



# Research on Intelligent Correction of Abnormal Data in Online Education System Based on Big Data Technology

Yana Tang<sup>1</sup>(✉) and Jing Zhao<sup>2</sup>

<sup>1</sup> South China Institute of Software Engineering GU, Guangzhou 510990, China

<sup>2</sup> Xi'an International Studies University, Xi'an 710128, China

**Abstract.** In the traditional correction method of abnormal data in online education system, the unreasonable selection of training samples by RBF neural network leads to a large error in the correction of abnormal data. Therefore, GA algorithm is used to improve RBF neural network, and a new intelligent correction method of abnormal data in online education system based on big data technology is proposed. The classification model of abnormal data in online education system is constructed by decision tree classification algorithm. The pretreatment of abnormal data is completed based on big data technology. The specific pre-processing steps include: data cleaning, data integration, data transformation, data reduction, dimension reduction, numerical reduction, data discretization and concept layering. GA-RBF neural network is used to correct abnormal data of online education system. By comparing the performance of this method with the traditional intelligent correction method of abnormal data in online education system, it can be seen that the prediction and filling accuracy of this method is higher than that of the traditional method, and the performance is improved.

**Keywords:** Big data technology · Online education system · Abnormal data · Smart correction

## 1 Introduction

For many years, there has never been a unified understanding of the definition of abnormal data. Researchers in different fields have given different definitions of abnormal data from different perspectives. Dixon regards abnormal data as “data that is suspected in data analysis”. Grubbs considers abnormal data to be those data points that deviate significantly from other sampled data. Elashoff stated that abnormal data are some extreme data points to some extent [1]. It was not until the mid-1980s that the definition of abnormal data became more complicated. Guttman regards abnormal data as pseudo data that does not obey the law of data distribution; Rohlf believes that points not in the data cloud can be considered as potential abnormal data; Hawkins defines abnormal data as: far away from most of the data and the true distribution density may be low Some of the data; Campbell defines multivariate abnormal data as: data points in the form of

relationships that are significantly different from important data [2]. In the 1990s, some authors began to describe abnormal data from a statistical perspective. Anscombe and Tukey believe that abnormal data should have larger residuals; Portnoy pointed out that the residuals of abnormal data are greater than the residuals of most other data, and these residuals are obtained by linear regression methods. And he further pointed out: If the residual error of a data is greater than five times the standard deviation, then the data is considered abnormal data; in 1983, Beckman and Cook pointed out that the definition of abnormal data is still very vague. Barnett further pointed out that “the important thing is not whether any observation data is outside  $F$ , but whether the largest observation data under  $F$  is abnormally huge”. Comrey believes that abnormal data is caused by incorrect measurement and data contamination [3]. There are also many abnormal data in online education systems. These abnormal data mainly refer to data that affect the operation of the online education system. For these abnormal data, it must be intelligently corrected to ensure the normal operation of the online education system.

At present, there are also some relevant research results, such as: Literature [4] is based on sliding basic Windows sampling (SBWB) and Gaussian process regression This paper presents a regression (GPR) model based on SBWS\_GPR prediction model for anomaly detection of uncertain multi-data in online education systems. In the historical data set collected based on time series, index number is introduced to cluster the historical data set of online education system, the mapping relationship between data set and index number is analyzed, and the input data obtained in real time is distributed through the sliding window to match, so as to realize the abnormal point detection and correction of single data of online education system. Literature [5] used fuzzy C-means clustering to obtain the clustering center curve of online education system data, and corrected the abnormal data according to the identified abnormal data types of online education system by combining the clustering center curve and the correction algorithm. However, the above methods have the problem of low accuracy of abnormal data prediction and filling.

Therefore, big data technology is applied to the intelligent correction of abnormal data in the online education system, and an online education system abnormal data based on big data technology is proposed. Smart correction method. Decision-tree classification algorithm is introduced to construct the classification model of abnormal data in online education system. The pretreatment of abnormal data is completed based on big data technology. GA-RBF neural network is used to correct abnormal data of online education system. The experimental results show that the accuracy of this method is higher than that of the traditional method, and its application performance is better than the traditional method.

