Complexity Problems Handled by Big Data Technology

Hindawi
Complexity
Volume 2019, Article ID 9090528, 7 pages
https://doi.org/10.1155/2019/9090528

Editorial

Complexity Problems Handled by Big Data Technology

Zhihan Lv[1], **Kaoru Ota**[2], **Jaime Lloret**[3], **Wei Xiang**[4] and **Paolo Bellavista**[5]

[1]*Qingdao University, Qingdao, China*
[2]*Muroran Institute of Technology, Hokkaido, Japan*
[3]*Universitat Politècnica de València, València, Spain*
[4]*James Cook University, Douglas, Australia*
[5]*Università di Bologna, Bologna, Italy*

Correspondence should be addressed to Zhihan Lv; lvzhihan@gmail.com

Received 29 October 2018; Accepted 30 October 2018; Published 9 January 2019

Big data needs new processing modes to own stronger decision-making power, insight discovery, the large volume and high growth of process optimization ability, and the diversified information assets. As the information technology of a new generation based on Internet of Things, cloud computing, and mobile internet, big data realizes the record and collection of all data produced in the whole life cycle of the existence and evolutionary process of things. It starts from the angle of completely expressing a thing and a system to express the coupling relationship between things. When the data of panorama and whole life cycle is big enough and the system component structure and the static data and dynamic data of each individual are recorded, the big data can integrally depict the complicated system and the emerging phenomena.

Viktor Mayer-Schönberger proposed the transformation of three thoughts in the big data era: it is not random samples but the whole data; it is not accuracy but complexity; and it is not causality but correlativity. "The whole data" refers to the transformation from local to overall thought, taking all data (big data) as analysis objects. "Complexity" means to accept the complexity and inaccuracy of data. The transformation from causality to correlativity emphasizes more on correlation to make data itself reveal the rules. It is closely related to the understanding of things by complex scientific thinking, which is also the integral thinking, relational thinking, and dynamic thinking.

The analysis technology of big data is the key to exploring the hidden value in the big data. The traditional scientific analysis method records the samples of the thing statuses, which is a method of small data, and perceives things based on small sample data, mathematical induction, and logical induction. But such a method cannot effectively solve complexity problems. In the big data era, the quantitative data description of complex huge system is no longer the mere experimental sample data but the full scene data of the overall state. Therefore, data analysis should adopt complex scientific intelligent analysis method for modeling and simulating, utilize and constantly optimize big data for machine learning, and analyze and study the self-organizing and evolving rules of complex systems.

In total, 35 papers were submitted to this special issue, 27 of which were accepted and published, constituting a 77% acceptance rate. The published papers addressed the following:

Burst growing IoT and cloud computing demands exascale computing systems with high performance and low power consumption to process massive amounts of data. Modern system platforms based on fundamental requirements encounter a performance gap in chasing exponential growth in data speed and amount. To narrow the gap, a heterogamous design gives us a hint. A Network-on-Chip (NoC) introduces a packet-switched fabric for on-chip communication and becomes the de facto many-core interconnection mechanism; it refers to a vital shared resource for multifarious applications which will notably affect system energy efficiency. Among all the challenges in NoC, unaware application behaviors bring about considerable congestion,

which wastes huge amounts of bandwidth and power consumption on the chip. In the paper titled "Hybrid Network-on-Chip: An Application-Aware Framework for Big Data," the authors propose a hybrid NoC framework, combining buffered and bufferless NoCs, to make the NoC framework aware of applications' performance demands. An optimized congestion control scheme is also devised to satisfy the requirement in energy efficiency and the fairness of big data applications. The authors use a trace-driven simulator to model big data applications. Compared with the classical buffered NoC, the proposed hybrid NoC is able to significantly improve the performance of mixed applications by 17% on average and 24% at most, decrease the power consumption by 38%, and improve the fairness by 13.3%.

