



金门开发介绍

IMPL 甘璐

Outline

- 金门概述
- 框架调度
- 策略开发



金门概述

- 用户意图推荐平台

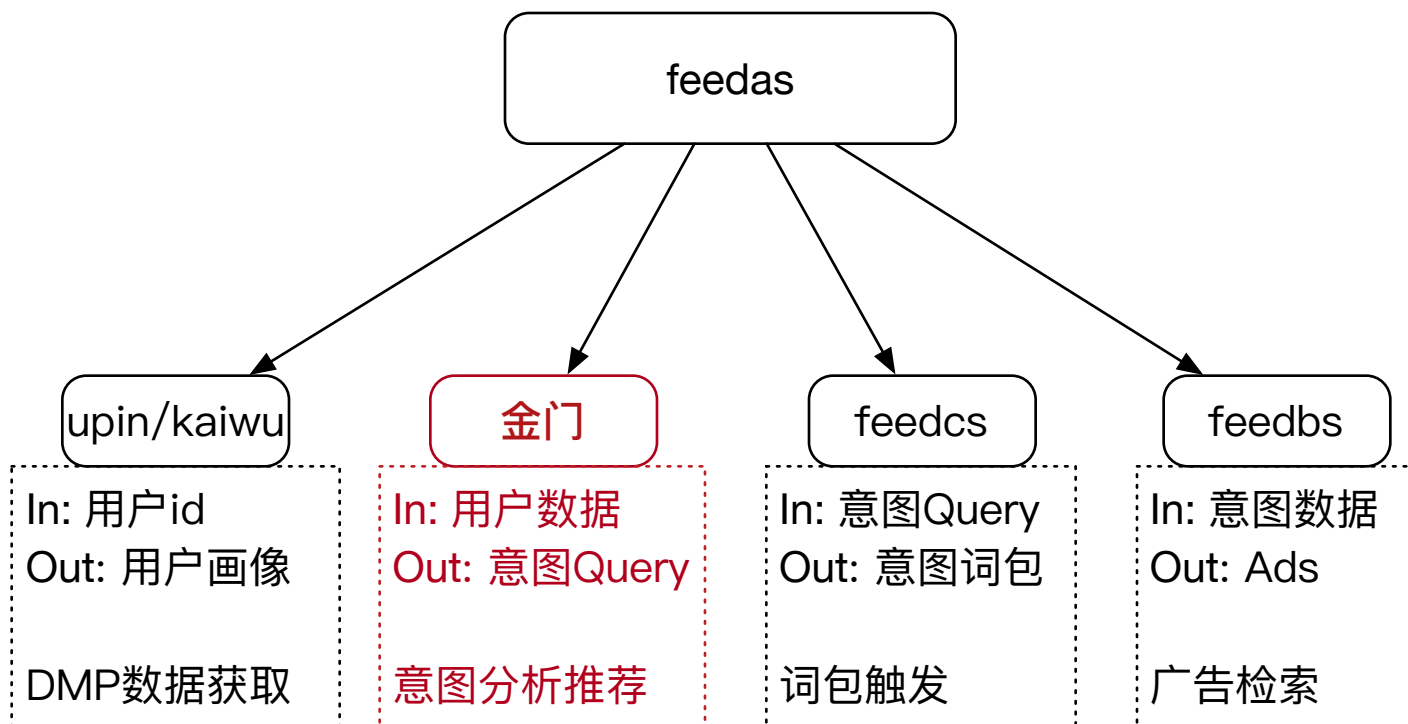
- 一期根据用户的个性化数据，推荐意图Query
- 目标是打通用户原始数据，离线挖掘策略到在线检索整个通路

