Failure Prediction for Temporal Dependency of Hard Drives

Xiaojian Li [1 +], Liyu Zhu [1], Cuiping Zhang [1], Haopeng Yang [1], Hailan Wang [1] and Jiajia Zhang [1]

[1] School of Computer Science & Information Engineering, Guangxi Normal University, Guilin, China

Abstract. Cloud service providers strive to predict hard drives' failure in advance to enhance user confidence in cloud storage resources. We explored the failure property from the self-monitoring, analysis, and reporting technology features of hard drives, finding that the long-term temporal changepoint dependency (LTCD) of hard drive failure creates new reconstruction challenges in failure prediction. The failure prediction for temporal dependency (FPTD) presented in this paper has three characteristics: primary identifying features and changepoint features and enhancing changepoint dependency, all of which make FPTD more sensitive to the failure of hard drives with LTCD. The experimental results show that the five evaluation metrics of FPTD are all above 94% while maintaining a low false alarm rate, among which Accuracy can reach 99.0% and Recall can reach 97.6% on average. In general, the FPTD has higher prediction quality and better stability, and is more suitable for predicting hard drive failures in the long-short temporal.

Keywords: SMART, temporal dependency, changepoint dependency, failure prediction.

1. Introduction

Since cloud storage services can be used by millions of users worldwide 24 hours a day, the high availability of services is essential to cloud systems. Although the goal of existing popular cloud service providers, such as Aliyun, Amazon, Microsoft Azure, Backblaze, and others, is high service availability, the monthly user uptime percentage is generally not less than 99.9%. Even if cloud service providers promise to provide users with high-availability services, accidental violations of Service Level Agreements (SLA) occur every year and cause serious issues. The Ponemon Institute published the report "Cost of Data Center Outages" in 2016 [1]. The institute assessed the unexpected downtime of 63 data centers and found that the most expensive cost of unexpected downtime exceeded \$17,000 per minute. The average cost per minute is close to \$9,000.

Large-scale storage systems usually comprise heterogeneous storage devices with significantly different failure rates. According to [2], sudden server failure caused by hardware failure is a widespread problem. Hard drive failure is the most common cause of cloud storage data loss, accounting for 81.84% of all failures [3]. Hard drive failures are likely to cause service providers to violate service level agreements and cause severe economic loss to service providers and users [4]. These losses are usually caused by underreporting traditional failure detection systems. Cloud service providers provide many resources needed to replace failed components, but this is only useful if these failures are detected. However, unexpected downtime in cloud systems often caused by subtle gray failures [5]. Even if the application exhibits signs of potential failure, the system failure detector may not notice the problem. To enhance user confidence in opaque cloud storage resources, service providers must predict hard drive failure before hard drive failure causes more severe damage to the cloud system.

The previous hard drive failure prediction [6-14] mostly takes a single snapshot with self-monitoring, analysis, and reporting technology (SMART) characteristics as the prediction input example. However, hard drive failure is not achieved instantly but is a time-varying process of the gradual decline of hard drive health. Furthermore, before the hard drive fails, it begins to report failure-prone logs, such as the reallocated sector count, raw read error rate, seek error rate, and so on. Therefore, hard drive failure is a gray failure, and hard drive failure prediction is a time-series prediction problem.

Some methods extract short-term information from SMART feature sequences to predict hard drive failure. However, these methods ignored the long-term changepoint dependence of SMART features, making it unable to accurately reflect the long-term time-varying process of the decline of SMART features. This paper aims to provide a method of FPTD that can give early warning methods and information for decision-making before hard drive failure, reduce the failure probability of a hard drive, and even avoid failure to maintain a highly reliable storage system. In summary, our major contributions presented in this paper are the following:

- We detect the failure of the temporal dependency of hard drives.
- We point out the challenges of long-term temporal changepoint dependency (LTCD) for FPTD.
- Our proposed methodology can improve the prediction quality by exploiting LTCD.

The remainder of the paper is organized as follows. First, we formulate the problem of predicting hard drive failures in Section 2. Then, we propose our methodology for failure prediction in Section 3. Afterward, we present our experiment and results in Section 4. Next, we survey the related work in Section 5, followed by the conclusions in Section 6.

2. Problem Formulation

We formulate the problem of predicting hard drive failures as a binary classification problem. Specifically, we use $T = \{(X, Y)\}$ to represent our training dataset, in which $X = \{X_1, \cdots, X_i, \cdots, X_m\}$ is a set of multidimensional observations collected in the m prediction horizon (PH), where $X_i = \{x_{i1}, \cdots, x_{ij}, \cdots, x_{in}\}$ represents the ith input sequence consisting of n SMART failure feature measurements. Here, $Y \in \{0,1\}$ is a binary response variable for each hard drive, where 0 indicates a healthy state, and 1 indicates a failed state.

