

# 微积分

——金融产品创新比赛

中国移动杭州研发中心  
2016.4

# 市场分析

# 征信应用产品市场分析

金融市场高速发展亟需信用市场的补充，国家开放企业准入个人征信市场，使征信业务进入快速发展阶段。

## 外部市场环境利好

### 01 国家政策支持，开放个人信用市场

国家加大对个人征信行业的支持，开放准入企业进入征信市场，其中八家获得了征信牌照

### 02 信用经济发展

信用交易在各种商业交易中的比重增加

### 03 互联网金融快速发展

互联网金融、P2P信贷等需要风险评估的应用蓬勃发展

## 内部应用价值凸显

### 01 国家信用体系的基石

国家亟待建立和完善标准化信用评估体系，需要多方数据支撑和整合

### 02 精确界定授信方可以接受的风险程度

越来越多应用领域需要评估风险程度

### 03 加速授信过程

用户对于授信效率的要求提升

信用市场对比	美国	中国
社会信用基础	信用体系完善	快速发展阶段
信用记录用户占比（信用卡用户）	80%	13%
移动用户中，无信用记录的人群总数	6000万	4亿
征信行业规模预测（2020年）	600亿	1075亿

- 存在评估难度大、评估需求旺盛的庞大群体；
- 市场潜力巨大

## 中国移动的优势：

- 覆盖最广、数量最多的用户群体
- 8亿用户的真实信息数据

# 初步构建中国移动大金融征信生态圈

查验类服务

信用类服务

初步建立移动用户信用支付体系

基本特质

还款及履约能力

价值观

社会资本

地理位置及行为偏好

朋友圈

用户信用管理

用户分级管理

个人风险评估

动态信用评估

内部数据

外部数据

坏帐风险预警

信用支付

基地专业  
公司数据

支付类数据

金融机构数据

差异化营销等  
延伸应用

衍生金融产品

中国移动通信数据

数据交换

第三方数据

授信服务

信用担保

提升行业信用  
评级

提升用户跨  
界运营

# 产品介绍

# 产品介绍

杭研微积分产品可为**金融机构、其他行业作为独立的第三方提供泛征信服务**。基于省移动和其他行业数据，对个人信用行为进行评估，为金融、保险等**强信用消费行业**的用户提供**信用评估服务**，帮助其**完善或建立用户信用体系**：

- 能有效补充央行征信中心覆盖面，提高信个人信用信息准确性，完善社会信用体系；
- 有效帮助保险等行业进行更全面的信用数据评估、反欺诈监控等提升其风险控制能力，优化业务处理流程；

## □ 信息验证服务

为保险公司新客户**提供单条/批量信息查验服务**，对客户身份信息进行甄别，可以防止用户进行**欺诈行为**，提前预防网络诈骗或犯罪。



## □ 反欺诈服务

防止投保人故意隐瞒有关保险标的的真实情况，诱使保险公司承保，或者利用保险合同内容，故意制造或捏造保险事故造成保险公司损害，以谋取**保险赔付金**。



## □ 信用评分服务

基于中国移动海量的个人通信数据及消费数据建立的**个人信用评分模型**，信用评估机构利用信用评分模型对消费者个人信用信息进行量化分析，以分值形式表述，是对其履行各种经济承诺的能力进行全面的判断和评估。



## 依托海量用户行为数据完善个人信用风险评估

### 个人信用风险评估

传统  
征信

账户	固定资产
单位	信贷历史
违约	.....

移动  
征信

朋友圈	地址变动
活动区域	访问行为
通信消费行为	.....



