12345市民热线智能工单系统

结项文档

2020年07月

目录

[1 文档说明 4](#_Toc46388780)

[1.1 编制说明 4](#_Toc46388781)

[1.2 适用范围 4](#_Toc46388782)

[1.3 术语定义 4](#_Toc46388783)

[1.4 参考资料 5](#_Toc46388784)

[2 需求描述 5](#_Toc46388785)

[2.1 建设需求 5](#_Toc46388786)

[2.2 数据要求 6](#_Toc46388787)

[2.3 交互协议 7](#_Toc46388788)

[2.4 运行环境 7](#_Toc46388789)

[2.4.1 软件说明 7](#_Toc46388790)

[2.4.2 硬件说明 8](#_Toc46388791)

[3 功能实现详细说明 8](#_Toc46388792)

[3.1 算法详细说明 8](#_Toc46388793)

[3.1.1 特征工程选型及结论 8](#_Toc46388794)

[3.1.2 算法对比及选型 11](#_Toc46388795)

[3.2 智能工单分类流程描述 11](#_Toc46388796)

[3.2.1 模型算法介绍 11](#_Toc46388797)

[3.2.2 模型训练 12](#_Toc46388798)

[3.2.3 模型测试 15](#_Toc46388799)

[3.3 智能工单派发流程描述 16](#_Toc46388800)

[3.3.1 模型算法介绍 16](#_Toc46388801)

[3.3.2 模型训练 19](#_Toc46388802)

[3.3.3 模型测试 20](#_Toc46388803)

[4 生产环境准确率测试结果 20](#_Toc46388804)

[4.1 智能工单分类 20](#_Toc46388805)

[4.2 智能工单派发 21](#_Toc46388806)

# 文档说明

## 编制说明

本文档用于对项目建设过程中采用的AI能力，所涉及到AI能力的需求范围、数据预处理、采用算法、算法评估方法和准确率指标等技术评估规范。

## 适用范围

## 术语定义

* **特征提取**：机器学习算法往往无法直接处理文本数据，文字必须转换成可以量化的特征向量。常用的方法为：One-Hot、n-gram、TFIDF、word2vec。
* **机器学习：**机器学习（MachineLearning）是一门涉及统计学、系统辨识、逼近理论、神经网络、优化理论、计算机科学、脑科学等诸多领域的交叉学科，研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能，是人工智能技术的核心。
* **深度学习：**建立深层结构模型的学习方法，典型的深度学习算法包括深度置信网络、卷积神经网络、受限玻尔兹曼机和循环神经网络等。
* **知识图谱：**知识图谱本质上是结构化的语义知识库，是一种由节点和边组成的图数据结构，以符号形式描述物理世界中的概念及其相互关系，其基本组成单位是“实体—关系—实体”三元组，以及实体及其相关“属性—值”对。
* **自然语言处理：**是计算机科学领域与人工智能领域中的一个重要方向，研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法，涉及的领域较多，主要包括机器翻译、机器阅读理解和问答系统等。
* **人机交互：**主要研究人和计算机之间的信息交换，主要包括人到计算机和计算机到人的两部分信息交换，是人工智能领域的重要的外围技术。

## 参考资料

# 需求描述

## 建设需求

上海12345市民服务热线是直接面向市民的综合政务服务一线窗口，代表了市热线办、全市140多家市政管理单位及公共事业机构。市民热线话务员良好的服务质量，直接向市民展现了政府的服务形象，体现了上海政务服务高品质。

12345市民热线自2012年开通运行以来，主体业务是受理市民来电，完成市民诉求的受理、转派、回访，与市民进行沟通互动的主要渠道是传统语音，在服务渠道方面，提供了电话语音、传真方式，受理市民诉求。围绕着为市民诉求解答、受理、反馈的闭环业务，热线中心形成了一套成熟的业务支撑平台。

目前，12345热线业务系统已建成话务系统、IVR系统、坐席系统、大屏幕系统、诉求单系统、知识库系统、考试培训系统、质检系统、统计报表系统、热线服务运营管理支撑系统等一系列的系统工具，为业务运营提供了一套功能全面、性能稳定的业务支撑系统平台。

为进一步提升服务渠道“运营智慧化”的服务能力，希望能够通过智能化手段，实现市民来访需求的快速记录，工单内容的智能分类及派发，关键热点问题的统计分析，市民情绪识别，以及工单处理的质量检测。达到加快现场话务员服务响应能效，提升整体服务质量与市民感知，提高热线整体运营数据分析的智能化程度，优化管理的效果。

