

推荐系统

快速入门

文档版本 01
发布日期 2024-09-07



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

目录

1 智能场景（猜你喜欢）	1
2 自定义场景（热度推荐）	7

1 智能场景（猜你喜欢）

RES提供了智能场景包括猜你喜欢、热门推荐和关联推荐。仅需要简单的配置和作业训练，即可获得推荐结果。

本章节以猜你喜欢为例，帮助您快速熟悉智能场景的使用过程。开始使用样例前，请仔细阅读**准备工作**罗列的要求，提前完成准备工作。使用智能场景获取推荐结果的步骤如下所示：

- **步骤1：准备数据**
- **步骤2：创建离线数据源**
- **步骤3：创建智能场景**
- **步骤4：获取预测结果**
- **步骤5：清除相应资源，避免产生费用**

准备工作

- 已注册华为帐号并开通华为云，且在使用RES前检查帐号状态，帐号不能处于欠费或冻结状态。
- 已在OBS服务中创建桶和文件夹，用于存放样例数据。确保您使用的OBS目录与RES在同一区域。

步骤 1：准备数据

RES在公共OBS桶中提供了猜你喜欢的示例数据集，命名为“test-data”，因此，本文的操作示例使用此数据集进行构建。您需要执行如下操作，将数据集上传至您的OBS目录下，即准备工作中您创建的OBS目录。

1. 单击**下载样例数据**，将“test-data”数据集下载至本地。
2. 在本地，将“test-data”压缩包解压。例如，解压至本地“test-data”文件夹下。
3. 在“test-data”文件夹下，将behavior.txt中的每条数据的actionTime字段的值修改到当前时间附近。将item.txt中的每条数据的publishTime字段的值修改到当前时间附近，将item.txt中的每条数据的expireTime字段的值修改成大于当前时间的值，避免数据因为过期被过滤掉。
 - a. 查询当前时间戳，网址<https://tool.lu/timestamp>。

图 1-1 查询当前时间戳



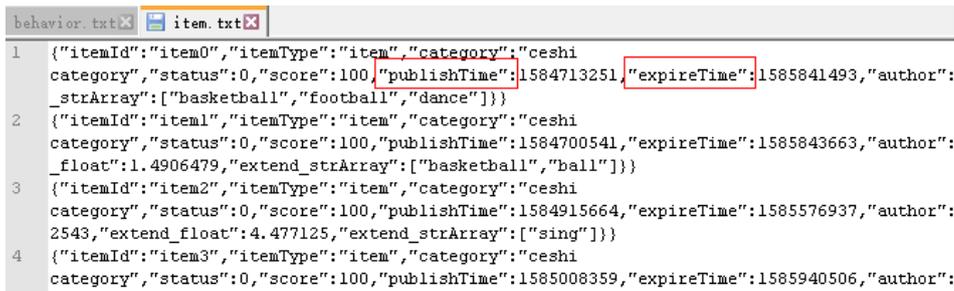
b. 将behavior.txt中的每条数据的actionTime字段的值修改到当前时间附近。

图 1-2 修改 behavior.txt 文件



c. 将item.txt中的每条数据的publishTime字段的值修改到当前时间附近，将item.txt中的每条数据的expireTime字段的值修改成大于当前时间的值。

图 1-3 修改 item.txt 文件



4. 将“test-data”文件夹下的所有文件上传至准备工作中您创建的OBS路径下。

步骤 2: 创建离线数据源

数据下载并上传至OBS后，您需要创建一个数据源用于后续的计算。具体操作步骤如下：

1. 登录RES管理控制台，在左侧菜单栏中选择“数据源”，进入“数据源”列表页面。
2. 单击“创建”，在创建数据源页面，输入数据源名称，选择离线计算规格，然后单击 选择数据的OBS存储路径。

图 1-4 创建离线数据源

3. 数据路径选择完成后单击“立即创建”。
4. 离线数据源创建完成后，在数据源列表页面单击目标数据源名称进行数据质量管理，具体操作请参见[推荐系统用户指南>数据源质量管理](#)，完成数据探索并生成数据质量报告，此步骤完成后创建的数据源才可用。

步骤 3：创建智能场景

1. 登录RES管理控制台，在左侧导航栏中选择“推荐业务>智能场景”，默认进入“智能场景”列表。
2. 在智能场景列表中，单击右上角“创建”，进入“创建智能场景”页面。
3. 在创建智能场景页面的“模板选择”页签，单击“猜你喜欢”进行参数配置。填写“基本信息”和“个性化配置”相关参数。