### 移动征信的优势

人群覆盖更广	信息更全面
维度更丰富	结果更可靠

微积分

信息验证

反欺诈

信用评分

# 微积分应用于“拍拍贷”

## 四类应用场景

**风控预警**：借助信用风控模型，节省成本、节省审批时间，建立金融行为、通信行为（中国移动数据）、生活行为（中国移动数据）的风险控制预警模型，对业务运营进行风控管理；

**存量客户提升**：借助中国移动对用户行为偏好的分析，发现潜在业务需求人群，并可借助中国移动的线上渠道进行精准化营销合作；

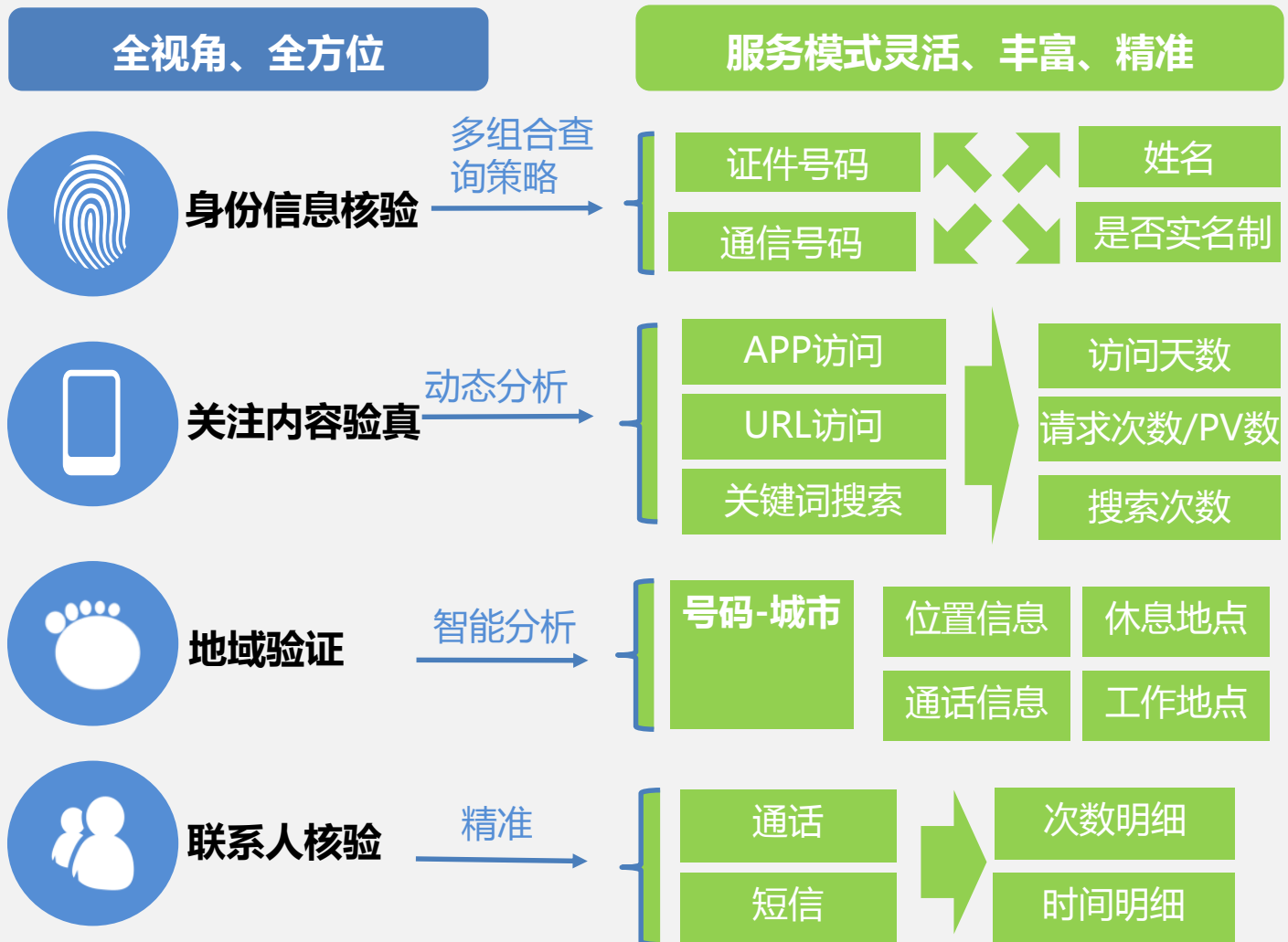


**新客户发展**：借助风控模型对**新客户**进行信息甄别，做为客户信用评估的参考依据，增加信用评估维度（增信的作用），**精准消息推送新客户**；

**提升客户获取精准度**：精准圈定客户，借助中国移动集团虚拟网数据或是相似度关联数据模型，获取客户的真实人群特征，如：职业特征的甄别；

# 应用方案一：信息查验

- 服务模式：提供个人不同维度的信息查验服务。
- 服务面向：包括政府部门、银行、信托、保险公司、小额贷款公司、汽车金融公司等。



## 应用方向

应用行业：P2P/银行等

**快速信贷**

快速信贷，对用户填写的身份信息等的核查过程，往往需要快速核查、分析。企业对新用户信息数据积累匮乏，有不能准确识别客户的风险。

**信用卡核放**

应用行业：银行/信用卡发卡机构等。

信用卡机构往往需要多方数据源，保证数据的真实、用户覆盖、有效等。

应用行业：保险等

**保险理赔/电子商务**

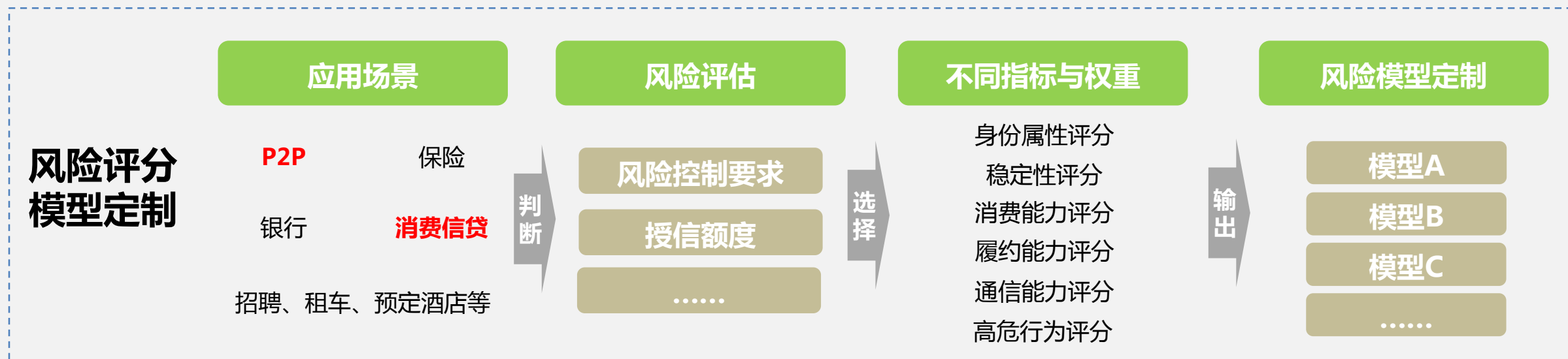
保险理赔时，信息不真实会带来巨大的损失。电子商务需要对注册者身份信息进行实名认证。



