Engineering Applications of Artificial Intelligence 68 (2018) 131–151

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Self-adaptive evolving forecast models with incremental PLS space updating for on-line prediction of micro-fluidic chip quality

Edwin Lughofer [a,*], Robert Pollak [a], Alexandru-Ciprian Zavoianu [a], Mahardhika Pratama [b], Pauline Meyer-Heye [c], Helmut Zörrer [c], Christian Eitzinger [c], Julia Haim [d], Thomas Radauer [d]

[a] Department of Knowledge-Based Mathematical Systems, Johannes Kepler University Linz, Austria
[b] School of Computer Science and Engineering, Nanyang Technological University, Singapore
[c] Profactor GmbH, Steyr-Gleink, Austria
[d] Stratec Consumables, Anif, Austria

A R T I C L E I N F O

Keywords:
Predictive maintenance
Time-series-based forecast models
Quality criteria
Batch process
Generalized evolving fuzzy systems
Non-linear dynamic PLS-fuzzy models
Incremental PLS space updating
IPLS-GEFS

A B S T R A C T

An important predictive maintenance task in modern production systems is to predict the quality of products in order to be able to intervene at an early stage to avoid faults and waste. Here, we address the prediction of the most important quality criteria in micro-fluidics chips: the flatness and critical size of the chips (in the form of RMSE values) and several transmission characteristics. Due to semi-manual inspection, these quality criteria are typically measured only intermittently. This leads to a high-dimensional batch process modeling problem with the goal of predicting chip quality based on the trends in these process values (time series). We apply time-series based transformation for dimension reduction to the lagged time-series space using of partial least squares (PLS), and combine this with a generalized form of Takagi–Sugeno (TS) fuzzy systems to obtain a non-linear PLS forecast model (termed as *PLS-fuzzy*). The rule consequent functions are robustly estimated by a weighted regularization scheme based on the idea of the elastic net approach. To address particular system dynamics over time, we propose dynamic updating of the non-linear PLS-fuzzy models using new on-line time-series data, with the options 1.) adapt and evolve the rule base on the fly, 2.) smoothly down-weight older samples to increase flexibility of the fuzzy models, and 3.) update the PLS space by incrementally adapting the loading vectors, where processing is achieved in a single-pass stream mining manner. We call our method *IPLS-GEFS (incremental PLS combined with generalized evolving fuzzy systems)*. We applied our predictive modeling approach to data from on-line microfluidic chip production over a time period of about 6 months (July to December 2016). The results show that there is significant non-linearity in the predictive modeling problem, as the non-linear PLS-fuzzy modeling approach significantly outperformed classical PLS for most of the targets (quality criteria). Furthermore, it is important to update the models on the fly *with* incremental updating of the PLS space and/or *with* down-weighting older samples, as this significantly decreased the accumulated error trends of the prediction models compared to conventional updating. Reliable predictions of flatness quality (with around 10% error) and of RMSE values and transmissions (with around 15% errors) can be achieved with prediction horizons of up to 4 to 5 h into the future.

1. Introduction

1.1. Motivation and state of the art

Predictive maintenance relies on real-time monitoring and diagnosis of system components, and process and production chains (Levitt, 2011). The primary strategy is to take action when items or parts show certain behaviors that usually result in machine failure, reduced performance or a downtrend in product quality. In contrast to classical quality control and condition monitoring (Montgomery, 2008; Lughofer et al., 2012), which basically operate in a kind of retrospective and reactive manner — for instance, by inspecting product parts for atypical aberrant appearance (Lughofer et al., 2009; Sannen and van Brussel, 2012; Demant et al., 1999) (as in Weigl et al., 2016; Pawell et al., 2015; Schwarzbauer et al., 2013 for microfluidic chips), predictive

* Corresponding author.
E-mail address: edwin.lughofer@jku.at (E. Lughofer).

https://doi.org/10.1016/j.engappai.2017.11.001
Received 28 April 2017; Received in revised form 21 September 2017; Accepted 1 November 2017

maintenance goes a step further and predicts impending problems at an early stage. The goal is to identify problems as early as possible (Aumi et al., 2012; Mobley, 2002) in order to have sufficient time to react properly (manually Wang and Gao, 2006 or automatically Permin et al., 2016), and to prevent severe quality downtrends and even risks in subsequent processing stages. This can bring considerable cost savings due to significantly reduced system downtime and reduced rejected parts. This also the case in microfluidic chip production, especially when taking into account that (1) the amount of defect chips (waste) is greatly reduced and (2) chip quality measurements require special equipment and are time intensive, as they are often manually conducted. Predictive maintenance in microfluidic chip production is thus concerned with the prediction of quality criteria that indicate the condition of the chips (e.g., shape, flatness and size), based on current process states (Wang and Gao, 2006) occurring in two essential production stages: injection molding and the bonding processes (Attia et al., 2009).

