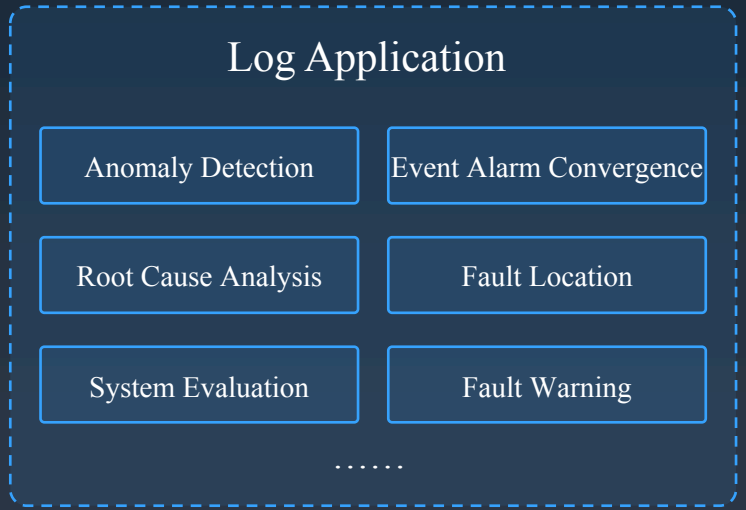
ITU AI/ML in 5G Global Challenge

Network anomaly detection based on logs

Chin：Qin YuKun，Song Yong，Xie Yuchen，Wei QiangShen

Logs are valuable for network maintenance with existing challenges

Huge chunks of data was generated by Network Equipment and these information can be used to detect existing faults to improve the network maintenance.