## 应用方案二：风险评分评估

服务模式：根据不同的应用场景，判断风控标准，选取不同的指标和权重，为客户量身定制风险评估模型。

服务对象：银行、P2P、贷款机构等



### 应用方向

应用行业：P2P/银行、  
租赁公司

#### 信贷/租赁服务

结合不同行业自有数据定制模型，根据风险分值判断是否发放贷款、提供租用服务等。

#### 授信额度确定

应用行业：银行、信用卡机构等

全面立体的了解和验证客户信息，多角度、全方位、综合性的判断客户的真实情况以及还款能力，根据不同客户还款能力的不同，给予不同额度的授信，从而实现真正的个性化授信

# 应用方案三：风险预警及追踪

服务模式：提供信用情况动态监控，分析其行为偏差值，提供风险预警；风险个体信息、位置等追踪服务。

服务对象：银行、P2P、贷款机构等

## 指标得分追踪

	1月	2月	3月	4月	5月	
指标1得分	4.95 ▲	-1.15	6.08 ▲	3.7 ▲	3.21 ▲	提升
指标2得分	-2.20	-0.34	-1.59	12.54 ▲	-4.90 ▼	
指标3得分	-6.50 ▼	0.15	-0.95	12.29 ▲	1.76	降低
指标4得分	2.91	-3.13 ▼	1.54	0.68	-0.09	
指标5得分	-14.71 ▼	-8.76 ▼	-1.68	-5.34 ▼	-9.51 ▼	降低
指标6得分	-12.27 ▼	-3.54 ▼	3.26 ▲	1.25	3.30 ▲	
指标7得分	-13.67 ▼	-3.50	0.64	-9.45 ▼	-9.44 ▼	降低
指标8得分	-6.80 ▼	-4.65 ▼	4.19 ▲	-13.05 ▼	-6.40 ▼	
指标9得分	-18.01 ▼	-2.71	-4.56 ▼	-9.40 ▼	-4.92 ▼	降低
指标10得分	-6.58 ▼	2.99	2.16	2.31	-13.29 ▼	

## 风险预警及追踪服务

- 贷后对贷款申请人的消费能力、还款能力等及时跟进分析，提前进行风险预警；
- 分数达到阈值主动推送风险信息。



信用卡  
风险追踪

商号	月	天	金额	03月11日	04月15日	星期四	21天	21天	04月30日	30天免息
招商银行	12	20	单休	03月12日	04月16日	星期六	-	-	04月12日	30天免息
招商银行	9	20	单休	03月09日	03月29日	星期六	-	-	04月09日	30天免息
招商银行	17	20	单休	03月17日	04月06日	星期日	2天	今天	04月17日	30天免息
招商银行	10	20	单休	03月10日	04月04日	星期四	-	-	04月10日	30天免息
招商银行	13	20	单休	03月13日	04月07日	星期一	3天	3天	04月13日	30天免息
招商银行	9	19	单休	03月09日	03月28日	星期四	-	-	04月09日	30天免息
招商银行	5	20	单休	03月05日	03月31日	星期四	-	-	04月05日	30天免息
招商银行	28	20	单休	03月28日	04月17日	星期四	13天	13天	04月	30天免息
招商银行	6	20	单休	03月06日	03月26日	星期四	-	-	04月06日	30天免息
招商银行	11	20	单休	03月11日	03月31日	星期四	-	-	04月11日	30天免息
招商银行	1	19	单休	04月01日	04月20日	星期四	16天	16天	04月01日	45天免息
招商银行	18	20	单休	03月18日	04月07日	星期一	3天	3天	04月18日	30天免息
招商银行	10	19	单休	03月10日	03月29日	星期六	-	-	04月10日	30天免息

