Applied Soft Computing 59 (2017) 77–87

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A study of overfitting in optimization of a manufacturing quality control procedure

Tea Tušar[a,*], Klemen Gantar[b], Valentin Koblar[c,d], Bernard Ženko[e], Bogdan Filipič[a,d]

[a] Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia
[b] Faculty of Computer and Information Science, University of Ljubljana, Slovenia
[c] Kolektor Group d.o.o., Idrija, Slovenia
[d] Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
[e] Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 18 January 2017
Received in revised form 24 April 2017
Accepted 15 May 2017
Available online 22 May 2017

Keywords:
Quality control
Machine vision
Machine learning
Optimization
Overfitting

ABSTRACT

Quality control of the commutator manufacturing process can be automated by means of a machine learning model that can predict the quality of commutators as they are being manufactured. Such a model can be constructed by combining machine vision, machine learning and evolutionary optimization techniques. In this procedure, optimization is used to minimize the model error, which is estimated using single cross-validation. This work exposes the overfitting that emerges in such optimization. Overfitting is shown for three machine learning methods with different sensitivity to it (trees, additionally pruned trees and random forests) and assessed in two ways (repeated cross-validation and validation on a set of unseen instances). Results on two distinct quality control problems show that optimization amplifies overfitting, i.e., the single cross-validation error estimate for the optimized models is overly optimistic. Nevertheless, minimization of the error estimate by single cross-validation in general results in minimization of the other error estimates as well, showing that optimization is indeed beneficial in this context.

1. Introduction

Quality control is essential for improving any manufacturing process. It encourages quality consciousness among workers, enables a more efficient utilization of resources and results in products of better quality at reduced production costs. It is especially crucial in processes with high quality requirements as is the case in automotive industry. There, in many cases only one part per million of supplied products is allowed be defective, which yields strict demands for the involved manufacturing processes as well as their quality control procedures.

This work is concerned with quality control of the manufacturing of graphite commutators (components of electric motors used in automotive fuel pumps) produced at an industrial production plant. More specifically, two different phases of the graphite commutator production process are considered. The first is the *soldering phase*, which consists of soldering the metalized graphite to the commutator copper base. The quality of the resulting copper-graphite joints is crucial since the reliability of end user applications depends on the strength of these joints. The second is the *turning phase* where the commutator mounting hole is formed. The diameter and roughness of the hole directly influence the force required to mount the commutator on the rotor shaft. The minimum and maximum force that can be used in the mounting operation are specified by the customer, which in turn defines the feasible values for the hole diameter and roughness.

Currently, the quality control for both phases is done manually. Automated on-line quality control would bring several advantages over manual inspection. For example, it could promptly detect irregularities making error resolution faster and consequently saving a considerable amount of resources. Moreover, it would not slow down the production line and would be cheaper than manual inspection. Finally, it would not suffer from fatigue and other human factors that can result in errors. This is why we aim for an automated on-line quality control procedure capable of determining:

- whether the joints are soldered well or have any of the known defects (the *soldering problem*), and
- the mounting hole roughness (the *roughness problem*).

* Corresponding author.
E-mail address: tea.tusar@ijs.si (T. Tušar).

http://dx.doi.org/10.1016/j.asoc.2017.05.027
1568-4946/

Measuring the mounting hole diameter is considered trivial and will be exempt from this study.

Such automation can be implemented on the production line by combining machine vision (MV), machine learning (ML) and optimization methods. It consists of assessing the quality of the soldered joints and mounting hole roughness from commutator images captured by a camera. Predictions are made by a ML model that needs to be previously trained on a database of commutator images with known soldered joints quality and mounting hole roughness. Attributes used by machine learning are extracted from the commutator images with MV methods. Most MV methods have parameters that greatly affect their outcome and are at the same time hard to set. This is why optimization is used to find the MV parameter settings that result in a ML model with a low error rate.

While previous work studied different setups for this automated quality control procedure (see [1–3] for the soldering problem and [4] for the roughness problem), this paper exposes the overfitting that emerges when searching for an accurate predictive model. Overfitting is shown for three ML methods with different sensitivity to it (trees, additionally pruned trees and random forests) and assessed in two ways (repeated cross-validation and validation on a set of unseen instances). The original contribution of the paper is the investigation of the effect of overfitting in such a procedure. We wish to test whether optimization can be beneficial despite using an overly optimistic error estimate. This work is an extended version of the initial overfitting study from [5] that included only the soldering problem, was limited to decision trees and used only repeated cross-validation to assess overfitting.

