



电力通信运行管理中典型业务数据的智能关联方法

吴桂龙, 杨志敏, 黄昱

(中国南方电网电力调度控制中心, 广东 广州 510525)

摘要: 电力通信运行管理过程中, 会产生和存储各类相对独立的业务数据(如故障工单、值班日志、检修工单、巡检记录等), 这些业务数据为电力通信网运行管理提供了重要支撑。目前大多数业务数据的统计过程相对独立, 后期较少人工加以关联。选取了电力通信运行管理中值班日志与故障工单两种典型的业务数据, 采用文本挖掘技术, 构建无监督召回和监督分类相结合的机器学习模型, 提出值班日志与故障工单的智能关联方法, 并利用电力通信运行管理系统中相关历史业务数据, 对智能关联方法进行实验验证, 达到较好的关联效果。

关键词: 电力通信运行管理; 典型业务数据; 值班日志; 故障工单; 智能关联方法

中图分类号: TP319

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2021014

Intelligent correlation method of typical business data in power communication operation management

WU Guilong, YANG Zhimin, HUANG Yu

Power Dispatching and Control Center of China Southern Grid Co., Ltd., Guangzhou 510525, China

Abstract: In the process of power communication operation management, various independent business data, such as trouble tickets, duty logs, maintenance tickets, and inspection records, are generated and stored. These business data provide important support for the operation management of the power communication network. At present, the statistical process of most business data is relatively independent, and there is less manual correlation in the later stage. Two typical business data of duty log and trouble ticket in power communication operation management were selected, text mining technology was used to build a machine learning model combining unsupervised recall and supervised classification, and the intelligent association method between duty log and trouble ticket was proposed. Besides, the relevant historical business data in the electric power communication operation management system was used to do the experimental verification of the intelligent association method. The results show that it can achieve positive effect.

Key words: power communication operation management, typical business data, duty log, trouble ticket, intelligent association method

1 引言

电力通信运行管理作为电力通信运行支撑系统中重要的组成部分,主要完成通信运维流程表单的信息化管控等功能。在电力通信运行管理过程中,会产生和存储各类相对独立的业务数据,比如故障工单、值班日志、检修工单、巡检记录等,这些业务数据为电力通信网运行管理提供了重要的支撑。目前大多数业务数据在统计过程中相对独立,后期也较少人工加以关联,比如大部分值班日志数据未关联至故障工单,若能将值班日志自动关联到相关故障工单,即可获取某项故障丰富的跟进日志信息,以更好地掌握故障的来龙去脉,有利于统计与分析故障。因此,开展电力通信运行管理中典型业务数据的智能关联方法研究对提升电力通信运行和管理效率,进一步保障电力通信网的安全稳定运行具有重要而实际的意义。

当前,电力通信运行管理在关联各类相对独立的业务数据时,一般采用人工关联方法,以值班日志与故障工单两种典型的业务数据为例,当需要关联与故障工单有关的值班日志,以获取与该故障工单有关的故障跟进信息时,只能通过人工检索关键词并加以判断进行关联,关联过程烦琐且耗费人力。电力通信运行管理中的业务数据主要是文本数据,当需要实现文本数据之间的智能自动关联时,可采用自然语言处理(natural language processing, NLP)技术,构建机器学习

模型的方法^[1-2]。目前,国内在电力文本数据挖掘方面已经取得了一定的进展^[3-4],实现了文本挖掘在电力领域的实际应用,如参考文献[5]提出的基于告警信号文本挖掘的电力调度故障诊断,参考文献[6]提出的基于卷积神经网络的电力设备缺陷文本分类模型研究。文本数据挖掘在电力领域的应用经验为研究电力通信运行管理中业务数据的智能关联方法提供了有益借鉴,比如在数据预处理时应采用电力通信专用词汇。另外,信息检索、搜索引擎等领域所采用的文本相似度计算方法^[7-8]也是本文所提出电力通信运行管理中业务数据的智能关联方法中的关键技术。

本文选取了电力通信运行管理中值班日志与故障工单两种典型的业务数据,采用文本挖掘技术,构建无监督召回和监督分类相结合的机器学习模型,提出值班日志与故障工单的智能关联方法,通过利用电力通信运行管理系统中相关的历史业务数据,对关联方法进行实验验证,证明了智能关联方法的有效性。

2 电力通信运行管理典型业务数据样本

2.1 值班日志数据样本

电力通信运行管理系统中的值班日志主要以记录值班事务以及各类故障、检修信息,是保证电力通信高效运行的重要手段,值班日志数据主要包含日志记录时间、日志内容、关联故障工单号等信息,表 1 展示了 3 条值班日志实例,其中关联故障工单号一列是值班调度员将值班日志

