JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Construction of an Internet of Things Industry Chain Classification Model Based on IRFA and Text Analysis

Zhimin Wang[*]

Abstract

With the rapid development of Internet of Things (IoT) and big data technology, a large amount of data will be generated during the operation of related industries. How to classify the generated data accurately has become the core of research on data mining and processing in IoT industry chain. This study constructs a classification model of IoT industry chain based on improved random forest algorithm and text analysis, aiming to achieve efficient and accurate classification of IoT industry chain big data by improving traditional algorithms. The accuracy, precision, recall, and AUC value size of the traditional Random Forest algorithm and the algorithm used in the paper are compared on different datasets. The experimental results show that the algorithm model used in this paper has better performance on different datasets, and the accuracy and recall performance on four datasets are better than the traditional algorithm, and the accuracy performance on two datasets, P-I Diabetes and Loan Default, is better than the random forest model, and its final data classification results are better. Through the construction of this model, we can accurately classify the massive data generated in the IoT industry chain, thus providing more research value for the data mining and processing technology of the IoT industry chain.

Keywords

Industrial Chain, IoT, Random Forest Algorithm, Text Analysis, Visualization

1. Introduction

The Internet of Things (IoT) industry chain can be subdivided into four segments: marking, sensing, information transmission, and data processing. Among which, data mining technology in the data processing segment is the focus of current research in the IoT industry chain [1]. From the sensing layer to the application layer of IoT, the type and quantity of various information increase exponentially. And the amount of data to be analyzed increases exponentially. How to dig out the hidden information and effective data information from the huge amount of data in a timely manner poses a great challenge to the data processing workers [2]. The rapid development of information technology has led to the explosive growth of data in the IoT and big data industries. Intelligent algorithm and mining technology are applied to data classification problems and advantage modeling of data classification. They cannot only better mine the information behind the data, but also promote the development of the IoT industry and big data [3]. The traditional random forest (RF) algorithm does not perform data prediction and

classification well in the face of many data features and unbalanced data samples [4]. Therefore, the study optimizes the traditional RF algorithm based on its voting method and reallocates the weights according to the classification performance of the classifier. And an optimized RF algorithm is proposed, aiming to improve the drawbacks of the traditional algorithm in data classification, so as to better explore various hidden information behind the big data of the IoT industry chain.

2. Related Works

The IoT is a new large-scale information carrier that aims to interconnect all common objects that can exercise independent functions through a network. The development of the IoT had led to a gradually large IoT industry chain system, and the development of various IoT industries was accompanied by the generation of many big data [5]. Supervised learning methods represented by random forests had a good performance in data classification and regression, and were widely used in various fields [6]. To investigate the relationship between the diversity of pore types and multi-scale characteristics of carbonate rocks and permeability, Zhang and Cai [7] designed three permeability prediction schemes by combining BP neural network and RF algorithm. And the experimental results showed that the method has better generalization ability. And it could be a reliable permeability prediction method for carbonate rocks. In order to have a more objective and accurate diagnosis of Alzheimer's disease, Yang et al. [8] applied the recurrent RF to the diagnosis of patients with mild cognitive impairment. The results showed that the screened imaging biomarkers could be used as a basis for diagnosis and the RF could also help healthcare professionals to diagnose the condition effectively. Pasinetti et al. [9] suggested a machine learning-based method for classifying user gait stages, in which the main method used was the Sigma-z RF classifier. The classifier was able to provide information on the classification of gait stages by taking into account the uncertainty associated with each feature set. The method was tested and found to have an average classification accuracy of 87.3%, outperforming the traditional RF classifier [9]. Yang et al. [10] used the RF for feature determination within the model. The stability and diagnostic properties of the RF model were evaluated and tested. It showed that the RF was able to determine relevant features well with an accuracy of 91.3% and a test sensitivity of 88.9%. Considering the differences in individual thermal sensations, a thermal sensation model based on the RF classification algorithm was investigated by Li et al. [11]. This model was on the basis of satisfying various factors such as passenger temperature comfort and energy consumption. The performance of the model was tested and found to have good functionality for intelligent temperature control selection.

