# ytluck—For you

Home About Me Big Data Data Mining

Items

OS

Program

# 携程模型部署-小记

# 模型的部署

模型部署是将预测解决方案有效地应用于实践的过程,

这项任务一般由与建模过程相分离的应用程序完成

服务部署主要包括构建、组装、测试和部署服务。于创建、发布、发现和调用服务的候选技术 部署说明

部署环境通常与预测解决方案要监控的系统和流程紧密集成

需要经常对模型进行更新, 所以必须关注互操作性和开放式标准需求。

预见性维护--防御型编程

部署类型

广告和推荐级别的部署

业务模型级别的部署

语言----R Python Scala Java

关注:模型以及模型的数据流

Published: 2017-05-13 20:05:00

By <u>ytwan</u>

In Program.

tags: Java

# 业务模型的上线流程部署场景和对应的方案

- 1. 对于离线的预测以及分析
  - 01. 直接采用Rscript model.r 或者 Python model.py的方式来预测 使用这种工作方式需要一个调度工具,可以用shell的Crontab做定时任务,当然如果 公司有统一的调度工具,则使用其规定的调度工具吧
  - 02. 写一下R的预测脚本,比如predict.R,是你的主预测的模型; 然后用shell封装成xx.sh,比如predict.sh,shell里面调用模型,存储数据 最后用Crontab来进行调度
  - 03. 使用Spark模型上线---将数据存储到数据库中--SQL
- 2. 实时响应--小批量数据
  - 01. Rserve---服务器上部署好R环境和安装好Rserve, 然后用JAVA写好SOA接口,调用Rserve来进行预测

提供: Rserve上线的文件只需要提供两个:

- 模型结果文件 (XX. Rdata)
- 预测函数 (Pred.R)
- 02. R model转换PMML的方式
- 01. Python----httpserver---部署好Python环境

然后用JAVA写好SOA接口,调用-httpserver-来进行预测

02. Python ---模型转换成PMML上线方法

- 3. 实时响应-大批量数据
  - 01. <1>训练模型--<2>将训练好的模型转换成PMML--<3>把模型封装成一个类,用Java调用这个类来预测需要提供模型文件和预测主类; --模型封装成类

Spark--放在spark集群,不脱离spark环境--打jar包

说明

2017/12/23 携程模型部署-小记

### 模型数据:

数据存放两个批次的数据,用来预防无法取到最新的数据,则用上一批次的数据来进行填充;

容灾---接口都需要有容灾机制

跟踪输入输出数据-log的查看和性能方面的监控--校验数据和检验API的稳定性

设置不同的key、不同的更新频次 和 不同的过期时间

数据流之间如何耦合和解耦,如何进行容灾、平台稳定、可用也是需要重点考虑的。

高性能、跨域

PMML 预测模型标记语言 (Predictive Model Markup Language, PMML)

是一种可以呈现预测分析模型的事实标准语言。Java 的JPMML ----生成pmm1文件

如果你的model training 和 prediction 都是同一种语言(或者工具),其实你没必要使用PMML。

因为无论是R, Python或者SparkML都有自己标准的输出格式可以直接Load。

个人觉得PMML的使用场景主要是prediction的环境和training环境不在一起。

离线训练模型之后,直接在prediction环境读取PMML获得训练后的模型,

而且没必要在prediction环境装相应的工具(R/Python/Spark等)。

通常是因为生产环境安装这些环境不方便。

SOA--- Service-Oriented Architecture (面向服务的体系结构)

接口---接口是干什么的、需要什么参数、可能返回的结果以及可能抛出的错误

#### 思维意识

-架构思维 ---如何进行架构设计是最合理的,

你要考虑是否需要字段冗余、行存储还是列存储、字段如何扩展最有效, 热数据和冷数据如何拆

数据质量---来源可靠性以及检验

危机意识--会数据分析、提炼结论、洞察数据背后的原因,了解业务,懂算法。

## 参考:

何为 PMML? https://www.ibm.com/developerworks/cn/opensource/ind-PMML1/

链接: https://www.zhihu.com/question/29793218/answer/137087671

干货 | 机器学习算法线上部署方法 http://mp.weixin.qq.com/s/KxClgDgPBy2fYA 4bCVebA

blogroll			social
Pelican	Python.org	yt	ye

Proudly powered by Pelican, which takes great advantage of Python.

The theme is by **Smashing Magazine**, thanks!