表 1 值班日志数据样本实例

日志记录时间	日志内容	关联故障工单号
2019-06-03 11:27:03	2019-06-03 08:25:05 报 A、B 变调度数据网设备异常,于 2019-06-03 11:25:56 启动故障流程处理	190603001
2019-06-04 13:06:33	C 电厂, D 换中兴设备板卡电压越限告警,已通知相关运维单位调集备品备件后报票进站更换	190401064
2019-06-05 16:15:39	调度数据网综合监视报 3 条通道异常,初步核实传输通道无异常,已通知超高压 E 局进站(A、B 变)处理	190305015



人工关联至相关的故障工单,表 1 中数据已做脱敏处理,其中变电站、电厂、换流站、供电局等名称均用大写字母代替。

2.2 故障工单数据样本

电力通信运行管理系统中的故障工单是处理光缆、设备等故障的信息化运维流程表单,故障工单数据主要包含故障工单号、故障名称、故障现象描述、故障发生时间等信息,表 2 展示了 3 条故障工单实例,表 2 中的数据已做脱敏处理,其中变电站、电厂、换流站、供电局等名称均用大写字母代替。

3 智能关联方法设计方案

3.1 整体方案

本文采用 NLP 技术对文本进行分析与处理,利用故障工单中故障名称或故障现象描述与值班日志内容的文本匹配程度,同时参考故障发生时间与日志记录时间的匹配程度,可得到与某项故障工单关联的所有跟进值班日志。本文对故障工单和值班日志中相关文本进行分析与处理时,采用了中文文本分词、去停用词以及词频-逆文档频率(term frequency-inverse document frequency, TF-IDF)^[9-10]、词向量 Word2vec^[11]、BM25 (best match 25, BM25) 算法^[12]等 NLP 技术,建模时采用了性能优异的机器学习模型,如梯度提升决策树(gradient boosting decision tree, GBDT)模型^[13]

以及 LightGBM (light gradient boosting machine, LightGBM)^[14]等。本节将以值班日志与故障工单两种典型的业务数据为例,具体阐述电力通信运行管理中业务数据智能关联方法的设计方案。

另一种方案是把关联问题转化为监督二分类问题,通常与无监督方法相比,监督方法可达到更高的精确率。监督二分类方法中,将故障工单与值班日志进行配对并打上标签,构造二分类训练集,具体来说,对于某条故障工单,分别跟与其关联的每条值班日志合并为特征记录,并打上标签 1,与其无关联的每条日志也合并为特征记录,并打上标签 0。但此方案存在的问题是在生成训练集时,每条故障工单需分别与所有的日志配对打标签,这将导致大量的特征记录标签为 0,造成正负样本比例不平衡,不利于二分类模型的训练。

本文提出的电力通信运行管理中典型业务数据的智能关联方法综合了上述两种方案的优点,本文所提智能关联方法的整体方案如图 1 所示,分无监督召回和监督分类两个部分。

首先对故障工单以及值班日志的中文文本做分词、去停用词处理,接着进行第一步的无监督召回,每条故障工单通过 3 种相似度衡量因子(TF-IDF、Word2vec 和 BM25)分别获取相似度最高的 30 条候选值班日志,接着对 3 种方法召回的各 30 条值班日志进行合并去重,得到故障工单

表 2 故障工单数据样本实例

故障工单号	故障名称	故障现象描述	故障发生时间
190603032	主干网 A 变第 13 槽至 B 变第 13 槽光路异常 (B 侧拉曼故障)	主干网 A 变第 13 槽至 B 变第 13 槽光路误码,导致 B 侧再生段帧丢失,该超长距光路不间断性误码,初步判断为 B 侧光迅拉曼光放器件工作异常	2015-3-13 13:24:43
190370934	主干调度数据网 X 平面 A 变思科设备异常	主干调度数据网 X 平面 A 变思科设备出现内存不足告警,与路由器 b 及交换机 a 互联通道不通	2015-5-25 10:08:00
190894508	主干传输网 C 换设备 7 槽板卡故障	巡视网管发现主干传输网发现 7:00 开始多条直流保护通道、C 换稳控通道反复出现 VC12 信号劣化告警后自动复归,与现场确认业务未受影响。经过检查分析发现告警电路均经过 C 换,检查 C 换设备无板卡异常告警	2015-9-17 07:30:54

