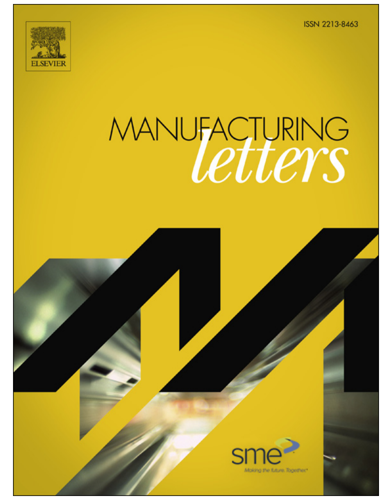

Process-Monitoring-for-Quality — A Model Selection Criterion

Carlos A. Escobar, Ruben Morales-Menendez

PII: S2213-8463(18)30004-X
DOI: https://doi.org/10.1016/j.mfglet.2018.01.001
Reference: MFGLET 124

To appear in: *Manufacturing Letters*

Received Date: 11 September 2017
Accepted Date: 10 January 2018

Process-Monitoring-for-Quality — A Model Selection Criterion

Carlos A. Escobar[†,‡,*], Ruben Morales-Menendez[‡]
[†] Global Research & Development, General Motors, Warren, MI, USA
[‡] Tecnológico de Monterrey, México

Thursday 11th January, 2018

Abstract

The new big data driven manufacturing quality philosophy, *Process Monitoring for Quality (PMQ)*, proposes Big Data — Big Models, a new modeling paradigm that includes a big data-driven learning process that requires many models to be created to find the final one. Since many candidates are created, the main challenges is to select the model that efficiently solves the tradeoff between complexity and prediction. Most mature manufacturing organizations generate only a few *Defects Per Million of Opportunities (DPMO)*; therefore, manufacturing-derived data sets for classification of quality tend to be highly unbalanced. The *Penalized Maximum Probability of Correct Decision (PMPCD)* is developed to solve the posed tradeoff. According to simulation and experimental results, the model selection criterion induces parsimony by selecting the model with the minimum number of features needed for an effective/efficient defect detection.

Keywords: Model selection criterion, Process monitoring for quality, Binary classification, Highly unbalanced data structures, Big data driven manufacturing.

1 Introduction

Manufacturing companies are intense users of *big data* bases, this industry generates/stores more data than any other [1]. *Machine Learning (ML)* techniques are applied to improve process monitoring, control and part quality [2–4]. Classification of quality is one of the most important applications [5, 6], where relevant characteristics of the process/product are related to an ordinal or binary output aimed at detecting defects [7, 8]. Most mature manufacturing organizations generate only a few *DPMO*; then, manufacturing-derived data sets for binary classification of quality are highly unbalanced.

The new big data driven manufacturing quality philosophy, *PMQ*, proposes Big Data – Big Models, a new modeling paradigm that includes a big data-driven learning process that often requires many models (classifiers) to be created to find the final one. Selecting the *best* single model is one of the most important challenges [7]. *Model Selection (MS)* methods are based on the principle of parsimony [9, 10]. Modeling and *MS* are essentially concerned with the art of approximation. Therefore, the *best* model can be defined as the one that efficiently solves the tradeoff between complexity – number of features – and prediction ability. In contrast with other industries, where prediction is the main goal, model interpretation is very important, since the learned characteristics of the system can be used to plan/design experiments to find optimal levels of process/product parameters, this situation highlights the importance of parsimonious modeling.

*All communications must be addressed to Carlos A. Escobar, carlos.1.escobar@gm.com.