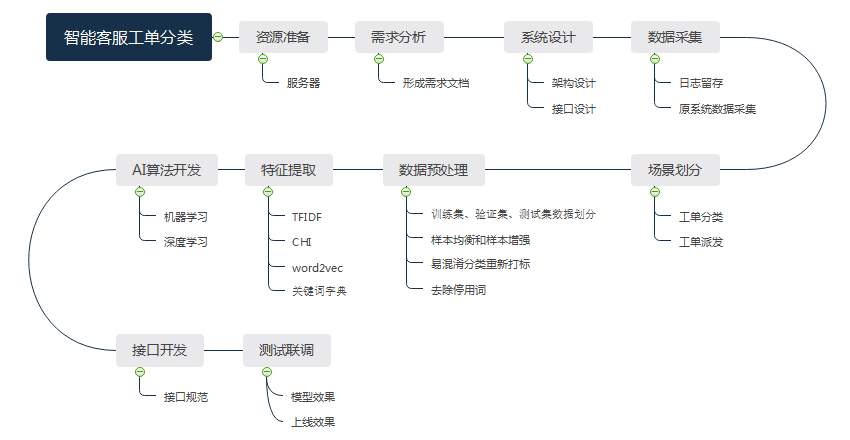
热线主要智能化应用规划包括基于语音识别、机器学习、深度学习、知识图谱、情绪识别以及智能预质检方法实现对市民通话表达诉求的快速识别、诉求工单的智能化生成及处理、实时服务质量监督、智能化运营管理分析。同时后续考虑开展基于智能问答的自助人机交互功能，将人工语音咨询逐步向自助咨询引流，提升服务效率和服务感知。

12345市民服务热线目前日均呼入量为25000通，而接通率为92%。电话诉求日均受理量已达到20000+件，话务员每日人均受理量达到60件。随着12345对市民服务定位的不断提升；对诉求受理范围与处理效率的不断提高要求，话务员的服务受理压力越来越大。

目前话务员在每次受理市民诉求时，需要花费时间去理解和分析整理市民电话口述过程中所提出的诉求内容，并需要大量时间去进行工单填写、分类和承办单位的转派协调。这部分工作非常耗费时间人力，导致服务效率不高。

经过已有的工单智能化经验，可以将工单智能化问题转化成自然语言处理问题，如关键信息抽取、文本分类、工单派发等。工单智能化需要解决的共有问题主要是：工单智能化问题采用文本分类、样本增强等逐步演进的方案逐步在12345客服落地AI能力，提升12345客服服务效率。

智能客服工单分类的流程如下图所示：



## 数据要求

以智能工单分类为例：

输入数据：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **参数名称** | **参数ID** | **说明** | **是否可空** | **备注** |
| 工单ID | wpid | String | 否 | 工单ID |
| 第一级分类 | class1 | String | 否 | 四级分类的第一级分类 |
| 第一级分类ID | class1Id | String | 否 | 四级分类的第一级分类ID |
| 第二级分类 | class2 | String | 否 | 四级分类的第二级分类 |
| 第二级分类ID | class2Id | String | 否 | 四级分类的第二级分类ID |
| 第三级分类 | class3 | String | 否 | 四级分类的第三级分类 |
| 第三级分类ID | class3Id | String | 否 | 四级分类的第三级分类ID |
| 第四级分类 | class4 | String | 否 | 四级分类的第四级分类 |
| 第四级分类ID | class4Id | String | 否 | 四级分类的第四级分类ID |
| 工单内容 | content | String | 否 | 客服总结的通话内容 |
| 电话ID | callId | String | 否 | 来电通话的ID |

输入样例：

{

"appID":"授权ID ",

"appKey":"授权密钥",

"content":"文本内容",

"num":"结果Top1" ,

"callId":"123456-123 "

}

输出数据：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **参数名称** | **参数ID** | **说明** | **是否可空** | **备注** |
| 第一级分类ID | class1Id | String | 否 | 四级分类的第一级分类ID |
| 第二级分类ID | class2Id | String | 否 | 四级分类的第二级分类ID |
| 第三级分类ID | class3Id | String | 否 | 四级分类的第三级分类ID |
| 第四级分类ID | class4Id | String | 否 | 四级分类的第四级分类ID |

出参样例：

{

"data":{

"result":

[

{"class1Id":"1","class2Id":"2","class3Id":"3","class4Id":"4"},

],

"status":1

},

"errmsg": "请求成功",

"errcode": "1000"

}

## 交互协议

相关功能接口开发完成后需要和电子工单进行联调测试相关功能的完备性，可通过http/https协议发送工单信息至平台，自动化平台接收数据并解析相关字段后存入后端数据库，开发人员调用待测试的相关功能接口，需求方验证新增或者优化的相关功能点是否满足需求。

## 运行环境

### 软件说明

* 操作系统：Windows server2012/python3.6
* 依赖包：jieba==0.39

pandas==0.23.4

Flask==0.11.1

Flask-Cors==2.1.2

jsonschema== 2.5.1

numpy==1.18.1

xlrd==1.0.0

joblib==0.14.1

xgboost==0.90

scikit-learn==0.20.0

### 硬件说明

服务器硬件：CPU20核，内存256GB，硬盘6TB；

# 功能实现详细说明

## 算法详细说明

### 特征工程选型及结论

本项目在进行NLP特征工程时，调研了以下四种算法。

#### ****1）One-hot****

**简介**:

One-hot encoding, 即one-hot编码，又称独热编码，使用n位状态寄存器来对n个状态进⾏行行编码，每个状态都由它自⼰独立的寄存器位，且在任意时候只有⼀一位有效。one-hot编码后是一个⼆二进制向量，采⽤0和1两个数字，将分类情况映射到这两个整数上。在一个向量中，所有元素都是0，除了了某个元素为1，该元素的索引即为编码样本对应的分类值（情况）。

**优势:**

1. one-hot编码对数据的分类更精确

2. 每个样本只对应于一个类别（即只在对应的特征处值为1，其余地方值为0），对于多分类的问题，我们的分类结果往往是隶属于某个类别的概率，这样在进行损失函数（例如交叉熵损失）或准确率计算时，变得非常方便。