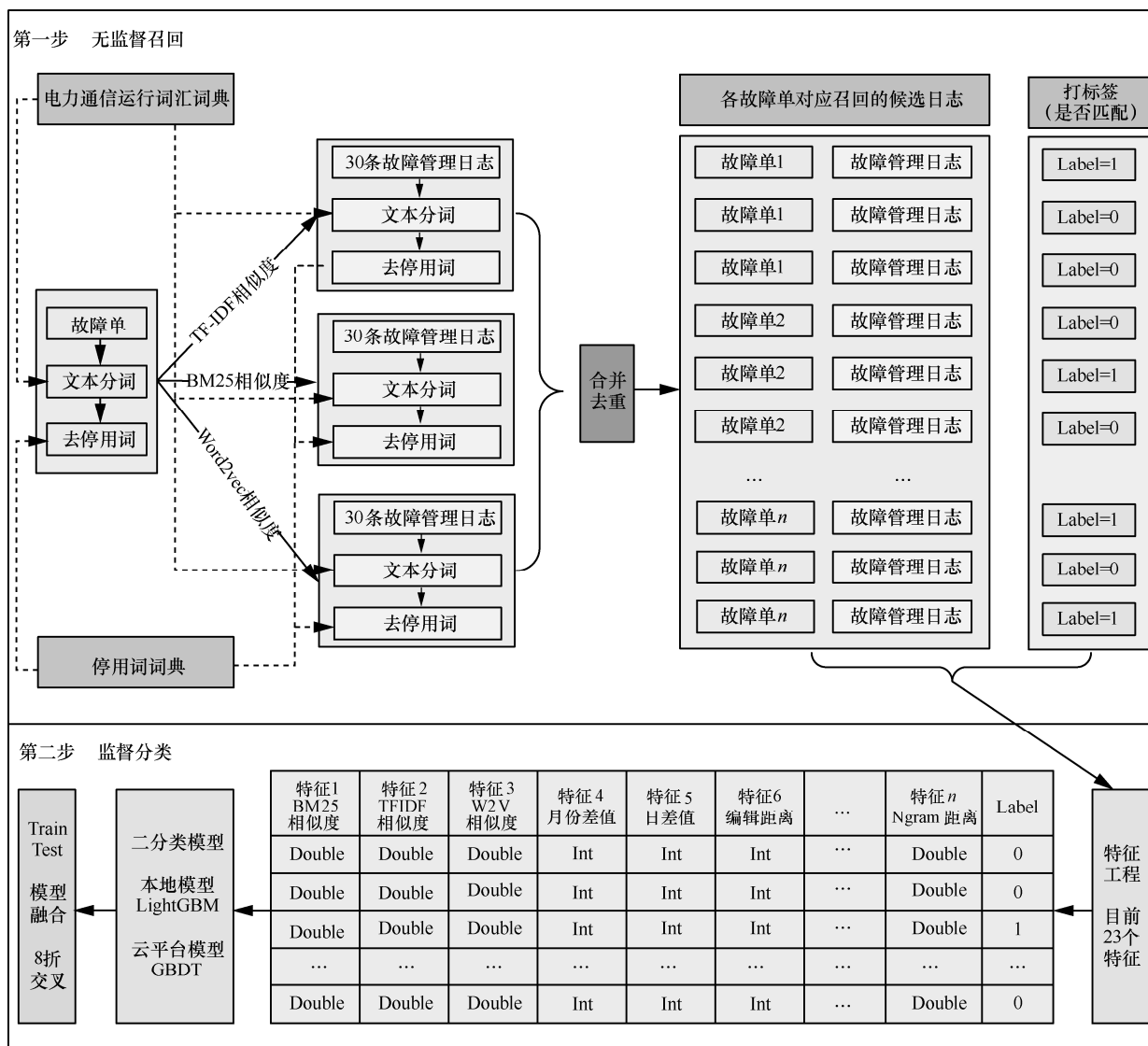


图1 智能关联方法整体方案示意图

的候选日志,并根据实际是否关联人工打上标签,得到可用于后续监督分类的标签数据集。

第二步的监督分类中,首先进行特征工程,根据任务需求构造出23个特征,接着将含有特征列和标签Label列的数据记录送进二分类模型中,在本地训练8折交叉的LightGBM模型,在南方电网调度云平台上训练GBDT模型,分别得到最终的二分类关联模型。

进行数据测试,即对某故障工单关联值班日志时,先为该故障工单召回候选的关联日志,接

着生成相应的特征列,经过二分类模型生成结果标签,其中标签为1的记录所对应日志即最终得到的关联值班日志。

3.2 数据预处理

数据预处理部分主要是对故障工单以及值班日志中的中文文本做分词、去停用词处理,其中中文分词采用了“结巴”分词,“结巴”分词是一个Python中文分词组件,可以对中文文本进行分词、词性标注、关键词抽取等,并且支持自定义词典^[15]。“结巴”分词主要通过词典进行分词,



正因如此,分词结果的优劣很大程度上取决于词典。针对电力通信运行管理中典型业务文本数据具有领域所独有词汇,本方案采用自定义词典的“结巴”分词。自定义的电力通信运行词汇词典主要包含了南方电网主干通信传输网各网元的名称、设备槽位、设备单盘、设备端口、各类专有名称以及英文代号,如E1、W1、VC4、VC12等。

