

深度学习技术在电力客户诉求自动分类业务中的应用

潘庆锋 陈玉娴 陈泽基

广东电网公司佛山供电局 广东佛山 528000

摘要: 电力客户诉求是反映供电企业服务质量的重要数据。对客户诉求进行问题类型分类,有利于准确把握供电服务的短板,有针对性地制定服务改进措施,从而提升服务质量。为解决人工识别为类型存在的效率低、易出错等问题,我们研究基于多层神经网络模型结合深度学习算法研究建立一套电力客户诉求问题类型识别的算法,实现客户诉求的问题类型在线自动识别。

关键词: 多层神经网络;深度学习;信息分类;客户诉求;自动识别;词干

Application of deep learning technology in automatic classification of power customers' demands

Pan Qingfeng, Chen Yuxian, Chen Zeji

Guangdong Power Grid Company Foshan Power Supply Bureau Guangdong Foshan 528000

Abstract: Power customer demand is important data reflecting the service quality of power supply enterprises. Classifying the problem types of customers' demands is conducive to accurately understanding the weaknesses of power supply service and formulating targeted service improvement measures to improve service quality. It does not solve the problems of low efficiency and error-prone which exist in manual identification as type. Based on the multi-layer neural network model combined with a deep learning algorithm, we study to establish a set of power customer demand problem type identification algorithms to achieve the automatic online identification of customer demand problem types.

Keywords: Multilayer neural network; Deep learning; Information classification; Customer appeal; Automatic identification; stems

引言

电力客户在消费电能的过程中,会提出各种用电问题,就形成了电力客户的用电诉求。电力客户用电诉求是反映供电企业服务质量的重要数据。对客户诉求进行问题类型分类和统计,有利于准确把握供电服务的短板,有针对性地制定服务措施有效提升供电服务质量。

目前,电力客户的诉求都会集中到电力服务热线95598座席受理并进行初步人工问题分类。由于多种因素影响,人工进行问题分类存在效率低和易出错等问题。

随着计算机神经网络技术的发展,利用计算机进行信息分类深度学习,形成稳定可靠的模型并应用于自动分类信息,已经得到成熟的应用。为此,本项目将研究建立电力客户诉求的问题类型自动识别的多层神经网络模型,并应用已经人工标注了问题类型的近30万宗历史诉求工单数据,对模型进行深度学习训练,并应用模型进行在线的、自动的诉求问题类型识别,从而减轻服务调度班组的工作压力,提升工作效率。

一、诉求问题类型识别需求分析

用电问题通常可归纳为若干类,以佛山供电局为例,

服务调度班组把客户用电问题归纳为13类,即:

问题类型编码	问题类型
0	变更类(杂项)业务
1	抄核收
2	电网建设/改造
3	电压质量
4	服务渠道
5	供电安全
6	故障停电
7	计量
8	新能源服务
9	业扩办理
10	有序用电
11	预安排停电
12	其他

表1 诉求问题类型表

目前,诉求的问题类型识别还需要人工识别标注。这里存在几个问题:

1.1 由于客户表达诉求不准确,可能导致人工标注的诉求问题类型不正确;

1.2 由于业务人员工作经验有差异,接受的培训程度不同,对一些问题的理解以及判断标准不完全一致,

也导致一些诉求的问题类型人工识别错误；

1.3 有些诉求的内容可能涉及不止一种问题类型，人工进行识别时容易混淆导致错误。

1.4 每次人工识别过程，需要认证阅读“来电内容”和“处理意见”的内容，人工识别的问题类型的效率比不高。

电力客户诉求的问题类型识别过程，其实是信息分类的过程。我们可以采用基于自然语言理解、多层神经网络模型和深度学习技术，构造和训练一个信息分类模型，能自动对新的客户诉求进行理解并自动识别诉求的问题类型。^[1]

二、技术可行性分析

由于电力客户诉求的问题类型识别，没有固定的规则，我们并不能单纯根据诉求内容中包含了哪几个关键字，就能准确判定该诉求的问题类型，所以无法将信息分类规则通过固化成程序代码的方式来实现诉求问题类型的识别。

2.1 提取词干

诉求问题类型由诉求的内容决定，而诉求内容由诉求内容文本的词组决定；

词组有有用的词组，也有无用的词组，我们将无用的词组清洗掉，剩下有用的词组；

通过对问题类型所包含的有用词组的出现次数和频率进行统计，可以得到每种问题分类哪些词组更重要，这些词组就是每种问题类型的关键词组；^[2]