## **2 Design an Intelligent Correction Method for Abnormal Data in an Online Education System Based on Big Data Technology**

### **2.1 Abnormal Data Detection**

First, the abnormal data detection of the online education system is carried out, and the classification model of the abnormal data of the online education system is constructed

through the decision tree classification algorithm, and the abnormal data of the online education system is detected through this model [4]. The steps to build an abnormal data classification model for the online education system are as follows:

The core part of the decision tree classification model construction is the split feature selection. The online education system abnormal data sample set is divided into different branches according to the abnormal data feature value of the online education system, so that each branch corresponds to the child node as “pure” as possible, that is, try to make The sample types in each subset are the same. The construction result of the decision tree is a tree model (binary or multi-branch). When the partition rule is Boolean logic or the classification target is binary classification, a binary tree model is generally generated [6]. The general flow of decision tree construction is shown in Fig. 1.

In the actual application scenarios of abnormal data detection in online education systems, the constructed decision tree classification model is often too “lush”. Each segmentation point is fully used, and the leaf nodes are almost completely “pure”. The model is too in line with the characteristics of the training sample, so the effect of classifying the training sample is excellent, but the effect of classifying the test data may be much worse, which is “overfitting”. The versatility of the over-fitting model is not strong, so it needs to be pruned to make it more simplified. Combining pre-pruning and post-pruning methods to pruning the constructed decision tree classification model.

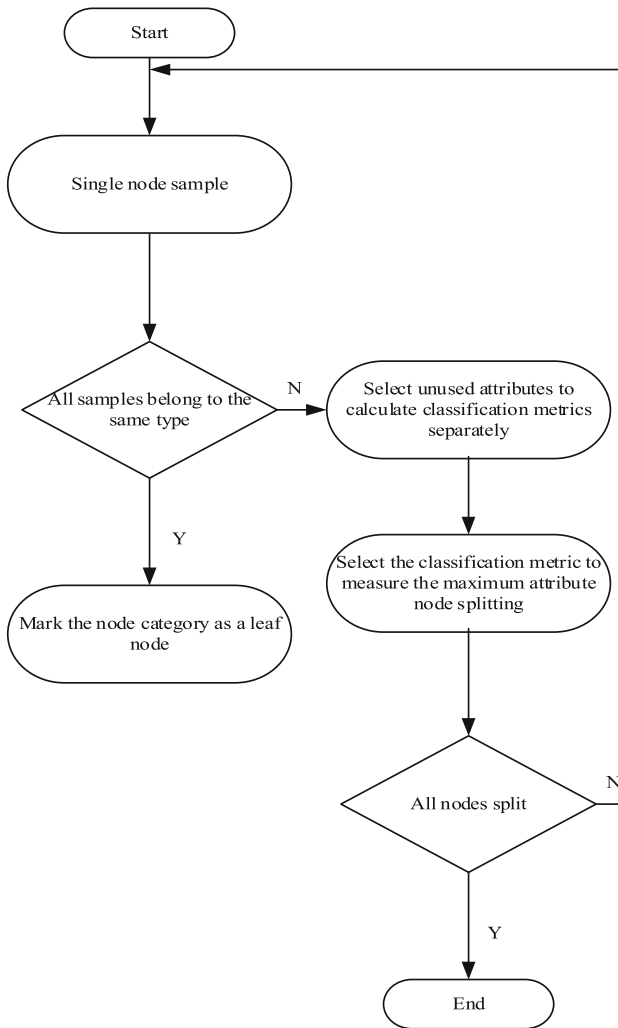
The pre-pruning method is as follows: Pre-pruning is the pruning performed during the construction of the decision tree. In order to avoid over-fitting when the number of nodes is too large, a condition to stop continuing to create branches can be set in advance. When samples of each category in a sample set of a node meet this condition, the splitting of the node will be stopped without waiting for all the splitting attributes All are used up, even if the node can be split more purely. Finally, use the category with the largest number of samples to mark the node category and use it as a leaf node. The pre-pruning algorithm is simple to use and suitable for use in scenarios where the decision tree model is relatively small.

