

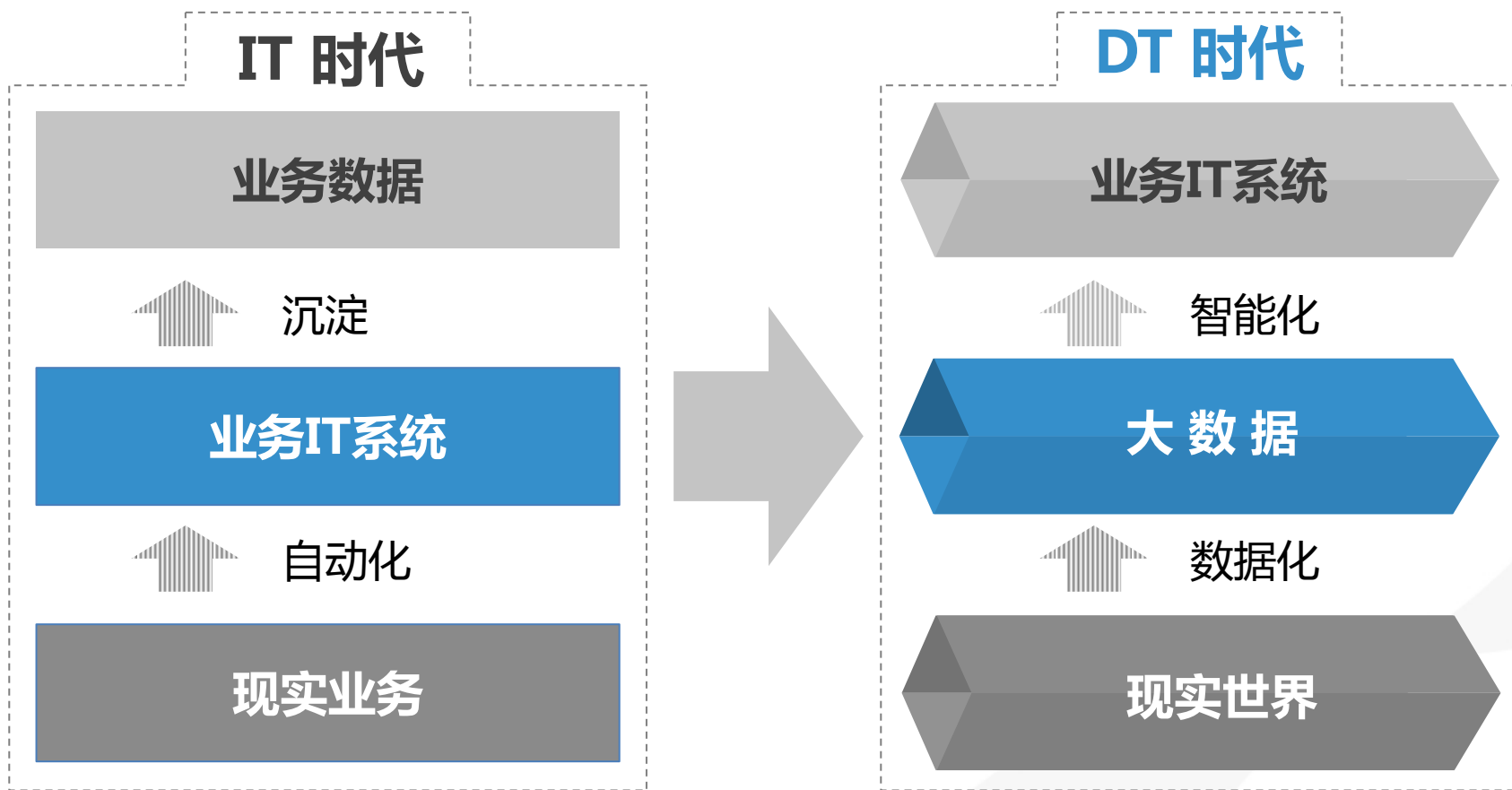
# 用户画像的构建及应用

百分点 2015.11

# 大数据与用户画像

# “Big Data” 之 “BIG”

大数据是信息化技术的自然延伸，意思是无处不在的数据



# 大数据时代需要将“人”数据化：“用户画像”

## Big Data = 无处不在的数据

### 信息化建设

- 全社会的信息化程度越来越高，越来越多的业务需要计算机应用，用户与这些应用交互产生大量数据

### 可穿戴设备

- 可穿戴设备甚至可植入设备将越来越多的出现在现实生活中
- 攻壳机动队（Ghost In The Shell）

### 信息网络

- 无处不在的网络将人和设备连接在一起，认识人、与人沟通的方法将发生本质性的变化

## 人类要学会从比特流中解读他人，更要教会机器从比特流中理解人类

- Eugene Goostman
- 信息诈骗
- 个性化推荐
- 个性化服务
- 智能理财
- 智能客服

# 理论

# 用户画像感性认识

身長八尺，面如冠玉，头戴纶巾，  
身披鹤氅，飘飘然有神仙之概



**13 詹姆斯·罗德里格斯**

10  180 cm/79 kg 哥伦比亚 1991.07.12

体形: 普通

5 3

花式动作 ★★★★★

球员等级 Lv.1 强化等级 +1

射门	71	反应	76	盘带速度	59
射门力量	72	弹跳	51	人盯人	26
弧线	77	体力	67	侵略性	67
远射	73	强壮	67	战术意识	54
凌空抽射	70	平衡	58	抢断	35
任意球	75	短传	75	铲断	30
罚点球	73	长传	74	GK扑救	17
头球	62	传中	73	GK手控球	17
站位	73	视野	78	GK大脚开球	17
速度	71	控球	77	GK反应	16
加速	73	盘带	78	GK防守站位	9
灵活	73			GK一对一	

