inspur 浪潮

Proactive Disk Failure Prediction System towards Reliable Data Centers

State Key Laboratory of High-end Server & Storage Technology, Beijing, China

Tuanjie Wang, Xinhui Liang, Qiang Li

27 OCT 2020

Outline

- ✓ **Problem Analysis**
- ✓ Solution Overview
- ✓ Data Analysis
- ✓ Preprocessing
- ✓ Model Training
- ✓ Model Ensemble
- ✓ Conclusion

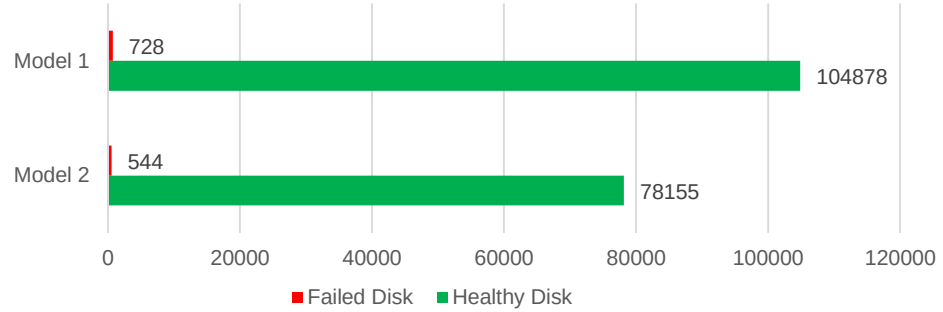
Problem Analysis

Problem:

- In large-scale data centers, disk is the component with the highest failure rate. Disk failure will seriously affect the stability and reliability of IT infrastructure.
- Given historical disk SMART data over a period of time, we need to predict whether each disk would fail or not within the next 30 days.
- Classification problem.
- Evaluation Metrics: F-score.

Challenges:

- The positive and negative samples are extremely imbalanced.
- The change of S.M.A.R.T. values is difficult to predict.
- The generalization ability of prediction model is challenging.



- Dataset Time Range： 2017-07-01 to 2018-07-31
- Healthy Disks : Faulty Disks = 143:1
- Negative Samples : Positive Samples = 40000:1

Outline

- ✓ Problem Analysis
- ✓ **Solution Overview**
- ✓ Data Analysis
- ✓ Preprocessing
- ✓ Model Training
- ✓ Model Ensemble
- ✓ Conclusion

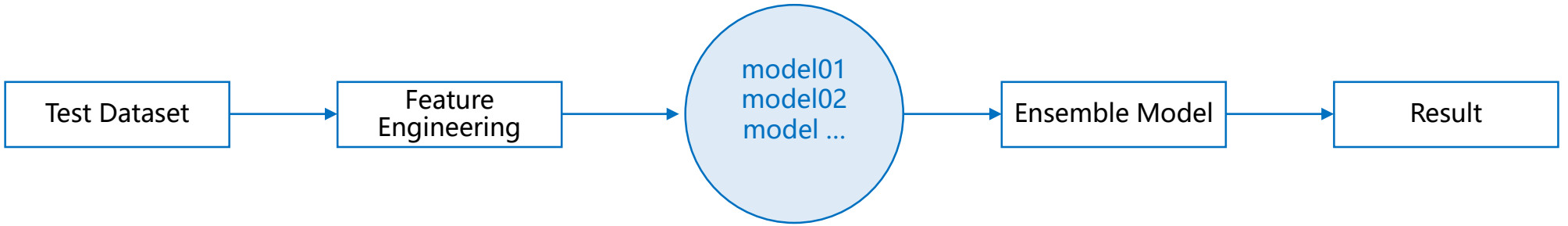
Solution Overview

Offline Training

Data Analysis	Sample Construction	Feature Extraction	Preprocessing	Model Training	Model Ensemble	Model Validation
range analysis	feature selection	original feature	filling missing values	AUC early stopping	soft voting	f-score max threshold
changepoint analysis	disk selection	time-series feature	sampling	hyperparameter tuning		
comparative analysis	split dataset		feature scaling	cost-sensitive learning		