图 1-5 创建智能场景

表 1-1 智能场景参数说明

参数名称	子参数	说明
基本信息	名称	自定义的名称。只能包含数字、大小写字母、下划线和中划线，长度不能超过64位且不能为空。 名称创建之后不支持修改。
	数据源	从右侧下拉框中选择RES系统中已有的数据源。当无可用数据源时，此下拉框为空。
	描述	对于该场景的描述信息。
场景规格	-	选择离线计算、实时计算、排序模型训练规格和在线并发数。
个性化配置	匹配特征对	匹配用户和物品特征，以便于筛选出该用户相关联的物品进行推荐。 <ul style="list-style-type: none">● 用户特征名：从下拉框中选择目标用户特征用于和物品特征进行匹配。● 物品特征名：从下拉框中选择目标物品特征用于匹配用户特征，更好的做出推荐。● 权重：取值为0.01-1。权重越高，该匹配特征所被优先推荐的概率越高。● 匹配个数度量：默认不勾选。当勾选之后表示满足的匹配程度。● 操作：选择删除某个匹配特征对。
	行为过滤	指定时间区间内用户对物品有过指定的行为类型记录时，将过滤掉相应物品。 其中排除行为过滤是指将有此行为类型的物品排除掉，不对其进行过滤。 例如，对于用户过去3天内有过曝光行为的物品过滤，但是可以对曝光过但是有购买行为的物品进行排除，不对其进行过滤。
	开启调度	开启调度，按照指定的调度策略定期执行作业。 <ul style="list-style-type: none">● “调度周期”：调度周期可选“天”或“周”。● “调度类型”：包括自定义和间隔调度。● “开始调度时间”：选择具体的调度时间。当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。● “时间间隔”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。

4. 配置完成后单击“立即创建”。

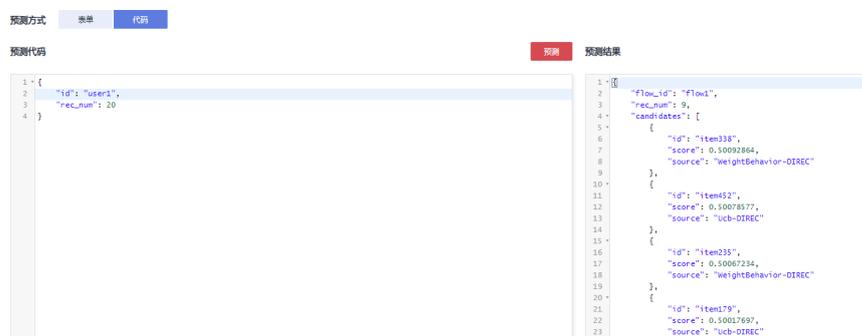
创建成功后会自动跳转到“智能场景”列表页面并提示您创建成功。针对创建好的智能场景，您可以单击操作列的“发布”。待场景状态为“运行中”时，您可以在智能场景列表页面单击目标场景名称进入详情页面进行“编辑”、“预测”和“效果评估”操作。

步骤 4：获取预测结果

智能场景发布运行成功后，当服务状态会显示“运行中”，表示状态正常。您可以通过预测功能测试推荐结果进一步调整作业参数，也可以通过预测接口来调用API，获取推荐结果。

1. 登录RES管理控制台，在左侧菜单栏中选择“推荐业务>智能场景”，进入智能场景列表页面。
2. 单击“已完成”状态的目标服务名称进入智能场景详情页面，单击“预测”页签，包括两种预测方式。此样例以页面预测代码为例，指导您获取推荐结果。
3. 选择“代码”方式，在左侧框内输入预测代码，单击“预测”后在右侧显示预测结果，如图1-6所示。输入用户ID和推荐数量，根据您的策略返回用户的预测结果。

图 1-6 代码预测



请求代码

“id”为需要请求的用户id，“rec_num”为推荐物品的数量。

```
{
  "id": "user894",
  "rec_num": 10
}
```

预测结果

“flow_id”代表流程id，“rec_num”为推荐物品结果的数量，“candidates”为候选集结果，字段“id”为推荐结果物料id，“score”为推荐物料的评分，“source”代表该推荐物料来源策略。

```
{
  "flow_id": "flow1",
  "rec_num": 10,
  "candidates": [
    {
      "id": "item235",
      "score": 0.58418584,
      "source": "WeightBehavior-DIRECT"
    },
    {
      "id": "item338",
      "score": 0.5653803,
      "source": "WeightBehavior-DIRECT"
    },
    {
      "id": "item709",
      "score": 0.55219465,
      "source": "WeightBehavior-DIRECT"
    },
    {

