

机智机器学习

操作指南

产品文档



腾讯云

## 【版权声明】

©2015-2016 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

## 【商标声明】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。

## 【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或模式的承诺或保证。

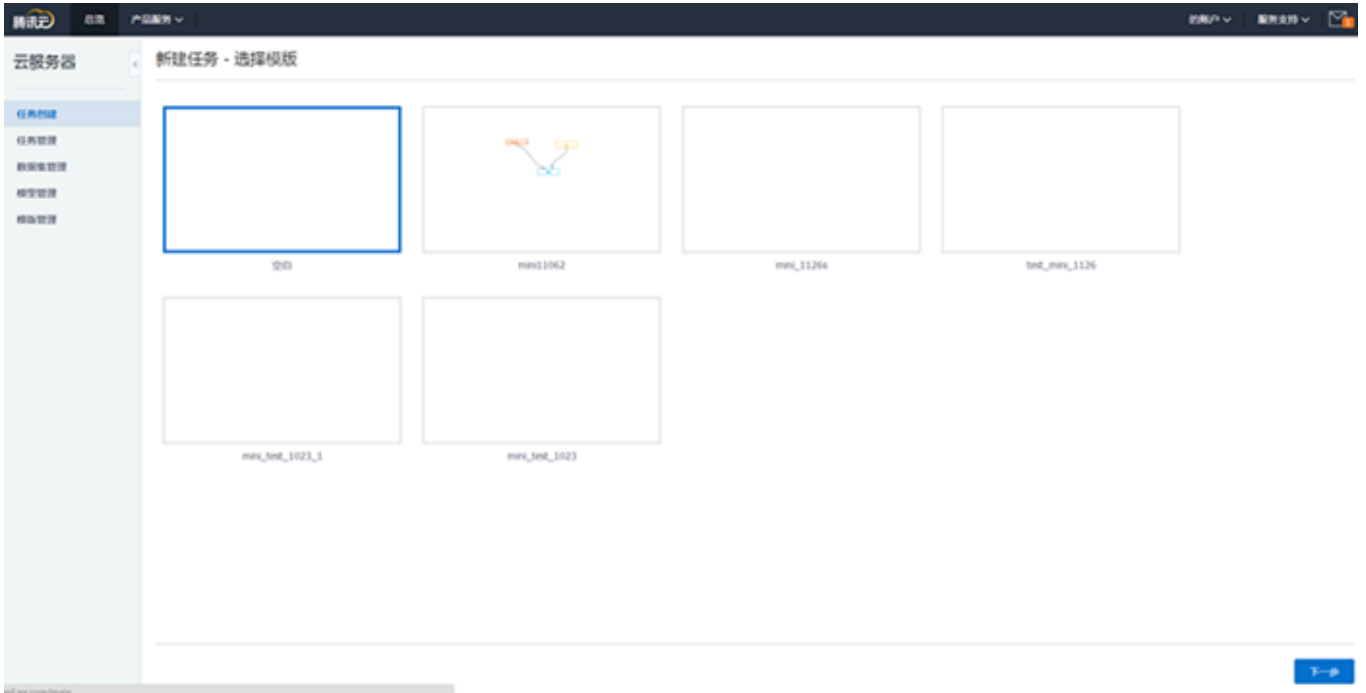
## 文档目录

文档声明.....	2
使用帮助 .....	4
使用场景 .....	9

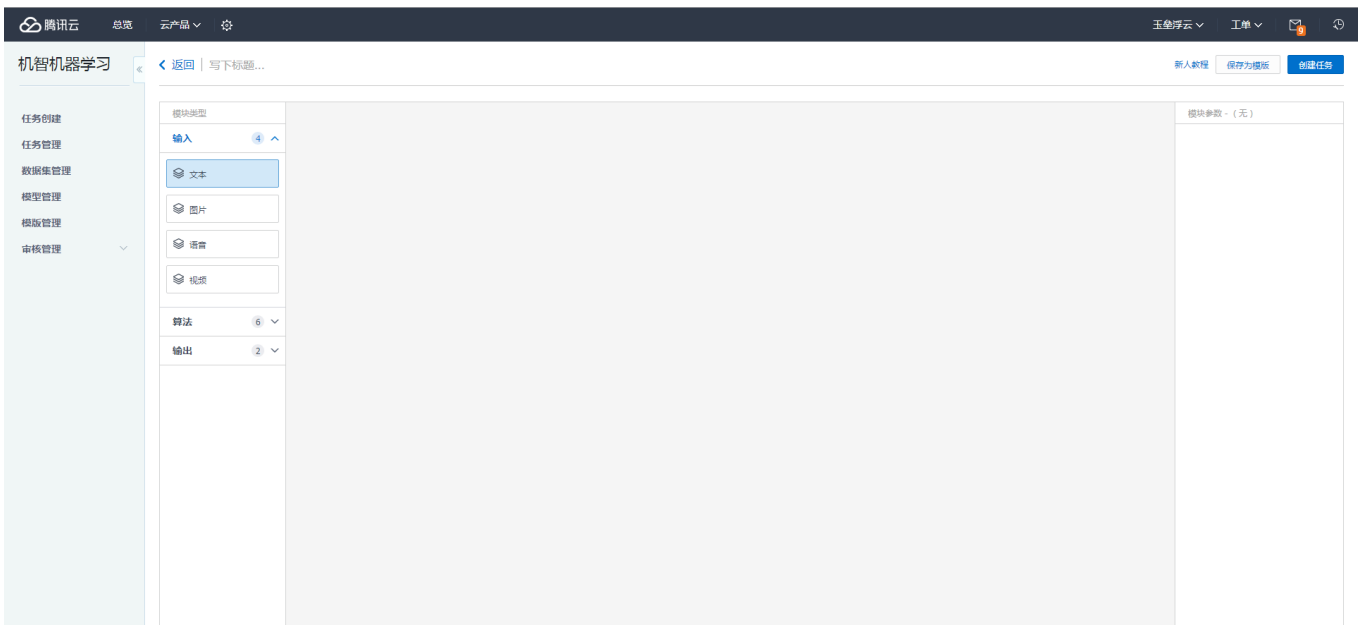
## 使用帮助

### 1 任务创建

首先选择一个模版，默认从空模版建立，也可以从已有模版建立任务，点击“下一步”（图1），跳到操作界面（图2）。新用户在没有历史任务的情况下，无此步骤，直接跳为图2。