# 用户画像感性认识

身高八尺，面如冠玉，头戴纶巾，  
身披鹤氅，飘飘然有神仙之概。



13 詹姆斯·罗德里格斯

10

180 cm/79 kg

体形: 普通

哥伦比亚

1991.07.12

RW 74

CFM 74

76

51

67

67

58

75

74

73

盘带速度

人盯人

侵略性

战术意识

抢断

铲断

GK扑救

GK手控球

59

26

67

54

35

30

17

17

站位

速度

加速

灵活

73

71

73

73

视野

控球

盘带

78

77

78

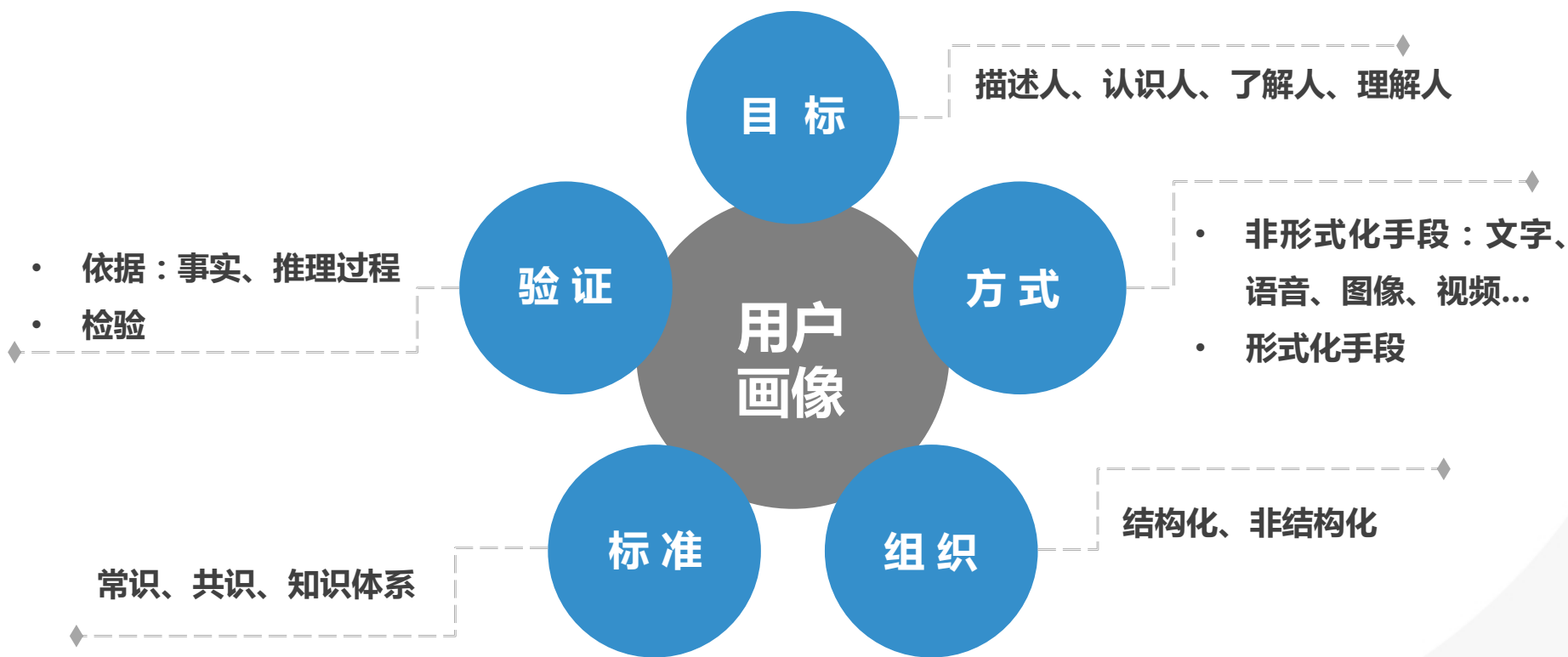
GK防守站位

GK一对一

16

9

# 用户画像的几个方面





# 大数据用户画像：定义

用户画像是对现实世界中**用户**的**数学建模**

用户画像是描述用户的**数据**，  
是符合**特定业务需求**的对用户的  
**形式化**描述

——  
源于现实，高于现实

用户画像是通过**分析挖掘**用户  
尽可能多的**数据信息**得到的

——  
源于数据，高于数据

## 大数据 + 洞察

# 大数据用户画像：构建原则

## 用户画像

业务知识体系

形式化

## 本 体

Wiki：

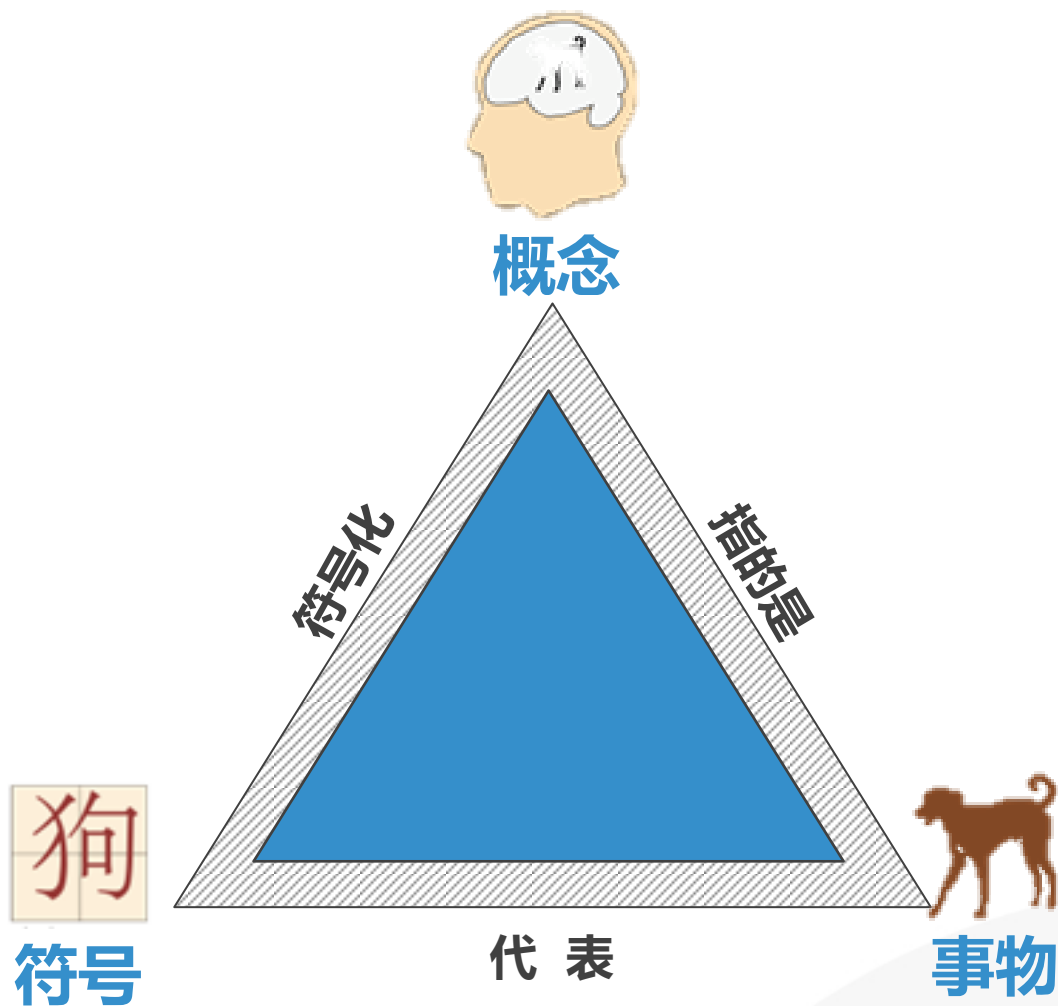
- 本体是一种形式化的、对于共享概念体系的明确而又详细的说明
- 本体提供的是一种共享词表，也就是特定领域之中那些存在着的对象类型或概念及其属性和相互关系
- 本体就是一种特殊类型的术语集，具有结构化的特点，且更加适合于在计算机系统之中使用
- 本体实际上就是对特定领域之中某套概念及其相互之间关系的形式化表达