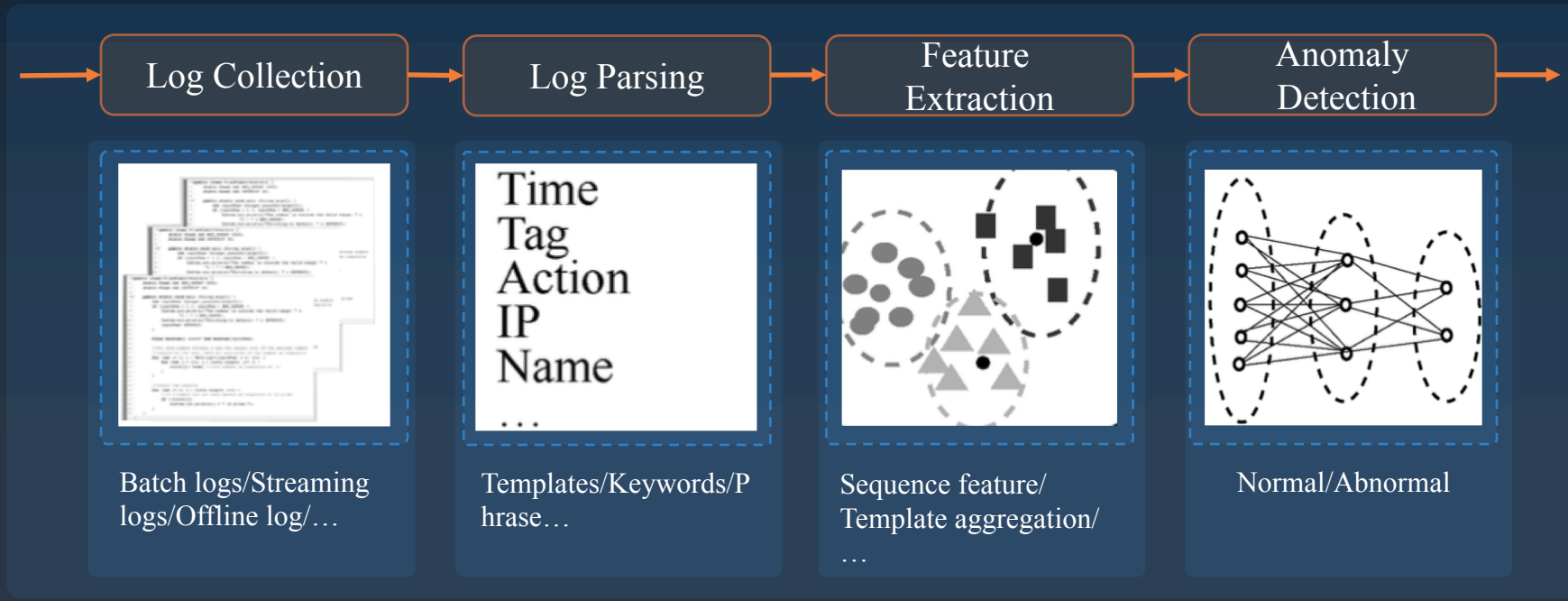


Challenges

- Increasingly vast amount（1GB/H+）
- Various types (Exchanger、Server、Router, etc.)
- Unstructured data
- Unsupervised / No-Label
- Tradeoff between effect and efficiency

Ensemble Solution: NLP & Abnormal Detection

Network Anomaly Detection based on Logs was composed by NLP and abnormal detection.

