

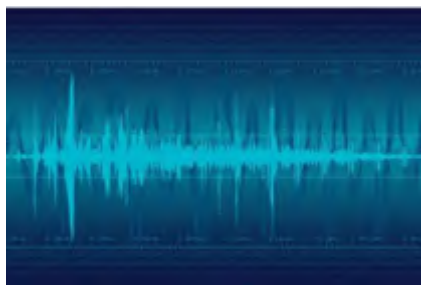
讯飞大数据的实践与思考

科大讯飞大数据研究院

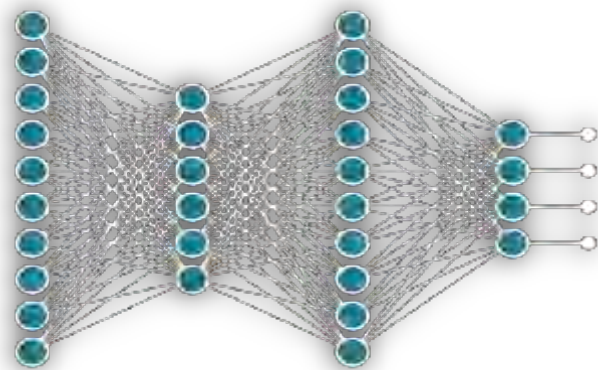
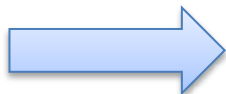
谭昶 博士

