Journal of Process Control xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Process Control

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

Statistical process monitoring as a big data analytics tool for smart manufacturing

Q. Peter He*, Jin Wang

Department of Chemical Engineering, Auburn University, AL 36849, United States

ARTICLE INFO

Article history:
Received 16 June 2016
Received in revised form 3 May 2017
Accepted 20 June 2017
Available online xxx

Keywords:
Statistical process monitoring
Big data
Smart manufacturing
Feature extraction
Internet of things

ABSTRACT

With ever-accelerating advancement of information, communication, sensing and characterization technologies, such as industrial Internet of Things (IoT) and high-throughput instruments, it is expected that the data generated from manufacturing will grow exponentially, generating so called 'big data'. One of the focuses of smart manufacturing is to create manufacturing intelligence from real-time data to support accurate and timely decision-making. Therefore, big data analytics is expected to contribute significantly to the advancement of smart manufacturing. In this work, a roadmap of statistical process monitoring (SPM) is presented. Most recent developments in SPM are briefly reviewed and summarized. Specific challenges and potential solutions in handling manufacturing big data are discussed. We suggest that process characteristics or feature based SPM, instead of process variable based SPM, is a promising route for next generation SPM and could play a significant role in smart manufacturing. The advantages of feature based SPM are discussed to support the suggestion and future directions in SPM are discussed in the context of smart manufacturing.

1. Introduction

With the emergence of the industrial Internet of Things (IoT) and ever advancing computing power and expansion of wireless networking technologies, a new generation of networked, information-based technologies, data analytics, and predictive modeling are providing unprecedented embedded computing capabilities as well as access to previously unimagined potential uses of data and information. These capabilities provide possibilities for new, radically better ways of doing manufacturing [1]. Although there are different names used to describe next generation manufacturing systems, such as smart/advanced manufacturing and Industry 4.0, the essence of these is the application of increasingly powerful and low-cost computation and networked information-based technologies in manufacturing enterprises. There is a general consensus that factories and plants connected to the Internet are more efficient, productive and smarter than their non-connected counterparts [2,3].

According to Smart Manufacturing Leadership Coalition (SMLC), smart manufacturing is the dramatically intensified and pervasive application of networked information based technologies throughout the manufacturing and supply chain enterprise [2,3]. One of the focuses of smart manufacturing is to create manufacturing intelligence from real-time data to support accurate and timely decision-making. Therefore, data analytics is expected to contribute significantly to the advancement of smart manufacturing.

Existing manufacturing process operation databases are already massive because of the use of process operation and control computers and information systems. With ever-accelerating advancement of IoT devices and other communication and sensing devices and technologies, it is expected that the data generated from future smart manufacturing systems will grow exponentially [4]. As shown in Fig. 1, 4 V's are often used to characterize the essence of big data [5,6]: Volume (the size/scale of the data), Variety (the form/format of the data), Velocity (the rate of the data being produced), and Veracity (the uncertainty/reliability of the data).

Big data is arguably a major focus for the next round of the transformation of advanced manufacturing. According to research by McKinsey Global Institute and McKinsey's Business Technology Office, the analysis of large datasets will become a key basis of competitiveness, productivity growth, and innovation [7].

Generally speaking, data analytics can be viewed as the science and engineering of examining data to uncover hidden patterns, unknown correlations and other useful information that can be used to make better decisions or to develop effective solutions. There are many applications of big data analytics that have been

* Corresponding author at: Auburn University, 212 Ross Hall, Auburn, AL 36849, United States.
E-mail address: qhe@auburn.edu (Q.P. He).

http://dx.doi.org/10.1016/j.jprocont.2017.06.012

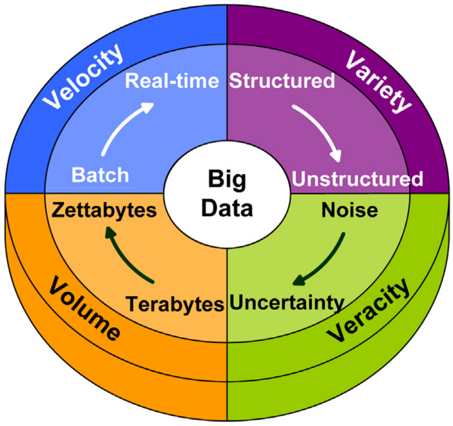

Fig. 1. The 4 V's of big data.

used in process industries, including fault detection and diagnosis, process optimization and control, predictive maintenance, etc. In this work, we focus on statistical process monitoring (SPM), and discuss the challenges and opportunities presented by the manufacturing big data. The rest of the paper is organized as the following: Section 2 briefly reviews the development of SPM and presents a roadmap of it. Section 3 discusses challenges and desired properties of next generation SPM driven by manufacturing big data. Some future directions in the application of SPM to smart manufacturing are explored in Section 4. And conclusions are drawn in Section 5.

