

# AnalyticDB向量检索+AI 实战: 单轮问答系统

---

- 1. 背景
- 2. 效果展示
- 3. 问答系统原理
- 4. 如何使用AnalyticDB搭建单轮问答系统
  - 4.0 创建插件
  - 4.1 建表
  - 4.2 建索引
  - 4.3 创建特征提取算法pipeline
  - 4.4 提取文本特征向量
  - 4.5 问题库导入AnalyticDB
  - 4.6 搜索最相似的问题
- 5. AnalyticDB介绍
- 6. 结语

## 1. 背景

传统的数据库对于管理和分析结构化数据有着很好地支持. 但传统数据库是对于图像, 视频, 文本等非结构化数据的分析能力是有所欠缺的. AnalyticDB的向量检索和AI算法能力能, 能够帮助用户轻松的搭建分析图像, 视频, 文本等复杂数据的应用. 今天我们来介绍如何通过AnalyticDB来搭建一个单轮问答系统. 内容包括: 单轮问答系统效果展示, 单轮问答系统实现原理, 如何使用AnalyticDB实现这样的系统, AnalyticDB介绍, 总结和演示系统的源码分享.

## 2. 效果展示

我们首先来介绍一下使用AnalyticDB搭建的单轮问答系统的效果. 我们预先收集了关于阿里云云服务器ECS的500多条问题和答案组成一个问题库, 如下图所示. 我们可以在搜索框中输入我们想要问的关于ECS的问题(例如"怎么改密码?", "公网ip能否更换?", "能不能退款?") 然后就可以快速找到与之最相似的问题和

对应的答案. 阿里云的文本特征提取模型基于海量数据训练可以适应多种不同的场景.

AnalyticDB 阿里云分析型数据库 非结构化数据分析 演示 产品页面

简介 AnalyticDB 演示 图像视频 语音文本 AnalyticDB Demos

**AnalyticDB 单轮问答系统**

单轮问答系统, 主要依赖两个关键模块. 首先是文本转特征模块, 通过基于海量文本数据训练的深度学习模型我们可以将不同长度的文本转化成固定长度的特征向量. 这个特征向量描述了文本的语义. 第二个模块是向量检索引擎, 可以在海量的特征向量中匹配与目标特征最相似的记录. 如下图所示, 我们使用特征提取模块, 离线提取问题库中所有问题的特征向量. 然后将问题库中的所有问题, 答案, 和问题的特征向量保存到数据库中. 然后当用于进行提问时, 我们首先提取问题的特征向量, 然后在数据库中检索与问题最为相似的问题并返回对应的答案. 在实际应用场景中我们可以根据目标问题与返回结果的特征距离来判断用户的提问是否存在于我们的问题库中.

**搜索结果**

问题	答案	距离(越小越相似)
如何在实例内部修改登录密码?	详细步骤请参见实例内部修改登录密码。	87.4777908325195
忘记了远程连接密码, 怎么办?	您可以修改远程连接密码, 详细步骤请参见修改远程连接密码。	100.970077514648
如何查询Linux实例的账号和密码?	云服务器ECS的账号和密码会在创建成功后发送到您账号绑定的手机和邮箱。Linux系统的默认管理员账号是root。	101.017272949219

返回个数:10 刷新

搜索问题: 怎么改密码? 搜索

关于阿里云ECS问题  
可以尝试以下问题  
忘记密码怎么办?  
怎么计费?  
如何升配?  
怎么开发票?  
实例无法连接怎么办?  
如何选择可用区?

提问"怎么改密码?"

AnalyticDB 阿里云分析型数据库 非结构化数据分析 演示 产品页面

简介 AnalyticDB 演示 图像视频 语音文本 AnalyticDB Demos

的话语. 第二个模块是向量检索引擎, 可以在海量的特征向量中匹配与目标特征最相似的记录. 如下图所示, 我们使用特征提取模块, 离线提取问题库中所有问题的特征向量. 然后将问题库中的所有问题, 答案, 和问题的特征向量保存到数据库中. 然后当用于进行提问时, 我们首先提取问题的特征向量, 然后在数据库中检索与问题最为相似的问题并返回对应的答案. 在实际应用场景中我们可以根据目标问题与返回结果的特征距离来判断用户的提问是否存在于我们的问题库中.

**搜索结果**

问题	答案	距离(越小越相似)
ECS实例转移和更换公网IP地址有什么限制?	转移和更换ECS实例的公网IP地址有以下限制: 经典网络 不支持公网IP跨账号转移. 实例创建后6小时内, 允许更换公网IP地址, 最多可以更换3次. 6小时后, 不能再更换公网IP地址。具体操作, 请参见修改公网IP地址。 使用高防IP业务后, 最多支持更换10次源站IP。具体操作, 请参见DDoS高防IP用户指南的更换ECS IP。专有网络 不支持公网IP及EIP跨账号转移。如果实例没有分配公网IP, 您可以为实例绑定EIP。您可以通过绑定新的EIP更换公网IP。如果实例已经分配了公网IP: 实例创建后6小时内, 允许更换公网IP地址, 最多可以更换3次。具体操作, 请参见修改公网IP地址。通过将公网IP转换为EIP的方式更换IP。具体操作, 请参见专有网络类型ECS公网IP转为弹性公网IP。如问题还未解决, 请提交工单。	72.0346221923828
除了更换公网IP地址, 我还可以更换私网IP地址吗?	专有网络类型ECS实例: 可以。详细步骤, 请参见修改私有IP地址。经典网络类型ECS实例: 不可以。	76.0608291625977