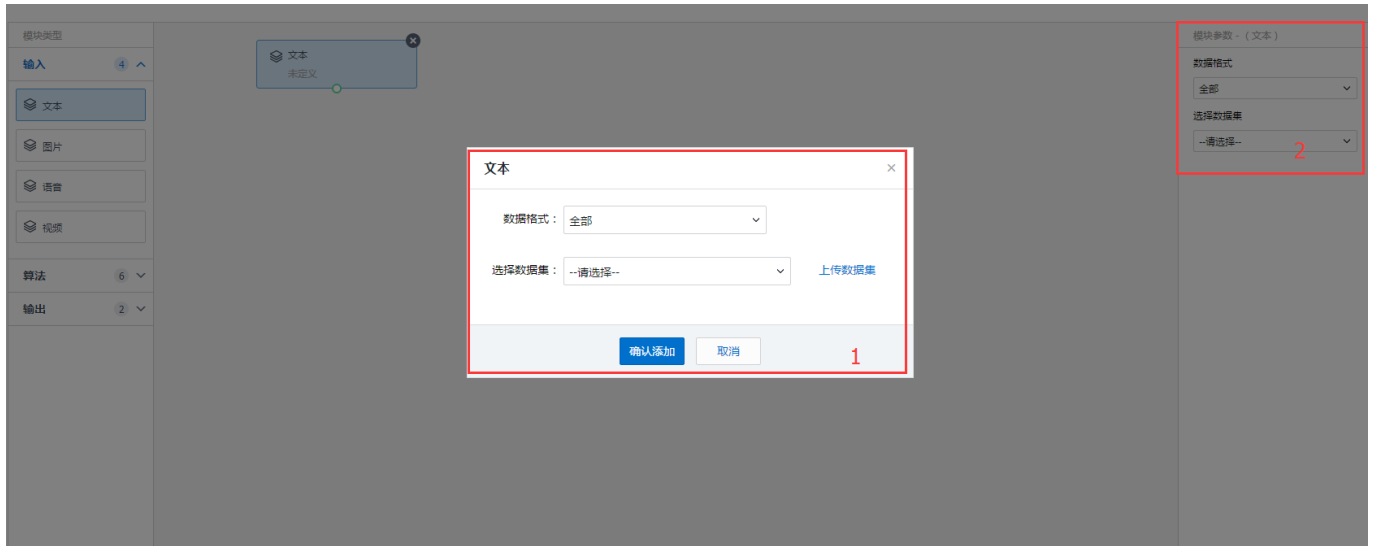


(图1)



(图2)

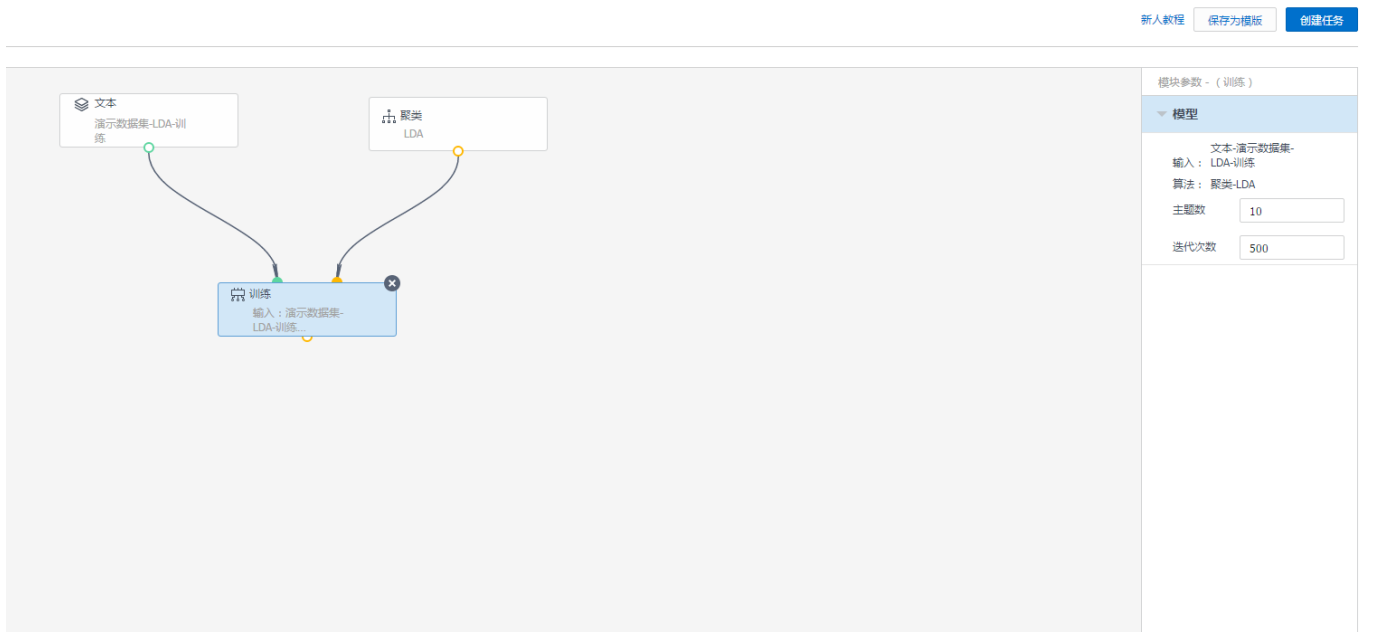
创建一个机器学习任务，需要有“输入”的数据、使用的“算法”、“输出”的动作三部分组成。因此，需要用户从操作界面左边的菜单栏中，拖拽出组成一个任务必要的三个模块，并填入好内容。如：图3示，拖拽出“输入” - “文本” 模块，填入文本相关数据。（可在1号窗口处填入，也可在2号窗口处进行数据的修改）。若数据库中未有所需处理的数据，可点击上传数据集。