从语音识别技术谈起

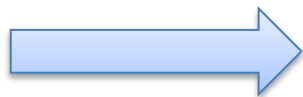
让世界听见我们的声音
让世界听见我们的声音



数万小时标注语音数据

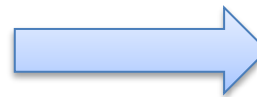


海量用户语音输入



讯飞开放平台

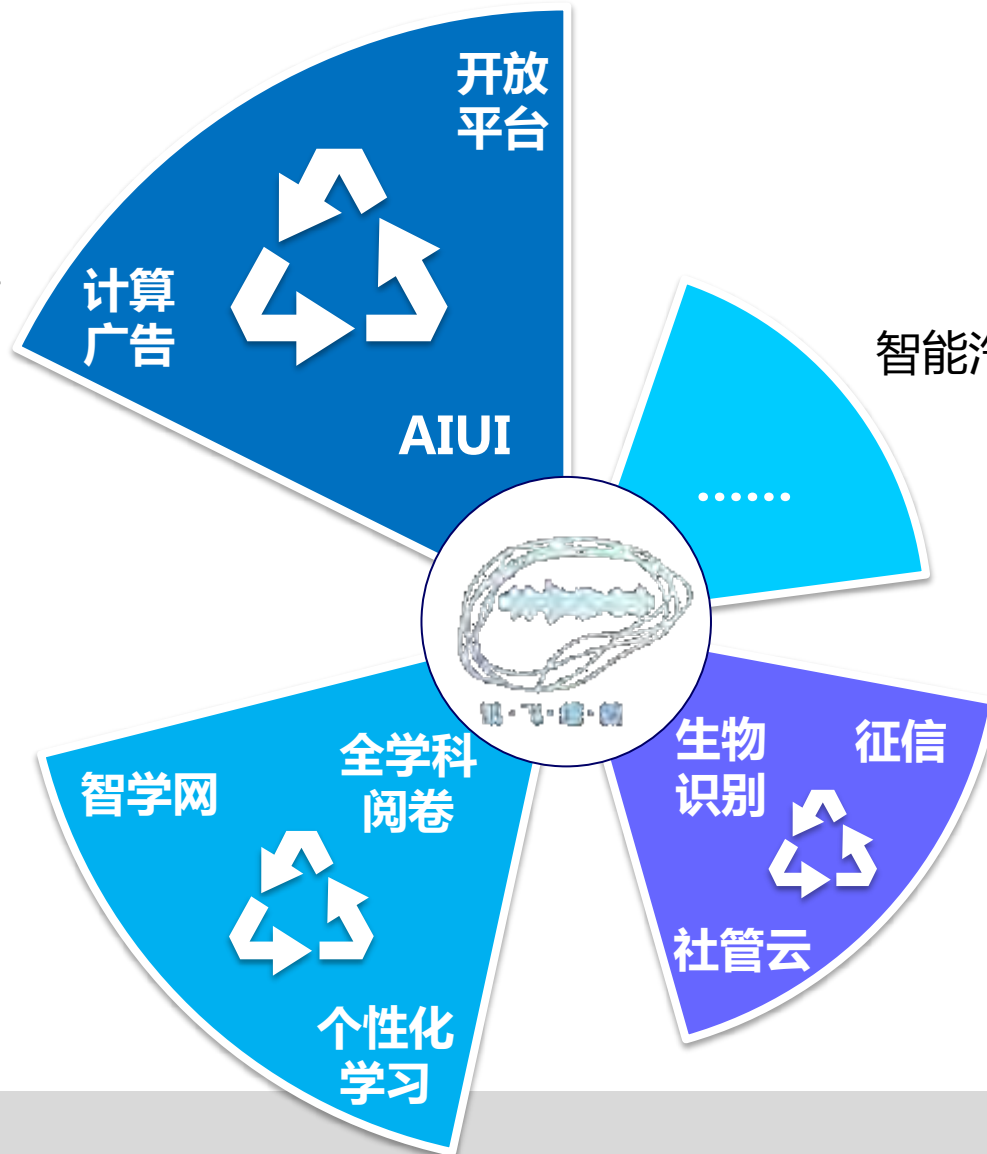
- 处理能力：30亿次/天
- 准确率：超过97%



大数据让人工智能更加智能
人工智能让大数据更有价值

人机交互

- 8.9亿用户日均交互 **30亿次**
- 输入法3.6亿用户，语音输入超过**12%**
- 广告算法决策较人工运营成本降低**1.2倍**



教育

- 覆盖全国26省106市近**10000所学校**
- CET考试作文阅卷 **超过人类专家**
- 个性化学习试点成绩平均提升**10%**

智慧城市

- 社管业务服务安徽、江西、吉林等**数十个城市**
- 前台办理人员减少**50%**，办理时间缩短为2-3天

海量真实的大数据资源

让世界聆听我们的声音
科大讯飞
iFLYTEK

海量用户



语音云

- 覆盖**8.9亿**用户，月活**2.36亿**
- 日交互次数**30亿次**



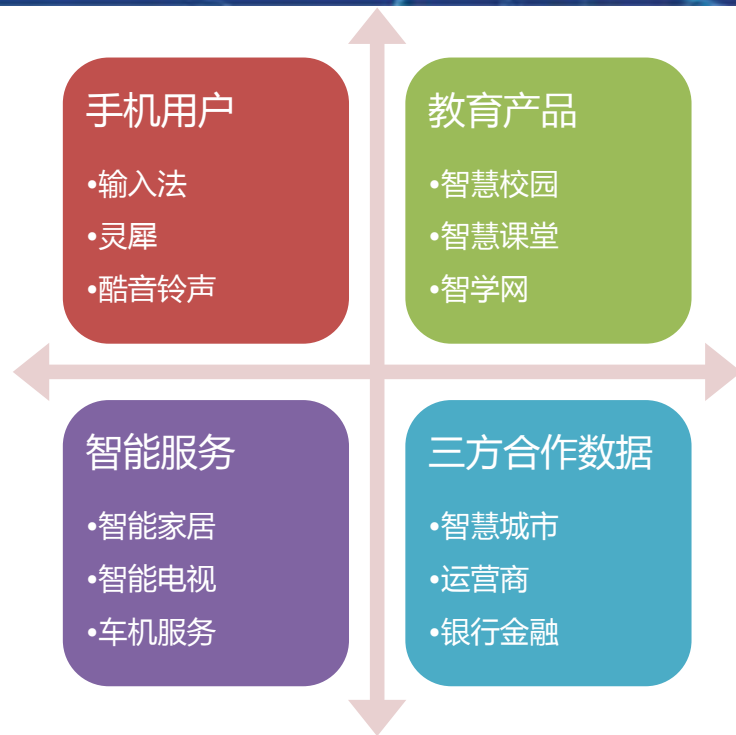
讯飞输入法

- **3.6亿**用户，月活超**1亿**



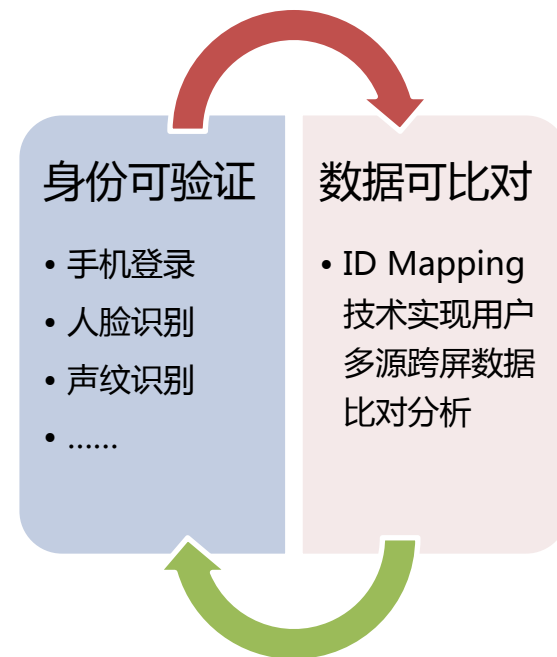
灵犀语音助手

- **1.2亿**用户，月活**5000万**



多源场景

真实可靠



讯飞语音云——自有大数据源泉

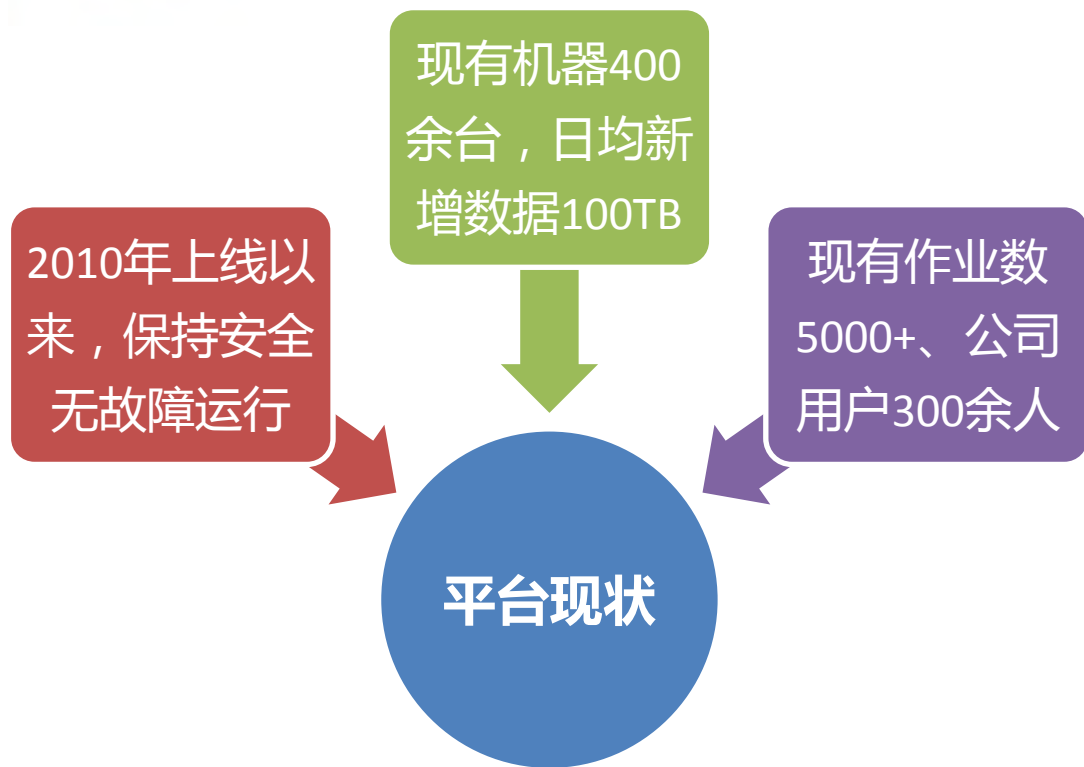
让世界聆听我们的声音
科大讯飞





Odeon(奏乐堂)

讯飞以语音为特色、希望大数据平台提供一个场所，把各种数据演奏出美妙的音乐



平台特色

以数据为中心

- 实现用户间数据隔离和授权访问，保障数据安全

整合人工智能能力

- 基于CPU+GPU的混合架构，整合了讯飞人工智能技术

允许私有化部署

- 平台已形成成熟产品，为公司客户实现私有化部署

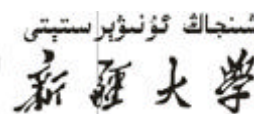
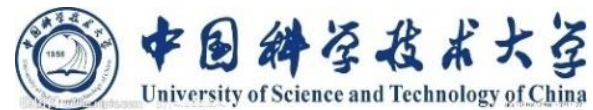
优秀的大数据技术团队



• 高水平的大数据研究和应用团队

- **大数据首席科学家**：陈恩红教授（中科大计算机学院副院长、杰青、CCF大数据专委会副主任）
- **讯飞大数据团队**：由近10位博士领衔，100余人的专业团队针对计算广告、个性化推荐、智慧城市、个性化学习、金融征信等领域开展大数据研究和应用工作。国内一流的大数据前沿研发队伍，帮助讯飞和客户共同成就大数据。

