



**SDCC 2016**

**中国软件开发者大会**

SOFTWARE DEVELOPER CONFERENCE CHINA

# 微博商业数据挖掘方法

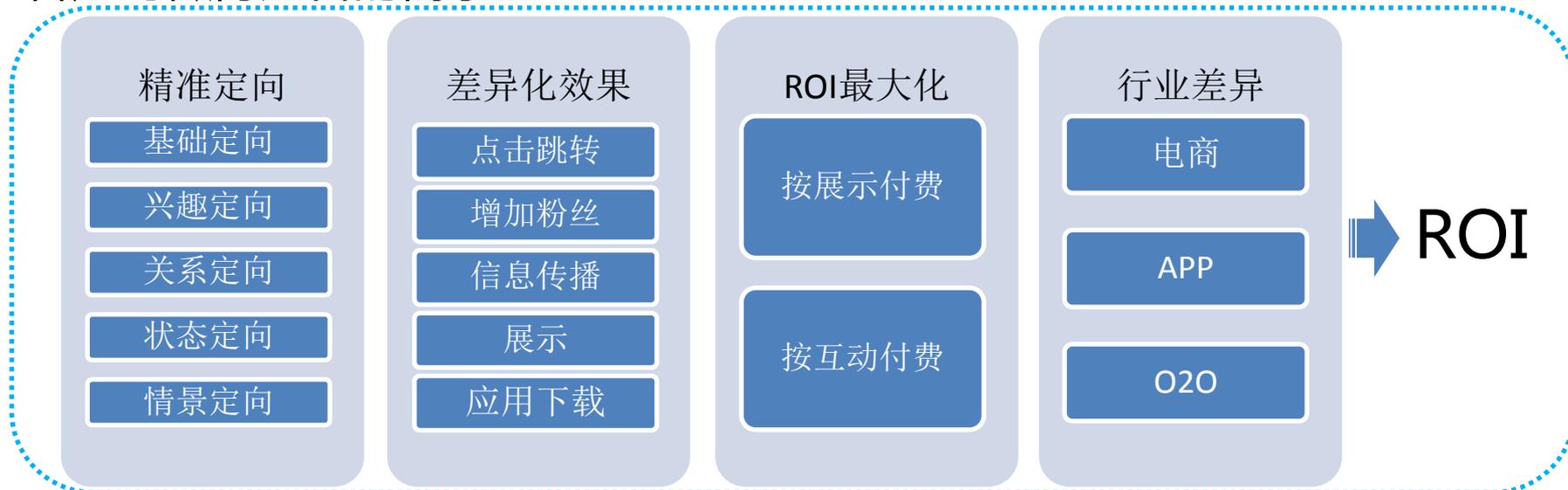
嘉宾：康乐

- 广告产品
- 数据体系
- 挖掘方法
- 总结

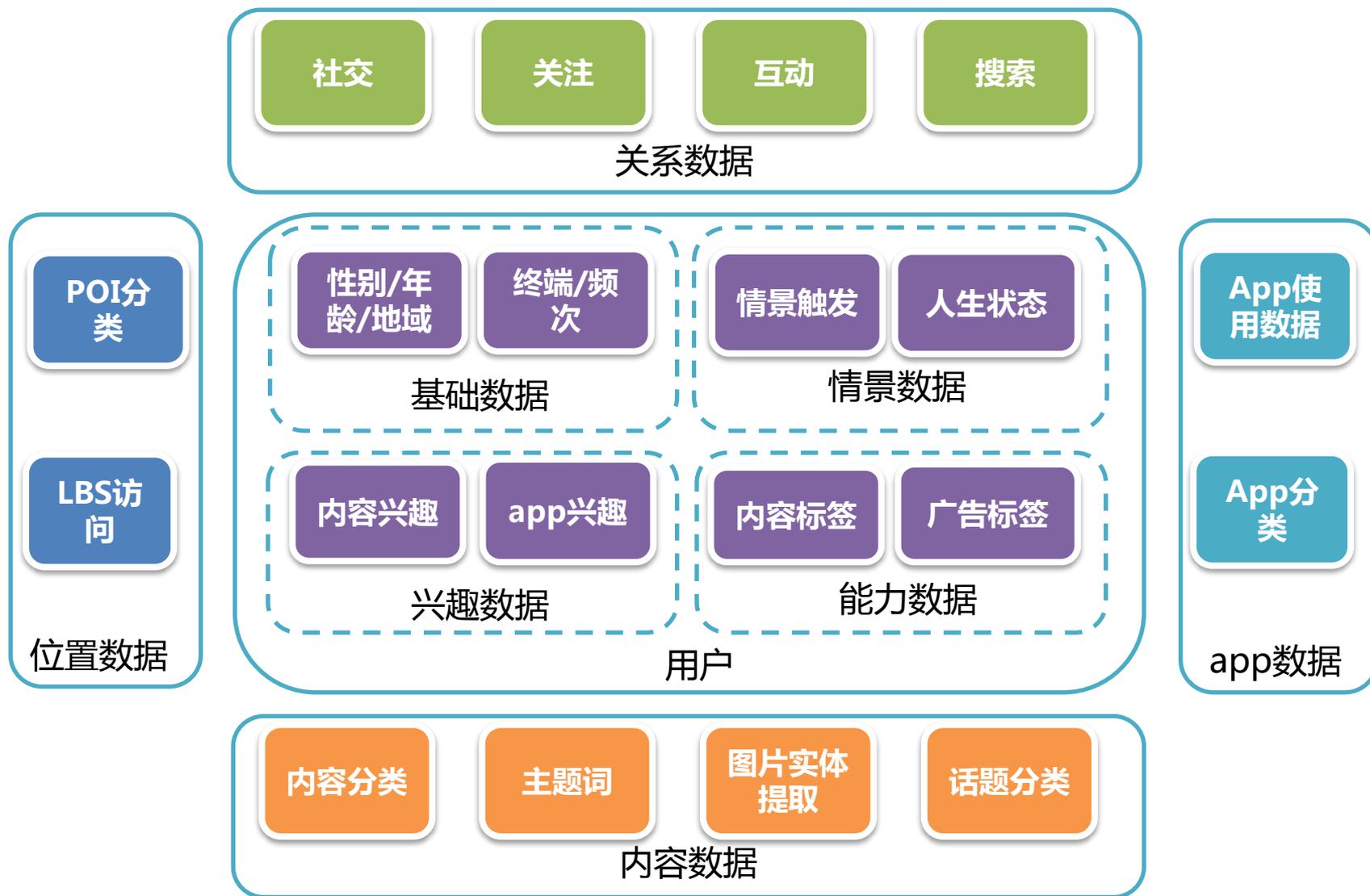
## 微博计算广告产品

产品	投放目标	特性
粉丝通	混合形式互动	基于Feed流
粉丝经济	混合形式互动	粉丝通lite版
应用家	应用下载，唤醒	基于Feed流并有特殊的广告模板
品牌速递	展示，跳转	传统tips和展示广告
WAX	混合形式互动	程序化投放

## 客户对微博广告的需求



- 广告产品
- 数据体系
- 挖掘方法
- 总结



求学状态

初中以下

高中

大学

研究生以上

婚恋状态

单身

恋爱未婚

已婚

育儿状态

准父母

婴幼儿父母

职业状态

求职中

白领

高管

房车状态

经济型车主

中档车车主

豪华车主

准备购房

旅行状态

正在旅行

刚旅行归来

高端商旅人士

常飞国外人士

中长期情景  
(人生状态)