From the above studies, it was clear that many data classification algorithms had been proposed and widely used in various fields [12]. The RF algorithm, represented by its excellent learning ability and classification capability, had a crucial role in the analysis and diagnosis of medical data [13]. The research and development of data mining technology was closely related to the future development of IoT industry clusters, which were characterized by a wide range of data types and complexity. The research content of this thesis would optimize the traditional RF algorithm and its advantages would be used in data classification to achieve the goal of mining and classification of IoT industry chain big data.

3. Research on the Classification Method of IoT Big Data Industry Chain based on IRF Algorithm and Text Analysis

3.1 Key Methodological Design for IoT Big Data Industry Chain Analysis

As the third revolution of information science and technology industry, IoT aims to achieve interconnection between all common physical objects. The IoT industry chain can be subdivided into four links: identification, sensing, information transmission, and data processing. The core technologies of each link mainly include radiofrequency identification technology, sensing technology, network and communication technology, data mining, and fusion technology. With the development and innovation of IoT and computer technology, IoT is also widely used in various fields, such as urban public safety, industrial production, environmental monitoring, health detection, smart home, etc. In information transmission between things and things, it not only generates a large amount of data, but also makes the big data industry and IoT industry gradually cross-pollinate [14,15]. In order to explore the deep value behind these data, the study first analyzes the IoT big data industry chain and classifies its industry data by using relevant integrated learning algorithms, and finally constructs the IoT big data industry chain classification model. The constructed model is used in practical situations to assist the relevant staff to complete the efficient classification of the IoT industry chain big data, so as to make full use of the value of all kinds of data.

Data mining technology has developed into a number of mature and diverse classification algorithms. Among them, decision trees are widely used in classification problems because of their high classification accuracy, high model efficiency, strong data processing capability and fast operation. The ID3 algorithm, the earliest decision tree algorithm, operates by selecting the optimal partitioning attribute through information gain at each node of the decision tree [16]. The degree of reduction in uncertainty of similar information for a given feature is called information gain $H(D)$.

$$H(D) = -\sum_{k=1}^{K} p_k \log_2(p_k). \tag{1}$$

The calculation of the information gain is shown in Eq. (1). D denotes the set of samples, and p_k denotes the samples proportion of type k in the set D.

$$H(D \mid A) = \sum_{i=1}^{n} \frac{|D_i|}{|D|} H(D_i) = -\sum_{i=1}^{n} \frac{|D_i|}{|D|} \sum_{k=1}^{K} p_{ik} \log_2(p_{ik}). \tag{2}$$

Eq. (2) shows the calculation of empirical conditional entropy. A denotes the branch nodes obtained by dividing the feature values into sample sets, $|D|$ denotes the weight of each branch node, and $|D_i|$ denotes the information entropy of each branch node.

$$Info\_Gain(D, A) = H(D) - H(D|A). \tag{3}$$

Eq. (3) shows the formula for calculating the information gain based on the discrete features that divide the branch node dataset D. Since the ID3 algorithm is prone to bias in selecting features that take more values in the information gain calculation, the C4.5 decision tree algorithm is used for improvement.

$$Info\_Gain\_Ratio(D|A) = \frac{Info\_Gain(D,A)}{H_A(D)}. \tag{4}$$

Eq. (4) is an optimization of the information gain calculation using the C4.5 decision tree algorithm. $H_A(D)$ denotes the entropy of the sample set with respect to the discrete features.

$$H_A(D) = -\sum_{i=1}^{n} \frac{|D_i|}{|D|} \log_2 \frac{|D_i|}{|D|}. \tag{5}$$

Eq. (5) is the formula for calculating entropy. According to its formula, the criterion for optimal feature classification is to select the feature with the highest rate of information yield. The decision tree algorithm does not require normalization of the data. It is easy to understand and easy to operate, and can handle both category and data features. Currently, many different integration algorithms have been derived based on different decision tree integration strategies, such as the RF based on the bagging method [17]. It consists of three main parts: bagging, decision tree construction, and out-of-bag estimation [18,19]. The sample instances belong to the bagging part, while the voting and the final structural analysis belong to the out-of-bag estimation part. Given a dataset B, the data is divided into l tuples and m attributes.