返回个数:10 刷新

搜索问题: 公网ip能否更换? 搜索

关于阿里云ECS问题  
可以尝试以下问题  
忘记密码怎么办?  
怎么计费?  
如何升配?  
怎么开发票?  
实例无法连接怎么办?  
如何选择可用区?  
重启失败怎么办?  
怎么配置安全组

提问"公网ip能否更换?"

AnalyticDB 阿里云分析型数据库非结构化数据分析 演示

产品页面

简介

AnalyticDB

演示

图像视频

语音文本

AnalyticDB Demos

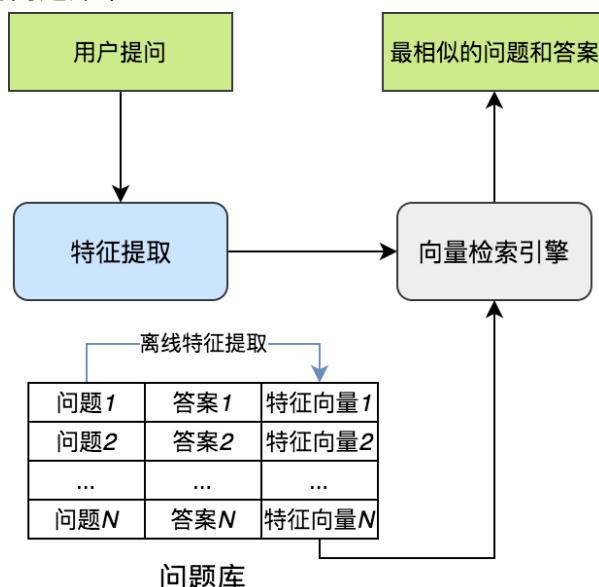
的语义. 第二个模块是向量检索引擎, 可以在海量的特征向量中匹配与目标特征最相似的记录. 如下图所示, 我们使用特征提取模块, 离线提取问题库中所有问题的特征向量. 然后将问题库中的所有问题, 答案, 和问题的特征向量保存到数据库中. 然后当用于进行提问时, 我们首先提取问题的特征向量, 然后在数据库中检索与问题最为相似的问题并返回对应的答案. 在实际应用场景中我们可以根据目标问题与返回结果的特征距离来判断用户的提问是否存在与我们的问题库中.

搜索结果		
问题	答案	距离(越小越相似)
SCU可以退款吗?	发生以下情况时, 您可以申请退款: 购买存储容量单位包SCU后, 五天内可申请无理由退款。说明 每个阿里云账号每年只有一次五天无理由退款SCU的机会, 即每个账号每年最多退款一次, 可退还的上限是一个存储容量单位包. 其他情况, 请提交工单联系阿里云客服处理。退款时会扣除掉已经消费的金额, 已使用的代金券等优惠折扣不退还。更多详情, 请参见退款规则及退款流程。	64.8935012817383
是否支持5天内无理由退款?	支持。更多详情请参见退款规则及退款流程。	83.5105438232422
退款时, 费用怎么结算?	如果可以退款, 镜像供应商将按照您的实际使用情况退款。	92.0893478393555
购买ECS实例后第一时间提出了退	如果您的ECS实例未满足五天无理由退款规则, 会按照剩余时间退款, 规则详情请参见退款规则及退款流程。	99.8048629760742
购买ECS实例后第一时间提出了退	发生以下情况时, 您可以申请退款: 购买预留实例券后, 五天内可申请无理由退款。每个账号每年只有一次无	

提问"能不能退款?"

### 3. 问答系统原理

上一节展示的单轮问答系统, 主要依赖两个关键模块. 首先是文本转特征模块, 通过基于海量文本数据训练的深度学习模型我们可以将不同长度的文本转化成固定长度的特征向量. 这个特征向量描述了文本的语义. 第二个模块是向量检索引擎, 可以在海量的特征向量中匹配与目标特征最相似的记录. 如下图所示, 我们使用特征提取模块, 离线提取问题库中所有问题的特征向量. 然后将问题库中的所有问题, 答案, 和问题的特征向量保存到数据库中. 然后当用于进行提问时, 我们首先提取问题的特征向量, 然后在数据库中检索与问题最为相似的问题并返回对应的答案. 在实际应用场景中我们可以根据目标问题与返回结果的特征距离来判断用户的提问是否存在与我们的问题库中.



# 4. 如何使用AnalyticDB搭建单轮问答系统

下面我们来介绍如何使用AnalyticDB来实现前两章节介绍的单轮问答系统. AnalyticDB提供了文本的特征向量提取模块和高效的向量检索引擎, 我们可以使用AnalyticDB轻松地实现一套单轮问答系统.

