

大数据在95598故障报修工单分析中的应用

郭小冰,黄小花,蒋海霞

(国网宁夏电力公司 电力科学研究院,银川 750002)

Application of large data in 95598 fault repair work order analysis

GUO Xiao-bing, HUANG Xiao-hua, JIANG Hai-xia

(Electric Power Research Institute, Sate Grid Ningxia Electric Power Company, Yinchuan 750002, China)

摘要:应用大数据分析技术,采取文本提取、分词等分析方法挖掘95598故障报修工单热点、难点和重点问题,展现客户诉求特征及变化规律,定位营销业务与工单之间的关联关系,查找内在原因,通过95598工单内容还原客户真实诉求。以大数据技术在宁夏95598故障报修工单分析中的应用为例,介绍分析方法,并依据分析结论提出相应的服务对策与措施,提出大数据分析技术在该应用上的改进方向。

关键词:大数据;故障报修;风险预警;热词;工单分析

Abstract: The application of large data analysis technology uses text extraction, hot words analysis methods to solve hot and difficult and key problems in 95598 fault repair service work order. It shows the customer demands characteristic and change rule, defines the relationship between marketing and work order to find the inherent reasons. Through 95598 work sheet content, customer's demands are found. With large data technology application in Ningxia 95598 fault repair work order analysis as an example, the article introduces the analysis method, and proposes the corresponding service according to the analysis conclusion countermeasures, the improvement direction in big data analysis application is put forward.

Key words: large data; power supply service; fault repair; risk pre-warning; hot words; work order analysis; fault reasons; power-off information

中图分类号:F407.61 文献标志码:C

1 故障报修工单分析研究的意义

用电业务的发展离不开数据的支撑,随着营销信息化工作的全面快速推动,客户用电基础信息不断完善,用电信息采集范围和采集成功率逐步扩大提高,用电信息采集数据、电费数据、95598数据等营销业务数据向海量规模发展,大数据特征日益明显,合理开发利用这些海量数据,可以为管理者提供明确的数据依据支撑。

在故障报修直派流程背景下,现有的故障服务热点分散在不同类型的95598工单中,不能集中反映当前供电服务的热点问题,同时95598工单依赖人工统计分析,服务热点问题的发现与处理存在较大的延迟与滞后,且人工统计分析不够全面,看待问题的观点具有局限性。本次分析研究将开展故障报修工单所引发的供电服务风险分析和预测,挖掘客户故障报修服务诉求与热点,实现故障报修服务的事前管控。

