Enterprise Railway Data Service Platform Based China Railway Safety Application

Zhongtieke (Beijing) Software Technology Co. Ltd
Introduction
The development of China Railway, the information systems, PB data of CR safety, the demand of application scene and the solution of "platform + application"

Railway Data Service Platform
The overall architecture, technical architecture, functional architecture and design features of RDSP

Application In China Railway Safety
The typical application scenario and the result obtained

Conclusion
Summarize the results of the application in CR safety and look forward to the future work from the platform and application

Zhongtieke (Beijing) Software Technology Co. Ltd
CHAPTER ONE

Introduction
The Rapid Development of China Railway

China has the world’s largest high-speed rail networks with nearly 20,000 km of track in service.
The Safety Information Systems of China Railway

More Than 60 Railway Safety Systems, eg. 5T, 6A, 6C

- **Rolling stock safety monitoring (5T)**
  - Trace Hotbox Detection System
  - Trackside Acoustic Detection System
  - Truck Performance Detection System
  - Trouble of Moving Freight Car Detection System
  - Train Coach Running Safety Diagnosis System

- **Locomotive vehicle safety protection (6A)**
  - Air brake/ fire-proofing/ high-voltage insulation / power supply/ running gear monitoring/ video monitoring record

- **Catenary and power supply detection (6C)**
  - High speed pantograph, pantograph pan, ground power supply detection
The Data Volume of China Railway Safety

- Sensors: 1,000,000+
- Operating Mileage: 124,000+
- Stations: 7000+
- Systems: 60+
The Requirement of China Railway Safety

Risk Early-warning

Hidden-danger identification

Qualitative Analysis of Accident and Failure

Overall Evaluation of Safety

Recommendation and Solution
The Solution-->

"Platform+Application"

**Risk early-warning and hidden danger identification**

**Qualitative analysis of accident and fault**

**Improvement and solution recommendation**

**Overall evaluation of safety**

---

**Railway Data Service Platform**

Zhongtieke (Beijing) Software Technology Co. Ltd
CHAPTER TWO

Railway Data Service Platform
The Technological Architecture of RDSP

Data Integration
- Unstructured Data
  - Streamsets/ftp/sftp
- Structured Data
  - Streamsets/sqoop

Data Storage
- HDFS/HBase
- Ceph/HBase
- SparkStreaming/Flink
- Spark/Mapreduce
- Unstructured data annotation
- Data cleaning and quality management

Data Analysis
- Spark MLib/Mahout
- Tensorflow
- Greenplum
- kylin
- Railway professional data analysis model

Data Service
- Greenplum
- Postgres
- Redis
- ElasticSearch
- NOSQL(Hbase Phoenix)
- DAAS
- JDBC/ODBC
- Restful API

Basic Data
- Master Data
- Geography Information
- Metadata

Zhongtieke (Beijing) Software Technology Co. Ltd
The Functional Architecture of RDSP

- **Data Service**
  - Data Integration
  - Data Analysis
  - Data Storage
  - Data Application

1. Data Collection
2. Data Storage
3. Data Analysis
4. Data Visualization

Safe operation center
Maintenance management
The Feature & Innovation of RDSP

- Storage and Computation are decoupled
- Distributed Computing Architecture
- Elastic Stretching Capacity
- Drag-and-drop Data Integration
- One-stop Self-help Analysis
- Container based Microservice
CHAPTER THREE

Application In China Railway Safety
1. Data Storage and Retrieval

- Electronic distributed storage and full-text retrieval, solving the problem of too many paper files when targeting one
- Efficient retrieval, almost real-time, and distributed storage can be extended
1. Data Storage and Retrieval

- almost real-time retrieval
- extended storage
2. Risk Early-warning and Hidden Danger Identification

Risk alert and hidden danger elimination are intelligent recognition of harmful factors and hazards, which may lead to hidden dangers, accidents or faults.
2. Risk Early-warning and Hidden Danger Identification

- **Flaw 1:** Cannot target hidden dangers precisely, and the examination has a wide scope, but no focus.

- **Flaw 2:** Due to different skill levels, people cannot find the real problem.

- **Flaw 3:** Only eliminate risks and hidden danger which can be seen, cannot find potential risks effectively.

*Current measurements:*

According to safety instructions, examine and eliminate one by one, taking historical record and expert experience into account.
2. Risk Early-warning and Hidden Danger Identification

Integrated multi-source data, big data platform analyzes and mines those data deeply, through statistical model, machine learning and deep learning algorithm, to target risks and hidden dangers more efficiently, and furthermore to predict accidents and faults happening in the future.

What Our Platform Do

1. Key points stand out
2. Target precisely
3. Identify potential risks and hidden dangers
2. Risk Early-warning and Hidden Danger Identification

- Project risk and hidden danger data with high dimension into low dimension
  - Linear method: PCA, LDA
  - Nonlinear method: Isomap, GDA

- Correlation among risks, hidden dangers and accidents
  - Correlation coefficient: Pearson
  - Covariance matrix
  - Information entropy

- Based on historical data, recognize seasonal pattern of hidden dangers, accidents and faults appearance
  - AR model
  - MA model
  - ARMA model

- Build a regression model, using risks and hidden dangers to predict accidents in the future
  - Linear & nonlinear regression
2. Risk Early-warning and Hidden Danger Identification

application within single major-Track

Through analysis of multi-source track data and GIS display, know comprehensively key risk points and hidden dangers along the railway.