## 4.0 创建插件

分别创建AnalyticDB的非结构化分析插件OpenAnalytic和向量检索插件fastann

```
1 CREATE EXTENSION IF NOT EXISTS open_analytic;
2 CREATE EXTENSION IF NOT EXISTS fastann;
```

## 4.1 建表

我们可以使用如下sql语句建表, 表中保存了图片的名称, 二进制文件和图像的特征向量, 商品的类目和商品的详细介绍.(用户也可以将图片的文件保存至阿里云的OSS对象存储服务上.这里不做详细介绍)

```
1 CREATE TABLE qa_table (
2     question TEXT NOT NULL, # 问题
3     answer BYTEA NOT NULL, # 答案
4     feature REAL[] NOT NULL, # 问题特征向量
5     PRIMARY KEY (question)
6 );
```

## 4.2 建索引

构建ANN索引可以高效的进行以图搜图的向量检索

```
1 CREATE INDEX qa_table_index
2 ON qa_table USING ann (feature) WITH (dim = 768);
```

## 4.3 创建特征提取算法pipeline

通过以下sql, 我们可以在数据库中创建文本特征提取的算法模型.

```
1 SELECT open_analytic.pipeline_create('text_feature_extractor');
```

## 4.4 提取文本特征向量

通过以下sql可以使用4.3创建的pipeline. 这个UDF的输入是pipeline名称和目标文本. 输出是一个包含文本特征向量的的JSON.

```
1 # 通过图像文件识别
2 SELECT open_analytic.pipeline_run_dist_random('text_feature_extractor',
3                                                 <目标文本>);
```

## 4.5 问题库导入AnalyticDB

在获取图像特征, 属性以后以后, 可以将图像数据导入4.5创建的 item\_search\_table表中.

```
1 INSERT INTO qa_table VALUES (<问题>, <答案>, <问题特征向量>);
```

## 4.6 搜索最相似的问题

通过以下sql可以检索与查询图片向量最相似且符合商品类别, 属性关键词的前10条记录. 然后我们可以根据距离判断用户所提出的问题是否已经保存在我们的问题库中.

```
1 SELECT question, answer, l2_distance(feature, <问题特征向量>)
2 ORDER BY feature <-> <问题特征向量>
3 LIMIT 10;
```

## 5. AnalyticDB介绍

分析型数据库(AalyticDB)是阿里云上的一种高并发低延时的PB级实时数据仓库，可以毫秒级针对万亿级数据进行即时的多维分析透视和业务探索。AnalyticDB for MySQL 全面兼容MySQL协议以及SQL:2003语法标准, AnalyticDB forPostgreSQL 支持标准 SQL:2003，高度兼容 Oracle 语法生态。

向量检索和非结构化数据分析是AnalyticDB的进阶功能。目前两款产品都包含向量检索功能，可以支持人脸，人体，车辆等的相似查询和推荐系统。AnalyticDB在真实应用场景中可以支持10亿级别的向量数据的查询，毫秒级别的响应时间。AnalyticDB已经在多个城市的重大项目中大规模部署。

在一般的包含向量检索的应用系统中，通常开发者会使用向量检索引擎(例如Faiss)来存储向量数据，然后使用关系型数据库存储结构化数据。在查询时也需要交替查询两个系统，这种方案会有额外的开发工作并且性能也不是最优。AnalyticDB支持结构化数据和非结构化数据(向量)的检索，仅仅使用SQL接口就可以快速的搭建起以图搜图或者图片+结构化数据混合检索等功能。AnalyticDB的优化器在混合检索场景中会根据数据的分布和查询的条件选择最优的执行计划，在保证召回的同时，得到最优的性能。AnalyticDB向量版采用了多项创新性技术，这些技术在我们的论文 *AnalyticDB-V: A Hybrid Analytical Engine Towards Query Fusion for Structured and Unstructured Data* 中有详细介绍。目前论文已经被数据库三大顶会之一的VLDB接受，具有技术领先性。

结构化信息+非结构化信息(图片)混合检索在实际应用中被广泛使用的。例如人脸门禁系统被部署在多个小区时，我们使用一张表存储了所有小区的人脸特征，在人脸检索时我们只需要检索当前小区的人脸特征。在这种情况下，使用AnalyticDB我们只需要在SQL中增加where 小区名 ='xxx' 就可以轻易实现。

AnalyticDB同时提供了先进的图像文本分析算法，能够提取非结构化数据的特征和标签，用户仅仅需要使用SQL就可以完成图像文本内容的分析。

更多信息可以参考文章: <https://zhuanlan.zhihu.com/p/82284704>

## 6. 结语

本文介绍了如何使用AnalyticDB来搭建单轮问答系统。演示系统的源码可以在 <https://github.com/aliyun/alibabacloud-AnalyticDB-python-demo-AI> 下载。AnalyticDB还支持其他多种多样人工智能算法如目标检测，声纹识别，基因识别等等。想了解更多请扫码加入AnalyticDB向量版交流群。



 扫一扫群二维码，立刻加入该群。