• 广泛的产学研合作



讯飞将大数据技术应用到多个领域

让世界聆听我们的声音
科大讯飞 FLYTEK



精准营销



个性化学习



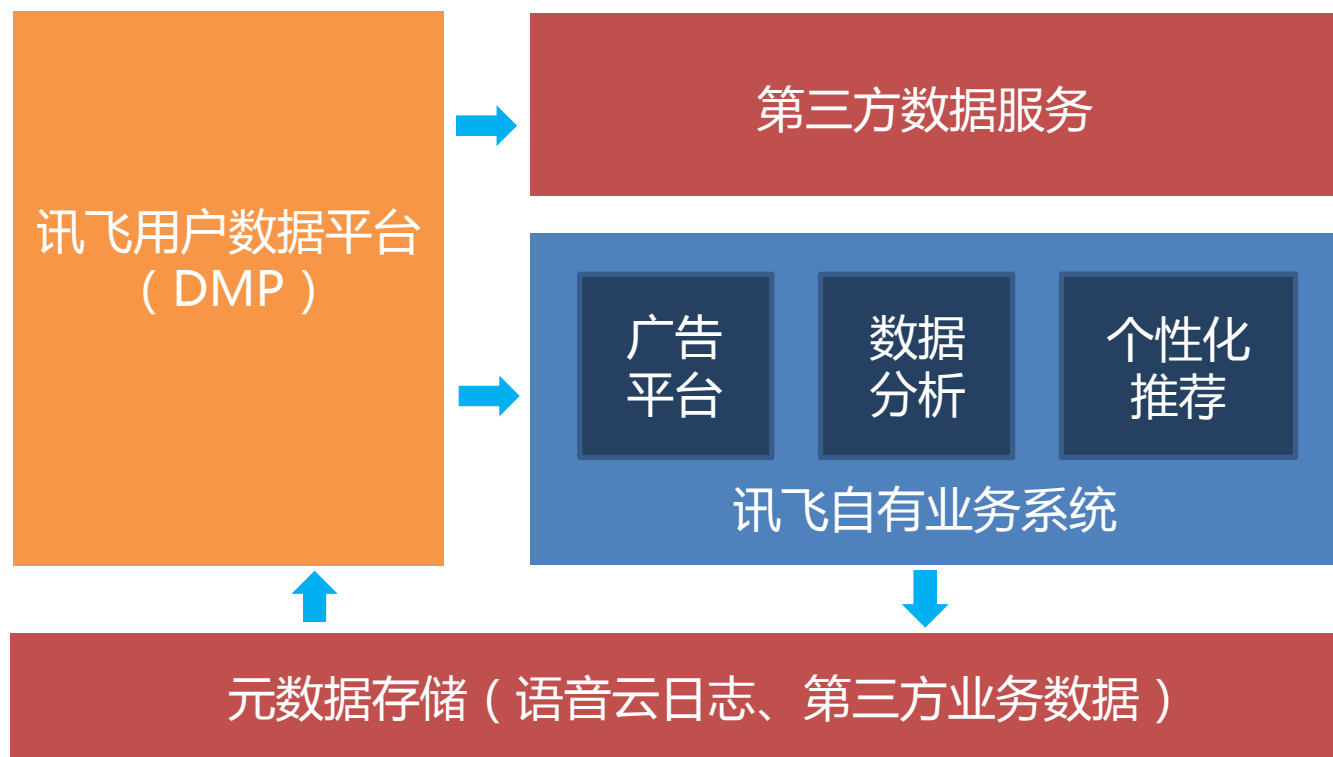
智慧城市

讯飞用户数据平台 (DMP)

让世界聆听我们的声音
科大讯飞 FLYTEK

能力输出

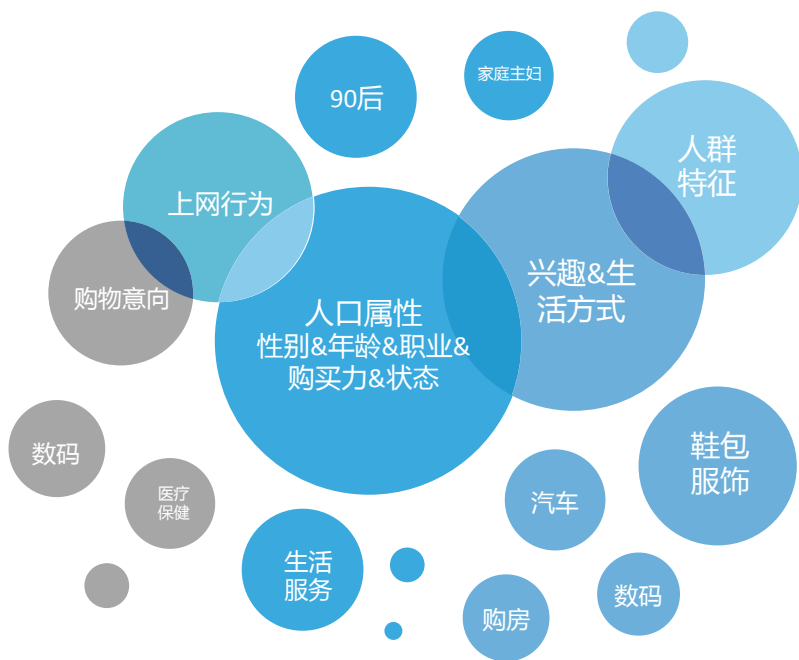
核心价值



用户画像实现数据增值

让世界聆听我们的声音
科大讯飞

结合讯飞大数据与人工智能技术构建的讯飞DMP平台。已经完成人生阶段、行业偏好、购物兴趣、媒介兴趣5个大类**1700个子标签**的用户填充工作。累计覆盖12亿终端设备。