$$B = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} & y_1 \\ x_{21} & x_{22} & \cdots & x_{2m} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{l1} & x_{l2} & \cdots & x_{lm} & y_l \end{bmatrix}. \tag{6}$$

Eq. (6) is an expression for a given dataset. Assuming that the RF model has k single classifiers, the decision trees is set as $T_i, i = (1,2,\dots,k)$. Bagging is done by taking the ith decision tree from the dataset B and labeling the lth tuple in a put-back manner. In RF, the splitting criterion is usually the Gini index.

$$Gin\,i(B) = 1 - \sum_{i=1}^{m} p_i^2. \tag{7}$$

In Eq. (7), p_i represents the probability of drawing a tuple from the dataset B. Dividing the dataset into two subsets, the Gini index is calculated as shown in (8):

$$Gini_A(B) = \frac{|B_1|}{|B|} Gini(B_1) + \frac{|B_2|}{|B|} Gini(B_2). \tag{8}$$

In Eq. (8), A represents the splitting attribute. B_1 and B_2 are two subsets of the dataset B, respectively. The optimal splitting criterion for each node is obtained by solving for the minimum value of the Gini index of the node. The sample X to be tested is fed into the model, and the final result of the corresponding classification sequence is shown in Eq. (9):

$$P = \{T_1(X), T_2(X), \cdots T_i(X)\}. \tag{9}$$

The RF makes a decision by counting the results of each decision tree to make a final prediction. The plurality of the predicted results is generally chosen as the final result.

$$T(X) = \arg\max \sum_{i=1}^{K} I(T_i(X) = Y). \tag{10}$$

In addition to this, the average method can be used to select the mean of the predicted classification results as the final classification result. The formula is shown in (11):

$$T(X) = \frac{1}{K} \sum_{i=1}^{K} T_i(X).$$ (11)

The overall performance of the traditional RF algorithm is defined by the categorization performance of each decision tree and the diversity of the forest. The better the performance of the decision trees and the lower the correlation between them, the better the performance of the RF. The training subset obtained by the traditional sampling method is prone to problems, such as unbalanced data features and poor data classification when features in the dataset is large. Considering that, the RF will be further improved in the following.

3.2 IoT Big Data Industry Chain Classification Model Construction based on IRF Algorithm and Text Analysis

The overall performance of the traditional RF algorithm is determined by the classification performance and forest diversity of each decision tree. The better the performance of the decision trees and the lower the correlation between them, the better the performance of the RF algorithm. When the number of features in the dataset is large, the training subset obtained by the traditional sampling method is prone to the problems of unbalanced data features and poor data classification. To solve the problems of traditional random forest algorithm, the study tries to optimize it and use the IRF algorithm in the classification of big data of IoT industry chain.

In the traditional voting session of the RF algorithm, the plurality or average of the predicted classification results is usually selected as the final classification result [20]. Such a classification approach does not take into account the existence of different classification performance among individual decision trees, resulting in some differences in the classification results. In this study, an IRF algorithm with optimized voting weights is proposed to optimize the traditional RF algorithm. The way to optimize the voting weights is to reallocate the weights based on the different categorization performance of the classifiers. The performance of the classifier in the test sample set of the training sample set is used as a specific score indicator, and this score indicator is used as the weight of each classifier in the voting session. The improved voting weights are calculated as shown in Eq. (12):

$$Weight(i) = \frac{Metric(i)}{\sum_{i=1}^{N} Metric(i)}.$$ (12)

Eq. (12) is the formula for calculating the optimized voting weight based on the mean value of the classification evaluation index. Where N denotes the number of base classifiers and $Metric(i)$ denotes the mean value of the performance metric of base classifier i. The optimized weight calculation is applied to the algorithm by setting the input training set, test set, number of features and base classifiers to D, T, d, and N, respectively. The output classification result is $H(x)$. N_{pos} means positive samples in the training set and their percentage of the total samples R_{pos} are calculated, and a training subset is created by automatic sampling input training set.

$$R_{pos} < 10\%.$$ (13)