With the diversification of pit mine slope monitoring and the development of new technologies such as multisource data flow monitoring, normal alert log processing system cannot fulfil the log analysis expectation at the scale of big data. In order to make up for this disadvantage, the paper titled "Ensemble Prediction Algorithm of Anomaly Monitoring Based on Big Data Analysis Platform of Open-Pit Mine Slope" will provide an ensemble prediction algorithm of anomalous system data based on time series and an evaluation system for the algorithm. This algorithm integrates multiple classifier prediction algorithms and proceeds classified forecast for data collected, which can optimize the accuracy in predicting the anomaly data in the system. The algorithm and evaluation system is tested by using the microseismic monitoring data of an open-pit mine slope over 6 months. Testing results illustrates prediction algorithm provided by this research can successfully integrate the advantage of multiple algorithms to increase the accuracy of prediction. In addition, the evaluation system greatly supports the algorithm, which enhances the stability of log analysis platform.

The paper titled "Investigation of Speed Matching Affecting Contrarotating Fan's Performance Using Wireless Sensor Network including Big Data and Numerical Simulation" describes the investigations performed to better understand two-stage rotor speed matching in a contra-rotating fan. In addition, this study develops a comprehensive measuring and communication system for contrarotating fan using ZigBee network. The investigation method is based on three-dimensional RANS simulations; the RANS equations are solved by the numerical method in conjunction with a SST turbulence model. A wireless measurement system using big data method is first designed, and then a comparison is done with experimental measurements to outline the capacity of the numerical method. The results show that when contra-rotating fan worked under designed speed, performance of two-stages rotors could not be matched as the designed working condition was deviated. Rotor 1 had huge influences on flow rate characteristics of contrarotating fan. Rotor 2 was influenced by flow rates significantly. Under large-flow-rates condition, the power capability of rotor 2 became very weak; under working small-flow-rates condition, overloading would take place to class II motor. In order to solve the performance mismatch between two stages of CRF under nondesigned working conditions, under small-flow-rates condition, the priority shall be given to increase of the speed of rotor 1, while the speed of rotor 2 shall be reduced appropriately; under large-flow-rates condition, the speed of rotor 1 shall be reduced and the speed of rotor 2 shall be increased at the same time.

In the paper titled "Research on the Effect of DPSO in Team Selection Optimization under the Background of Big Data," team selection optimization is the foundation of enterprise strategy realization; it is of great significance for maximizing the effectiveness of organizational decision-making. Thus the study of team selection/team foundation has been a hot topic for a long time. With the rapid development of information technology, big data has become one of the significant technical means and played a key role in many researches. It is a frontier of team selection study by the means of combining big data with team selection, which has the great practical significance. Taking strategic equilibrium matching and dynamic gain as association constraints and maximizing revenue as the optimization goal, the Hadoop enterprise information management platform is constructed to discover the external environment, organizational culture, and strategic objectives of the enterprise and to discover the potential of the customer. And, in order to promote the renewal of production and cooperation mode, a team selection optimization model based on DPSO is built. The simulation experiment method is used to qualitatively analyze the main parameters of the particle swarm optimization in this paper. By comparing the iterative results of genetic algorithm, ordinary particle swarm algorithm, and discrete particle swarm algorithm, it is found that the DPSO algorithm is effective and is preferred in the study of team selection with the background of big data.

In the paper titled "Intelligent Method for Identifying Driving Risk Based on V2V Multisource Big Data," risky driving behavior is a major cause of traffic conflicts, which can develop into road traffic accidents, making the timely and accurate identification of such behavior essential to road safety. A platform was therefore established for analyzing the driving behavior of 20 professional drivers in field tests, in which overclose car following and lane departure were used as typical risky driving behaviors. Characterization parameters for identification were screened and used to determine threshold values and an appropriate time window for identification. A neural network-Bayesian filter identification model was established and data samples were selected to identify risky driving behavior and evaluate the identification efficiency of the model. The results obtained indicated a successful identification rate of 83.6% when the neural network model was solely used to identify risky driving behavior, but this could be increased to 92.46% once corrected by the Bayesian filter. This has important theoretical and practical significance in relation to evaluating the efficiency of existing driver assist systems, as well as the development of future intelligent driving systems.

