Prediction of Large-Scale Instrument Usage Based on Catboost Algorithm for Science and Education Integration in Local Universities

Lin Tao[1], Jiading Bao[2*], Bao Zhu[3]

[1]Department of State-owned Assets Management, Guilin University of Electronic Technology, Guilin 541004, China
[2]Faculty of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin, 541004, China
[3]Faculty of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin, 541004, China

TaoLin@guet.edu.cn
*jdbao@guet.edu.cn
1459899709@qq.com

Abstract. Under the background of the "Double First-Class" university plan, the sharing mechanism of large-scale instruments plays an indispensable role in science and education integration in universities. This paper first collects usage data (from September 2020 to September 2022) of large campus-shared devices in a real-world setting. Second, based on the machine learning algorithm, CatBoost, this paper constructs a method to predict the usage of devices. The results show that the best prediction is achieved when the input time step is 3 (with MAE, MSE, and RMSE being 56.10, 7445.43, and 86.28, respectively). Based on the obtained prediction results, corresponding policy recommendations are proposed further. Finally, taking Guilin University of Electronic Science and Technology as an example, this paper illustrates the use cases of large campus-shared devices in the real world. The method in this paper provides concrete decision support for large campus-shared equipment managers to estimate the busy and idle periods of equipment usage and develop equipment maintenance plans.

Keywords: machine learning; time series; forecasting; science education integration

1 Introduction

The integration of science and education was put forward under the background of China's major decision and plans of "building first-class universities and first-class disciplines (together known as Double First-Class)" [1-2]. In October 2015, China's Overall Plan for Promoting the Construction of First-Class Universities and First-Class Disci-

H. Wu et al. (Eds.): CSIEDE 2022, ACSR 103, pp. 728–738, 2022.
https://doi.org/10.2991/978-94-6463-108-1_81

plines pointed out that it is necessary to promote the integration of science and education and improve the training mechanism for high-level scientific research to support top-notch innovative talents [3-4]. In June 2018, the Guiding Opinions on Accelerating the Construction of Double First-Class in Colleges and Universities pointed out that a collaborative training mechanism for the integration of science and education should be established to promote the organic combination of knowledge learning, scientific research, and ability training [5].

Due to the concept of "emphasizing scientific research and ignoring education," large-scale instruments and equipment have long been used more as high-quality scientific research resources in colleges and universities [6-7]. However, in the context of building China into a powerful country in education, it is important to promote the transformation of scientific research resources into teaching resources and develop the training mechanism for the integration of science and education [8-10]. At present, fully opening the access to large-scale instruments and equipment for teaching sharing is an important means for local colleges and universities to achieve the integration of science and education.

With the rapid development of higher education in China, local colleges and universities have invested a lot of money in purchasing large-scale instruments and equipment. The number of large-scale instruments and equipment has grown rapidly, but it has not changed the current situation of low participation in the undergraduate training process of large-scale instruments and equipment. With the gradual expansion of the unbalanced development of large-scale instruments and equipment among colleges and universities, the gap between general local colleges and universities and key colleges and universities in the sharing of large-scale instruments for teaching has become more prominent, which seriously limits the comprehensive development of integration of science and education. The dilemmas they face are as follows:

Relatively Insufficient Resources of Large-scale Instruments and Equipment. Take Guilin University of Electronic Science and Technology (a local university in China) as an example. The school has 1,354 large-scale instruments and equipment for teaching and scientific research that are worth above 100,000 yuan, with a total value of 489 million yuan, accounting for 47.7% of the total value of instruments and equipment; 298 large-scale instruments and equipment above 400,000 yuan, with a total value of 301 million yuan, accounting for 29.3% of the total. In Guangxi, about half of the large-scale instruments and equipment worth above 400,000 yuan in colleges and universities are concentrated in the only Project 211 institution (Guangxi University) in the province.

Prevailing Concept of "Research First." Local colleges and universities use large-scale instruments and equipment for scientific research purposes far more than teaching, which can be seen from the output of large-scale instruments. Take Guangxi as an example. In 2019, large-scale equipment undertaking teaching projects accounted for about 20% of the total projects undertaken, while scientific research projects accounted for about 70% of the total. This indicates that colleges and universities in Guangxi typically lay more emphasis on scientific research rather than teaching.

The Backward Idea of Sharing. Local colleges and universities are generally unwilling to open and share scarce large-scale instruments and equipment. Even if there are, they are open to scientific research teams only and completely closed to undergraduate teaching. In addition, the reform direction of teaching in some colleges and universities is simply to transfer scientific research equipment from scientific use to teaching use and does not allow students to master scientific research results, scientific research trends, and scientific research methods by sharing scientific research equipment.