- 和feedcs的区别

- 技术上通用，但是金门和广告主业务无关



Feed中的金门

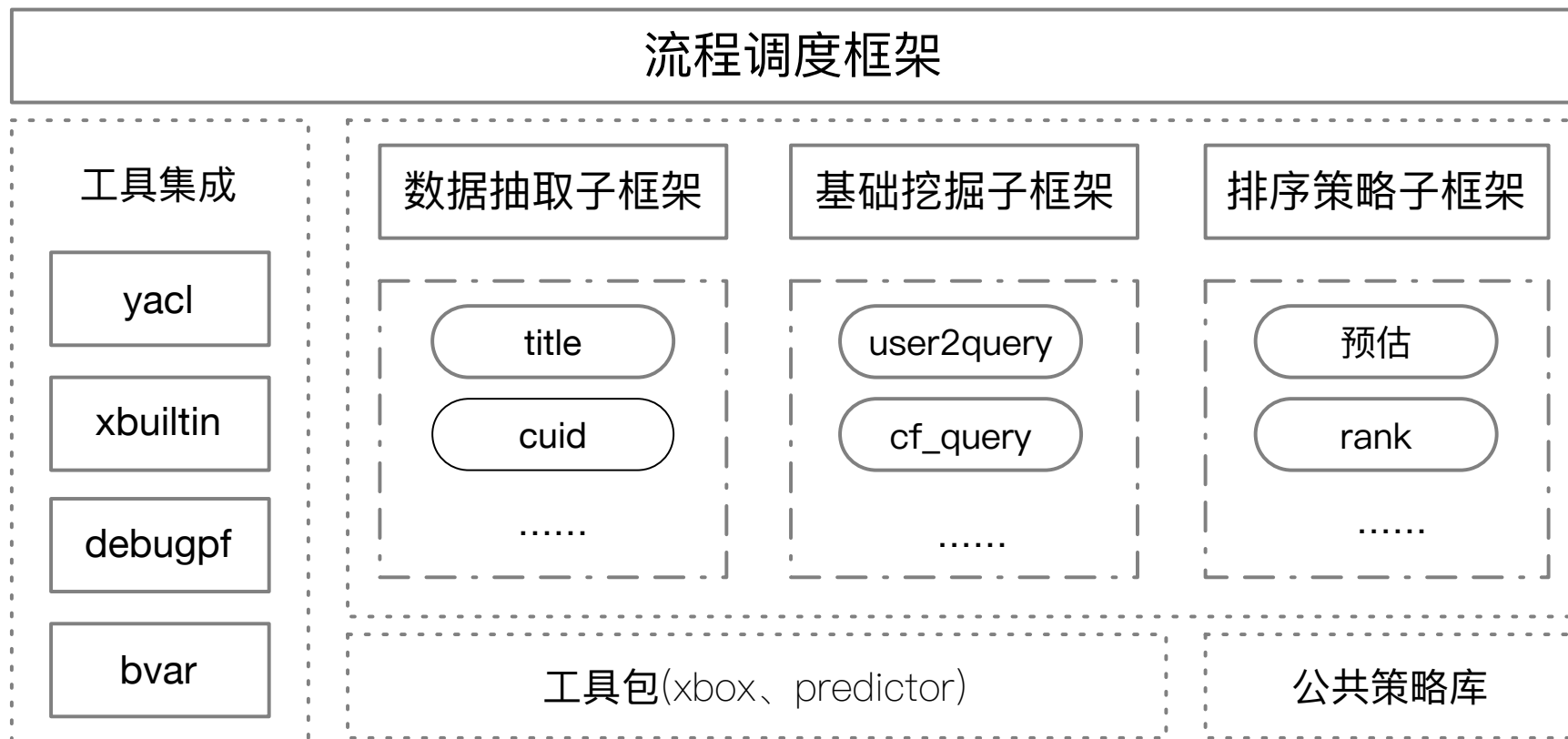


Outline

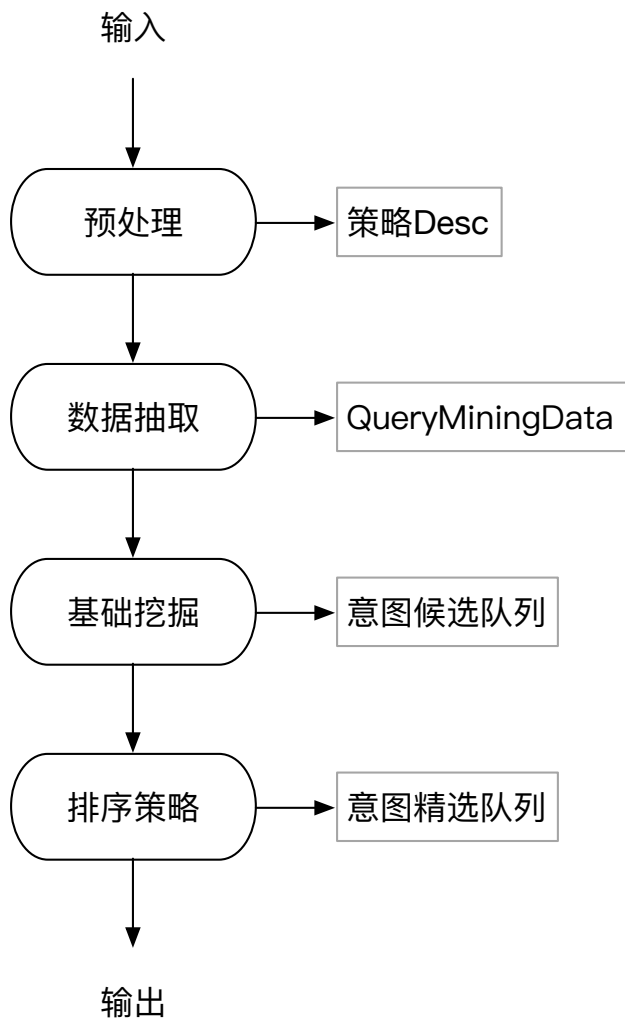
- 金门概述
- 框架调度
- 策略开发



整体框架



整体流程调度