2. A roadmap of statistical process monitoring

We propose a roadmap of statistical process monitoring (SPM) as shown in Fig. 2, which divides the development of SPM into three generations: 1st generation: statistical process control (SPC); 2nd generation: multivariate statistical process monitoring (MSPM); and 3rd generation: yet to be properly defined and named.

2.1. Statistical process control (SPC): the 1st generation

The first generation SPM is called statistical process control (SPC), which was pioneered by Walter Shewhart at the Bell Laboratories in the 1920s [8]. Shewhart developed the first control chart in 1924 with an attempt to differentiate "assignable" or "special" source of variation from "common" sources. The common sources of variations result in a normal distribution of samples/observations, where the mean ($\bar{m}$) and standard deviation (σ) can be estimated. Any observation outside of $\bar{m} \pm 3\sigma$ deserves further investigation to see if there is an assignable cause and any correction or adjustment needs to be made. The underlying statistical model for a SPC chart is simply a normal or Gaussian distribution model, which is determined by sample mean and variance or standard deviation. It is worth noting that, before SPC, there was no model based quality control – adjustments were made by comparing each individual sample to a specified target – if a sample measurement is off target, an adjustment would be made with the attempt to bring the next sample measurement to the target, which is unnecessary and often worsens the product quality. One of the most notable example of SPC's contribution to manufacturing is the great success of Japanese manufacturing in the 1970s [8].

2.2. Multivariate statistical process monitoring (MSPM): the 2nd generation

In the 1980s, some limitations of univariate SPC charts were revealed [9], and a major one is illustrated in the following simulated example where two properties, x_1 and x_2, were measured for a product. In this simulated example, the univariate SPC charts, shown in Fig. 3(a), fail to detect the faults occurred in the last six samples, as shown in Fig. 3(b). The multivariate chart, Fig. 3(b), clearly shows that x_1 and x_2 of the normal measurements are positively correlated, while the last six samples do not follow that positive correlation.

Meanwhile, it was recognized that process variables (*e.g.*, process temperature and pressure) and their correlations to product quality, which were not utilized by SPC, could provide extra benefit to process monitoring [9]. The utilization of both process and quality variables led to the birth of the 2nd generation SPM or multivariate SPM (MSPM). Principal component analysis (PCA), partial least squares (PLS), and their variants in many ways form the basis of latent variable based MSPM.

MSPM was a significant leap forward in terms of fault detection performance as compared to SPC, especially for large and complex chemical processes where process variables are highly correlated due to the physical and chemical laws and principles that govern the process such as mass and energy balances, thermodynamics and chemical kinetics. As a result, MSPM methods have become the industrial standard methods for process monitoring – they have been widely implemented in process industries and numerous successful stories have been reported [10–12]. More in-depth discussion on various MSPM methods and their applications can be found elsewhere [13].

2.3. Limitations of the 2nd generation MSPM and recent developments in addressing them

Despite their significantly improved monitoring performance compared to the 1st generation methods, the 2nd generation MSPM methods have their limitations. The underlying assumption made for the PCA and PLS based MSPM methods is that the

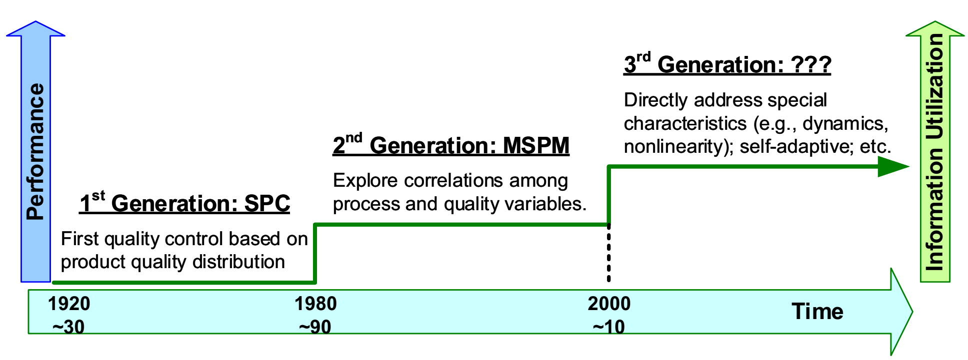

Fig. 2. A roadmap of statistical process monitoring (SPM) technology.

Please cite this article in press as: Q.P. He, J. Wang, Statistical process monitoring as a big data analytics tool for smart manufacturing, J. Process Control (2017), http://dx.doi.org/10.1016/j.jprocont.2017.06.012