Eq. (13) is the judgment criterion for self-sampling to stop. Self-sampling is performed on the number of positive samples in the input training set until the above conditions are met. The performance metrics of each base classifier are calculated and summarized by using Eq. (12).

$$E = \{h_1(x), h_2(x), \cdots, h_i(x)\}. \tag{14}$$

Eq. (14) is the summary result of each base classifier. Where, $h_i(x)$ is the classification result of the base classifier obtained using the training subset.

$$H(x) = \omega_{\text{arg max}j} \sum_{i=1}^{T} \frac{1}{2} v_{i,j}\big(Weight(i) + Weight\ (i)\big) \tag{15}$$

Equation (15) is the formula for calculating the final classification result obtained by the IRF algorithm under the optimized weight method. Where, $Weight(i)$ denotes the weight of base classifier i calculated by the test set, and $Weight_o(i)$ denotes the weight of base classifier i on the out-of-bag sample. Using Eq. (15), the results of different voting methods can be obtained, and the voting weights of each base classifier are combined to select the category label with the highest score as the final result.
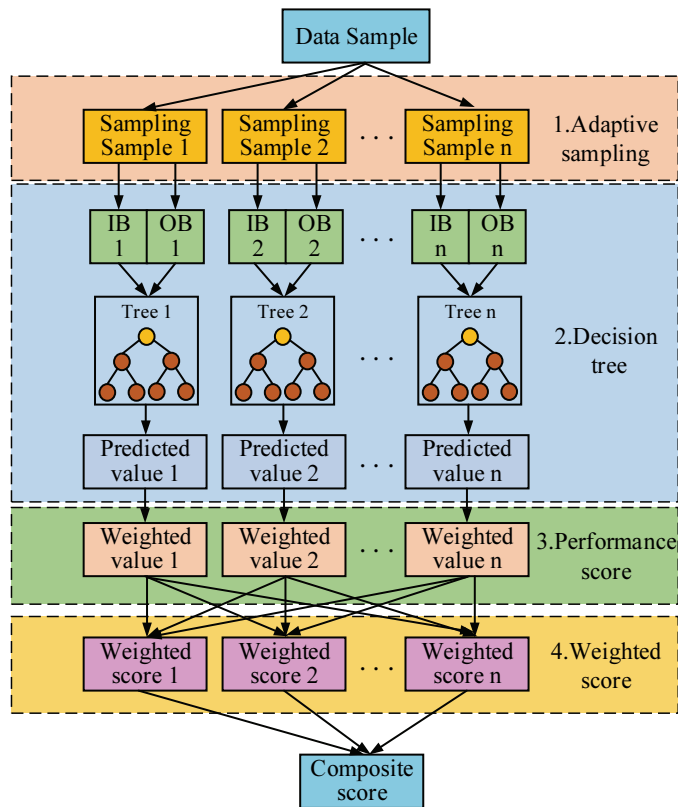

Fig. 1. Overall flowchart of the improved random forest model.

Combining the above classifier weight-based optimization method with the traditional RF model, the overall flowchart of the IRF algorithm is obtained as shown in Fig. 1. As can be seen from Fig. 1, the

whole IRF model is divided into four parts, which are adaptive sampling, decision tree classification, decision tree classification performance scoring and weighted scoring. In order to better classify each data accurately, the study applies the text analysis module to the decision tree classification, and tries to further optimize the decision tree classification module by using the idea of text classification.

4. Performance Test Results and Analysis of Classification Models based on IRFA and Text Analysis

4.1 Experimental Development Environment and Data Processing

To ensure the proper conduct of the experiment, the simulation environment for this experiment is shown in Table 1. In order to verify the performance of the classification model constructed in this article, this experiment selected four types of datasets from the Kaggle dataset, namely Ecoli, P-I Diabetes, B-C-Wisconsin, and Loan Default Dataset. These four types of datasets cover the medical, health, and economic fields. These four types of datasets are preprocessed in the text analysis module using spaCy to delete unrelated text and merge the same text. The preprocessed text is placed into the spaCy high-level natural language library, and the long text is matched by the PhraseMatcher in spaCy to use a classification model to select appropriate data for classification testing. By testing the performance of the model on different datasets, it is more possible to demonstrate the performance of the IoT big data industry chain classification model constructed in this study from different aspects.