Meanwhile find potential risks and hidden dangers, to instruct risk control and hidden danger elimination.
2. Risk Early-warning and Hidden Danger Identification application within single major-Track

Safety Risk → Hidden danger → Accident Failure

- Rail Cracks
- Tiny Cracks
- Rail Gap
- Rail Inpurity
- Weld Joint
- Rail Irregularity
- Ballast
- Roadbed
- Joint Component
- Temperature
- Rail Fatigue

Rail Break → Rail Bruise → Rail Expand

Train Derailment

Speed Limitation

Stop

Control → Elimination → Prevention
2. Risk Early-warning and Hidden Danger Identification application across majors-Signal & Communication

- Signal and communication
- Various type of equipment, such as track circuit, track and so on

Possible causes
When one equipment have a failure, it may be not caused by that equipment itself, but by problems from other majors such as one failure in track.
2. Risk Early-warning and Hidden Danger Identification application across majors—Signal & Communication

Officer on duty in signal and communication finds the track red light appearance.

Signal and communication
Filed check: no vehicle or foreign body
Track circuit: normal
This morning just maintain fasteners.

Finally it is because a short circuit between fasteners and the basic track.
Record ASAP

Signal and communication

Zhongtieke (Beijing) Software Technology Co. Ltd
2. Risk Early-warning and Hidden Danger Identification

Application across majors-Signal & Communication

1. Phenomenon of accidents or failures
   - the track red light appearance

2. Analysis on big data platform
   - monitoring data of track circuit
   - maintaining data of track circuit

3. Trace responsibility of accidents or failures
   - 80% caused by wet
   - 15% caused by insulation damaged
   - 5% caused by such as rain and thunder

4. Improvement and maintenance
   - Find the corresponding responsible unit to repair and improve it

Zhongtieke (Beijing) Software Technology Co. Ltd
3. Qualitative Analysis of Accident and Failure

Deeply investigate accidents and failures, focus on weak aspects such as employees, management, equipment and so on.

- Help supervisors to trace responsibilities.
- Establish knowledge base of accident & failure causes and corresponding measurements, to achieve a better before-event proof and after-event maintenance.
3. Qualitative Analysis of Accident and Failure

Parse report file, locate and extract the name of accidents or failures, the location, causes and so on.
- Name entity recognition (NER)

Intelligent classification of causes
- Semantic similarity analysis
- Semantic classifiers

Calculate possibilities of causes

Investigate and trace responsibility of accidents or failures, based on possibilities of causes

Accident Name: Cable trench cover invasion
Cause: Strong wind and heavy rain

$p(y, x) = \prod_{t=1}^{T} p(y_t | y_{t-1}) p(x_t | y_t)$

$P(Y|X) = \frac{1}{Z} \prod_{i=0}^{n-1} \Psi_i(Y_i, Y_{i+1}|X)$
4. Improvement and Solution Recommendation

When finding risks and hidden dangers, or after accidents or failure appearance, improvement and maintenance are needed. Traditional methods usually come from paper documents, manuals, or staffs’ experience, which leads to difficulties and inefficiency.

Current Problems

- Too many paper files, hard to find
- Limited experience
- Various methods
- Effective methods hard to popularize
4. Improvement and Solution Recommendation

Intelligent QA

Through organizing non-order materials and information, intelligent question and answer system is building a knowledge classification model, to offer information of railway improvement, maintenance and method.

- anywhere, anytime
- always updated

functions
- related QA
- top questions
- hot words
- question intelligent completion

approaches
- build knowledge base and knowledge map through natural language processing (NLP)
- full-text retrieval of QA
The track red light appearance is a usual accident. The main causes are the following: 1) wet-80%; 2) insulation damaged by human-15%; 3) weather such as rain and thunder-5%.

Measurements: 1) dry; 2) change the damaged part; 3) test and make sure track circuit normal.
5. Overall Evaluation of Safety

Use a dashboard to evaluate the overall safety condition and trend.

Bring advantages for supervisors and corresponding units to know the overall situation about safety.
5. overall evaluation of railway safety

- Accidents and failures
  - Safety evaluation standards
    - Build standards for railway bureau, majors and sections
  - Effects of maintenance
- Hidden dangers
  - Evaluating content
    - Risks
      - Level of risks, risk number, no. of risks under monitoring
    - Hidden dangers
      - Level of hidden dangers, number, no. of eliminated dangers
- Effects of maintenance

Zhongtieke (Beijing) Software Technology Co. Ltd
CHAPTER 4

Conclusion
Conclusion

By the RDSP, accidents and failures are reduced by nearly 20%.

Before (without the RDSP)

- Safety data is printed, unstructured
- Query and statistics are difficult
- Identification of risk and hidden danger by human being
- Passive safety, Rectification after the accident

After (with the RDSP)

- Safety data is digitized, structured
- Full text retrieval and big data analysis
- Hidden danger and risk identification by Big data Analysis
- Active safety, Risk early-warning Overall evaluation
The Future Work

Enrich platform functions and modules

Improve the usability and performance of the platform

Expand the application to passenger and freight service

Encapsulate railway professional data analysis model

Platform & Application
The RDSP is Moving Forward

Thank You!