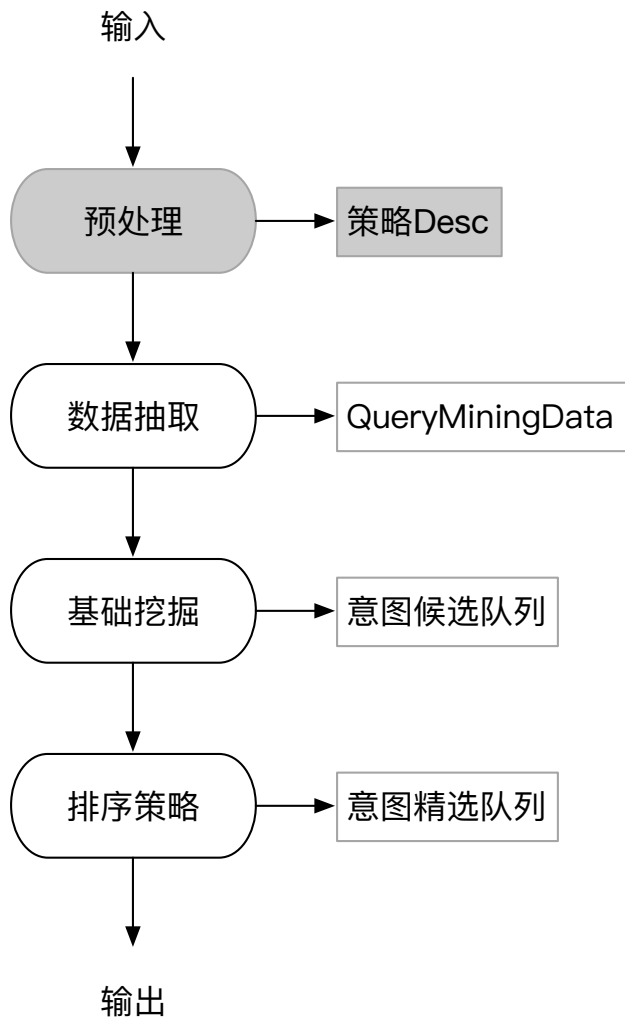


输入 用户数据（请求信息、upin、kaiwu）

输出 意图query

- ① 预处理产出策略Desc数据
- ② 数据抽取产出QueryMiningData
- ③ 基础挖掘产出候选意图QueryList
- ④ 排序策略产出精选意图QueryList