精准营销实现数据价值变现：金融

让世界聆听我们的声音
科大讯飞 飞



客户

某金融理财产品推广



人群定向

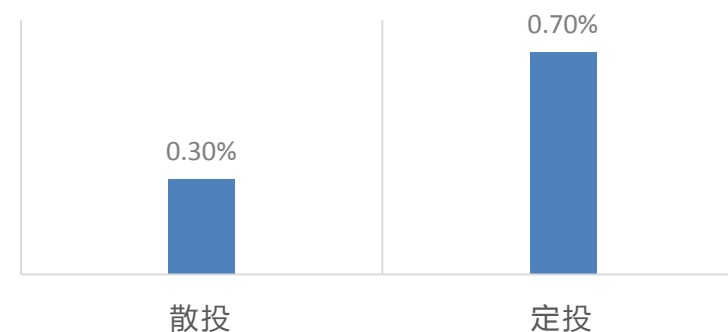
男性
广州地区
第三方网贷投资类产品用户群体



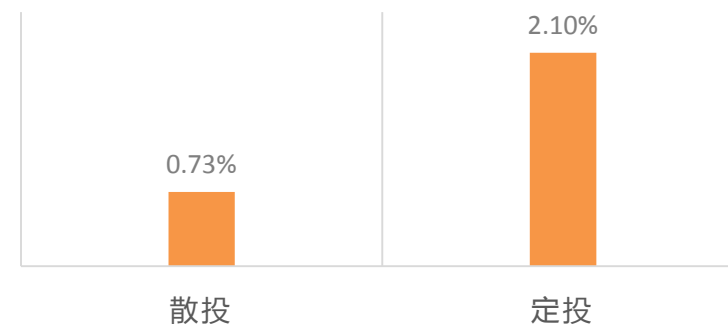
转化方式

有效注册（预留电话）

页面点击率



有效注册转化



点击率提升 **130%**，有效注册转化提升 **187%**

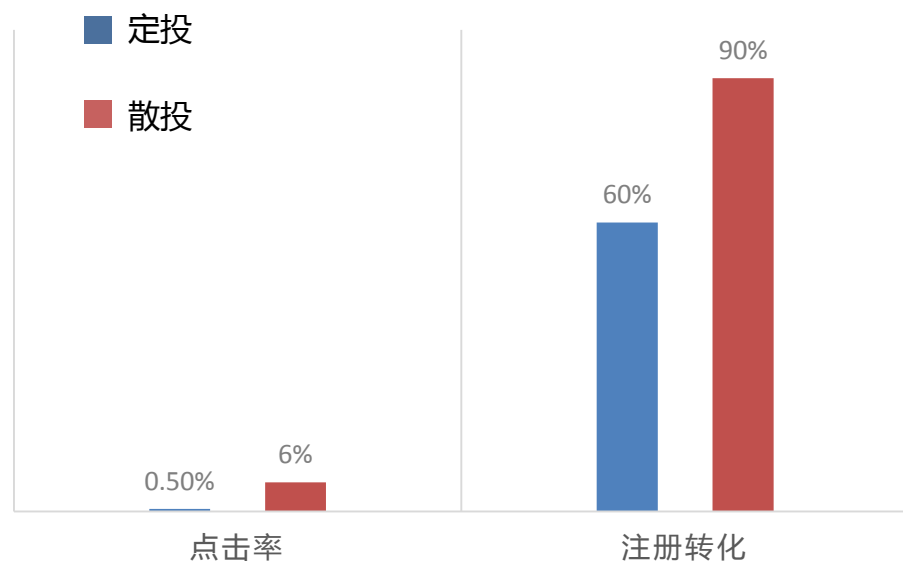
精准营销实现数据价值变现：游戏

让世界聆听我们的声音
科大讯飞 FLYTEK

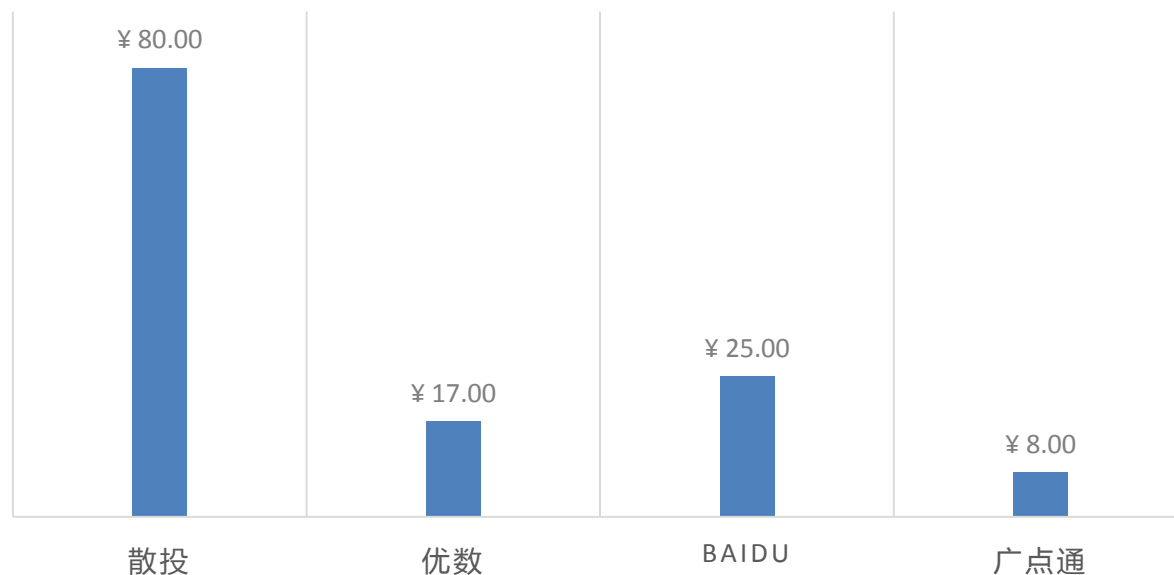
优化策略：针对5家游戏媒体，6个效果类投放，按游戏类别针对游戏人群进行投放。

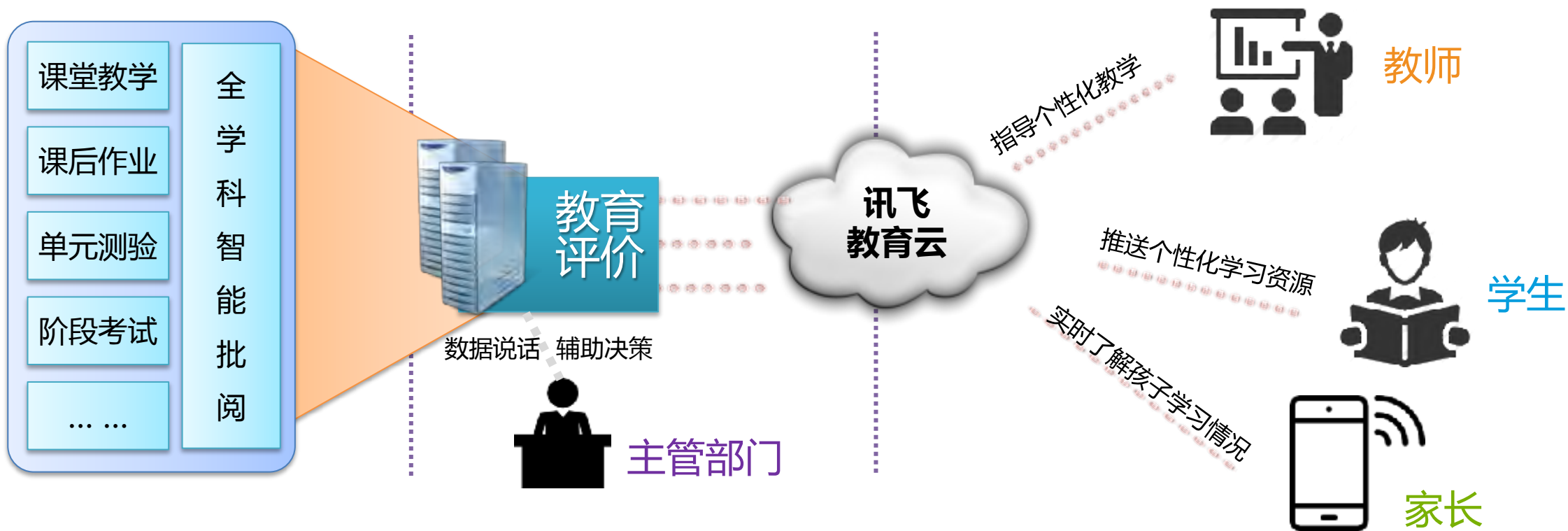
优化效果：日匹配曝光PV**3600万**，匹配点击率**提升10倍**，注册转化**提升50%**，单用户注册成本**降低4倍**。

点击与注册转化



注册成本降 4倍





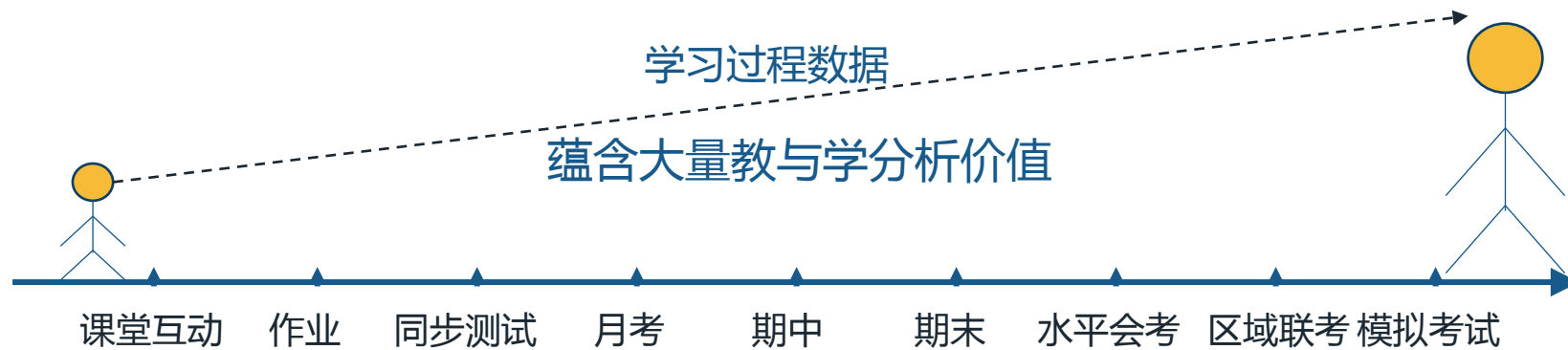
1 教学过程数据
采集与汇聚

2 教育大数据
自动分析与评价

3 应用大数据实现
个性化教学

教学大数据的伴随式采集

让世界聆听我们的声音
科大讯飞



全国29省200市近6000所学校使用，受益师生1000万+



日常作业	随堂检测	假期作业	校园考试	区域联考
教辅作业	答题卡 卡卷合一	线上录播 线上直播	答题卡	答题卡

人工智能实现教学过程数据全量采集

让世界聆听我们的声音
 让世界看见我们的努力

扫描试卷

识别结果

成绩



2015年7月合肥会考和安庆会考中英文作文评分技术试点成效

相关度	机器	专家1	专家2	专家3	现场评分
合肥	0.786	0.769	0.758	0.744	0.760
安庆	0.937	0.929	0.923	0.915	0.900

相关度	机器	专家1	专家2	专家3	现场评分
合肥	0.882	0.825	0.814	0.798	0.800
安庆	0.930	0.920	0.920	0.916	0.843

全学科阅卷让学习过程数据全量采集成为现实

智学网打造个性化学习平台

让世界聆听我们的声音
科大讯飞 FLYTEK

姓名：王小明
学号：121302

数据标签

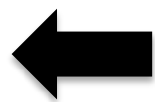
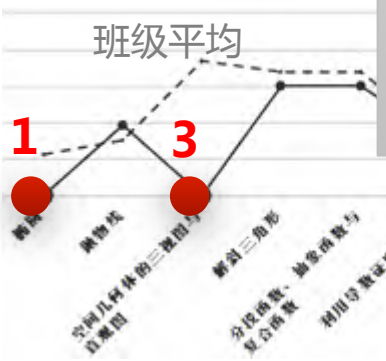
作答情况
难度
知识点
解题方法
.....