# 基于本体论的知识表示方法

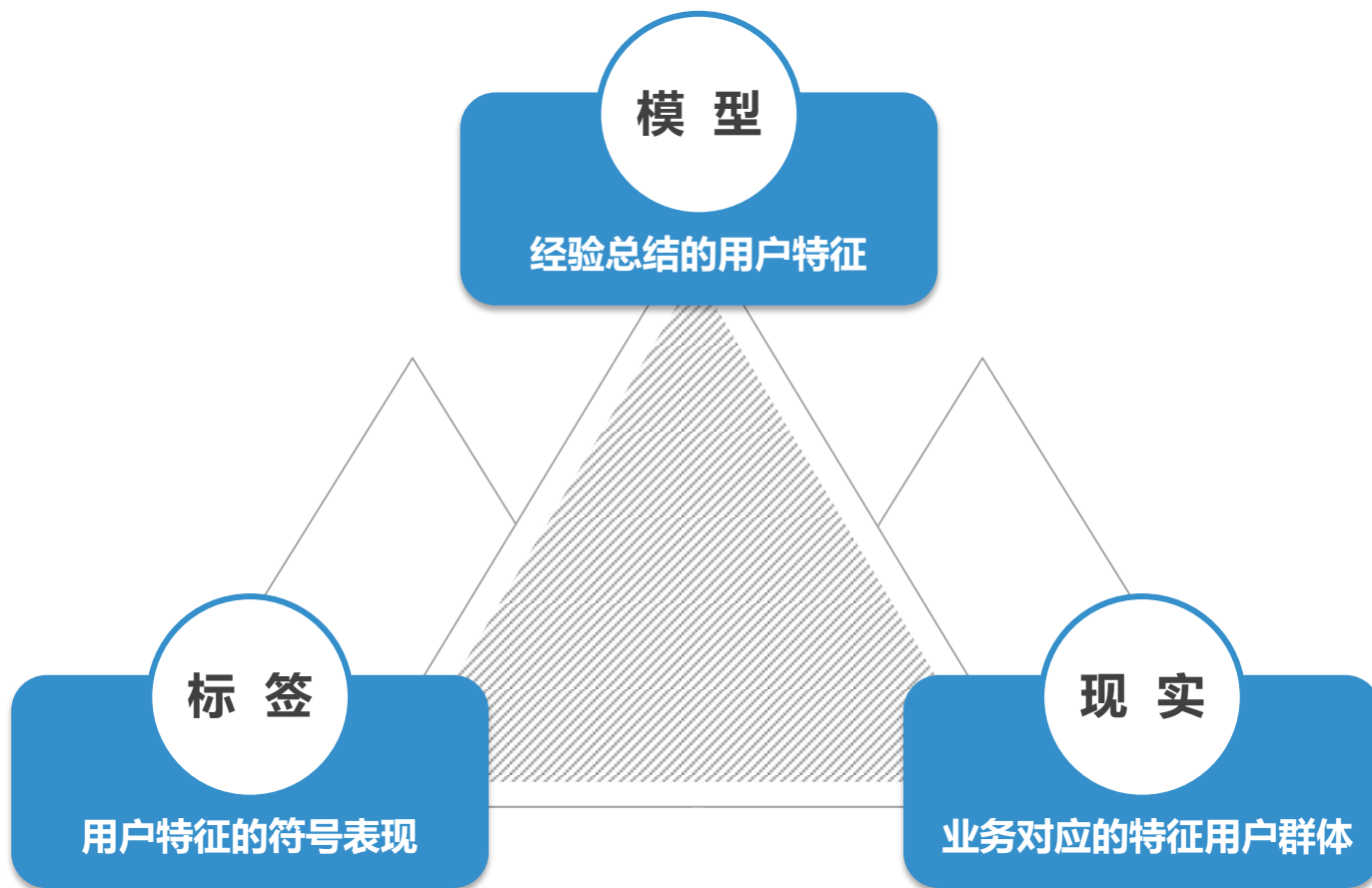


本体通常采用谓词逻辑作为描述语言（符号表示、形式化表现）

# 朴素的知识表示方法：符号-概念



# 朴素的用户特征表示方法：标签-模型



# 大数据用户画像：“标签体系”方法

标签是某一种用户特征的**符号**表示

## 标签体系

### 化整为零

每个标签都规定了我们观察、认识和描述用户的一个角度

### 化零为整

用户画像是一个整体，各个维度不孤立，标签之间有联系

用户画像可以用**标签的集合**来表示

# 大数据用户画像：验证

用户画像是对现实世界中用户的**数学建模**

## 模型是否反应了现实

- 逻辑：可靠性
- 机器学习：准确率
- 搜索：查准率
- ...

## 现实是否在模型中体现

- 逻辑：完全性
- 机器学习：召回率
- 搜索：查全率
- ...

标签（体系）准不准？

标签（体系）全不全？

# 大数据用户画像：验证

用户画像是对现实世界中用户的**数学建模**

## 模型是否反应了现实

- 逻辑：可靠性
- 机器学习：准确率
- 搜索：查准率
- ...

## 现实是否在模型中体现

- 逻辑：完全性
- 机器学习：召回率
- 搜索：查全率
- ...

无法同时满足

标签（体系）准不准？

标签（体系）全不全？



# 大数据用户画像：验证

用户画像是对现实世界中用户的**数学建模**

模型是否反应了现实

- 逻辑：可靠性
- 机器学习：准确率
- 搜索：查准率
- ...

现实是否在模型中体现

- 逻辑：完全性
- 机器学习：召回率
- 搜索：查全率
- ...