Techniques from the fields of forecasting (Box et al., 1994) and prognostics (Ekwaro-Osire et al., 2017) form the core components in predictive maintenance systems. These techniques can build on analytical (Liu, 2008), knowledge-based (Fonseca, 0000) or purely data-driven models (Liao and Wang, 2013) that map process trends and states to some quality information content (e.g., the health index) (Adams, 2007). While both analytical and knowledge-based models require time-intensive derivation and development phases and thus significant man-power and furthermore often are often restricted to very specific applications and product settings (Kluska, 2009), data-driven models can be generated more or less automatically (with some support from machine learning experts) from data (typically time-series-based process data) recorded at the system and stored. In the context of time-series data, the particular data-driven models are usually termed as time-series based prediction or forecast models (Gaxiola et al., 2015).

Several approaches have been proposed for quality control and predictive maintenance in chip production, for instance, in Pawell et al. (2015) and Preininger and Sauer (2003). These rely on analytical and knowledge-based models, which require adequate (physical or chemical) knowledge of the (relations/dependencies in the) system and whose derivation is very time- and effort-intensive for experts; further, they are not sufficiently flexible to automatically adapt to (on-line) process changes and often require complete re-development phases for different product settings, variants and/or charges. Another, more automated approach that addresses the supervision of process parameters from injection molding machines was proposed in Park et al. (2016). It applies basic statistical concepts, such as linear regression analysis, to historical data, but integration of more complex soft computing and machine learning models to capture model non-linearities in the system is insufficient, and it does not address any changes in characteristics of the production process over time. Similar considerations apply to the approaches in Zhang et al. (2015), which relies on linear principal component regression (PCR) that are built once and not adapted over time to address system dynamics. Other approaches, such as those presented in Fu et al. (2011); Yu et al. (2014), perform an influence analysis of process parameters on product quality, but do not provide a real predictive modeling procedure for forecasting future quality or recognizing problems at an early stage. In Kano and Nakagawa (2008), conventional partial least squares regression (PLSR) models are used for transforming the high-dimensional process value space in order to reliably predict product quality in steel industry; again, these models can reflect only linear behaviors sufficiently well and cannot autonomously adapt to process changes and system dynamics.

Dynamics are an important aspect of chip production systems, as they can become significant due to varying process cycles, charges, events or even non-stationary environmental influences (such as temperature and air pressure changes). In fact, in case of micro-fluidics chip production, we analyzed this dynamics (see the experimental results section) and realized that the predictive performance of static models trained once on training samples deteriorate significantly after several weeks. Such cases typically require model maintenance cycles (Wise and Roginski, 2015), in which the model is re-calibrated based on data collected from the new (changed) states/influences. Ideally, such model updates should be achieved automatically and on-line in order to avoid time-consuming off-line model re-design phases that require an expert in (data-driven) modeling. This can be enabled by self-adaptive model adaptation techniques, such as incremental learning of parameters and on-the-fly evolution of structural components. None of the approaches discussed above have such techniques embedded in their quality control and predictive models.

Further, (longer) trends of (measured) process values are usually required to make reliable predictions of product (chip) quality with a sufficiently distant horizon. This is because individual samples may be too affected by noise and may not contain the necessary information over time for accurately predicting a particular quality. Hence, we are faced with a high-dimensional time-series-based forecast problem, since entire trends of time series that reflect the progress of the process affect the quality of future chips. As the quality of the chips are recorded intermittently, it leads to a *batch process forecast modeling* problem, see Section 2. Appropriate dimension reduction techniques (Carreira-Perpinan, 1997) are to be used to avoid the curse of dimensionality (Hastie et al., 2009), because high input dimensionality is a major cause of over-fitting. This becomes especially severe for small sample sizes, as is usually the case in our application, since we are dealing with quality criteria sampled periodically only a few times per day (leading to around a 100 samples in total per month). To properly address system dynamics in the form of changes in the influence/importance of variables for a reliable quality prediction, an appropriate combination of (incremental) dimension reduction and model adaptation is required, which can also deal with batch process modeling problems.

1.2. Our approach

From the application point of view, the main novelty of our approach lies in its holistic approach towards delivering a highly performant, robust and purely data-driven (i.e., general) solution for predicting/forecasting the most important quality criteria in on-line micro-chip production (with certain dynamics). As such, our method is designed to handle important general and domain-specific complications associated with data-driven modeling like:

- A rather prohibitive cost of (manually) sampling and analyzing new target data (i.e., quality criteria) that strongly favors prediction models based on the online (single-pass) learning paradigm.
- A very complex modeling environment that is dynamic and is also likely to feature non-linear targets.
- A very high number of process values (i.e., read-only/diagnostics information) that can be inexpensively and inherently measured but are not assured to be useful for modeling the targets.

It is noteworthy that the first two expected complications form an apparent modeling predicament as (i) a sparser sampling strategy can obfuscate the true dynamics and non-linearity of a production process and (ii) a highly complex and dynamic production process should be more extensively sampled in order to produce good (i.e., usable) prediction models. Our approach aims to efficiently (i.e., automatically and without expert-based supervision) make use of the continuously recorded process values and on-line updates in order to accommodate on the one side the dynamics and complexity of the production process as well as the cost-efficient sampling requirements on the other side.

In order to effectively tackle the aforementioned challenging characteristics of the micro-chip production process, the predictive maintenance strategy we propose is grounded in five major aspects that are motivated both by strong theoretical and practical considerations: