Artificial Intelligence for Engineering Design, Analysis and Manufacturing (2018), *32*, 122–130.
© Cambridge University Press 2017 0890-0604/17
doi:10.1017/S089006041600055X

TECHNICAL BRIEF PAPER

Data-driven process planning for shipbuilding

JINSONG BAO,[1,2] XIAOHU ZHENG,[1,2] JIANGUO ZHANG,[1] XIA JI,[1] AND JIE ZHANG[1]

[1]School of Mechanical Engineering, Donghua University, Shanghai, China
[2]Shanghai Engineering Center of Process and Equipment for Aerospace Devices Manufacturing, Shanghai, China

(RECEIVED December 20, 2015; ACCEPTED October 31, 2016)

Abstract

Erection planning in shipbuilding is a highly complex process. When a process change happens for some reason, it is often difficult to identify how many factors are affected and estimate how sensitive these factors can be. To optimize the planning and replanning of the shipbuilding plan for the best production performance, a data-driven approach for shipbuilding erection planning is proposed, which is composed of an erection plan model, identification of major factors related to the erection plan, and a data-driven algorithm to apply shipbuilding operation data for creating plans and forecasting, for plan adjustment, future availabilities of shipyard resources including machines, equipment, and man power. Through data clustering, the relevant factors are identified as a result of plan change, and critical equipment health management is carried out through data-driven anomaly detection. A case study is implemented, and the result shows that the proposed data-driven method is able to reschedule the shipbuilding plans smoothly.

Keywords: Anomaly Detection; Industrial Big Data; Optimization; Reschedule; Shipbuilding Plan

1. INTRODUCTION

Shipbuilding industry is a typical discrete manufacturing industry featuring small batch production and high-level process complexity. The manufacturing processes of shipbuilding include preprocessing, fabrication, assembly, painting, preerection, and block erection. The block erection process is the final stage of shipbuilding and involves constructing a ship by assembling blocks transported to docks by cranes. However, it is the starting point for production planning and constitutes the basis of the flow of manufacturing plans.

Production planning management is the core of shipbuilding management. In an actual shipbuilding process, it is necessary to carry out shipbuilding standard time estimation and develop a design plan, a production plan, and material requirement plans. The main problem in this process is the frequent changes of the production plan, with the change rate reaching from 30% to 70%. As a result of such frequent changes, the manufacturing performance can be significantly reduced, and sometimes the whole process may fail. Therefore, it is important to carefully devise a production plan control system before the total manufacturing process starts. The system must be able to adjust the plans during the block erection, if necessary, in order to prevent any avoidable losses. To develop such a production plan control system, both factory modeling and manufacturing process visualization are highly effective approaches. Such models for the block erection process should have the power to express and analyze highly complex planning and scheduling processes involving all the relevant factors that may affect the manufacturing process.

Current approaches to process planning in shipbuilding usually cannot identify how many factors are affecting changes and which factors are more sensitive than others. Therefore, the shipbuilding plans are mostly created and managed by humans. Although the human-based approach can take too much time and be erratic, the planning relies on the experience of the human experts in predicting possible future happenings that may affect the plan being developed. The key challenge for developing a systematic approach to shipbuilding planning is how one can predict, with sufficient accuracy, the future availability of resources, including equipment, machines, cranes, locations/spaces, and time slots.

Recent advancement in research on data mining and big data provides a promising approach to predicting the future. The basic idea of this paper has two parts: first, frequent changes of block erection planning can be dealt with properly if one can predict the future availability of all relevant resources; and second, the historical shipbuilding process data of a shipyard can serve as a big data set for predicting

Reprint requests to: Xiaohu Zheng, School of Mechanical Engineering, Donghua University, Number 2999 North Renmin Road 201620, Songjiang District, Shanghai, China. E-mail: xhzheng@dhu.edu.cn

https://doi.org/10.1017/S089006041600055X Published online by Cambridge University Press

future availabilities of all relevant resources. The following sections described the details that elaborate this basic idea.

In recent years, the research of industrial big data has become a hot spot in many industrial areas. Data collection and data analyses are both important aspects of data mining. Deutsch (2014) pointed out that data collection infrastructure-centric technology are growing faster than data analytic models. Knowledge discovery method is the key aspect in big data technology. The algorithms about knowledge-extraction involve artificial intelligence, machine learning, and data visualization (Choudhary et al., 2008). In order to improve the use of big data technology in manufacturing industry, researchers should pay much attention to data collection, processing, cleaning, transformation, and decision making based on big data (Stockton et al., 2012). The more valuable data collection and analysis will lead to more valuable decisions (Jain et al., 2011). The current data-related challenges in manufacturing systems are poor-quality data and lack of data analysis models (Wang et al., 1995; Lee et al., 2014). Visualization is an efficient method for decision making (Esmaeilian et al., 2016). Three-dimensional display technology, immersive computing technologies, and virtual reality technology will provide great support to the data visualization and manufacturing plan (Kim et al., 2006).

Jin et al. (2015) made a survey of the big data, and pointed out the prospects of big data in data complexity, computational complexity, and system complexity. The accuracy of big data model is studied by Ateca-Amestoy and Prieto-Rodriguez (2013). The research of big data in the manufacturing industry focused on using different algorithms to optimize scheduling to assess the program. Basak and Albayrak (2014) proposed a real-time scheduling and control decision model based on object-oriented Petri net for automobile manufacturing. However, this algorithm is not applicable for batch mode adaptation and discrete manufacturing. Lin and Chen (2015) studied the scheduling algorithm for mixed semiconductor assembly lines, and evaluated the performance of that algorithm with a simulation method. However, this research only considered scheduling simulation but did not mix the plans together effectively. Azab and Naderi (2015) proposed a scheduling and modeling method for configurable manufacturing system, which is based on a small-scale model. Lee et al. (2013) recently forecasted the current research progress and the tendency on the application of the big data technology in the manufacturing system. Huang et al. (2015) conducted the relevant research on the supply chain and logistic scheduling based on the use of radiofrequency identification.

Hey (2012) pointed out that the current data volume is increasing, the data structure is more complex, and the traditional experiment (empirical), theory (theoretical), and computing (computational) of the three paradigms have become difficult to fit new research areas. He put forward the "fourth data" (the fourth paradigm). This method, based on the data exploration, uses small models, guiding research for using "small model, big data."

Shipbuilding planning is complex. With the continuous development of information systems such as enterprise resource planning and product life cycle management, the management mode is gradually transferred to the fine management. Especially in the manufacturing process, the accumulation of a large amount of data or information during the shipbuilding process can be used to determine the core factors of the ship manufacturing plan and to help designers and managers to make decisions so as to make continuous improvement in the manufacturing process possible.

In the following, the problem statement is given in Section 2, followed by erection plan modeling in Section 3. Section 4 describes a proposed data-driven approach to erection planning, and a case study is presented in Section 5. Finally, concluding remarks are drawn in Section 6.

2. PROBLEM STATEMENT

2.1. The process of shipbuilding

There are three types of plan in shipbuilding: long-term plan, middle-term plan, and short-term plan. The long-term plan (shipbuilding master plan) gives the deadline of key stages, such as deliver date of ship, completion time of ship design, makespan of main equipment, and so on. Those milestones are not changeable normally. Therefore, it is the primary coordinate for various detailed plans and for implementation of shipbuilding process control. The middle-term plan (block, dock erection, and outfitting construction plan) involves manufacturing main processes of ship construction. The short-term plan (weekly plan) is developed based on the long-term and middle-term plans, and it schedules all detailed manufacturing activities. Approximately 50% of ship construction time is occupied by the erection activities (gray area in Fig. 1).

To provide effective and efficient support for shipbuilding, an adaptive shipbuilding planning system must be able to generate block erection sequences that do the following:

1. Minimize the total time and cost for the erection process and guarantee the assembly accuracy and smoothness.
2. Identify and use manual labors and machines efficiently.
3. Avoid needs for unplanned maintenance. Failing to know certain manufacturing conditions often produces unplanned maintenance, which can be so expensive that it costs around $200 billion in the United States in 1990 (Baydar & Saitou, 2001).

2.2. Major factors of erection plan

Shipbuilding plans are the main basis for both the production and production management in shipyards. There are many uncertain factors in the implementation of the shipbuilding plans. Any change with any of these factors may lead to significant plan delays. An analysis of the main factors affecting the building plans of a ship is shown in Table 1.

All of the factors listed in Table 1 have a serious impact on the shipbuilding plan. The lean production of a ship can be realized if the historical data concerning these factors can be applied to model, analyze, and forecast the building plans and associated man power, equipment, and machine availability changes.

https://doi.org/10.1017/S089006041600055X Published online by Cambridge University Press

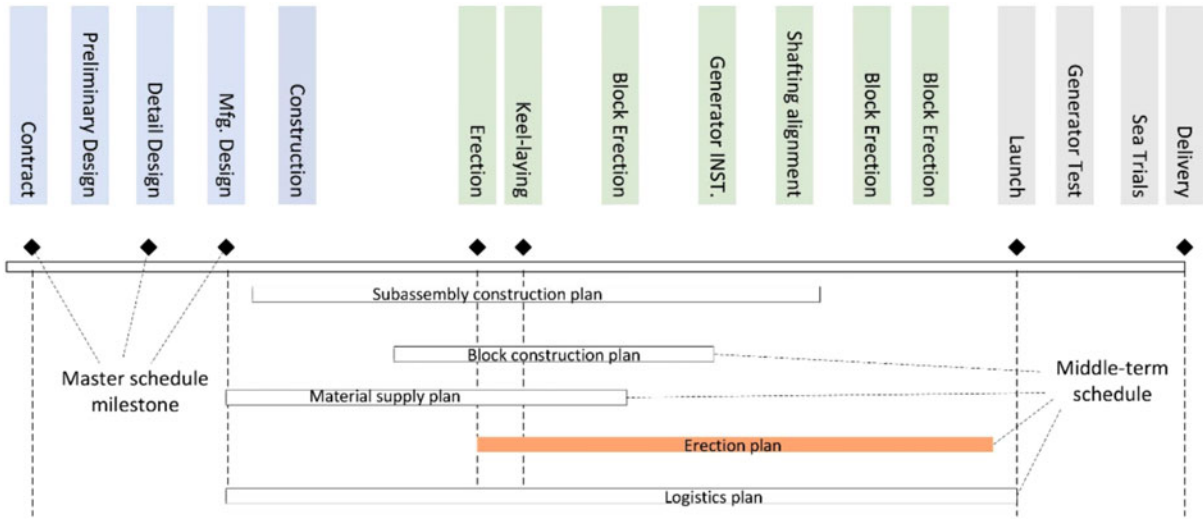


Fig. 1. Plans and primary activities for single shipbuilding.

Table 1. *Relative major factors*

Content	Parameters	Weight	Freq.	Type	E/I
Manufacturing capacity	Device capacity	High	Low	TS	E
	Logistic capacity	High	Low	S	E
Admission date (primary devices)	Engine and shafts	High	Low	TS	E
Prior block availability	WIP percentage complete	Middle	High	S	I
	Rework or correction	High	High	TS	I
Equipment availability	Busy or idle	Middle	High	S	E
	Failure or maintenance		High	S	I
Yard space availability	Position, weight of block	High	High	S	I
Human-resource availability	Number, work shift	Low	High	S	E
Weather allowance	Temperature & weather	Middle	Low	TS	E
ECN	Process & sequences	High	Low	S	I
Work plan	Number of plan & change	High	High	S	E

Note: TS, Time series; S, static; E, explicit; I, implicit; WIP, workpiece in progress; ECN, engineering change notice.

3. ERECTION PLAN MODELING

One major problem with the current shipbuilding planning is the mismatch between the utilization ratio of the dock and the availability of yard and manufacturing equipment. The unreasonable equipment scheduling for yard, dock, and hull closure has resulted in many "free operations," taking a long time as a result of the delay of the plan. Figure 2 illustrates a layout of the restriction map for hull layout and the order of lifting segments, including erection devices (e.g., cranes) and storages. Erection devices include two types: gantry crane and portal crane. Storages include the field of dock and the platform near the dock. Figure 2 shows an example of the resources of a dock. The status of the cranes (gantry crane: M1; portal cranes: M2, M3) may be busy or idle and the status of the storages (dock field: B0–B3; platform field: A0–A3) may be occupied or empty.

The ship segments, built at the platform, were taken to the yard and assembled into a total section. Finally, these segments were lifted to the dock for hull closure. The platform is divided into seven zones. The yard is divided into three districts, composed of four cranes, two plate transport vehicles, seven operating zones, and the yard's three operation areas. This paper focuses on erection plan. An abstraction of the weekly erection plan is illustrated in Figure 3.

Plans ($1 \sim N$) are made by different erection processes. Each plan has some changeable subplans, and its duration is defined as $t_{i,j}$. These activities are often constrained by sets of resources ($R_{i,j}$) and precedence orders. In addition, the subplans are also limited by the master or middle-term schedules. The planning problem is defined in this paper as to minimize Eq. (1) subject to Eqs. (2) through (5):

$$Z = \sum_{i=1}^{n} w_i C_i, \tag{1}$$

$$\forall_{1 \le t \le S} \quad \sum_{i=1}^{n} O_{it} \le M, \tag{2}$$

https://doi.org/10.1017/S089006041600055X Published online by Cambridge University Press

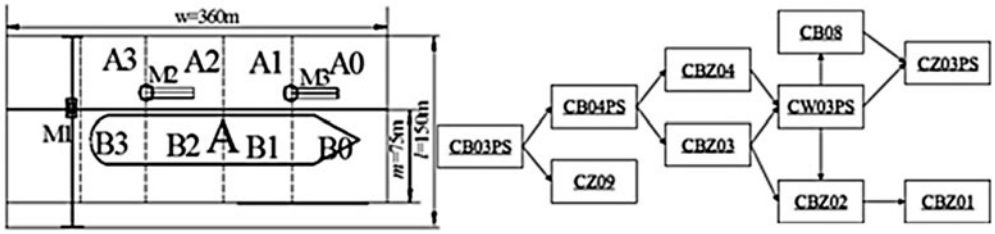


Fig. 2. Diagram of shipyard layout and its resource, precedence sequence.

$$\forall_{1\le i\le n}\ \forall_{j\in D_i} \quad t_j + p_j \le t_i, \tag{3}$$

$$S = \max\{t_i, (1 \le i \le n)\}, \tag{4}$$

$$t_i, t_j \in Z^+ \ (1 \le i, j \le n), \tag{5}$$

where Z is the total completion time considering every plan's weight factor, w_i is the weight of plan I, C_i is the completion time of plan I, n is the total number of plans, O_{it} is the total number of devices that are occupied by plan i in time t, M is the total number of devices, t_i is the starting time of plan I, p_j is the processing time of plan j ($j \in D_i$), and S is the probable completion time of the final plan.

The objective of scheduling is to minimize the weighted total completion time Z, as shown in Eq. (1). Every block is assigned a weighting factor according to its importance. One machine can process only one task at any time. Some blocks need only one machine to complete, while some blocks require two machines simultaneously. The total number of occupied machines at particular time t cannot exceed the total capacity of all machines M. The crane machine capacity constraints are represented by Eq. (2). A task cannot be started until its direct preceding tasks are finished, and a task requires a specific amount time for processing. The precedence constraints are the precedence of different blocks to be processed. Equation (3) is the formulation of task precedence constraints. Here, t_i is the decision variable; S, C_i, and O_{it} are the dependent variables. Equation (5) indicates that the start time of a task is a positive integer.

4. A DATA-DRIVEN APPROACH

4.1. Data in the shipbuilding erection plan

The smooth implementation of the established production plan not only depends on the material, equipment, labor, location, capital, and information and other resources, but also depends on how to configure and use these resources. A scientific method is adopted to optimize the operation plan of the

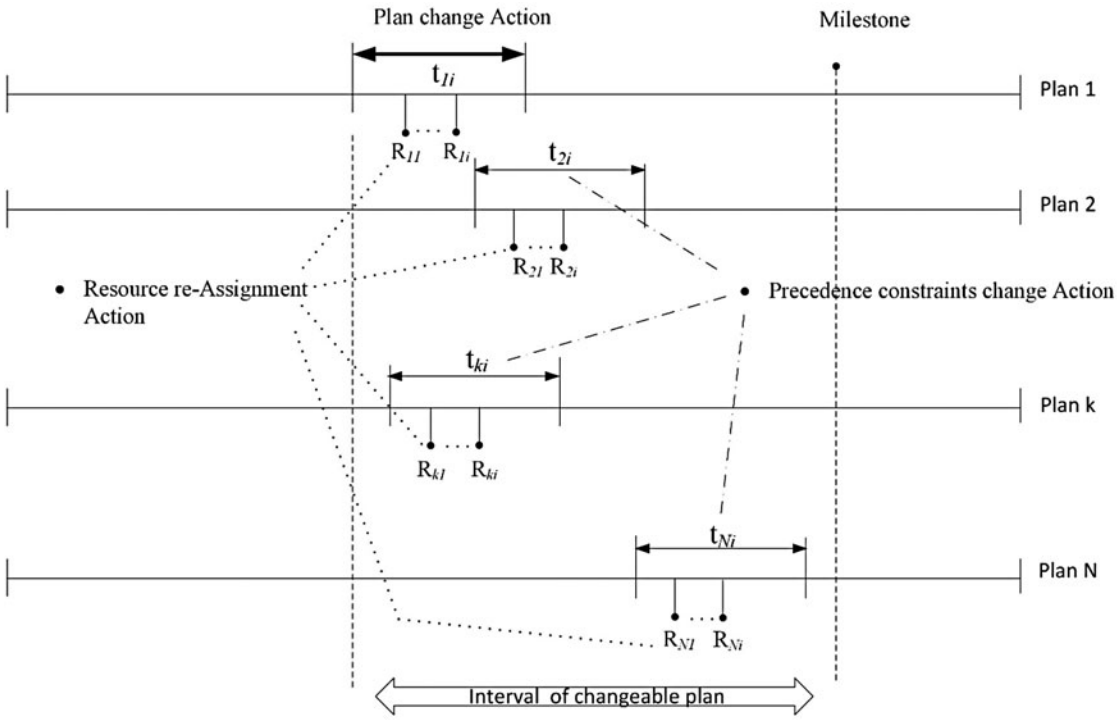


Fig. 3. Definition of erection plan.

https://doi.org/10.1017/S089006041600055X Published online by Cambridge University Press

building, and the work plan is decomposed and transformed into efficient daily production activities.

There are two major types of data in shipbuilding: contextual data and content. Content is the data related to the ship manufacturing concerning management and design procedures. Contextual data is more critical, and collected from the manufacturing process, product-related manufacturing equipment, testing equipment, and auxiliary equipment, as shown in Table 2.

The shipbuilding industry is different from other manufacturing industries due to its manufacturing chain length. The monitoring and forecasting of abnormal condition of equipment is to collect the operation status data and maintenance a log for the key equipment (crane, welding machine, cutting machine, cutting machine, etc.), find out the mode of the abnormal occurrence of the equipment, monitor and predict the future failure probability, and then plan the future, taking into account the effect of the repair time on the schedule. These data are generated each second and grows massive in size after a long period.

Unstructured data is raw and unorganized, and organizations store all of it. Not all types of unstructured data can easily be converted into a structured model. Only 10%–20% of context-relevant information originates in an unstructured form, primarily monitor videos, pictures, and geometric models. Between 60% and 70% of content information is structured data, which are stored in RDMS or other NoSQL. Meanwhile, the balance is semistructured data, such as process logs. In this research, all of the information is organized into structured data using nested *key-value* pairs. A typical document is as follows:

{
id: 1,
name: "Crane 10-3,"
location: 10.125L,45.023W
specs: {
manufacturer: "ZPMC,"
capacity: 160T,
speed: 1.6,
Weight: 150
},
Description: "1# Crane,"
Runtime: 42.2h,
Current status: Occupied
Last HM Date:2015-2-5 12:34
}

Table 2. *Data set of shipbuilding erection*

Catalog	Data Source & Characters
Content	Design data: CA*x* and simulations, etc.
	Management data: ERP, PDM, CRM, etc.
Context	SCADA: cutting tools, devices status, sensors, smart devices, etc.
	Collected by MES: manufacturing process, NC, machines, robots, etc.
	Process control: job site control, product status information
	Log: work plan log, device warning log, etc.
	Real-time communication: connected devices

4.2. Data-driven system

Figure 4 is an illustration of the data-driven system developed to deal with the main problems that the shipbuilding industry is facing with erection planning. The historical shipbuilding data is analyzed, organized, and utilized in the process of generating the shipbuilding plan. The process includes the following: generation and loading of structured, unstructured, or semistructured data; data cleaning, data flow organizing, compression, and storage to Hadoop or NoSQL; MapReduce flow data analysis and calculation (Moseley, 2011); data utilization; and data visualization and decision-making support.

4.3. Clustering based on K-means

The big data generated by the shipbuilding process is composed of real-time data that represent the status of the whole shipbuilding system. These data are generated by the sensors of the system and then converted to the information of the various components. By processing these data, the working state of the shipbuilding system is identified, and the future status of the shipbuilding machines and facilities is forecasted by comparing the real-time data with the historical data. The affected factors related to the change of shipbuilding plan are discovered through the K-means clustering method. If the sample set is divided into C categories, the algorithm can be described as follows:

1. selection of initial centers for C classes properly;
2. in the k iteration, for any samples, calculate the distance to the C center, and put the sample to the shortest distance of the center;
3. to update the central values of the class by means of the mean value; and
4. for all C clustering centers, if the value remains unchanged after using the iterative method, then the stop the iteration, otherwise continue the iteration.

More details can be found in the case study section.

4.4. Health management based anomaly detection

Three broad categories of anomaly detection techniques exist. Unsupervised anomaly detection techniques detect anomalies in an unlabeled test data set under the assumption that the majority of the instances in the data set are normal by looking for instances that seem to fit least to the remainder of the data set. Supervised anomaly detection techniques require a data set that has been labeled as "normal" and "abnormal" and involves training a classifier (the key difference to many other statistical classification problems is the inherent unbalanced nature of outlier detection). Semisupervised

https://doi.org/10.1017/S089006041600055X Published online by Cambridge University Press

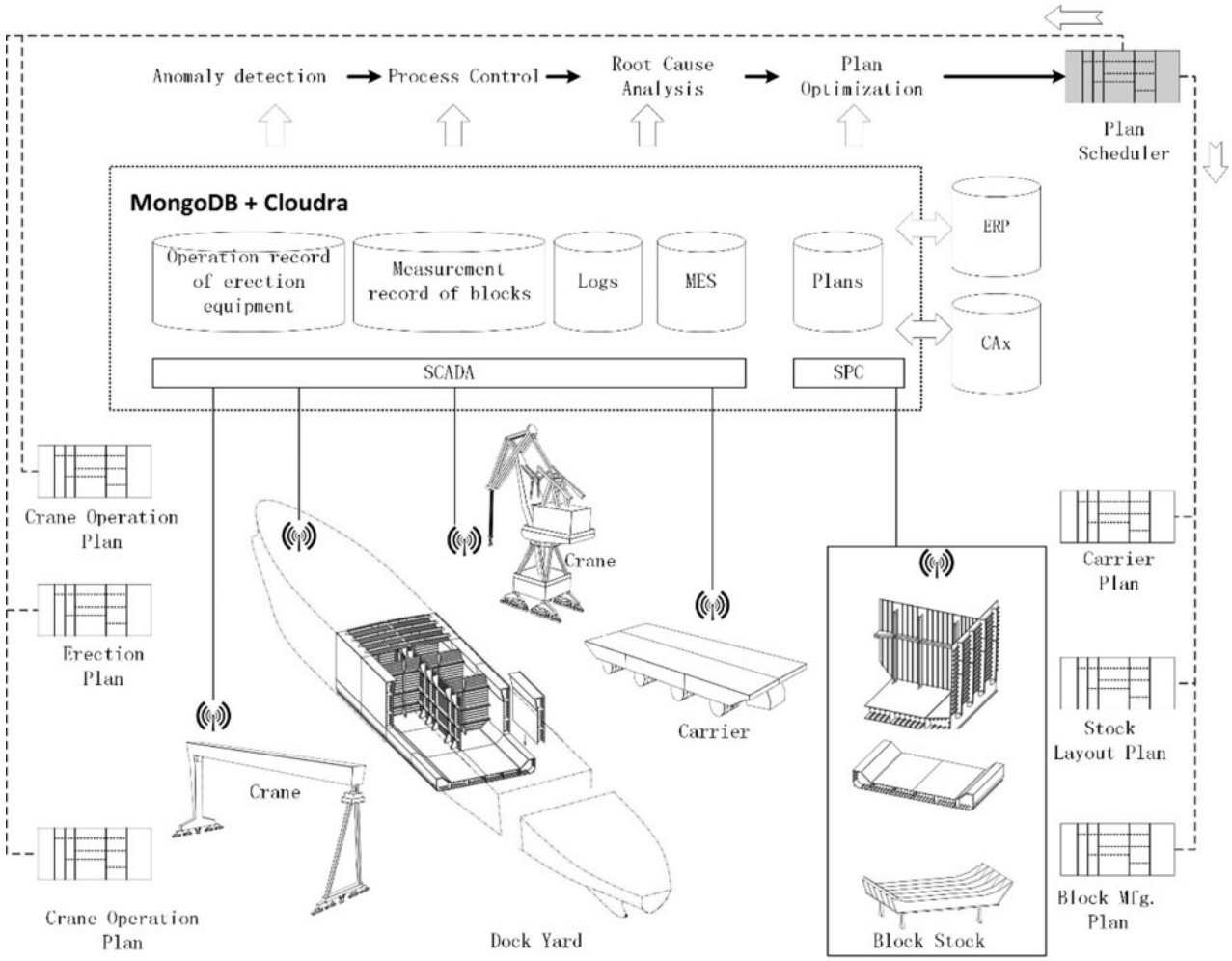


Fig. 4. A data-driven system for shipbuilding manufacturing planning.

anomaly detection techniques construct a model representing normal behavior from a given normal training data set, and then testing the likelihood of a test instance to be generated by the learned model. This paper adopted a simple method: distance-based method. There are two steps in this method:

Step 1. Compute neighborhood for each data record.

Step 2. Analyze the neighborhood to determine whether data record is an anomaly.

5. CASE STUDY

Six months of related manufacturing data are gathered continuously, stored in MongoDB and Cloudra. The total data volume is about 1865GB for erecting 53 blocks of a 53xx TEU. The erection network and precedence sequence of 5xxx TEU is shown in Figure 5. As mentioned above, there are two types of data, as shown in Table 3: content data and contextual data.

As described in Section 4.3, the K-means algorithm was used to discover the affected factors related to the change of the shipbuilding plan. However, the variables of erection plan are usually high-dimensional data. Therefore, it is important to find a way to reduce its dimension. Principal component analysis (PCA) is a technique to emphasize variation and bring out strong patterns in a data set. It is often used to make data easy to explore and visualize. The PCA method improves K-means clustering solutions and the cluster structures are embedded in the first K – 1 principal components. In this study, the adopted method runs the PCA algorithm many times with similar plans and provides visualization of the clustering structure of the obtained components. Furthermore the K-means algorithm was conducted to carry out factor clustering.

Defining the ith principal component (like a new variable or column) as PC_i, it can be calculated by Eq. (6):

$$\mathrm{PC}_i = \sum_{j=1}^{p} (a_{ij} X_j), \tag{6}$$

where X_j is the ith origin variable, $PC_i = a_i^{\mathrm{T}} \Sigma a_i$, such that $a_i^{\mathrm{T}} a_i = 1$, where Σ is the covariance matrix. PC_1 is the first principle component.

https://doi.org/10.1017/S089006041600055X Published online by Cambridge University Press

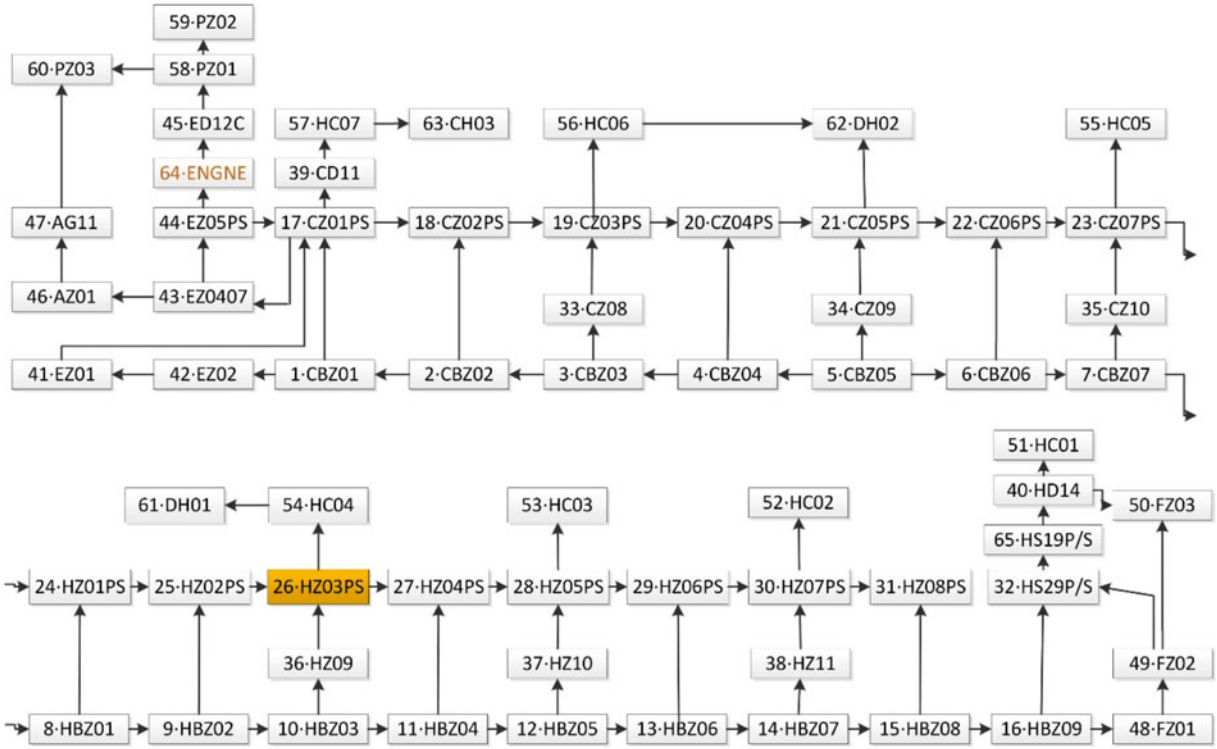


Fig. 5. Erection network and precedence sequence of 5xxx TEU ship.