全部问题类型的关键字的并集，就是用于全部诉求问题类型识别的中心词，叫做词干。也就是说，无论什么问题类型的诉求，都是以该诉求所含的词干的值作为识别其问题类型的依据。

这就是从大量的历史诉求工单中提取问题分类词干的过程。

2.2 确定词干值

确定了词干后，词干的值如何确定？词干的值，可以是词干所在诉求中出现的次数或频率（频率为次数/该诉求有用词组数）。

2.3 生成诉求图谱

接下来，我们把所有诉求的词干值与诉求的问题类型对应起来，建立一个二维矩阵。它的每一行，表示一宗诉求的各个词干的值，每一列表示一个词干在不同诉求中出现的次数。这个矩阵已基本展示了词干在不同问题类型的诉求中出现的规律，如果增加词干的数量，将更加清晰描述这个问题类型诉求里词干的分布情况。这个二维矩阵叫做“诉求图谱”，

2.4 建立和训练模型

如果我们创建一个多层神经网络原始模型，并且使用诉求图谱的数据来训练模型，让模型通过深度学习拟合模型参数，抽象出词干分布与诉求问题类型的非线性关系，那么这个训练后的成熟模型就可以用来预测（识

别）新诉求的问题类型。^[3]

也就是说，如果新增诉求的问题类型是未知的话，我们只要生成新诉求的诉求图谱，然后把新的图谱输入成熟的模型，就能输出新诉求的问题类型，从而实现对新增诉求的问题类型进行识别。

为此创建一个多层感知机，其输入层的神经元数量与词干数量相同，输出层神经元数量为问题类型数量并通过 softmax 激活函数输出分类值，中间则设置多层全连接层，并在各层选择合适的激活函数，形成一个完整的多层神经网络。

然后，使用近 30 万宗历史诉求的诉求图谱对原始模型进行训练，输出一个稳定的、成熟的多层神经网络模型。

2.5 自动识别新诉求的问题类型

当业务系统有新的诉求工单时产生时，获取新诉求的诉求内容文本，计算词干的值，生成新诉求的图谱，并输入成熟的多层神经网络模型，让成熟模型输出新诉求的问题类型值。^[4]

三、算法设计

算法流程图如下图所示。整个算法分为两部分，左半部分，是模型训练的流程；右半部分是在线自动识别新增客户诉求问题类型的流程。

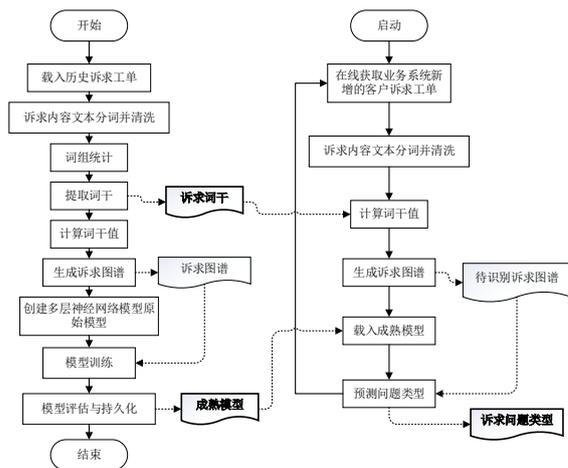


图 1. 算法流程图

模型训练的部分，主要从大量历史诉求中提取并输出客户诉求词干，以及输出训练后的成熟模型；新增客户诉求自动问题类型识别的部分，主要输出新诉求的问题类型。

3.1 模型训练流程：

3.1.1 载入历史诉求工单：此模块近几年佛山供电局辖下客户的 30 万宗用电诉求工单载入系统。

3.1.2 诉求内容文本分词并清洗：对每宗诉求的来电内容字段和处理意见字段的文本内容，进行分词运算。并使用停用词库将诉求中无意义的词组过滤掉，如常用的助词、连词、介词、称呼词等。经过这些处理后，剩下的词叫做“有用词”。

3.1.3 词组统计：因为在各类诉求工单中，出现次数多或者出现频率高的词，是对该诉求分类比较重要的词组，因此，我们需要从诉求的问题类型、诉求工单等不同纬度，对所有有用词组进行词组出现次数和出现频率的统计。

3.1.4 提取词干：根据词组统计的结果，按照各问题类型所包含的词组的出现次数和频率，确定该问题分类的关键词，然后各问题类型的关键词集合，组成全部诉求的词干。词干作为诉求的中心词，或者说是描述每宗诉求特征的词组。