model
preprocessing parameters
prediction threshold

Online Prediction



Contents

Data Analysis

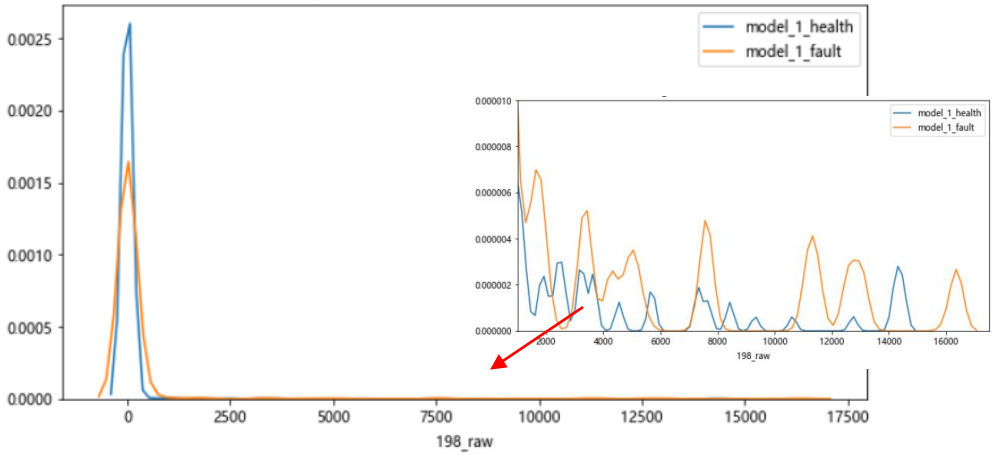
Feature Selection

- Number of valid（Not All Empty）SMART features: 48.
- Probability density distribution and KL divergence were calculated for healthy and faulty disks on each SMART feature.
- Raw SMART features selected finally include 5，187，192，193，197，198，199.

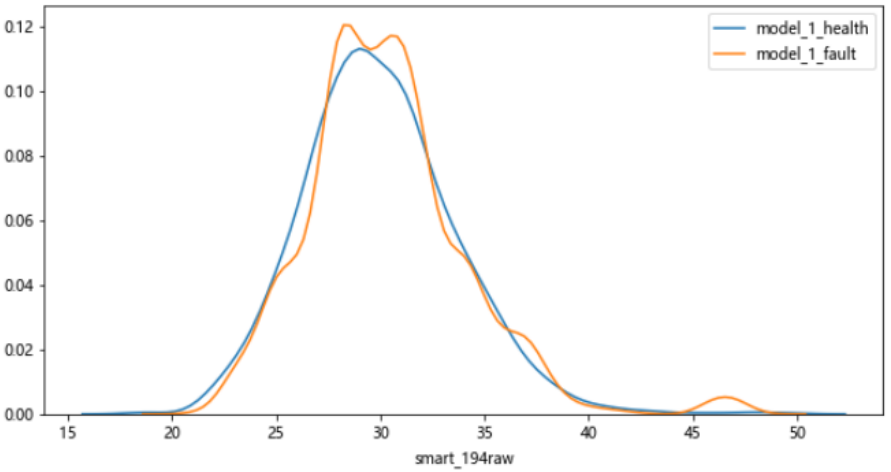
SMART IDS	Description
05-Reallocated Sector Count	Count of reallocated sectors.
187-Reported Uncorrectable Errors	The count of errors that could not be recovered using hardware ECC.
192-Unsafe Shutdown Count	Number of power-off or emergency retract cycles.
193-Load Cycle Count	Count of load/unload cycles into head landing zone position.
197-Current Pending Sector Count	Count of "unstable" sectors.
198-Reallocation Retries	The total count of uncorrectable errors when reading/writing a sector.
199-UltraDMA CRC Error Count	The count of errors in data transfer via the interface cable as determined by ICRC

https://en.wikipedia.org/wiki/S.M.A.R.T.