```

```
"id": "item960",
"score": 0.55219465,
"source": "WeightBehavior-DIREC"
},
{
"id": "item469",
"score": 0.55219465,
"source": "WeightBehavior-DIREC"
},
{
"id": "item236",
"score": 0.55219465,
"source": "WeightBehavior-DIREC"
}
]
}
```

步骤 5：清除相应资源，避免产生费用

为避免产生不必要的费用，在完成试用后，建议您删除相关资源，如数据源、智能场景作业及其OBS数据。

- 删除数据源：在“数据源”页面，单击操作列的“删除”。
- 删除智能场景作业：在“推荐业务>智能场景”页面，单击操作列的“删除”。
- 删除数据：前往OBS，删除上传的数据，然后删除文件夹及OBS桶。

2 自定义场景（热度推荐）

RES提供了推荐算法，让用户能够根据场景自定义推荐策略，可以基于RES提供的多种召回、排序算法等进行自定义的推荐场景创建。

本章节介绍热度推荐场景的搭建样例，该场景常见于电商或者视频网站首页的排行榜或者畅销榜等。

上述推荐场景在RES的自定义场景通过简单配置和计算，即可完成典型的热度推荐功能，通过RES内置综合行为热度排序算法，使用户能够快速够获取到时下最流行的信息。

开始使用样例前，请仔细阅读[准备工作](#)罗列的要求，提前完成准备工作。

准备工作

- 已注册华为帐号并开通华为云，且在使用RES前检查帐号状态，帐号不能处于欠费或冻结状态。
- 已在OBS服务中创建桶和文件夹，用于存放样例数据。确保您使用的OBS目录与RES在同一区域。

使用自定义场景获取推荐结果的步骤如下所示：

- [步骤1：准备数据](#)
- [步骤2：创建离线数据源](#)
- [步骤3：创建自定义场景](#)
- [步骤4：发布自定义场景](#)
- [步骤5：获取预测结果](#)
- [步骤6：清除相应资源，避免产生费用](#)

步骤 1：准备数据

RES在公共OBS桶中提供了猜你喜欢的示例数据集，命名为“test-data”，因此，本文的操作示例使用此数据集进行构建。您需要执行如下操作，将数据集上传至您的OBS目录下，即准备工作中您创建的OBS目录。

1. 单击[下载样例数据](#)，将“test-data”数据集下载至本地。
2. 在本地，将“test-data”压缩包解压。例如，解压至本地“test-data”文件夹下。

- 在“test-data”文件夹下，将behavior.txt中的每条数据的actionTime字段的值修改到当前时间附近。将item.txt中的每条数据的publishTime字段的值修改到当前时间附近，将item.txt中的每条数据的expireTime字段的值修改成大于当前时间的值，避免数据因为过期被过滤掉。
 - 查询当前时间戳，网址<https://tool.lu/timestamp>。

图 2-1 查询当前时间戳



- 将behavior.txt中的每条数据的actionTime字段的值修改到当前时间附近。

图 2-2 修改 behavior.txt 文件



- 将item.txt中的每条数据的publishTime字段的值修改到当前时间附近，将item.txt中的每条数据的expireTime字段的值修改成大于当前时间的值。

图 2-3 修改 item.txt 文件



- 将“test-data”文件夹下的所有文件上传至准备工作中您创建的OBS路径下。

步骤 2：创建离线数据源

数据下载并上传至OBS后，您需要创建一个数据源用于后续的计算。具体操作步骤如下：

1. 登录RES管理控制台，在左侧菜单栏中选择“数据源”，进入“数据源”列表页面。
2. 单击“创建”，在创建数据源页面，输入数据源名称，选择离线计算规格，然后单击  选择数据的OBS存储路径。

图 2-4 创建离线数据源



名称

离线作业计算规格 16CU 64CU 128CU

用户属性表 

物品属性表 

用户操作行为表 

3. 数据路径选择完成后单击“立即创建”。
4. 离线数据源创建完成后，在数据源列表页面单击目标数据源名称进行数据质量管理，具体操作请参见[推荐系统用户指南>数据源质量管理](#)，完成数据探索并生成数据质量报告，此步骤完成后创建的数据源才可用。

步骤 3：创建自定义场景

1. 在RES管理控制台，在左侧导航栏中选择“推荐业务 > 自定义场景”，然后单击“创建”进行场景的配置。
2. 在“基本信息”页面，配置“基本信息”、“数据区间”和“场景规格”信息。完成参数配置后单击“创建并下一步：召回策略”，进入召回策略的配置。
 - “场景名称”：自定义您的场景名称，自定义样例为“hot-scene”。
 - “场景类型”：选择基于用户推荐物品。
 - “服务类型”：选择推荐引擎。
 - “数据源”：选择[步骤2](#)中创建完成的数据源。
 - “数据区间”：选取离线数据中最近N天的数据（从数据中最近的有效时间之前30天）。
 - “场景规格”：由于此样例中使用的测试数据量较小，此处场景规格均默认选最小配置，“在线并发规格”是指在线服务获得推荐的最高调用次数，单位为次每秒，取值范围为5-500。

图 2-5 自定义场景基本信息

基本信息

* 场景名称

* 场景类型 ? 基于用户推荐物品 基于用户推荐用户 基于物品推荐物品 基于物品推荐用户

* 服务类型 ? 推荐引擎 排序引擎

* 数据源 ?
注意：创建完场景后数据源不能修改

数据区间

时间单位 天 小时

* 数据选择 ?

物品类别 ?

设置黑名单 ?

场景规格

* 离线作业计算规格 ? 16CU 64CU 128CU

* 实时作业计算规格 ? 4CU 16CU 32CU

* 排序模型训练规格 ? GPU-V100 8CU

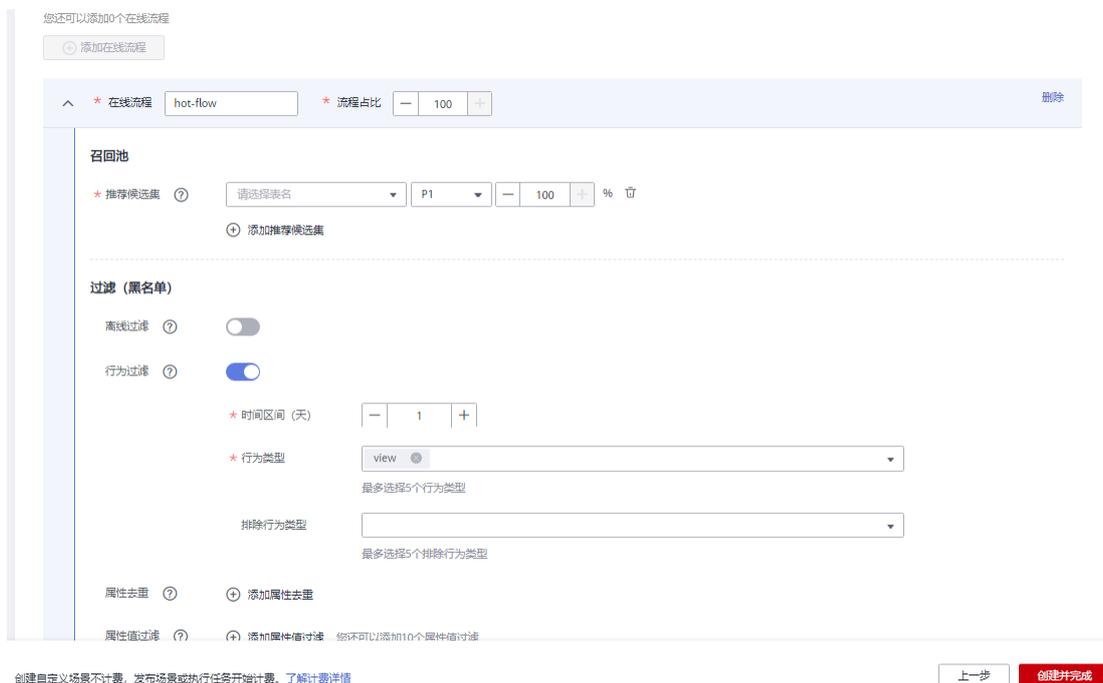
* 在线服务并发规格 ?