The automated quality control procedure is explained in more detail in Section 2, while Section 3 discusses overfitting. Afterward, Section 4 presents previous work in this domain and other related work. Performed experiments and their results are detailed in Section 5. Finally, Section 6 summarizes the paper.

2. The automated quality control procedure

This section starts with a general overview of the proposed automated quality control procedure followed by more details for the two separate problems.

2.1. Overview

The goal of this automated quality control procedure is to accurately predict the quality of the soldered joints (or mounting hole roughness) from images of the commutator. This entails the following three steps (see Fig. 1):

1. Preprocess the original image (adjust its position, extract the regions of interest etc., see Sections 2.2 and 2.3 for details).
2. Extract image attributes from the preprocessed image using machine vision with the given settings.
3. Predict the quality of soldered joints or mounting hole roughness from image attributes using the given ML model.

The inputs to this prediction are, beside the original image, the settings to MV methods and the ML model. These are retrieved from the optimization procedure presented in Fig. 2.

The goal of optimization is to find the MV settings that yield the best ML model (the one with the lowest prediction error, which is estimated on a set of original images). In optimization terminology, the MV settings represent one *solution* to this optimization problem, while the ML model error is the objective function to be minimized. The optimization procedure starts by preprocessing the whole set of original images. Then, an evolutionary optimization algorithm is used to search for the best MV settings. It starts with

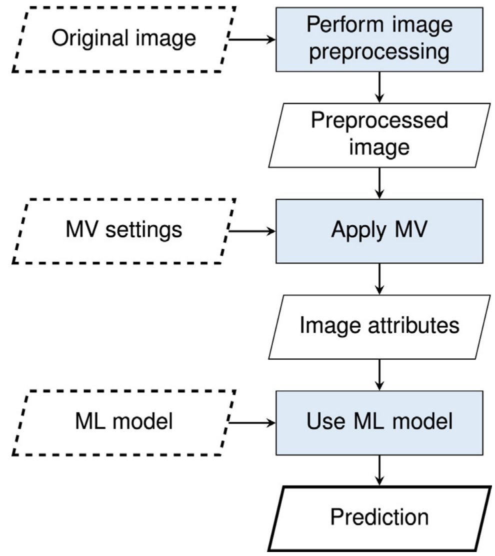

Fig. 1. The procedure for predicting the quality of soldered joints (or mounting hole roughness) from an image of the commutator using the given MV settings and ML model.

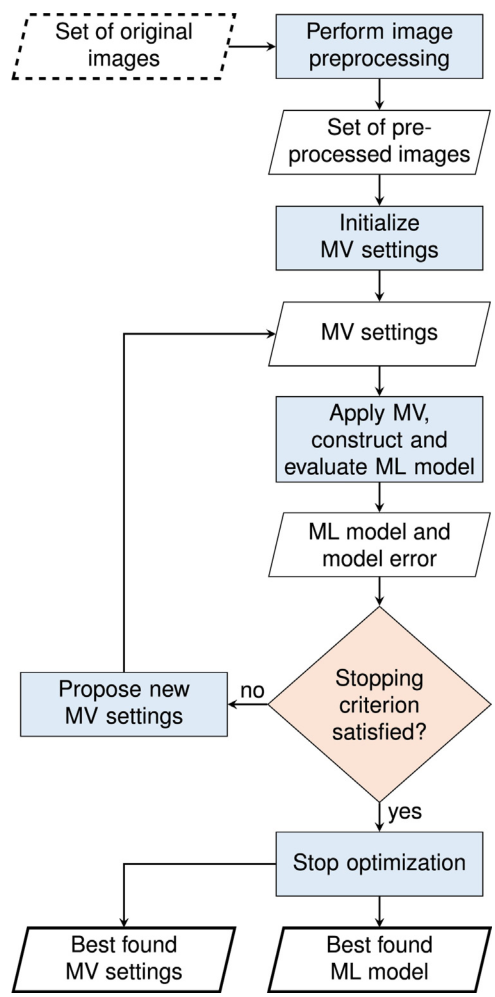

Fig. 2. The optimization procedure that searches for the MV settings that result in the best ML model (the one with the lowest error on the given set of original images).