Table 1. Computer equipment information

Name	Configuration
Graphics card	GTX 1080ti
CPU	Inter Xeon E5
GPU-accelerated library	CUDA 10.0
Memory	32 GB
Operating system	Windows 10
Deep learning framework	TensorFlow 1.8

4.2 Model Performance Testing

Firstly, the performance of the algorithm in the classification model is tested. The performance of the traditional RF algorithm and the IRF algorithm proposed in the study were analyzed using accuracy indicators.

The accuracy performance of the different algorithmic models on each dataset is shown in Fig. 2. Fig. 2(a)–2(d) show the four datasets of Ecoli, P-I Diabetes, B-C-Wisconsin, and Loan Default, respectively. Comparing the performance of RF and IRF on the four datasets. The IRF model performs better than the traditional RF model in terms of accuracy on all four datasets. The final accuracy of the RF model on the Ecoli, P-I Diabetes, B-C-Wisconsin, and Loan Default datasets is 83.78%, 68.70%, 98.01%, and 93.27%, respectively. The final accuracy of the IRF model on the four datasets is 83.87%, 71.28%, 98.33%, and 93.46%, respectively.
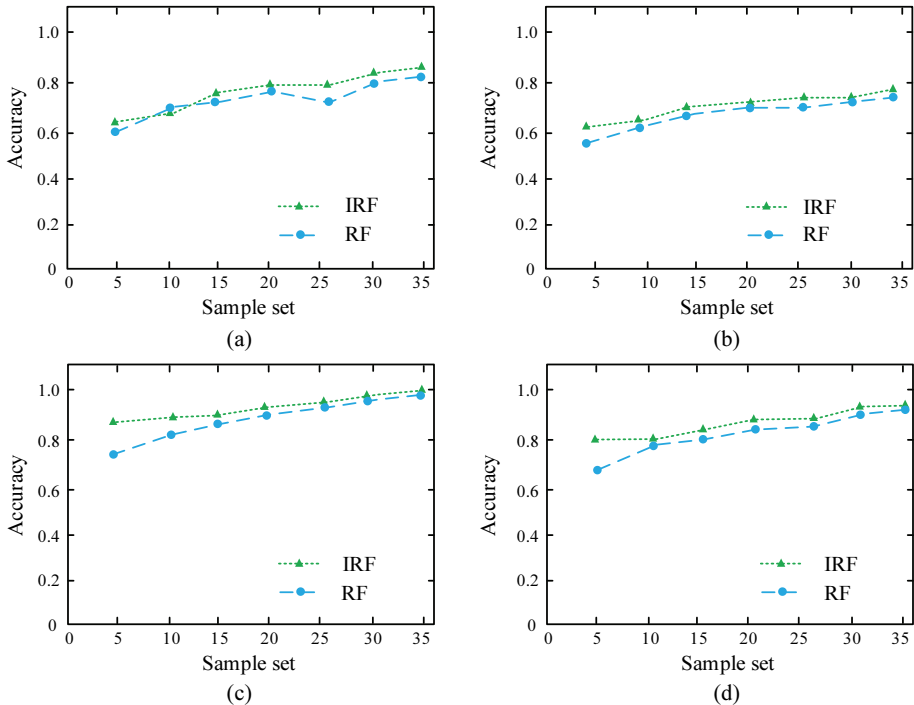
Fig. 2. Accuracy performance of different models on each dataset: (a) Ecoli, (b) PI Diabetes, (c) B-C-Wisconsin, and (D) Loan Default.

4.3 Analysis of Application Results

The classification effects of different classification models on each dataset are shown in Fig. 3. Comparing the classification effects of RF and IRF on the four datasets, it can be seen that the IRF model has better classification effects than the RF model on the Ecoli, B-C-Wisconsin, and Loan Default datasets. And the classification effects on the P-I Diabetes dataset are not much different from RF. The AUC values of IRF model on Ecoli, B-C-Wisconsin, and Loan Default datasets are larger than those of RF model, which are 0.942, 0.879, and 0.664, respectively. It can be seen that IRF model has better classification effect on IoT industry big data, and can identify different types of IoT industry data more accurately and classify them accurately.