3. one-hot编码对特征进了扩充

4. 经过编码后，从⼀一个权重变为多个权重，提⾼高了了模型的⾮线性能⼒5. 不不需要多参数进⾏归一化处理

6. 随着将⼤权重拆分成⼏个⼩权重管理理特征，降低了异常值对模型的影响，增加了模型稳定性

7. 计算⽅便快捷，表达能⼒强

**劣势:**

1. 生成了了较大的稀疏矩阵，对于复杂的对象，样本的特征向量维数很大

2. 没有考虑元素（词）之间的顺序和相互之间的关系

#### ****2）TF-IDF****

**简介**:

TF-IDF(Term Frequency-Inverse Document Frequency, 词频-拟文件频率) 是一种⽤用于资讯检索与资讯探勘的常用加权技术。TF-IDF是一种统计方法，用以评估一字/词对于一个文件集或一个语料库中的其中一份文件的重要程度。字词的重要性随着它在⽂文件中出现的次数成正比增加，但同时会随着它在语料库中出现的频率成反比下降，i.e.，一个词语在一篇文章中出现次数越多, 同时在所有⽂文档中出现次数越少,越能够代表该文章。

**优势:**

1. 实现简单快速，容易理解

2. 能有效避免常用词对关键词的影响，提高了关键词与文章之间的相关性。

**劣势:**

1. 仅以“词频”度量词的重要性，后续构成文档的特征值序列，词之间各自独立，无法反映序列信息

2. 易受数据集偏斜的影响，如某一类别的文档偏多，会导致IDF低估。IDF的简单结构并不能使提取的关键词十分有效地反映单词的重要程度和特征词的分布情况，使其无法很好地完成对权值调整的功能。尤其是在同类语料库中，这一方法有很大弊端，往往一些同类文本的关键词被掩盖。例如：语料库D中教育类文章偏多，而文本j是⼀一篇属于教育类的文章，那么教育类相关的词语的IDF值将会偏小，使提取文本关键词的召回率更更低

3. 没有考虑类内、类间分布偏差（被用于特征选择时）。比如只有2类数据，文档总数200，类1类2各100个文档；term1只出现在类1的所有100个文档，在类1出现总次数500；term2在类1出现次数也是500，但是类1和类2各有50个⽂文档出现term2；此时对类1，计算两个term得到的TF-IDF结果是一样的；无法反映term1对类1的重要性

#### ****3）Word2vector****

**简介**:

使用one-hot编码忽略了词与词之间的关系，且维度太高的向量作为模型输入会导致模型变的复杂。word embedding 是一种更加合理的对词的向量表示，将高维稀疏的向量降维成稠密的低维向量。（一种高维到低维的映射），而word2vector就是一种高效实现word embedding 的工具。

**优势:**

1.Hierarchical Softmax 去掉了隐藏层，输出层改用Huffman树，大大减少了集中在隐藏层和输出层之间的矩阵向量运算和输出层上的softmax归一化运算

2.邻近的词之间有很强的关联性

3.考虑到语义，语义相似的词关联性更强

**劣势:**

无法解决多义词的问题，因为每一个词只有用一个向量表示。

#### ****4）CHI****

**简介**:

卡方检验最基本的思想就是通过观察实际值与理论值的偏差来确定理论的正确与否。CHI特征选择算法利用了统计学中的"假设检验"的基本思想：首先假设特征词与类别直接是不相关的，如果利用CHI分布计算出的检验值偏离阈值越大，那么更有信心否定原假设，接受原假设的备则假设：特征词与类别有着很高的关联度。

**优势:**

**CHI特征选择方法，综合考虑文档频率与类别比例两个因素。**

**劣势:**

只统计文档是否出现词，而不管出现了几次，夸大了低频词的作用。

通过调研对比以上四个特征工程方法，在通常的自然语言处理中，Word2vector通常应用到对语境/上下文敏感的场合，对一些应用特定特征单词进行分类的场合会产生混淆信息，使得问题变的更复杂。结合实际的数据集，每一个“工单”内容是客服根据业务理解对双方沟通的要点总结，包含了相关业务描述、诉求以及类别的高频单词/标记，使用考虑了每个词的TF-IDF值与CHI值特征进行特征提取将更为有效。因此本模型选用TF-IDF值与CHI值特征进行特征提取。

在本项目对使用JIEBA分词技术，结合12345的业务世纪，在分词字典中增加诸如“随申码、健康码、市民云、政务大厅、水污染、环评、办事效率、拼多多、新冠、新型冠状病毒、冠状病毒、核酸检测、新冠肺炎、外籍人士、少儿基金、互助基金”等词汇，使得分词效果和算法结果更显著。

项目中以工单文本为输入，通过JIEBA分词加载12345客服的字典库，采用CHI的方式统计每个词的词频特征，然后通过每个词在各文档中出现的次数统计出每个词的逆文档频率特征，在此基础上计算出每个词以及每个文档的TF-IDF特征；最后将选取到特征词向量化作为模型的输入。