流程调度 - 预处理



输入 product + srcid

输出 策略Desc数据

- ① 依赖的个性化数据
- ② 要执行的挖掘branch
- ③ 要执行的策略strategy

- 依次找寻srcid对应的配置，以第一个匹配的为准

- 没找到srcid会报FATAL

流程调度 - 预处理

依赖的个性化数据(1)

[@branch]

branch_name: session_query

plugin_name: dmp_query

[.plugin_params]

@key_type: upin_fcr_ss_info

@key_type: upin_fcr_gs_wise_ss_info

@key_type: upin_fcr_gs_pc_ss_info

依赖的个性化数据(2)

[@branch]

branch_name: intent_model

plugin_name: intent_xbox_query

[.plugin_params]

[..@intent_xbox_params]

table_name: user2query

table_type: 2

key_type: baiduid

truncate_num: 30

流程调度 - 预处理

依赖的branch

[@query_mining_desc]

product: feedas

src_id: 1092

@branch_name: original_query

@branch_name: session_query

@branch_name: query_profile

@branch_name: intent_model

@branch_name: sequence_word

依赖的Strategy

[@query_strategy_desc]

product: feedas

src_id: 1092

@strategy_name: query_cpm_trade

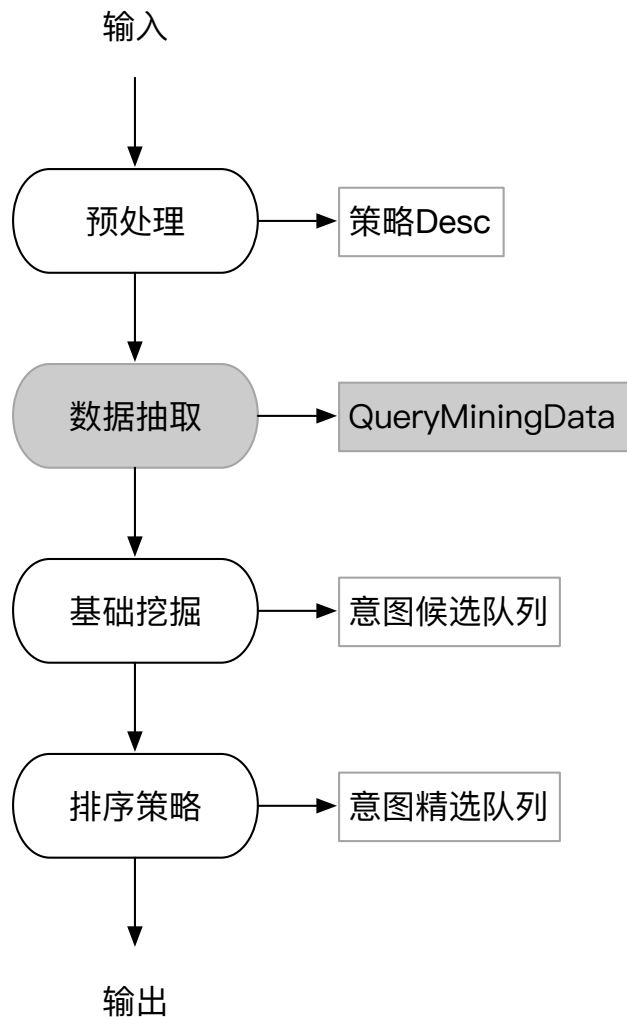
@strategy_name: feed_predictor

@strategy_name: feed_compute

@strategy_name: query_sort

@strategy_name: query_filter

流程调度 - 数据抽取



输入 请求数据 + 依赖的key

输出 QueryMiningData

- ① 配置驱动
- ② 以组件为单位进行数据抽取
- ③ 一个key对应一个抽取plugin
- ④ 抽取的数据输出到 QueryMiningData(PB)
- ⑤ 后续个性化数据的使用屏蔽具体数据结构
- ⑥ `Utils::get_querymining_data(key)`