In the paper titled "Big Data Digging of the Public's Cognition about Recycled Water Reuse Based on the BP Neural Network," reuse of recycled water is very important to both environmental benefits and economic benefits, while public cognition degree towards recycled water reuse also

plays a key role in this process, and it determines the acceptance degree of the public towards recycled water reuse. Under the background of the big data, Hadoop platform was used to collect and save data about the public's cognition towards recycled water in one city and use BP neural network algorithm to construct an evaluation model that could affect the public's cognition level. The public's risk perception, subjective norm, and perceived behavioral control regarding recycled water reuse were selected as key factors. Based on multivariate clustering algorithm, MATLAB software was used to make real testing on massive effective data and assumption models so as to analyze the proportion of three evaluation factors and understand the simulation parameter scope of the cognition degree of different group of citizens. Lastly, several suggestions were proposed to improve the public's cognition of recycled water reuse based on the big data in terms of policy mechanism.

With the development of technologies such as multimedia technology and information technology, a great deal of video data is generated every day. However, storing and transmitting big video data require a large quantity of storage space and network bandwidth because of its large scale. Therefore, compression method of big video data has become a challenging research topic at present. Performance of existing content-based video sequence compression method is difficult to be effectively improved. Therefore, in the paper titled "Parallel Fractal Compression Method for Big Video Data," the authors present a fractal-based parallel compression method without content for big video data. First of all, in order to reduce computational complexity, a video sequence is divided into several fragments according to the spatial and temporal similarity. Secondly, domain and range blocks are classified based on the color similarity feature to reduce computational complexity in each video fragment. Meanwhile, fractal compression method is deployed in a SIMD parallel environment to reduce compression time and improve compression ratio. Finally, experimental results show that the proposed method not only improves the quality of the recovered image but also improves the compression speed compared with existing compression algorithms.

To appropriately realize the performance of a web service, it is essential to give it a comprehensive testing. Although elastic test could be guaranteed in traditional cloud testing systems, the geographical test that supports real user behavior simulation remains a problem. In the paper titled "Distributed Testing System for Web Service Based on Crowdsourcing," the authors propose a testing system based on crowdsourcing model to carry out distributed test on target web server automatically. The proposed crowdsourcing-based testing system (CTS) provides a reliable testing model to simulate real user web-browsing behaviors with the help of web browsers scattered all around the world. In order to make the entire test process the same as the real situation, two test modes are proposed to simulate real user activity. By evaluating every single resource of web service automatically, tester can not only find out internal problems but also understand the performance of the web service. In addition, complete geographical test is available with the performance measurements coming from different regions in the world. Several experiments are performed to validate the functionality and usability of CTS. It is demonstrated that CTS is a complete and reliable web service testing system, which provides unique functions and satisfies different requirements.

In the present big data background, how to effectively excavate useful information is the problem that big data is facing now. The purpose of this study is to construct a more effective method of mining interest preferences of users in a particular field in the context of today's big data. We mainly use a large amount of user text data from microblog to study. LDA is an effective method of text mining, but it will not play a very good role in applying LDA directly to a large number of short texts in microblog. In today's more effective topic modeling project, short texts need to be aggregated into long texts to avoid data sparsity. However, aggregated short texts are mixed with a lot of noise, reducing the accuracy of mining the user's interest preferences. In the paper titled "CLDA: An Effective Topic Model for Mining User Interest Preference under Big Data Background," the authors propose Combining Latent Dirichlet Allocation (CLDA), a new topic model that can learn the potential topics of microblog short texts and long texts simultaneously, avoid the data sparsity of short texts by aggregating long texts to assist in learning short texts, and reuse short text filtering long text to improve mining accuracy, making long texts and short texts effectively combined. Experimental results in a real microblog data set show that CLDA outperforms many advanced models in mining user interest, and we also confirm that CLDA also has good performance in recommending systems.

Sites for web-based shopping are winding up increasingly famous these days. Organizations are anxious to think about their client purchasing conduct to build their item deal. Internet shopping is a method for powerful exchange among cash and merchandise which is finished by end client without investing a huge energy spam. The goal of the paper titled "Ranking Analysis for Online Customer Reviews of Products Using Opinion Mining with Clustering" is to dissect the high recommendation web-based business sites with help of collection strategy and swarm-based improvement system. At first it gathered the client surveys of the items from web-based business locales with a few features and afterwards utilizing Fuzzy C Means (FCM) grouping strategy to group the features for less demanding procedure. From the execution, the outcomes demonstrate the greatest exactness rate that is 94.56% and best E-commerce sits as "Amazon" of our proposed display and diverse features based positioning additionally dissected; it is contrasted with existing grouping and improvement systems.

