Out of Service(OOS) Alarm Prediction

NKU-Excavator

Tianyu Zheng, Chao Zhou, Meijun Jiang

Outline

1. Background
2. Data Processing
3. Model Optimization
4. Conclusion

PART 01

Background

01 Background

BS investment is increasing year by year.

Construction cost   Operating cost   Maintenance cost

The three major operators of China have more than 6.8 million base stations.

China Mobile's total revenue in 2019 was 700 billion, of which **1/3** was used for base station operation and maintenance.

In 2021, China will have a **400 billion** operation and maintenance market, and the global market will be nearly **1 trillion**.

01 Background

The Traditional Maintenance Method

STEP1

The base station is out of service, affecting the normal communication of users.

1

STEP2

2

The system issues maintenance instructions.

STEP3

3

Workers go to the base station to solve the fault.

STEP4

4

The base station resumes operation and records the cause of the failure.

- Lack of real-time performance
- High maintenance cost
- Low efficiency
- Passive

The Intelligent Network Maintenance

The intelligent network operation and maintenance achieves fault prediction through data analysis and continuous self-learning, which can effectively improve processing efficiency and accuracy

Big Data

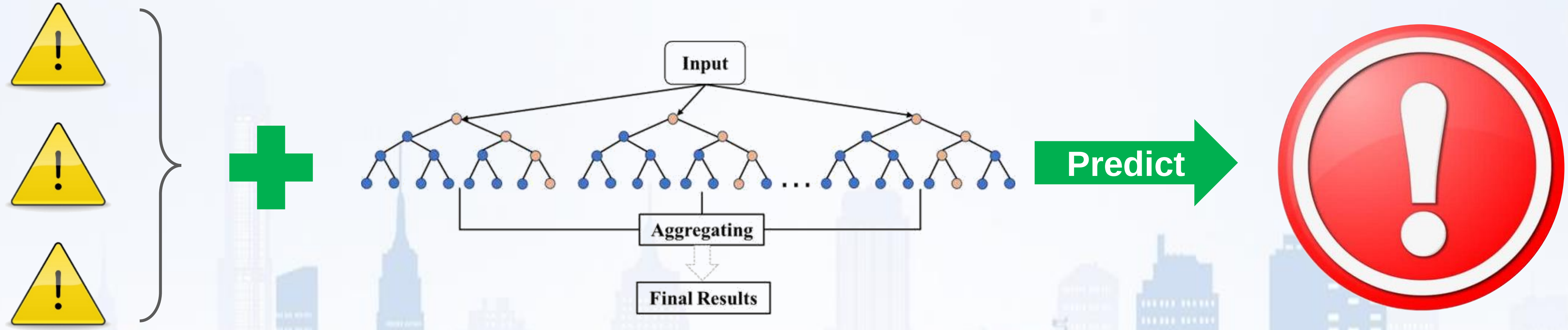
Artificial Intelligence

Maintenance Experience

Base Station Out-of-Service Alarm Prediction

When the network base station malfunctions, an alarm will be generated, which will cause network interruption and affect normal communication of users.

We can build an AI model to predict the probability of out-of-service alarms in the future through minor alarms, which will benefit maintenance personnel to deal with faults in advance and effectively avoid the base station from out-of-service.



Minor Alarms   AI Model   Out-of-Service Alarms

PART 02

Data Processing

02 Data Processing

Table 1. Original data

告警开始时间	基站名称	告警名称
2019-10-10 9:23	ACZDoAAEUAA	用户面故障告警
2019-10-10 9:24	ACZDoAAEUAA	用户面故障告警
2019-10-10 9:26	ACZDoAAEUAA	网元连接中断
2019-10-10 9:28	ACZDoAAEUAA	时钟参考源异常告警
2019-10-10 9:32	ACZDoAAEUAA	时钟参考源异常告警
2019-10-11 8:42	ACZDoAAEUAA	基站S1控制面传输中断告警
2019-10-11 8:44	ACZDoAAEUAA	网元连接中断
2019-10-11 8:50	ACZDoAAEUAA	时钟参考源异常告警
2019-10-12 9:47	ACZDoAAEUAA	传输光接口异常告警

Complex and disordered

Succinct and clear

target

Table 2. Train data

time	基站名称	网元连接中断	小区不可用告警	label
20200322	ACZDoAAEEAAAI1wABz	0.0	0.0	NaN
20200322	ACZDoAAEEAAAI1wAC9	0.0	0.0	NaN
20200322	ACZDoAAEEAAAI1wACU	0.0	0.0	NaN
20200322	ACZDoAAEEAAAI1wACh	1.0	0.0	NaN

ID Feature Label

- ID: Distinguish base station
- Feature: Training samples
- Label: Prediction target

02

Data Processing

Innovation 1: Generate labels accurately

2. Shift up the alarm columns

Judging by the warnings of the next day,
Use **Shift(-1)** function to shift up。

4. Eliminate forecast data

Exclude the data of March 22nd and March 30th.

