

# 大数据技术在三好网的应用实践

卫向军 2017/8/19

# 目录

- 三好网介绍
- 大数据平台在三好网的发展历程
- 第一阶段：个性化搜索、推荐
- 第二阶段：多维数据分析提升平台运营效率
- 第三阶段：人工智能提升服务效果和质量
- 经验总结



成立于3年前，是全球首家C2B2C教学服务平台，年初获得A+轮融资，面向家长、学生、老师，提供中小学在线1对1课外辅导服务，目前月流水数千万。

## 线下教育痛点：

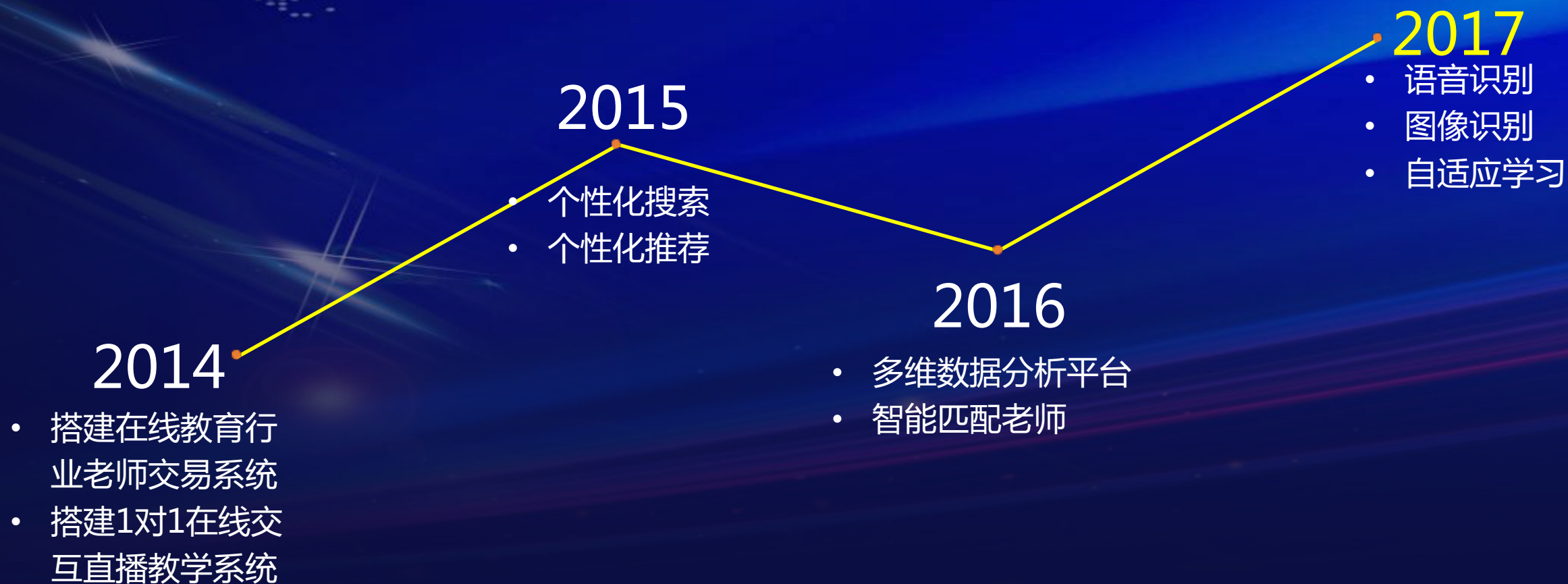
上课效率低、效果差 => 在线教室

教育资源不均衡 => 在线平台对接老师、学生

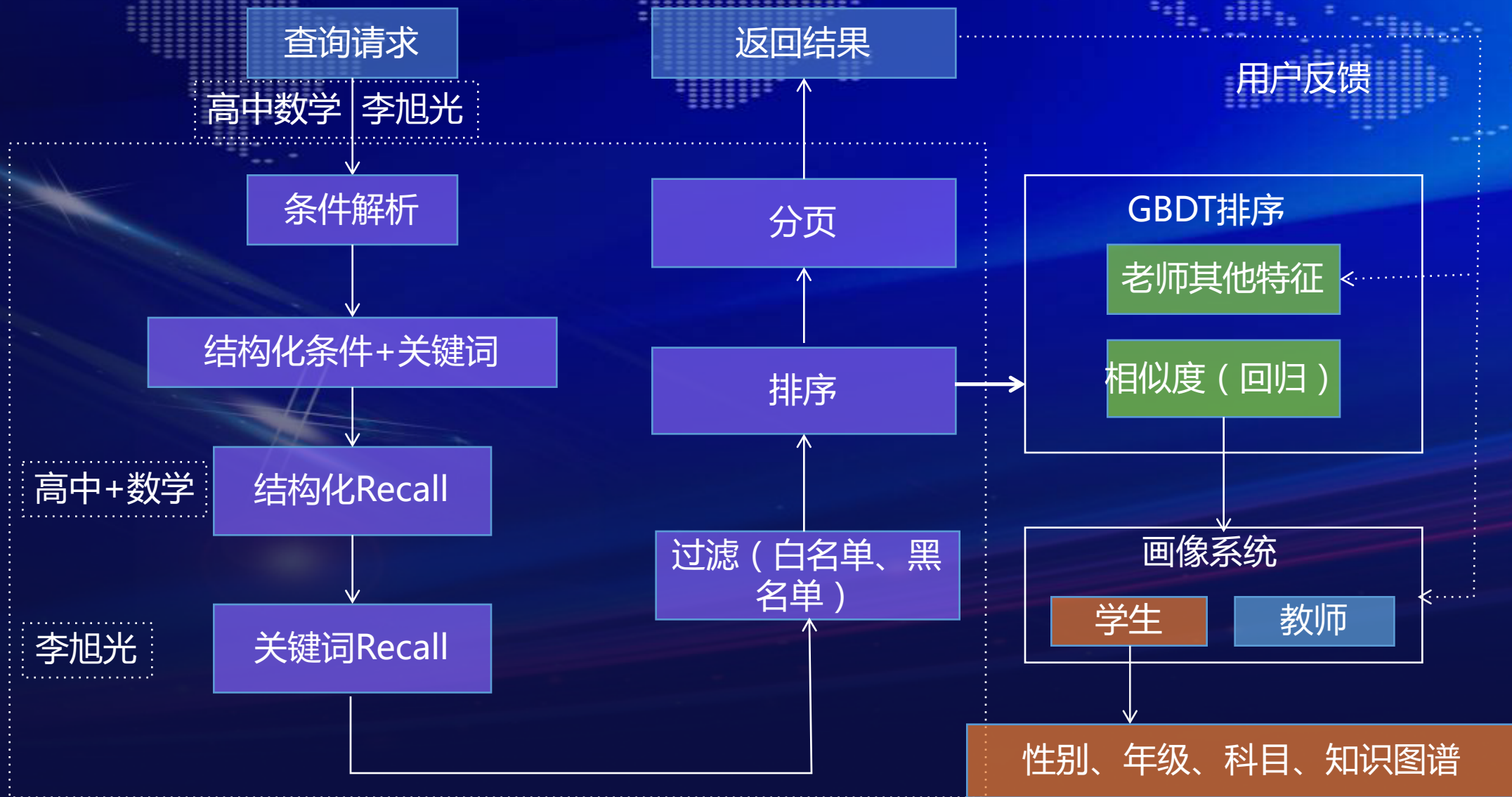
招生和场地租金高 => 在线低成本



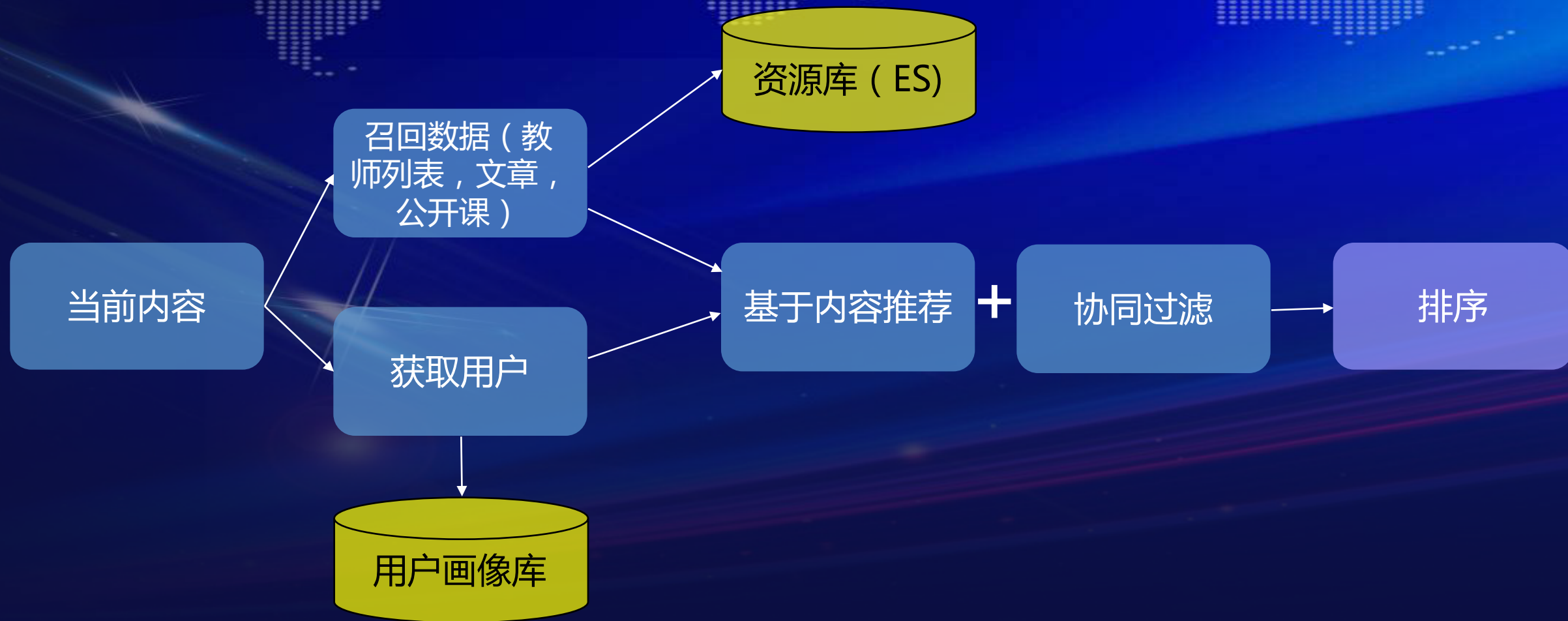
# 大数据平台在三好网的发展与经验总结



# 个性化搜索（结构化搜索+大搜索）



# 推荐引擎 = 内容推荐 + 协同过滤





# 教师匹配



# 搜索推荐系统的几点经验

- 结构化搜索+大搜索在电商平台的应用
- 按产品维度垂直拆分（1对1、公开课、教育资讯）
- 推荐场景(扩科、续费)，扩科准确度更高
- 人工匹配教师->系统匹配教师（结构化搜索）->精准匹配教师（大数据）
- 业务系统先行，大数据迭代替换



