



Analysis of production cycle-time distribution with a big-data approach

Xu Tan^{1,2} · Lining Xing^{1,2} · Zhaoquan Cai³ · Gaige Wang⁴

Received: 4 August 2019 / Accepted: 29 January 2020 / Published online: 5 February 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

In production planning, one of the most crucial issues involves cycle time forecasting and distribution. Particularly, the parameter aids in realizing high delivery reliability. In the production planning process that involves computer component manufacturing, an estimation of the tasks' cycle time offers an important basis for dispatching control, material purchase, and due date assignment. In this study, a big-data approach was proposed and examined to determine how it could be used to predict cycle time distribution. Also, the research context involved computer components manufacturing systems. Indeed, the motivation was to determine how the proposed mechanism could improve delivery reliability in manufacturing systems. Regarding the implementation and design of the CT forecasting system, with the proposed DP-RBFN framework being a model to be implemented in computer components manufacturing, components of the system constituted three major parts. The first part, being the basic platform, played the role of Hadoop series software installation. This installation had its role lie in enabling the parallel computing of big data. Another part of the framework design and implementation involved data preprocessing. In this case, the role of the data preprocessing procedure lay in the extraction, transformation, and loading of data to CTF. The third part that followed the basic platform design and data preprocessing procedure involved CT forecasting. Results demonstrated that the proposed model performs superiorly than the contrast or other comparative methods on both the computer components manufacturing system dataset and benchmark datasets. From the findings, the proposed framework (DP-RBFN) exhibited superior performance compared to previous performance outcomes that had been reported relative to the use of the RBFN algorithm. These findings held for both MAD and SD—relative to the selected datasets.

Keywords DP-RBFN · RBFN · Cycle time (CT) · Computer components manufacturing · CT forecasting · Parallel training

Introduction

To improve manufacturing processes and overcome the currently rigid planning, key elements that have been proposed involve real-time data analytics methods. In particular,

these methods fulfill the aforementioned tasks by deriving measures, detecting patterns, and analyzing historical data, eventually counteracting the issues (Zhengcai et al. 2012). To optimize, assist, and improve process flows, the key element establishes virtual product representations on shop-floors, emerging in the form of the digital shadow or digital twin (Gopalakrishnan et al. 2013). To use the collected data, high data quality is required, pointing to the criticality of developing measures through which the data's correctness could be verified (Gröger et al. 2012). It is also worth noting that digital manufacturing reflects a new emerging technology whose evolution has been informed by the need to increase productivity. In manufacturing firms, shop-floor data is collected in digital forms via sensor technologies and Manufacturing Execution Systems (MES) (Hyndman and Khandakar 2008). The eventuality is that voluminous data is collected at high velocities and, upon scaling up to a factory level or a produc-

✉ Lining Xing
liningxing518172@gmail.com

¹ School of Transportation and Logistics, Central South University of Forestry and Technology, Changsha 410004, People's Republic of China

² School of Software Engineering, Shenzhen Institute of Information Technology, Shenzhen 518172, People's Republic of China

³ Department of Information Science and Technology, Huizhou University, Huizhou 516007, People's Republic of China

⁴ Department of Computer Science and Technology, Ocean University of China, Qingdao 266100, People's Republic of China

tion system level, it translates into big data (Li et al. 2011; Roser and Nakano 2015).

Given that most of the data grow exponentially, having been acquired from the machines, it becomes imperative to ensure that data science is leveraged in a manner that fosters the enhancement of manufacturing processes, upon which the decision making processes could be data-driven (Shao 2015). For these forms of decision making (that are data-driven) to be realized, the need for companies to adopt informative analytical algorithms through which fast-moving and high-volume data could be turned into meaningful insights cannot be overstated (Subramaniyan et al. 2016; Lee et al. 2013). The latter trend calls for research about some of the data analytics responsible for the realization of effective and efficient data extraction (from raw information), upon which new insights and knowledge could be derived (Michael et al. 2016; Wuest et al. 2016; Luo et al. 2015; Yang et al. 2016). Furthermore, the data extraction is likely to aid in the introduction of intelligence into production process control, besides yielding improvements in the manufacturing enterprises' system-level operations (Zhang et al. 2015; Zhong et al. 2015).

In production planning, one of the most crucial issues involves cycle time forecasting and distribution. Particularly, the parameter aids in realizing high delivery reliability. In the production planning process that involves computer component manufacturing, an estimation of the tasks' cycle time offers an important basis for dispatching control, material purchase, and due date assignment. This study proposes a big data-driven approach through which the cycle time distribution of computer component manufacturing processes could be predicted, especially that which prompts a delivery reliability of the manufacturing system. The technique comes in the form of Density Peak based Radial Basis Function Network (DP-RBFN). Also, the proposed approach involves parallel computing in which a numerical experiment is conducted for forecasting accuracy and training time of the technique, eventually evaluating and discussing its performance.

Methodology

With the proposed model, which reflects a cycle time forecasting (CTF) system, this section describes the structure of the model, its peak clustering-based learning method, and the training steps aimed at discerning the method's time complexity. As mentioned earlier, the proposed framework for CT forecasting entails a DP-RBFN model. In this model, three layers are established. Having described the proposed model's network structure (above), the next step involves determining the leaning method. In particular, the model attracts a density peak clustering-based technique. This technique provides room for the separate learning of the RBFN

model's output layer and the hidden layer. To achieve rapid training of the RBFN, a parallel learning method is established within the hidden layer. To establish the density peak, the process involves massive data. Another step involves the implementation of a least square technique within the output layer, a process whose role lies in the estimation of the respective hidden unit outputs' weights.

The above process culminates into the radial basis function networks' parallel training. In particular, the Hadoop platform is employed for DP-RBFN parallel training. Initially, network coefficients are adjusted relative to the designed training steps. This process is followed by the training procedure's time complexity analysis. In turn, steps that are deemed the most time-consuming are parallelized with the help of a MapReduce framework.

Regarding the implementation and design of the CT forecasting system, with the proposed DP-RBFN framework being a model to be implemented in computer components manufacturing, components of the system constitute three major parts. The first part, being the basic platform, plays the role of Hadoop series software installation. This installation has its role lie in enabling the parallel computing of big data. In the Hadoop components, the notable, efficient, and highly effective tool that is selected towards big data analysis, hence CT forecasting, entails MapReduce. Another part of the framework design and implementation involves data preprocessing. In this case, the role of the data preprocessing procedure lies in the extraction, transformation, and loading of data to CTF. Upon ensuring that the manufacturing execution system's raw data is extracted, processes that follow include the transformation and loading of the raw data into the forecasting model. Imperative to note is that the process of transforming the data involves cleaning and formatting procedures. On the one hand, the role of raw data formatting constitutes reformatting it to achieve uniform units that are, in turn, normalized to realize a canonical distribution. On the other hand, the processing of cleaning the raw data ensures that any abnormal and missing records are fixed. Indeed, missing records are predicted to arise from errors linked to raw dataset collection, transformation, and storage.

The third part that will follow the basic platform design and data preprocessing procedure involves CT forecasting. The proposed DP-RBFN model strives to aid in CT prediction in the context of computer components manufacturing processes involving big data, especially with the need to discern the effectiveness and efficiency of the framework. The implementation of the CT forecasting process will involve the initialization of the model, the training of the network, and the CT prediction procedure.

Indeed, the experimental setup constituted four main parts. These parts included the parallel CTF model, CMI-based feature selection, data pre-processing, and the construction of candidate feature sets. Regarding the section or phase

constituting candidate feature set constructing, CT as an experimental parameter was treated as the time that had elapsed between the last manufacturing cycle process and the beginning of the first process. As such, this parameter, CT, entailed the waiting time and the processing time. From the previous literature, the main challenging facing CT estimation involves the uncertain nature of the waiting time (Hyndman and Khandakar 2008; Li et al. 2011; Roser and Nakano 2015; Shao 2015). As such, three major factors were considered at this phase. These factors included high dimension, variety, and volume. Whereas volume was treated as the data amount such as the number of samples in the data recorders, variety involved the number of types of data (such as the type of percentage due to machine utilization, the type of data time due to the processing time, and the type of numerical due to the size of the manufacturing lots). Lastly, high dimension as a CT factor constituted the candidate feature quantity relative to the manufacturing stations and huge process.

For data pre-processing, raw data-sets were transformed into proper structures or formats through which analyses or queries could be done. Indeed, raw data contained millions of records regarding the computer components manufacturing. A manufacturing execution system was used to extract this (raw) data. Some of the specific data that was selected in its raw form before being transformed into proper structure or format included the equipment name, hold time, the track out time, and the track in time, as well as the operation step number.

Upon completing the pre-processing stage, the next step involved feature selection. The aim of this step was to establish the main features. In particular, the interest of the study was on CT-related features. The aim was to determine how the proposed model for CT prediction would choose key features from a given sample of data sets from the computer components manufacturing database. Also, the feature selection process sought to discern how the proposed framework, compared to other models, would estimate and depict how the parameter of CT in the manufacturing context would be affected by the nature of candidate features. It is also notable that this phase strived to address the difficulty that had been reported previously regarding how models could preselect regression functions that are well suited to discern the relationship between CT and candidate features.

It is also imperative to highlight that the discretization process was implemented because sets of candidate features in manufacturing systems tend to have both continuous and discrete variables. To ensure that the relationship among variables was measured uniformly, this study implemented a discretization technique to ensure that all data points were classified in the form of continuous factors. Hence, all the features' data points were divided into different intervals independent of any prior knowledge. Also, unique values

were used to label the acquired intervals. To ensure that there was the same number of data points for the respective intervals, the continuous candidates in this study were discretized using an equal frequency discretization procedure.

To measure the relationship between candidate features and CT, parameters that were used included MI, conditional entropy, joint entropy, and basic entropy formulas. In feature selection, the study applied a CMI feature technique. Specifically, this method of feature selection was used to determine the efficiency of the proposed model, upon which improvements, if any, could be made to the algorithm.

Results and discussion

To determine the degree to which the proposed model, the DP-RBFN framework was likely to be effective relative to CT forecasting (hence system delivery improvements) in computer components manufacturing, three experiments were implemented. The experimental procedures included computer components manufacturing process case study, standard datasets' CT forecasting performance experiment, and an experiment targeting the proposed model's training time. Imperatively, the initial experiment that targeted the DP-RBFN model's parallel training sought to unearth the extent to which the proposed method would save time, should it be employed in CT prediction among production planning processes that involve big data. Indeed, the selected datasets were of different sizes. The motivation was to find out the parallel training technique's accelerating ability. Relative to the second experimental process that aimed at determining the proposed model's CT prediction performance, upon exposure to different computer components manufacturing scenarios, the selected standard datasets exhibited varying properties. Indeed, the second experimental process focused on four standard datasets while the first procedure involved seven selected datasets, which exhibited varying size. Finally, the performance of the DP-RBFN framework in relation to CT forecasting was evaluated based on four major datasets linked to the computer components manufacturing system. The process culminated into an examination of the scope within which the proposed model could function efficiently and effectively.

Results on the proposed DP-RBFN model's training time

The parallel DP-RBFN, which was based on MapReduce, had its efficiency tested relative to the parameter of speedup. In theory, many scholarly studies affirm that the speedup should coincide with the time complexity ratio.

The results of the DP-RBFN model's training time experiment are summarized in the Table 1.

Table 1 The proposed model's results regarding the training time experiment

Selected dataset	Speedup (ratio of computing time without parallelization over computing time with parallelization)	Parallelization-related computing time in m/s	Computing time independent of parallelization in m/s	Number of selected records	Number of selected features
Computer components manufacturing data	1.4500	213,510	310,074	1850	78
Winequality-white	1.7300	2,997,219	5,198,205	4800	12
Winequality-red	1.6700	161,791	270,941	1600	12

Imperatively, the ratio between computing time without and computing time with parallelization led to the realization of speedup values relative to the Hadoop platform. From Table 1, it is evident that the proposed model, which was in the form of the parallel DP-RBFN framework, proves effective regarding its capacity to reduce the amount of time spent on the training process; with particular reference to computer components manufacturing system datasets. This inference is informed by the outcomes in which the speedup values that were achieved were closer to theoretical values, suggesting that the proposed parallel technique can save more than 50% of the training time. For big data, the framework's speedup was also found to outweigh that of the experiment's alternative datasets. From the previous literature, a factor that could explain this trend or outcome is that when small datasets are used, the model training's communication time is unlikely to be negligible (Guo et al. 2015; Jeschke et al. 2017; Wang et al. 2018; Shang et al. 2019). On the other hand, the case of big data holds that communication time is likely to be less dominant compared to the case of the saving processing time (Guo et al. 2015). Hence, in this study, it was established that the proposed DP-RBFN model provides better speedup.

Results on the proposed DP-RBFN's CT forecasting

The second phase of the experiment involved evaluating the degree to which the proposed framework would prove effective regarding CT prediction. To discern possible deviations or greater effectiveness, the outcomes were compared to those that had been documented by previous investigations that relied on the standard RBFN model, enabled by a backpropagation framework. As avowed by many studies, the latter algorithm has gained application in numerous evaluations targeting prediction problems (such as CT forecasting) and how they could be enhanced (Zhang et al. 2015; Zhong et al. 2015; Gröger 2016; Guo et al. 2015). Given some selected datasets for benchmarking, the proposed model's capacity towards effective CT forecasting was evaluated through the examination of standard deviation values, as well as the mean absolute deviation (MAD). The Table 2 provides

a comparative analysis of the CT forecasting outcomes relative to the implementation of the RBFN versus the proposed DP-RBFN framework.

Based on the results presented in Table 2, this study established that the proposed DP-RBFN model exhibits superior performance compared to the use of RBFN, a standard algorithm. The inference was informed by how the target parameters compared. For instance, the case of the CMB dataset saw the SD and MAD of the proposed DP-RBFN model merge as much lower compared to the case of the standard RBFN. Whereas this advantage was weaker when datasets EE, CCS and ISE were used (from the SD and MAD points), a consistent trend was that the proposed framework outperformed the standard RBFN algorithm. It is also notable that the case of dataset CCS saw the RBFN algorithm's SD prove lower than at of the proposed DP-RBFN model, implying that RBFN exhibited a more stable performance compared to the proposed model. Despite the mixed outcomes, the proposed framework (DP-RBFN) outperformed the standard algorithm, RBFN.

Focusing on the results for the computer components manufacturing case scenario

In the third and final phase of the experiment, candidate factors were evaluated, and the results used to discern the efficiency of the proposed model of CT prediction. Specific factors that were assessed included the size of the waiting queue for the respective stations in the computer components manufacturing system, the priority of the respective dataset lots, the exploitation or utilization of various stations, and the processing time of various operations on their associated processing routes. As mentioned in the methodology, the data preprocessing procedure constitutes the transformation of raw data into a stat aligned to candidate factor datasets. Specific processes included key factor selection, formatting, and cleaning. Specific raw data that was cleaned involved that with logic, variance, redundancy, and null value. In situation, where some raw datasets were missing in values, an adjacent

Table 2 A comparison of CT forecasting results between the RBFN and DP-RBFN models

Type of dataset	SD RBFN	SD Proposed DP-RBFN	MAD RBFN	MAD Proposed DP-RBFN	Number of records	Number of features
EE	2.8	2.27	3.5700	3.57	768	8
CCS	6.98	9.64	8.7000	8.7	1030	8
ISE	0.0071	0.0065	0.0076	0.0076	536	7
CMB	0.0008	0.0002	0.0003	0.0003	11,934	16

record, having been selected manually, aided in filling the missing values with their associated attributes.

After implementing the data preprocessing stage, the CT forecasting capability of the proposed model was assessed. Similar to the previous phase, the results or performance of the framework were compared to the standard RBFN algorithm. From the findings of the experimental outcomes for the case scenario in a selected computer components manufacturing system, the proposed framework (DP-RBFN) exhibited superior performance compared to previous performance outcomes that had been reported relative to the use of the RBFN algorithm. These findings held for both MAD and SD—relative to the selected datasets. Whereas the MAD and SD of the proposed DP-RBFN framework for the case of the CMD dataset stood at 0.0002 and 0.0003 respectively, the MAD and SD of the standard RBFN algorithm with which the model was compared stood at 0.0008 and 0.004 respectively (also for the case of the CMB dataset in the case scenario). From the findings collected in relation to the computer components manufacturing system's data, the proposed DP-RBFN model was superior to the standard RBFN algorithm concerning MLR and BPN; with the MAD and SD values aiding in discerning and establishing this consistent trend. Overall, the results demonstrated that the proposed model performs superiorly than the contrast or other comparative methods on both the computer components manufacturing system dataset and benchmark datasets.

From the perspective of comparative analysis, the performance of the proposed framework, DP-RBFN, was compared to that which had been documented for the case of models such as mRMR_D, DISR, and CMIM. The following figures summarize the findings that were obtained when the performance of the model was compared to that of other frameworks. The specific parameter that was used to determine how DP-RBFN rated or compared with other models regarding CT prediction in computer components manufacturing included prediction or classification accuracy relative to feature selection, as well as the number of features. The decision to use the number of features as an independent variable and the prediction or classification accuracy as the dependent variable was to give insight into the extent which the number of features in the given manufacturing dataset was likely to affect the performance of DP-RBFN and, if so,

how the variation in the number of features in the dataset was likely to affect the performance of the proposed model compared to other frameworks mentioned above (mRMR_D, DISR, and CMIM) (Zhang et al. 2019). Indeed, the figures below summarize the comparative analyses and results that were obtained when the performance of the proposed framework, DP-RBFN, was compared to that of the other three models—regarding their CT prediction accuracy in relation to varying numbers of selected features (Figs. 1, 2, 3).

Notably, findings saw the DP-RBFN framework exhibit superior performance when applied to large- and medium-scale datasets—relative to CT prediction performance. A possible explanation for this outcome is that at lower-scale datasets, the small size would make it too tiny to select and predict CT-associated features in computer components manufacturing processes. However, the prediction accuracy of the proposed model was observed to improve significantly with an increase in the size of the manufacturing data. However, it is also worth contending that in situations where large-scaled datasets were affected by noise, the CT prediction stability and accuracy of the proposed framework, DP-RBFN, tended to be hampered (Wang et al. 2018; Yi et al. 2019; Ren et al. 2017).

Despite the mixed outcomes above, an emerging theme that this experimental study revealed is that as different datasets in the computer components manufacturing are presented, there is likely to be a significant variation in the CT prediction performance of the DP-RBFN model. Given that the estimation of relationships among massive data in the computer components manufacturing is employed in CT prediction, the eventuality is that the proposed framework for predicting the CT of manufacturing systems is data-dependent.

This study established further that in situations where there is a delicate change to the flow of materials in the computer components manufacturing system, having considered CT-related features in the entirety, the proposed model is capable of capturing the impact of these delicate changes on the projected CT. Apart from being data-dependent, another emerging theme that was established is that the proposed framework (DP-RBFN) tends to apply or prove efficient to large-scale datasets. Compared to the traditional techniques used to predict the CT parameter, which achieve this objec-

Fig. 1 Comparing the CT prediction performance of the proposed model versus mRMR_D, DISR, and CMIM on 16 features

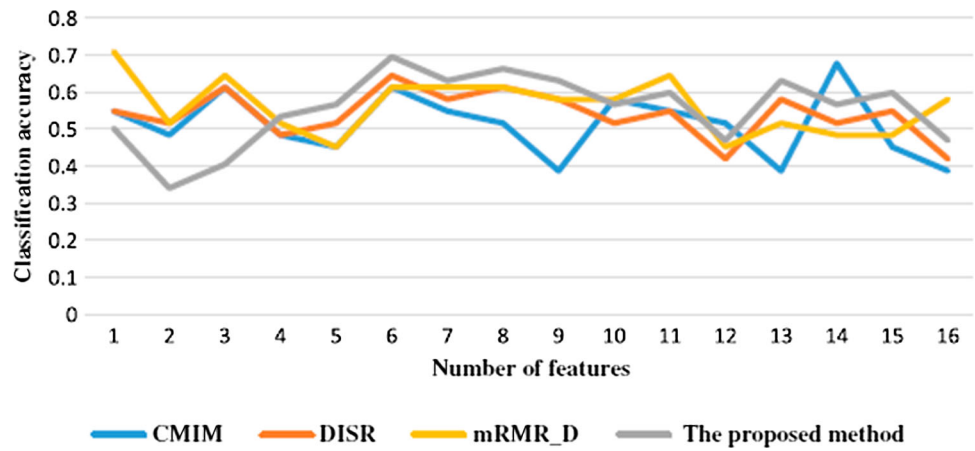


Fig. 2 Comparing the CT prediction performance of the proposed model versus mRMR_D, DISR, and CMIM on 10 features

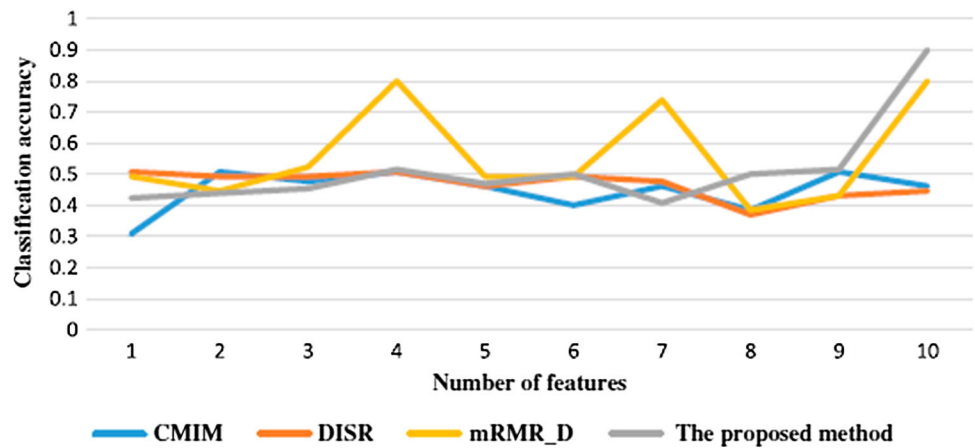
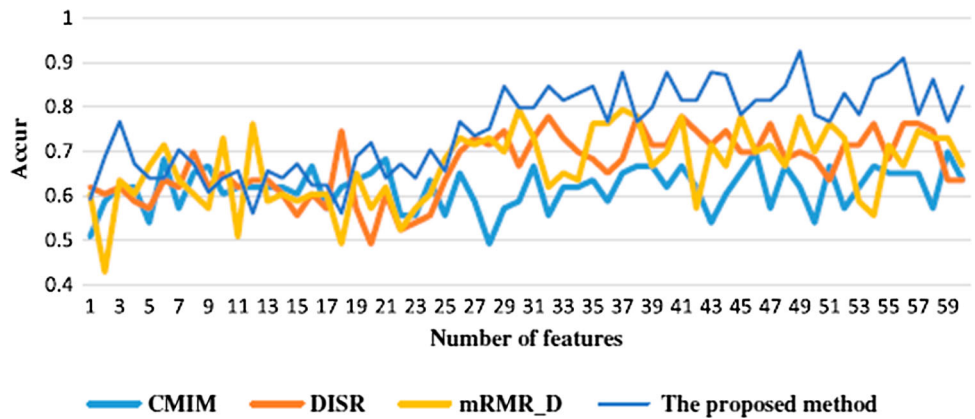


Fig. 3 Comparing the CT prediction performance of the proposed model versus mRMR_D, DISR, and CMIM on 59 features



tive via manufacturing system modeling and analysis, the role of DP-RBFN model was found to lie on CT prediction via the analysis of the interaction and correlation with massive data. As such, the framework was observed to prove highly competitive and worth applying for CT prediction in situations where computer components manufacturing systems are marked by large-scale complex problems that prove too difficult to have their CT predictions achieved via traditional methods (Ren et al. 2019).

Indeed, a density peak clustering technique aided in the training of the proposed DP-RBFN model. Given that many computer components manufacturing systems continue to collect large data amounts, most of the current literature contends that there is a growing need to embrace data-driven CT distribution and forecasting. In this investigation, the proposed CT predictor is big data-driven (Hyndman and Khandakar 2008; Lee et al. 2013; Michael et al. 2016; Wuest et al. 2016). The density peak clustering technique was selected in this study because it had been documented as one

that could predict the clustered, diversified, and large-scale manufacturing data's cycle time accurately. To estimate various parameters of the system network, the learning method that was embraced operated by finding the density peaks. Without necessarily pre-determining the shape and number of categories, the method would detect data classes in their entirety and, in turn, pave the way for the proposed DP-RBFN model to focus on agglomerative data and achieve better CT prediction performance in relation to computer components manufacturing systems (Wu et al. 2018, 2019). To ensure that the DP-RBFN model would be adjusted quickly, the study employed a parallel training process in the form of MapReduce-based technique. In turn, the training procedure's time complexity was analyzed before ensuring that training steps that were deemed to be the most time-consuming were parallelized; with particular reference to the Hadoop platform.

Therefore, there was a successful presentation and evaluation of the DP-RBFN framework relative to its capacity to predict the CT of computer components manufacturing systems. Four major parts through which the objectives of the study were achieved included parallel CTF or data prediction, CMI-based feature selection or data analysis procedure, data pre-processing, and data gathering or the procedure of candidate feature set constructing. The construction of the proposed model was motivated by the need to handle big data in computer components manufacturing, especially after traditional methods had been documented to be limited relative to CT prediction capacity in situations involving large-scale datasets. The big data manufacturing context to which the framework was applied and its results compared to the performance previously documented for other models was characterized by high volume, variety, and dimensionality. With the motive of reducing input feature dimensionality, the proposed model posed improvements to the conditionals mutual data-based method of feature selection. To determine the relationship between continuous variables and discrete variables uniformly, the study implemented a discretization procedure, which ensured that all data points were classified in the form of a continuous factor. Given the large number of the experimental data and the need to deal with the same, the CT in parallel prediction was achieved via an implementation of a parallel CTF framework. Overall results demonstrated that the DP-RBFN model outperforms the other models with which it was compared regarding CT prediction, should a computer components manufacturing system contain large-scale data (Wu et al. 2016, 2018; Jiao et al. 2019).

From the results obtained, the resulting inference is that the proposed DP-RBFN framework is quantitatively superior in terms of predicting the cycle time. To compare the CT prediction performance of the proposed model with that of other techniques, several metrics were examined. The metrics included the selected frameworks' easiness to

use, data required, speed, and accuracy. A factor that was the easiest of these parameters regarding the comparison between the CT prediction performance of the proposed model and that of other frameworks that had been documented earlier concerned the accuracy dimension. Some of the quantitative indicators that were available and made the comparison among models easy (based on the dimension) included RMSE and MAPE. Compared to previous studies that had documented the performance of frameworks such as time-series and statistical, MFLC, analytical, AI, and hybrid model analyses (Jiao et al. 2019; Cai et al. 2019).

For the case of PS analysis, it is worth indicating that the framework calls for constantly updated and large databases. On the other hand, hybrid model analyses required teams of experts whose role would lie in the fuzzy interval correct assessment, especially if a fuzzy logic is adopted (Subramaniyan et al. 2016; Lee et al. 2013; Michael et al. 2016). As such, the two families of frameworks, compared to the proposed model, were found to be the most demanding relative to the issue of the required data. For the case of statistical methods and AI, most of the previous literature contends that the techniques are causal and would only gain application to the same data types, including WIP level-related information, as well as fab utilization rates and the length of the queues. In relation to the case of analytical methods, which were also compared regarding what the previous literature had reported about their prediction performance (and how they would compare with the proposed model in this study), the methods (analytical) are seen to consider the manufacturing machine utilization rate and other parameters but do not prove efficient when applied to scenarios of single-job specific levels. As mentioned earlier, the proposed model's performance was compared to the case of time-series techniques that have been observed to gain application in manufacturing systems' CT prediction. Indeed, the time-series techniques apply to the previous CT themselves and other high-level data, but prove the cheapest to be developed (Hyndman and Khandakar 2008).

Away from the dimension of accuracy, another attribute through which the proposed model's CT prediction performance was compared with the performance of other frameworks involved easiness-to-use. On this parameter, most of the previous scholarly findings contend that the most difficult approaches to use for CT prediction in manufacturing systems include hybrid approaches, followed by AI frameworks. Indeed, the case of AI models is more pronounced is they are applied to situations involving a fuzzy logic (Hyndman and Khandakar 2008; Li et al. 2011; Roser and Nakano 2015; Shao 2015). Whereas the proposed model proved easier to use compared to the aforementioned CT prediction frameworks, a technique that was also found to be easier to use was MFLC. However, when manufacturing scenarios involve uncomplicated fabs, an ideal technique, based

on the comparative analysis, was found to involve analytical methods.

It is also notable that this study contributes to and extends the work of several previous scholarly studies. For instance, Lou (2018) developed a data-driven approach for the discernment of customer requirements and concluded that due to its ability to manage vagueness via the use of intuitionist fuzzy sets, the model is feasible. In another investigation, Ji et al. (2019) focused on the enriched Distribution Process Planning and proposed a big data analytics based optimization technique for determining machining conditions, selecting cutting tools, and selecting machine tools. In the results, the study established that the proposed algorithm optimizes machining processes and also enhances the original DPP functionality. It is also worth noting that the current study extended the work of Wang (2018), who proposed a data-driven approach for fault detection. In the results, the proposed model was found to reduce manufacturing costs and also improve system reliability, having outperformed other methods of fault detection—on benchmark datasets.

Another study with which this study's findings concur in relation to the area of big data is that which was conducted by Wang (2007). In the latter study, the proposed method involved a novel approach of fusing Game Theory and data mining. Based on the results obtained after testing the model on real-world manufacturing datasets, it was concluded that the proposed approach is superior and that it could be applied to complex engineering system analysis. Lastly, the current study's findings extend the work of Choudhary et al. (2009). In the latter investigation, the objective was to determine the efficacy of using data mining applications in manufacturing systems. Indeed, the investigation culminated into the implementation of a novel text mining technique, with the conclusion demonstrating that in manufacturing systems, which constitute big data, data mining applications, especially through the proposed novel text mining technique, improved efficiency in terms of detecting faults. Therefore, this study extended the work of the aforementioned studies' results and is poised to lay a foundation for steering improvements in the context of computer components manufacturing.

Conclusion

Despite the informative nature of the results, which indicated that the proposed model performs superiorly than the contrast or other comparative methods on both the computer components manufacturing system dataset and benchmark datasets, the study was prone to a few limitations. For instance, the study assumed the existence of a perfect situation in the target computer components manufacturing system. This limitation pointed to the need for future scholarly investigations to consider how the proposed model of big

data-driven CT prediction could perform (when compared to the standard RBFN) in situations marked by manufacturing system damages. Another limitation of the study was that the investigation concentrated on the computer components manufacturing system's attribute of production planning, failing to give insight into how the proposed DP-RBFN framework could perform if implemented on a manufacturing system in the entirety. Despite these limitations, the proposed DP-RBFN model was found to be better placed to determine how real-time could be reduced in computer components manufacturing systems; hence, system efficiency.

Acknowledgements This work is supported by the National Natural Science Foundation of China (61773120). This research is also supported by the Foundation for the Author of National Excellent Doctoral Dissertation of China (2014-92).

References

- Cai, Z., Chen, G., Xing, L., et al. (2019). Evaluating hedge fund downside risk using a multi-objective neural network. *Journal of Visual Communication of Image Representation*, 59, 433–438.
- Choudhary, A. K., Harding, J. A., & Tiwari, M. K. (2009). Data mining in manufacturing: a review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, 20(5), 501.
- Gopalakrishnan, M., Skoogh, A., & Laroque, C. (2013). Simulation based planning of maintenance activities in the automotive industry. In R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, & M. Kuhl (Eds.), *Proceedings of the 2013 winter simulation conference* (pp. 342–353).
- Gröger, C. et al. (2016) The data-driven factory—Leveraging big industrial data for agile, learning and human-centric manufacturing. In *18th international conference on enterprise information systems* (pp. 40–52).
- Gröger, C., Niedermann, F., & Mitschang, B. (2012). Data mining-driven manufacturing process optimization. In *Proceedings of the world congress on engineering* (Vol. III, pp. 0–6). London, UK.
- Guo, Z. X., Ngai, E., Liang, X., et al. (2015). An RFID-based intelligent decision support system architecture for production monitoring and scheduling in a distributed manufacturing environment. *International Journal of Production Economics*, 159, 16–28.
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1–22.
- Jeschke, S., et al. (2017). *Industrial Internet of Things. Springer Series in Wireless Technology*. Cham: Springer.
- Ji, W., Yin, S., & Wang, L. (2019). A big data analytics based machining optimisation approach. *Journal of Intelligent Manufacturing*, 30(3), 1483–1495.
- Jiao, B., et al. (2019). Graph sampling for Internet topologies using normalized Laplacian spectral features. *Information Sciences*, 481, 574–603.
- Lee, J., Lapira, E., Bagheri, B., & Kao, H. A. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters*, 1, 38–41.
- Li, L., Chang, Q., Xiao, G., & Ambani, S. (2011). Throughput bottleneck prediction of manufacturing systems using time series analysis. *Journal of Manufacturing Science and Engineering*, 133, 1–8.
- Lou, S., Feng, Y., Zheng, H., Gao, Y., & Tan, J. (2018). Data-driven customer requirements discernment in the product lifecycle man-

- agement via intuitionistic fuzzy sets and electroencephalogram. *Journal of Intelligent Manufacturing*, <https://doi.org/10.1007/s10845-018-1395-x>.
- Luo, H., Fang, J., & Huang, G. Q. (2015). Real-time scheduling for hybrid flow shop in ubiquitous manufacturing environment. *Computers & Industrial Engineering*, *84*, 12–23.
- Michael, W., Noessler, P., & Metternich, J. (2016). Development of bottleneck detection methods allowing for an effective fault repair prioritization in machining lines of the automobile industry. *Production Engineering*, *10*(3), 329–336.
- Ren, T., Li, S., Zhang, X., & Liu, L. (2017). Maximum and minimum solutions for a nonlocal p-Laplacian fractional differential system from eco-economical processes. *Boundary Value Problems*, *2017*(1), 1–15.
- Ren, T., Xiao, H., Zhou, Z., Zhang, X., Xing, L., Wang, Z., et al. (2019). The iterative scheme and the convergence analysis of unique solution for a singular fractional differential equation from the eco-economic complex system's co-evolution process. *Complexity*, *2019*(3), 1–15.
- Roser, C., & Nakano, M. (2015). A quantitative comparison of bottleneck detection methods in manufacturing systems with particular consideration for shifting bottlenecks. *IFIP International Federation for Information Processing*, *2015*(1), 273–281.
- Shang, X., Lining, X., Ling, W., et al. (2019). Comprehensive learning pigeon-inspired optimization with tabu list. *Science China Information Sciences*, *62*(7), 07028.
- Shao, G. et al. (2015). Data analytics using simulation for smart manufacturing. In A. Tolk, S. Diallo, I. Ryzhov, L. Yilmaz, S. Buckley, & J. Miller (Eds.), *Proceedings—Winter simulation conference* (pp. 2192–2203).
- Subramanian, M., Skoogh, A., Gopalakrishnan, M., et al. (2016). Real-time data-driven average active period method for bottleneck detection. *International Journal of Design & Nature and Ecodynamics*, *11*(3), 428–437.
- Wang, J., Yang, J., Zhang, J., Wang, X., & Zhang, W. (2018a). Big data driven cycle time parallel prediction for production planning in wafer manufacturing. *Enterprise Information Systems*, *12*(6), 714–732.
- Wang, R., Lai, S., Wu, G., et al. (2018b). Multi-clustering via evolutionary multi-objective optimization. *Information Sciences*, *450*, 128–140.
- Wang, T., Qiao, M., Zhang, M., Yang, Y., & Snoussi, H. (2018). Data-driven prognostic method based on self-supervised learning approaches for fault detection. *Journal of Intelligent Manufacturing*, <https://doi.org/10.1007/s10845-018-1431-x>.
- Wang, Y. (2007). Combining data mining and Game Theory in manufacturing strategy analysis. *Journal of Intelligent Manufacturing*, *18*(4), 505–511.
- Wu, G., Mallipeddi, R., & Suganthan, P. N. (2019). Ensemble strategies for population-based optimization algorithms—A survey. *Swarm and Evolutionary Computation*. <https://doi.org/10.1016/j.swevo.2018.08.015>.
- Wu, G., Pedrycz, W., Li, H., et al. (2016). Coordinated planning of heterogeneous earth observation resources. *IEEE Transactions on System Man Cybernetics: System*, *46*, 109–125.
- Wu, G., et al. (2018). Ensemble of differential evolution variants. *Information Sciences*, *423*, 172–186.
- Wuest, T., et al. (2016). Machine learning in manufacturing: Advantages, challenges, and applications. *Production & Manufacturing Research*, *3277*(2017), 1–23.
- Yang, Z.-X., Zhang, P., & Chen, L. (2016). RFID-enabled indoor positioning method for a real-time manufacturing execution system using OS-ELM. *Neuro-Computing*, *174*, 121–133.
- Yi, J.-H., Xing, L.-N., Wang, G.-G., et al. (2019). Behavior of crossover operators in NSGA-III for large-scale optimization problems. *Information Science*. <https://doi.org/10.1016/j.ins.2018.10.005>.
- Zhang, J., Wang, L., & Xing, L. (2019). Large-scale medical examination scheduling technology based on intelligent optimization. *Journal of Combinatorial Optimization*, *37*(1), 385–404.
- Zhang, Y., Xu, J., Sun, S., & Yang, T. (2015). Real-time information driven intelligent navigation method of assembly station in unpaced lines. *Computers & Industrial Engineering*, *84*, 91–100.
- Zhengcai, C., Jijie, D., Min, L., & Yongji, W. (2012). Bottleneck prediction method based on improved adaptive network-based fuzzy inference system (ANFIS) in semiconductor manufacturing system. *Chinese Journal of Chemical Engineering*, *20*(6), 1081–1088.
- Zhong, R. Y., et al. (2015). A two-level advanced production planning and scheduling model for RFID-enabled ubiquitous manufacturing. *Advanced Engineering Informatics*, *29*(4), 799–812.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.