Probability Density Distribution of SMART 198 raw
KL divergence +inf



Probability Density Distribution of SMART 194 raw
KL divergence 0.015



inspur 浪潮

Data Analysis

Range Analysis

- There are 5,584 healthy disks with none-empty values, which need more attention.

Changepoint Analysis

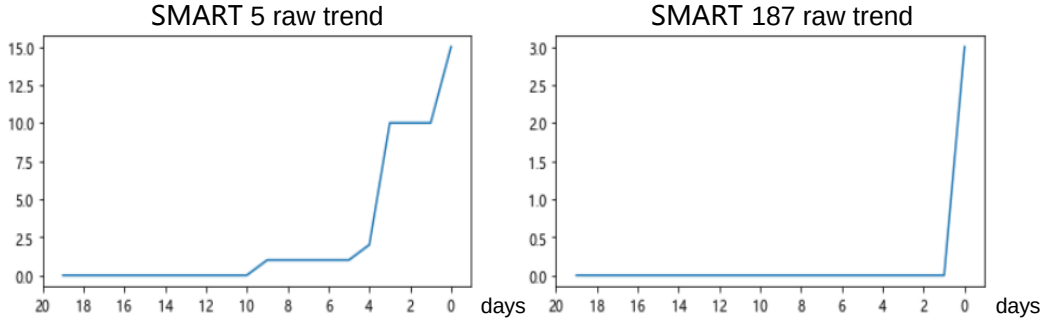
- Even in the last 7 days of the faulty disks, the values of 50%-75% of features such as SMART 5,187,197,198,199 are zeros.
- SMART value of faulty disks will not change significantly until the last two weeks.
- The closer to the end of life, the more likely sudden change will occur.

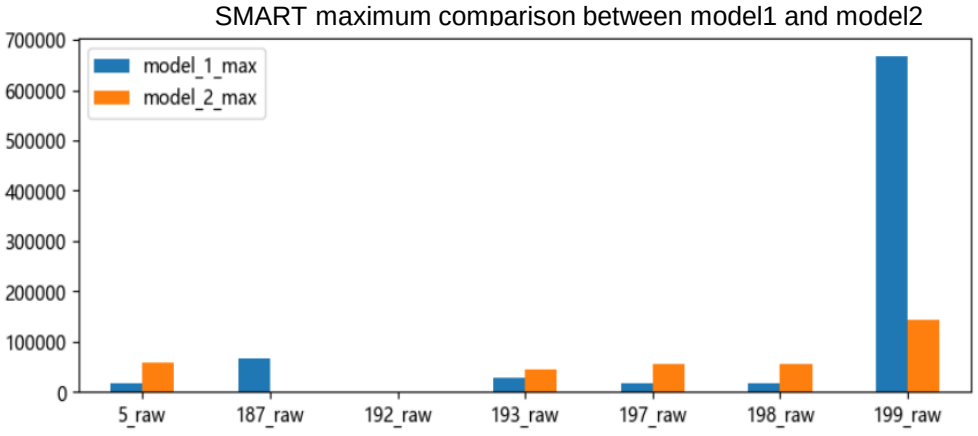
Comparison of different disk models

- The difference in the value range of each SMART feature between model 1 and model 2 is significant.

Conclusion

- Healthy disks with none-empty values should be used.
- Label the last few days of the faulty disks as positive.
- Time series feature extraction with sliding window between 3 and 7.
- The SMART values of different models should be scaled to the same range.