文本分词后进行停用词的过滤,停用词主要包括英文字符、数字、数学字符、标点符号及使用频率特高的单汉字等,比如中文的语气助词、副词、介词、连接词等。但对于本文特定任务,某些数字和英文字符因具有特殊含义,具有较强的特征属性,需保留,因此本文在停用词过滤中并未整体过滤所有的数字和英文字符,而是采用自定义字典的形式保留部分数字和英文字符。此外,本文的停用词词典还包含了一些会影响关联性能的词语,比如调度员姓名等词语。以实际文本为例,对文本“2014-09-09 17:50:38 传输网/A 水电站, 05 盘 3 端口发生连接信号丢失(LINK_LOS)”进行分词、去停用词后,得到分词文本“传输网 A 水电站 5 盘 3 端口 发生连接 信号 丢失 LINKLOS”。

3.3 无监督召回方案

完成文本数据预处理后,初步分析故障工单及值班日志数据相关列信息,发现故障名称列较故障现象描述列更能准确地描述故障内容,因

此,本文采用故障名称文本列代表故障工单,对值班日志进行候选召回,日志则用日志内容文本列代表。

无监督召回中,第1种召回方法采用TF-IDF,将故障名称及值班日志内容文本转化为TF-IDF。具体地,一条故障名称或日志内容记录均可认为是一篇文档,用文档中每个词的TF-IDF组成的词袋向量表示该文档,再根据余弦相似度计算某条故障工单与所有日志的相关性,按照余弦相似度的大小对日志进行排序,取相似度最大的前30条日志作为该故障工单的召回候选日志。选取一条故障工单为例,该故障工单名称为“主干网A变-B中调光路异常”,表3展示了采用TF-IDF召回的相似度排序靠前的3条候选日志,从结果可以看出召回的候选日志在内容上基本与故障工单有关,说明采用TF-IDF可有效召回。

第2种召回方法采用Word2vec,将故障名称以及日志内容文本作为语料库,利用Word2vec模型训练出每个词的词向量,一条故障名称或日志内容记录均可认为是一篇文档,接着把文档中每个词的词向量加以平均得到该文档的向量表示,根据余弦相似度计算某条故障工单与所有值班日志的相关性,接着按照余弦相似度的大小对日志进行排序,同样取相似度最大的前30条日志作为该故障工单的召回候选日志。

同样选取一条故障工单作为例子,该故障工

表3 TF-IDF召回候选日志实际效果示例

故障名称	日志内容	故障与日志相似度
主干网A变-B中调光路异常	2019-10-24 05:05:26,报主干网A变-B中调光路异常,于2019-10-24 04:49:16启动故障处理流程	0.870 65
主干网A变-B中调光路异常	调度员2019-10-24 05:03:09,对编号为190603019的故障进行了派工,故障名称为:主干网A变-B中调光路异常,B中调负责处理故障	0.813 11
主干网A变-B中调光路异常	2019-10-24 04:57:00,B中调至220 kV K变光缆故障,导致主干传输网B中调至S变2.5 Gbit/s光路、主干传输网B中调至A变10 Gbit/s光路先后中断,无重要业务受影响	0.379 37

单名称为“主干传输网 D 中调至 E 变光路异常”，采用 Word2vec 召回的相似度排序靠前的 3 条候选日志见表 4，从结果可以看出召回的日志在内容上基本与故障工单有关，说明采用 Word2vec 召回的有效性。另外，发现部分召回的值班日志在内容上虽与故障工单相似，但时间上却有差别，这个问题可通过后续构造的时间特征加以校正。

第 3 种召回方法采用 BM25，利用 BM25 算法计算故障名称和每条日志内容文本之间的相关性，并按照 BM25 相关性大小对日志进行排序，取相关性最大的前 30 条日志作为该故障工单的召回候选日志，采用 BM25 方法同样可有效召回日志。本文采用 3 种不同的相似度衡量因子，目的是利用文本表示方法的不同方向寻找相似文本，以保证较高的召回率，取相似度靠前的 30 条日志，可同时保证较为平衡的正负样本比例和较高的召回率。

最后将上述 3 种方法召回的所有候选值班日志进行合并去重，并根据实际是否关联匹配打上标签，若故障工单与召回日志存在关联关系，则打标签 1，若无关联则打标签 0，得到用于后续监督分类的标签数据集。经实验计算，通过本召回方法，关联日志的召回率可达 90% 以上，即整体上 90% 以上实际存在关联的日志已存在候选日志中，且标签数据集正负样本比例约为 1:10，比例较为平衡，适合进行下一步的二分类。