1. Fill in missing dates

Fill in all time range data:
2019.10.01-2020.03.30

3. Generate label

Add the two columns and determine whether it is non-zero, then **label=1**

time	基站名称	网元连接中断		小区不可用告警
20200316	ACZDoAAEEAAAI1wACh	0.0		0.0
20200318	ACZDoAAEEAAAI1wACh	$\boxed{2.0}$	+	$\boxed{0.0}$

Fill in 20200317

Label=1

Innovation 2: Threshold adjustment

Unbalanced data set——label 0 : label 1 ≈ 5 : 1

Oversampling

Repeat the sampling of the data sample with label 1.

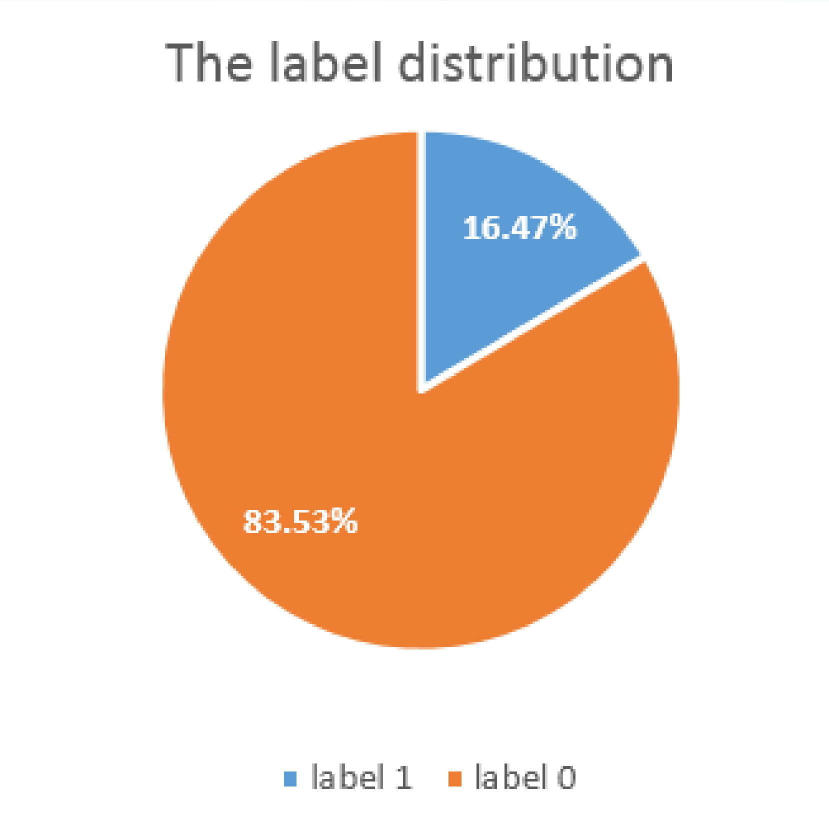
Undersampling

Randomly sample the data samples with label 0.

Threshold adjustment

Adjust the threshold for dividing label according to the ratio of 0 to 1.

datum：0.165



Innovation 3: Multi-dimensional features

Basic features

- Statistics of alarm times
  - Each type of alarm times
  - alarm_times per day
  - sum_each_times in a week
  - sum_times in N days(N<=7)
  - avg_times in N days(N<=7)
- Important alarm statistics
  - imp_alarm_times per day
  - time_interval(days/seconds)
  - time_interval_mean/std/max/min
  - imp_alarm_times in N days(N<=7)
  - sum_imp_time in N days(N<=7)

