the process may be strongly nonlinear, and the number of labeled samples is limited, which are two major difficulties in developing high-accuracy soft sensors. To cope with this issue, a novel soft sensing method based on semi-supervised selective ensemble learning strategy is proposed. In the proposed method, the process is initially localized by an adaptive process state partition approach, where the information of unlabeled samples can be incorporated. Subsequently, a new distance to model (DM) criterion is defined for selective ensemble learning, which can overcome the drawback of the *k* nearest neighbor method. The newly defined DM criterion along with the incorporation of unlabeled samples can help to describe the relationships between query samples and local models more accurately, and therefore is able to provide higher estimation accuracy. The parameters of the proposed method are finalized automatically by the particle swarm optimization technique. The proposed method is investigated using three real-life benchmark datasets, and the simulation results demonstrate the effectiveness of the proposed method in dealing with nonlinear regression problems in the process system and in the urban pollutant monitoring area.

Keywords: nonlinear soft sensor, semi-supervised ensemble learning, selective ensemble learning, distance to model, adaptive process state partition, particle swarm optimization.

1. Introduction

In process industries, a large amount of hardware sensors are installed to provide data for the purpose of monitoring and control [1]. However, many quality-related key variables can not be measured online by these

.