(图3)

拖拽出所有模块后，需要根据图4示例，将模块进行连接。在“输出” 模块选择状态下，右侧会展现模块相关信息，并且有使用连接算法的参数设置功能（每个算法已经默认填好了参数，可以直接运行）。然后点击右上角的“创建任务” 进行任务的运作，或者“保存为模版”。

注意：算法和用于该算法的数据必须统一，因为不同的算法支持的数据格式不同



(图4)

填入任务名称和任务描述，任务描述为非必填。任务或模版保存成功后，跳转到对应的“任务管理” 或“模版

管理”界面进行管理。

### 保存任务 ×

将会另存为一个新的任务，不覆盖原有任务

任务名称：

添加描述：

审核通过后，自动运行任务

保存 取消

(图5)

## 2 任务管理

对创建的任务进行管理。可以对生成的任务进行“停止”、“查看”、“编辑”和“删除”的操作。

[停止](#) | [查看](#) | [编辑](#) | [删除](#)

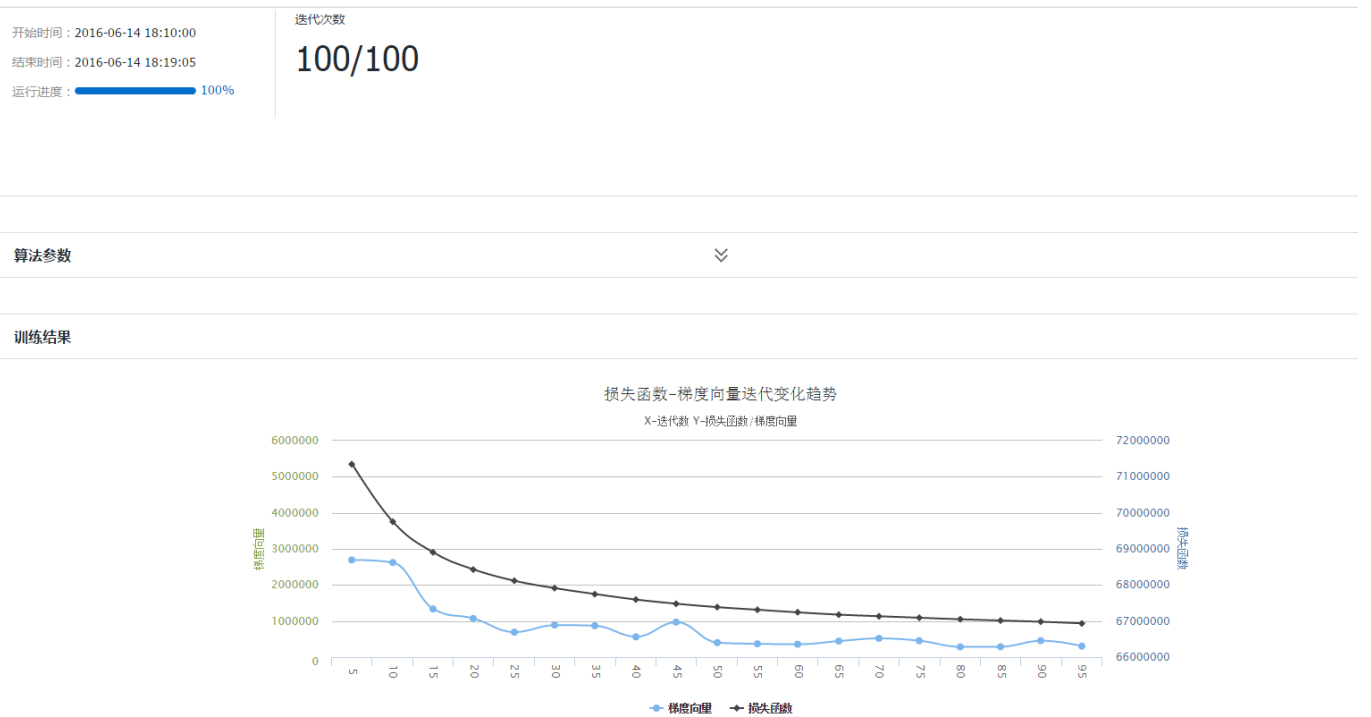
(图6)

