Postprint of Time-Frequency System Fault Prediction

Authors: Wang Lingling, Gou Wei

Date: 2021-09-29T00:00:00+00:00

Abstract

The hydrogen maser clock and time comparison system (hereinafter referred to as the time-frequency system) of VLBI observation stations provides a stable and reliable time and frequency reference for the stations, and the quality of the system signals directly affects observation data quality. The hydrogen maser clock is the critical equipment of the time-frequency system; once a failure occurs, its impact on data is devastating. Therefore, the VLBI station time-frequency system was designed and constructed with real-time monitoring and recording of the hydrogen maser clock and its surrounding parameter states; upon detection of any anomaly, technical personnel must respond promptly. Since its establishment, the Chinese VLBI Network (CVN) currently comprises five stations, each equipped with two or more hydrogen maser clocks in its time-frequency system. The system has accumulated substantial data through long-term operation. By analyzing these data, we can investigate the relationships among hydrogen maser clock status changes, performance variations, and environmental changes, thereby establishing failure prediction models for equipment fault forecasting. This paper describes the process of creating a time-frequency system data warehouse and utilizing data mining techniques to predict system failures. Test results demonstrate that clustering models can be effectively employed for prediction in this project.

Full Text

Timing System Failure Prediction

WANG Lingling[1,2]**, GOU Wei**[1]

[1] Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China

[2] Shanghai Key Laboratory of Space Navigation and Positioning Technology, Shanghai 200030, China

Abstract

The hydrogen maser and time comparison system (timing system) of VLBI stations provides a stable and reliable time and frequency reference, and the quality of its signals directly affects observation data quality. As the most critical equipment in the timing system, the hydrogen maser' s failure can be catastrophic to data integrity. Therefore, VLBI station timing systems have implemented real-time monitoring and recording of hydrogen maser status and surrounding parameters since their initial design and construction, enabling rapid technician response upon detecting anomalies. The China VLBI Network (CVN), currently comprising five stations each equipped with two or more hydrogen masers, has accumulated substantial operational data over more than a decade of continuous operation. Analyzing this data allows us to investigate relationships between hydrogen maser state changes, performance variations, and environmental fluctuations, thereby establishing failure prediction models for equipment fault forecasting. This paper describes the testing process of creating a timing system data warehouse and utilizing data mining techniques for system failure prediction. Test results demonstrate that clustering models can effectively predict failures in this application.

Keywords: timing system; hydrogen maser; failure prediction; data-driven

1. Introduction

The timing system provides VLBI stations with a stable and reliable time standard, and the quality of its output signals directly impacts observation quality. The system comprises hydrogen masers and other time comparison equipment, with the hydrogen maser serving as the most critical component—a high-precision frequency standard whose microwave resonant cavity output frequency is highly susceptible to external environmental interference. Consequently, its performance is easily affected by environmental conditions. Current research on timing system state and performance changes primarily focuses on environmental impacts on hydrogen maser performance, with numerous contributions from domestic and international scholars [1-4]. These studies show that hydrogen maser frequency stability significantly degrades under poor environmental control, making effective environmental management a prerequisite for ensuring reliable operation of hydrogen masers and the entire timing system.

From a user perspective, beyond improving equipment reliability itself, advance failure prediction represents an effective method for reducing failures and enabling early fault handling. Hydrogen maser operation is a dynamic process involving complex physical and chemical state changes, some of which ultimately lead to performance degradation or failures. Observing and analyzing these relationships facilitates failure prediction, which requires equipment status monitoring, followed by analysis of relevant data to derive prediction models for accurate equipment state forecasting.

2. Data Collection

Failure prediction technology is a multidisciplinary comprehensive information processing technique that extends beyond traditional fault diagnosis [5]. It predicts potential failure trends using historical data and relevant technologies while equipment remains operational, incorporating current state information [6]. Prediction methods can be broadly classified into three categories: model-based, data-driven, and statistical reliability approaches [5]. The China VLBI Network (CVN), established over a decade ago, has accumulated extensive operational data, providing a solid foundation for such research. This paper employs data-driven failure prediction technology, which has gained significant attention and rapid development in recent years, becoming an important research focus [7].

The primary equipment in VLBI station timing systems includes hydrogen masers, frequency dividers/station clocks, time interval counters, switches, frequency multipliers, frequency/pulse distributors, GPS/BD time servers, and supporting infrastructure such as UPS systems, constant temperature and humidity air conditioners, indoor digital thermometers, and outdoor weather stations for uninterrupted power supply and environmental monitoring. According to the "Practical Hydrogen Maser Technical Manual," active hydrogen masers used at VLBI stations provide a status monitoring system similar to compact clocks [8], enabling data collection via serial ports for clock status analysis. Data from other devices can similarly be collected through built-in serial ports or network protocol interfaces using corresponding acquisition software, or through specialized serial-to-network converters when necessary. The acquisition frequency is determined by requirements.

Timing system monitoring data encompasses hydrogen maser internal parameters, clock differences, clock room temperature and humidity, external environmental temperature and humidity, atmospheric pressure, wind direction, wind speed, and more, sourced from different devices. This data provides firsthand material for prediction model research, ensuring feasibility. However, after years of construction, continuous maintenance, and upgrades, numerous devices have been operated across timing systems, including multiple hydrogen maser models such as MHM2010, VCH-1003M, SHOM-3, SHOM-4, and SHOM-A, with specific device identifiers also changing over time. Data storage content and formats vary significantly—some stored as databases, others as files—requiring consolidation for analysis.

3. Data Organization

Typical data-driven failure prediction methods include artificial neural networks, fuzzy systems, and other computational intelligence approaches [5]. This study employs a prediction algorithm based on data mining principles.

The timing system's primary equipment has been described previously. Through business analysis, the conceptual model design for the timing system data warehouse is illustrated in [Figure 1: see original paper]. Following data warehouse construction procedures, logical and physical modeling were subsequently performed, with the logical model shown in [Figure 2: see original paper] and partial data table designs presented in through .

Due to data diversity and complexity, organization requires several steps including data cleaning and format conversion. Data cleaning involves removing erroneous and invalid data from raw datasets, followed by format conversion for unified storage. Timing system data records exist in both database and text file formats. During cleaning, special attention must be paid to time format differences between Beijing Time and Universal Time—all data are converted to Universal Time for consistency. Equipment replacements, including model upgrades or same-model substitutions, have occurred at stations, with associated information sometimes incomplete, erroneous, or missing. Missing values are filled with nulls, while errors require correction. Data records may also be mixed, with the same table potentially containing data from different periods or devices of different models or identifiers, necessitating splitting and re-entry based on log files. Additionally, due to large source data volumes, uniform sampling reduction was applied for testing convenience.

4. Data Analysis

Among all data, hydrogen maser status data most intuitively reflects timing system state. This paper focuses on SHOM-4 hydrogen maser status data and corresponding environmental temperature data as an example. According to the technical manual, the maser has 32 conventional status parameters, with all but one spare parameter and eight isolation amplifier data points capable of real-time operation status reflection. [Figure 3: see original paper] shows 2019 data from maser #88, with the vertical axis representing parameter values.

These status parameters can be categorized by physical module into power supply, temperature control, ion pump, receiver, and isolation amplifier sections. The receiver' s intermediate frequency signal (IFL) value represents the clock signal—values below 1.5 indicate clock output signal loss, constituting a critical failure. This parameter is the primary analysis focus. After data refinement and based on the hydrogen maser technical manual combined with historical experience, 11 parameters were identified as closely related to IFL values, as shown in [Figure 4: see original paper]. Due to large value differences, data are displayed in two panels for clarity: the upper panel shows OSCI values, while the lower panel displays the other ten parameters. Further refinement reveals additional details—for instance, [Figure 5: see original paper] suggests strong correlation based on curve shapes, with the curves bending in approximately opposite directions. When temperature rises, OVN1 and OVN2 values decrease, and when temperature falls, they increase, indicating the highest correlation between OVN1, OVN2, and temperature data. This occurs because OVN1 and OVN2 represent voltage values at the outer furnace position in the maser' s temperature control section, directly affected by environmental temperature changes.