无法同时满足

标签（体系）准不准？

标签（体系）全不全？

无法完全

# 大数据用户画像：准确性验证

## 有事实标准

数据+学习，可以验证结果

- 训练集 + 测试集

### 示 例

- 以注册填写性别为标注集，用ML算法摸索用户行为与性别之间的关系

## 无事实标准

假设+实现，只能验证过程

- 计算过程是否合乎逻辑

### 示 例

- 流失用户 = 半年未交易用户
- 忠诚度 = 若干综合指标评分

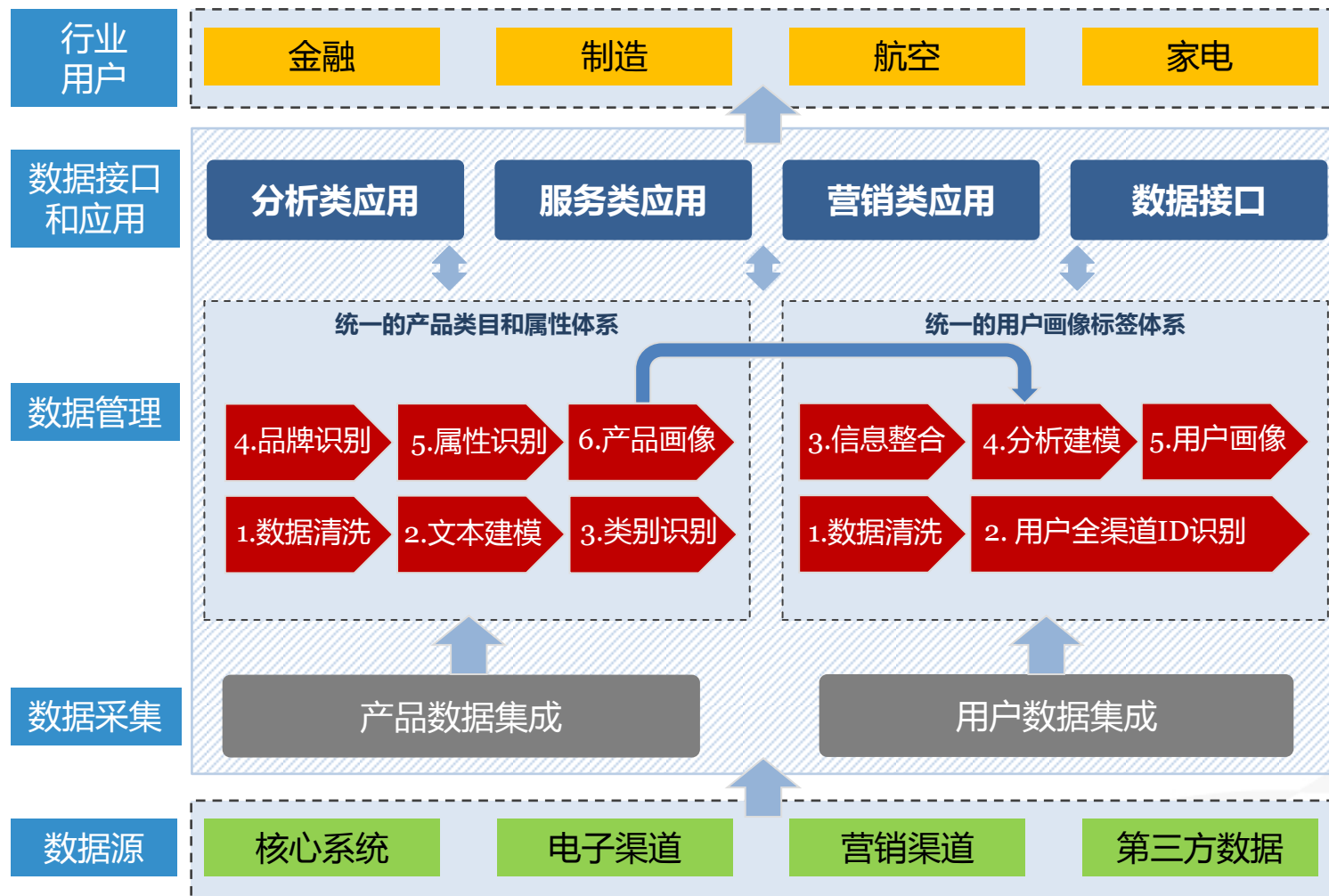
## 实践检验

### 证伪主义

- 试错、A/B Test、数据闭环、自我完善

# 实 践

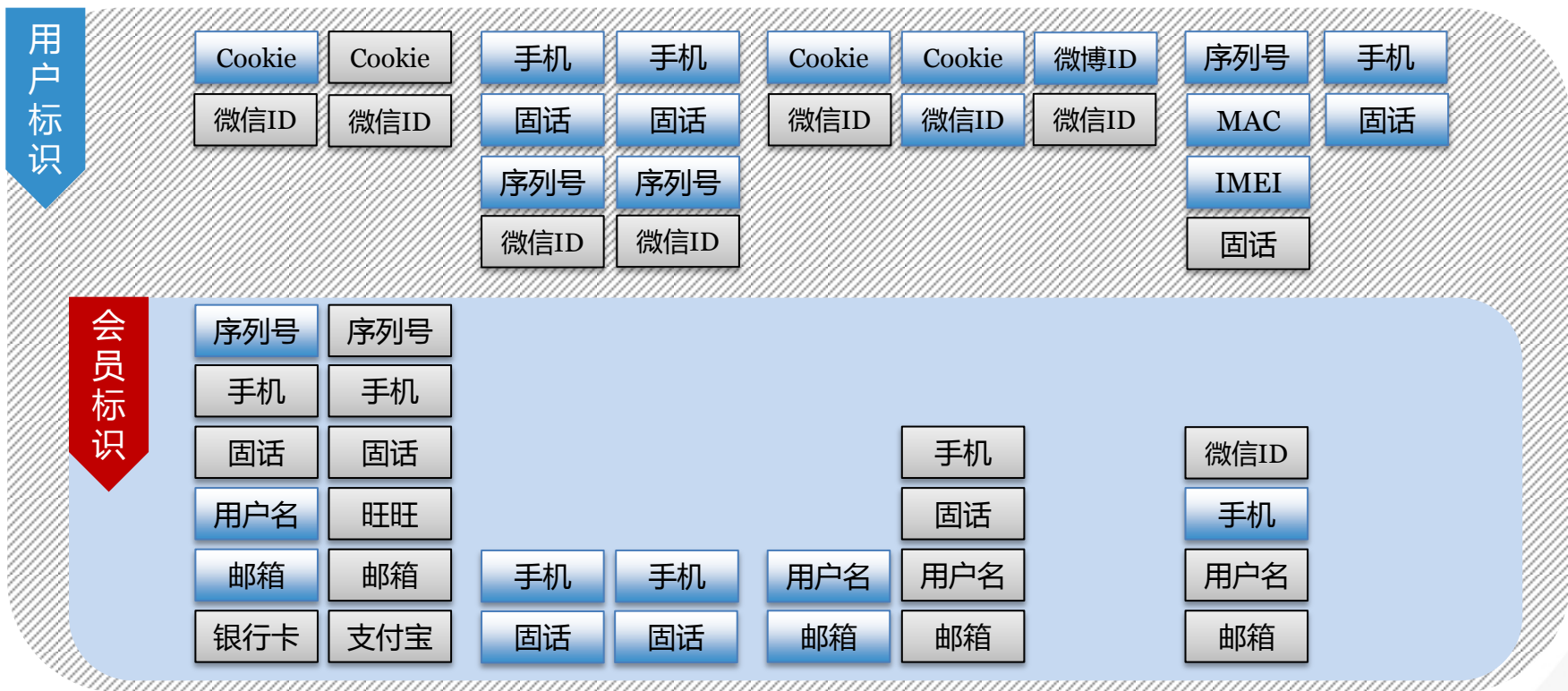
# 用户画像生产和应用：逻辑架构



# 构建用户画像的关键难题



## 企业用户触点举例



一定能得到的标识

有可能得到的标识

官网  
商城