The post-pruning method is as follows: Post-pruning is pruning after the construction of the decision tree is completed. The specific method is to merge all the node samples rooted at a certain node. If the increase in the amount of information of the sample set after the merging is less than a certain threshold, the subtree is replaced with a leaf node, and the category of the leaf node is subject to a “majority vote” Decision [7, 8]. Typical post-pruning algorithms include error rate reduction pruning and pessimistic pruning. The post-pruning operation is relatively complicated, which is suitable for use in scenes with large-scale models and high precision requirements.

## 2.2 Preprocessing of Abnormal Data

Then, based on big data technology, preprocess the abnormal data of the online education system. Specific preprocessing steps include: data cleaning, data integration, data transformation, data reduction, dimensionality reduction, numerical reduction, data discretization, and concept layering.

The purpose of the data cleaning routine is to fill in missing values, smooth noise and identify outliers, and correct inconsistencies in the data [8]. First fill in missing values with the mean of the attributes. Then smooth the noise data, the specific steps



**Fig. 1.** The general flow of decision tree construction

are as follows: the outliers can be detected by clustering, and similar values can be organized into groups or “clusters”. Intuitively, values that fall outside the cluster set are considered outliers. Then perform data cleaning, the first step is deviation detection. When discovering noise, outliers, and unusual values that need to be investigated, you can use existing knowledge about the nature of the data. This kind of knowledge or “data about data” is called metadata. Examine the domain and data type of each attribute, the acceptable value of each attribute, and a length range of the value; examine whether all values fall within the expected range and whether there are known dependencies between attributes; grasp the data Trends and recognition anomalies, such as values that are more than two standard deviations away from the mean of a given attribute may be marked

as potential outliers. Another type of error is the inconsistency of the source code and the inconsistency of the data representation. Field overload is another source of error. Observing the data must follow the uniqueness rule, continuity rule and null value rule. Other external materials can be used to manually correct some data inconsistencies [9, 10]. For example, errors in data input can be corrected using paper records. But most errors require data transformation.

Step 2: Correct the deviation. That is, once deviations are found, it is usually necessary to define and use (a series of) transformations to correct them. Choose to write a customized program for this step of the data cleaning process. The two-step process of deviation detection and correction is performed iteratively. As the knowledge of the data increases, it is important to continuously update the metadata to reflect this knowledge. This helps speed up the data cleaning of future versions of the same data store.

In data integration, there are many problems to be solved. The first is the problem of pattern integration and object matching. The matching of real-world equivalent entities from multiple information sources involves entity recognition. Determine whether the customer id in one database is the same attribute as the cust number in another database. The metadata of each attribute can be used to help avoid schema integration errors. Metadata can also be used to help transform data.

Redundancy is another important issue. An attribute may be redundant if it can be “derived” by another or another set of attributes. Inconsistent attribute or dimension naming can also lead to redundancy in the result data set. Some redundancy can be detected by correlation analysis. Given two attributes, this analysis can measure the extent to which one attribute implies the other based on the available data. For numerical attributes, the correlation between attributes A and B is calculated to estimate the correlation between these two attributes.

In addition to detecting redundancies between attributes, duplicates should also be detected at the tuple level. The use of denormalized tables may also lead to data redundancy. Inconsistencies usually occur between different copies, due to incorrect data input, or due to part of the updated data, but not all occurrences.

The third important issue of data integration is the detection and processing of data value conflicts. For example, for the same entity, attribute values from different data sources may be different. This may be due to different representations, scales or encodings. When matching the attributes of one database with another during data integration, the structure of the data should be considered to ensure that the attribute functional dependencies and reference constraints in the original system match those in the target system. The heterogeneity and structure of data semantics pose a huge challenge to data integration. Careful integration of data from multiple data sources can help reduce and avoid redundancy and inconsistencies in the resulting data set.

The purpose of data transformation is to transform data into a unified form. Data transformation mainly involves the following contents:

- (1) Smooth: Remove the noise in the data. This technique includes binning, regression, and clustering.
- (2) Aggregation: aggregate or aggregate data.
- (3) Data generalization: Use concept layering to replace low-level or “original” data with high-level concepts.

- (4) Standardization: Scale the attribute data proportionally to make it fall into a small specific interval.
- (5) Attribute construction (or feature construction): New attributes can be constructed and added to the attribute set. Attribute construction is the construction of given attributes and adding new attributes to help improve the understanding of high-dimensional data structures.