Our goal is to employ the best method to learn the function $f : X \to \{0,1\}$, which minimizes the loss function $l(h(X); Y)$. To solve this problem, we propose a prediction model consisting of two dense fully connected layers, followed by a convolutional neural network (CNN) layer and a long short-term memory (LSTM) layer. Note that we aim to predict whether the hard drive will fail, not the exactly when the hard drive will fail in the next future.

3. Methodology for Failure Prediction

3.1. Overview

This section presents FPTD, the proposed methodology that can improve prediction quality by exploiting the failure of temporal dependency of hard drives. Fig. 1 illustrates the overall workflow of the proposed methodology. In the figure, the ellipse represents the sub-activities of FPTD, and the arrow indicates the data transmission. The associated activity starts execution after all inputs are satisfied at the same time.
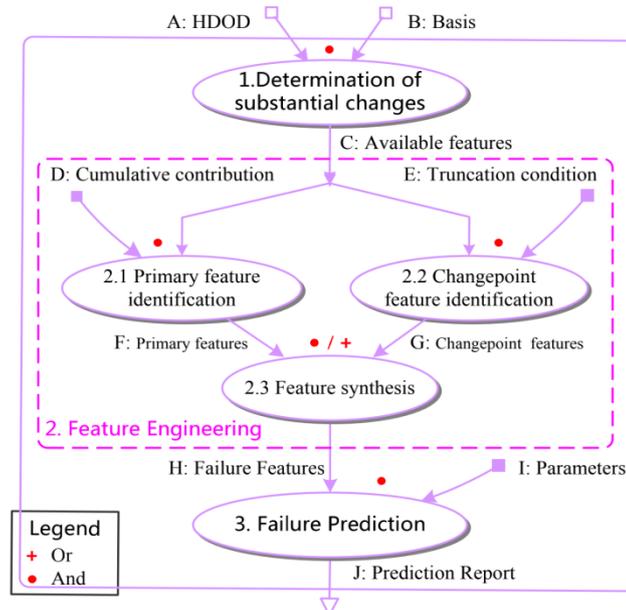

Fig. 1: The overview of the proposed methodology.

The FPTD consists of three phases: 1) determining a substantial change from the historical data on the healthy and failed drives, 2) feature engineering by identifying primary features (PF) and changepoint features (CPF) and synthesizing the feature weight, and 3) predicting failure using the CNN and LSTM.

3.2. Determination of Substantial Changes

This paper selected ST4000DM000 as the critical research object. We eliminated hard drives with multiple failures, too few sample records, and severely missing data records. As of the first quarter of 2019, Backblaze had 124 columns of SMART feature data (i.e., 62 smart IDs). Next, we interpreted the meaning of the 124 SMART features involved in this dataset. After a series of preprocessing, such as eliminating features with no value and no substantial change in value, identifying and filling missing values in the feature sequence, and rearranging features, we have identified 40 features. These features include 1 date feature, 1 label feature, and 38 SMART features.

3.3. Feature Engineering

Feature engineering consisting of PF identification, CPF identification, and feature synthesis.

3.3.1. Primary Feature Identification

After preprocessing the SMART features of the hard drive, 38 available SMART features remain, and the data dimension is still high. Suppose the SMART feature data are abused, and the information in the hard drive decay process is ignored. In that case, the evaluation and prediction of the prediction model are affected, and the prediction quality is negatively affected. Therefore, the primary failure characteristics of hard drive degradation must be determined and characterized as computable methods. The main failure feature determination algorithm is composed of feature transformation and strong correlation determination.

(1) Feature Transformation

To reflect the comprehensiveness of the system analysis and the evaluation from different aspects, we choose the principal component analysis (PCA) algorithm [15] for feature transformation. The PCA is a standard dimensionality reduction method based on unsupervised learning, which is often used in multivariate statistical analysis. We implement the PCA based on the singular value decomposition covariance matrix. Under the condition that the cumulative contribution rate is 80%, the method transforms 38 linearly correlated available SMART features into 14 linearly independent principal failure components. It sorts the contributions of each element according to the ratio of the variance of each principal component to the total variance.

(2) Strong Correlation Determination

According to the Pearson correlation coefficient method, we trace the available SMART features, which are strongly related to the 14 principal components (PC), to concentrate the feature information (Fig. 2). We take the first three available SMART features with larger absolute values of correlation coefficients of the PC of 1 to 3, the first two available SMART features with larger fundamental values of correlation coefficients of the PC of 4 to 7, and the available SMART features with the most significant absolute value of correlation coefficients of the PC of 8 to 14. Because the raw and normalized features of the same SMART ID have a strong correlation, we only retain the feature with the highest correlation coefficient when they appear simultaneously. The PF set contains 19 available SMART features, as shown in Table 1.
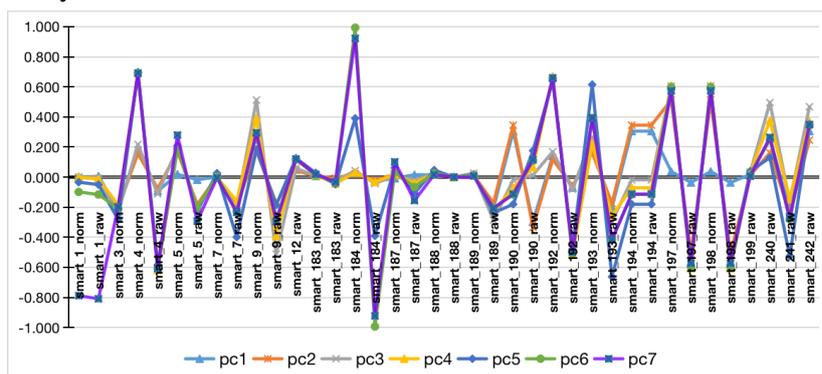

Fig. 2: Correlation between SMART features and PC of 1-7.