Aiming at the braking energy feedback control in the optimal energy recovery of the two-motor dual-axis drive electric vehicle (EV), the efficiency numerical simulation model based on the permanent magnet synchronous motor loss was established. At the same time, under different speed and braking conditions, based on maximum recovery efficiency and data calculation of motor system, the optimization motor braking torque distribution model was established. Thus, the distribution rule of the power optimization for the front and rear electric mechanism was obtained. The

paper titled "Energy Recovery Strategy Numerical Simulation for Dual Axle Drive Pure Electric Vehicle Based on Motor Loss Model and Big Data Calculation" takes the Economic Commission of Europe (ECE) braking safety regulation as the constraint condition, and finally a new regenerative braking torque distribution strategy numerical simulation was developed. The simulation model of Simulink and Carsim was established based on the simulation object. Moreover, it is 9.95% higher than the maximum braking energy recovery strategy of the front axle. Finally, through the driving behavior of the driver obtained from the big data platform, we analyze how the automobile braking force matches with the driver's driving behavior. It also analyzes how the automobile braking force matches the energy recovery efficiency. The research results in this paper provide a reference for the future calculation of braking force feedback control system based on big data of new energy vehicles. It also provides a reference for the modeling of brake feedback control system.

In the era of big data, the efficient use of idle data in reinforced concrete structures has become a key issue in optimizing seismic performance evaluation methods for building structures. In the paper titled "Evaluation of Seismic Performance of Reinforced Concrete Frame Structures in the Context of Big Data," based on the evaluation method of structural displacement seismic performance and based on the characteristics of high scalability and high fault tolerance of cloud platform, the open source distributed and storage features of Hadoop architecture cloud platform are introduced as a subproject of Apache Nutch project, Hadoop cloud platform. With features such as high scalability, high fault tolerance, and flexible deployment, the storage platform is secure, stable, and reliable. From the evaluation of the seismic performance of newly built buildings and existing damaged buildings, according to the structural strength-ductility theory of the structure, the building structure resists earthquakes with its strength and ductility, and buildings are divided into four categories. Due to the influence of time or seismic damage on the structure of reinforced concrete frame structures, their material properties are often deteriorating. Using the distributed computing design concept to efficiently process big data, a dynamic evaluation model for the seismic performance of reinforced concrete frame structures is established. A project of a 10-story reinforced concrete frame structure was selected for calculation and analysis, and the engineering example was used to verify the accuracy and efficiency of the model, and the seismic performance of the floor was analyzed. It can be seen that the initial stiffness index of the structure is not sensitive to the damage location of the structure. The platform based on the concept of distributed computing big data processing can effectively improve the efficiency and accuracy of the evaluation of reinforced concrete frame structures.

In the paper titled "The Performance Study on the Long-Span Bridge Involving the Wireless Sensor Network Technology in a Big Data Environment," the random traffic flow model which considers parameters of all the vehicles passing through the bridge, including arrival time, vehicle speed, vehicle type, vehicle weight, and horizontal position, as well as the bridge deck roughness, is input into the vehicle-bridge coupling vibration program. In this way, vehicle-bridge coupling vibration responses with considering the random traffic flow can be numerically simulated. Experimental test is used to validate the numerical simulation, and it had the consistent changing trends. This result proves the reliability of the vehicle-bridge coupling model in this paper. However, the computational process of this method is complicated and proposes high requirements for computer performance and resources. Therefore, this paper considers using a more advanced intelligent method to predict vibration responses of the long-span bridge. The PSO-BP (Particle Swarm Optimization-Back Propagation) neural network model is proposed to predict vibration responses of the long-span bridge. Predicted values and real values at each point basically have the consistent changing trends, and the maximum error is less than 10%. Hence, it is feasible to predict vibration responses of the long-span bridge using the PSO-BP neural network model. In order to verify advantages of the predicting model, it is compared with the BP neural network model and GA-BP neural network model. The PSO-BP neural network model converges to the set critical error after it is iterated to the 226th generation, while the other two neural network models are not converged. In addition, the relative error of predicted values using PSO-BP neural network is only 2.71%, which is obviously less than predicted results of two other neural network models. We can find that the PSO-BP neural network model proposed by the paper in predicting vibration responses is highly efficient and accurate.