The data specification first requires data cube aggregation: the data cube stores multi-dimensional aggregation information. Each unit stores an aggregate value, which corresponds to a data point in a multi-dimensional space. Each attribute may have a conceptual hierarchy, allowing data analysis at multiple abstract layers. The data cube provides quick access to pre-calculated summary data, so it is suitable for intelligent correction of abnormal data in online education systems.

The cubes created at the lowest level of abstraction are called primitive cubes. The basic cube should correspond to the individual entity of interest. That is, the lowest level should be available or useful for analysis. The cube at the highest level of abstraction is called a vertex cube [11]. The data cubes created for different abstraction layers are called cubes, so the data cubes can be regarded as the grid of cubes. Each higher level of abstraction will further reduce the size of the resulting data. When performing intelligent correction of abnormal data in the online education system, the smallest available cube associated with a given task should be used.

Then proceed to attribute subset selection. The heuristic methods for attribute subset selection include the following:

- (1) Step by step forward selection: The process starts with an empty attribute set as the reduction set, determines the best attribute in the original attribute set, and adds it to the reduction set. In each subsequent iteration, the best attribute in the remaining original attribute set is added to the set.
- (2) Delete backward step by step: The process starts with the entire attribute set. At each step, delete the worst attribute still in the attribute set.
- (3) Combination of forward selection and backward deletion: It is possible to combine the stepwise forward selection and backward deletion methods, each step selects a best attribute, and deletes a worst attribute from the remaining attributes [12].
- (4) Decision tree induction: The decision tree algorithm was originally used for classification. Decision tree induction constructs a structure similar to a flowchart, in which each internal (non-leaf) node represents a test of an attribute, and each branch corresponds to an output of the test; each external (leaf) node represents a class prediction. At each node, the algorithm selects the “best” attribute and divides the data into categories. When decision tree induction is used for attribute subset selection, the decision tree is constructed from the given data. All attributes that do not appear in the tree are assumed to be irrelevant. The attributes appearing in the tree form a reduced attribute subset. The end criteria of the method can be different. The process can use a degree threshold to decide when to stop the attribute selection process.

Dimensional reduction uses data encoding or transformation to obtain a reduced or “compressed” representation of the original data. The method used is principal component analysis.

Principal component analysis is to search for  $k$   $n$ -dimensional orthogonal vectors that best represent the data, where  $k \leq n$ . In this way, the original data is projected into a much smaller space, leading to dimensionality reduction. PCA works by creating a replacement, smaller set of basic elements of “combined” attributes. The original data can be projected into this smaller set. PCA often reveals previously undetected connections and therefore allows interpretation of unusual results. The basic process is as follows:

- (5) Normalize the input data so that each attribute falls into the same interval. This step helps to ensure that attributes with larger domains do not dominate attributes with smaller domains [13].
- (6) PCA calculates  $k$  orthonormal vectors as the basis of normalized input data. These are unit vectors, and each direction is perpendicular to the other. These vectors are called principal components. The input data is a linear combination of principal components.
- (7) Arrange the principal components in descending order of “importance” or intensity. The principal components basically act as the new axis of the data, providing important information about the variance. In other words, sort the coordinate axes so that the first coordinate axis displays the largest variance of the data, the second displays the second largest variance, and so on.
- (8) The principal components are arranged in descending order of “importance”, and the size of the data can be reduced by removing the weaker components (that is, the variance is smaller). Using the strongest principal component should be able to reconstruct a good approximation of the original data [14].

PCA has low computational overhead, can be used for ordered and disordered attributes, and can handle sparse and skewed data. Multidimensional data with more than 2 dimensions can be handled by reducing the problem to a 2-dimensional problem. Principal components can be used as input for multiple regression and cluster analysis.

Numerical reduction technology refers to the selection of alternative, “smaller” data representations to reduce the amount of data. The method used is clustering: clustering technology treats data tuples as objects [15, 16]. It divides objects into groups or clusters, so that objects in one cluster are “similar” to each other, but “different” from objects in other clusters. Generally, similarity is based on a distance function, defined by the “closeness” of objects in space. The “mass” of a cluster can be expressed by its diameter, which is the maximum distance between any two objects in the cluster. Centroid distance is another measure of cluster quality, defined as the average distance from the cluster centroid (representing the “average object”, or average point in the cluster space) to each cluster object [17].