3.3.2. Changepoint Feature Identification

The failure of a hard drive is a long-term and time-varying process in which the health state gradually declines. Moreover, CPF identification detects the changepoint of the SMART feature sequence of the failed hard drive to determine the failure features with changepoint characteristics and to capture the LTCD from the SMART feature sequence of the failed hard drive.

The changepoint problem includes two sub-problems: the test of the existence of the changepoint and the estimation of the changepoint position [16]. In this paper, the Bayesian changepoint detection algorithm based on recursion proposed by Fearnhead et al. [17] is used to solve multiple points in SMART time-series data. According to the different distributions of SMART feature data of a hard drive, the algorithm can segment the time-series data and form multiple recognizable subsequences. The boundary between partitions is the changepoints. A changepoint is defined as a location where the statistical characteristics of the observations of the sequence abruptly change unexpectedly. In this paper, the changepoint place represents the time point of the changepoint event, which is the focus of the CNN model enhancement. The changepoint can also help staff trace a specific date during the hard drive operation and determine the cause of the failure of the hard drive. In addition, Changepoint feature identification has two essential functions.

(1) Determination of failure property with changepoints.

In [18], several specific performance aspects of SMART features before hard drive failure include shock type, steep increase type, steep decrease type, and quiet type, but not all features represent the failure characteristics of the hard drive. Moreover, [19] also used Bayesian changepoint detection to select features. Changepoint detection can detect unexpected changes in the time-series distribution function. In most failed hard drive samples, if the value of a SMART feature constantly changes suddenly before failure, which is a common property of mutation, this feature is usually essential and must be selected.

(2) Capture the LTCD from SMART feature sequence.

The Bayesian changepoint detection algorithm can detect the change point of the decline distribution of hard drive failure data and determine when the probability of significant change is the largest. For example, the changepoint detection results of the smart_197_raw of the failed hard drive are shown in Fig. 3. The x-axis represents the number of days before the final failure, where $x = 0$ is 100 days before the final failure and $x = 100$ is 1 day before the absolute failure. The y-axis of the upper half represents smart_197_raw. The y-axis of the lower half represents the probability of the changepoint appearing on a particular day. The figure reveals that the most likely time for the hard drive to reach a changepoint is about 30 days before failure. From the experimental results, the decay characteristics of the hard drive can be traced to 30 days or even 70 days ago, which proves that the SMART feature sequence has LTCD.

Through the changepoint detection of SMART features of the failed hard drive, 13 CPF are selected in our paper, as shown in Table 1.
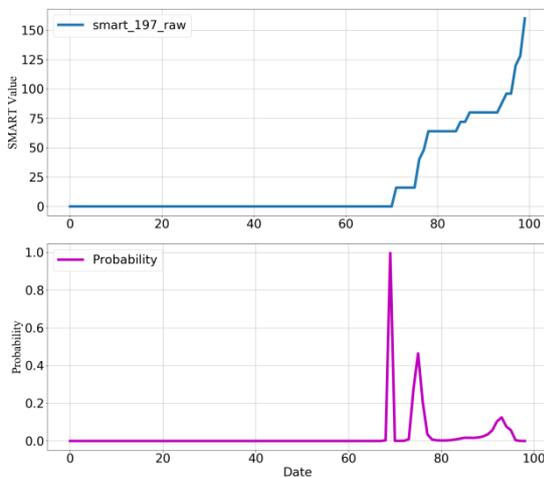

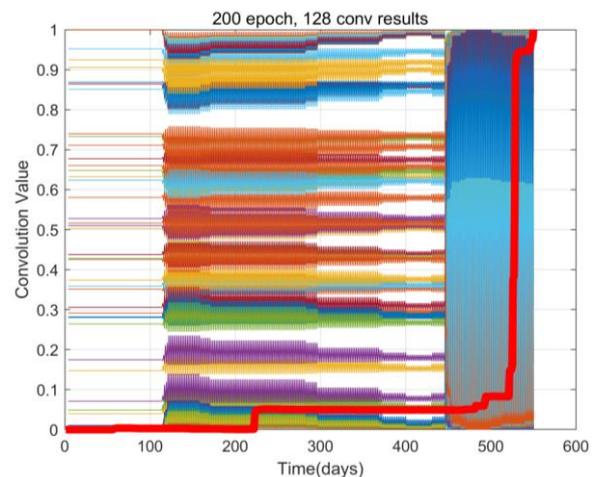Fig. 3: Changepoint detection of smart_197_raw.