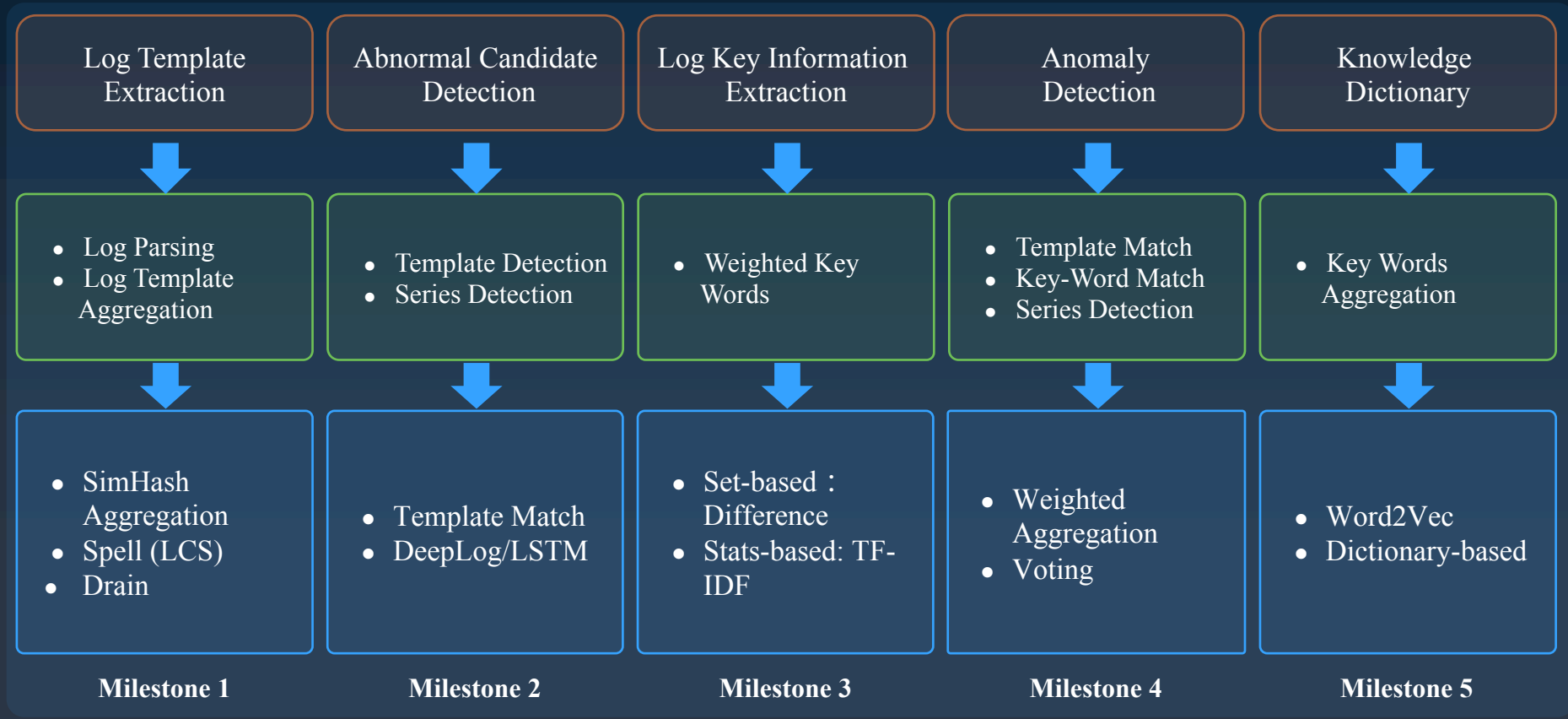


Key-word Dictionary Construction is the most important part

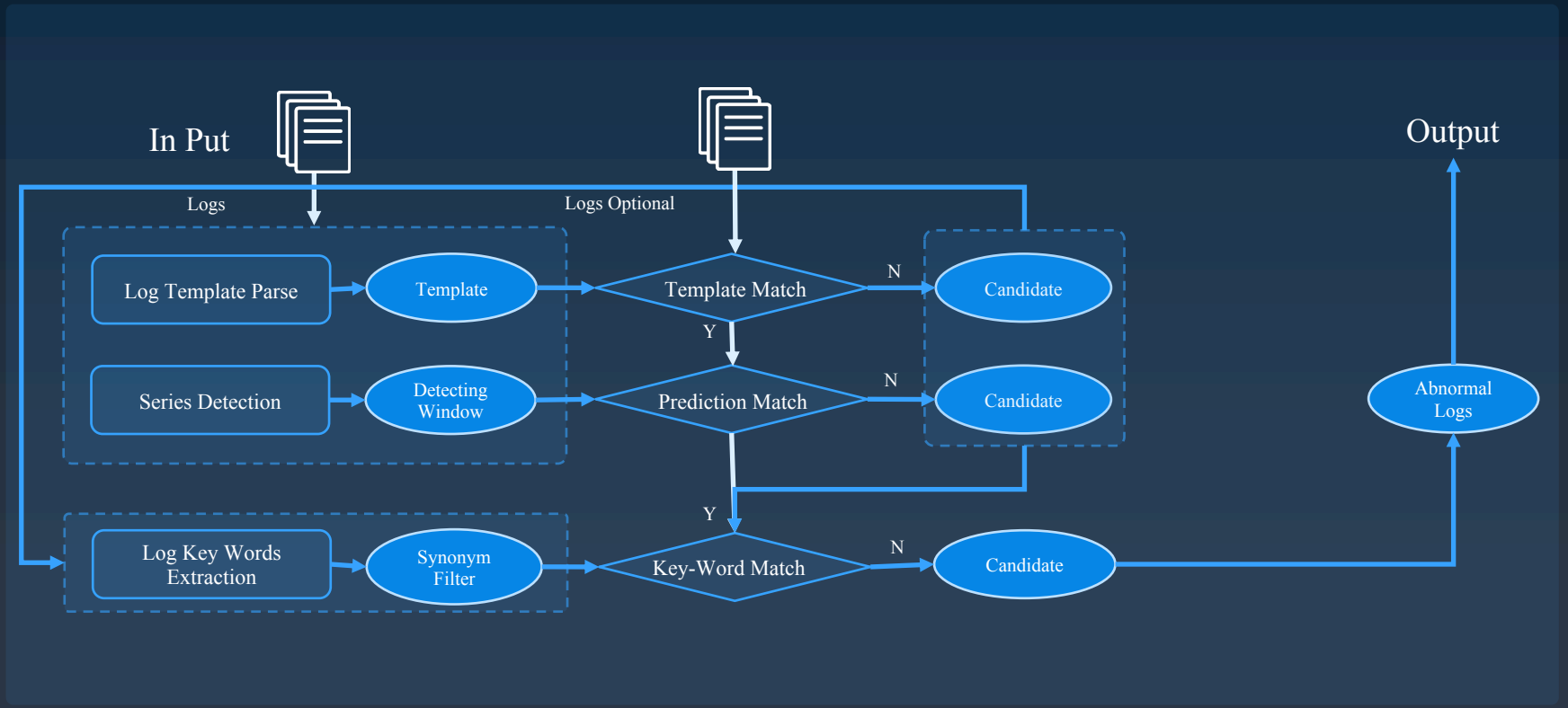
Although there are plenty of AI algorithms such as Deep Learning involved in log anomaly detection, all methods without key words performs not good enough.