Outline

- ✓ Problem Analysis
- ✓ Solution Overview
- ✓ Data Analysis
- ✓ **Preprocessing**
- ✓ Model Training
- ✓ Model Ensemble
- ✓ Conclusion

Preprocessing

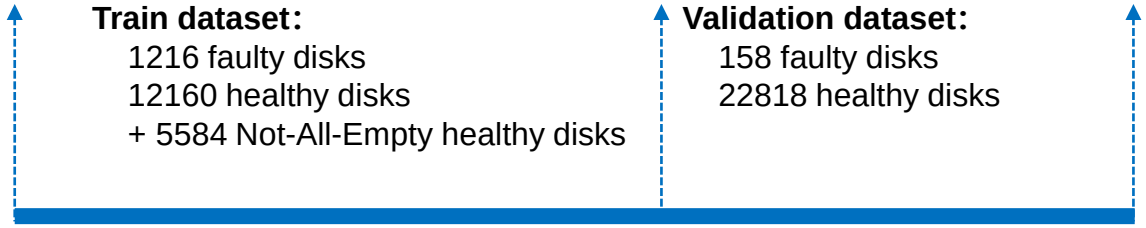
1. **Fill NaN：** backforward filling with zero

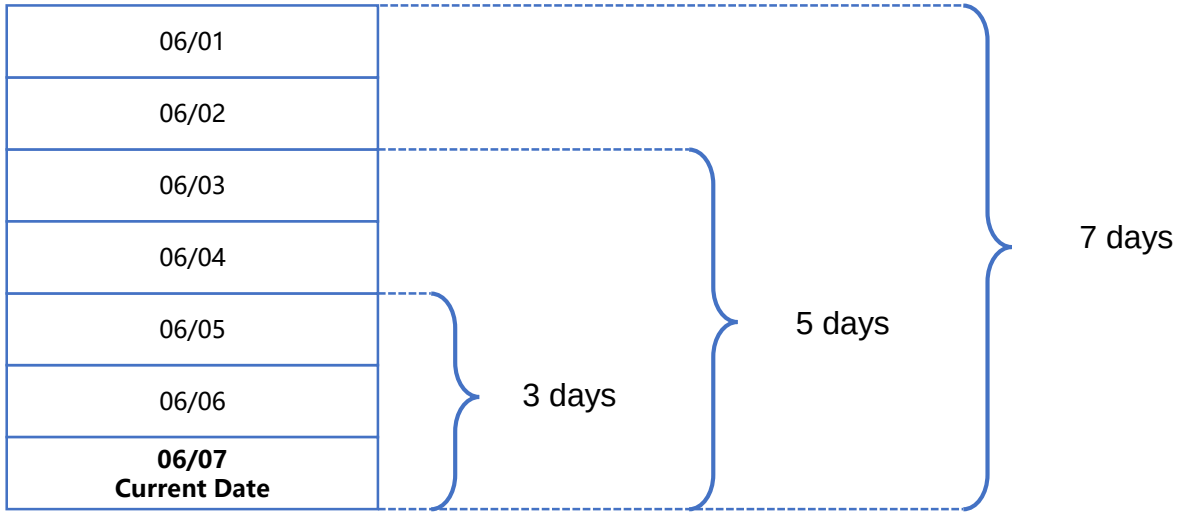
2. **Sample and Label：**
   - Use last 7 days and last 30th, 40th, 50th, 60th days.
   - Lable last 7 days of faulty as positive.

3. **Time Series Features Extraction：**
   ① Extract Windows：3，5，7
   ② Extract Functions：
      - change_time
      - change_rate
      - Std

4. **Scale Value Range of Different Disk Models:**
   Scale value range of disk model2 to the same as disk model1.

5. **Unified Scale：**
   standard scaler





Time Series Feature Extraction Windows

Outline

- ✓ Problem Analysis
- ✓ Solution Overview
- ✓ Data Analysis
- ✓ Preprocessing
- ✓ **Model Training**
- ✓ Model Ensemble
- ✓ Conclusion

Model Training

Algorithm：XGBoost

- The number of samples and the number of features are small.
- The hyperparameters of XGBoost are easy to adjust, and XGBoost is not easy to overfit.

