Journal of Manufacturing Systems 43 (2017) 187–194

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

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

Big data analytics based fault prediction for shop floor scheduling

Wei Ji [a,b], Lihui Wang [a,*]

[a] KTH Royal Institute of Technology, Stockholm, 100 44, Sweden
[b] Harbin University of Science and Technology, Harbin, 150080, China

ARTICLE INFO

Article history:
Received 27 March 2017
Accepted 29 March 2017

Keywords:
Big data analytics
Fault prediction
Shop floor
Scheduling

ABSTRACT

The current task scheduling mainly concerns the availability of machining resources, rather than the potential errors after scheduling. To minimise such errors in advance, this paper presents a big data analytics based fault prediction approach for shop floor scheduling. Within the context, machining tasks, machining resources, and machining processes are represented by data attributes. Based on the available data on the shop floor, the potential fault/error patterns, referring to machining errors, machine faults and maintenance states, are mined for unsuitable scheduling arrangements before machining as well as upcoming errors during machining. Comparing the data-represented tasks with the mined error patterns, their similarities or differences are calculated. Based on the calculated similarities, the fault probabilities of the scheduled tasks or the current machining tasks can be obtained, and they provide a reference of decision making for scheduling and rescheduling the tasks. By rescheduling high-risk tasks carefully, the potential errors can be avoided. In this paper, the architecture of the approach consisting of three steps in three levels is proposed. Furthermore, big data are considered in three levels, i.e. local data, local network data and cloud data. In order to implement this idea, several key techniques are illustrated in detail, e.g. data attribute, data cleansing, data integration of databases in different levels, and big data analytic algorithms. Finally, a simplified case study is described to show the prediction process of the proposed method.

1. Introduction

The modern manufacturing industry is characterised by high quality and fast delivery, and current manufacturing systems are aimed for good performance, high stability and high repeatability. Among many others, predictability, an important function, offers a good potential to future manufacturing, such as smart manufacturing and intelligent manufacturing. Here, two potensial scenarios are related to the predictability: (1) before the decision making of shop floor scheduling, an effective prediction of potential faults can avoid task delay or other unnecessary loss; and (2) during machining, the patterns of real-time monitoring signals may indicate possible faults, leading to the rescheduling of the remaining tasks. Therefore, fault prediction on shop floor plays an important role in optimising resource allocation, reducing manufacturing cost, and improving manufacturing efficiency.

Nowadays, highly dynamic and integrated methods are employed in scheduling, which makes fast rescheduling possible should there be faults occurred during machining. However, today's scheduling systems lack prediction capability with regard to errors or potential faults of planned or ongoing tasks on the shop floor. It is also a big challenge on how to predict potential faults, and what the error patterns are before scheduling. In recent years, big data analytics has emerged, which offers a good potential for fault prediction.

Within the context, this paper presents a big data analytics based fault prediction approach for shop floor scheduling. The remainder of this paper is organised as follows. A short literature review on both scheduling and big data research is presented in Section 2. Section 3 describes an overview of big data analytics, together with a concept of big data analytics based fault prediction for scheduling. Section 4 reveals the architecture of big data analytics based fault prediction to be embedded in task scheduling, and several technologies and procedures are also introduced. A simplified proof-of-concept case study is carried out to validate the prediction processes in Section 5. Finally, Section 6 concludes the paper and outlines our future work.

2. Literature review

The research on scheduling started with static scheduling useful for the development of shop floor scheduling systems for mass

* Corresponding author.
E-mail addresses: weiji@kth.se (W. Ji), lihui.wang@iip.kth.se (L. Wang).

http://dx.doi.org/10.1016/j.jmsy.2017.03.008
0278-6125/

production. A mathematical model based method was developed by Manne [1]. Laporte et al. [2] analysed two integer linear programming formulas to address job sequencing and tool switching problems, in which branch-and-cut and branch-and-bound algorithms were developed and compared. Asokan et al. [3] used adaptive genetic algorithm and particle swarm optimisation to obtain optimal schedules and storage assignments.

The concept of integration of both process planning and scheduling was developed by Khoshnevis and Chen [4]. In their method, the feasible process for a given feature of a part was not found in the shop, in case of which, the real-time system was then applied to generate process plans and schedules together. Within the context, Chen and Khoshnevis [5] also presented some methods for the integrated system and the performance of the algorithm. Tan and Khoshnevis [6] further extended the approach. Mohapatra et al. [7] proposed an improved controlled elitist non-dominated sorting genetic algorithm to reduce scheduling objectives, for instance makespan, cost, idle time and efficiency for the integration of process planning and scheduling. Freitag and Hildebrandt [8] proposed a simulation-based multi-objective hyper-heuristic to develop optimisation dispatching rules for complex manufacturing systems. Li at al. [9] pointed out that the integration of process planning and scheduling would be developed towards multi-objectives, dynamic and hybrid algorithm application. Rajkumar et al. [10] applied greedy randomised adaptive search procedures algorithm to the integration of process planning with production scheduling with regard to the process problems having multi-objectives of makespan, maximum workload, total workload, tardiness and total flow time. In addition, scheduling was also carried out for some of special objectives, e.g. sustainable development. Gahm et al. [11] suggested developing energy-efficient scheduling for manufacturing companies. Other related work on integrated process planning and scheduling can be found in [12,13].