1. RISTO V, PIHELGAS M. LogCluster — a data clustering and pattern mining algorithm for event logs[C]// Conference on Net-work and Service Management (CNSM). 2015: 1-7.
2. HEPJ, ZHUJM, HESL, etal. Towards automated log parsing for large-scale log data analysis[J]. IEEE Transactions on Dependable and Secure Computing, 2018, 15(6): 931-944.
3. STUDIAWAN H, SOHEL F, PAYNE C. Automatic event log ab- straction to support forensic investigation[C]// ACSW 2020. 2020: 1-9.
4. STUDIAWAN H, PAYNE C, SOHEL F. Automatic graph-based clustering for security logs[C]// Advanced Information Networking and Applications(AINA). 2019: 914-926.
5. DAI H, LI H, CHEN C S, et al. Logram: efficient log parsing using n-gram dictionaries[R]. 2020.
6. DU M, LI F F. Spell: online streaming parsing of large unstructured system logs[J]. IEEE Transactions on Knowledge and Data Engi- neering, 2019, 31(11): 2213-2227.

Industry-level solution is empowered by evolution

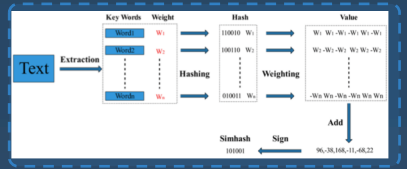


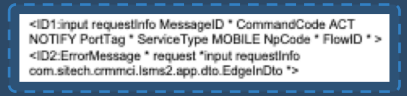
Pre-detection is used to update key word dictionary for detection fixing



Fast template extraction is utilized to make pre-detection based on LSTM much more efficient