5.1. Comparison of Mining Algorithms

In today' s information society, the growing volume of data contains substantial hidden value, and data mining technology has emerged to extract useful information and knowledge from large amounts of random data. While numerous algorithms have been developed, each has limitations, making algorithm fusion a development trend for solving specific problems [9].

Using 2019 data as an example, this study analyzed and processed several typical algorithms and compared their performance. Based on traditional algorithm applications and project-specific scenarios, decision tree, clustering, association, and time series algorithms were selected for testing. Algorithm parameters are adjustable—for example, in the decision tree model, the dependency network between parameters shows that IFL is linked to OVN1, OVN2, FLUX, DIO, and temperature data. When adjusted to display only strong links, the relationship with DIO is disconnected ([Figure 6: see original paper]). In association algorithms, multiple parameters are ignored due to being continuous data. Time series algorithms require continuous data without gaps; since the selected test data contained missing values, gap-filling methods had to be specified using previous values, averages, or specific constants. The mining models created by each algorithm are shown in [Figure 7: see original paper].

Model accuracy can be evaluated using lift charts ([Figure 8: see original paper]), where the blue line represents the ideal model. The X-axis shows the percentage of test datasets used for comparison, while the Y-axis shows the percentage of accurate predictions. The ideal line is a diagonal indicating that in 50% of data, the model accurately predicts 50% of cases (the expected maximum). All models approach the ideal model, with the decision tree model performing best.

However, careful examination of prediction results reveals that while most values were predicted successfully, an actual anomaly was not predicted—the value 0.38 selected in [Figure 9: see original paper], which is an actual anomaly because it is less than 1.5. This value was not predicted by any of the three models, likely because the anomaly occurred only once in 15,130 data points, making the test data insufficiently representative. Consequently, actual prediction performance would be poor, and the time series algorithm tests exhibited the same issue. Therefore, subsequent tests used reselected data.

5.2. Model Optimization and Prediction

New test data were selected from maser #43 in 2013, with IFL data shown in [Figure 10: see original paper], demonstrating multiple signal loss events during the period. This data segment was extracted for model construction demonstration.

In this model construction test, the clustering model scored higher than the decision tree, reaching 0.88 ([Figure 11: see original paper]). Therefore, this model was used to predict the latter half of the test data, with predicted IFL values set to NULL. The prediction classification matrix ([Figure 12: see original paper]) shows uniform prediction results for various IFL data categories, consistent with actual data. The model predicted 15 data points with a value of 0.06 (failure value), very close to the actual 14 data points, with prediction probabilities ranging from 0.51 to 0.857 and high accuracy. This demonstrates successful failure prediction using the clustering model.

This project also tested time series algorithms, which predict future values in continuous data sequences and require a time key—our time-series collected data meets this requirement. Actual prediction results are shown in [Figure 13: see original paper], where the solid line on the left represents actual values and the dashed line on the right represents predicted values. Comparing with actual data from [Figure 10: see original paper] shows that data around March 19, 5:17:10 were indeed around 3, indicating highly accurate predictions. However, when selecting a different prediction starting point, results were unsatisfactory ([Figure 14: see original paper]), likely due to poor data periodicity, making the time series algorithm less than ideal. This requires further analysis and research.

6. Conclusion

This paper explores methods for timing system failure prediction, analyzing and testing historical data using data mining techniques. The testing process demonstrates that combining equipment status data with operational environment data can analyze and predict equipment state trends, thereby providing early warning of potential failures.

The primary challenge lies in algorithm selection and design, with model construction being an ongoing optimization process. Testing shows that clustering models can effectively predict failures and should be validated with additional data. The next step involves integrating this prediction model into timing system software to enable real-time failure prediction.

Additionally, prediction models require extensive original data for training—the more comprehensive the training data, the more accurate and robust the model. However, efficiently processing these data under limited computational power presents a challenge that warrants dedicated research.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.