

携程模型部署-小记

模型的部署

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

By [ytwan](#)In [Program](#).tags: [Java](#)

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

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