5. Discussion

The results of this study are as follows.

– The traditional RF algorithm is optimized, and an IRF algorithm with optimized voting weights is proposed. The IRF algorithm can take into account the differences between the performance of each decision tree, thus optimizing the classification performance.

– A text analysis module is constructed with the spaCy high-level natural language library. And it is used in the decision tree classification module of the IRF algorithm to construct the final classi-fication model. The classification model is used in IoT industry classification, and it can classify the data according to the pre-defined classification criteria and achieve good classification results.
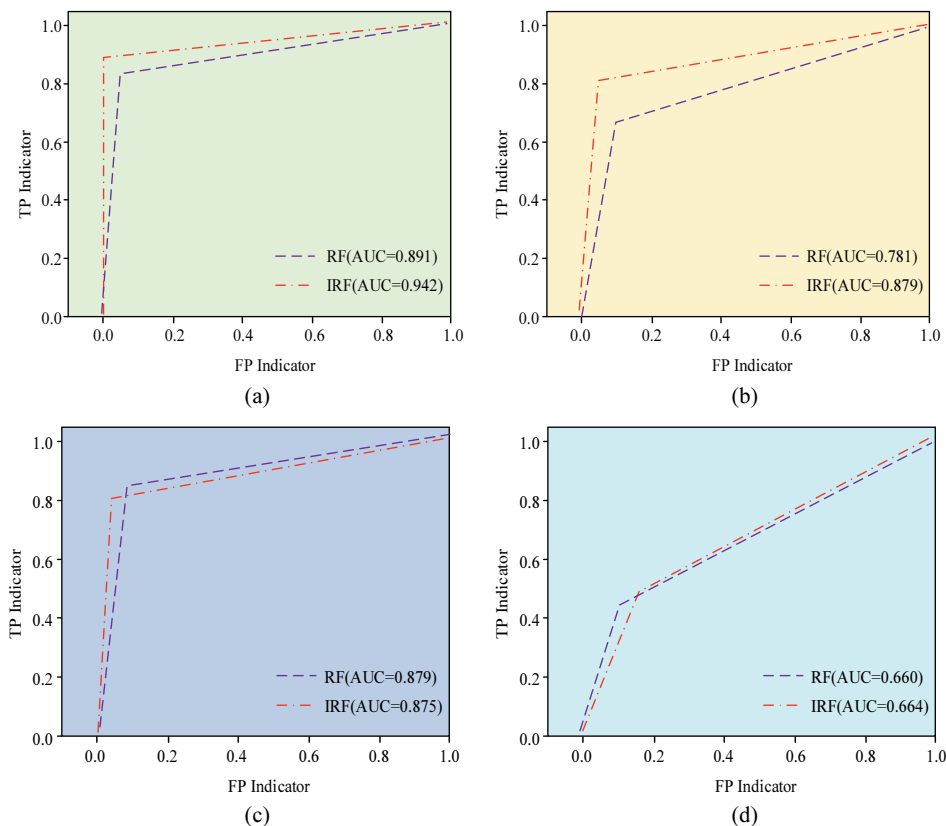
Fig. 3. Classification results of different models on each dataset: comparison of the AUC values of (a) Ecoli, (b) PI Diabetes, (c) B-C-Wisconsin, and (D) Loan Default.

6. Conclusion

The development of the IoT and the big data industry has led to the continuous generation of various new data in the IoT industry chain. In the current industrial chain of the IoT, various data mining technologies are the focus of research. This study constructs a classification model of the IoT industry chain based on IRF and text analysis. The final constructed algorithm model was tested to compare the accuracy, recall, and AUC values of traditional RF algorithms and optimized RF algorithms on different datasets. The experimental results show that the IRF model outperforms other traditional models in all four datasets. The experimental results show that the final accuracy of the IRF model on the four datasets is 83.87%, 71.28%, 98.33%, and 93.46%, respectively, which is superior to the RF model. The AUC values of the IRF model for the three datasets were higher than those of the RF model, with values of 0.942, 0.879, and 0.664, respectively. By applying the classification model to practical applications, data in the industrial chain of the IoT can be well classified. Therefore, the information behind the data can be mined based on different data characteristics to promote the development of the IoT industrial cluster. Due to the complex structure and huge system of the IoT, subsequent research needs to analyze other modules of the IoT industry chain, such as sensing technology, wireless transmission, and various RFID technologies.