02 Data Processing

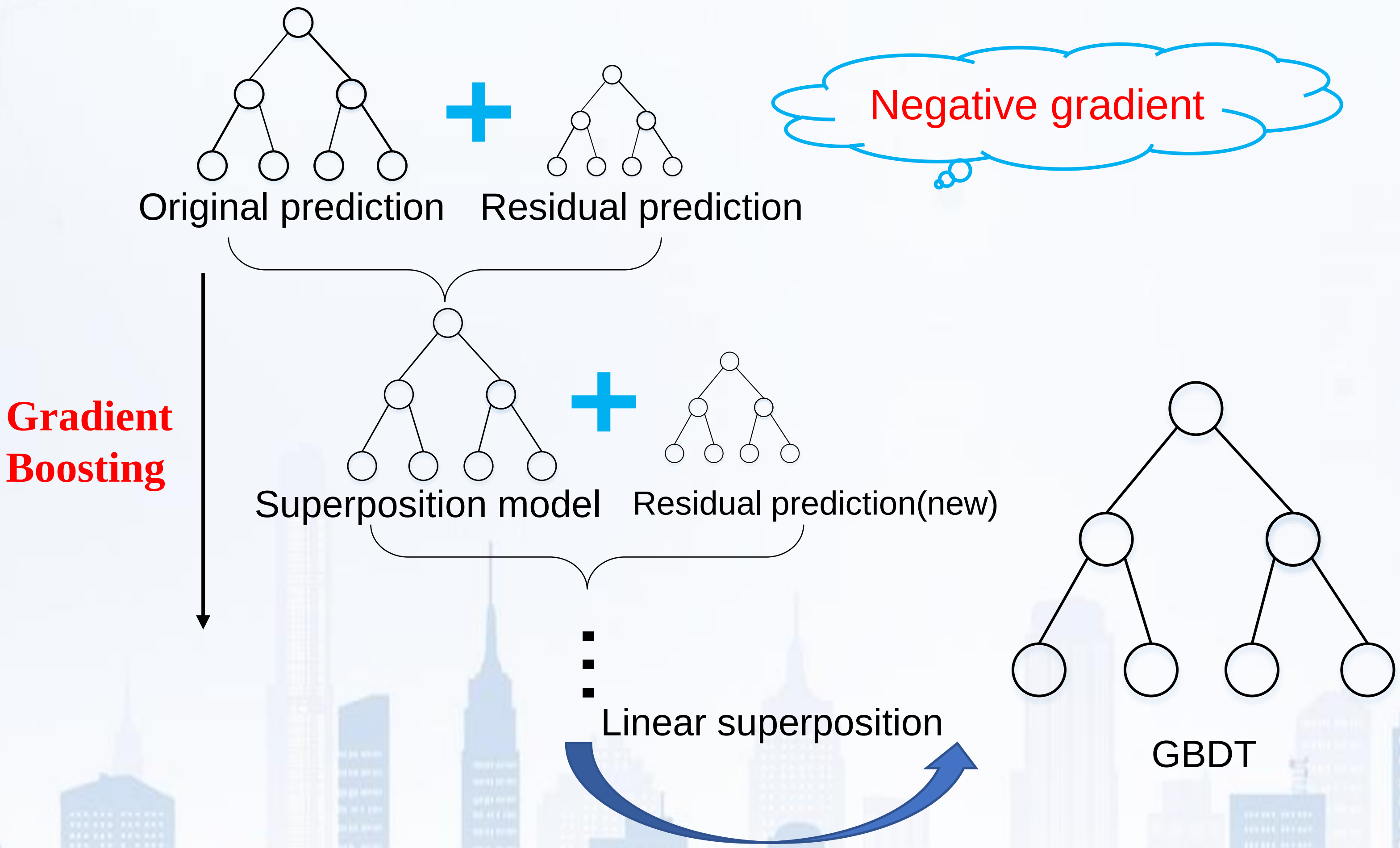
Innovation 3: Multi-dimensional features

- Alarm type statistics
  - malfunction_times
  - abnormal_alarm_times
  - failure_alarm_times
  - community_alarm_times
  - radiofrequency_alarm_times
  - BBH_alarm_times
  - RRU_alarm_times
  - RHUB_alarm_times
- Other features
  - daytime_alarm_numbers/tpyes
  - night_alarm_numbers/tpyes
  - alarm_time_earliest/latest per day
  - alarm_days in the previous 7 days
  - alarm_types in the previous 7 days
  - alarm_types_times in 7 days

PART 03

Model Optimization

Gradient Boosting Decision Tree (GBDT)



LightGBM is one of the implementation frameworks of the GBDT model.

- ✓ Training fast.
- ✓ High accuracy
- ✓ Low memory usage.