The fundamental reason is that the lack of a science and education integration system makes it difficult to open up large-scale instruments and equipment to the whole process of undergraduate training through an effective collaborative sharing platform. As a result, teachers cannot apply advanced experimental methods to undergraduate teaching and integrate scientific research resources into the practical teaching process.

To better promote the usage of large-scale instruments, a prediction method based on the CatBoost algorithm was proposed in this paper, which was validated on a real university campus usage dataset and showed good adaptability. It can provide decision support for instrument managers in local universities.

The rest of this paper is organized as follows: Section 2 provides a detailed discussion of the methodology of this paper. The obtained results are discussed in Section 3. Finally, Section 4 summarizes the full paper and suggests future research directions.

2 Materials and methods

2.1 Categorical boosting

Categorical Boosting (CatBoost) is an improved gradient-boosted decision trees method proposed by Dorogush et al. in 2018 [11]. The other two variants of gradient boosting algorithms are eXtreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM). Though XGBoost and LightGBM have achieved massive advantages, CatBoost has emerged as a more promising algorithm in terms of accuracy and generalization ability.

Given a dataset $D = \{(x_k, y_k)\}_{k=1}^{n}$, where $x_k = (x_k^1, \cdots, x_k^{d'})$ is a d-dimensional feature vector and $y_k \in \mathbb{R}$ is the corresponding label, the trees in CatBoost are constructed by recursively partitioning the entire feature space. For convenience, decision tree *h* can be expressed as a superposition of estimated values of all regions:

$$h(x) = \sum_{j=1}^{J} b_j \pi_{x \in R_j} \tag{1}$$

where $\pi_{x \in R_j}$ is an indicator function:

$$\pi_{x \in R_j} = \begin{cases} 1 \text{ if } x \in R_j \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

In the iteration process, the idea of the greedy algorithm is used to minimize the expected loss:

$$F^t = F^{t-1} + \alpha h^t \tag{3}$$

where α is the step size and h^t is a tree selected from a series of H functions in the t-th iteration:

$$\begin{aligned} h^t &= \arg\min_{h \in H^\delta} (F^{t-1} + h) \\ &= \arg\min_{h \in H} EL(y, F^{t-1}(x) + h(x)) \end{aligned} \tag{4}$$

After N iterations, we obtain a series of approximate functions $F^t (t = 0,1,\cdots,N)$ and sum them to get the final results:

$$F(x) = \sum_{t=1}^{N} h^t \tag{5}$$

In the paper, we use the ordered boosting method to develop CatBoost, as shown in Algorithm 1.

Algorithm 1: Building a tree in CatBoost

Input: M, $\{(x_k, y_k)\}_{k=1}^{n}$, α , L , $\{h_i\}_{i=1}^{s}$,
$grad \leftarrow CalcGradient(L, M, y)$;
$r \leftarrow random(1, s)$;
$G \leftarrow (grad_{r,h_r(i)-1}(i)$ for $i = 1,\cdots,n)$;
$T \leftarrow$ empty tree;
For each step of the top-down procedure, **do**
  For each candidate split *c*, **do**
$T_c \leftarrow$ add split c to T;
For *i=1,…,n*
$\Delta(i) \leftarrow avg(grad_{r,h_r(i)-1}(p)$ for
p: $leaf_r(p) = leaf_r(i)$, $h_t(p) < h_r(i))$
$loss(T_c) \leftarrow \cos(\Delta, G)$
$T \leftarrow \arg\min_{T_c} (loss(T_c))$
$M_{r',j}(i) \leftarrow M_{r',j}(i) - \alpha \mathrm{avg}(grad_{r',j}(p)$ for
$p : leaf_{r'}(p) = leaf_{r'}(i), h_{r'}(p) \le j)$ for r'=1,...,s,
$i = 1,..,n, j \ge h_{r'}(p) - 1$;
Return T, M

2.2 Large instrument usage dataset of local universities

To study the usage of large instrument sharing platforms for science-education integration in local universities, we obtained a real-world dataset from September 1, 2020, to

August 31, 2022. The research data obtained for this paper are presented in detail in Figs. 1-2.