Fig. 4: Convolution visualization of smart_192_raw.

Table 1: Failure Feature

NO	SMART Name	Description	PF	CPF	NO	SMART Name	Description	PF	CPF
1	smart_1_raw	Raw_Read_Error_Rate	✓	✗	11	smart_190_normalized	Airflow_Temperature_Cel	✓	✗
2	smart_4_raw	Start_Stop_Count	✓	✓	12	smart_192_raw	Power-Off_Retract_Count	✓	✓
3	smart_5_raw	Reallocated_Sector	✓	✓	13	smart_193_raw	Load_Cycle_Count	✓	✓
4	smart_7_normalized	Seek_Error_Rate	✓	✓	14	smart_194_raw	Temperature_Celsius	✓	✗
5	smart_12_raw	Power_Cycle_Count	✓	✓	15	smart_194_normalized		✓	✗
6	smart_183_raw	Runtime_Bad_Block	✓	✓	16	smart_197_raw	Current_Pending_Sector	✓	✓
7	smart_184_raw	End-to-End_Error_ Detection_ Count	✓	✓	17	smart_198_raw	Offline_Uncorrectable_ Sector_Count	✓	✓
8	smart_187_raw	Reported_Uncorrect _Errors	✓	✓	18	smart_241_raw	Total_LBAs_Written	✓	✗
9	smart_188_raw	Command_Timeout	✓	✓	19	smart_242_raw	Total_LBAs_Read	✓	✗
10	smart_189_raw	High_Fly_Writes	✓	✓					

3.3.3. Feature Synthesis

Thus far, we have obtained the PF and CPF. Feature synthesis can accumulate the weights of the SMART features to capture various failure properties of the SMART features, which include both PF and CPF, so that the prediction model can focus on the failure features with a strong correlation and changepoint characteristics. Therefore, the hard drive failure features include PF, CPF, and the combined PF and CPF group (P-CPF). We consider different input datasets to evaluate the effectiveness of feature engineering.

3.4. Failure prediction

According to the principle of the CNN [20] and LSTM [21] models, the prediction model is designed as illustrated in Fig. 5. The failure prediction model consists of an input layer, enhancement layer, temporal layer, and output layer. In the enhancement layer, the Conv1D layer interprets snapshots. The max-pooling layer consolidates and abstracts the interpretation to obtain the features. A 1D convolution is good at processing sequence data and is used to process natural language. In this article, the convolution enhances the changepoint events in the SMART feature sequence X_i to help LSTM learn the failure property of more extended time series. For example, the Convolution visualization of the smart_192_raw of the failed hard drive with serial number Z305D6PY is shown in Fig. 4. Fig. 4 is the cumulative map obtained after the smart_192_raw sequence is convolved with 128 convolution kernels. In Fig. 4, the x-axis represents the number of days before the final failure, and the y-axis represents the convolution value. The red curve represents the sequence value of smart_192_raw. Most importantly, the figure reveals that CNN has enhanced the changepoint events that occurred 20-70 days before the failure of smart_192_raw. The temporal layer predicts failure for the time series of changepoints. It captures the temporal dependence in the SMART feature sequence. The last layer outputs the prediction results and forms the prediction report through the full connection.
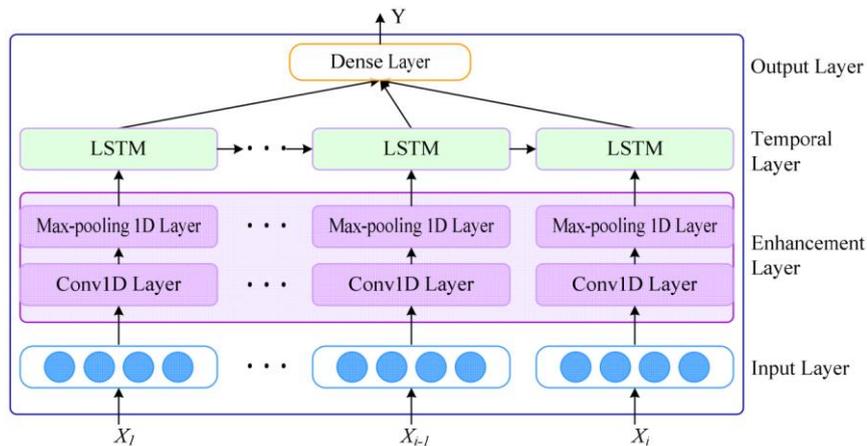

Fig. 5: Structure of CNN+LSTM.