In data reduction, the actual data is replaced with the cluster representation of the data. The effectiveness of this technique depends on the nature of the data. If the data can be organized into different clusters, this technique is much more effective. In the database system, the multidimensional index tree is mainly used for quick access to data. It can also be used to reduce hierarchical data and provide multi-dimensional clustering of data. This can be used to provide approximate answers to queries. For a given set of data objects, the index tree recursively divides the multidimensional space, and its root node represents the entire space. Usually, this kind of tree is balanced and consists of internal nodes and leaf nodes. Each parent node contains keywords and pointers to child nodes, and the child nodes together represent the space represented by the parent node [18]. Each leaf node contains a pointer (or actual tuple) to the data tuple it represents.

In this way, the index tree can store aggregate and detailed data at different resolutions or abstract layers. It provides hierarchical clustering of data sets, in which each cluster has a label to store the data contained in the cluster. If we regard each child of the parent node as a bucket, the index tree can be seen as a hierarchical histogram [19]. Similarly, each bucket is further divided into smaller buckets, allowing data to be aggregated at a finer level. The use of a multidimensional index tree as a form of data reduction depends on the order of attribute values on each dimension. Two-dimensional or multi-dimensional index trees include R-trees, quad-trees and variants of Chuangmen. They are very suitable for processing sparse data and skewed data [20, 21].

**Data discretization and concept stratification:** Among them, the method used for data discretization is as follows: Based on the discretization: The discretization is one of the most commonly used measures of discretization. Di-based discretization is a supervised, top-down splitting technique. It uses the class distribution information when calculating and determining the split point (the data value that divides the attribute interval) [22]. For discrete numerical attribute A, select the value of A with the smallest direct value as the split point, and divide the result interval recursively to obtain hierarchical discretization. This discretization forms a conceptual hierarchy of A.

The method used for concept hierarchy is as follows: The partial order of attributes is explicitly stated by the user or expert at the schema level: Generally, the conceptual hierarchy of classification attributes or dimensions involves a set of attributes. Users or experts can easily define the concept hierarchy by explaining the partial order or total order of attributes at the pattern level [23].

### 2.3 Intelligent Correction of Abnormal Data

The abnormal data detected in the online education system is corrected by GA-RBF neural network, and the specific process of correcting abnormal data is shown in Fig. 2.

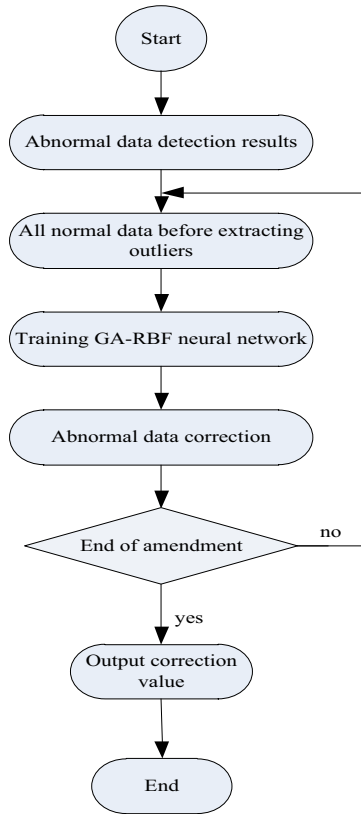


Fig. 2. Intelligent correction process of abnormal data

### 3 Experimental Verification

#### 3.1 Experimental Design

The performance of the designed method of intelligent correction of abnormal data in the online education system based on big data technology is experimentally verified. Intelligent correction method of abnormal data of online education system based on big data technology is used to intelligently correct abnormal data of an online education system. The course data structure of the system is shown in Table 1.