### 挖掘层

位置情景

话题情景

微博互动情景

App使用情景

短期情景

### 情景引擎

LBS数据

APP使用数据

话题数据

微博互动

关注

用户信息

分组信息

分组信息

粉丝通

应用家

粉丝经济

WAX

基础数据

兴趣数据

情景数据

基础画像

用户内容兴趣

用户App兴趣

短期情景

中长期情景

基础画像修正

兴趣统计模型

内容兴趣预估

App兴趣定向

App召回预估

情景挖掘

内容分类体系

广告特征库

下载激活预估

App分类体系

App安装预估

情景引擎

相关性模型

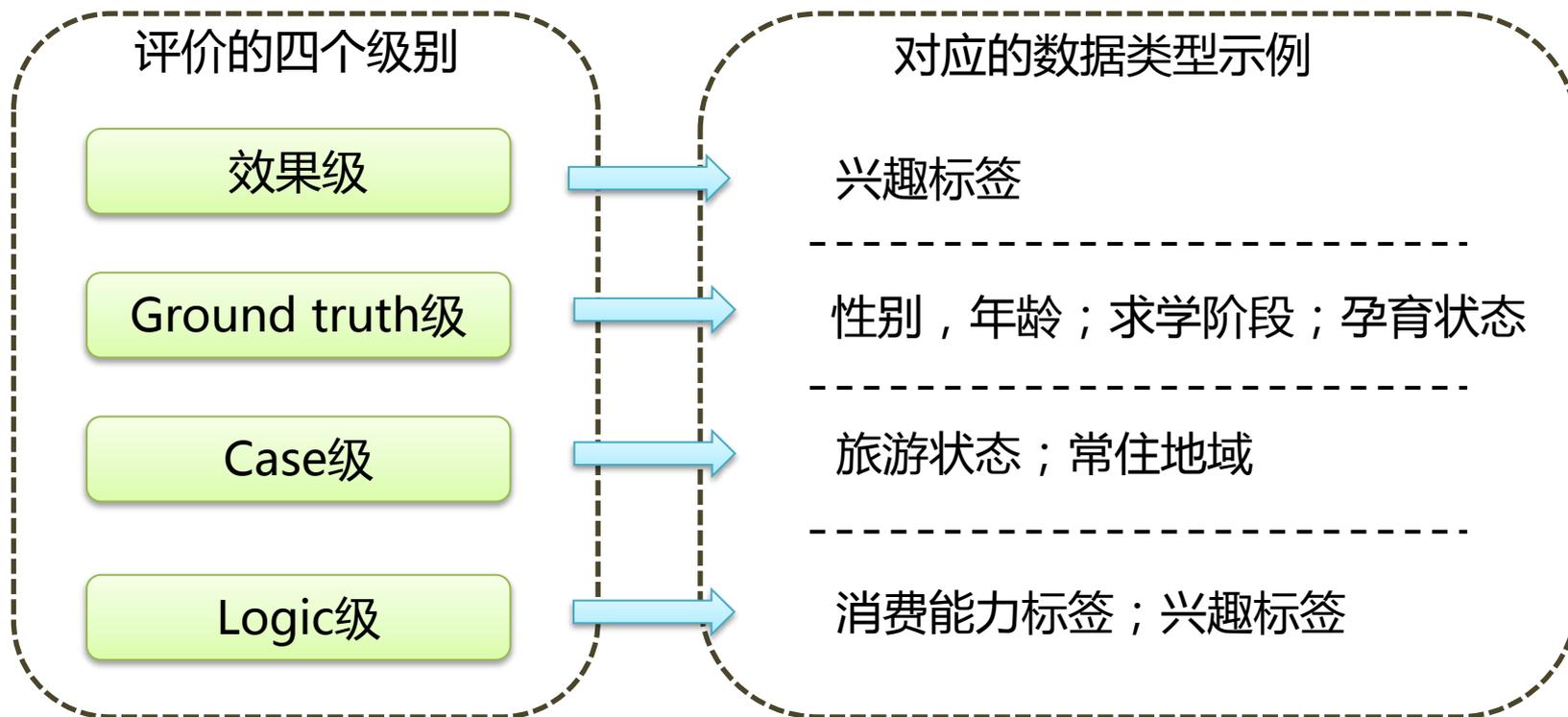