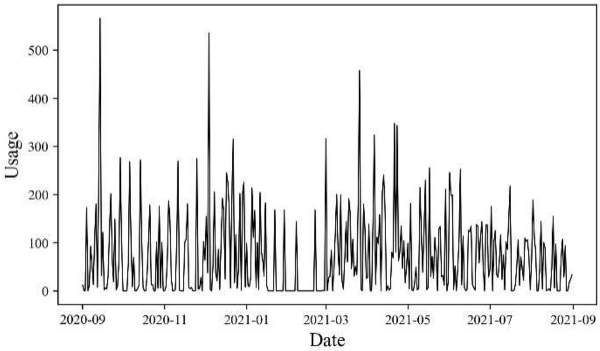

Fig. 1. Daily usage of large-scale instruments and equipment (2020/09/01-2021/09/01) [draw myself]

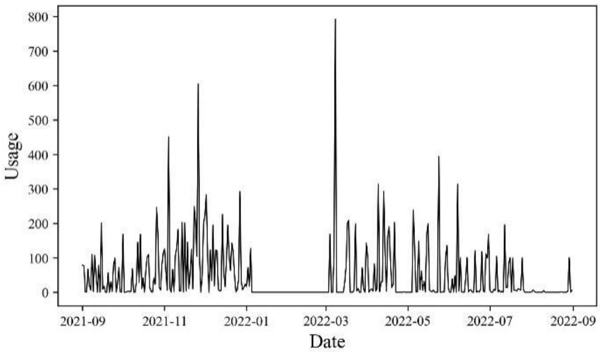

Fig. 2. Daily usage of large-scale instruments and equipment (2021/09/01-2022/09/01) [draw myself]

As seen in Figs. 1-2, there is a high degree of uncertainty in the use of large-scale instruments and equipment. For example, there is a long gap period between January 2022 and April 2022 and a peak after this gap. Notably, there is a peak and a gap period approximately every month. Therefore, the forecast of large instrumentation usage data can accurately assess the idle and busy periods of equipment in a certain period. This can provide an important reference for equipment managers to devise maintenance plans and usage specifications.

2.3 Evaluation metrics

To better evaluate the model prediction performance, we used diverse evaluation metrics, such as Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE), as shown in Eqs. (6-8).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | \hat{y}_i - y_i |, \tag{6}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2, \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}, \tag{8}$$

where $\hat{y}_i$ is the predicted value, y_i is the actual value, and n is the number of observations.

3 Results and discussions

3.1 Predicting results

The dataset was divided into a training set and a test set, with 70% of the data used for training and 30% for testing. Then, we evaluated the model effects when the input time steps were 3, 5, and 7, respectively. Table 1 and Table 2 respectively show the in-sample (i.e., training) and out-of-sample (i.e., testing) forecast performance.

Table 1. Training performance

Input Step	MAE	MSE	RMSE
3	34.94	2086.90	45.68
5	23.81	1047.22	32.36
7	18.19	666.76	25.82
9	13.15	362.19	19.03

Table 2. Testing performance

Input Step	MAE	MSE	RMSE
3	56.10	7445.43	86.28
5	57.00	7424.89	86.16
7	56.02	8324.16	91.23
9	55.82	8537.41	92.39

From Tables 1-2, it can be seen that the best prediction is achieved when the input time step is 3. When the input time step is 9, the model shows the risk of overfitting (i.e., the in-sample error is low, but the out-of-sample error is exceptionally high).

Fig. 3 shows the in-sample prediction results in detail. It can be seen that satisfactory training results have been achieved.

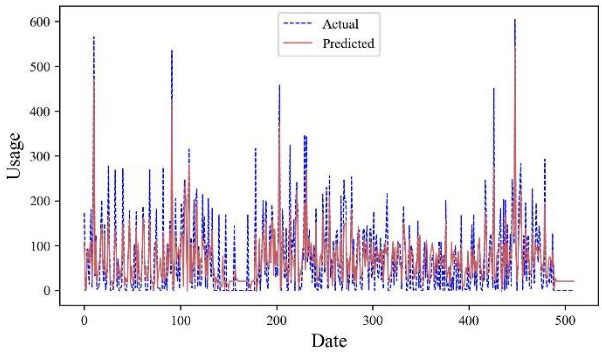

Fig. 3. In-sample prediction performance [draw myself]

3.2 Policy suggestions

The key for large-scale instruments and equipment to give full play to the function of "integration of science and education" lies in the reform of the traditional large-scale instruments and equipment management sharing mode. Local colleges and universities have phased in large-scale instruments and equipment mainly for undergraduate training. However, due to the lack of research on the uniqueness of individual colleges and universities, most of them copied the model of key colleges and universities, which made it difficult to implement. To fix this problem, the traditional management system of local colleges and universities should be reformed in a targeted manner, and improvement measures should be proposed in terms of teaching training, asset management, personnel assessment, and equipment operation, so as to establish a shared platform system for large-scale instruments and equipment that can be implemented in local colleges and universities.