In the paper titled "Differential Diagnosis Model of Hypocellular Myelodysplastic Syndrome and Aplastic Anemia Based on the Medical Big Data Platform," the arrival of the era of big data has brought new ideas to solve problems for all walks of life. Medical clinical data is collected and stored in the medical field by utilizing the medical big data platform. Based on medical information big data, new ideas and methods for the differential diagnosis of hypo-MDS and AA are studied. The basic information, peripheral blood classification counts, peripheral blood cell morphology, bone marrow cell morphology, and other information were collected from patients diagnosed with hypo-MDS and AA diagnosed in the first diagnosis. First, statistical analysis was performed. Then, the logistic regression model, decision tree model, BP neural network model, and support vector machine (SVM) model of hypo-MDS and AA were established. The sensitivity, specificity, Yoden index, positive likelihood ratio (+LR), negative likelihood ratio (-LR), Area Under Curve (AUC), accuracy, Kappa value, positive predictive value (PV+), and negative predictive value (PV-) of the four-model training set and test set were compared, respectively. Finally, with the support of medical big data, using Logistic regression, decision tree, BP neural network, and SVM four classification algorithms, the decision tree algorithm is optimal for the classification of hypo-MDS and AA and analyzing the characteristics of the optimal model misjudgment data.

The prediction of stock premium has always been a hot issue. By predicting stock premiums to provide a way for companies to respond to financial risk investments, companies can avoid investment failures. In the paper titled

"Research on Decision-Making of Complex Venture Capital Based on Financial Big Data Platform," under the financial big data platform, Bootstrap resampling technology and Long Short-Term Memory (LSTM) are used to predict the value of the stock premium within 20 months. First, the theme crawler, Jsoup page parsing, Solr search, and Hadoop architecture are used to build a platform for financial big data. Second, based on the block Bootstrap resampling technology, the existing data information is expanded to make full use of the existing data information. Then, based on the LSTM network, the stock premium in 20 months is predicted and compared with the values predicted by support vector machine regression (SVR), and the SSE and R-square average indicators are calculated, respectively. The calculation results show that the SSE value of LSTM is lower than SVR, and the R-square value of LSTM is higher than SVR, which means that the effect of LSTM prediction is better than SVR. Finally, based on the forecast results and evaluation indicators of the stock premium, we provide countermeasures for the company's financial risk investment.

As the information technology of a new generation based on Internet of Things (IoT), big data realizes the record and collection of all data produced in the whole life cycle of the existence and evolutionary process of things. Thus, the amount of data in the network is more huge. However, reprogramming is needed to process big data; big data codes need to be transmitted to all nodes in the network. But the diffusion of codes is an important processing technology for big data networks. In previous scheme, data analysis was conducted for small samples of big data; complex problems cannot be processed by big data technology. Due to the limited capacity of intelligence device, a better method is to select a set of nodes (intelligence device) to form a Connected Dominating Set (CDS) to save energy, and constructing CDS is proven to be a complete NP problem. However, it is a challenge to reduce the communication delay and complexity for urgent data transmission in big data. In the paper titled "Construction Low Complexity and Low Delay CDS for Big Data Code Dissemination," an appropriate duty cycle control (ADCC) scheme is proposed to reduce communication delay and complexity while improving energy efficiency in CDS based WSNs. In ADCC scheme, the method for constructing CDS is proposed at lower complexity. Nodes in CDS are selected according to the degree of nodes. Then duty cycle of dominator nodes in CDS is higher than that of dominated nodes, so the communication delay in the proposed scheme is far less than that of previous scheme. The duty cycle of dominated nodes is small to save energy. This is because the number of dominator nodes in CDS is far less than the number of dominated nodes whose duty cycle is small; thus, the total energy consumption of the network is less than that of the previous scheme.