3.4 监督分类方案

监督分类方案包括特征工程和模型训练两个过程，首先进行特征工程，对无监督召回得到的记录数据，根据任务需求构造出有利于区分标签的 23 个特征，包括故障名称与候选日志内容之间的 TF-IDF 余弦相似度、Word2vec 词向量余弦相似度、Word2vec 词向量曼哈顿距离、Word2vec 词向量欧氏距离、BM25 相关性；故障现象描述与候选日志内容之间的 TF-IDF 余弦相似度、Word2vec 词向量余弦相似度、Word2vec 词向量曼哈顿距离、Word2vec 词向量欧氏距离、BM25 相关性；故障工单发生时间与日志记录时间的年份差、月份差、日差；未作分词的故障名称与候选日志内容之间的文本编辑距离^[16]（包括 Levenshtein 距离和 Jaro-Winkler 距离）；未作分词的故障现象描述与候选日志内容之间的文本编辑距离；分词后的故障名称与候选日志内容之间的文本编辑距离；分词后的故障现象描述与候选日志内容之间的文本编辑距离；分词后的故障名称与候选日志内容之间的 2-gram 距离^[17]；分词后的故障现象描述与候选日志内容之间的 2-gram 距离等特征。图 1 中的第二步即展示了无监督召回、打标签和特征工程后的训练数据示例。

接着，将含有特征列和标签列的数据送进二分类模型中，在本地训练 8 折交叉的 LightGBM 模型，在南方电网调度云平台上训练 GBDT 模型，

表 4 Word2vec 召回候选日志实际效果示例

故障名称	日志内容	故障与日志相似度
主干传输网 D 中调至 E 变光路异常	2018-12-08 16:10:13, 报主干传输网 D 中调至 E 变光路异常, 于 2018-12-08 16:10:17 启动故障处理流程	0.966 21
主干传输网 D 中调至 E 变光路异常	调度员于 2018-12-08 16:24:21, 对编号为 190606054 的故障进行了派工, 故障名称为: 主干传输网 D 中调至 E 变光路异常, D 中调、超高压 F 局负责处理故障	0.923 84
主干传输网 D 中调至 E 变光路异常	2019-04-29 18:40:12, 报主干传输网 D 中调 19 槽至 J 变 19 槽 10 Gbit/s 光路异常, 于 2019-04-29 18:45:15 启动故障处理	0.920 43



分别得到最终的二分类关联模型。由于无监督召回方案可达到 90%以上的召回率,因此理论上如果二分类模型拟合效果足够好,最终的关联日志召回率可达到无监督方案的召回率,但不会超过这个值。

4 实验结果与分析

4.1 实验数据集

本文进行智能关联方法的相关实验时,采用了电力通信运行管理系统中 2010—2019 年共 10 年的故障工单数据和值班日志数据,其中,故障工单数据共 1 291 条,值班日志数据共 6 632 条,数据所含具体信息见表 1 和表 2,10 年的数据中,故障工单与值班日志有关联的不同记录共有 5 384 条,均由人工进行关联。本文实验将划分 2010—2018 年的数据作为训练集,2019 年的数据作为测试集。

4.2 实验结果分析

为方便对实验结果进行对比与分析,本文引入召回率和精确率两个常见的指标,其中,召回率反映了正样本被预测正确数占原正样本的比例,针对本文具体场景,即:

$$\text{Recall} = \frac{TP}{P} \quad (1)$$

其中,TP 表示正样本被预测正确,也即关联正确的数量;P 表示原正样本,即原存在关联的样本数量。

精确率反映了正样本被预测正确的数量占被预测为正样本数的比例,针对本文具体场景,即:

$$\text{Precision} = \frac{TP}{P'} \quad (2)$$

其中,TP 表示正样本被预测正确,也即关联正确的数量;P'表示被预测为正样本的数量。

本地实验分数据预处理、无监督召回和监督分类 3 个步骤进行,首先对 10 年内的 1 291 条故障工单数据进行无监督召回,通过 TF-IDF、

Word2vec 和 BM25 3 种相似度衡量因子分别召回 38 730 条值班日志,经过合并去重后得到 68 077 条值班日志数据,其中故障工单与日志有关联的记录(Label 为 1)的数量为 4 923 条,即无监督方案召回率为 91.44%。

实际训练需要将 2010—2018 年无监督召回的 62 874 条记录与 2010—2018 年原本存在关联的 4 965 条记录(Label 为 1)进行合并去重,得到 63 309 条训练集,正负样本比为 4 965: 58 344=1:11.75。

本地实验首先采用 LightGBM 单模型进行训练,对 2019 年的数据进行测试,本地实验不同迭代次数下 LightGBM 单模型的测试性能如图 2 所示,横轴表示迭代次数,纵轴表示精确率或召回率。可以看到当迭代次数为 2 500 时,模型可以达到较好的测试性能。