### 算法对比及选型

为了保证数据预测的准确率和训练时间。本项目在进行文本分类建模时，调研了**以下几种，并选取10万条第一级数据做训练集，11500条数据做为预测集，进行了算法实验，实验结果如下：**

|  |  |
| --- | --- |
| 算法 | 第一级准确率 |
| 贝叶斯算法 | 68% |
| MLP算法 | 66% |
| Xgboost算法 | 86.7% |
| 随机森林算法 | 62% |
| TextCNN算法 | 89.7% |
| Logistics Regression算法 | 80% |

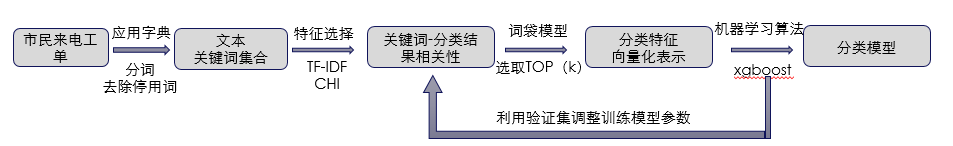
其中TextCNN以及Xgboost两个算法的训练速度最快，准确率也较其他算法的准确率高。因此，根据两种算法进行逐级级测试，最终发现在工单分类进行到第四级时，TextCNN算法在第三级第四级的准确率都低于Xgboost算法。将四级的准确率相乘发现Xgboost算法的准确率高于TextCNN算法。因此本项目中选取的算法为Xgboost算法。

另外，在本项目开发过程中，发现Bert预训练模型能够提升工单分类的准确率，但是由于前期GPU服务器制约原因，在本项目过程中仅使用Bert算法进行了工单分类实验，未采纳Bert作为本项目中的文本分类算法。

## 智能工单分类流程描述

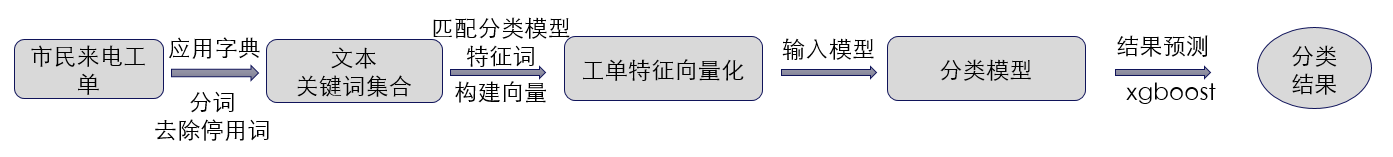
### 模型算法介绍

分类模型训练流程如下图：



1. 将有分类标签的工单训练集使用JIEBA应用字典进行分词，去除停用词，得到相应工单和分类标签对应的文本关键词结合；
2. 使用TF-IDF和CHI进行分类之间相关性的关键词集合选择；
3. 对每个类别的关键词结合选取Top（K），在分类模型经过反复验证，选取Top 2000为最优。
4. 将训练接工单的分词应用分类特征，形成工单分析向量和分类标签表示，应用到xgboost算法，生成初步的分类模型；
5. 根据测试集进行分类模型的准确率测试，调整相关的特征词集合及其它模型参数，生成最终的分类模型。

分类预测使用流程图如下图：



1. 将需要预测类别的工单进行分词，并去掉停用词，得到工单关键词结合
2. 匹配分类模型特征词，构建工单特征的向量化表示
3. 将工单向量输入分类模型，预测分类结果。

### 模型训练

#### ****训练流程****

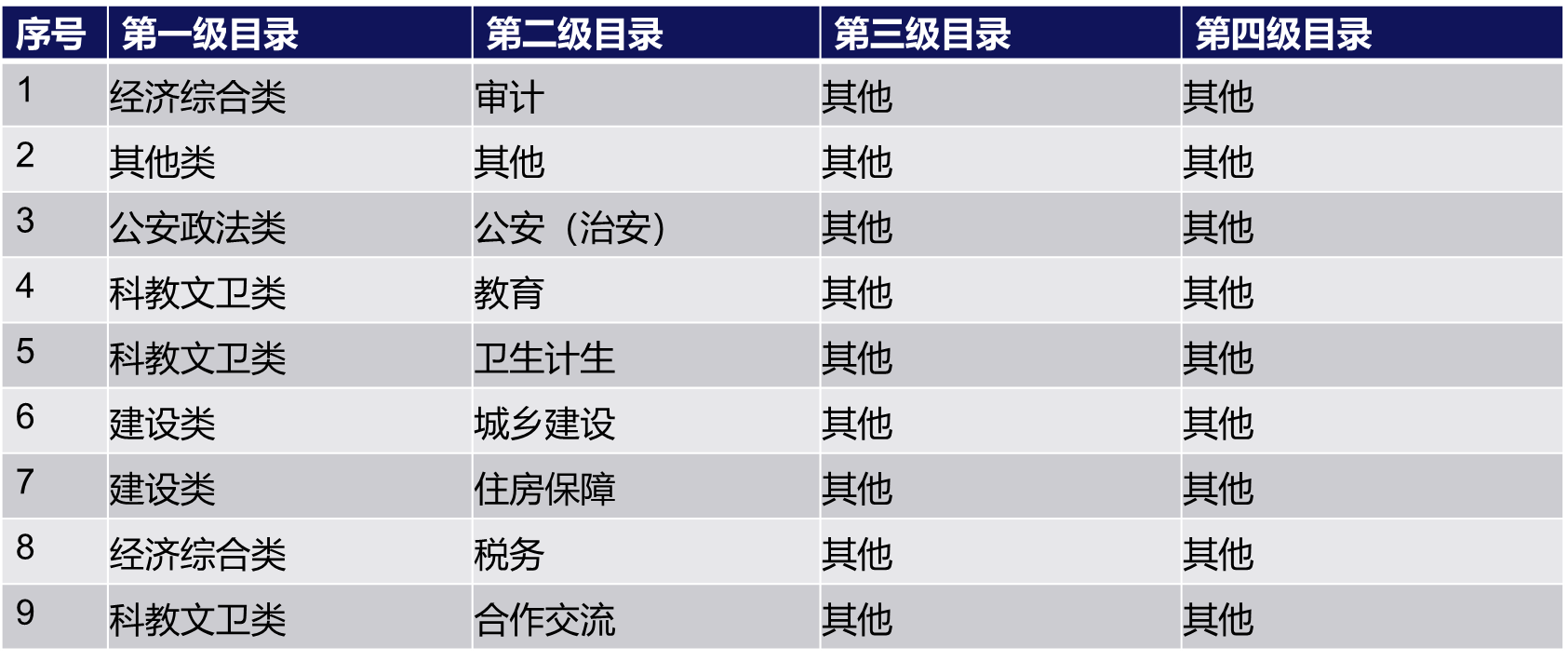
为有效的进行工单分类模型的构建，主要选取了4-5月份的工单数据，经过对工单分类数据分析，将明显分类错误，存在歧义的、无效的工单数据进行加工清洗；针对部分分类数据条目<20条，难以有效提取特征情况，从其它月份的工单数据进行抽取补充，使得各分类的有效工单数据条目数满足>=20条的要求，得到可以使用的工单数据共437624条，其中模型训练工单数据357500条，模型测试工单数据80124条。