## 第三方线上渠道

客户经理

## 第三方渠道

论坛社区

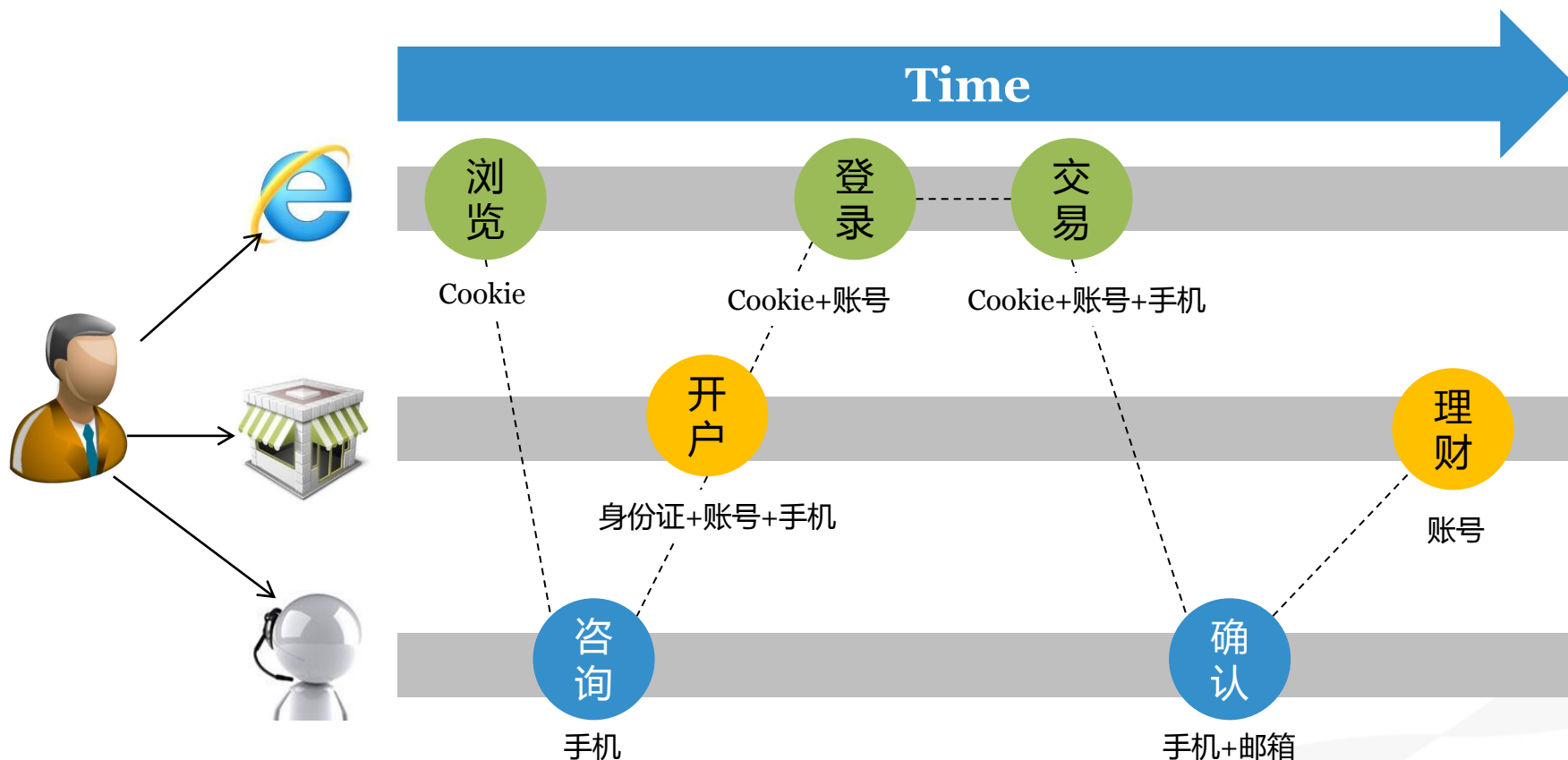
微信

微博

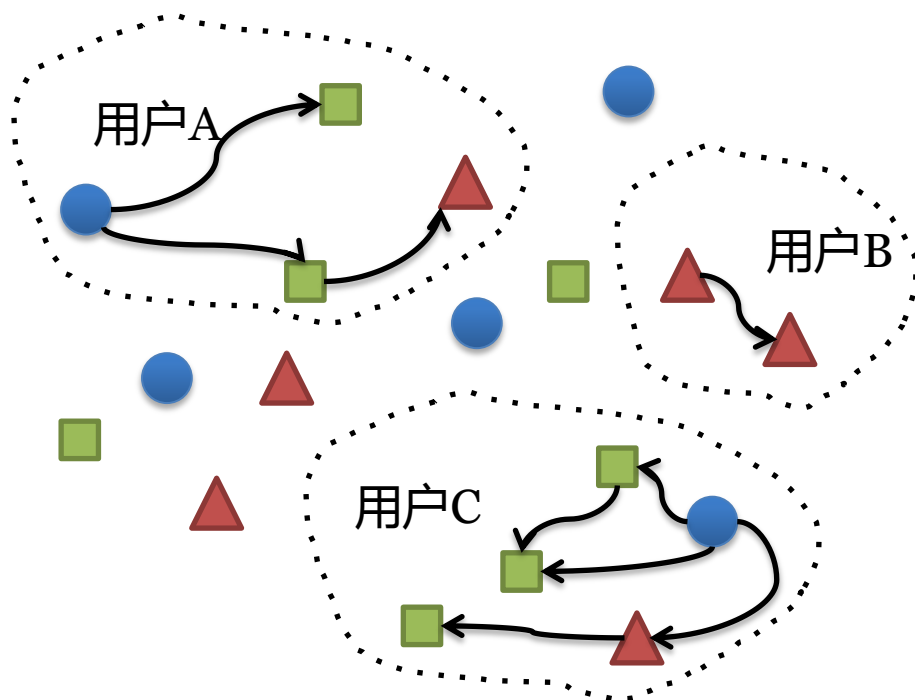
智能应用

客服

# 什么要打通：大数据时代我们需要上帝视角！



# 用户全渠道ID打通



## 图中

- 有三类ID
- ID间有相互联系
- 相互联系的ID反映出它们很有可能代表同一个用户

## 低密级业务

- 可以仅使用多重ID中的任意一个
- 最大程度打通，跨平台一致体验

## 高密级业务

- 使用特定ID，或者多种ID的组合
- 保证数据的准确和安全



# 如何拉通：标签体系拉通（1）

## 人工手段

业务人员手工映射

人工成本昂贵

标准不好统一

难以大规模开展

采用

## 自动手段

机器学习算法

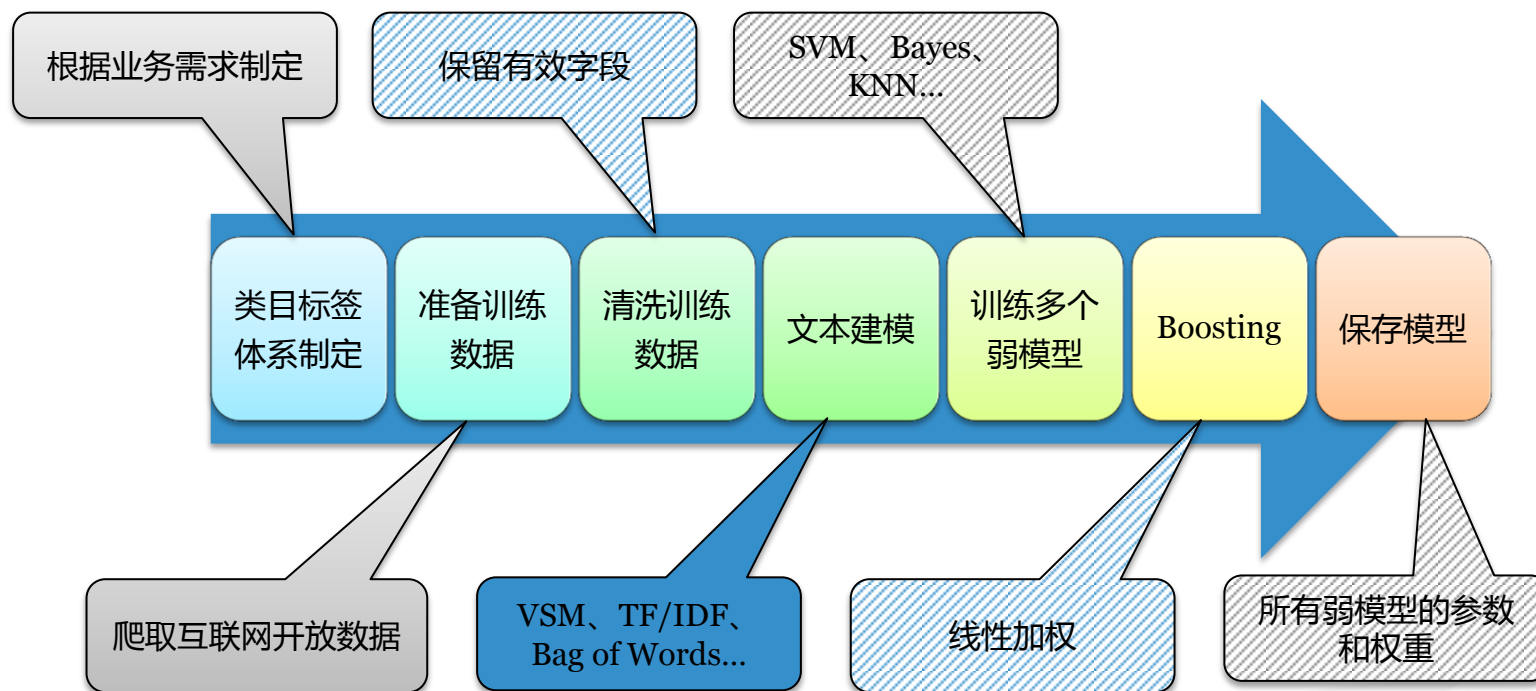