In the era of big data, group division in online social network analysis is a basic task. It can be divided into the group division based on static relationship and the group division based on dynamic relationship. Compared with the static group division, the users express their semantic information in all kinds of social network behaviors, and users tend to interact with other users who have the same idea and attitude; this is how different groups are formed. In the paper titled "Tibetan Weibo User Group Division Based on Semantic Information in the Era of Big Data," aiming at the issue where some Tibetan users use Chinese to publish microblogs on social platforms, a group division method based on semantic information of Tibetan users under the big data environment is proposed. When dividing a large number of Tibetan user groups in a social network, a large amount of semantic information of Tibetan users in the social network is first analyzed. Then, based on the semantic similarity between users, we aggregate the Tibetan users with high similarities into one group and thus achieve the final group division. The experimental results illustrate the effectiveness of the method of analyzing Tibetan user semantic information in the context of big data for group partitioning.

The RAMS (Reliability, Availability, Maintainability, and Security) of the air braking system is an important indicator to measure the safety performance of the system; it can reduce the life cycle cost (LCC) of the rail transit system. Existing safety analysis methods are limited to the level of relatively simple factual descriptions and statistical induction and fail to provide a comprehensive safety evaluation on the basis of system structure and accumulated data. In the paper titled "RAMS Analysis of Train Air Braking System Based on GO-Bayes Method and Big Data Platform," a new method of safety analysis is described for the failure mode of air braking system, GO-Bayes. This method combines the structural modeling of the GO method with the probabilistic reasoning of Bayes methods, introduces the probability into the analysis process of GO, performs reliability analysis of the air braking system, and builds a big data platform for the air braking system to guide the system maintenance strategy. An automatic train air-braking system is taken as an example to verify the usefulness and accuracy of the proposed method. Using Extend-Sim software shows the feasibility of the method and its advantages in comparison with fault tree analysis.

The large-scale and resourceful utilization of solid waste is one of the important ways of sustainable development. The big data brings hope for further development in all walks of life, because huge amounts of data insist on the principle of "turning waste into treasure." The steel big data has been taken as the research object in the paper titled "Systematic Research on the Application of Steel Slag Resources under the Background of Big Data." Firstly, a big data collection and storage system has been set up based on Hadoop platform. Secondly, the steel slag prediction model based on the convolution neural network (CNN) is established. The material data of steel-making, the operation data of steel-making process, and the data of steel slag composition are put into the model from the Hadoop platform, and the prediction of the slag composition is further realized. Then, the alternatives for resource recovery are obtained according to the predicted composition of the steel slag. And considering the three aspects of economic feasibility, resource suitability, and environmental acceptance, the comprehensive evaluation system based on AHP is established to realize the recommendation of the optimal resource approach. Finally, taking a steel plant in Hebei as an example, the alternatives

according to the prediction of the composition of steel slag are blast furnace iron-making, recycling waste steel, and cement admixture. The comprehensive evaluation values of the three resources are 0.48, 0.57, and 0.76, respectively, and the optimized resource of the steel slag produced by the steel plant is used as the cement admixture.

In the paper titled "Research on Application of Big Data in Internet Financial Credit Investigation Based on Improved GA-BP Neural Network," the arrival of the era of big data has provided a new direction of development for Internet financial credit collection. First of all, the article introduced the situation of Internet finance and traditional credit industry. Based on that, the mathematical model was used to demonstrate the necessity of developing big data financial credit information. Then, the Internet financial credit data is preprocessed, the variables suitable for modeling are selected, and the dynamic credit tracking model of BP neural network based on adaptive genetic algorithm is constructed. It is found that both LM training algorithm and Bayesian algorithm can converge the error to 10e-6 quickly in the model training, and the overall training effect is ideal. Finally, the rule extraction algorithm is used to simulate the test samples. The accuracy rate of each sample method is over 90%, and some accuracy rate is even more than 90%, which indicates that the model is applicable to the credit data of big data in Internet finance.