Table 3. *Details of data*

Data Source	Format or System	Freq (s)	Volume (G)
CAD	Tribon	1	26
CAE	Ansys	1	129
ERP	SAP	1	380
PLM	PTC Windchill	1	76
SCADA			
Cutting devices	BSON	1 Hz	30 (4 devices)
Welding devices	BSON	2 Hz	96 (42 devices)
Cranes	Videos + BSON	0.5 Hz	106 (6 cranes)
Carriers	Videos + BSON	0.5 Hz	75 (7 carriers)
Blocks stock	Videos + BSON	0.1 Hz	146 (4 stocks)
Worker tracker	BSON	2 Hz	76 (1250 workers)
Ship yard monitor	Videos	1/24s	12 (retain 1 week)
painting devices	Videos + BSON	1 Hz	13 (7 devices)
RFID on blocks	BSON	0.1 Hz	86 (53 blocks)
Logistics of mfg.	BSON	1 Hz	267 (8 workshops)
Workshops monitor	Videos + Image	1/24 s	151
Process control: job site control, product status information	BSON	0.01 Hz	21
Log: work plan log, device warning log, etc.	BSON + .log text	1 Hz	6
SPC	BSON + images + PLY	—	146 (12,965 check point)
Intermediate result	BSON, images, DB files	—	23

In general, it essentially rotates the data (via an N-dimensional rotation matrix) in such a way to maximize the variance along the new axis $PC_1, PC_2, \ldots, PC_N$, sorted from largest to smallest.

The calculation process consists of the following steps:

STEP 1. Parameters for the PCA estimation algorithm are selected.

https://doi.org/10.1017/S089006041600055X Published online by Cambridge University Press

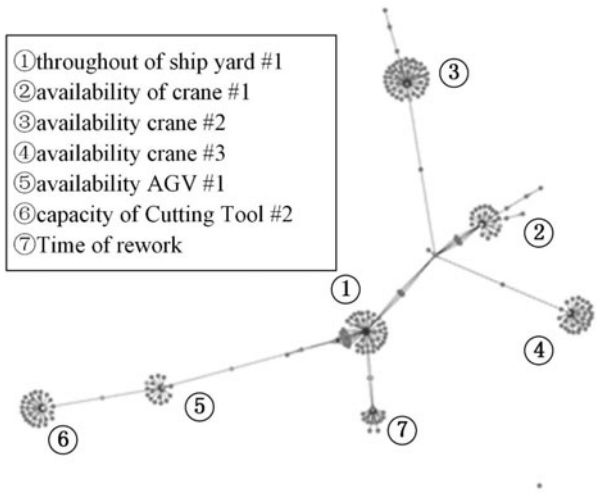


Fig. 6. Sensitive factors identified by clustering and visualization of Plan 26.

STEP 2. Translate the data so that the center is at the origin.

STEP 3. Calculate the covariance matrix and distance using K-means.

STEP 4. Find the principal components after clustering computing.

STEP 5. Reduce the data using the selected principal components.

STEP 6. Visualize by nonlinear two-dimensional projection (multidimensional scaling suggested by Gordon, 1987) using R.

Taking the Block HZ03-PS erection plan, for example, during processing this plan (Plan 26, orange node in Fig. 5), cranes' availabilities and other parameters are collected. Points in the diagram are nodes (including resources required and its weight) defined in the erection network, and edges are precedence relations in the plan. The availability of equipment and other factors are highlighted using a clustering method in our data-driven planning system with visualization

Table 4. *Erection plan execution by block*

Plan Execution for 53xx TEU With Big Data			Plan Execution for Same Type Ship 52xx TEU With Manual		
Block ID	Complete Time	Planned Time	Block ID	Complete Time	Planned Time
EZ11PS	53	54	EZ11PS	74	59
EDZ11PS	54	56	EDZ11PS	69	62
HZ11PS	58	60	HZ11PS	78	58
HZ12PS	60	60	HZ12PS	79	72
HBZ01	61	62	HBZ01	86	54
HZ13PS	62	62	HZ13PS	87	65
HBZ02	64	65	HBZ02	88	70
ABZ02	64	61	ABZ02	88	67
AZ13PS	66	65	AZ13PS	92	82
AZ12PS	72	74	AZ12PS	93	70

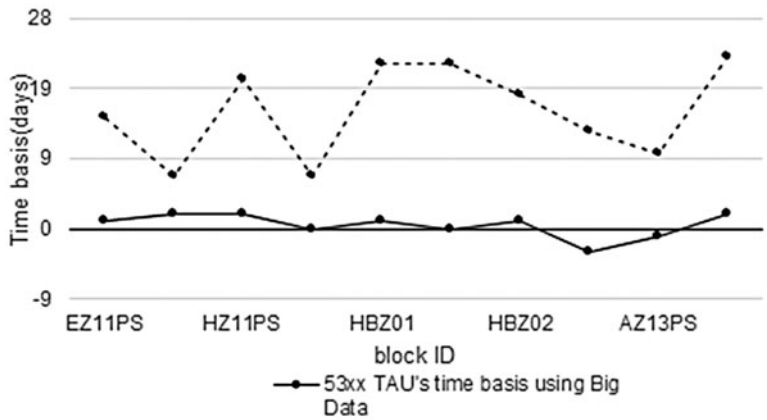


Fig. 7. Time basis between manual and using big data.

https://doi.org/10.1017/S089006041600055X Published online by Cambridge University Press

shown in Figure 6. Meanwhile the device failure detections were taken into consideration in plan adjustment.

To demonstrate the advantage of the data-driven planning approach, the result of the system was compared to the result with the 10 blocks' schedule attainment of a similar ship (52xx TEU) with the same erection network. The comparison is shown in Table 4.

In Figure 7, the plan execution of Ship 52xx (broken lines in Fig. 7) was manually adjusted when the plan is changed, while the data-driven planning-based adjustment is automated by the system (solid lines in Fig. 7). As indicated in Figure 7, data-driven planning-based adjustment yields a much smaller range of time variation, less than 5 days, when compared with the manual adjustment result.