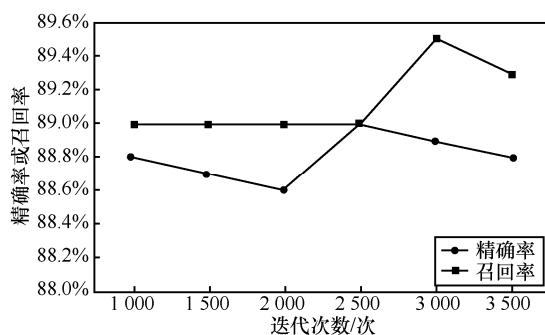


图2 本地实验不同迭代次数下模型测试性能

接着采用 n 折交叉验证的训练方法,即将训练集平均分为 n 份数据,训练 n 次模型,并对测试集进行 n 次测试,将 n 次测试结果进行平均得到最终的测试结果。 n 折交叉验证每次训练时,依次将均分的 $n-1$ 份数据作为训练集,剩下 1 份数据作为验证集。迭代次数为 2 500 时,不同交叉折数下的模型测试性能如图 3 所示,横轴表示模型交叉折数,纵轴表示精确率或召回率。从实验结果可以看到,从 2 折交叉到 8 折交叉,模型的测试性能,包括精确率和召回率都是逐步上升的,而 10 折交叉与 8 折交叉的测试性能基本一致,因

此, 最终选择性能最佳的 LightGBM 的 8 折交叉验证模型。

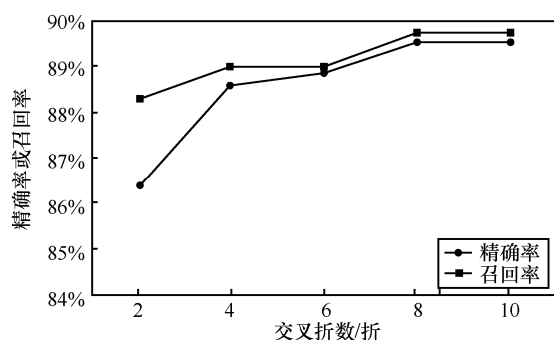


图3 本地实验不同交叉折数下模型测试性能

为了推进智能关联方法的实际应用, 在南方电网调度云平台开展了 GBDT 模型实验, 将本地处理完成的二分类特征数据同步至南方电网调度云平台的对象存储 OSS(object storage service)中, 映射到 ODPS(open data processing service)后在机器学习 PAI(platform of artificial intelligence)平台进行模型训练与预测, 采用 GBDT 二分类单模型训练, 树的数目设置为 400, 关联的精确率可达到 88.1%, 召回率达到 90.2%, 从预测结果可以看出, GBDT 二分类基本上可以达到与本地实验 LightGBM 同样的效果。

4.3 实验拓展

为了更加充分证明所提出智能关联方法的普适性, 本文拓展了电力通信运行管理中检修单与值班日志的关联实验, 实验选取了 2011—2013 年的检修单与值班日志数据, 其中检修单数据共 1 661 条, 值班日志数据共 3 863 条, 值班日志具体信息见表 1, 检修单具体信息见表 5, 主要包括

检修单号、检修内容、检修结束时间等信息, 表中数据已作脱敏处理, 其中变电站、电厂、换流站、供电局等名称均用大写字母代替。

检修单和值班日志已由人工进行关联, 本文实验将划分 2011—2012 年的数据作为训练集, 2013 年的数据作为测试集。

采用上述故障工单和值班日志智能关联实验中性能最佳的迭代次数为 2 500, LightGBM 的 8 折交叉验证模型, 对检修单与值班日志的智能关联进行实验, 同样分无监督召回和监督分类两个部分, 实验预测结果中, 关联的精确率可达到 94.3%, 召回率达到 83.4%, 此处召回率未达到 90% 是由检修日志的数据特性所决定的, 部分值班日志数据实际上有关联至检修单, 但日志内容格式相对统一, 可辨性低, 例如对于多数检修单, 都会有记录工作结束的日志, “值班员于 2013-12-04 08:21:04 跟进批复编号为 2013.5529 的检修票, 执行事件记录如下: 11 月 27 日 10:30 工作结束, 业务恢复正常。”单从数据文本内容来看, 无法精确关联至具体的检修单, 因此这部分日志利用文本处理方法较难精确召回, 后续可通过规则匹配检修单号的方法进行关联。整体上, 本文所提出的电力通信运行管理中典型业务数据的智能关联方法在应用至检修单与值班日志数据时, 同样具有较好的关联效果。

5 结束语

本文针对电力通信运行管理过程中产生的业务数据在统计上相对独立、关联性不强、人工关

表 5 检修单具体信息

检修单号	检修内容	检修结束时间
2011.3446	500 kV AB 甲线、500 kV AB 乙线专网应急通道进行路由调整 (配合 110 kV BC 线消缺工作)	2011-08-25 17:00:00
2012.4066	(1) -48 V 通信电源 A 蓄电池组核对性放电; (2) -48 V 通信电源 B 蓄电池组核对性放电	2012-11-27 22:00:00
2013.3804	E 开关站传输系统 2 Mbit/s 业务综合配线改造	2013-08-14 23:55:00



