ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THPV036

LASER DRIVER STATE ESTIMATION ORIENTED DATA GOVERNANCE

J. Luo†, L. Li, Z. G. Ni, X.W. Zhou, Institute of Computer Application, China Academy of Engineering Physics, Mianyang City, China

Abstract

Laser driver state estimation is an important task during the operation process for the high-power laser facility, by utilizing measured data to analyze experiment results and laser driver performances. It involves complicated data processing jobs, including data extraction, data cleaning, data fusion, data visualization and so on. Data governance aims to improve the efficiency and quality of data analysis for laser driver state estimation, which focuses on 4 aspects — data specification, data cleaning, data exchange, and data integration. The achievements of data governance contribute to not only laser driver state estimation, but also other experimental data analysis applications.

INTRODUCTON

Laser driver state evaluation is an important part of the business process of a laser shooting experiment, which mainly includes physical experiment results evaluation and driver performances evaluation. The energy, pulse power waveform and near-field results of the experimental process are extracted by using the measurement data obtained by acquisition equipments such as energy card meters, oscilloscopes and CCDs in the diagnostic system, and the energy dispersion, power imbalance and other results of a shooting experiment, as well as the gain capacity, output capacity and output stability of the driver are calculated, to evaluate the effectiveness of physical experiments and the performance status of the laser driver.

By adopting the Oracle relational database platform, all the experimental measurement data of the facility are stored [1, 2]. The early database design defined the physical structure and logical structure specification of experimental data storage, but there were no constraints on the storage format of experimental data, so there existed inconsistencies in the storage format of the same type of experimental data. Secondly, there are invalid measurements or missing measurements in the process of experimental data measurement, but such measurement data are not screened when the experimental data are stored in the database, resulting in experimental data quality problems such as missing or abnormal data items. In addition, with the laser facility putting into experimental operations, a large number of experimental measurement data are accumulated, and multiple application systems involving experimental data processing and storage are generated. For these late emerging application systems, most of them do not follow the early database design specifications in terms of data storage. Therefore, the experimental data relationship of the whole facility is complex.

The operational process of a shooting experiment.is described in Figure 1. As an important asset of the facility, the increasingly accumulated experimental data has played an important role in the operation control optimization of the facility in recent years, and the state evaluation of laser driver is one of the main contents. However, limited by the above problems such as inconsistent experimental data storage format, missing experimental data items, abnormal experimental data and complex experimental data relationship, data preprocessing consumes a lot of time, and application systems involving experimental data processing and analysis reveal common problems, such as slow development progress, repeated processing of experimental data, low computational efficiency, and uncertain analysis results, which coincides with the current situation of data preparation mentioned in literature [3].

Data governance aims to improve the efficiency and quality of data analysis and plays an important role in the whole big data analysis process [4]. In recent years, with the rapid development of big data analysis and application research, data governance has played a very significant role in enterprise big data mining, government public data sharing, industrial big data analysis, scientific research data management and other industries [5-12]. The state estimation of a laser driver involves a lot of jobs, like experimental data extraction, transformation, joint calculation and visualization. Considering the current situation of the experimental data quality and the problems faced by the data processing application system, we utilize a framework consisting with data governance and data visualization to implement the laser driver state estimation system. This paper mainly introduces and summarizes our work about the data governance part.

The next section provides an overview of laser driver state estimation. Section III illustrates the data governance. The computation of state parameters of laser driver is discussed in section IV. Finally, we conclude this paper in section V.

LASER DRIVER STATE EVALUATION

The main purpose of state estimation of laser driver is to evaluate the compliance of physical experiment and the performance of the laser driver. The energy, pulse power waveform and near-field results of the experimental process are extracted from the original measurements. And the key technical parameters such as power imbalance, output beam quality, output stability, gain capability and third harmonic efficiency of the laser driver are calculated. Furthermore, their evolution law is visually analyzed.