Frequency-based Template Extraction





Tokenization

Frequency Filter

Wildcard Replace

Simhash Aggregation

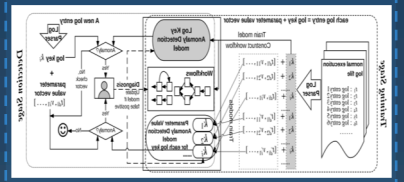
Template

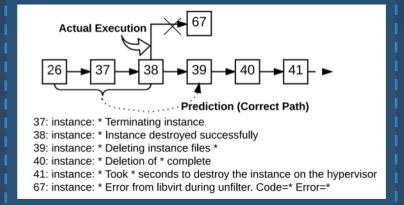
Series Anomaly Detection by LSTM

Workflow

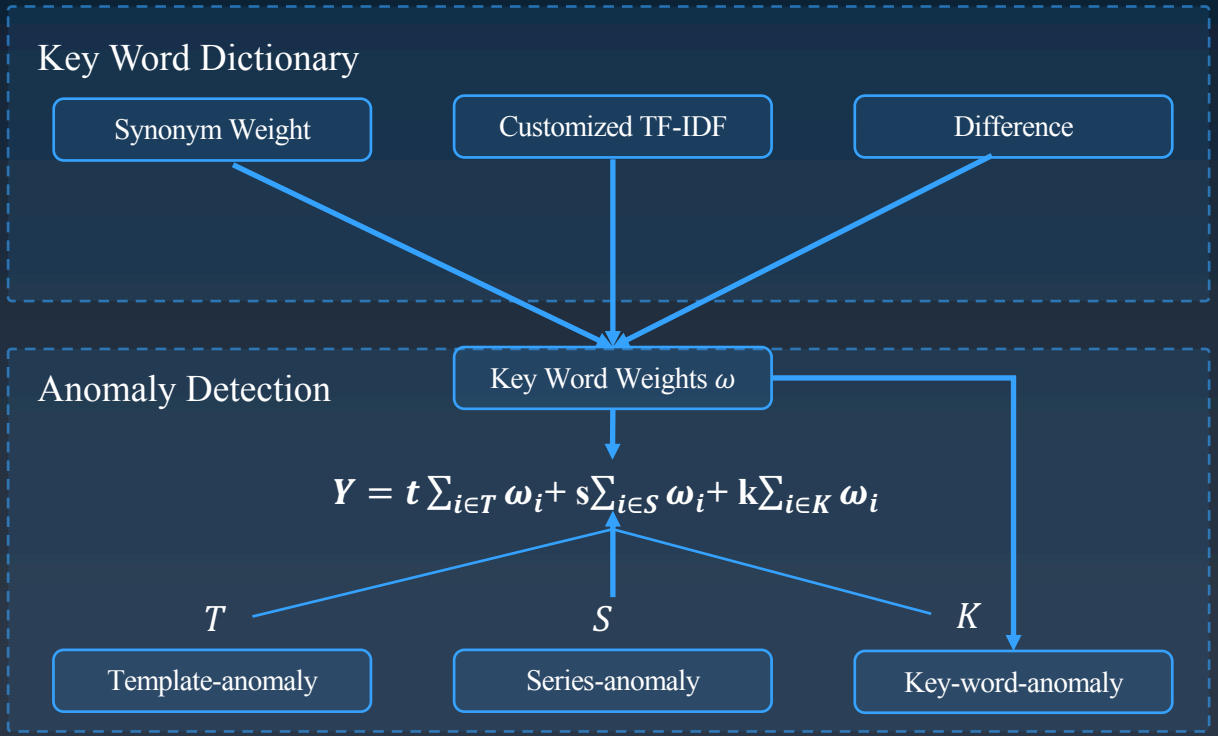
LSTM

Anomaly Candidate





Weighted detection with key-words fix generates the final anomaly



The only 100 out of 100 score in competition dataset

Precision=100%；Recall=100%；***F1=1.00***

Top 1/379 at China Unicom 5G+AI Network Application Special Competition 。

- Advantages

① More efficient template extraction；

② Increasingly evolved knowledge dictionary and potential knowledge graph construction；

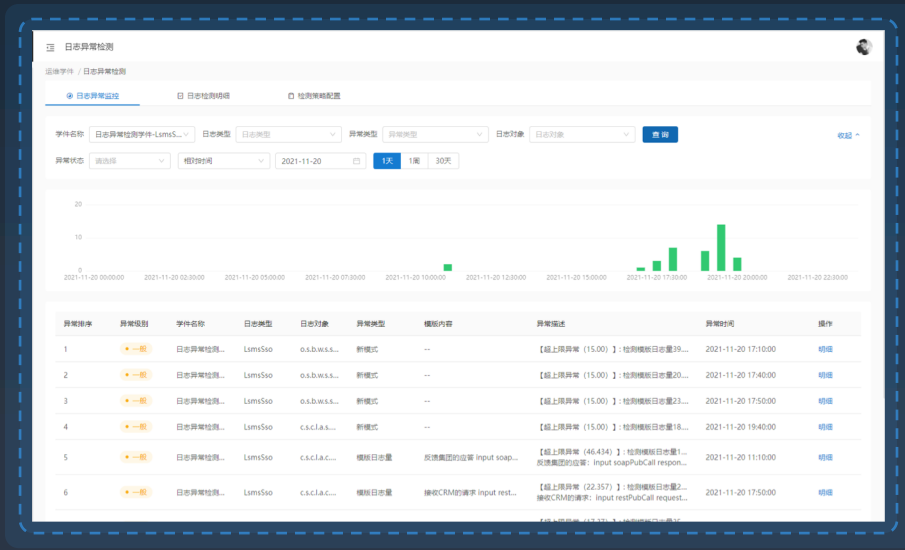
③ Perfect tradeoff between time consumption and performance；