少量人工辅助

统一数据模型

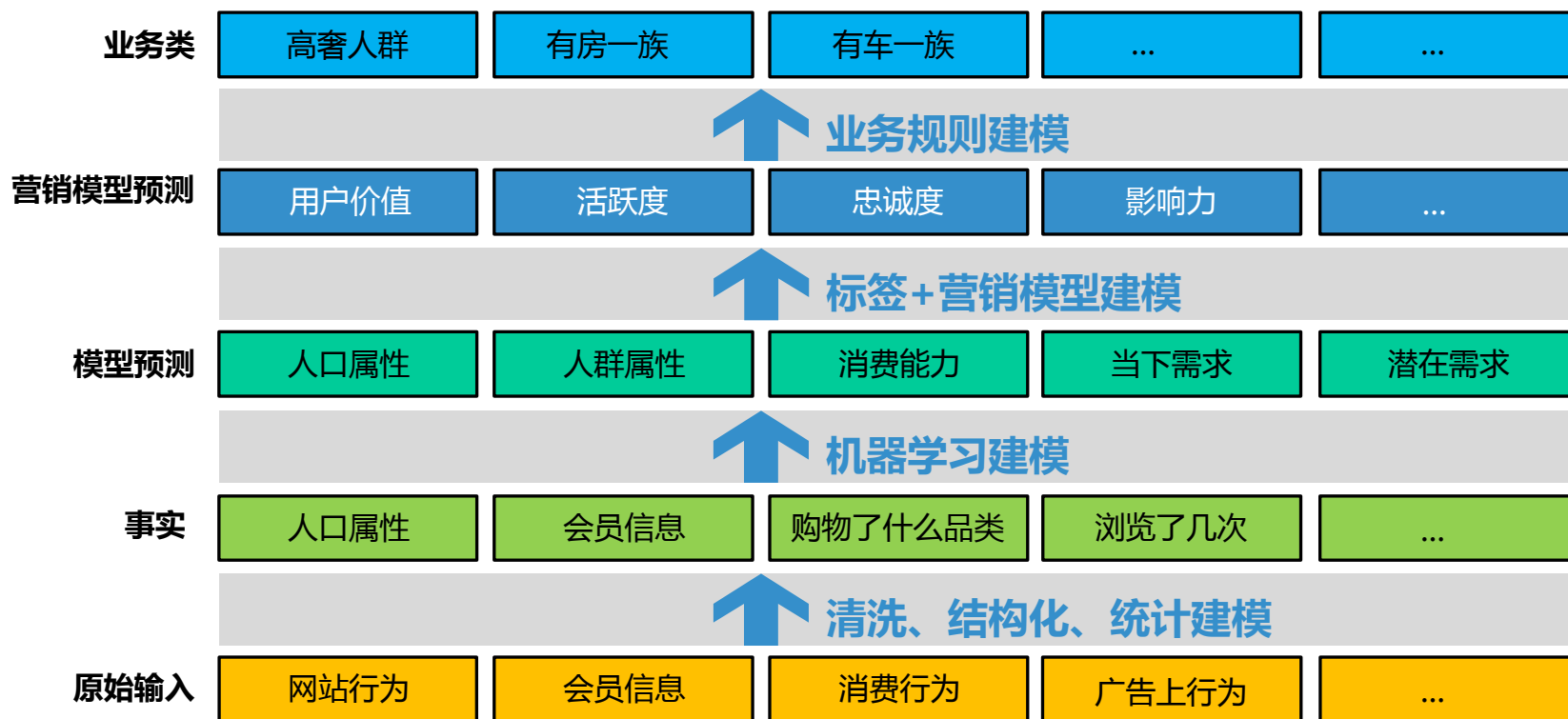
适合海量数据

# 如何拉通：标签体系拉通（2）

以拉通类目体系为例，构建基于机器学习的自动分类模型



# 用户画像标签层级



# 应用

# 用户画像之应用



# 数据应用能力——典型应用：“售前”精准营销



# 某知名制造企业支持营销应用

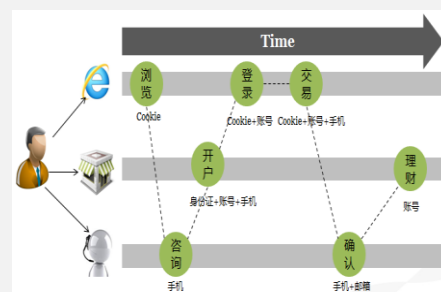
## 目标

- 建立互联网全触点用户经营模式
- 拉通和建立消费者统一的用户数据平台
- 利用用户数据驱动业务改进

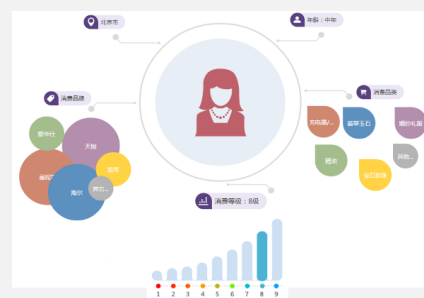
## 解决方案

- 建设用户中心大数据平台
- 收集、拉通企业内、外部消费者用户数据
- 建立消费者用户画像，对用户进行细分，识别用户的兴趣爱好等特征
- 基于用户画像完成营销增强和精准营销