3.1.5 计算诉求词干值：确定诉求的词干后，我们还需要计算每宗诉求中这些词干的值，即词干在每宗诉求中出现的次数和频率。

3.1.6 生成诉求图谱：把所有历史诉求的词干值组成一个二维矩阵，就形成了历史诉求图谱。

创建多层神经网络模型原始模型：根据词干的数量，创建相应维度的输入层、根据问题类型的数量（13项）创建含13个输出项的softmax函数输出层、以及多个隐含的全连接多层神经网络原始模型，为每一层设置优化函数、损失函数以及测量方法。

3.1.7 模型训练：将诉求图谱作为训练数据集、已标注的诉求问题类型作为相应的训练标签（Label），输入到原始进行深度学习训练。

3.1.8 模型评估与持久化：模型训练完毕，需进行评估。使用验证数据集和相应的验证标签输入经过训练的模型，已检验模型对未知数据进行预测的准确率。如果验证得出的准确率满足业务标准则认为模型是成熟的模型，将此模型固化。若准确率未达标准，则需要重新调整模型的结构、参数，并重新训练。

3.2 自动识别问题类型：

3.2.1 在线获取业务系统新增的客户诉求工单：每隔一定时间（例如10分钟），查询业务系统有无新的客户诉求工单，如果有，则提取这些新增诉求的工单信息。

3.2.2 诉求内容文本分词并清洗：与模型训练流程的“诉求内容文本分词并清洗”环节一样，按照相同的规则对每宗诉求的来电内容字段和处理意见字段的文本内容，进行分词运算，并过滤掉停用词。

3.2.3 计算词干值：根据模型训练流程提取的词干，计算新增诉求的词干值。

3.2.4 生成诉求图谱：将新增诉求的词干值组成诉求图谱矩阵，生成待识别问题类型的诉求图谱。

3.2.5 载入成熟模型：把模型训练后固化的问题类型识别成熟模型，载入系统。

3.2.6 预测问题类型：将待识别问题类型的诉求图谱作为输入数据，应用成熟模型，对新增的诉求的问题类型进行识别，识别结果输出，然后循环继续。

问题类型自动识别流程小结：这部分流程主要在线获得新增的客户诉求工单，自动计算和生成待识别问题类型的诉求图谱，然后应用以训练的成熟模型自动识别问题类型并输出，从而实现了在线的、自动的、准实时的客户新增诉求问题类型识别。

三、应用情况

本算法成果——《佛山供电局客户诉求问题类型自动识别软件》，已于2021年11月份开始试运行，并在线准实时自动对2022年1月份至今（7月份）的电力客户诉求工单共75330宗进行了问题类型自动识别，总体识别准确率约为92.7%，其中新能源服务、预安排停电类问题的识别准确率超过99%；故障停电、抄核收类工单的识别准确率超过96.5%；电压质量、业扩办理和变更类（杂项）类问题识别率都超过90%。成果的应用，推动客户服务领域能更准确地瞄准服务短板，在开展业务分析、指导服务提升等应用中发挥了积极作用。

四、结束语

本文介绍了在大数据、多层神经网络、自然语言处理等技术支撑下，电力客户诉求问题类型自动识别的算法设计以及实现方式，以及初步应用情况。由于实现了电力客户诉求在线的、自动的、准实时的问题类型识别，节省了服务调度人员、服务分析人员的人工分类时间，减少了工作量，提高了工作效率。

今后，我们还将加大电力营销领域、电力服务领域的大数据价值挖掘，进一步深入研究把成熟的数字化技术应用到更多的业务场景，使数字化更全面、更深入渗透到日常业务处理、业务分析工作中，提升业务智能化、自动化处理水平，助力提升业务效率，减少业务差错，为客户提供更加优质高效的供电服务。

参考文献：

[1] 作者：[日]太田满久 须藤广大 黑泽匠雅 小田大辅，《TensorFlow 开发入门》，出版社：水利水电出版社，国际标准书号 ISBN：9787517092803

[2] 作者：郑泽宇，《TensorFlow: 实战 Google 深度学习框架》，出版社：电子工业出版社，国际标准书号 ISBN：9787121330667

[3] 作者：何晗，《自然语言处理入门》，ISBN 编号：9787115519764

[4] 作者：涂铭、刘祥、刘树春，《Python 自然语言处理实战：核心技术与算法》，ISBN：9787111597674