Hui-fang WANG, Zi-quan LIU, 2019. An error recognition method for power equipment defect records based on knowledge graph technology. *Frontiers of Information Technology & Electronic Engineering*, 20(11):1564-1577.
https://doi.org/10.1631/FITEE.1800260

An error recognition method for power equipment defect records based on knowledge graph technology

Key words: Error recognition; Power equipment defect record; Knowledge graph; Machine learning

Corresponding author: Hui-fang WANG
E-mail: huifangwang@zju.edu.cn
ORCID: http://orcid.org/0000-0002-1483-364X

Motivation

- Power equipment defect records are short texts with limited information, most of which do not contain redundant information. It is almost impossible to use information redundancy in a single text.
- The complexity of power equipment structure, defects, and colloquial phenomena in defect records result in diverse texts, making it difficult to comprehensively consider the error conditions of defect records by manual analysis or even power industry specifications based on expert knowledge.

Main idea

- With the interpretable graph structure, a knowledge graph can express the complex relationships among information contained in texts, which makes it possible to identify key information by knowledge reasoning.

- By constructing a knowledge graph of power equipment defects from the existing defect records and a graph search algorithm for error recognition, errors in new records can be identified with the aid of the knowledge graph.

Method

1. Knowledge graph of power equipment defects

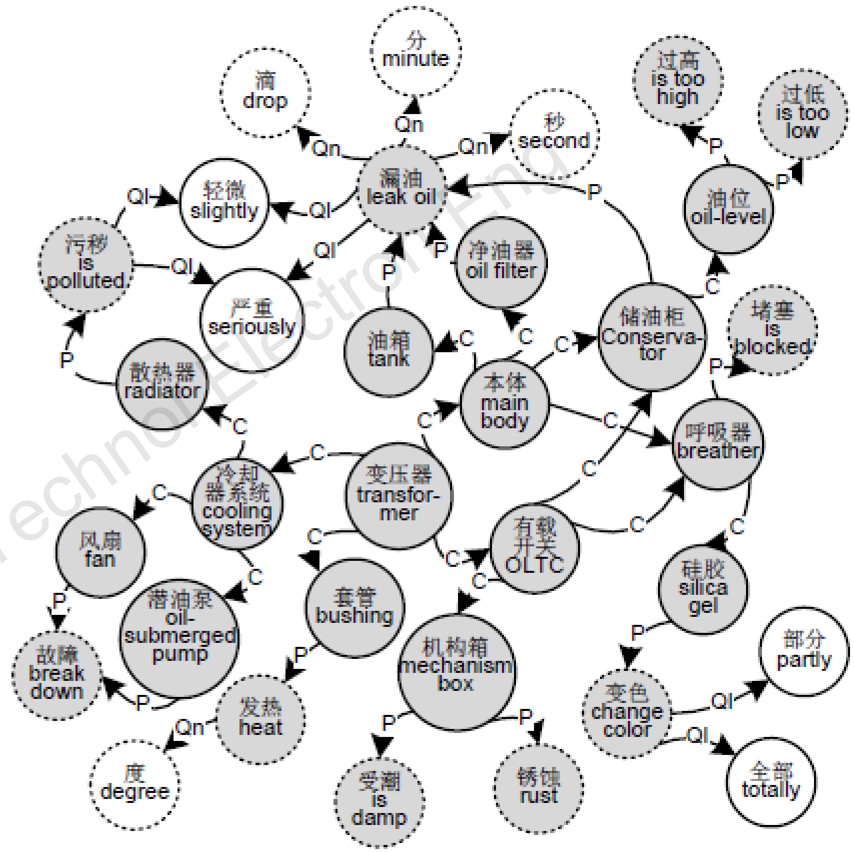

Fig. 5 Part of the transformer defect knowledge graph

A node with a solid line and gray bottom is an En node, a node with a dotted line and gray bottom is a Pv node, a node with a solid line and white bottom is a Pad node, and a node with a dotted line and white bottom is a Pq node