在任务正常运行过程中、运行完成时期，可以随时查看中间结果和最终结果。

任务管理

ID	名称	描述	修改日期	类型	运行状态	操作
10000	LDA模型test		2015-12-01 14:29:53	预测	待审核	运行   编辑   删除
178	ONN模型预测	ONN模型预测	2015-11-27 16:59:33	训练	运行中	停止   编辑   删除
171	mini1062		2015-11-26 14:41:35	训练	已完成	查看详情   编辑   删除
169	mini_1126s		2015-11-26 14:38:36	训练	已审核	运行   编辑   删除
168	test_mini_1126	test	2015-11-26 14:32:29	训练	已停止	运行   编辑   删除
102	mini_test_1023_1	cuowu	2015-10-23 10:11:44	训练	已停止	运行   编辑   删除
99	mini_test_1023	数据和算法不匹配	2015-10-23 10:00:11	训练	已停止	运行   编辑   删除

(图7)



(图8)

### 3 数据集管理

在此界面中，可以对自身数据进行管理，可下载或删除数据集，也可在上传新数据。

数据集管理

+ 新建

名称	描述	创建日期	操作
234	54	2015-09-24 15:03:38	删除   下载
123	123	2015-09-24 15:01:36	删除   下载

(图9)

## 4 模型管理

训练任务完成后，会生成模型文件，此界面用于对模型文件的管理。可以“下载”、“预览”和“删除”模型文件。

## 5 模版管理

当任务成功创建保存后，会自动生成一个以任务为原型的模版，用于后续同质任务的快速创建。

欢迎加入腾讯机智官方交流qq群：252119476



## 使用场景

### 1 LDA

应用场景比较广泛，适合但不限于以下场景：

- 1) 对无标注的文本进行聚类，适用于做topic model，比如新闻热点、热门应用等的挖掘；
- 2) 由于主题模型挖掘了文本的语义关系，也可用在文本分类等领域做feature；
- 3) 得到对词的主题分布聚类也可用于标签自动生成等。

### 2 CNN

CNN主要应用于多媒体数据(图片、音频)的识别，特征提取等，典型的应用场景有：人脸识别，手写体识别。

### 3 LR

如：用户购买行为预测

以预测用户对品类的购买偏好为例，介绍逻辑回归应用。该问题可以转换为预测用户在未来某个时间段内是否会购买某个品类商品，如果把会购买标记为1，不会购买标记为0，就转换为一个二分类问题。我们用到的特征包括用户浏览，购买等历史信息，见下表

类别	特征
用户	购买频次，浏览频次，时间，地理位置 ...
品类	销量，购买用户，浏览用户 ...
交叉	购买频次，浏览频次，购买间隔 ...

人工把相似的小品类聚合成18个较为典型的品类集合。如果用户在给定的时间内购买某一品类集合，就作为该品类分类器的正例，没有购买的则作为负例。有了训练数据后，使用LR算法对每个品类训练一个二分类模型，用于预测在各个品类上的购买概率。预测的结果则会用于推荐等场景。

### 4 Word2Vec

Word2Vec在NLP应用广泛，输出的词向量用作词性分析、聚类、同义词等。

欢迎加入腾讯机智官方交流qq群：252119476