The paper titled "Research on Workshop-Based Positioning Technology Based on Internet of Things in Big Data Background" first analyzes the data collection and data management of the workshop, obtains the data of the workshop changes with time, and accumulates the data. There are bottleneck problems such as big data being difficult to be fully used. Then, the concept of the Internet of Things was introduced into the workshop positioning to realize the comprehensive use of the big data in the workshop. Finally, aiming at the positioning problem of manufacturing workshop items, the Zigbee positioning algorithm, the received signal strength indication algorithm RSSI, and the trilateration algorithm are applied, and the trilateral positioning algorithm is applied to the CC2430 wireless MCU, and the positioning node is designed and implemented. The three-sided localization algorithm was used to locate and simulate the horizontal and vertical comparisons of six groups of workshop terminals. The results showed that the difference between the simulated position and the actual position did not exceed 1m, which was in line with the positioning requirements of the workshop.

In the paper titled "Dynamic Prediction Research of Silicon Content in Hot Metal Driven by Big Data in Blast Furnace Smelting Process under Hadoop Cloud Platform," the content of [Si] in molten iron is an important factor in the temperature of molten iron and blast furnace anterograde. At present, the time series based statistical model and machine learning algorithm model are used to predict the content of [Si] in molten iron and guide the actual production of blast furnace. The SVM algorithm is a machine learning algorithm based on maximum interval theory and structural risk principle, which can effectively avoid dimension disaster and enhance the generalized performance of the algorithm, and has significant effect in the application of big data sample set. Because the big-scale blast furnace smelting process is a continuous nonlinear large time delay process, the factors influencing the content of [Si] in molten iron are complicated and the synchronous correspondence is complex, and the amount of data generated during the smelting process is extremely large. In order to explore a dynamic prediction model with good generalization performance of the content of [Si] in molten iron, an improved SVM algorithm is proposed to enhance its practicability in the big data sample set of the smelting process. Firstly, the basic principle of SVM algorithm is studied. On the basis of clarifying that the SVM solution is a quadratic programming problem, we propose a parallelization scheme to design SVM solution algorithm based on MapReduce model under Hadoop platform to improve the solution speed of the SVM on big data sample set. Secondly, based on the characteristics of stochastic subgradient projection, the execution time of the SVM solver algorithm does not depend on the size of the sample set, and a structured SVM algorithm based on neighbor propagation algorithm is proposed, and, on this basis, a parallel algorithm for solving the covariance matrix of the training set and the parallel algorithm of the tth iteration of the random subgradient projection are designed.

In the paper titled "Intrinsic Mode Chirp Multicomponent Decomposition with Kernel Sparse Learning for Overlapped Nonstationary Signals Involving Big Data," the authors focus on the decomposition problem for nonstationary multicomponent signals involving big data. We propose the kernel sparse learning (KSL), developed for the T-F reassignment algorithm by the path penalty function, to decompose the instantaneous frequencies (IFs) ridges of the overlapped multicomponent from a time-frequency representation (TFR). The main objective of KSL is to minimize the error of the prediction process while minimizing the amount of training samples used and thus to cut the costs interrelated with the training sample collection. The IFs first extraction is decided using the framework of the intrinsic mode polynomial chirp transform (IMPCT), which obtains a brief local orthogonal TFR of signals. Then, the IFs curves of the multicomponent signal can be easily reconstructed by the T-F reassignment. After the IFs are extracted, component decomposition is performed through KSL. Finally, the performance of the method is compared when applied to several simulated micro-Doppler signals, which shows its effectiveness in various applications.

In the paper titled "Analysis of Converter Combustion Flame Spectrum Big Data Sets Based on HHT," the characteristics of the converter combustion flame are one of the key factors in the process control and end-point control of steelmaking. In big data era, it is significant to carry out high-speed and effective processing on the data of frame spectrum. By installing data acquisition devices at the converter mouth and separating the spectrum according to the wave length, high-dimension converter flame spectrum big data sets are achieved. The data of each converter is preprocessed after information fusion. By applying the SM software, the correspondence with the carbon content is obtained. Selecting the relative data of the two peak ratios and the single-peak absolute data as a one-dimensional signal, due to the obvious nonlinear and nonstationary characteristics, using HHT to

do empirical mode decomposition and Hilbert spectrum analysis, the variation characteristics after 70% of converter steelmaking process are obtained. From data acquisition and data preprocessing to data analysis and results, it provides a new perspective and method for the study of similar problems.