6. CONCLUSIONS

With the development of industrial big data technology, the rapid growth of semistructured and nonstructured data, the analysis and application of industrial data give more production and management support for realizing "intelligent shipbuilding." This study proposed a data-driven approach for shipbuilding erection planning, and the planning and replanning of the shipbuilding plan for the best production performance were optimized. Through data clustering, the relevant factors are identified as a result of plan change, and critical equipment health management is carried out through data-driven anomaly detection. In the course of this research, two major challenges were addressed. First, the complex shipyard situation and complex block erection scheduling had made it difficult to do pure data mining. Utilizing PCA has led to reduced dimension and visualization (human in the loop), and a blended PCA plus K-Means approach has been proven effective to carry out factor clustering tasks in this domain. Second, the complex erection plan is composed of many changeable subplans, which made it difficult to build a proper process model for planning. Introducing the weighted total completion time Z has simplified the problem and made the objective of scheduling to minimize Z.

The approach described in this technical brief paper still relies on human experience and human decisions. Future work will apply machine learning techniques to acquire human knowledge and reduce the need for human involvement.

ACKNOWLEDGMENT

This research was supported by the National Science Foundation of China (51475301, 51435009), Shanghai Leading Academic Discipline Project B602, and the Fundamental Research Funds for the Central Universities.

REFERENCES

Ateca-Amestoy, V., & Prieto-Rodriguez, J. (2013). Forecasting accuracy of behavioural models for participation in the arts. *European Journal of Operational Research 229(1)*, 124–131.

Azab, A., & Naderi, B. (2015). Modelling the problem of production scheduling for reconfigurable manufacturing systems. *Procedia CIRP 33*, 76–80.

Başak, O., & Albayrak, Y.E. (2014). Petri net based decision system modeling in real-time scheduling and control of flexible automotive manufacturing systems. *Computers & Industrial Engineering 86*, 116–126.

Baydar, C.M., & Saitou, K. (2001). Automated generation of robust error recovery logic in assembly systems using genetic programming. *Journal of Manufacturing Systems 20(1)*, 55–68.

Choudhary, A.K., Harding, J.A., & Tiwari, M.K. (2008). Data mining in manufacturing: a review based on the kind of knowledge. *Journal of Intelligent Manufacturing 20(5)*, 501–521.

Deutsch, T. (2014). *Analytics Hype: The Next Wave in Big Data Backlash*. Accessed at http://www.ibmbigdatahub.com/blog/analytics-hype-next-wave-big-data-backlash

Esmaeilian, B., Behdad, S., & Wang, B. (2016). The evolution and future of manufacturing: a review. *Journal of Manufacturing Systems 39*, 79–100.

Gordon, A.D. (1987). A review of hierarchical classification. *Journal of the Royal Statistical Society 150(2)*, 119–137.

Hey, T. (2012). *The Fourth Paradigm—Data-Intensive Scientific Discovery*. Berlin: Springer.

Huang, G.Q., Zhong, R.Y., & Tsui, K.L. (2015). Special issue on "big data for service and manufacturing supply chain management." *International Journal of Production Economics 165*, 172–173.

Jain, S., Triantis, K.P., & Liu, S. (2011). Manufacturing performance measurement and target setting: a data envelopment analysis approach. *European Journal of Operational Research 214(3)*, 616–626.

Jin, X., Wah, B.W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research 2(2)*, 59–64.

Kim, Y.-S., Yang, J., & Han, S. (2006). A multichannel visualization module for virtual manufacturing. *Computers in Industry 57(7)*, 653–662.

Lee, J., Kao, H.-A., & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia CIRP 16*, 3–8.

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(1)*, 38–41.

Lin, J.T., & Chen, C.M. (2015). Simulation optimization approach for hybrid flow shop scheduling problem in semiconductor back-end manufacturing. *Simulation Modelling Practice & Theory 51*, 100–114.

Moseley, B., Dasgupta, A., Kumar, R., & Sarlós, T. (2011). On scheduling in map-reduce and flow-shops. *Proc. 23rd Annual ACM Symp. Parallelism in Algorithms and Architectures*, pp. 289–298. New York: ACM.

Stockton, D.J., Khalil, R.A., & Mukhongo, L.M. (2012). Cost model development using virtual manufacturing and data mining: part II—comparison of data mining algorithms. *International Journal of Advanced Manufacturing Technology 66(9–12)*, 1389–1396.

Wang, R.Y., Storey, V.C., & Firth, C.P. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering 7(4)*, 623–640.

Jinsong Bao is an Associate Professor at Donghua University. He received his PhD degree at Shanghai Jiaotong University. Dr. Bao's research interests are in intelligent manufacturing systems and computer-integrated manufacturing systems.

Jianguo Zhang is a Lecturer at Donghua University. He received his PhD degree at Shanghai Jiaotong University in 2014. Dr. Zhang's research interests are in intelligent manufacturing and coated tools.

Xia Ji is a Lecturer at Donghua University. He received his PhD degree at Shanghai Jiaotong University in 2014. Dr. Ji's research interests are in intelligent manufacturing and green manufacturing technology.

Xiaohu Zheng is a Lecturer at Donghua University. He received his PhD degree at Shanghai Jiaotong University. Dr. Zheng's research interests are in intelligent manufacturing systems and precision manufacturing technology.

https://doi.org/10.1017/S089006041600055X Published online by Cambridge University Press