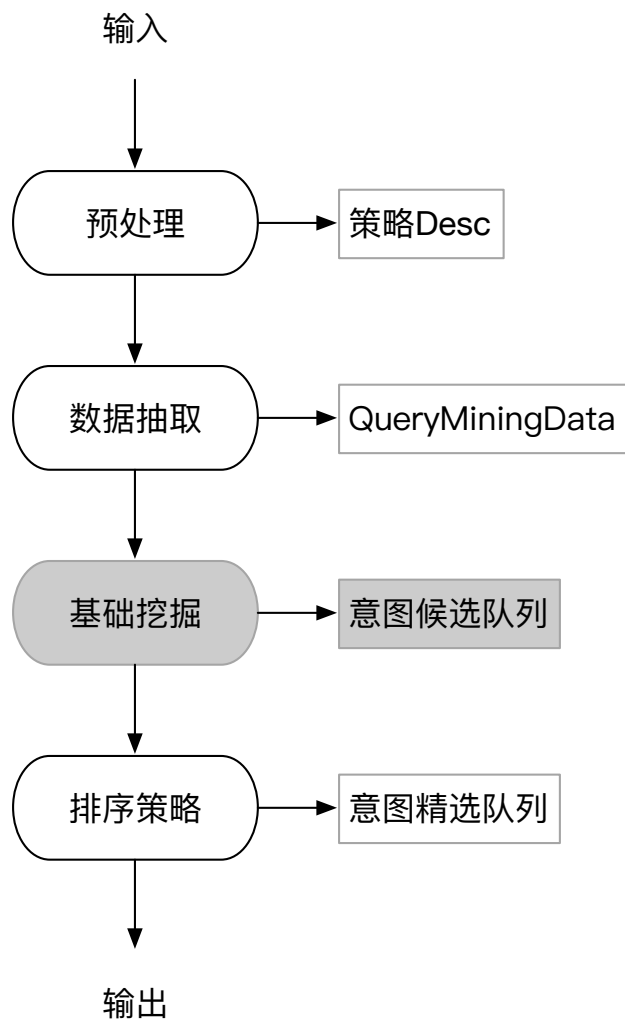
流程调度 - 数据抽取

QueryMiningData (PB)

- 统一接口，全部为repeated bytes
- 数据分为三类
 - 请求携带数据
 - upin数据
 - kaiwu数据
- PB方便利用反射机制实现配置驱动

```
message QueryMiningData {  
    // req  
    repeated bytes baiduid = 1;  
    repeated bytes original_query = 4;  
  
    // upin  
    repeated bytes upin_fcr_ss_info = 100;  
    repeated bytes upin_intent_query = 102;  
  
    // kaiwu  
    repeated bytes kaiwu_partial_query =  
200;  
    repeated bytes kaiwu_qp_intent = 201;  
};
```

流程调度 - 基础挖掘



输入 策略Branch + QueryMiningData

输出 意图候选Querylist

- ① 配置驱动, 组件复用
- ② DMP分支从QueryMiningData获取意图
- ③ IntentXbox分支Xbox结果中获取意图
- ④ Xbuiltin分支从词表中获取意图
- ⑤ ANN分支通过语义检索获取意图

流程调度 - 基础挖掘

branch = plugin + params

DmpPlugin

- 根据配置从QueryMiningData取数据

[@query_mining_desc]

product: feedas

src_id: 1092

@branch_name: original_query

@branch_name: session_query

[@branch]

branch_name: original_query

plugin_name: dmp_query

[.branch_params]

branch: 0

rank: 1

weight: 1.018

[.plugin_params]