# 数据化运营-高效运营的基石

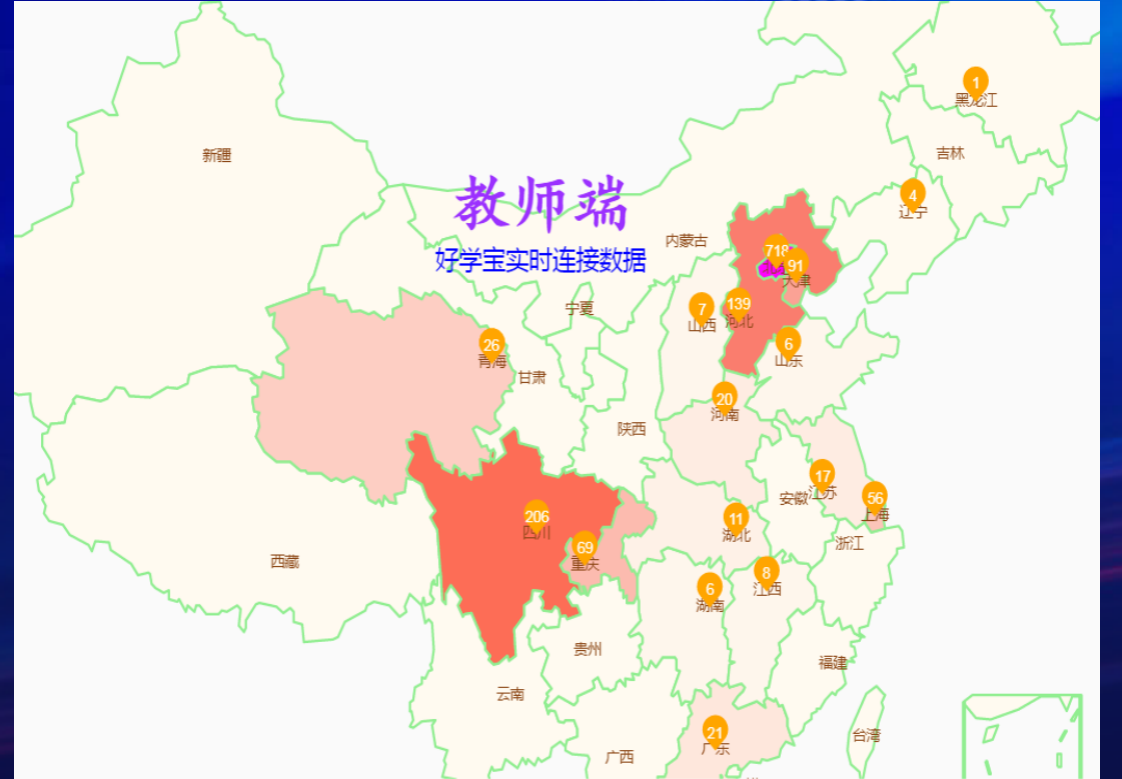
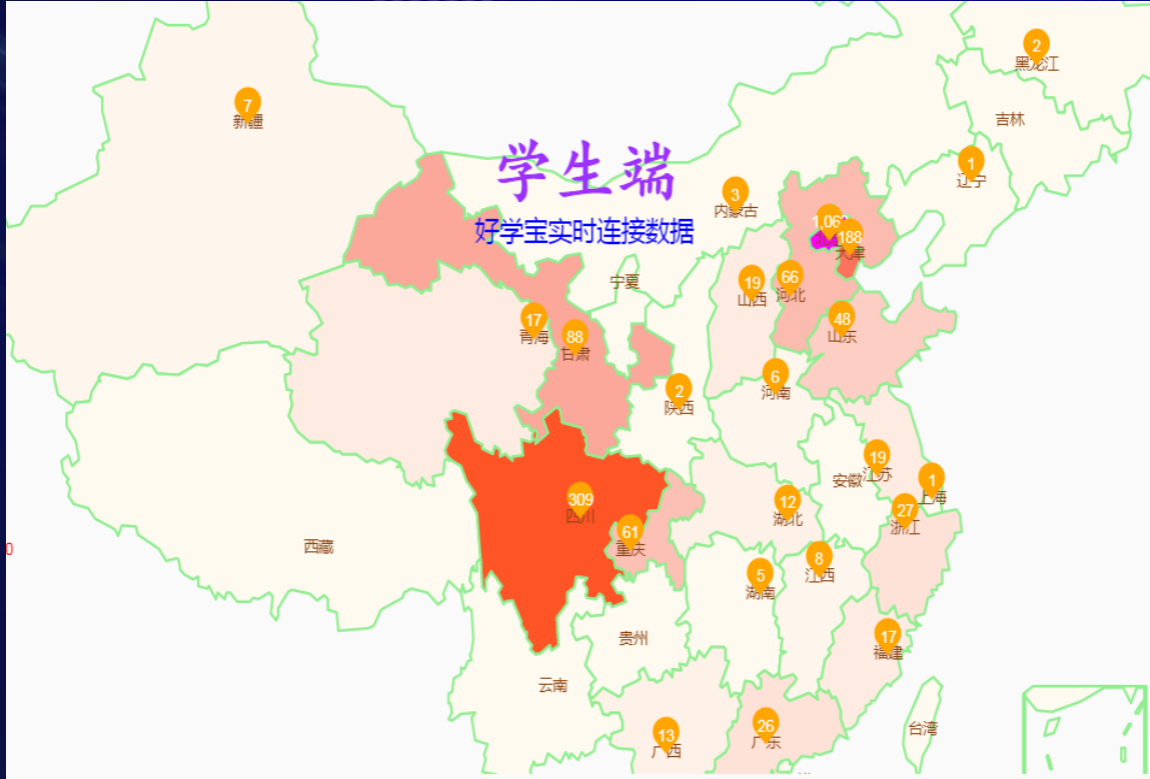
- 高效运营建立在精细化的数据统计分析上，运营效率的提升必须有相应系统支撑
- KPI体系建立（管理层视角）
- 用户转化漏斗模型（业务流程，从结果到过程的拆解）
- 运营数据统计分析（中层视角）
- 业务从粗放式运营到精细化运营的转变
- 通过数据发现市场、销售、服务三个环节的问题

## 转化率案例——一点都不简单

- 付费转化率 = 当月付费用户/当月注册用户 对不对？
- 互联网产品的留存率如何计算的？
- 转化周期，数据统计分析结果和业务实际结果比较，转化的时效是否及时。
- 注册/体验转化率，一个指标拆解为四步：电话接通率、预约体验课、体验课出勤率、体验课成功率。
- 数据指标的质量：按照渠道统计、按照地区统计