训练数据共分为592个四级分类和一个“无效无效”类，592个分类类别共有297985条工单数据，“无效无效”类别有59515条训练数据。

“无效无效”类为在12345业务分类以外新建的类，包括以下几种类型的数据：

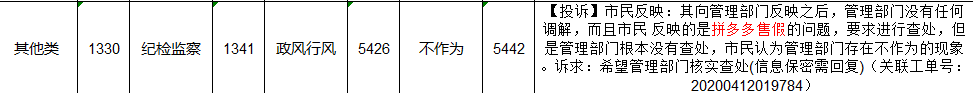
1）四级分类中的（\*\*-其他-其他）数据

\*\*-其他-其他类数据比较分散，而且与具体的分类有数据重叠。在进行分类时，很难被分到具体的其他-其他类，但是该类在工单类目中又是不可或缺的。而\*\*-其他-其他类共有60个，其工单数据量占总工单数据量的9%。



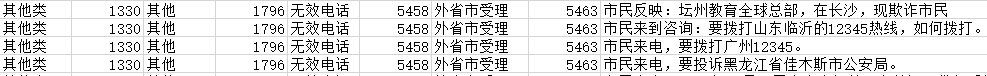
2）涉及态度相关类难以分类

市民来电反映时，如果涉及到事件投诉或工作人员的投诉，工单需要对具体事件进行描述，会误导工单分类。例如：



3）存在较难进行数据建模的类

工单分类类目中，存在一些咨询相关的类，而咨询类工单较短，关键词少且咨询的范围分散，较难建模。例如：





在模型预测为“无效无效“类时，不对客服进行提示。此时需要客服手动进行选择该工单的四级分类。

#### ****算法模型生成****

本次训练生成模型文件和特征词文件各250个，详细的模型文件和特征词文件已部署生产环境，部分模型文件和特征词文件列表如下图：

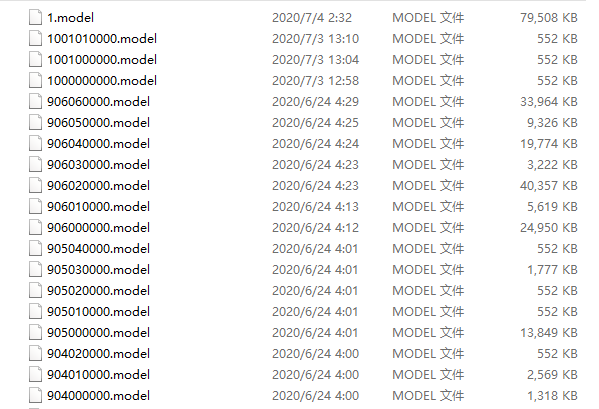


图2 工单分类模型文件

本模型建模依据多层级建模方法，根据第一层分类目录建立第一层模型如图中的1.model；根据第二层分类目录创建第二层模型如图中的100000000.model、200000000.model、…900000000.model、1000000000.model；根据第三层分类目录创建第三层模型如图中的101000000.model、…906000000.model、1001000000.model等；根据第四层分类目录创建第四层模型如图中的906010000.model、1001010000.model等。

模型加载的流程是，首先加载第一层级模型1.model，将分类类别号x1保存在内存中，再加载x100000000.model (x=1,2…,9)模型文件进行第二层级预测分类，将第二层级分类结果类别号x2保存在内存中，再加载x10x2000000.model (x=1,2…,9;y=1,2…)模型文件进行第三层级预测分类，将第三层级分类结果类别号x3保存在内存中，再加载x10x20x30000.model模型文件进行第四层级预测分类，并将第四层级的分类结果号x4保存到内存中，最终x1、x2、x3、x4分别为四层分类模型的分类结果，返回前端即可。