联低效的问题,以文本挖掘技术为基础,提出了典型业务数据的智能关联方法。本文选取了故障工单和值班日志两种典型的业务数据,设计了无监督召回和监督分类相结合的机器学习模型,最终达到了较高的精确率与召回率,实现了故障工单与值班日志数据的准确关联。

下一步,将根据需要继续完善南方电网调度云平台上的智能关联模型,同时与系统数据库互联,逐步实现数据的统一调管与模型的统一运维。另外,将电力通信运行管理中典型业务数据的智能关联方法通用化,以解决如在运行监控、资源管理以及资产管理中的信息关联问题。

参考文献:

- [1] DAN J, JAMES H M. Speech and language processing (3rd ed. draft)[EB].
- [2] 李生. 自然语言处理的研究与发展[J]. 燕山大学学报, 2013, 37(5): 377-384.
LI S. Research and development of natural language processing[J]. Journal of Yanshan University, 2013, 37(5): 377-384.
- [3] 王慧芳, 曹靖, 罗麟. 电力文本数据挖掘现状及挑战[J]. 浙江电力, 2019, 38(3): 1-7.
WANG H F, CAO J, LUO L. Current status and challenges of power text data mining[J]. Zhejiang Electric Power, 2019, 38(3): 1-7.
- [4] 邱剑. 电力中文文本数据挖掘技术及其在可靠性中的应用研究[D]. 杭州: 浙江大学, 2016.
QIU J. Research on power Chinese text data mining technology and reliability application[D]. Hangzhou: Zhejiang University, 2016.
- [5] 汪崔洋, 江全元, 唐雅洁, 等. 基于告警信号文本挖掘的电力调度故障诊断[J]. 电力自动化设备, 2019, 39(4): 126-132.
WANG C Y, JIANG Q Y, TANG Y J, et al. Fault diagnosis of power dispatching based on alarm signal text mining[J]. Electric Power Automation Equipment, 2019, 39(4): 126-132.
- [6] 刘梓权, 王慧芳, 曹靖, 等. 基于卷积神经网络的电力设备缺陷文本分类模型研究[J]. 电网技术, 2018, 42(2): 644-650.
LIU Z Q, WANG H F, CAO J, et al. A classification model of power equipment defect texts based on convolutional neural network[J]. Power System Technology, 2018, 42(2): 644-650.
- [7] 王春柳, 杨永辉, 邓霏, 等. 文本相似度计算方法研究综述[J]. 情报科学, 2019, 37(3): 158-168.
WANG C L, YANG Y H, DENG F, et al. A review of text similarity approaches[J]. Information Science, 2019, 37(3): 158-168.
- [8] 沈斌. 基于分词的中文文本相似度计算研究[D]. 天津: 天津财经大学, 2006.
SHEN B. Study on chinese text similarity computing based on word segmentation[D]. Tianjin: Tianjin University of Finance and Economics, 2006.
- [9] JONES K S. A statistical interpretation of term specificity and its application in retrieval[J]. Journal of Documentation, 1972, 28(1): 11-21.
- [10] SALTON G, YU C T. On the construction of effective vocabularies for information retrieval[C]//Proceedings of ACM SIGIR Forum. New York: ACM Press, 1973: 48-60.
- [11] MIKOLOV T, CORRADO G, CHEN K, et al. Efficient estimation of word representations in vector space[C]//Proceedings of the International Conference on Learning Representations (ICLR 2013). [S.l.:s.n.], 2013.
- [12] ROBERTSON S E, ZARAGOZA H. The probabilistic relevance framework: BM25 and beyond[J]. Foundations and Trends in Information Retrieval, 2009, 3(4): 333-389.
- [13] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. The Annals of Statistics, 2001, 29(5): 1189-1232.
- [14] KE G, MENG Q, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree[C]//Advances in Neural Information Processing Systems (NIPS). [S.l.:s.n.], 2017.
- [15] 曹小芹. 基于Python的中文结巴分词技术实现[J]. 信息与电脑, 2019(18): 38-39, 42.
CAO X Q. Technology implementation of Chinese jieba segmentation based on Python[J]. China Computer & Communication, 2019(18): 38-39, 42.
- [16] 陈正铭, 霍英. 编辑距离算法在中文文本相似度计算中的优化与实现[J]. 韶关学院学报, 2019, 36(12): 8-12.
CHEN Z M, HUO Y. Optimization and implementation of the edit distance algorithm in chinese similarity calculation[J]. Journal of Shaoguan University, 2019, 36(12): 8-12.
- [17] 余小军, 刘峰, 张春. 基于 N-Gram 文本特征提取的改进算法[J]. 现代计算机, 2012(23): 3-7.
YU X J, LIU F, ZHANG C. Improved text feature extraction algorithm based on N-Gram[J]. Modern Computer, 2012(23): 3-7.