† luoj1987@caep.cn

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THPV036

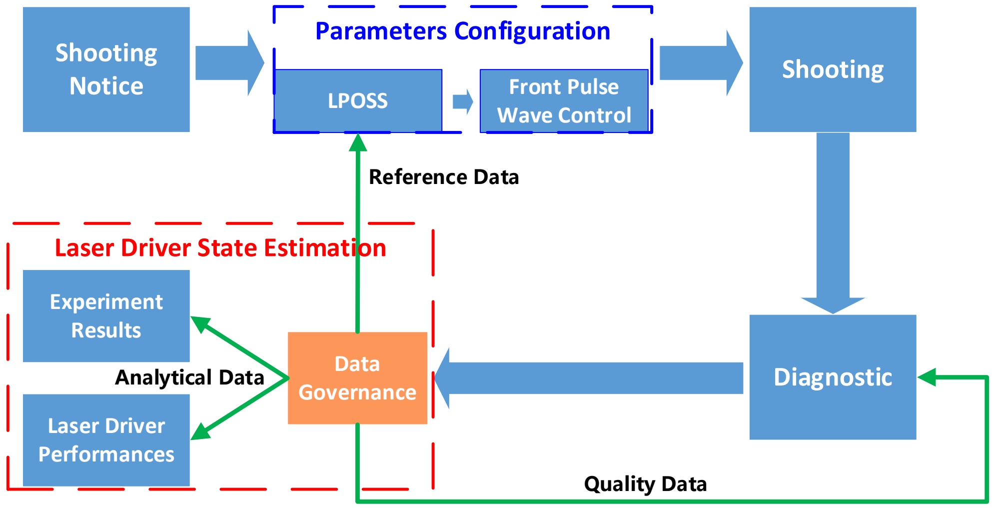

Figure 1: Business process of a shooting experiment.

Evaluation Contents

The state evaluation of the laser driver mainly includes two aspects: the evaluation of physical experiment results and the evaluation of laser driver performance.

a. State of physical experiment results: including output energy and pulse power waveform of each laser beam, total output energy, energy dispersion of each beam set, total average power and power imbalance. Among them, energy dispersion and power imbalance are two important criterion to judge whether the physical experiment meets the target or not, and they are also key factors to indicate the energy and power balance of the laser driver.

b. State of laser driver performance: including driver gain capability, output capability and output stability. Among them, the gain capability includes the main amplification gain and pre amplification gain, the output capability mainly refers to the peak power of each laser beam, and the output stability includes the frequency conversion efficiency and the energy deviation of each laser beam.

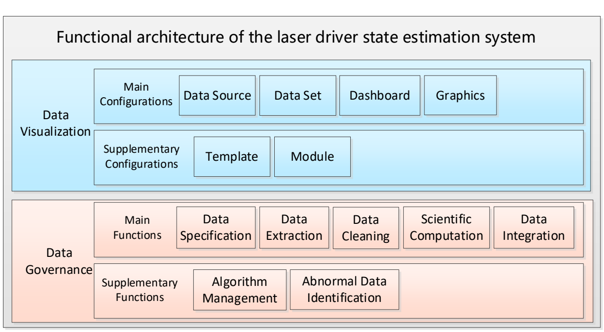

Figure 2: Functional architecture of the laser driver state estimation system.

Functional Architecture

The laser driver state estimation system mainly consists with two functional parts, namely data governance and data visualization, as shown in Figure 2.

Data visualization is mainly responsible for the interface display of driver state parameters, while data governance realizes the whole process of data extraction, cleaning, calculation, transformation and storage integration, as well as the functions of analysis and processing algorithm management and anomaly detection.

Technical Architecture

The laser driver state estimation system is composed of a hierarchical architecture, including data application layer, data management layer, data computation layer and data sources layer. The technical architecture is shown in Figure 3. Data governance mainly covers the data computing layer and the data management layer.

a. Data sources: in the process of experimental operation, the original experimental data is generated and stored in Oracle database. The data forms are diversified, including numerical value, text, picture and other data.

b. Data computation: extracting the original measurement data from the data source system, then cleaning the source data, detecting the abnormality and mark the quality of the measured data. Finally, combined with the statistical analysis algorithm, the state evaluation result data of the laser driver are obtained.

c. Data management: according to certain physical structure, logical structure and data format specifications, the analyzed state parameters are stored in the data warehouse for centralized management. In the data management layer, data is stored in Oracle relational database.

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THPV036

d. Data application: providing the data in the data warehouse system to the end user in the form of chart through the web interface. According to the analysis needs of users, report, query, multi-dimensional analysis and other functions are used to display data.

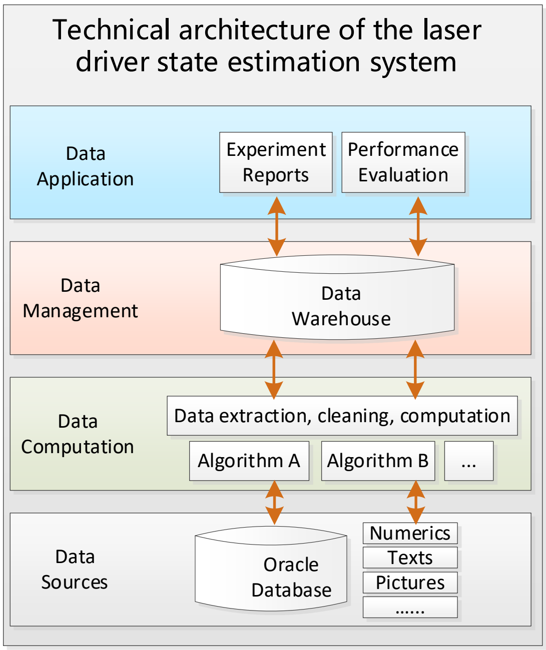

Figure 3: Technical architecture of the laser driver state estimation system.