App安装数据挖掘

Look alike模型

数据仓库

数据合作

- 广告产品
- 数据体系
- 挖掘方法
- 总结

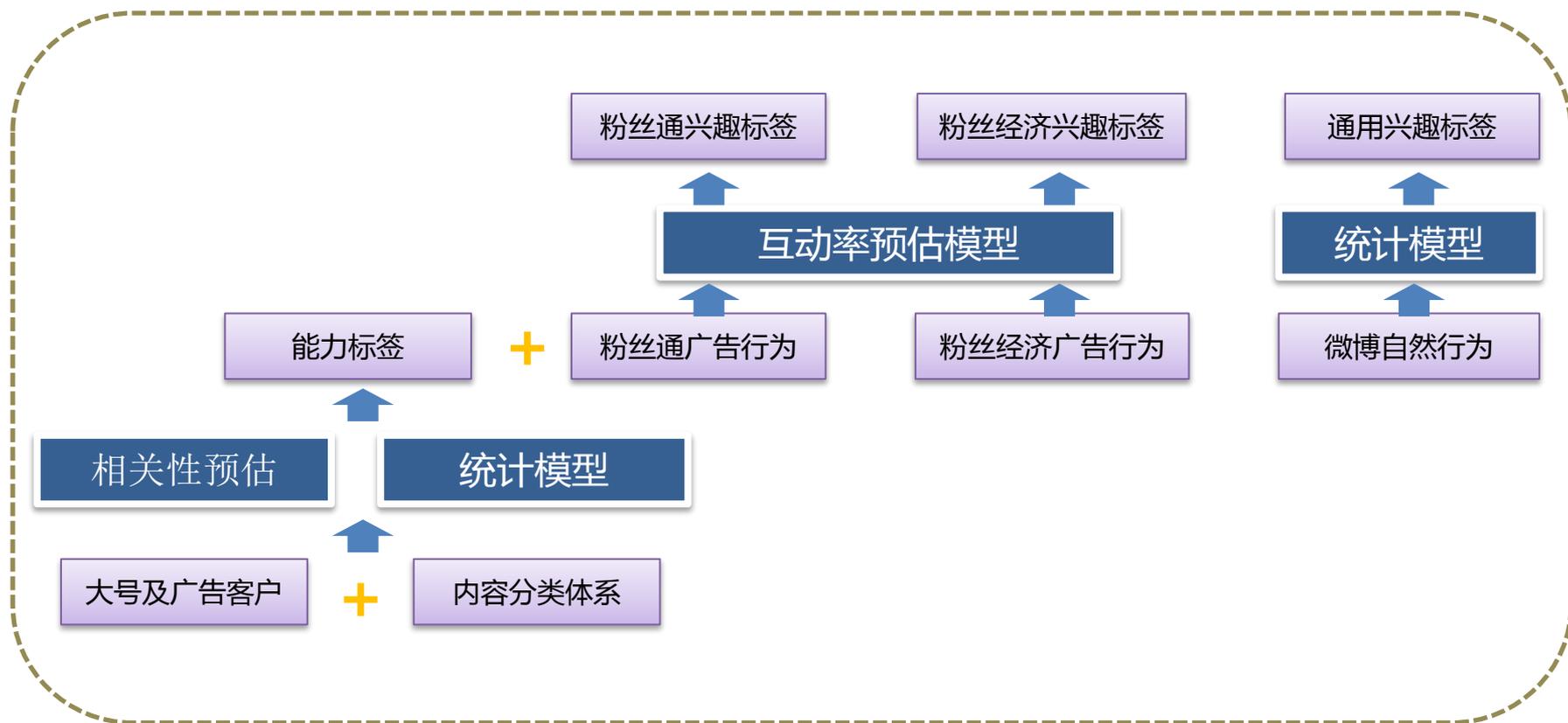


## 原则

- 1.任何挖掘工作必须首先确定评价方式
- 2.让相关业务人员理解并接受评价方式
- 3.尽量把数据的评价级别往上推

## 兴趣计算

- 兴趣：指定广告场景下用户对某个粒度的广告发生目标行为的概率
- 分广告产品，按粒度分层
- 通用兴趣标签：统计



## 内容分类

- 根据平台内容特性来确定分类
- 用户-词-类目