@key_type: original_query

[@branch]

branch_name: session_query

plugin_name: dmp_query

[.branch_params]

branch: 0

rank: 1

weight: 1.018

[.plugin_params]

@key_type: upin_fcr_ss_info

流程调度 - 基础挖掘

branch = plugin + params

IntentXboxPlugin

- 根据配置请求xbox
- step1. 集中所有xbox请求, 填充请求数据
- step2. 一并发送, 一并接收
- step3. 解析请求, Query加入候选队列

[@query_mining_desc]

product: feedas

src_id: 1092

@branch_name: intent_model

[@branch]

branch_name: intent_model

plugin_name: intent_xbox_query

[.plugin_params]

[..**@intent_xbox_params**]

table_name: user2query

table_type: 2

key_type: baiduid

truncate_num: 30

[..**@intent_xbox_params**]

table_name: user2query

table_type: 2

key_type: cuid

truncate_num: 30

流程调度 - 基础挖掘

branch = plugin + params

HotQuery & TiebaQuery

- 从xbuiltin词表获取Query

[@query_mining_desc]

product: feedas

src_id: 1092

@branch_name: hot_query

@branch_name: bar_name

[@branch]

branch_name: hot_query

plugin_name: hot_query

enable: 1

[.branch_params]

branch: 4

rank: 8

weight: 0.1

max_num: 1

[@branch]

branch_name: bar_name

plugin_name: tieba_query

enable: 1

[.branch_params]

branch: 7

rank: 2

weight: 1

max_num: 1

流程调度 - 基础挖掘

branch = plugin + params

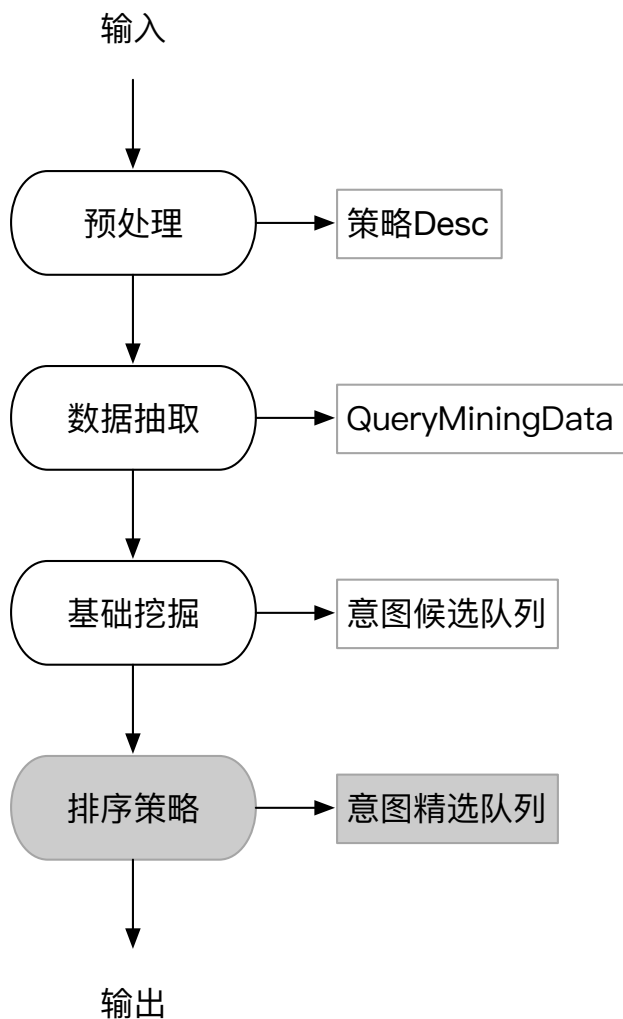
AnnSearch

- 语义检索
- 所需配置自行定义使用

```
[@query_mining_desc]
product: feedas
src_id: 1092
@branch_name: ann_search
```

```
[@branch]
branch_name: ann_search
plugin_name: ann_search
enable: 0
[.branch_params]
branch: 23
rank: 4
weight: 1.153
max_num: 5
[.plugin_params]
@key_type: feed_session
predictor_phase_name: feedsimq
max_key_num: 3
```