4. Experiments and Results

4.1. Experimental Setups

4.1.1. Dataset

We used the public dataset [22] provided by the Backblaze data center to verify our method. The dataset has been taking a snapshot of the running hard drives every day since 2013, and the data sampling frequency is once a day. The snapshot consists of the basic drive information and S.M.A.R.T. statistics. The basic drive information includes five parts: date, serial number, model, capacity bytes, and failure. If the failure field value is 0, the drive is healthy; if the failure field value is 1, the value of the date is the last day that the drive ran before it failed. According to [23], each SMART attribute consists of a 6-byte raw value and a 1-byte normalized value, and the normalized value is converted from the original value. In this paper, the raw features and normalized features of each SMART attribute are called SMART features.

We calculated the failure rate of all manufacturers' hard drives in the Backblaze dataset from July 2017 to June 2019 and found that the Seagate hard drive failure rate accounted for 87.22% of all failures. During this period, the failure rate of ST4000DM000 was the highest among all Seagate hard drives, which is about 41.04%. The failure rate of each manufacturer/model of a hard drive is listed in Fig. 6. Because some manufacturers/models of hard drives have a tiny population and few available SMART attributes, this paper selected ST4000DM000, which contains nearly 35,000 hard drives, as the critical research object.
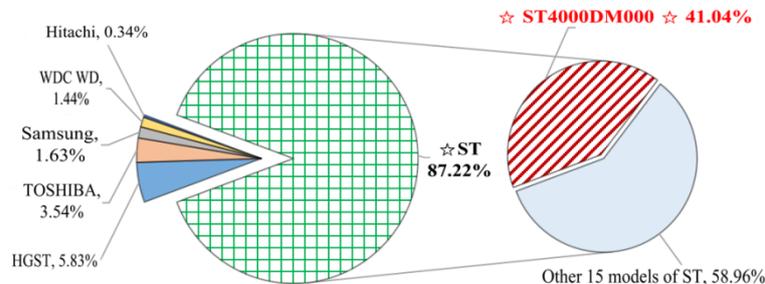

Fig. 6: Failure rate of hard drives by Serial Number / Model.

4.1.2. Hyper-parameters

In the hyper-parameter tuning phase, we conduct a grid search to build and evaluate models for each combination of hyper-parameters. The goal is to find the best variety with the highest performance. We empirically set the same learning rate of 0.001 for the LSTM and CNN+LSTM models, and we put the drop-out rate to 0.25. Then, we run experiments with different prediction horizon. For LSTM-based models, we build an LSTM model with four layers and 100 nodes, but the model's prediction quality for PH still fluctuates wildly. For CNN+LSTM, we set the number of convolution kernels to 128. To avoid overfitting the models, we observe the changing trend of the training and validation loss function as the epoch and batch-size increases. Finally, we choose 32 epochs and 64 batch-sizes for LSTM. For CNN+LSTM, we choose 200 epochs and 128 batch-sizes.

4.1.3. Evaluation Metric

The ST400DM000 dataset is a highly imbalanced distribution of healthy and failed drives. Therefore, to measure our prediction model in the dataset, we used five different evaluation metrics: accuracy, precision, recall, F-measure, and the Matthews correlation coefficient (MCC). In the problem of hard drive failure prediction, the accuracy rate represents the proportion of correctly predicted as failed hard drives to healthy hard drives among all hard drives. The precision rate indicates the balance of true predicated failed hard drives to all predicted failed hard drives. The recall rate is the proportion of true predicated failed hard drives to all true failure hard drives. The F-measure [24][25] represents the harmonic average of the precision rate and recall rate, and its value range is [0,1]. A more considerable value indicates a better model. The MCC[26] is the correlation coefficient between the real and predicted labels. Its value range is [1,-1], with 1 indicating that the prediction label is entirely consistent with the real label, -1 indicating that the prediction label is entirely inconsistent with the real label. These metrics calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall\left(TPR\right) = \frac{TP}{TP + FN} \tag{3}$$

$$F\text{-}measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{\left(TP + FN\right)\left(TP + FP\right)\left(TN + FP\right)\left(TN + FN\right)}} \tag{5}$$

4.2. Experimental Results

We use 5-fold cross-validation learning models [27][28], judge how models perform to testing dataset and avoid the overfitting issue. Using the three feature groups and five evaluation metrics we proposed, we observed several exciting phenomena.

Fig. 7 shows the influence trend of the PH on five metrics (Accuracy, Precision, Recall, F-measure, MCC), where the pink line and the blue line represent measuring the quality of the CNN+LSTM and LSTM models, respectively. It can be seen from Fig. 7 that when PH ={40, 50, 60, 70, 80, 90, 100}, the value range of the five metrics of the LSTM model is [60%, 95%], and the model's prediction quality for PH still fluctuates wildly. The five metrics values of the CNN+LSTM model range in [87%, 99%], which are stable above 95%. Compared with other work [29-36], their PH is lower than 40 days, and our proposed FPTD can still achieve higher prediction quality when the PH is 100 days. FPTD is more sensitive to hard drive failures with LTCD.