风险分值分析

警告：  
姓名：张三  
信息：张三风险分值已超过警告值，存在盗刷风险，请注意。地点：XXX商场

应用方向

### 贷款再次发放、追踪

应用行业：P2P/银行等

个体等需要持续的贷款，或者中长期贷款；个人多次、持续的消费贷款也逐渐增多；需要了解个人或个体当前的风险状况，是否适合发放贷款等

### 反信用欺诈、反盗

应用行业：银行、贷款机构等

风险分值出乎意料的变化；或有关个体的记录的错误可以起着极重要的报警信号作用，显示用户的身份可能已经，或很快就将被盗用，银行和个体都将受损。

# 应用方案四：个人信用评级及信用生活

**服务模式：**根据用户的消费能力、欠费记录、好友信用度等属性，建立**信用模型**，对消费者个人信用信息进行量化分析，以分值的形式反映个人信用的高低，信用评级是对电信数据的综合利用。

面向社会的信用服务体系，客观呈现个人的信用状况，通过连接各种服务，让每个人都能体验信用所带来的价值。

## 信用分：

- 个人可查询信用分，展现自己的信用状况

## 信用分的七个维度：

- 履约能力、移动行为偏好、人脉信用、活动区域偏好、基本特质、通讯消费能力、忠诚度



## 信用生活：

- 用户—可凭借个人信用分享受商家优惠服务
- 商家—入驻信用评级APP，为高信用分的用户提供便利，引入客流

## 应用方向



贷款利率  
优惠多



租车订酒店  
免押金、免预授权



机票极速退款  
无需漫长等待



出境免签证  
全球随便跑



购物  
先试用后付款



其他延伸应用

# 微积分产品界面

## 01 主页—API类别展示

01

### 主页—API类别展示

- 基本信息**  
通过API获取用户注册信息、手机号码、身份证号、银行卡号、地址等信息。支持批量查询。
- 消费信息**  
根据用户在网时长、消费金额、套餐类型等信息，获取用户在网时长、消费金额、套餐类型等信息。
- 位置信息**  
通过API获取用户在网时的位置信息，支持批量查询。
- 社交信息**  
通过API获取用户在网时的社交信息，支持批量查询。
- 价值信息**  
根据用户在网时长、消费金额、套餐类型等信息，获取用户在网时长、消费金额、套餐类型等信息。
- 风控**  
通过API获取用户在网时的风控信息，支持批量查询。

02

### 基本信息类接口展示

- 姓名手机号匹配验证**  
103456
- 黑名单查验**  
56034
- 停机状态验真**  
12743
- 手机号正常使用验证**  
5836
- 资费等级**  
4422
- 在网时长查询**  
3141

03

### 接口详情页

姓名手机号匹配验证 意见反馈

提供方：贯众 | 服务类别：服务器 | 适用账户：企业账户

根据用户在中国移动注册的信息，验证合作方用户提供的手机号码和姓名是否匹配。

[试用](#)