流程调度 - 排序策略



输入 QueryList + 策略Strategy

输出 精选意图QueryList

- ① 配置驱动，组件复用
- ② 基于QueryList进行基础策略
- ③ 依次执行计算、过滤、预估、排序

流程调度 - 排序策略

strategy = plugin + params

predictor

- 根据指定的观星阶段，执行Q预估

compute

- 根据配置的参数，进行score计算

[@query_strategy_desc]

product: feedas

src_id: 1092

@strategy_name: feed_predictor

@strategy_name: feed_compute

@strategy_name: query_sort

@strategy_name: query_filter

[@strategy]

strategy_name: feed_predictor

plugin_name: predictor

[.strategy_params]

predictor_phase_name: query_strategy

[@strategy]

strategy_name: feed_compute

plugin_name: compute

[.strategy_params]

weight_cpm: 32

weight_intentq: 1000

t_cpm: 1.0

Outline

- 金门概述
- 框架调度
- 策略开发



基础工具包

Xbox并行请求框架

- 封装对xbox请求的管理
- 自行定义请求数据，结果解析
- 多个xbox并行请求

不论是单次请求，还是多个请求，强烈建议使用该工具包！

Predictor工具包

- 封装对观星的请求
- 自定义访问条件，特征数据，预估数据，Q解析
- 多个观星并行请求

引入新个性化数据

需求

接入kaiwu的wise_query数据

升级点

- ① 升级interface（可选）
- ② 新增数据抽取组件wise_query
- ③ 数据添加至QueryMiningData

使用

DmpPlugin指定wise_query

配置驱动

```
[@branch]  
branch_name: kaiwu_wise_query  
plugin_name: dmp_query  
[.plugin_params]  
@key_type: wise_query
```

新增xbox词表

需求

新增xbox策略词表的访问

复用已有xbox交互proto

使用

配置新增xbox表

配置驱动

[@branch]

branch_name: intent_model

plugin_name: intent_xbox_query

[.plugin_params]

[..@intent_xbox_params]

table_name: **intent_xbox_name**

table_type: 2

key_type: baiduid

truncate_num: 30

新增xbox词表

需求

新增xbox策略词表的访问
但请求proto不复用

升级点

- ① 实现XboxRequest
- ② 注册到对应的Pool中

使用

配置新增xbox表、指定请求类型

配置驱动

```
[@branch]
branch_name: intent_model
plugin_name: intent_xbox_query
[.plugin_params]
[..@intent_xbox_params]
table_name: intent_xbox_name
request_type: intent_new
key_type: baiduid
truncate_num: 30
```


新增挖掘策略

需求

以ANN为例

新增ANN推荐策略

升级点

- ① 实现ANN策略Plugin
- ② 注册到对应的管理类

使用

配置使用该plugin，传入参数

配置驱动

```
[@branch]
branch_name: ann_search
plugin_name: ann_search
[.branch_params]
branch: 23
rank: 4
weight: 1.153
[.plugin_params]
@key_type: feed_session
predictor_phase_name: feedsimq
max_key_num: 3
```

新增过滤策略

需求

cpm过滤独立

升级点

- ① 实现cpm过滤组件
- ② 注册到对应的管理类

使用

指定参数定义该策略分支

配置驱动

```
[@strategy]
strategy_name: query_filter_by_cpm
plugin_name: cpm_filter
[.strategy_params]
cpm_thr: 0.99
```

实验上线

- 开发路径: baidu/im-goldengate-deploy/feedgoldengate
- 爱迪生模块名称: goldengate (待升级)
- prod上线目录: feedgoldengate
- 实验平台: 3个D平台
- 上线流程: RB nodiff观察2天
- goldengate-user群: 1594210



THANKS