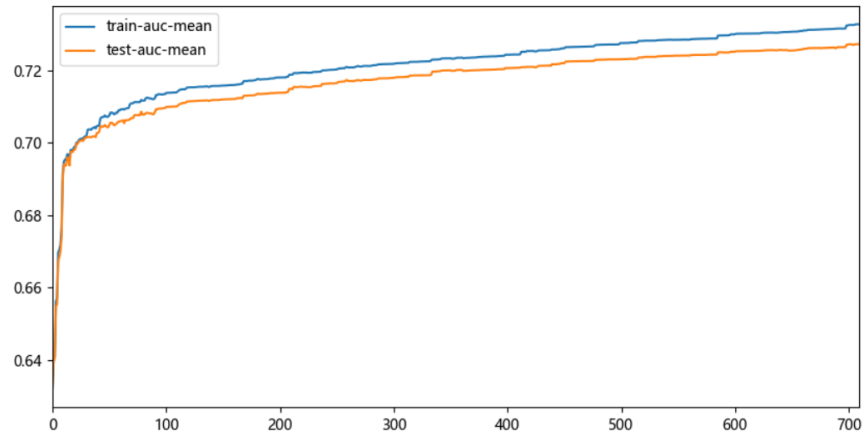
Hyperparameter Turning：

- CV: n_iterations is determined by 3-folder cross-validation , with AUC as the early stopping evaluation function.
- GridSearch: CART tree and regularization parameters are tuned with grid search.
- Cost-sensitive Learning: tune parameter scale_pos_weight to address the imbalanced distribution of healthy and faulty disks.

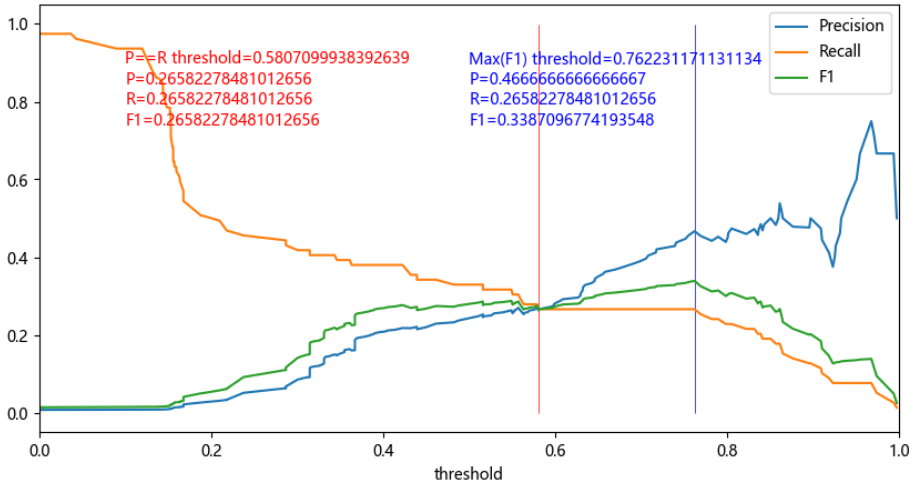
Prediction Threshold：

- Select the threshold with maximum F-score on validation dataset as prediction threshold.

AUC learning curve



F-score, Recall and Precision with different threshold



inspur 浪潮

Outline

- ✓ Problem Analysis
- ✓ Solution Overview
- ✓ Data Analysis
- ✓ Preprocessing
- ✓ Model Training
- ✓ **Model Ensemble**
- ✓ Conclusion

Model Ensemble

Selecte 6 sub-models that perform well on the validation set.