04

### 接口试用

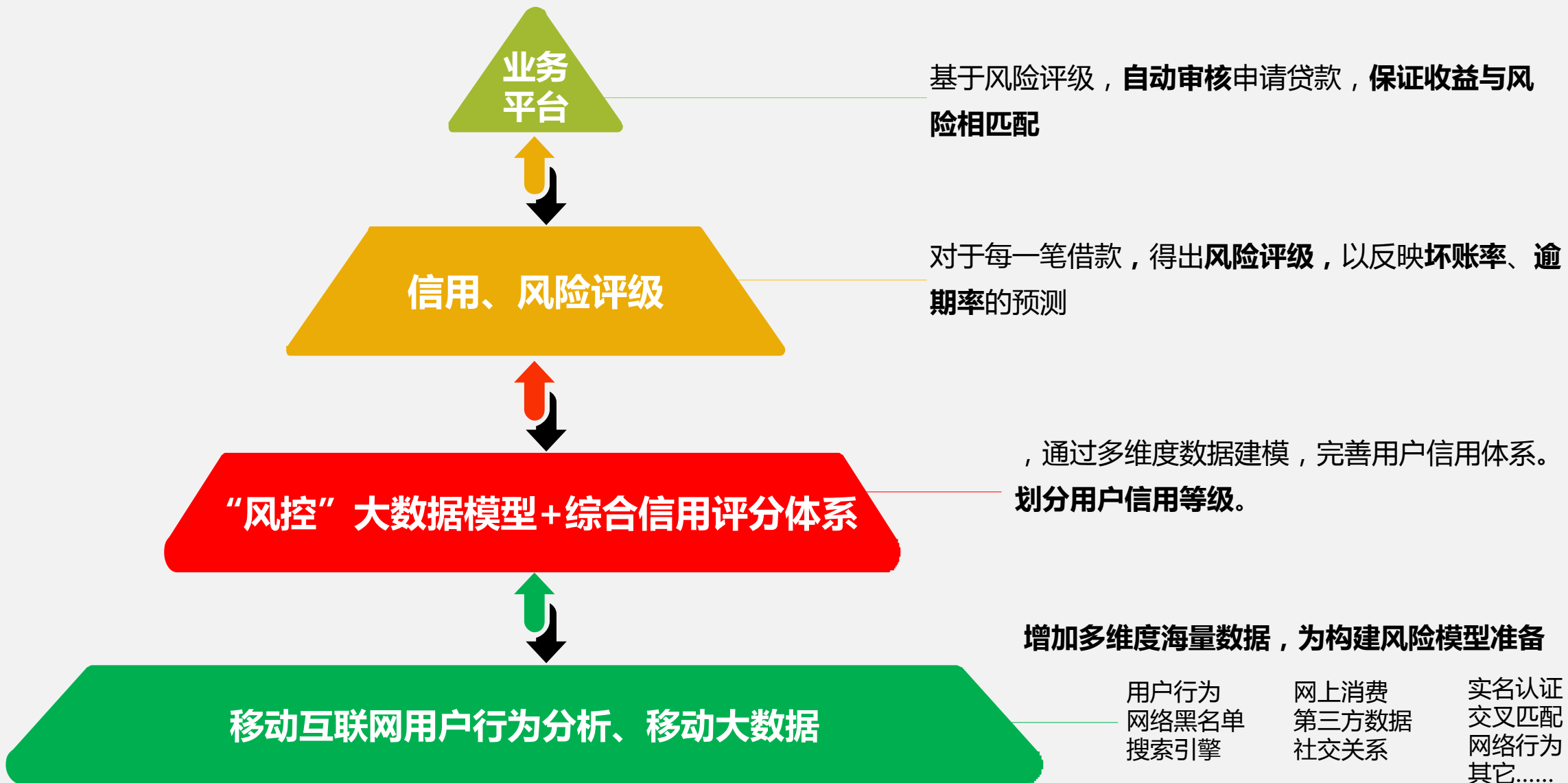
姓名手机号匹配验证

手机号码:

姓名:

客户授权码:

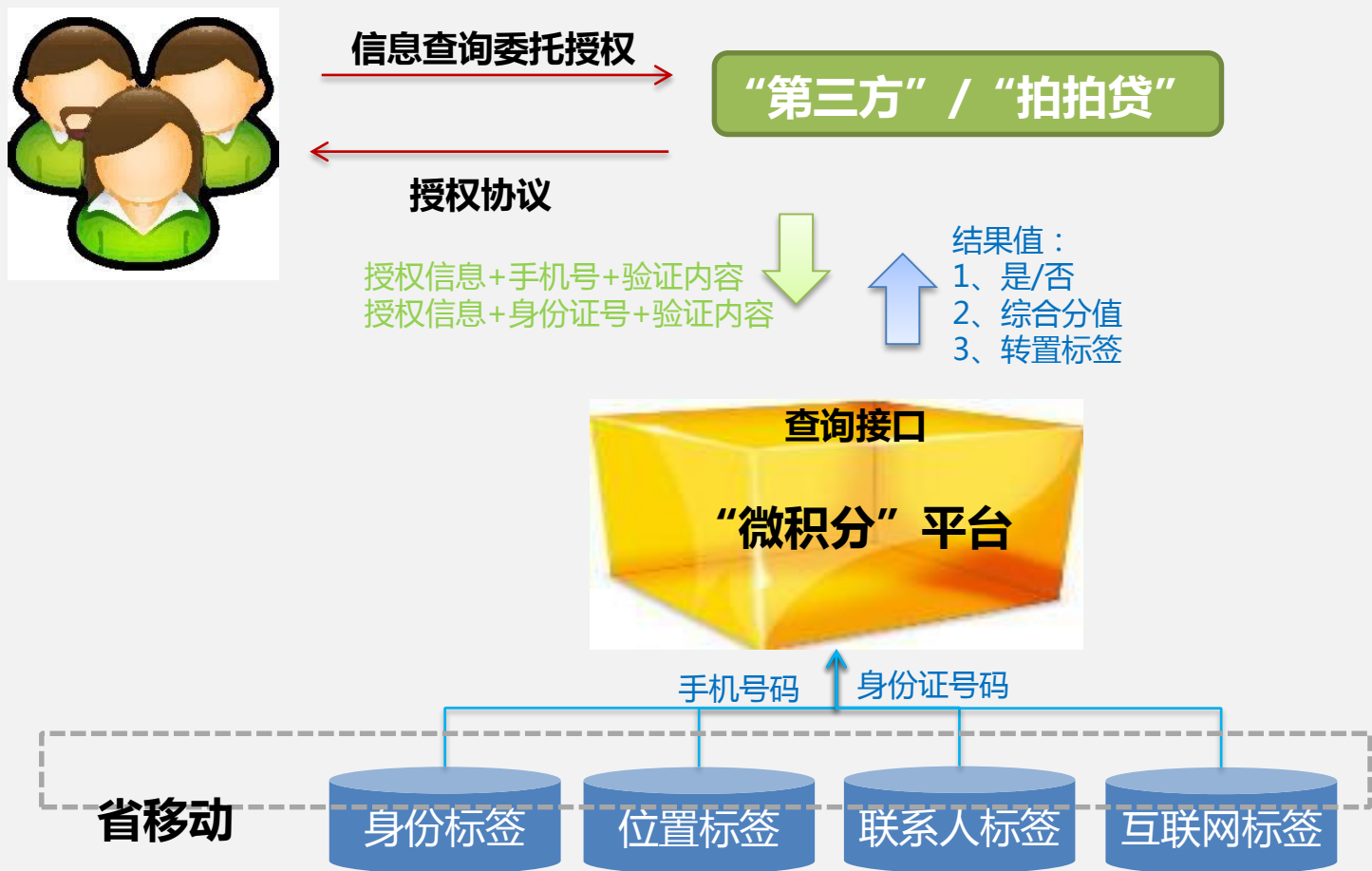
# 微积分产品体系架构



# 产品方案

# 信息验证服务——产品方案

信息验真服务：依托身份证号码/手机号码作为查询的**唯一标识**，由验真客户发起“**唯一标识+验真内容**”的**实时/批量**查询请求，征信平台验证请求的真实性后，在不侵犯用户隐私的基础上以结果值的方式将验证结果返回。