Obtain the abnormal data prediction filling accuracy data of the online education system abnormal data intelligent correction method based on big data technology as the experimental data. In order to enhance the comparison of the experimental results, the original two online education system abnormal data intelligent correction methods were used as the comparison methods in the experiment to conduct comparative experiments. Including the intelligent correction method of online education system abnormal data based on AR model and the intelligent correction method of online education system abnormal data based on neural network algorithm. These two methods are also used to conduct intelligent correction experiments of abnormal data of experimental online

**Table 1.** Course data structure of the system

| Type of data            | Field         | Type of data | Field interpretation            | Is it empty | Data length |
|-------------------------|---------------|--------------|---------------------------------|-------------|-------------|
| Course information data | Name          | Varchar      | Course title                    | No          | 20          |
|                         | Summary       | Varchar      | Course summary                  | Yes         | 500         |
|                         | Category      | Int          | Content classification          | No          | 11          |
|                         | Num           | Int          | Total number of videos          | No          | 11          |
|                         | Id            | Number       | Course id                       | No          | 10          |
|                         | Id of teacher | Number       | Teacher id                      | No          | 10          |
|                         | Id of resouce | Number       | Course resource id              | No          | 10          |
|                         | Time          | Number       | Duration                        | No          | 10          |
| Teaching video data     | Index         | Int          | Video sequence number           | No          | 11          |
|                         | Path          | Varchar      | Video path                      | No          | 100         |
| Video associated data   | Time          | Varchar      | Point in time                   | No          | 20          |
|                         | Title         | Varchar      | Associate video title           | No          | 100         |
|                         | Summary       | Varchar      | Introduction to related courses | Yes         | 200         |
|                         | Relatedid     | Int          | Associated video id             | No          | 11          |

education system courses, and the abnormal data prediction filling accuracy data of these two methods are obtained as comparative experimental data.

### 3.2 Analysis of Experimental Results

There is an intelligent correction method of abnormal data in online education system based on big data technology, a method in literature [4] and a method in literature [5]. The comparison between the prediction accuracy of abnormal data and the filling accuracy of experimental data is shown in Table 2.

According to the comparison between the accuracy of the prediction and filling of abnormal data in Table 2 and the experimental data, it can be seen that the prediction and filling accuracy of the intelligent correction method of online education system based on big data technology is higher than that of the traditional method.

**Table 2.** Comparison of experimental data for prediction and filling accuracy of abnormal data

| The number of abnormal data test samples (a) | Abnormal data prediction filling accuracy rate (%) |                           |                           |
|--|--|---------------------------|---------------------------|
|  | Method based on big data technology                | Methods in literature [4] | Methods in literature [5] |
| 500  | 99.35  | 94.20                     | 95.80                     |
| 600  | 99.32  | 94.18                     | 95.74                     |
| 700  | 98.65  | 94.15                     | 95.64                     |
| 800  | 98.28  | 94.13                     | 95.58                     |
| 900  | 98.17  | 94.08                     | 95.47                     |
| 1000   | 98.04  | 93.95                     | 95.37                     |
| 1100   | 97.98  | 93.85                     | 95.24                     |
| 1200   | 97.92  | 93.74                     | 95.05                     |
| 1300   | 97.81  | 93.28                     | 95.01                     |
| 1400   | 97.80  | 93.10                     | 95.00                     |

## 4 Concluding Remarks

The intelligent correction method of abnormal data in online education system based on big data technology realizes the improvement of abnormal data prediction and filling accuracy. This method is of great significance to the operation and promotion of online education system. In practical application, the application of this method can optimize the operating quality of the online education system, and the correction of abnormal data can improve the efficiency of the online education system.

## References

1. Zou, X., Azam, M., Islam, T., et al.: Environment and air pollution like gun and bullet for low income countries: war for better health and wealth. *Environ. Sci. Pollut. Res.* **23**(4), 3641–3657 (2020)
2. Liu, F., Choi, D., Xie, L., et al.: Global spectral clustering in dynamic networks. *Proc. Natl. Acad. Sci. U.S.A.* **115**(5), 927–932 (2018)
3. Parisi, C.A., Ritchie, C., Petriz, N., et al.: Chronic urticaria in a health maintenance organization of Buenos Aires, Argentina - new data that increase global knowledge of this disease. *An. Bras. Dermatol.* **93**(1), 76–79 (2018)
4. Helma, C., Cramer, T., Kramer, S., et al.: Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds. *J. Chem. Inf. Comput.* **35**(4), 1402–1411 (2018)
5. Liu, Z., Liu, C., Ding, J., Zhang, H.: Characterization and modelling of high temperature flow behaviour of V modified 2.25Cr-1Mo heat resistant steel plate. *J. Wuhan Univ. Technol.-Mater. Sci. Ed.* **35**(1), 192–199 (2020). <https://doi.org/10.1007/s11595-020-2243-3>
6. Kenselyn, O., Boff, A.L., Rangel, B.R.: Verrucous hemangioma and histopathological differential diagnosis with angiokeratoma circumscriptum nevirforme. *An. Bras. Dermatol.* **93**(5), 712–715 (2018)