However, the static scheduling does not have the capacity to handle the situation with growing products of both small batch and wide variety, in particular unexpected faults. Therefore, dynamic scheduling method was developed, where the decisions can be made with rapid response automatically. Real-time scheduling of a manufacturing system involves scheduling and revised scheduling [14]. Vieira et al. [15] proposed a rescheduling method based on a wide variety of experimental and practical approaches. In their method, two common strategies were introduced, dynamic scheduling and predictive–reactive scheduling. Zhong and Xu [16] presented a job-shop scheduling model which converted the real-time captured data from physical Internet-based manufacturing shop floor. Ham et al. [17] proposed a three-stage flexible job-shop scheduling method to deal with unpredictable system disturbances. Iwamura et al. [18] introduced an estimation of future status based real-time scheduling approach for holonic manufacturing systems (HMS). In their method, the future status of an HMS is predicted by applying a neural network model based simulation model. In addition, an agent-based service-oriented architecture was presented for real-time distributed shop floor scheduling [19]. Semi-Markov decision models were also applied to real-time scheduling by Yih and Thesen [20]. Ant colony optimisation was applied to two dynamic job scheduling by Zhou et al. [21]. For real-time scheduling, real-time decision making is the key. There are many algorithms. By constructing a decision tree, Metan et al. [22] proposed a new scheduling system for selecting dispatching rules in real-time. The proposed scheduling system was developed by combining the techniques of simulation, data mining, and statistical process control charts. Bayesian algorithm was used to discover priority dispatching rules from a large amount of structured or unstructured data for the single machine scheduling problem [23]. In addition, real-time monitoring methods were also developed. Kohn et al. [24] proposed repair-control of manufacturing systems using real-time RFID information. Through applying RFID on shop floor, the real-time information of objects including operators, machines and materials, can be automatically captured, bound and synchronised with manufacturing orders [25–27].

According to the literature, dynamism, flexibility and adaptability are the important features in modern scheduling, and a scheduling system should be able to perform task rearrangement in case of unexpected events. Nevertheless, job delay and fault in manufacturing are still hard to predict if not unavoidable. Although limited in achievements, big data analytics shed lights in fault prediction. In big data research, many reported work focused on applications of big data in production lifecycle and supply chain management [28]. Köksal et al. [29] introduced data mining to quality improvement. Mavridou et al. [30] developed a mass-customisation recommender system using data mining of automotive industry customer data. Zhang et al. [31] proposed a big data-based analytics for product lifecycle, supply chain management and maintenance of complex products, where big data analytics and service-driven patterns were used. Big data analytics was also used in product lifecycle management [32], and the authors investigated big data in manufacturing phases. Chen et al. [33] proposed an integrated model by combining K-means clustering, feature selection and the decision tree method into a single evaluation model to address evaluation problem of suppliers in the supply chain. Zhong et al. [34] reviewed the state of big data technology used in services and manufacturing supply chain management, including six aspects of challenges, opportunities and future perspectives: data collection, data transmission, data storage, processing technologies, big data-enabled decision-making models, and big data interpretation and applications. Babiceanu and Seker [35] reviewed the relevant research, and they insisted that big data analytics will be used for manufacturing cyber-physical systems. In addition, Zhong et al. [36] proposed a big data approach to logistics trajectory discovery. In their method, RFID-Cuboids were used to establish data warehouse, and then abstract data can be converted into meaningful information, based on which, mined knowledge and associated indexes were worked out for different manufacturing goals. Woo et al. [37] developed a big data analytics platform for manufacturing systems, in order to create prediction models specific for target machine tools. In conclusion, big data analytics has been used widely, and shows the potential of fault prediction for shop floor scheduling.

3. Big data analytics based fault prediction

Big data analytics is the process of examining large data sets to reveal hidden patterns, unknown correlations, market trends, customer preferences and other information. It has been applied to a number of areas by many companies [38].

3.1. Algorithms used in big data analytics

Methods employed in big data analytics are developed from the traditional approaches for data analytics. The principal methods are summarised as follows.

(1) Cluster analysis: it captures the natural structure of data. Originated in anthropology in 1932 and introduced to psychology in 1938 [39,40], it was then used for trait theory classification of personality psychology in 1943 [40].
(2) Factor analysis: it is a statistical method used to describe variability among observed and correlated variables in terms of a potentially lower number of unobserved variables called factors [41].