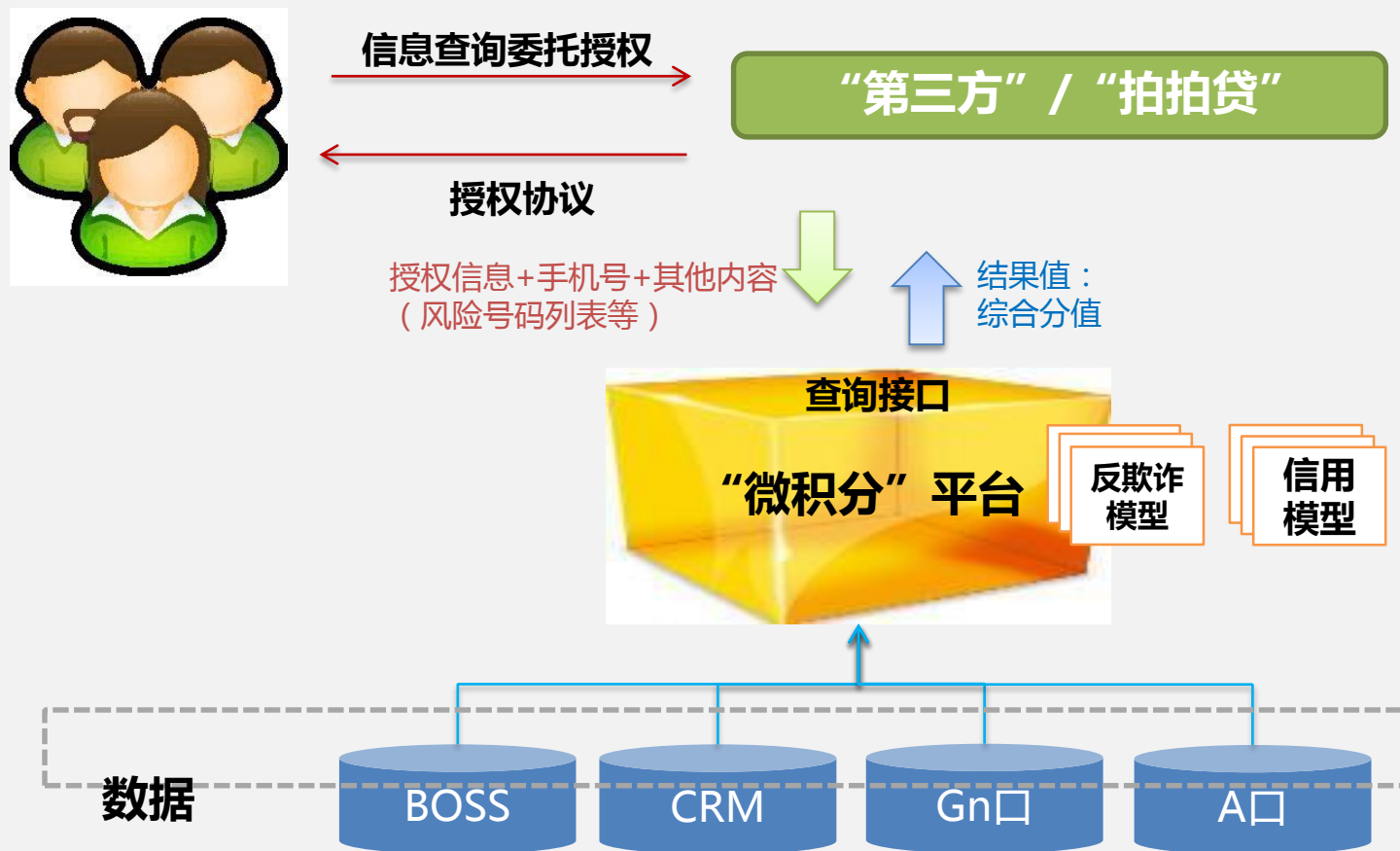


# 反欺诈服务——产品方案

保险欺诈是指在保险活动中采取欺骗手段侵犯保险合同当事人合法权益的违法犯罪行为。

保险欺诈主要表现为：1、投保人在投保时不如实告知；2、虚构或者伪造索赔；3、多次重复索赔等。

面对投保人的欺诈行为，承保人通过大数据应用开放平台，利用用户信息真实性和用户风险行为的验证，提前对用户的欺诈行为进行预判，避免欺诈事故的发生。





## 信用评级服务——产品方案

保险公司在开展业务时，需要针对个人信用信息进行评估，根据信用评级结果，来决定是否承保。

传统征信数据依赖于银行信贷数据及个人的固态数据，且覆盖人群范围狭窄；基于移动的海量个人通信数据及消费数据建立的个人信用评级模型，不仅直接反映用户信用好坏，而且覆盖人群广泛，结果更有指导意义。