Acknowledgement

References

[1] S. Sharma, N. Chhimwal, K. K. Bhatt, A. K. Sharma, P. Mishra, S. Sinha, A. Raj, and S. Tripathi, "FCS-fuzzy net: cluster head selection and routing-based weed classification in IoT with MapReduce framework," *Wireless Networks*, vol. 27, pp. 4929-4947, 2021. https://doi.org/10.1007/s11276-021-02723-x

[2] D. P. Penumuru, S. Muthuswamy, and P. Karumbu, "Identification and classification of materials using machine vision and machine learning in the context of Industry 4.0," *Journal of Intelligent Manufacturing*, vol. 31, pp. 1229-1241, 2020. https://doi.org/10.1007/s10845-019-01508-6

[3] L. Zhang and N. Ansari, "Optimizing the operation cost for UAV-aided mobile edge computing," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 6085-609, 2021. https://doi.org/10.1109/TVT.2021.3076980

[4] L. Liu, E. G. Larsson, P. Popovski, G. Caire, X. Chen, and S. R. Khosravirad, "Guest editorial: massive machine-type communications for IoT," *IEEE Wireless Communications*, vol. 28, no. 4, pp. 56-56, 2021. https://doi.org/10.1109/MWC.2021.9535445

[5] J. G. Wieringa, "Comparing predictions of IUCN Red List categories from machine learning and other methods for bats," *Journal of Mammalogy*, vol. 103, no. 3, pp. 528-539, 2022. https://doi.org/10.1093/jmammal/gyac005

[6] A. Beniiche, A. Ebrahimzadeh, and M. Maier, "The way of the DAO: toward decentralizing the tactile Internet," *IEEE Network*, vol. 35, no. 4, pp. 190-197, 2021. https://doi.org/10.1109/MNET.021.1900667

[7] Z. Zhang and Z. Cai, "Permeability prediction of carbonate rocks based on digital image analysis and rock typing using random forest algorithm," *Energy & Fuels*, vol. 35, no. 14, pp. 11271-11284, 2021. https://doi.org/10.1021/acs.energyfuels.1c01331

[8] J. Yang, H. Sui, R. Jiao, M. Zhang, X. Zhao, L. Wang, W. Deng, and X. Liu, "Random-forest-algorithm-based applications of the basic characteristics and serum and imaging biomarkers to diagnose mild cognitive impairment," *Current Alzheimer Research*, vol. 19, no. 1, pp. 76-83, 2022. https://doi.org/10.2174/1567205019666220128120927

[9] S. Pasinetti, A. Fornaser, M. Lancini, M. De Cecco, and G. Sansoni, "Assisted gait phase estimation through an embedded depth camera using modified random forest algorithm classification," *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3343-3355, 2020. https://doi.org/10.1109/JSEN.2019.2957667

[10] C. Yang, Z. K. Jiang, L. H. Liu, and M. S. Zeng, "Pre-treatment ADC image-based random forest classifier for identifying resistant rectal adenocarcinoma to neoadjuvant chemoradiotherapy," *International Journal of Colorectal Disease*, vol. 35, pp. 101-107, 2020. https://doi.org/10.1007/s00384-019-03455-3

[11] Q. Y. Li, J. Han, and L. Lu, "A random forest classification algorithm based personal thermal sensation model for personalized conditioning system in office buildings," *The Computer Journal*, vol. 64, no. 3, pp. 500-508, 2021. https://doi.org/10.1093/comjnl/bxaa165