Fig. 8 indicates the influence trend of the PH on the false alarm rates (FARs) and failure detection rates (FDRs) of CNN+LSTM, where the purple line and the blue line represent the FDR and FAR indicators respectively. It can be seen from the above figure that when the PH is 70 days, the CNN+LSTM model has the best effect in distinguishing between healthy hard drives and failed hard drives, the prediction quality is the best when the FDR is as high as 98%, the FAR is about 0.8%. When PH=50 days, although the FDR of the CNN+LSTM model can reach approximately 97.5%, the FAR value at this time surpasses 1.7%, which far exceeds the FAR requirement of the actual data center.
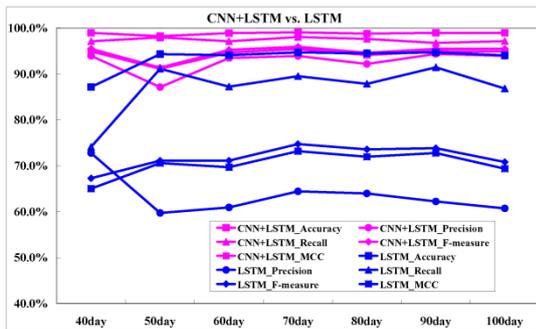

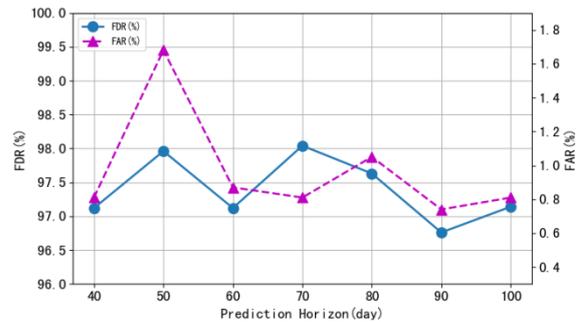Fig. 7: Model prediction quality with different models.

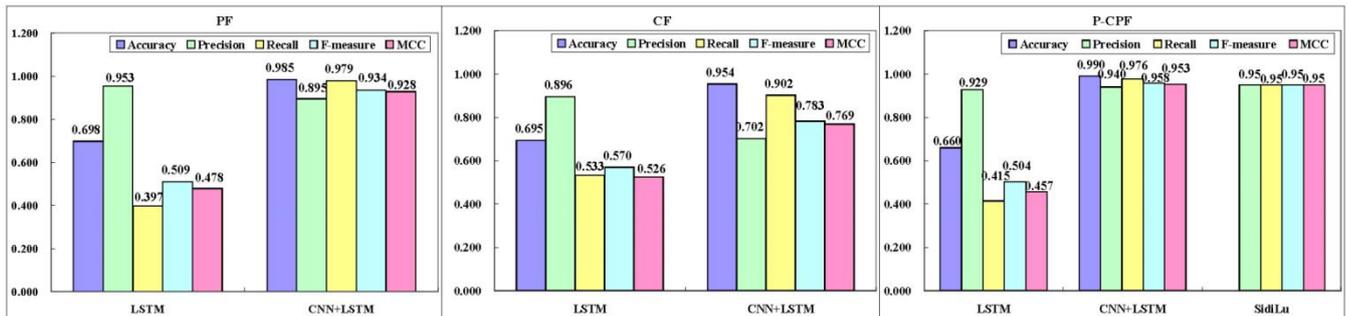Fig. 8: Model prediction quality of CNN+LSTM with different PH.

Fig. 9: Model prediction quality with different groups of PF, CF, P-CPF.

Fig. 9 saw that when the three feature groups (PF, CPF, P-CPF) are input, both LSTM and CNN+LSTM models are more sensitive to the P-CPF group and have higher prediction quality, which proves the effectiveness of our feature engineering. Meanwhile, it shows no matter which feature group is input, the prediction quality of CNN+LSTM model performs better than that of LSTM model. When inputting P-CPF, the prediction quality of the CNN+LSTM model is the best. The five evaluation metrics of CNN+LSTM model are above 94% while maintaining a low false alarm rate, among which Accuracy can reach 99.0% and Recall can reach 97.6% on average. Sidi Lu [36] also used the CNN+LSTM model to predict hard drive failure. Our experimental results are better than those of his four evaluation metrics except "Precision" metrics. Our work shows that the FPTD has a higher prediction quality and better stability, and it is more suitable for predicting hard drive failures in the long-short temporal.

5. Related Work

Previous hard drive failure prediction methods [6-14] primarily use a single snapshot with SMART characteristics as the prediction input example to evaluate the health status of hard drive and predict the failure of a hard drive, without considering the dependence of different health conditions of the hard drive in the time range. However, hard drives usually do not fail suddenly but do so gradually. That is, SMART features are part of a time series. Capturing the information in SMART feature sequences, each snapshot in the sequence should not be regarded as an individual item but should be combined. In addition, due to the influence of measurement noise, a single observation is insufficient to provide reliable results on the health state of a hard drive. Therefore, considerable work must extract information from a SMART time series to predict hard drive failure.

