特征提取模块搭建

* 环境部署
* 全量历史生成
* 增量部署

模型机器：

SSP: [work@180.76.162.30](mailto:work@180.76.162.30)

BES: work@180.76.135.242

离线环境

内网IP：192.168.0.5

外网IP：180.76.135.242

* 环境部署

代码：<https://github.com/bayescom/bayesalgo>

1. 将adfea\_control复制到需要部署的目录，并在adfea\_control下mkdir bin,conf,script,history\_adfea、model。
2. 将adfea/bin、ftrl/bin下的文件都拷贝到adfea\_control/bin下。
3. 将adfea/conf下的feature list配置文件拷贝到adfea\_control/conf下。

* 全量历史生成

1. 生成历史：使用adfea模块将历史的shitu\_log提取特征，假设历史的shitu\_log合并后文件名为shitu\_log\_dmssp，提取的特征文件名rain\_dmssp\_ins\_exchange。将这2个文件mv到history\_adfea下。
2. 生成done\_file：done\_path下生成done\_file文件：格式如下：

|  |  |
| --- | --- |
| 全量的最后的时间（包含） | end\_time\_stamp=2016091811 |
| 全量提取的特征文件 | adfea\_result\_exchange=./history\_adfea/train\_shaibo\_ins |
| 全量的视图日志 | shitu\_log=./history\_adfea/shitu\_batch\_shaibo |

* 增量部署

1. cd adfea\_control，编辑启动脚本script/adfea\_train\_dmssp.sh,主要修改

|  |  |
| --- | --- |
| batch\_adfea\_done="${done\_path}/adfea\_dmssp\_done.last" | done\_file文件 |
| shitu\_path=/home/work/xuyichen/shitu\_log2.0 | 视图日志来源 |
| adfea\_run\_temple="./conf/run\_temple.conf" | 配置scheme、hash\_num的adfea提取配置 |
| train\_run\_temple="./conf/ftrl\_temple\_${exchange}.conf" | 配置模型训练的参数 |
| bidder\_machine\_new="192.168.13.177" | 模型推送的线上机器 |

.