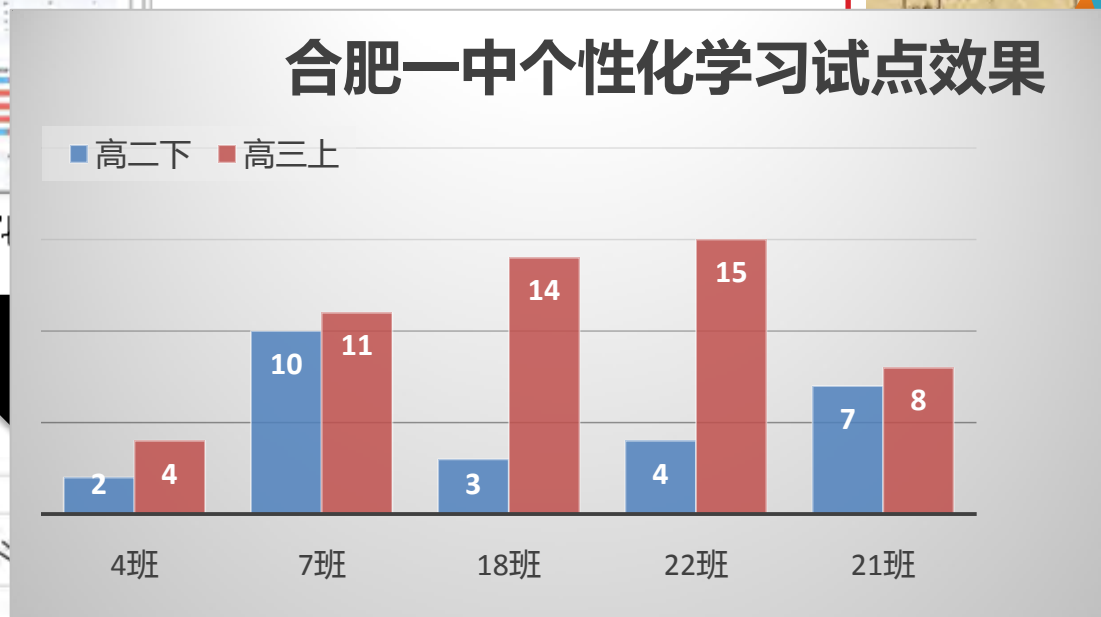
过程化

粗心?
不熟?
不会?