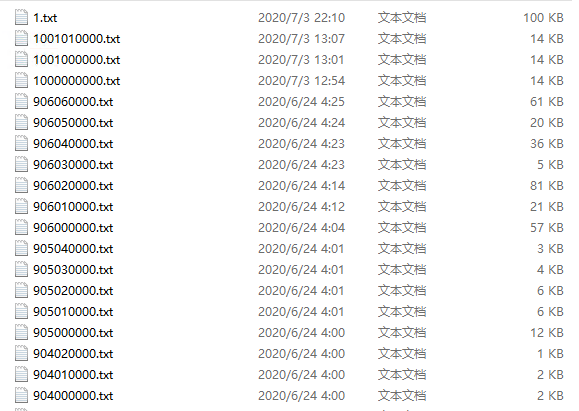


图3 工单分类特征词文件

本模型建模依据多层级建模方法，根据第一层分类目录提取第一层分类关键词如图中的1.txt；根据第二层分类目录提取第一层分类关键词如图中的100000000.txt、200000000.txt、…900000000.txt、1000000000.txt；根据第三层分类目录提取第一层分类关键词如图中的101000000.txt、…906000000.txt、1001000000.txt等；根据第四层分类目录提取第一层分类关键词如图中的906010000.txt、1001010000.txt等。

### 模型测试

#### 测试流程：



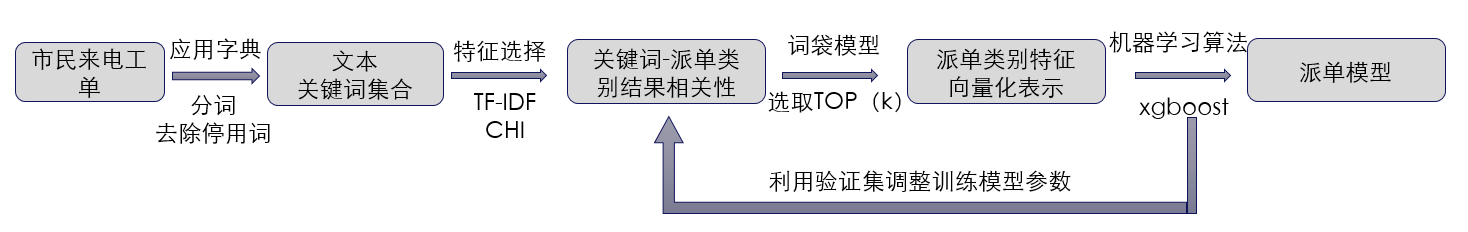
#### 测试结果：

模型测试共有工单数据80124条，经过模型的预测，将实际工单分类和预测分类进行对比，其中分类准确的工单数为59425条，模型预测准确率=59425/80124\*100%=74.17%。

## 智能工单派发流程描述

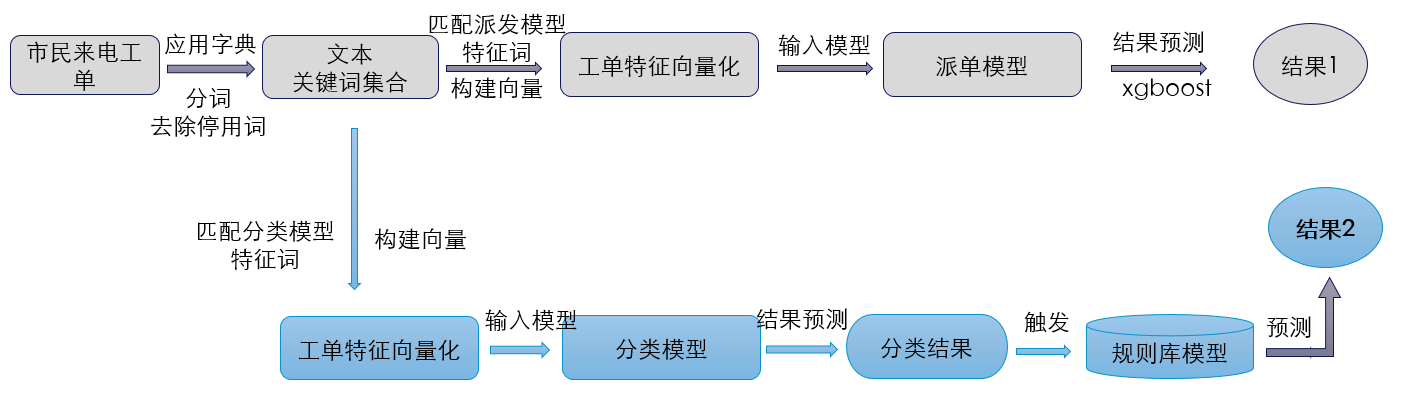
### 模型算法介绍

针对工单派发的模型训练流程图：



1. 将有派发类别标签的工单训练集使用JIEBA应用字典进行分词，去除停用词，得到相应工单和派单类别标签对应的文本关键词结合；
2. 使用TF-IDF和CHI进行派单类别之间相关性的关键词集合选择；
3. 对每个类别的关键词结合选取Top（K），在派单类别模型经过反复验证，选取Top 1600为最优。
4. 将训练接工单的分词应用分类特征，形成工单分析向量和派单类别标签表示，应用到xgboost算法，生成初步的派单模型；
5. 根据测试集进行分类模型的准确率测试，调整相关的特征词集合及其它模型参数，生成最终的派单模型。