个人信用评级：指信用评估机构利用信用评级模型对消费者个人信用信息进行量化分析，以分值形式表述，是对其履行各种经济承诺的能力进行全面的判断和评估。



# 初步搭建了以位置、金融风控为核心的大数据应用模型库

## 应用模型

### 位置类模型

用户通勤模型

**职住分析模型**

**客流预警**

商旅人士判定

常驻、流动人口分析

客流定点预测

### 用户行为识别模型

**购物行为分析模型**

内容行为分析模型

App特征识别模型

搜索行为分析模型

### 用户轨迹预测模型

**热门路线发现模型**

用户区域偏好模型

### 金融风控模型

**手机号是否正常使用模型**

异地盗刷卡识别模型

**贯众和信分**

用户国际漫游分析模型

朋友圈消费能力模型

劣质客户识别模型

**联系人亲密度模型**

### 基站识别模型

异常小区聚类模型

违规手机识别模型

基站位置定位模型

同源系列化模型

### 精准营销模型

兴趣偏好识别模型

**协同过滤推荐算法**

关联规则推荐算法

### 用户价值刻画模型

**高价值用户识别模型**

经济能力评估模型

### 文本分析模型

正文文本抽取算法

关键词抽取算法

## 能力模型

### 基础层模型

数据预处理算法

**泰森多边形算法**

自然语义处理算法

假设检验方法

加密方法

数值计算方法

### 挖掘层模型

**分类算法**

**回归算法**

聚类算法

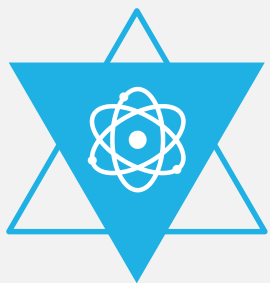
时间序列算法

频繁模式算法

推荐算法

# 模型算法库在金融风控领域的应用实践

- 对运营商自有的**用户自然信息**、**运营商基本信息**、**通信消费信息**、**通话/短信数据**、**上网行为数据**、**位置数据**建立起一套丰富而全面的指标体系，并已经在金融风控方向做了大量的尝试和实践，开发的**个人信用评分**，**欺诈风险评估**，**信息验真**及**贷后关键指标监控**等核心模型已经应用于合作金融机构的贷前贷后环节，逐步实现移动数据大数据的在金融征信上的深度应用。



## 敏感信息核验

可提供用户身份/消费/位置/社交/价值观5大维度，包括工作地验证，手机活跃度验证等33个关键指标验证，提供api及web界面两种形式数据传输。已接入上海及湖北两省数据，潜在用户包括浦发银行等金融机构



## 欺诈风险评估

结合金融机构场景欺诈场景，基于移动真实的用户身份，位置等关键数据以及杭研积累的黑名单数据，提供虚假材料申请欺诈识别，盗刷卡识别，黑名单关联等欺诈识别服务，帮助金融机构评估用户交易，信贷申请等关键环节欺诈评估。



## 贷后风险变化监控

针对银行贷后无法获取用户位置，生活状态变化等关键信息的痛点，结合运营商数据提供用户手机状态，用户位置状态等24项贷后关键指标监控，判断用户指标是否出现异常，当异常指标达到阈值时，即推送预警信息至银行负责人。



## 个人信用评分

通过对移动运营商数据建模，得到上百个与个人信用评分可能相关的指标变量，结合银行和金融机构已知的个人信用度，以优化的逻辑回归算法训练，得到能够识别信用度的规则，并用一个综合评分来度量用户的信用度，以此作为信用好坏的参考值。