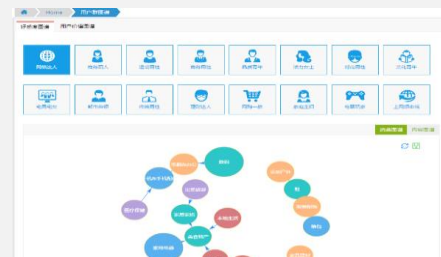
## 价值展示



全触点打通



用户微观画像

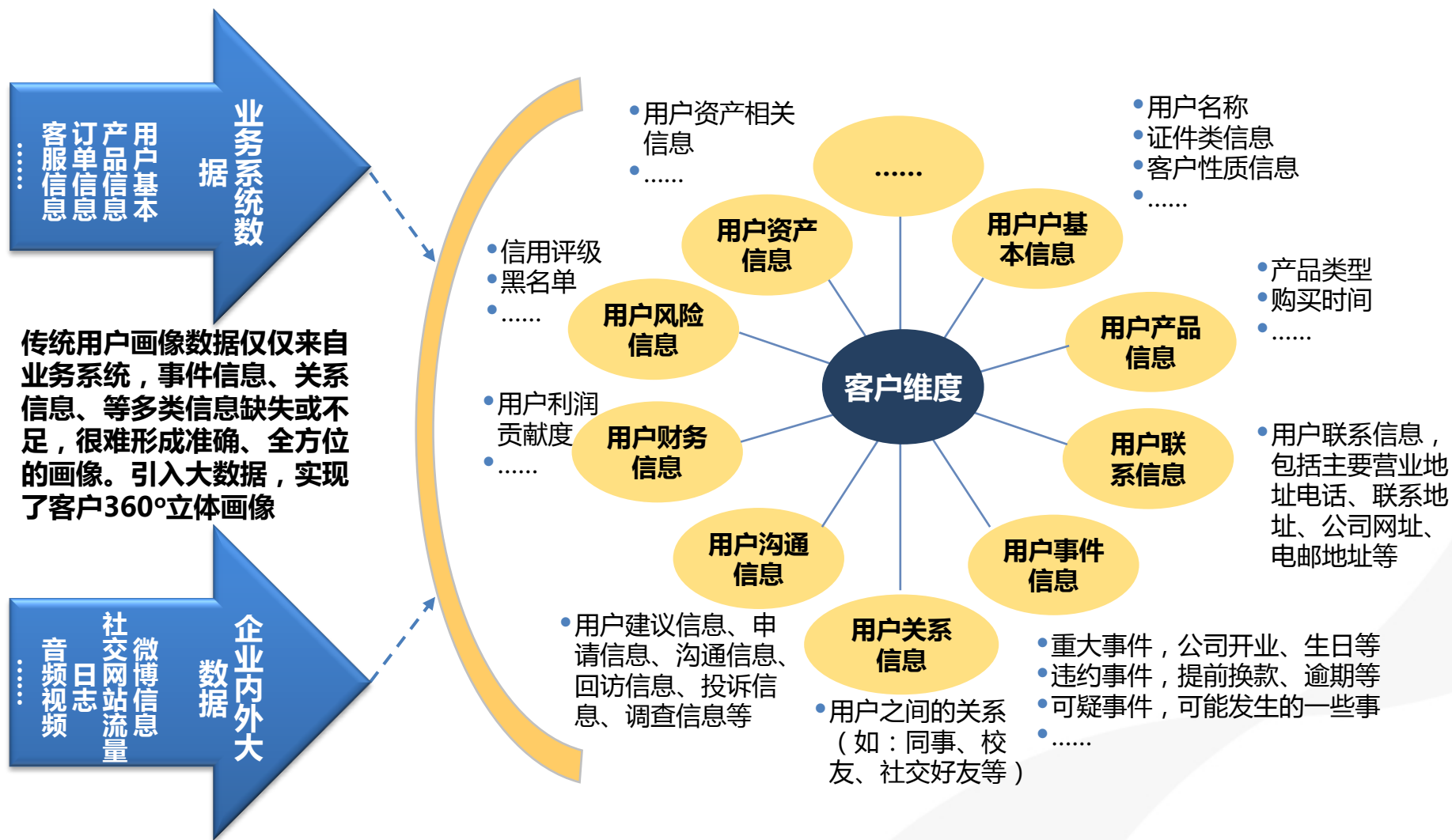


用户行为偏好



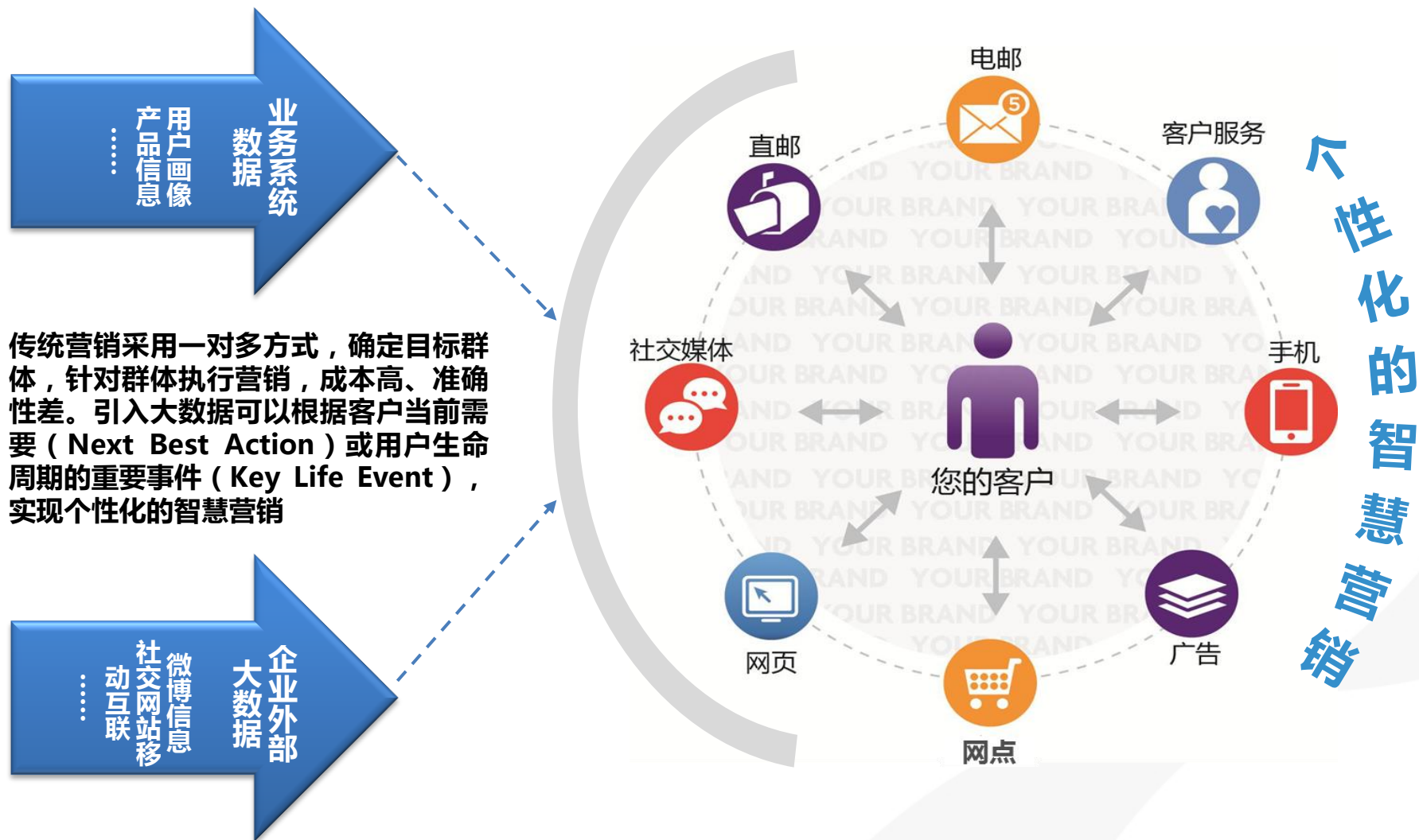
精准广告营销

# 构建360°用户画像体系





# 执行个性化精准营销



# 执行个性化精准营销



# 应用于社会化营销

## 某家电制造企业，新品发布时招募粉丝

### 问题和需求

- 期望通过短信和邮件方式，从老用户中找出最有可能参加活动的粉丝

### 分析和解决方案

- 愿意参加活动的一定是对品牌认同和忠诚度最高的用户
- 利用企业的CRM、客服、销售等数据，对用户忠诚度进行综合评定并挑选忠诚度最高的用户作为招募目标
  - 接触渠道的多少
  - 购买品类的多少
  - 投诉次数的多少
  - ...

### 效果

- 带来了超过一半的粉丝，但成本只有以往的40%

# 数据应用能力——典型应用：“售中”个性化推荐



# 百分点个性化推荐引擎 (BRE)



# 应用于个性化推荐

## 某团购网站，应用百分点推荐引擎优化案例

### 问题和需求

- 下单率低



