京东Recommender简介

# 概述

本文档旨在简要介绍京东推荐服务中recommender的功能、组成及工作流程和主要的开发需求。

# Recommender简介

## Recommender功能

Recommender能接受一个推荐请求，解析推荐请求中的参数，获取其中用户行为、偏好等信息，并根据解析出的信息通过数据召回接口召回推荐数据并解析，并根据一定的逻辑进行过滤、排序等操作，最后组装推荐结果并返回。

## Recommender框图

一个典型的Recommender框图如下。



## Recommender模块

### 调用

Recommender现采用http请求调用。用户请求前端服务器后，如果有推荐需求，则前端服务器会通过http请求调用推荐服务。推荐配置（现为divineradmin.jd.com）中有部分配置逻辑（用于实验、分流、服务器参数配置等），最终推荐调用传入的推荐请求会包含有http请求和配置信息。

一般推荐请求的入口为RecommendResult getRecommendations(RecommendRequest request)方法。

### 参数解析

推荐请求中会有target, cookie及配置等信息，从中能解析出用户PIN，IP地址所在地区等参数。

解析方式见RecommendRequest类及具体Recommender和前端服务器的接口协议。

### 获取用户信息

用户信息目前分为用户行为和用户偏好。

用户行为是指：订单、点击、收藏、搜索，分别对应Order, Click, Focus, Search类。前三类保存有用户行为的SKU，搜索保存有用户搜索词。通过调用相关接口，能返回指定天数内用户行为信息。

用户偏好是指：用户的CID3权重表、产品词、品牌词。用户偏好信息通过调用Zeus系统接口（架构组提供）获得。

## 数据召回

通过不同的用户信息，调用相应的接口能够召回不同的离线数据（例如通过SKU召回、通过品牌词召回）。

离线数据接口视具体Recommender而不同，部分通过调用已有接口完成，部分需要自行搭建。

## 离线召回数据生成

使用不同的工具（HIVE, python, R等）和不同的算法，能够召回不同的离线数据。

具体离线数据生成需要对所在推荐位有深入理解，结合业务开发相应模型，才能获取满足业务需求的离线数据。

该部分属于Recommender线下流程。

## 召回数据处理

通过调用数据召回接口召回数据后，根据其格式（如JSON）进行解析，具体解析逻辑视具体召回数据接口而定。

解析得到数据，还需要根据当前推荐位进行推荐结果过滤（如用户已购买、已有相关展示类目）、根据权重排序等操作，得到合适的推荐结果，并将其写入到RecommendResult类对象中，并返回该推荐结果到前端用于展示。

# 开发需求

本节介绍一个推荐位的主要开发需求。具体推荐位可能有其特殊要求。

### 参数解析

和前端协定接口，配置divineradmin.jd.com中参数，并根据接口和配置进行参数解析。

### 获取用户信息

不同用户信息可召回不同的数据。因此，需要理解具体推荐位业务，挑选合适的信息用于召回。

### 数据召回接口

根据推荐位挑选合适的召回接口。必要时需要自行搭建。

### 离线数据生成

需要理解该推荐位业务，选用合适的算法，从HIVE, JSHOP等数据源中提取数据，并利用这些数据训练算法模型得到离线数据。

该部分属于Recommender线下流程。

### 结果解析、过滤、组装

根据挑选的接口返回数据接口进行解析。

不同推荐位有不同的推荐逻辑（如部分推荐位在某些时期会要求不展示店铺号大于某个数的活动），需要根据此些业务需求进行过滤。

同时，召回数据可能未必满足当前推荐逻辑，因此也需要进行过滤。

如果召回数据不够，还需要调用冷启接口获取冷启数据。