Wang et al. [31] proposed a two-step parametric time probability model that uses the current observations and its neighborhood to track the degradation information for hard drives over time. Zhao et al. [32] regarded the SMART feature series as a time series and proposed a failure prediction model based on the hidden Markov model. Goldszmid designed a failure detector comprising the hidden Markov model and a logistic regression model [33]. The hidden Markov model is used to capture the time correlation of the signals. However, the above three works are all based on the correlation of a short time series in a SMART feature sequence and do not capture long-term hard drive decay time information.

Xu Chang et al. [29] proposed a hard drive failure prediction model based on RNNs, and divided the SMART feature sequence of the hard drive within 28 days into six subsequences with unequal time series, namely, six health degrees. Lima et al. [34] proposed an approach to predict drive failures based on the LSTM method. In particular, LSTM, which includes a storage unit that can keep information for a relatively long time, can solve long-term back-propagation. However, the literature [30] has pointed out that neither the Markov model nor the RNN can extract long-term temporal dependency from SMART feature sequences due to the inherent limitations of these models.

Hard drive SMART feature sequence data comprise different distribution functions. For example, a solid or weak relationship exists between health status, making it difficult for an RNN-based model to capture the strong connection in the hard drive SMART feature sequence. Moreover, for the data in the same distribution, it is not necessary to equally focus on samples because paying attention to too many ordinary samples may dilute the weight of the changepoint events quickly. Therefore, we should consider enhancing the weight of changepoint events. Considering the LTCD of SMART sequences, we integrated the CNN and LSTM into a unified CNN+LSTM model [35] because the CNN enhances the changepoint events in the SMART feature sequence, which can help the LSTM model learn the failure property from long time series.

6. Conclusion

This paper explored the failure properties from the SMART features of hard drives, identifying the PF and CPF from the running data. We found that the LTCD on hard drives failure creates a new challenge for reconstruction in failure prediction.

To make the prediction model focus on the failure features with a strong correlation and changepoint characteristics, we synthesized PF and CPF as the new input of the prediction model. The experimental

results proved the effectiveness of our feature engineering. Meanwhile, we combined the CNN with the LSTM because the CNN enhances the changepoint events in the SMART feature sequence to help LSTM learn the failure property of more extended time series. Our work shows that CNN+LSTM is more suitable for predicting hard drive failures in the long-short temporal.

The experiments on real-world datasets reveal that the five evaluation metrics we used can evaluate the highly unbalanced data more comprehensively. The five evaluation metrics of FPTD proposed in our paper are above 94% while maintaining a low false alarm rate. Accuracy can reach 99.0% and Recall can reach 97.6% on average. Compared with other works, the FPTD system has a higher prediction quality and better stability, and it is more sensitive to hard drive failures with LTCD.

7. Acknowledgments

8. References

[1] Ponemon Institute. Cost of Data Center Outages[EB/OL]. https://www. vertivco.com/globalassets/documents/ reports/2016-cost-of-data-center-outages-11-11_51190_1.pdf.

[2] Oppenheimer D L, Ganapathi A, Patterson D A. Why Do Internet Services Fail, and What Can Be Done About It?[C]. *In Proceeding of the 4th Conferencse on Usenix Symposium on Internet Technologies & Systems(Usenix ' 03)*, 2003, Vol.67.

[3] Wang G, Zhang L, Xu W. What Can We Learn from Four Years of Data Center Hardware Failures?[C]. I*n Proceedings of the 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 2017: 25-36.

[4] David A Patterson. A Simple Way to Estimate the Cost of Downtime[C]. *Large Installation System Administration*. 2002: 185-188.

[5] Huang P, Guo C, Zhou L, et al. Gray Failure: The Achilles' Heel of Cloud-Scale Systems[C]. *In Proceedings of the 16th Workshop on Hot Topics in Operating Systems*. ACM, 2017: 150-155.

[6] Hamerly G, Elkan C. Bayesian approaches to failure prediction for disk drives[C]. *In ICML*. 2001,Vol.1: 202-209.

[7] Hughes G F, Murray J F, Kreutz-Delgado K, et al. Improved disk-drive failure warnings[J]. IEEE *Transactions on Reliability*, 2002, 51(3): 350-357.

[8] Murray J F, Hughes G F, Kreutz-Delgado K. Machine Learning Methods for Predicting Failures in Hard Drives: A Multiple-Instance Application[J]. *Journal of Machine Learning Research*, 2005, 6(1): 783-816.

[9] Yu Wang, Qiang Miao, and M. Pecht. Health monitoring of hard disk drive based on Mahalanobis distance[C]. *In Proceedings of the Prognostics and System Health Managment Conference*. 2011: 1-8.

[10] Farzaneh Mahdisoltani, Ioan Stefanovici, and Bianca Schroeder. Improving Storage System Reliability with Proactive Error Prediction[C]. *In Proceedings of the 2017 USENIX Conference on Usenix Annual Technical Conference*. USENIX Association, 2017: 391-402.