2. Graph search algorithm for error recognition

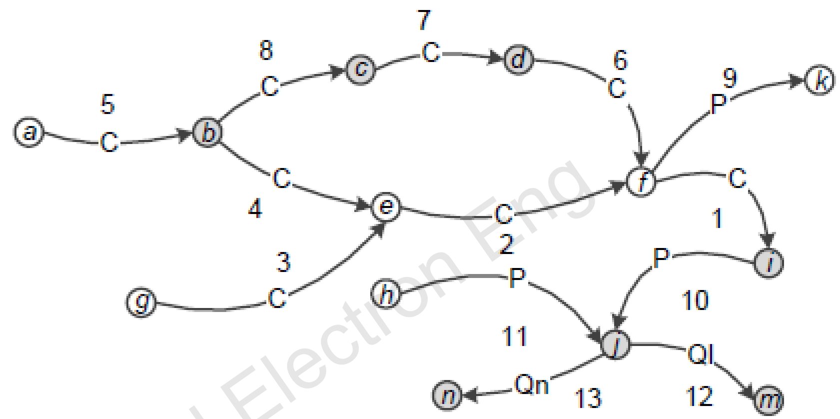

Fig. 4 An example of the knowledge graph

Nodes *a*–*i* represent entities with POS "En," nodes *j* and *k* represent properties with POS "Pv," node *m* represents property with POS "Pad," and node *n* represents property with POS "Pq." An edge with "C" means that the relation between *p* and *q* (*p* points to *q*, the same as below) is "*p* contains *q*," an edge with "P" means that the relation between *p* and *q* is "*q* is a defect phenomenon of *p*," an edge with "Ql" means that the relation between *p* and *q* is "*q* is a qualitative description of *p*," and an edge with "Qn" means that the relation between *p* and *q* is "*q* is a quantitative description of *p*"

Major results

Table 3 Statistical results of different models

Model	P (%)	R (%)	F (%)	A (%)	t (ms)
LR	100.00	20.86	34.51	92.21	212.69
MLSVM	70.06	64.44	67.13	93.79	3879.63
SVMR	73.97	62.30	67.63	94.13	6975.86
WRF	70.10	72.73	71.39	94.26	349.13
KG1	87.50	97.33	92.15	98.37	7885.31
KG2	88.19	97.86	92.78	98.50	10 228.56

The machine learning models and knowledge graph models were employed to recognize incorrect records in the test set.

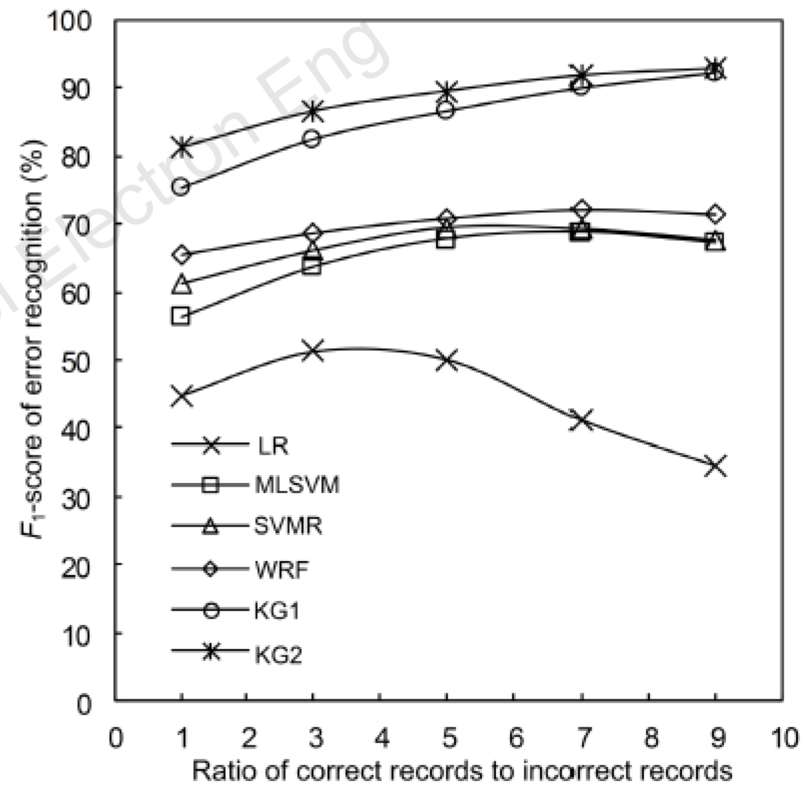

Fig. 6 F_1-score curves of different models under different data skewness

Conclusions

- The construction of a knowledge graph of power equipment defects has been described in detail.
- Based on the knowledge graph, the graph search algorithm has been proposed to recognize errors in power equipment defect records.
- Results and analysis of examples indicated the remarkable superiority in recognition accuracy (98%) and efficiency (10 s for 3798 records) of the knowledge graph model. So, the error prompt can be properly given in real time when an incorrect defect record is entered by the inspector, which will ensure the quality of defect records from the source.