7. Schuemie, M.J., Hripcsak, G., Ryan, P.B., et al.: Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proc. Natl. Acad. Sci. U.S.A.* **115**(11), 2571–2577 (2018)
8. Kwietniewski, L., Schreyögg, J.: Profit efficiency of physician practices: a stochastic frontier approach using panel data. *Health Care Manag. Sci.* **21**(1), 76–86 (2016). <https://doi.org/10.1007/s10729-016-9378-y>
9. Palumbo, J.R., et al.: Assessing associations between indoor environment and health symptoms in Romanian school children: an analysis of data from the SINPHONIE project. *Environ. Sci. Pollut. Res.* **25**(9), 9186–9193 (2018). <https://doi.org/10.1007/s11356-018-1568-3>
10. Ho, L.S., Wheeler, E.: Using program data to improve access to family planning and enhance the method mix in conflict-affected areas of the democratic Republic of the Congo. *Glob. Health Pract.* **6**(1), 161–177 (2018)
11. Liu, S., Glowatz, M., Zappatore, M., et al. (eds.): *e-Learning, e-Education, and Online Training*, pp 1–374. Springer, Heidelberg (2018)
12. Liu, S., Li, Z., Zhang, Y., et al.: Introduction of key problems in long-distance learning and training. *Mob. Netw. Appl.* **24**(1), 1–4 (2019)
13. Liu, S., Bai, W., Zeng, N., et al.: A fast fractal based compression for MRI images. *IEEE Access* **7**, 62412–62420 (2019)
14. Bag, S., Aich, P., Islam, M.A.: Behavioral intention of “digital natives” toward adapting the online education system in higher education. **45**(3), 158–166 (2020)
15. Jiang, Y., Li, X.: Intelligent online education system based on speech recognition with specialized analysis on quality of service. *Int. J. Speech Technol.* (1), 12–24 (2020)
16. Al-Odeh, M.: What universities in the Middle East can learn from the American online education system. *Int. J. Inf. Commun. Technol. (IJ-ICT)* **9**(1), 31–40 (2020)
17. Guan, Y., Wang, L., Meng, Z.M.: Accurate attitude determination based on adaptive UKF and RBF neural network using fusion methodology for Micro-IMU applied to rotating environment. *Math. Probl. Eng.* **20**(1), 20–31 (2020)
18. Chen, X., Chen, H., Xu, H.: Vehicle detection based on multifeature extraction and recognition adopting rbf neural network on ADAS System. *Complexity* **2020**(2), 1–11 (2020)
19. Yang, W., Guo, J., Liu, Y., et al.: Multi-objective optimization of contactor’s characteristics based on RBF neural networks and hybrid method. *IEEE Trans. Magn.* 1–4 (2019)
20. Mohammadnezhad, M., Ghaffari, A.: Hybrid routing scheme using imperialist competitive algorithm and RBF neural networks for VANETs. *Wirel. Netw.* **25**(5), 2831–2849 (2019). <https://doi.org/10.1007/s11276-019-01997-6>
21. Ting, X.: Chemistry course network teaching based on key information search and big data cloud platform. *J. Intell. Fuzzy Syst.* (3), 1–12 (2020)
22. Ding, Y.: Performance analysis of public management teaching practice training based on artificial intelligence technology. *J. Intell. Fuzzy Syst.* **40**(5), 1–14 (2020)
23. Li, H.: Analysis of computer teaching pattern based on outlier data mining and machine learning. *J. Intell. Fuzzy Syst.* (2), 1–11 (2020)