- 广义客户：所有广告客户+有一定影响力的内容号

### 词模型



广义客户

广告博文

分组标签

v用户信息

其它人工加工过的信息

相关性  
计算

客户输入定  
向词

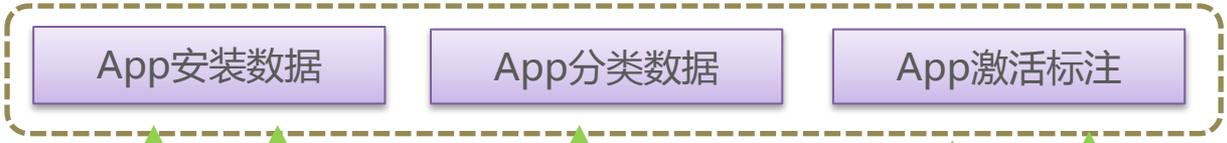
聚类后  
产品化

客户商业  
类目

精准兴趣

请添加兴趣关键词，如有多个请以逗号分隔

添加

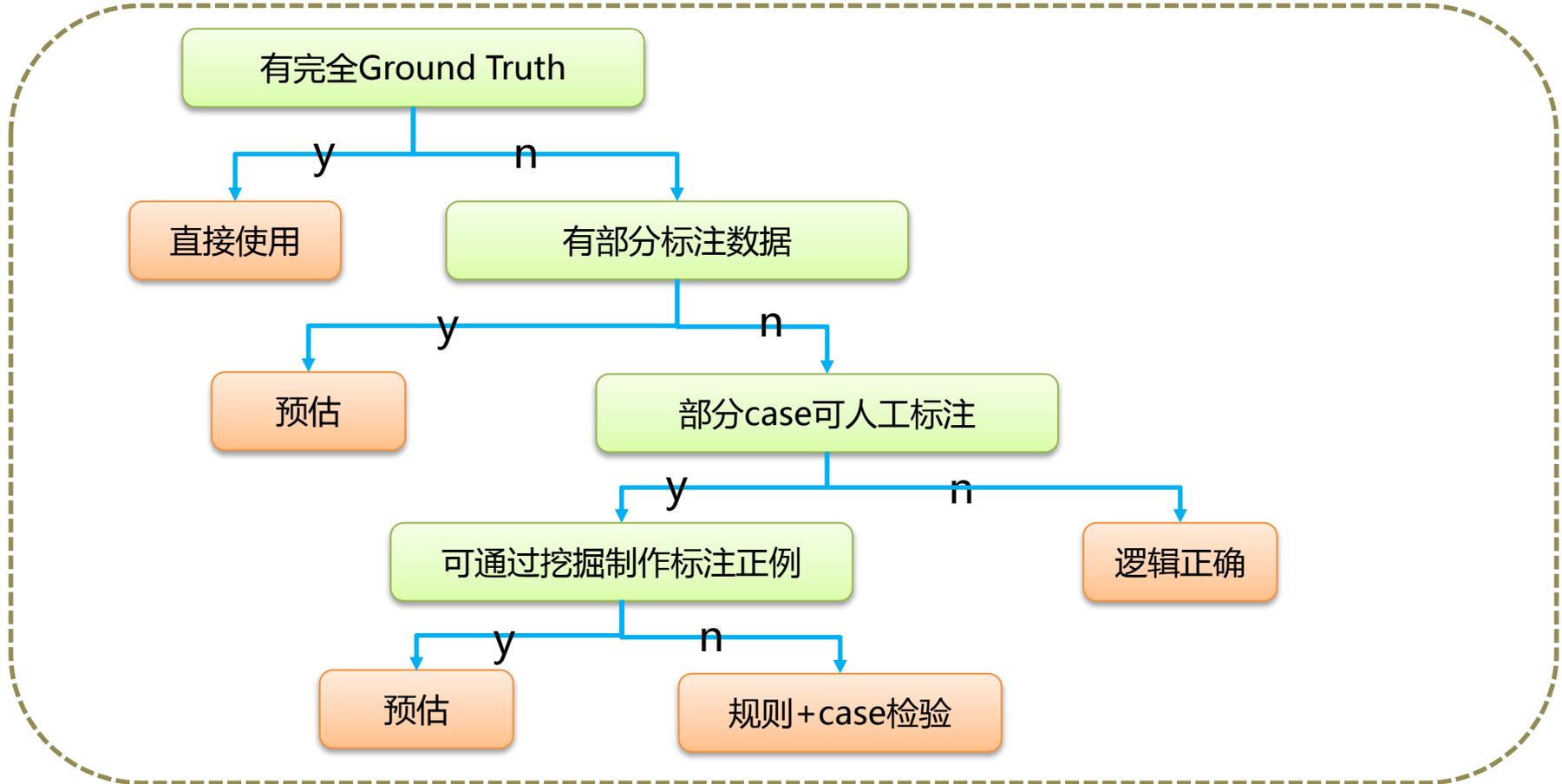


数据合作体系

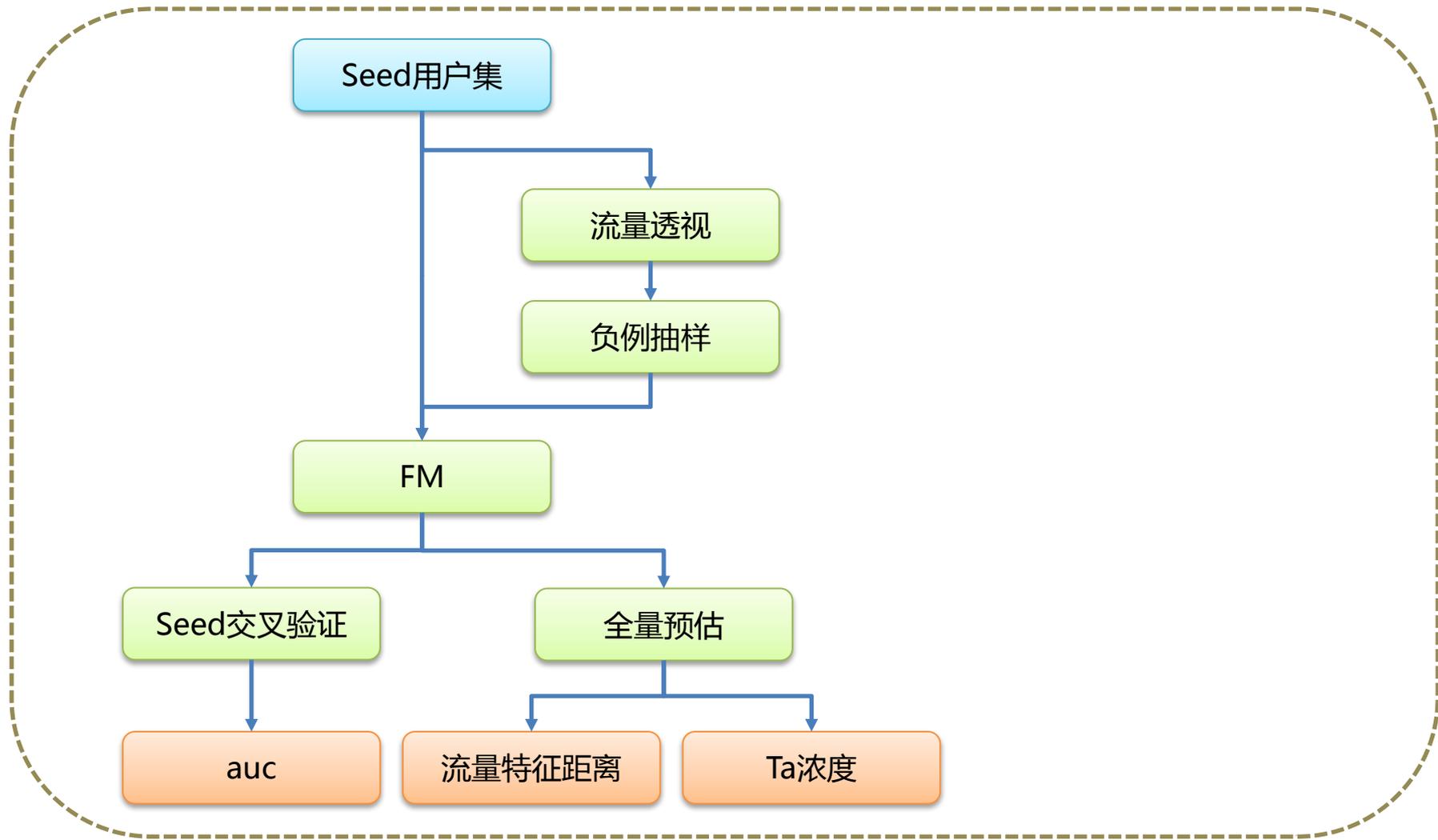
### 应用家数据算法体系

1. 数据获取, 合作
2. 数据清洗, 对齐
3. App分类
4. 标注预估, 安装预估, 标签预估

## 挖掘方法流程



## Look alike计算



- 广告产品
- 数据体系
- 挖掘方法
- 总结

## 紧贴业务

- 花大量的时间沟通
- 让数据和业务都能活起来

## 建设评价体系

- 迭代评价方式
- 获取评价数据
- 优化业务，简化工作

## 数据工作

- 数据 > 特征 > 模型
- 多层次，低耦合的算法模块
- 规则和ML结合，但禁用带加权参数的规则



**SDCC 2016**

**中国软件开发者大会**

SOFTWARE DEVELOPER CONFERENCE CHINA

**谢谢！**