03 Model Optimization

- ✓ **Whether offline AUC is improved**
- ✓ **Whether the feature importance increases**
- ✓ **Whether the category ratio is consistent**

Cross-validation

5-fold cross-validation improves model stability and prevents overfitting

Grid Search

Find the optimal parameters of the model by grid search

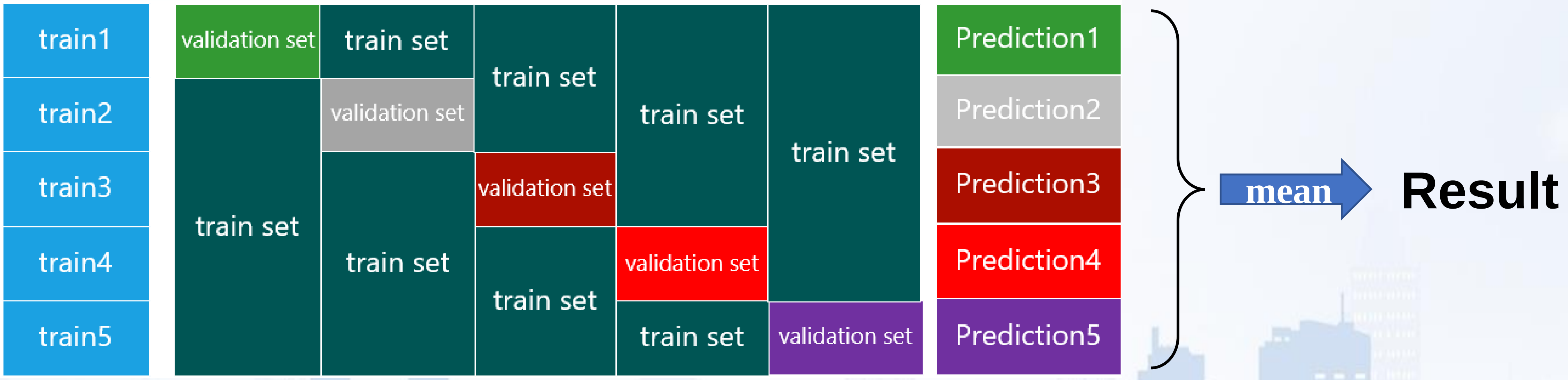
Model Merging

Through the integration of the advantages of each sub-model, the optimal model is generated by superposition.

03 Model Optimization

◆ Cross-validation

The 5-fold cross-validation method can make full use of the data set to prevent overfitting.



03 Model Optimization

◆ Grid Search

It traverses the parameter matrix to find the optimal parameters of the model.

Parameters	Grid search range	Optimal parameters
max_depth	-1、6、8、10	-1
n_estimators	100、500、1000	1000
num_leaves	16、32、51、61	32
learning_rate	0.01、0.02、0.05	0.02
feature_fraction	0.7	0.7
bagging_fraction	0.7	0.7
bagging_freq	5	5
feature_fraction_seed	7	7

03 Model Optimization

◆ Final F1-score

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

74.6667

基站退服告警预测排行榜

排名	参赛团队	指标得分	提交时间
1	黑白双煞	74.6667	2020.8.04 13:54
2	网络天眼	73.2943	2020.8.10 17:12
3	AI_STORM	72.5064	2020.6.23 11:34
4	tys	72.2924	2020.7.19 17:18
5	DB2	70.7365	2020.7.27 17:50
6	监控鹰眼	70.6221	2020.7.19 13:13

PART 04

Conclusion and

04

Conclusion

High prediction accuracy

Online F1-score is 0.747, 1% higher than the second place.

Training fast

Model training only takes 3-5 minutes, which is much faster than LSTM model.

Generalization ability

Collect two-year operating data of thousands of base stations to train the model.

Strong expandability

Multi-dimensional predictions can be made by adding features such as power consumption.

Feedback

- What are the benefits of BS out of service alarm prediction? why is it so important?

  1)Background: The base station maintenance cost is huge and needs to be solved urgently.
  2)Strength： Reduce maintenance costs by 25%-35%.
  Equipment life can be extended by 10%-15%.

- How to improve the generalization ability of the model?

  1)Add more base station alarm data for model training.
  2)Model merging: stacking merging and weighted merging.

- What are the advantages of your scheme?

  ★1)Generate labels accurately and automatically------achieve accurate prediction.
  2)Training fast------real-time prediction capability.
  3)Strong expandability

04

Conclusion

Future Development Direction

Dynamic model optimization

Comparing real-time data and prediction results. Achieve model self-optimization through continuous loop self-detection.

Alarm root cause location

Integrate AI algorithms with expert experience to realize intelligent fault location and improve the accuracy of root cause identification.

Base station energy saving

Add features such as power and throughput to predict future network traffic. Energy saving can be realized by means such as carrier sleep.

Build a cloud database

Gather fault data and algorithm parameters from various places to form a cloud sharing model to improve model accuracy and generalization capabilities.

Thanks