[作者简介]



吴桂龙(1993—), 男, 中国南方电网电力调度控制中心工程师, 主要从事电力通信运行等工作。

杨志敏(1982—), 男, 博士, 中国南方电网电力调度控制中心高级工程师, 主要从事电力通信运行及技术支持系统研究等工作。

黄昱(1980—), 男, 中国南方电网电力调度控制中心高级工程师, 主要从事电力通信运行管理等工作。

《电信科学》第五届编辑委员会

主任委员:

韦乐平 中国电信集团有限公司

副主任委员:

朱 峰 中国通信学会

刘华鲁 人民邮电出版社有限公司

彭木根 北京邮电大学

编委: (按姓氏笔画排序)

山世光 中国科学院计算技术研究所

王 宇 中国科学院电子学研究所

王继业 国家电网有限公司信息通信部

方景龙 杭州电子科技大学

龙 腾 北京理工大学

毕 军 清华大学

卢光跃 西安邮电大学

曲 桦 西安交通大学

吕文俊 南京邮电大学

刘永祥 国防科技大学

刘建明 工业和信息化部产业发展促进中心

闫 江 北方工业大学

江 涛 华中科技大学

孙晓颖 吉林大学

阳小龙 北京科技大学

纪越峰 北京邮电大学

杜百川 国家广播电视总局科学技术委员会

李 丹 清华大学

李长海 中国工信出版传媒集团

李玉峰 上海大学

杨小康 上海交通大学

吴 巍 中国电子科技集团公司第五十四研究所

余少华 中国信息通信科技集团有限公司

沈连丰 东南大学

宋令阳 北京大学

张云勇 中国联合网络通信集团有限公司

张成良 中国电信集团有限公司

张同须 中国移动通信有限公司研究院

张宏莉 哈尔滨工业大学

陈 巍 清华大学

陈山枝 中国信息通信科技集团有限公司

陈芳炯 华南理工大学

陈铁明 浙江工业大学

陈章渊 北京大学

易东山 北京信通传媒有限责任公司

金 石 东南大学

周 涛 中国电信股份有限公司上海西区电信局

周一青 中国科学院计算技术研究所

周华春 北京交通大学

孟维晓 哈尔滨工业大学

赵军辉 华东交通大学

赵慧玲 工业和信息化部通信科学技术委员会

段晓东 中国移动通信有限公司研究院

高 鹏 中国移动通信集团设计院有限公司

高新波 重庆邮电大学

唐雄燕 中国联合网络通信有限公司研究院

曹卫平 桂林电子科技大学

曹蓟光 中国信息通信研究院

章坚武 杭州电子科技大学

梁海滨 北京信通传媒有限责任公司

董振江 高新兴科技集团股份有限公司

蒋 力 中国电信股份有限公司研究院

蒋林涛 中国信息通信研究院

程 方 南京信息工程大学

奚 笠 中国铁塔股份有限公司通信技术研究院

蔡 康 中国电信股份有限公司研究院

收录声明

本刊已被中国核心期刊遴选数据库(万方)、中国学术期刊数据库(知网)全文收录,被本刊录用的文稿,将一律由编辑部统一纳入中国核心期刊遴选数据库(万方)、中国学术期刊数据库(知网)。若作者不同意被收录,将在来稿时向本刊说明,本刊将做适当处理。若无特别声明,视同将录用文稿的电子网络版、汇编或合订本的使用权授予本期刊社,本刊支付的稿费中已包含上述费用。

The articles of *Telecommunications Science* have been embodied by WANFANG DATA and CNKI. The accepted articles will be unified into WANFANG DATA and CNKI. If the author does not agree to his or her article to be included, the author must illustrate when the article is submitted. The editorial department will properly handle. If the authors do not give a special announcement, the editorial department believes that the authors have granted the right of their articles to *Telecommunications Science*. The expenses have been included in the fee paid by editorial department.

道德声明

本刊发表的文稿是作者独立取得的原创性研究成果,无一稿多投;文稿内容不涉及国家机密;未曾以任何形式用任何文种在国内外公开发表过;文稿内容不侵犯他人著作权和其他权利。若发生一稿多投、侵权、泄密等问题,文稿作者将承担全部责任。

The authors of *Telecommunications Science* guarantee that their submitted articles are original and contain nothing confidential. The said article is only submitted to *Telecommunications Science*. The said article has not been published before and has not been submitted elsewhere for print or electronic publication consideration. The said article is in no way whatever a violation or an infringement of any existing copyright or license from the third party. Otherwise, the authors of the said article shall take the blame for the violation or infringement of the related copyright and the leakage of secrets.

电信科学

Telecommunications Science



邮发代号: 2-397 国外代号: M841 定价: 68.00元

ISSN 1000-0801