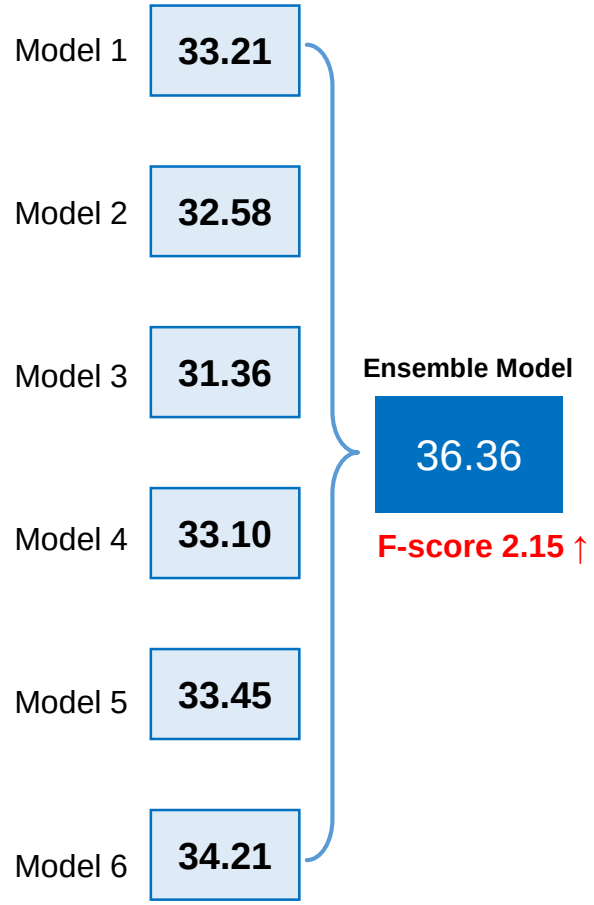
All the sub-models use the XGBoost as Algorithm.

Differences:

- SMART ids
- Feature extraction methods
- Feature extraction windows
- Positive positions
- Sampling positions

Model	SMART ids	Feature extraction methods	Feature extraction windows	Positive positions	Sampling positions
model_01	5_raw, 187_raw, 197_raw, 198_raw, 199_raw	std_values	3, 5, 7	< 7	0, 1, 2, 3, 4, 5, 6
model_02	5_raw, 187_raw, 192_raw 197_raw, 198_raw, 199_raw	change_rate change_time	3, 5, 7	< 3	0, 1, 2 30, 60
model_03	5_raw, 187_raw, 193_raw 197_raw, 198_raw, 199_raw	change_rate change_time	3, 5, 7	< 5	0, 1, 2, 3, 4 30, 40, 50, 60
model_04	5_raw, 187_raw, 192_raw 197_raw, 198_raw, 199_raw	change_rate change_time	3, 5, 7	< 3	0, 1, 2 30, 40
model_05	5_raw, 187_raw, 197_raw 199_raw	std_values	3, 5	< 7	0, 1, 2, 3, 4, 5, 6
model_06	5_raw, 187_raw, 192_raw 197_raw, 198_raw, 199_raw	std_values change_rate change_time	3, 5, 7	< 4	0, 1, 2, 3 30, 40, 50

Mean of Probability → **Final Probability**



F-score on Validation Dataset

Outline

- ✓ Problem Analysis
- ✓ Solution Overview
- ✓ Data Analysis
- ✓ Preprocessing
- ✓ Model Training
- ✓ Model Ensemble
- ✓ **Conclusion**

Conclusion

Contributions

Feature Engineering

- Extract time series features as a supplement.

Two-stage Normalization

- Eliminate the difference in the distribution of each disk model.

Hyperparameter Tuning

- CV with AUC early stopping and grid search.
- Cost-sensitive gradient boosting with XGBoost.

Ensemble Learning

- Make the prediction model more stable.

Semi-Finals Score：

4 /39.98/52.42/32.31

More Works

- Applying transfer learning algorithm to solve the problem of insufficient samples of faulty disks.
- Using ranking algorithms to make further improvements.
- Analyzing disks that are not reported in time or reported wrongly.

THANKS

for your listening