2 故障报修工单分析研究思路

故障报修分析以宁夏电力公司2014—2015年故障报修工单为基础数据,根据工单数据中显示的工单内容进行多维度组合分析。

本次故障数据均为有效数据,由于数据中缺失客户联系电话数据,故“重复申告客户分析”和“故障与其他工单关联关系分析”无法开展,其它分析均可实现。

故障报修分析具体思路如图1所示。

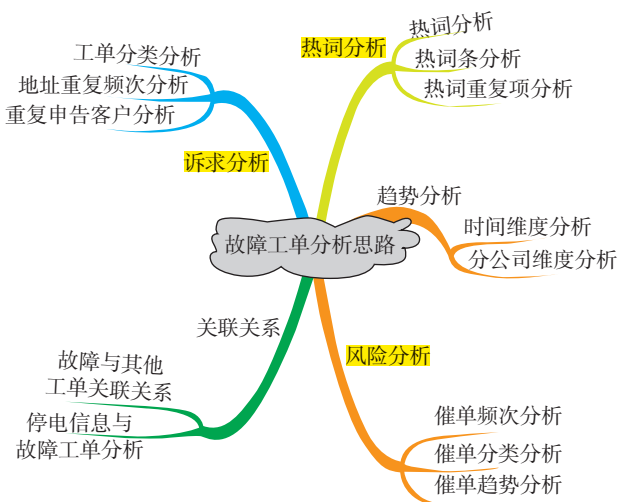


图1 故障报修分析思路

3 故障报修工单分析研究主题

3.1 故障分类分析

抽取2014—2015年故障报修工单总数据,按整

收稿日期:2016-07-01

体故障工单分类、地市供电公司(精确到县)故障类型工单分类识别宁夏公电力司故障具体类型。

在原有故障分类的基础上,通过对受理文本内容的识别,根据客户的描述将故障报修分类进行细化,得到图2详细分类数据。

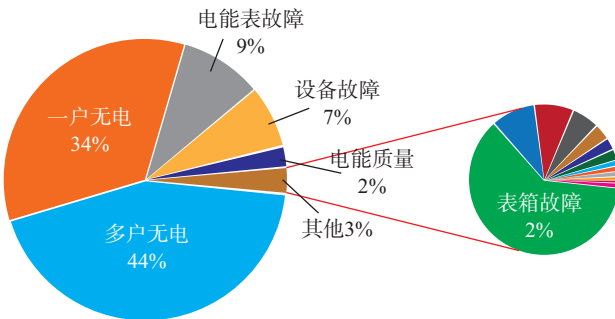


图2 故障报修工单类别识别

3.2 故障趋势分析

抽取2014—2015年故障报修工单总数据,清理后按照日、月、周、季、年、时段、地市维度,进行趋势分析。从走势图了解全年故障高发期和低谷期,快速识别故障工单发展趋势,掌握全年故障发生特征。如图3。

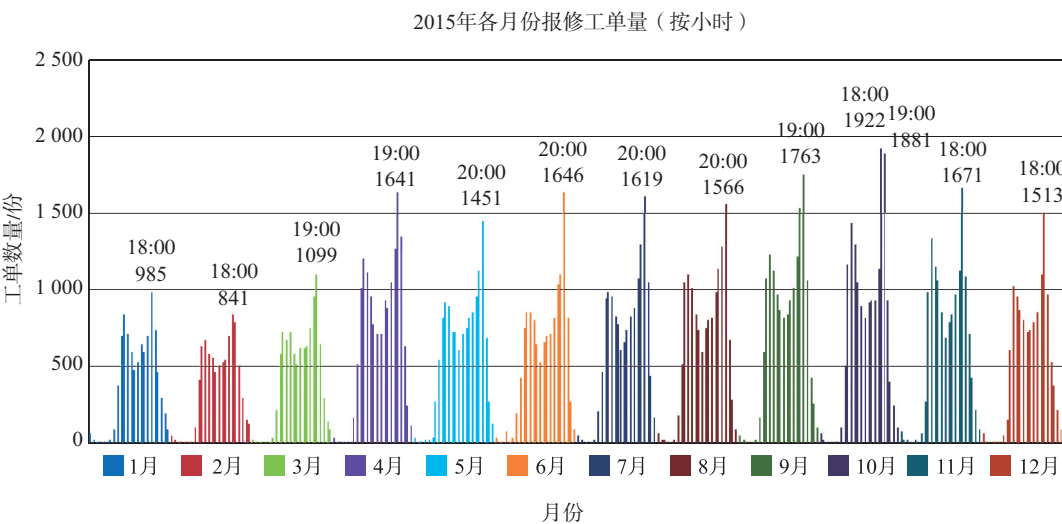


图3 故障报修时段趋势数据

趋势分析在原有日、月、年常规分析基础上,对故障报修受理工单的时段进行挖掘分析,查找出不同月份各时间段故障报修业务趋势。

3.3 故障热词分析

抽取2014—2015年故障报修工单受理内容和处理内容两大文本,利用R分词工具分别找出故障热词,再将热词代入原文本中与宁夏95598业务专家确认热词条,最后分析热词及热词条词频,掌握客户故障报修诉求。

热词分析在于利用“R分词”技术,将工单的受理文本进行词组划分,从划分出的词组出现频次来识别客户故障报修的热点问题。如表1。

表1 故障报修热词分析数据

序号	2015年故障受理热词条重复项	出现频率/条	占比/%	热词数量/个
1	“多户”“实习”	11 603	7.25	2
2	“一户”“无法判断” “资产归属”	2 983	1.86	3
3	“低压居民”“一户” “无法判断” “资产归属”	28 905	18.07	4
4	“低压居民” “一户”“实习” “无法判断” “资产归属”	6 172	3.86	5
5	“低压居民”“一户” “实习”“电表”“无法判断” “资产归属”	9	0.01	6
合计		49 672	31.05	

3.4 故障关联分析

抽取2014—2015年故障报修工单、故障催单、停电信息数据,分析故障重复频次和催单频次,再将停电信息数据与故障工单进行匹配,分析故障与投诉的关联、与催单的关联。最终识别出故障报修风险点。