班级平均



合肥一中个性化学习试点效果



十万套密卷
400万道试题

习题精练



页

“解析几何”
出问题需要从
“二元二次方程”补起

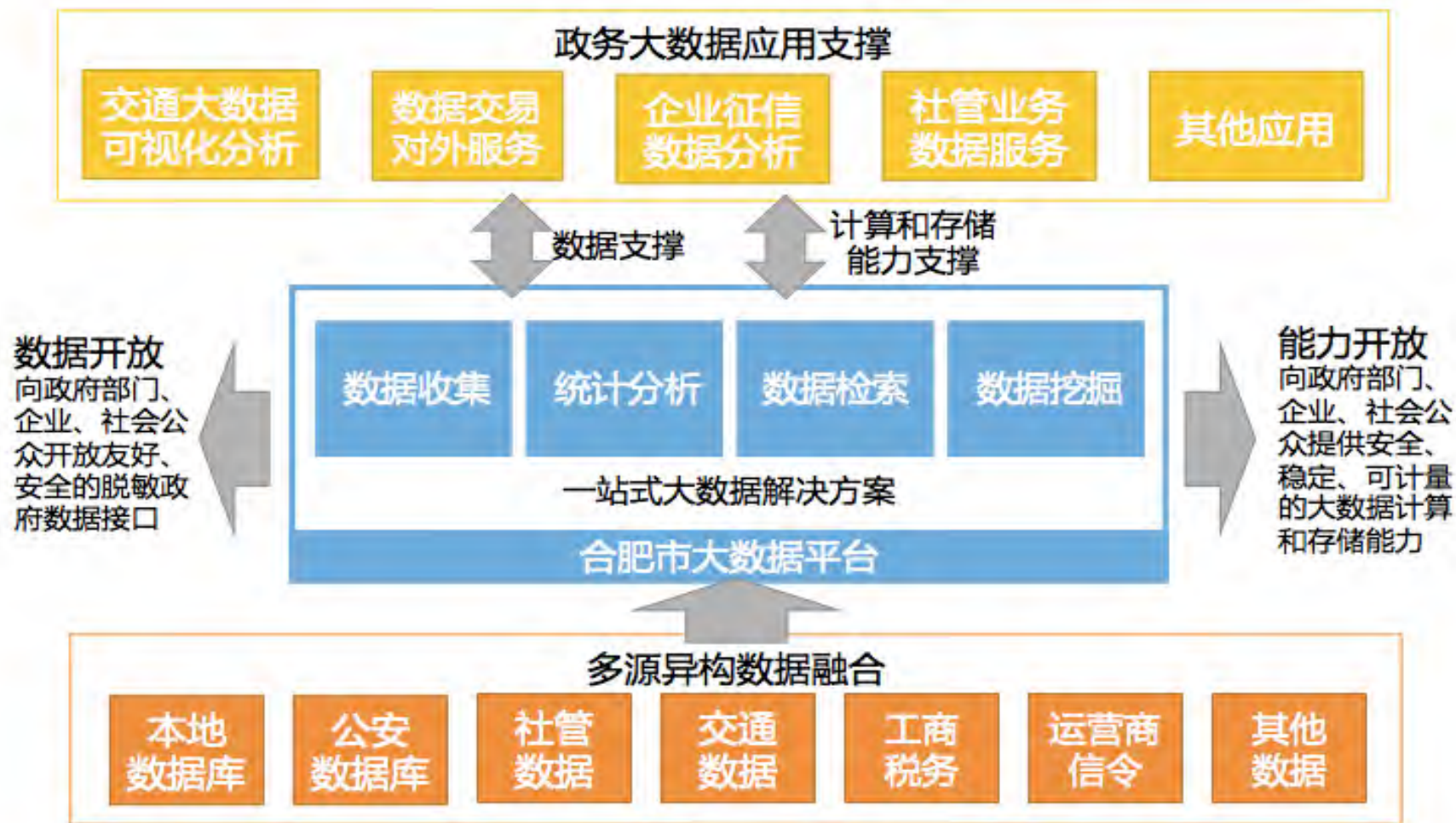
3、7、11、13
>圆锥曲线>>椭圆
>空间几何体的三视图与直观图

7、函数与导数>>导数的综合运用
11、统计与概率>>排列组合与二项式定理
13、不等式>>线性规划

建设城市大数据平台实现数据贯通

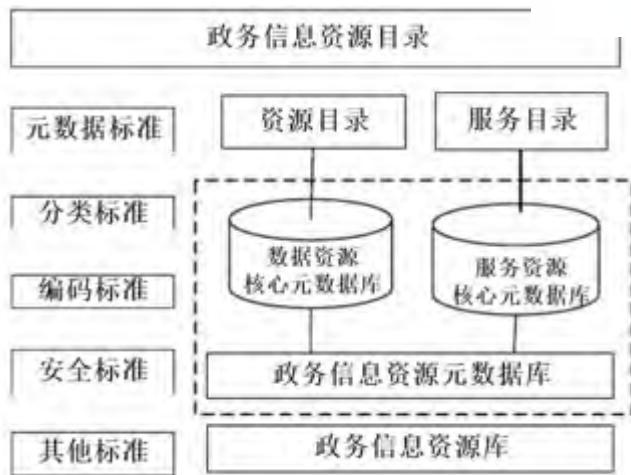
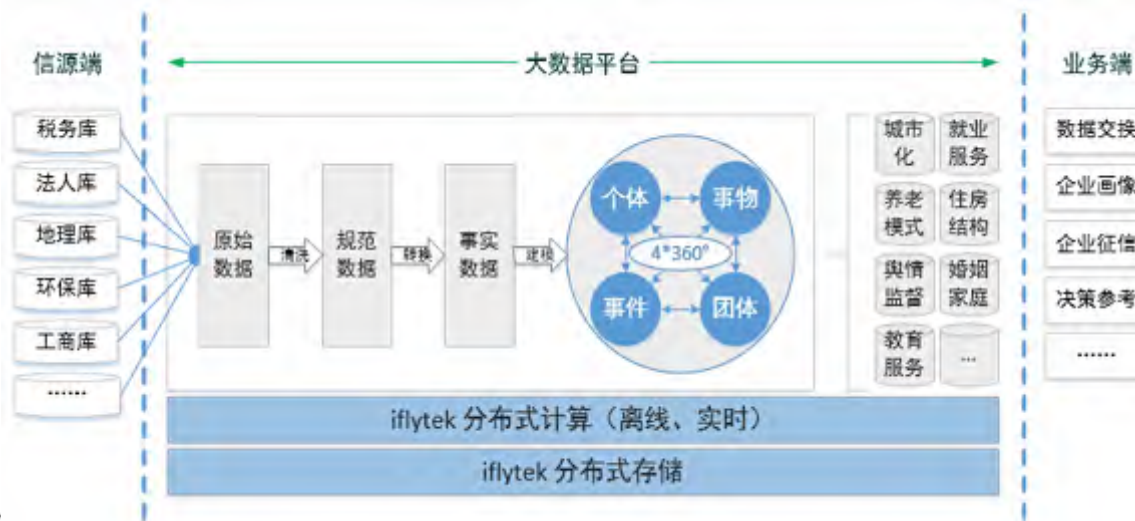
让世界聆听我们的声音
科大讯飞 FLYTEK

- 数据汇聚
- 行业互通
- 数据脱敏
- 数据清洗
- 数据加工
- 数据挖掘
- 数据支撑



建立数据规范体系实现数据共享

让世界聆听我们的声音
科大讯飞 FLYTEK



数据组织模型(4360)



目录体系

共享交换体系

大数据提升政府服务效率

让世界聆听我们的声音
让世界了解我们的声音

便民、便企数据全部打通，政府权利清单全面覆盖

- 业务：
 - 建立便民、便企权力清单库，为民、为企事项6727项。
 - 在全国20余地市，汇集503个部门5870类政务数据资源。
- 效率：
 - 窗口由原来8-10个减少到现在2-3个，数量减少70%，前台办理人员减少50%
 - 事项办理由原先10-30天，缩减到现在1-5天，办结率99.7%，越来越多的事项当场办结



政务数据打通实现“零上传、秒办理、一站受理、一点办结”

大数据提高市场监管水平

让世界聆听我们的声音
科大讯飞

企业画像：企业的标签体系、信用评级、风险等级、企业关系图谱以及企业评估报告的展现，对企业在政务监管、金融机构的风险评估等层面均能提供能力支撑。

企业360°信息一站全览

企业关系图谱深度洞察

企业奖惩信用一处查看

企业风险事件

企业画像

企业标签体系

- 企业舆情风险
- 企业经营风险
- 企业金融风险
- 失信企业
- 僵尸企业

监管

- 风险实时展示
- 多维统计分析
- 自定义预警

协同监管 联合惩戒

金融机构 工商 税务 质监 食药监 环保

监管状态、结果

大数据实现交通精细分析

让世界聆听我们的声音
科大讯飞 FLYTEK

海量用户时空数据

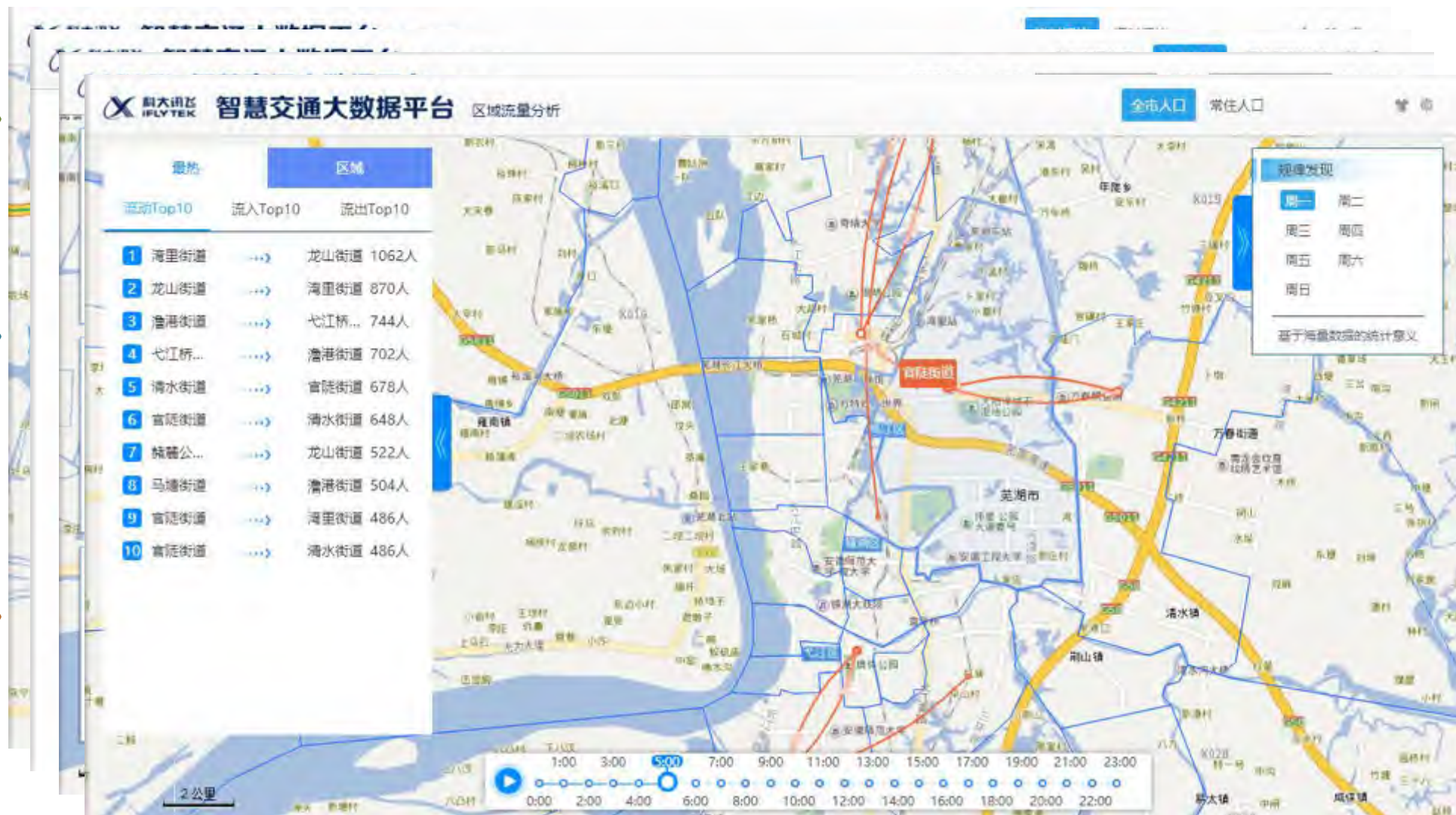
- 某地级市30%人口覆盖
- 每天1.2亿条时空日志
- 每天500万条移动轨迹