[12] X. Deng, K. Milligan, R. Ali-Adeeb, P. Shreeves, A. Brolo, J. J. Lum, J. L. Andrews, and A. Jirasek, "Group and basis restricted non-negative matrix factorization and random forest for molecular histotype classification and Raman biomarker monitoring in breast cancer," *Applied Spectroscopy*, vol. 76, no. 4, pp. 462-474, 2020. https://doi.org/10.1177/00037028211035398

[13] J. Wang, Z. Jiang, Y. Wei, W. Wang, F. Wang, Y. Yang, H. Song, and Q. Yuan, "Multiplexed identification of bacterial biofilm infections based on machine-learning-aided lanthanide encoding," *ACS Nano*, vol. 16, no. 2, pp. 3300-3310, 2022. https://doi.org/10.1021/acsnano.1c11333

[14] L. Yu, W. Jiang, Z. Ren, S. Xu, L. Zhang, and X. Hu, "Detecting changes in attitudes toward depression on Chinese social media: a text analysis," *Journal of Affective Disorders*, vol. 280, pp. 354-363, 2021. https://doi.org/10.1016/j.jad.2020.11.040

[15] O. Kulkarni, S. Jena, and V. Ravi Sankar, "MapReduce framework based big data clustering using fractional integrated sparse fuzzy C means algorithm," *IET Image Processing*, vol. 14, no. 12, pp. 2719-2727, 2020. https://doi.org/10.1049/iet-ipr.2019.0899

[16] M. Macnee, E. Perez-Palma, S. Schumacher-Bass, J. Dalton, C. Leu, D. Blankenberg, and D. Lal, "SimText: a text mining framework for interactive analysis and visualization of similarities among biomedical entities," *Bioinformatics*, vol. 37, no. 22, pp. 4285-4287, 2021. https://doi.org/10.1093/bioinformatics/btab365

[17] M. Mahendran, D. Lizotte, and G. R. Bauer, "Describing intersectional health outcomes: an evaluation of data analysis methods," *Epidemiology*, vol. 33, no. 3, pp. 395-405, 2022. https://doi.org/10.1097/EDE.0000000000001466

[18] J. Zhou, Q. Mao, J. Zhang, N. M. Lau, and J. Chen, "Selection of breast features for young women in northwestern China based on the random forest algorithm," *Textile Research Journal*, vol. 92, no. 7-8, pp. 957-973, 2022. https://doi.org/10.1177/00405175211040869

[19] Y. J. Yoo and K. S. Cho, "Development of cost-effective IoT module-based pipe classification system for flexible manufacturing system of painting process of high-pressure pipe," *The International Journal of Advanced Manufacturing Technology*, vol. 119, pp. 5453-5466, 2022. https://doi.org/10.1007/s00170-021-08478-1

[20] G. Shirazinejad, M. J. V. Zoej, and H. Latifi, "Applying multidate Sentinel-2 data for forest-type classification in complex broadleaf forest stands," *Forestry*, vol. 95, no. 3, pp. 363-379, 2022. https://doi.org/10.1093/forestry/cpac001

Zhimin Wang https://orcid.org/0009-0006-8730-6185

He is a professor and master tutor of Zhongyuan University of Technology. He has presided over five provincial projects such as Henan Provincial Social Science Planning and Consulting Project, Henan Provincial Soft Science Project, Henan Provincial Government Decision-making Consulting Project, and seven departmental projects such as Henan Provincial Department of Education Social Science Planning Project, and as the main author of 15 national and provincial projects such as National Natural Science Foundation of China, National Soft Science, Henan Provincial Soft Science, etc. The Henan Provincial Soft Science Project presided over won the second prize of Henan Provincial First Natural Science, and the Middle Plain Economic Zone Special Project presided over won the prize of Henan Provincial Department of Education Humanities and Social Science Research Results. The project hosted by him won the second prize of the first natural science of Henan Province, and the special project hosted by him in the Central Plains Economic Zone won the second prize of the humanities and social science research results of Henan Provincial Education Department. He has published five textbooks and books, and 36 academic papers in CSSCI source journals and Chinese core journals, such as *Economic Vertical*, *Statistics and Decision*, *Price Theory and Practice*, and *Enterprise Economy*.