DATA GOVERNANCE FOR LASER DRIVER STATE EVALUATION

The data governance is the process of reviewing and transforming data in order to make data better serving data analysis. It is the most energy consuming part in the whole data analysis process [3]. Although the original motivation of conducting the data governance is to conveniently serve the laser driver state evaluation system, once the data governance is carried out, it becomes a task of dynamically constructing data processing process for various business systems. Data governance pre-processes data for specific problems, adopts different data processing algorithms for different problems, and some data governance achievements can be shared among multiple applications. The final result of data governance for laser driver state estimation can serve all relevant data users in the whole facility, as shown in Figure 1.

Data governance mainly includes four aspects: data specification, data cleaning, data exchange and data integration. The work in these four aspects is described in detail below.

Data Specification

Aiming at the problem that the data formats of the same type of data are inconsistent, the data formats of pulse power waveform data and CCD image data are designed uniformly.

Pulse power waveform data is represented by [time column, power column] two-dimensional double array. The background measurement results outside the effective time window in the original measurement waveform are no longer retained, and only the part within the effective time window in the original measurement waveform is selected, as depicted in Figure 4.

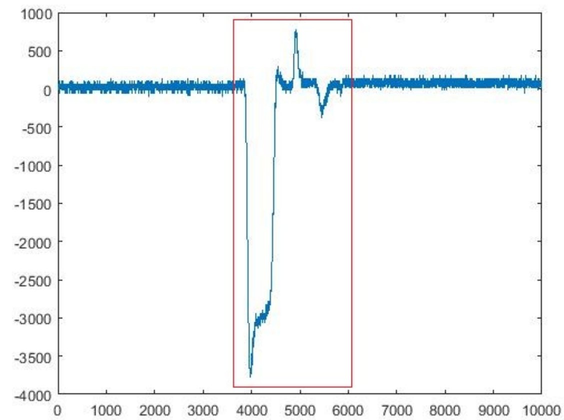

Figure 4: Example of pulse waveform data (red rectangle indicates the effective time window).

CCD image data is represented according to integer pixel matrix. Each element in the matrix represents the pixel gray-scale value at the corresponding position of the image. In order to facilitate subsequent data analysis, CCD image data is divided into two types, original measured and intended clipped. The matrix dimensions of these two types are different. The original measurement is fixed at 1024 * 1024 scale, and the pixel matrix dimension of the clipped image depends on the effective illumination range during actual imaging, as shown in Figure 5.

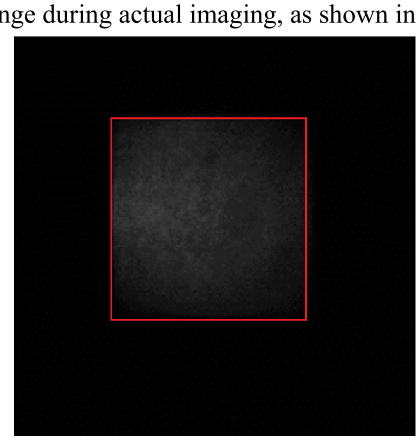

Figure 5: Example of CCD image data (red rectangle indicates effective illumination range).

Data Cleaning

Data cleaning is a necessary step for most data-driven tasks. Data cleaning refers to detecting and correcting possible errors from data to ensure the quality of data and

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THPV036

comply with domain related integrity constraints. The data cleaning mainly refers to the integrity and anomaly detection of the original measurement data. Integrity detection is to judge whether the laser beam participating in the experiment has energy, waveform and near-field measurement results at all positions based on the experimental configuration parameters. Anomaly detection is to judge whether there is a serious mismatch between the measurement results and the actual output, which will significantly affect the state evaluation results.

Data Exchange

A data governance service named "report and performance calculation process" is established to exchange data through the database read-write interface.

The service reads the experimental measurement data from the original database of the data source layer through the database interface, realizes the functions of data extraction, data cleaning, analysis and calculation in combination with the algorithm, and stores the analytical results in the data warehouse of the data management layer, for the visualization function module to directly read the data and display the charts.

The "report and performance calculation process" service is integrated into the main operation workflow. It is called by the centralized control system. After each shooting experiment, the experimental data governance of the shot is started to calculate the state parameters of the laser driver.

Data Integration

Data integration is the physical or logical integration of heterogeneous, distributed and autonomous data to provide users with a unified access interface, so that they can access data sources transparently, so as to realize comprehensive data sharing [3,4]. There are two common schemes for data integration, one is pattern integration method, and the other is data replication method, such as data warehouse. This paper utilizes the data replication method, based on the logical model of experimental number, laser beam number, experimental configuration parameters, experimental measurement results and state evaluation results, constructs the data warehouse of driver state evaluation. The visualization function module directly reads data from the data warehouse and displays a variety of charts.

CALCULATION OF LASER DRIVER STATE PARAMETERS

The contents of laser driver state estimation are described in Section 2.1, mainly including physical experiment results and laser driver performances. In order to calculate these data, we need to focus on two important issues: the data flow and the calculation algorithm.

Data Flow

The relationship between experimental configuration parameters, original measurement data and state parameters data is shown in Figure 6.

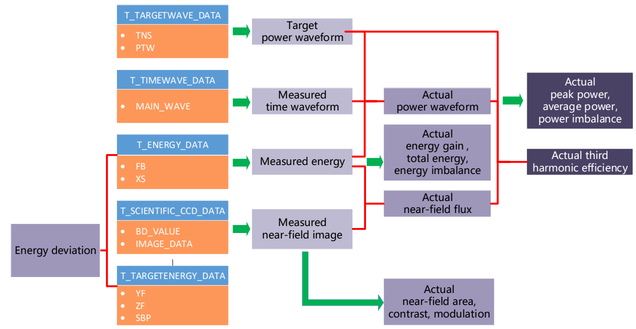

Figure 6: Data flow of calculating state parameters.

Agorithm Integration

The data governance service for the laser driver state estimation follows the software architecture and integration specification of the centralized control system, and is coded in C++. Data governance service involves various data processing algorithms in data extraction, data cleaning, analysis and calculation. By weighing the difficulty of algorithm implementation and the time efficiency of calculation, the matlab is mostly used. These algorithms are encapsulated into dynamic link library files for data governance services to call. According to the data relationship as shown in Figure 6, the data input and output interfaces of each algorithm are almost fixed. In case of algorithm upgrade, we only need to replace the corresponding algorithm dynamic link library file and rebuild the data governance service.

CONCLUSION

This paper summarizes the data governance work for the application of laser driver state estimation from four aspects: data specification, data cleaning, data exchange and data integration. The final result can serve all relevant data users in the whole laser facility.

Data governance involves various data processing algorithms. As a supplement to traditional algorithms, artificial intelligence algorithms increasingly show a big potential. Our future research will be carried out in the development and integration of intelligent algorithms to gradually optimize the effect of data governance.

REFERENCES

[1] Xiaowei Zhou, Daojian Yao, *et al.*, “Centralized control software system of XX unit - database design report”.

[2] Daojian Yao, “Design of computer centralized control system for large laser device”, National Defense Technology (GF) report database of Chinese Academy of physics, 2013.

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THPV036

[3] Xiaoyong Du, Yueguo Chen, Ju Fan, Wei Lu "Data collation - Key Technology of big data governance [J]", Big Data, 2019, 3:13-22.

[4] Xindong Wu, Bingbing Dong, Xinzheng Du, Wei Yang, "Data governance technology [J]", Journal of Software, 2019, 30 (9): 2830-2856.

[5] Jian Yin, Huaijie Zhu, Jianxing Yu, Shuang Qiu, "Panoramic framework of big data governance [J]", Big Data, 2020, 2:19-26.

[6] Xiaoou Ding, Hongzhi Wang, Shengjian Yu, "Quality management of industrial time series big data [J]", Big Bata, 2019, 6:19-29.

[7] Yike Guo, Wei Pan, Simiao Yu, Chao Wu, Shicai Wang, "Big Data for Science [J]", Journal of the Chinese Academy of Sciences, 2016, 6 (2): 599-607.

[8] Yaping Yang, Yi Wang, Yan Bai, Xiafang Le, Jia Du, Yongqing Bai, Jiulin Sun, "Development and practice of national earth system science data center [J]", Journal of Agricultural Big Data, 2019, 1 (4): 5-13.

[9] Senlin Xiong, Ziming Zou, Xiaoyan Hu, Zhen Ji, "Spatial science data product organization model [J]", Journal of Agricultural Big Data, 2019, 1 (4): 30-36.

[10] Guomin Zhou, Jingchao Fan, "Design and implementation of agricultural science observation data aggregation management platform [J]", Journal of fan, 2019, 1 (3): 38-45.

[11] Guoliang Li, JiannanWang, Yudian Zheng, Franklin M. J., "Crowdsourced data management: a survey[J]", IEEE Transactions on Knowledge and Data Engineering, 2016, 28(9): 2296-2319.

[12] Daniel E. O'Leary, "Embedding AI and Crowdsourcing in the Big Data Lake[J]", IEEE Intelligent Systems, 2014, 29(5): 70-73.