工单派发预测使用流程图：



1. 将需要预测派单方向的工单进行分词，并去掉停用词，得到工单关键词结合
2. 匹配派单模型特征词，构建派单方向特征的向量化表示
3. 将工单向量输入派单模型，预测派单结果1。
4. 文本关键词另外通过分类模型预测出相应的四级分类结果，根据规则库进行判断，如果规则库判断无结果，则以派单模型的结果为准，否则以规则库的结果优先。

工单派发模型关键点：

1、根据历史工单数据，建立工单派发模型。

2、建立工单派发规则库，通过工单四级分类结果触发规则库调用模块，得出相应的派发结果。其规则库样例如下表所示。

|  |  |  |  |
| --- | --- | --- | --- |
| **四级分类** | **现象** | **规则** | **派单方向** |
| 建设交通类\_住房保障\_物业服务管理\_群租现象 | 群租 | <p style="LINE-HEIGHT: 1.5em">“群租”类来电的派单方向：</p><p style="LINE-HEIGHT: 1.5em">1、现涉及“群租”的<strong><span style="COLOR: #ff0000">政策咨询类</span></strong>来电<strong><span style="COLOR: #00b0f0">先查询知识库后回复</span></strong>，需派单的转送至<strong>市房屋管理局</strong>。</p><p style="LINE-HEIGHT: 1.5em">2、涉及“群租”的<strong><span style="COLOR: #ff0000">投诉来电</span></strong>派单至“群租”所在地的<strong>区政府</strong>。</p> | 政策资讯类#市房屋管理局|投诉来电#区政府 |
| 安全监管类\_食品药品安全\_安全管理\_食品安全 | 食品安全 | <p>涉及<strong><span style="COLOR: #ff0000">食品安全</span></strong>的投诉，转送<strong>市市场监管局（原</strong><strong>市食药监局）</strong>处理。</p> | 市市场管理局 |
| 社会管理类\_知识产权\_知识产权\_侵权行为 | 专利侵权 | <p style="LINE-HEIGHT: 1.5em"><span style="COLOR: #ff0000">专利侵权</span>纠纷问题，派发<strong>市知识产权局</strong>。</p><p style="LINE-HEIGHT: 1.5em">（注：根据《上海市人民政府关于浦东新区知识产权工作有关事项的决定》，涉及<strong>浦东新区范围内</strong>的<span style="COLOR: #ff0000">商标、版权、专利等侵权纠纷问题</span>，均由浦东新区知识产权局负责处理，这类工单请直接派发<strong>浦东新区</strong>。）</p> | 专利侵权纠纷#市知识产权局|浦东新区#浦东新区 |
| 建设交通类\_交通港口\_出租车\_黑车 | 四轮黑车、克隆出租车 | <p style="LINE-HEIGHT: 1.5em">如市民来电举报黑车，按以下方式处理：<br/>询问是否为<strong><span style="COLOR: #c0504d">四轮机动车</span></strong><br/>如市民举报的黑车是四轮机动车，需询问具体区域：<br/>（1）如市民来电举报黑车（<strong><span style="COLOR: #4f81bd">非浦东新区</span></strong>范围内）问题，根据《上海市查处车辆非法客运办法》，市交通行政管理部门负责本市行政区域范围内车辆非法客运查处的管理。记录相关要素，派单至<strong><span style="COLOR: #4f81bd">交通委</span></strong>。<br/>（2）如市民来电举报<strong><span style="COLOR: #8064a2">浦东新区</span></strong><strong><span style="COLOR: #8064a2">范围</span></strong>内黑车问题，根据《上海市人民政府关于扩大浦东新区城市管理领域相对集中行政处罚权范围的决定》，记录相关要素，派单至<strong><span style="COLOR: #8064a2">浦东新区</span></strong>。<br/>（3）如市民来电反映<strong><span style="COLOR: #9bbb59">机场内</span></strong>涉及黑车治理、规划设计的问题，派单至<strong><span style="COLOR: #9bbb59">交通委</span></strong>。</p><p>★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★</p><p style="LINE-HEIGHT: 1.5em">如市民来电举报克隆出租车：</p><p style="LINE-HEIGHT: 1.5em">&nbsp;1、市民要求相关部门<strong><span style="COLOR: #c0504d">立即到现场</span></strong>进行查处的，<strong><span style="COLOR: #4f81bd">转接</span></strong>110专席<strong><span style="COLOR: #4f81bd">22022616</span></strong>，进行三方通话，工单记录归档；</p><p style="LINE-HEIGHT: 1.5em">2、<strong><span style="COLOR: #c0504d">非即时性</span></strong>的举报，记录相关要素，派单至<strong>交通委</strong>。</p><p style="LINE-HEIGHT: 1.5em">&nbsp;</p><p style="LINE-HEIGHT: 1.5em"><strong>工单要素：</strong>车辆牌照、车身颜色、顶灯、车辆停放点（营运区域）。</p> | 黑车#交通委|浦东新区黑车#浦东新区|黑车治理规划设计#交通委|克隆出租车#交通委|克隆出租车 立即到场#110专席 22022616 |