### 分析和结论

- 步长偏短。PV为1的用户占65+%，PV为2~5用户比例逐步下滑，7~10开始稳定，11步以上又开始略有上升
- 主动访问用户比较少，仅占8%左右，大多数用户通过导航和搜索过来
- 用户首次进入和离开的品类变化率比较低
- 用户首次进入商圈和下单商圈较接近
- 用户的忠诚度比较低
- 大多数用户都是有强烈需求过来
- 用户具有区域性购买特征

# 应用于个性化推荐

## 某团购网站，应用百分点推荐引擎优化案例

### 解决方案

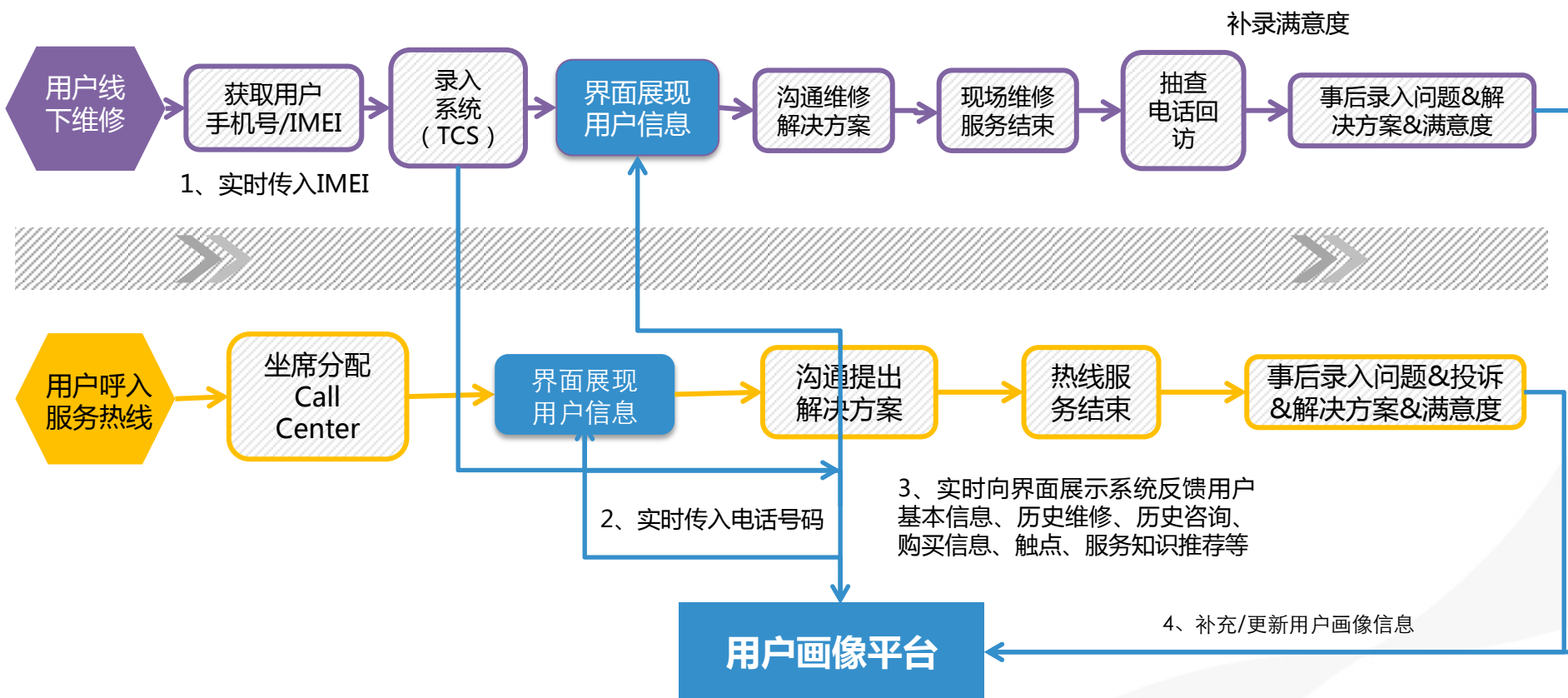
- 改进召回：使用用户画像中的品类偏好、商圈偏好、消费能力等标签优化召回
- 去除用户反感：利用用户标签衰减、权重清零等机制，进行品类过滤，避免给用户进行过力营销
- 利用百分点覆盖多行业多客户的全网数据特点，构建用户全网的需求潜在需求标签：解决冷启动问题

### 效果

- 点击率平均提升18.23%
- 平均步长提升45.8%
- 影响下单提升30%
- 直接下单提升86.95%



# 数据应用能力——典型应用：“售后”增值服务





# 小 结

用户画像不是数学游戏，  
而是严肃的业务问题，  
是业务与技术的最佳结合点，  
是现实与数据化的最佳实践



Big Data Practitioner