# 通过改进的逻辑回归算法建立信用模型

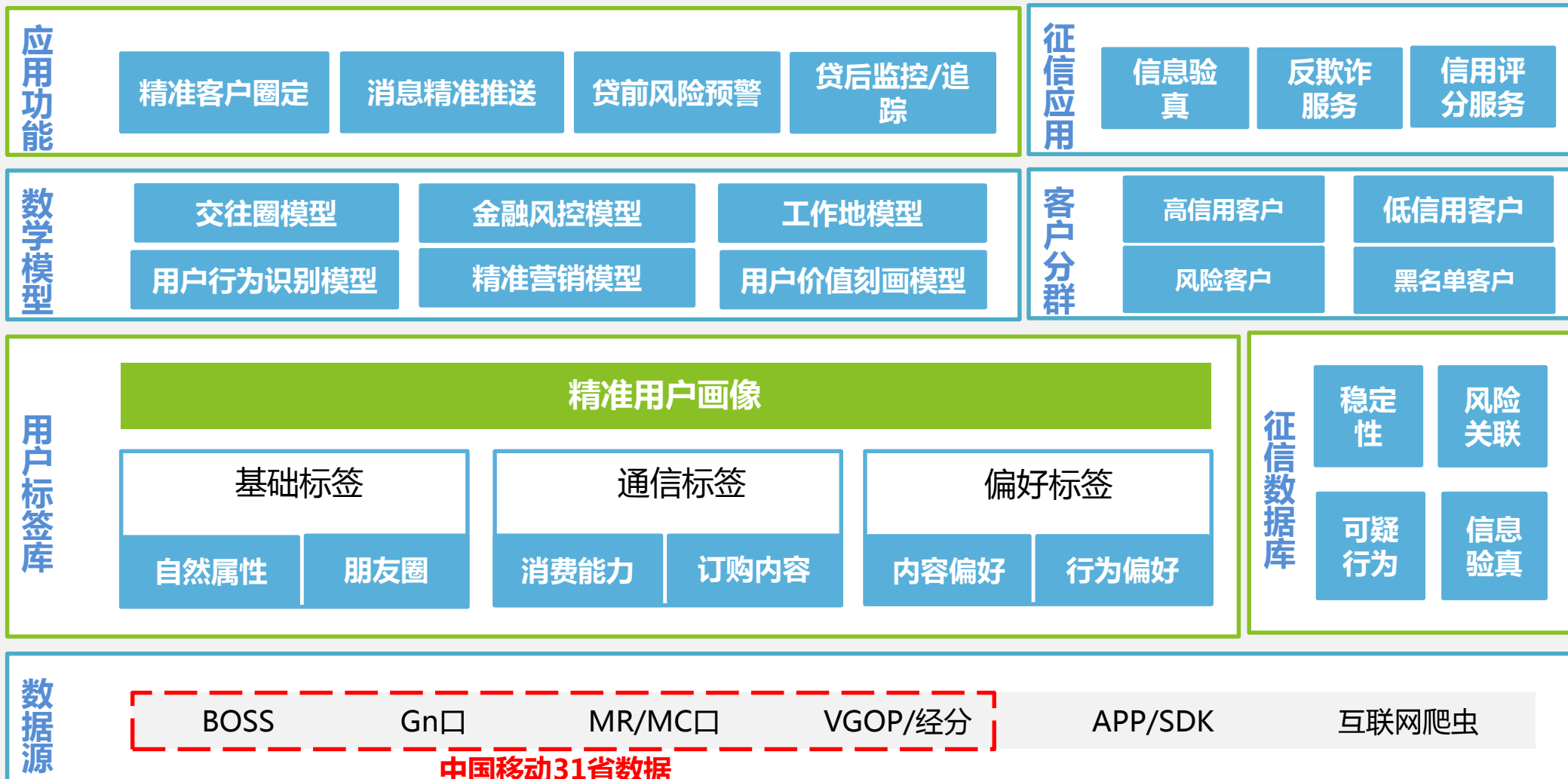
以运营商自有的用户自然信息、运营商基本信息、通信消费信息、通话/短信数据、上网行为数据、位置数据的上百个指标变量为基础，通过建模得到一套全面刻画用户行为面貌的画像体系，用于**金融风险评估**；通过数据挖掘中特征选取算法提炼出**身份特质、忠诚度、通信消费能力、固定资产、线上消费能力、线下消费能力、价值观、社交人脉**等8个与消费能力、信用评估强相关特征维度；通过优化后的**逻辑回归算法**进行训练，得到识别信用度的规则，并使用一个综合评分来度量。

$$Y (\text{信用分}) = F (X_1, X_2, \dots, X_n)$$



# 微积分数据平台功能架构

- 利用基于运营商位置数据、DPI数据、网络爬虫数据、自有客户端SDK数据等搭建动漫底层标签库，业务驱动模式建立数学模型，并基于场景划分用户群。



# 核心数据积累

## 爬虫网站

- 已爬取视频、阅读、动漫、音乐、APP、新闻、房产 等7大主题61个网站
- 新增爬取旅游、生活、游戏、汽车、招聘、征信等6大主题40个网站
- 目前已爬取3200w条有效数据，新增爬取3000w条数据
- 每个网站做到每2周刷新一次最新网站数据信息同步至数据库

## 用户标签库

- 开发多种用户标签算法，给用户从12类维度上打标签

## DPI解析

- 已解析7大主题下52个主流APP的83个用户操作行为
- 例如：优酷视频APP，用户“收藏”笑傲江湖电影行为
- 增加对5大主题下，50个主流APP的用户操作行为

## 标签、关键词库

- 已开发并调试TF-IDF标签、关键词算法，沉淀用户标签、关键词信息
- 尝试多种算法，并调优，沉淀用户标签、关键词信息

# 已对接中国移动31省公司用户数据

## 用户基本数据

号码  
身份信息  
缴费信息  
消费信息



## 移动位置数据

基站信息  
移动信息  
滞留信息  
.....



## 信令数据

主叫信息  
被叫信息  
IP信息  
位置信息



## 终端数据

Mac地址  
终端ESN  
APP信息  
.....



## 网络日志数据

移动 / 固网源地址  
目的地IP  
目的地URL  
.....



## 用户触点数据

10000  
网厅  
.....



## 中国移动 数据优势

- 覆盖广：**对接中国移动31个省，包括8亿多的移动用户数**
- 种类多：包括位置信令数据、用户网络数据、呼叫记录数据、终端数据、视频数据、...
- 体量大：中等规模本地网2G/4G用户日上网记录达到数TB以上
- 质量高：最真实的客户、产品、资源和订单等数据

谢谢!

未经许可，不得转载或商用!