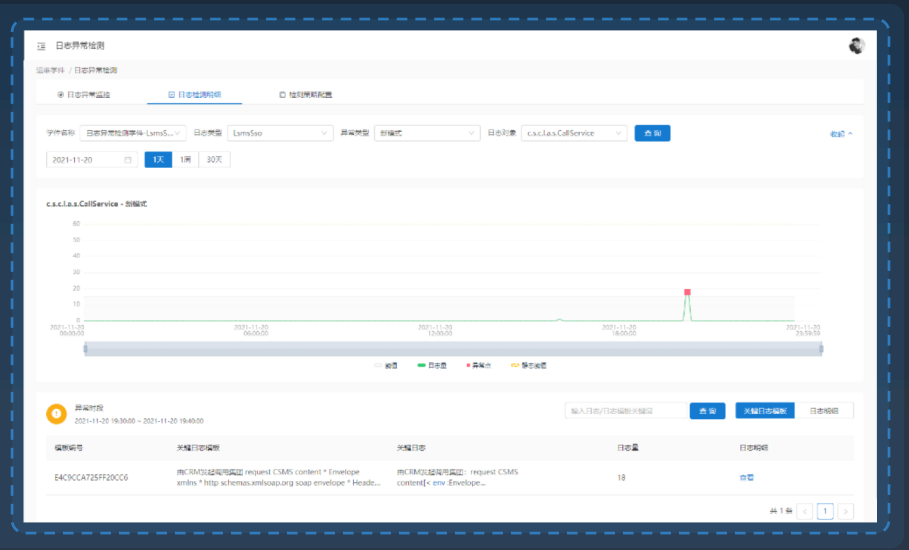
④ Optional fast detecter with dictionary

- Result

Industrial Application is perfect

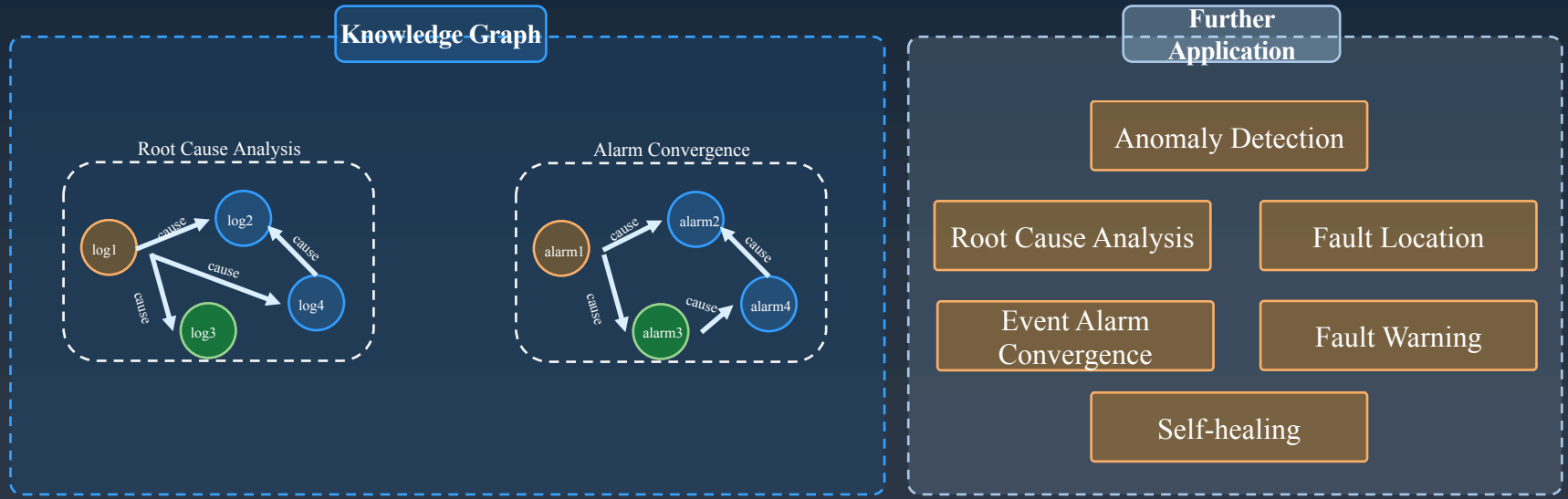
- Productization in AIOps
- Deployed for more than 20 systems involved in both BSS/OSS
- Daily processed 0.92TB data with 92% accuracy





Network will be more intelligent thanks to log anomaly detection

Knowledge Graph
Root Cause Analysis



Thank you

Q & A