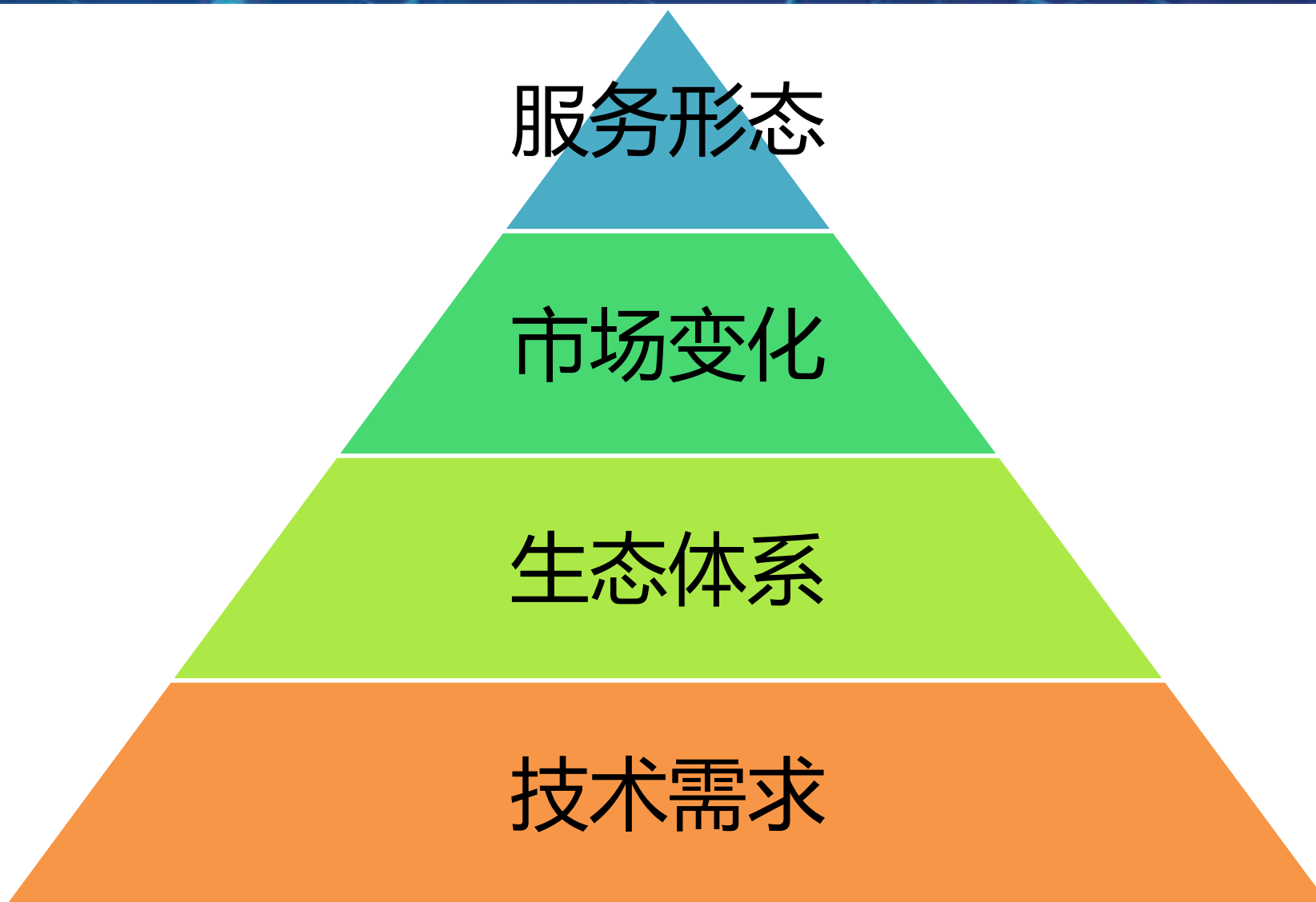
实时上报位置数据

- 打通交警、公交、出租数据
- 汇集运营商数据、讯飞自有数据
- 位置数据实时上传

实时流式计算平台

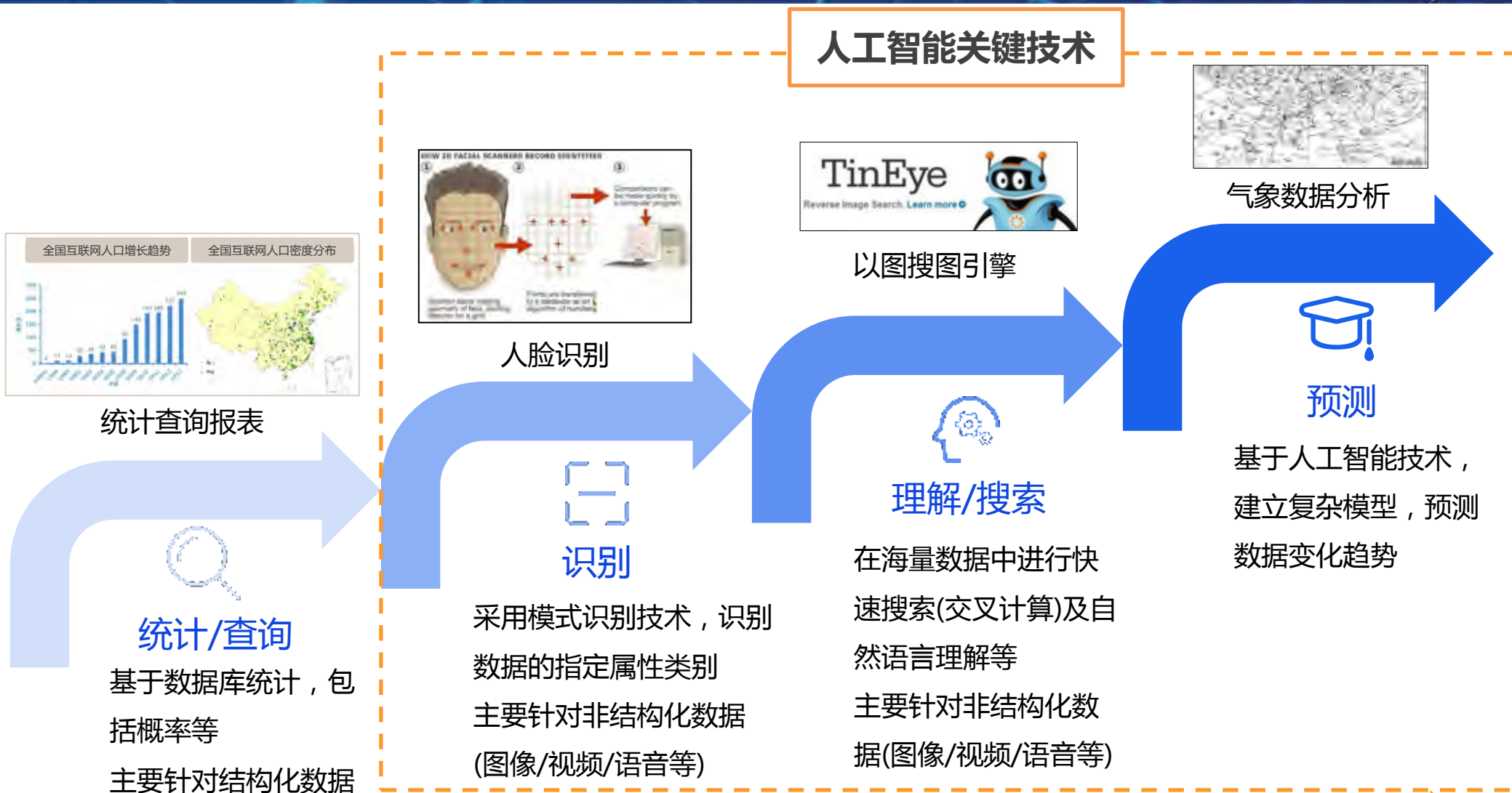
- 100GB/小时计算能力
- 每5分钟更新全市数据
- 支持1-3年历史数据回溯