In the paper titled "FDM Rapid Prototyping Technology of Complex-Shaped Mould Based on Big Data Management of Cloud Manufacturing," in order to solve the problem of high cost and long cycle in traditional subtractive material manufacturing process of complex shaped mould, by using the technology of FDM rapid prototyping and combining with the global service idea of cloud manufacturing, the information of various kinds of heterogeneous forming process data produced in the process of FDM rapid prototyping is analyzed; meanwhile, the transfer and transformation relation of each forming process data information in the rapid manufacturing process with the digital model as the core is clarified, so that the FDM rapid manufacturing process is integrated into one; the digital and intelligent manufacturing system of complex shaped mould based on the cloud manufacturing big data management is formed. This paper takes the investment casting mould of spur gear as an example; through the research on the forming mechanism of jet wire, the factors affecting forming quality and efficiency are analyzed from three stages: the pretreatment of the 3D model, the rapid prototyping, and the postprocessing of the forming parts; the relationship between the forming parameters and the craft quality is established, and the optimization schemes at each stage of this process are put forward through the study on the forming mechanism of jet wire. Through rapid prototyping test, it is shown that the spur face gear master mould based on this technology can be quickly manufactured with critical surface accuracy within a range of 0.036mm-0.181mm and surface roughness is within the range of 0.007-0.01μm by only 1/3 processing cycle of traditional subtractive material manufacturing. It lays a solid foundation for rapid intelligent manufacturing of products with complex shaped structure.

There is a huge amount of data in the opportunity of "turning waste into treasure" with the arrival of the big data age. Urban layout is very important for the development of urban transportation and building system. Once the layout of the city is finalized, it will be difficult to start again. Therefore, the urban architectural layout planning and design has a very important impact. The paper titled "Optimization of Planning Layout of Urban Building Based on Improved Logit and PSO Algorithms" uses the urban architecture layout big data for building layout optimization using advanced computation techniques. Firstly, a big data collection and storage system based on the Hadoop platform is established. Then the evaluation model of urban building planning based on improved logit and PSO algorithm is established. The PSO algorithm is used to find the suitable area for this kind of building layout, and then, through five impact indicators, land prices, rail transit, historical protection, road traffic capacity, and commercial potential, has been established by using the following logit linear regression model. Then the bridge between logit and PSO algorithm is established by the fitness value of particle. The particle in the particle swarm is assigned to the index parameter of logit model, and then the logit model in the evaluation system is run. The performance index corresponding to the set of parameters is obtained. The performance index is passed to the PSO as the fitness value of the particle to search for the best adaptive position. The reasonable degree of regional architectural planning is obtained, and the rationality of urban architectural planning layout is determined.

In the paper titled "Research on Optimization of Big Data Construction Engineering Quality Management Based on RNN-LSTM," construction industry is the largest data industry, but with the lowest degree of datamation. With the development and maturity of BIM information integration technology, this backward situation will be completely changed. Different business data from construction phase and operation and maintenance phase will be collected to add value to the data. As the BIM information integration technology matures, different business data from the design phase to the construction phase are integrated. Because BIM integrates massive, repeated, and unordered features text data, we first use integrated BIM data as a basis to perform data cleansing and text segmentation on text big data, making the integrated data "clean and orderly" valuable data. Then, with the aid of word cloud visualization and cluster analysis, the associations between data structures are tapped, and the integrated unstructured data is converted into structured data. Finally, the RNN-LSTM network was used to predict the quality problems of steel bars, formworks, concrete, cast-in-place structures, and masonry in the construction project and to pinpoint the occurrence of quality problems in the implementation of the project. Through the example verification, the algorithm proposed in this paper can effectively reduce the incidence of construction project quality problems, and it has promotion. And it is of great practical significance to improving quality management of construction projects and provides new ideas and methods for future research on the construction project quality problem.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We would like to thank all the authors for their valuable contributions in this special issue as well as all the reviewers who are all experts on the theme.

Zhihan Lv
Kaoru Ota
Jaime Lloret
Wei Xiang
Paolo Bellavista