To this aim, this paper first constructed a prediction system for large device usage by the CatBoost algorithm and validated it on real-world data from a university. The results showed that the MAE, MSE, and RMSE in the test set were 56.10, 7445.43, and 86.28, respectively, for an input step of 3, which was better than other methods. In addition, based on the prediction results, this paper provided suggestions for building a device management platform. Finally, a case study of an actual equipment operation management system at Guilin University of Electronic Science and Technology was conducted. The method in this paper provides a reference for the construction of a shared equipment management platform for large campuses and offers practical advice for managers to estimate the busy and idle periods of equipment and develop maintenance plans, etc.

Leverage the value of forecasting

As mentioned earlier, the use of large equipment on campus has great uncertainty, making it difficult to manage and for students to use. Therefore, a prediction system based on the usage data of large equipment can effectively help managers grasp the busy and idle periods of equipment and make management decisions such as maintenance plans and usage adjustments.

Incorporate the top-level design of teaching training

Colleges and universities can include the opening and sharing of large-scale scientific research equipment for the integration of science and education in the top-level design of undergraduate teaching and designate it as an important assessment indicator of scientific research platforms, as well as guide undergraduates to participate in scientific research practice to create a strong atmosphere for scientific research and education. All research institutions are required to list open laboratories, the proportion of undergraduates participating in scientific and educational collaborative projects should be more than 10%, and that of graduation projects (thesis) coming from scientific research projects should be over 50%. More than 80% of large-scale instruments and equipment in the laboratory should be open to undergraduates every year and cover more than 30% of the counterpart majors.

Expand the asset sharing system

Colleges and universities should identify and expand the sharing scope of large-scale instruments and equipment as much as possible to resolve the bottleneck of insufficient equipment. Within the school budget, instruments and equipment worth more than 100,000 yuan for teaching purposes and those more than 400,000 yuan for scientific research purposes can be included in the school sharing system. For equipment purchased with extra-budgetary funds, those worth more than 400,000 yuan will also be included in the school's sharing system, and those below 100,000 yuan can either be included in the school's sharing system or the unit's sharing system.

Develop assessment incentives

Higher education institutions should encourage more educators to join the rank of sharing large-scale instruments and equipment in scientific research and teaching, so as to change the traditional concept of "no opening and no sharing." Teachers with high professional titles can be encouraged to reform experimental classes for undergraduates through job title evaluation and merit evaluation. For managers of large-scale instruments and equipment, the operation assessment, management assessment, and benefits assessment of equipment shall be included in the job title evaluation.

Introduce paid sharing mechanism

Sharing large-scale instruments and equipment can be hard for local colleges and universities as they need financial support for operation and maintenance. This is why scientific research platforms with limited funding refuse to provide large-scale instruments and equipment for teaching activities free of charge. To solve the shortage of maintenance funds, a paid sharing mechanism must be introduced for equipment maintenance, performance development, and subsidy incentives. The paid sharing system should focus on the application of "science and education collaboration" of large-scale instruments and equipment. Once the instruments and equipment are included in the undergraduate experiment curriculum system, the cost of equipment usage will be subsidized to the equipment provider in full. In addition, those who recruit undergraduates to join the scientific research platform and guide them to use large-scale instruments and equipment for testing will be subsidized the margin from the cost of equipment usage.

4 Practices of scientific research and teaching collaborative platform for large-scale instruments in local universities

In recent years, Guilin University of Electronic Science and Technology has continuously explored the mechanism of building a scientific research and education collaborative platform for large-scale instruments and equipment in local colleges and universities. The school has implemented specific measures such as the "First-Class Undergraduate Education Action Plan" and "Large-scale Instruments and Equipment Open and Shared Management Measures" and launched a management and service platform for scientific research instruments sharing network.

The platform has the conventional functions of a large-scale equipment sharing system, such as equipment information management, user information management, reservation management, experimental information, and statistical management. More importantly, the platform possesses three core modules: "instrument usage statistics," "paid sharing," and "benefit assessment and evaluation," according to the improved plan of Guilin University of Electronic Science and Technology on the scientific research and education collaborative platform for large-scale instruments and equipment, which can shed some light on large-scale equipment operation.