通过将“计划停电”数据和“故障报修”数据进行关联分析,快速分析出“计划停电”客户在故障报修工单中占比。如图4所示。

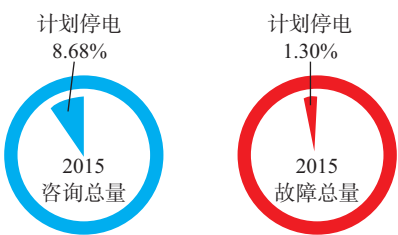


图4 故障报修与计划停电的关系

4 故障报修工单分析运用技术

采用大数据分析研究相关常用技术手段,整合 95598 业务支持系统中的报修业务工单信息,并结合工单受理文本信息、处理内容文本信息、停电信息等多维度数据,通过大数据分析技术对其进行预处理、清洗,并利用数据挖掘聚类、分类、关联规则等算法工具,围绕分析主题构建分析模型,完成数据的深度挖掘分析,并对结果进行可视化呈现。

4.1 文本获取

将故障报修业务工单进行提取,保留工单受理文本信息和处理文本信息,形成文本数据集。

4.2 文本处理

使用 R 分词工具将工单中的受理文本信息和处理文本信息切割成单个词汇,将文本转化为了数学语言。

在分词之前,首先结合电力业务对分词词库进行补充和完善,比如:词汇“电量异常”,在中文分词工具包自带的词库中将该词分为“电量”和“异常”2 个词,但是在 95598 工单中“电量”和“异常”2 个词在很多情况下是同时出现在受理内容和处理情况中的,因此需要将自定义的行业专有词加入热词库。

进行分词时,分词工具会优先将“电量异常”识别为一个词,从而增加分词结果的准确率。同时由于不同记录人员在语言习惯上存在差异,因此在对同一问题的表述上可能呈现多种描述,因此在分词后,词汇中存在一定的同义词和近义词。

4.3 热词提取

剔除对分类无贡献的词汇,分拣出工单中的高频高指向性词,并提取主要特征词(如:专业名词、特殊词汇)。最终将提取出的热词与宁夏 95598 业务专家进行研讨确认,形成宁夏 95598 热词库。

5 客户诉求分析

5.1 聚类分析

运用 K-最近邻参照聚类法对工单进行聚类分析,聚类结果反映了客户诉求的类别。通过聚类分析首先可以发现与某工单相似的一批工单,发现需求特征等;其次,将一群工单聚类成若干个类;再次,生成分类器对文档进行分类,并以时间、地市等纬度展示分析结果。

5.2 分类分析

对工单样本进行分类分析,遍历一个分类体系来寻找需要的信息,达到发现问题的目的,并以时间、地市等纬度展示分析结果。

5.3 热点分析

运用柏拉图对客户诉求分类出现的次数从高到低排序,找出所有诉求中的热点内容。

5.4 分类关联分析

产生工单的原因多种多样,其中营销业务引发的工单情况较为常见,因此分析工单关联,即从热点事件工单出发,运用关联规则算法分析需求之间的关联关系(如:故障报修、咨询查询可能会向投诉业务转化)。一是分析基于文本挖掘得到的客户需求之间的关联关系;二是基于工单现有人工分类,分析客户需求之间的关联关系,并以时间、地市等纬度展示分析结果。

5.5 内在因素关联分析

运用关联规则的算法,分析故障报修、投诉举报等工单中热词的关联关系,了解客户各类诉求的内在因素。

6 结束语

故障报修工单的深入分析与研究拟从“客户诉求”出发,打破原有数据分析方法,深入工单受理内容,挖掘客户的真实诉求与申报原因,展开所引发的供电服务风险分析和预测,挖掘客户故障报修服务诉求与热点,以实现故障报修服务的事前管控和事中追踪,从而有效缓解故障报修服务压力,扩展服务深度,提升服务精度,不断强化 95598 客户服务中心的服务管理水平。

参考文献:

- [1] 中国电机工程学会信息化专委会. 中国电力大数据发展白皮书[Z]. 2013.
- [2] 李贵兵,罗洪. 大数据下的智能数据分析技术研究[J]. 科技资讯,2013(30):11-12.
- [3] 孙柏林. 大数据技术及其在电力行业中的应用[J]. 电气时代,2013(8):18-23.
- [4] 戴镭. 电力营销服务质量管理体系要求及工作流程和具体实现过程的探究[J]. 现代营销,2012(8):61.
- [5] (英)维克托·迈尔-舍恩伯格. 大数据时代——生活、工作与思维的大变革[M]. 盛杨燕,周涛,译. 杭州:浙江人民出版社,2012.

(本栏责任编辑 徐文红)