表2 派单规则库样例

### 模型训练

为有效的进行派单模型的构建，主要选取了4-5月份的派单数据，经过对工单数据分析，将明显分类错误，存在歧义的、无效的工单数据进行加工清洗；针对部分分类数据条目<20条，难以有效提取特征情况，从其它月份的工单数据进行抽取补充，使得各分类的有效工单数据条目数满足>=20条的要求，得到可以使用的派单数据共161381条，其中模型训练派单数据131381条，涵盖136个派单方向，模型测试派单数据30000条。

**模型生成**

本次训练生成模型文件和特征词文件各8个，详细的模型文件和特征词文件已部署生产环境，其模型文件和特征词文件列表如下图:

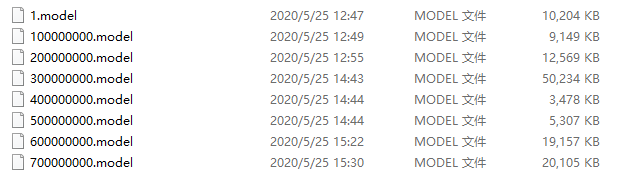


图5 派单模型文件

本模型建模依据多层级建模方法，根据第一层派单目录建立第一层模型如图中的1.model；根据第二层派单目录创建第二层模型如图中的100000000.model、200000000.model、…700000000.model。

模型加载的流程是，首先加载第一层级模型1.model，将派单类别号x1保存在内存中，再加载x100000000.model (x=1,2…,9)模型文件进行第二层级预测派单，将第二层级派单结果类别号x2保存在内存中，最终x1、x2分别为两层派单模型的分类结果，返回前端即可。

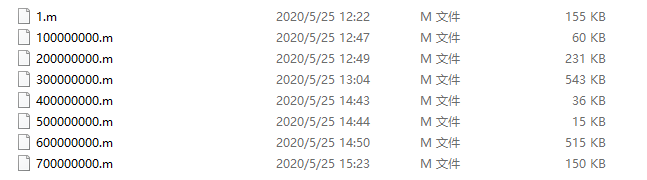


图6 派单特征文件

本模型建模依据多层级建模方法，根据第一层派单目录提取第一层派单关键词如图中的1.m；根据第二层派单目录提取第一层派单关键词如图中的100000000.m、200000000.m、…700000000.m。

### 模型测试

模型测试共有派单数据30000条，经过模型的预测，将实际派单结果和预测派单结果进行对比，其中派单准确的工单数为24011条，模型预测准确率=24011/30000\*100%=80.04%。

# 生产环境准确率测试结果

## 智能工单分类

工单分类功能自2020年7月1日上线进行试运行，共有4个智能客服坐席，本次测试选取7月2日至7月8日共8天的所有智能工单接口的调用数据，共有工单分类数据5844条。

1. **测试标准**

以模型分类结果和客服分类结果进行对比，两者结果相同则判定为正确，两者结果不相同则判定模型分类结果错误。

1. **测试结果**

共有工单分类数据5844条，其中模型分类结果和客服分类结果相同的工单数为4340条，工单分类平均准确率=4340/5844\*100%=74.26%。

其中每一天的准确率详细情况如下表所示：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **日期** | **工单条目数** | **正确条目数** | **错误条目数** | **准确率** |
| 20200702 | 822 | 550 | 272 | 66.91% |
| 20200703 | 732 | 538 | 194 | 73.50% |
| 20200704 | 930 | 714 | 216 | 76.77% |
| 20200705 | 444 | 337 | 107 | 75.90% |
| 20200706 | 1021 | 775 | 246 | 75.91% |
| 20200707 | 846 | 641 | 205 | 75.77% |
| 20200708 | 1049 | 785 | 264 | 74.83% |
| 总计 | 5844 | 4340 | 1504 | 74.26% |

表 工单分类每天准确率详细结果

## 智能工单派发

本次准确率测试选取7月1日至7月8日共8天的所有涉及到工单派发的工单数据，共有工单分类数据79553条。

1. **测试标准**

以模型派发结果和客服选择的派发结果进行对比，两者结果相同则判定为正确，两者结果不相同则判定模型派发结果错误。

1. **测试结果**

共有工单派发数据79553条，其中模型派发结果和客服选择的派发结果相同的工单数为63396条，工单派发平均准确率=63396/79553\*100%=79.69%。

其中每一天的准确率详细情况如下表所示：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **日期** | **总条数** | **正确条数** | **错误条数** | **准确率** |
| 20200701 | 13193 | 10519 | 2674 | 79.73% |
| 20200702 | 11130 | 8931 | 2199 | 80.24% |
| 20200703 | 11409 | 8955 | 2454 | 78.49% |
| 20200704 | 8403 | 6904 | 1499 | 82.16% |
| 20200705 | 6117 | 4929 | 1188 | 80.58% |
| 20200706 | 11737 | 9053 | 2684 | 77.13% |
| 20200707 | 10202 | 8200 | 2002 | 80.38% |
| 20200708 | 7362 | 5905 | 1457 | 80.21% |
| 总计 | 79553 | 63396 | 16157 | 79.69% |

表 工单派发每天准确率详细结果