[11] Zhu B, Wang G, Liu X, et al. Proactive drive failure prediction for large scale storage systems[C]. *In Proceeding of the 29th Symposium on Mass Storage Systems and Technologies (MSST)*. IEEE, 2013: 1-5.

[12] S. Pang, Y. Jia, R. Stones, G. Wang, et al. A combined Bayesian network method for predicting drive failure times from SMART attributes[C]. *In Proceedings of the International Joint Conference on Neural Networks*. IEEE, 2016: 4850-4856.

[13] Li J, Ji X, Jia Y, et al. Hard Drive Failure Prediction Using Classification and Regression Trees[C]. *In Proceedings of the 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*.

IEEE, 2014: 383-394.

[14] Li J, Stones R J, Wang G, et al. Hard drive failure prediction using Decision Trees[J]. *Reliability Engineering & System Safety*. 2017, 164: 55-65.

[15] Zhao Q. A Review of Principal Component Analysis[J]. Software Engineering, 2016, 19(6): 1-3.

[16] Li Yaguang. Several Problems on Multiple Change Points Analysis of Complex Data[D]. *University of Science and Technology of China*, 2018.

[17] Fearnhead P. Exact and efficient Bayesian inference for multiple changepoint problems[J]. *Statistics and computing*, 2006, 16(2): 203-213.

[18] Yang Hongzhang. Proactive Fault Tolerance Based on "Collection—Prediction—Migration—Feedback" Mechanism[J]. *Journal of Computer Research and Development*, 2020, 57(2): 306-317.

[19] Sun X, Chakrabarty K, Huang R, et al. System-level hardware failure prediction using deep learning[C]. *In Proceeding of the 56th ACM/IEEE Design Automation Conference (DAC)*. IEEE, 2019: 1-6.

[20] Rippel O, Snoek J, Adams R P. Spectral representations for convolutional neural networks[C]. *Advances in neural information processing systems*. 2015: 2449-2457.

[21] Sepp Hochreiter, Jürgen Schmidhuber. Long short-tesrm memory[J]. *Neural Comput*, 1997, 9(8): 1735-1780.

[22] Backblaze. Hard Drive Data and Stats[EB/OL]. https://www.backblaze.com/b2/hard-drive-test-data.html# overview-of-the-hard-drive-data.

[23] Allen B. Monitoring hard disks with smart[J]. *Linux Journal*, 2004, 2004(117): 9.

[24] Hripcsak G, Rothschild A S. Agreement, the f-measure, and reliability in information retrieval[J]. *Journal of the American medical informatics association*, 2005, 12(3): 296-298.

[25] Rish I. An empirical study of the naive Bayes classifier[C]. *IJCAI,* 2001, 3(22): 41-46.

[26] Boughorbel S, Jarray F, El-Anbari M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric[J]. *PloS one*, 2017, 12(6): e0177678.

[27] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection[C]. *IJCAI*, 1995, 14(2): 1137-1145.

[28] Rodriguez J D, Perez A, Lozano J A. Sensitivity analysis of k-fold cross validation in prediction error estimation [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2009, 32(3): 569-575.

[29] Xu C, Wang G, Liu X G, et al. Health Status Assessment and Failure Prediction for Hard Drives with Recurrent Neural Networks[J]. *IEEE Transactions on Computers*, 2016, 65(11):1-1.

[30] Wang J, Bao W, Zheng L, et al. An Attention-augmented Deep Architecture for Hard Drive Status Monitoring in Large-scale Storage Systems[J]. *ACM Transactions on Storage*, 2019, 15(3): 1-26.

[31] Wang Y, Ma E W M, Chow T W S, et al. A Two-Step Parametric Method for Failure Prediction in Hard Disk Drives[J]. *IEEE Transactions on Industrial Informatics*, 2014, 10(1): 419-430.

[32] Ying Zhao, Xiang Liu, Siqing Gan, et al. Predicting disk failures with HMM-and HSMM-based approaches[C]. *In Proceedings of the 10th Industrial Conference on Advances in Data Mining*. Springer, 2010: 390-404.

[33] Moises Goldszmidt. Finding soon-to-fail disks in a haystack[C]. *In Proceedings of the 4th USENIX Conference on Hot Topics in Storage and File Systems*. USENIX Association, 2012: 8.

[34] Lima F D D S , Amaral G M R , Leite L G D M , et al. Predicting Failures in Hard Drives with LSTM Networks[C]. *Brazilian Conference on Intelligent Systems*. IEEE Computer Society, 2017: 222-227.

[35] Sainath T N, Vinyals O, Senior A, et al. Convolutional, long short-term memory, fully connected deep neural networks[C]. *International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2015: 4580-4584.

[36] Sidi Lu, Bing Luo, Tirthak Patel, Yongtao Yao, Devesh Tiwari, and Weisong Shi, Making Disk Failure Prediction SMARTer!. *In Proceedings of the18th USENIX Conference on File and Storage Technologies (FAST '20)*, 2020:151-167.