大数据价值挖掘由易变难

让世界聆听我们的声音
让数据说话



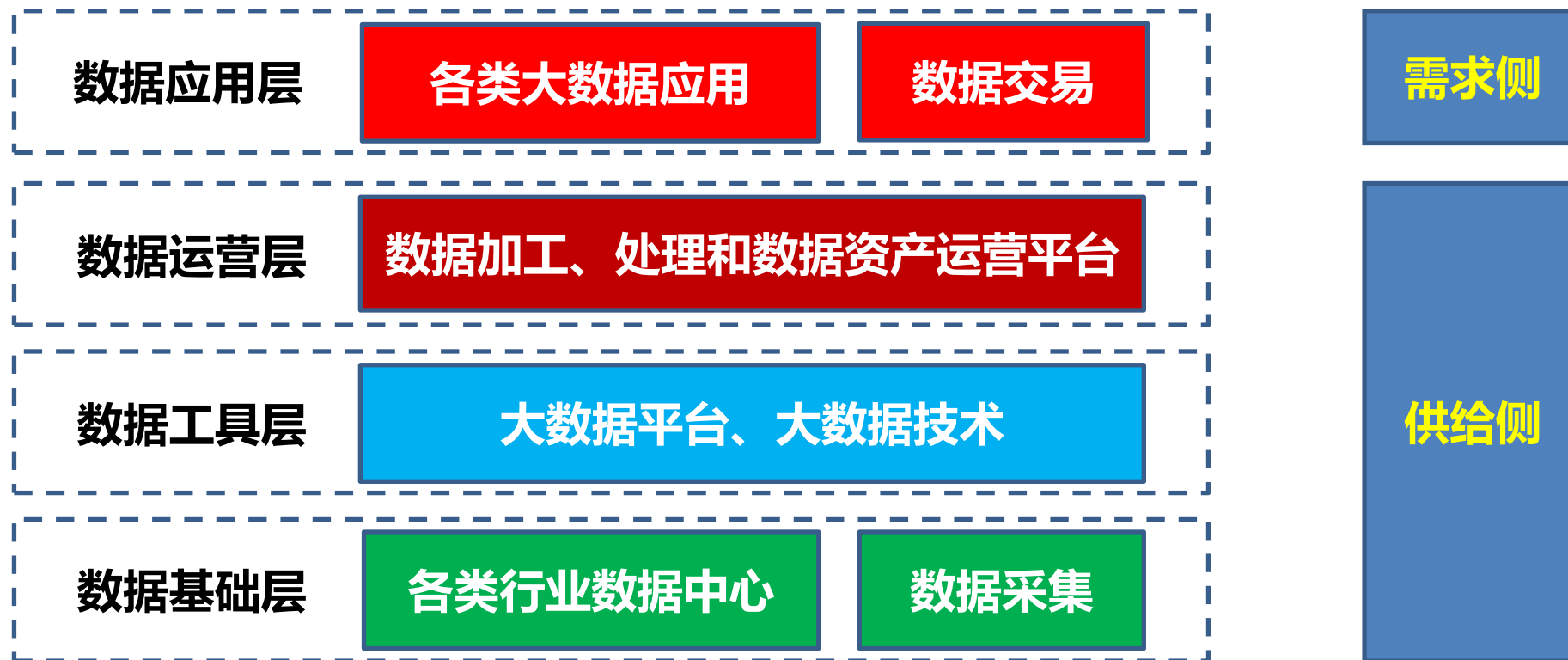
人工智能关键技术

易

难

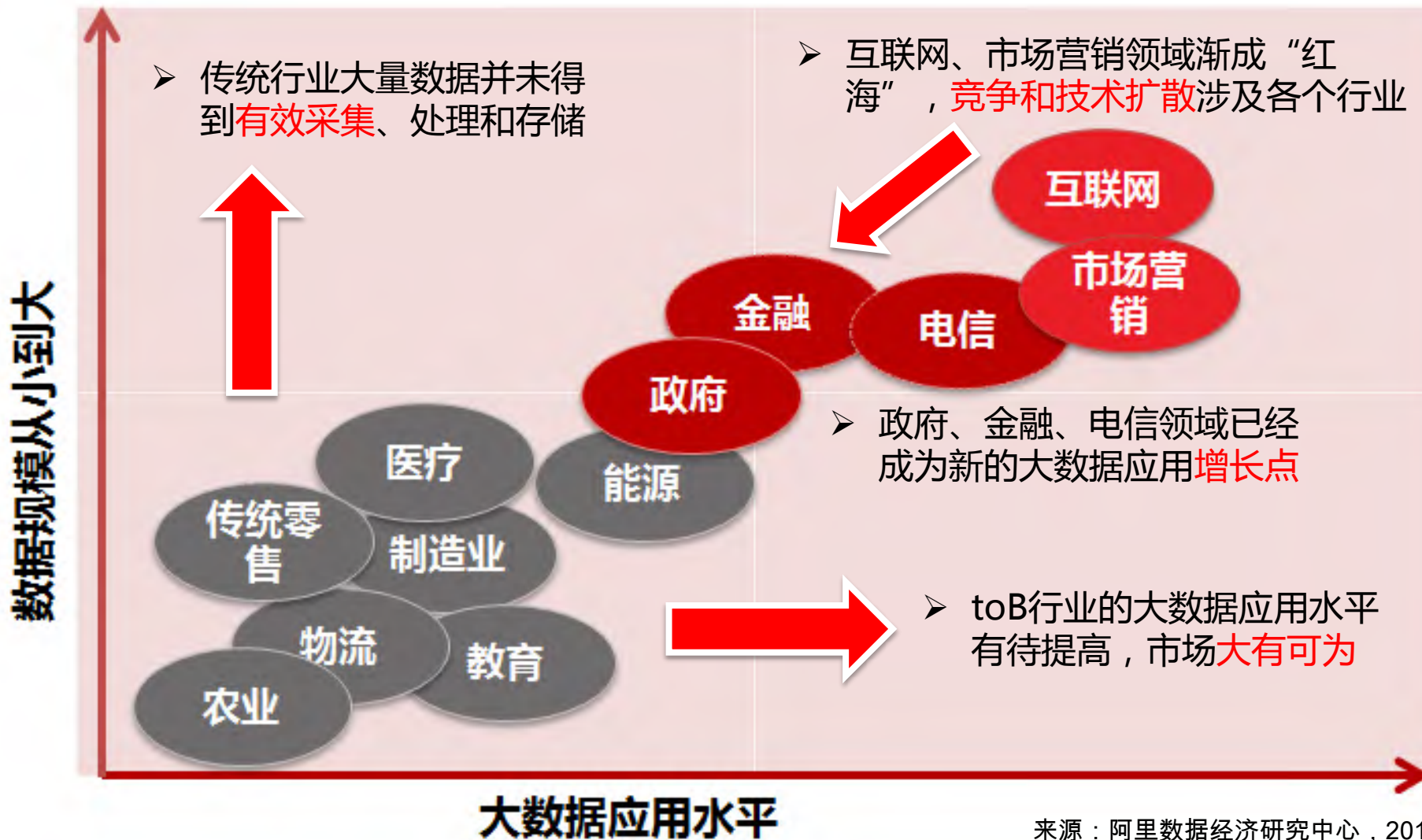
大数据产业生态逐渐完善

让世界聆听我们的声音
科大讯飞



传统产业大数据市场尚待深挖

让世界聆听我们的声音
科大讯飞 FLYTEK

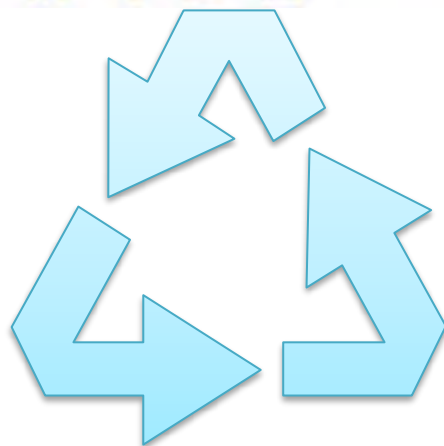


来源：阿里数字经济研究中心，2015



讯·飞·超·脑

- 人工智能的自我学习需要海量数据的输入
- 大数据的价值发现需要高效的人工智能方法



- 人工智能技术的进步离不开计算能力的不断增长
- 云计算让人工智能服务无处不在，触手可及

- 大数据的存储处理需要云计算基础设施的支撑
- 云计算环境生成和收集海量原始数据



未来的信息化服务将从海量数据中自我优化、
在日常生活中触手可及的智能化信息服务

大数据时代 不可不为，大有可为！

谭昶 博士

科大讯飞大数据研究院 副院长