4.1 Instrument usage statistics module

Statistics on the use of large-scale instruments and equipment can reflect the utilization rate of large-scale instruments and equipment and the sharing rate for teaching applications and provide essential data support for the decision-making of equipment managers. Managers can enter the personal name or select the department name for retrieval and check the instrument usage status of the designated person or designated department by year, month, week, and day, including the usage time, frequency of use, method of use, etc.

4.2 Paid sharing module

The billing function for large-scale instruments and equipment provides the basis for the reward and subsidy of the cost of equipment usage to the users and managers. This module mainly includes shared service information such as instrument reservation information, user affiliation/research group information, estimated price, contact information, expected completion time, and reservation order status.

4.3 Benefits assessment and evaluation module

This module collects and counts the annual benefits of instruments and equipment to reflect the utilization rate, direct economic income, and related discipline construction and teaching achievements, so as to provide data support for benefits assessment and evaluation. It is based on the annual benefits assessment form of valuable instruments and equipment in colleges and universities formulated by the Ministry of Education while conforming to the actual situation of local colleges and universities.

5 Conclusion

Large-scale instruments and equipment have gradually become important carriers for universities as they transition from scientific research advantages to educational advantages, and the integration of science and education for large-scale instruments and equipment has also become an important way for the implementation of "collaborative education with science and education" in colleges and universities. In response to the "Double First-Class" university plan, local colleges and universities, as a wider body of higher education institutions, represent the overall pace of China's higher education reform in "collaborative education with science and education." How to deepen the development of "collaborative education with science and education" through effective sharing of large-scale instruments and equipment has become the theme of the next reform. Currently, the sharing mechanism is in its infancy, which entails educators continuously innovating and promoting it in practice in the new era.

To address the above questions, this paper first collected usage data (2020-09 to 2022-09) of large campus-shared devices in a real-world setting. Second, based on the CatBoost algorithm, this paper constructed a method to predict the usage of devices. The prediction results showed that the best prediction was achieved when the input time step was 3 (MAE, MSE, and RMSE are 56.10, 7445.43, and 86.28, respectively). Based on the obtained prediction results, this paper further proposed corresponding policy recommendations. Finally, the Guilin University of Electronic Science and Technology was taken as an example to illustrate the use of large campus-shared devices in the real world. The method in this paper provides concrete decision support for large campus-shared equipment managers to estimate the busy and idle periods of equipment usage and develop equipment maintenance plans.

Acknowledgements

This research was supported by the Guangxi Higher Education Undergraduate Teaching Reform Project under Grant 2017JGB240 and 2022JGB191.

References

1. Z. Liu. Studying and Implementing the Spirit of the 19th National Congress of the Communist Party of China Intensively to Speed up the Construction of a Powerful Nation in Education [J]. Educational Research, 2017, 38(12): 4-13.
2. J. Yang. Responsibility, Mission and Action: First-Class University Constructure in the New Era [J]. Academic Degrees & Graduate Education, 2018(9): 1-5.
3. G. Chen, H. Zhang. Predicaments, Countermeasures, and Practices of Research-nurtured Cultivation in the Context of "Double First-Class" Construction [J]. Journal of Nanjing University of Science and Technology (Social Sciences), 2019, 32(5): 67-71.

4. L. Shan, L. Xu, W. Xia, et al. Discussion on the key role of coordinated education of science and education in Zhejiang University [J]. China University Science & Technology, 2019(08): 46-49.
5. W. Xu, R. Yao, H. Ning. Management of university instruments and equipment for undergraduate experimental teaching [J]. China University Science & Technology, 2019(5): 24-26.
6. H. Zhang, G. Li, G. Liang. Exploration and research on open sharing management system of large-scale instruments and equipment [J]. Experimental Technology and Management, 2017, 34(3): 257-259.
7. L. Tao, X. Zhu, Y Duan. Research on sharing management of large-scale instruments and equipment for undergraduates in local colleges and Universities [J]. Laboratory Science, 2018, 21(5): 197-199,203.
8. L. Wang, Y. Hou. Exploration and practice of innovating large-scale instruments and equipment management systems [J]. Experimental Technology and Management, 2016, 33(6): 259-262.
9. X. Liang, W. Li, Z. Wang, et al. Practice and exploration of opening and sharing large scientific research instruments to the society [J]. Experimental Technology and Management, 2019(9): 272-275.
10. S. Wu. Construction of sharing platform for large-scale precision instruments and equipment in colleges and universities [J]. China Management Informationization, 2018, 21(4): 190-191.
11. Dorogush A V, Ershov V, Gulin A. CatBoost: gradient boosting with categorical features support[J]. arXiv preprint arXiv:1810.11363, 2018.