3. 在“召回策略”页面，单击“添加召回策略”选择“基于综合行为热度推荐”，进入策略详情页进行参数配置。
 - “名称”：自定义您的作业名称，自定义样例为“hot-recall”。
 - “行为次数统计方法”：选择“uv”，即相同的行为记录会进行去重。“pv”的意思是同样行为记录不会去重。
 - “用户分群”：数据源类型包括用户特征和物品特征，根据数据源筛选数据，选出需要的属性。只能选择进行用户分组还是物品分组，分组内可配置多个特征。默认关闭。
 - “最大推荐数结果数”：指定召回的结果数量。
 - “开启时间跨度”：不开启取全部数据，开启则指定从数据源中取最近天数或小时数的行为数据计算相似度。默认关闭。
 - “开启调度”：开启调度，按照指定的调度策略定期执行作业。默认关闭。

图 2-6 创建召回策略

4. 配置完成召回策略参数之后单击“确认”，创建召回策略成功，继续单击“下一步”，跳过可选步骤过滤策略和排序策略，进入“在线服务”页面，进行在线服务的配置。
5. 在“在线服务”配置页面，进行在线流程配置，配置完成后单击“创建并完成”。
 - “在线流程”：自定义在线流程名称，此样例命名为“hot-flow”。
 - “推荐候选集”：选择步骤3配置的召回策略生成的召回结果集“hot-recall-DIREC”。
 - “过滤”：非必选，此样例进行黑名单过滤配置来完成在线的推荐结果过滤。
 - “行为过滤”：配置“时间区间”为“3”，“行为类型”选择“物品曝光”即为在线服务生成的结果会过滤近三天内用户浏览过的物品，可以提高结果集的丰富度，尤其是在短视频领域，达到推荐的视频列表不包含近期看过的视频，避免重复推送。

图 2-7 添加在线流程



步骤 4：发布自定义场景

1. 自定义场景创建完成，在自定义场景列表，选择目标场景，即**步骤3：创建自定义场景**创建的“hot-scene”。
2. 单击目标场景“操作”列的“发布”即可以启动场景的运行。
3. 待自定义场景列表，目标场景状态为“运行中”时，代表发布成功。

图 2-8 发布服务



步骤 5：获取预测结果

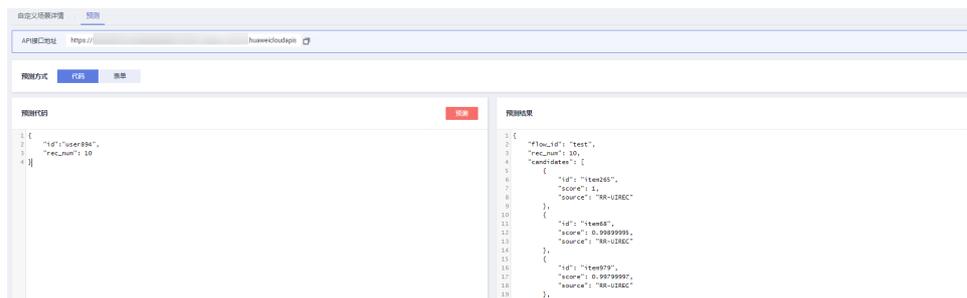
在线服务发布执行成功后，您可以进入在线服务，发起预测请求进行测试。

1. 在“自定义场景”列表页面，单击目标场景名称，进入“自定义场景详情”页面。
2. 单击“预测”页签，进入预测页面。
3. 预测方式选择“代码”，在“预测代码”部分输入预测代码。单击“预测”，右侧出现排序后的预测结果。

说明

由于此处是测试服务，为保证测试效果，此处选用测试数据中的用户ID进行预测，建议为user894。

图 2-9 预测结果



— 预测代码

```
{
  "id": "user894",
  "rec_num": 10
}
```

— 预测结果

```
{
  "flow_id": "hot-flow",
  "rec_num": 10,
  "candidates": [
    {
      "id": "item332",
      "score": 1,
      "source": "hot-recall-DIREC"
    },
    {
      "id": "item709",
      "score": 0.995,
      "source": "hot-recall-DIREC"
    },
    {
      "id": "item338",
      "score": 0.99,
      "source": "hot-recall-DIREC"
    },
    {
      "id": "item960",
      "score": 0.98499995,
      "source": "hot-recall-DIREC"
    },
    {
      "id": "item469",
      "score": 0.97999996,
      "source": "hot-recall-DIREC"
    },
    {
      "id": "item236",
      "score": 0.97499996,
      "source": "hot-recall-DIREC"
    }
  ]
}
```

步骤 6：清除相应资源，避免产生费用

为避免产生不必要的费用，在完成试用后，建议您删除相关资源，如自定义场景、数据源及其OBS目录。

- 删除数据源：在“数据源”页面，单击操作列的“删除”。
- 删除自定义场景作业：在“推荐业务>自定义场景”页面，单击操作列的“删除”。删除自定义场景后，其对应的召回策略和在线服务都将被删除。

- 删除数据：前往OBS，删除上传的数据，然后删除文件夹及OBS桶。