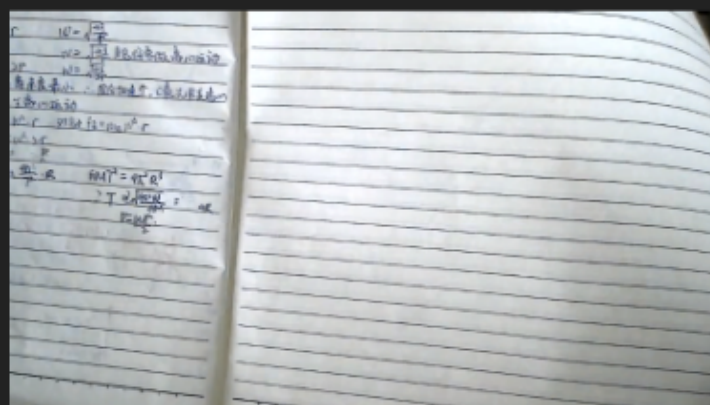
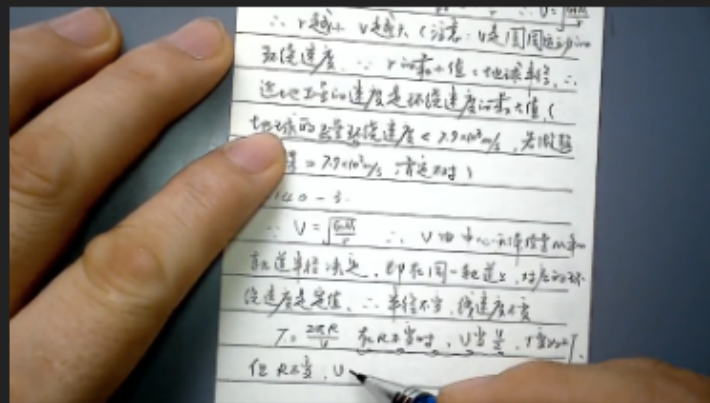
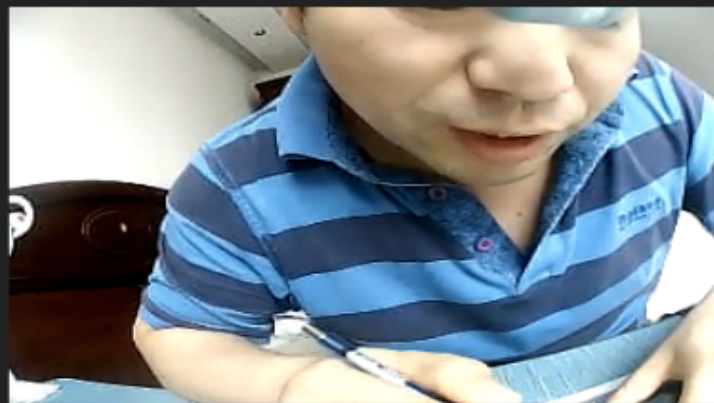
# 全国上课服务实时监控





# 单节课实时监控

三好网课堂监控中心 另一窗口打开



# 上百万小时的音视频数据

- 上课一对一音视频数据  
约200万小时音视频数据
- 学情报告音频数据  
110万份学情报告音频数据
- 咨询电话录音数据  
109万小时电话录音数据



# AI在教学场景中的应用

## 语音识别

师生  
互动

上课  
知识点挖  
掘

电话  
销售质  
检

## 图像识别

教师  
上课  
状态

上课  
背景  
替换

手势  
识别

教师  
美颜



# 大数据架构

数据采集平台	数据服务平台	数据挖掘平台	深度学习平台	数据分析可视化平台
教师数据	推荐搜索	学生模型	语音识别	数据展现
学生行为数据	内容推荐	教师模型	图像识别	数据分析
学生数据	试题推荐	知识模型	情感分析	业务预测
音视频数据	模拟考试			

Flume+ Ptail	搜索(ES)	处理平台(Spark)	深度学习 Tensorflow
--------------	--------	-------------	--------------------

存储平台(Hadoop/HDFS,Hbase,MySQL)

# 总结

- 绝大多数创业公司，业务先行，积累数据，越往后，大数据的价值越大
- 推荐引擎、数据平台、AI，大数据平台发展的三个阶段
- 大数据，必须通过技术手段，挖掘数据的价值，要么提升产品体验，要么为运营决策提供弹药。

The background is a deep blue gradient. In the upper portion, a world map is formed by white binary code (0s and 1s). Several bright, multi-colored light streaks (blue, green, and red) sweep across the lower half of the image, creating a sense of motion and digital energy.

**Thanks !**