AI筑巢:机器学习在 百度凤巢的深度应用

刘斌新

百度 商务搜索智能交互部负责人

2017.7





CNUTCon 2©17 全球运维技术大会

上海·光大会展中心大酒店 | 2017.9.10-11

智能时代的新运维

大数据运维

DevOps

安全

SRE

Kubernetes

Serverless

游戏运维

AlOps

智能化运维

基础架构

监控

互联网金融









斯达克学院(StuQ), 极客邦旗下实践驱动的IT教育平台。通过线下和线上多种形式的综合学习解决方案 ,帮助IT从业者和研发团队提升技能水平。





















10大职业技术领域课程 http://www.stuq.org

SPEAKER INTRODUCE

刘斌新

- 现负责凤巢核心策略、产品团队,包括CTR预估、动态内容、富媒体广告等方向。
- 2010年正式加入百度,作为负责人先后完成移动凤巢、应用商店变现、联盟流量、信息流变现等关键业务的攻坚;
- 带领团队连续4年6个项目入围最高奖TOP 10 并两获最高奖(百万美金大奖)。

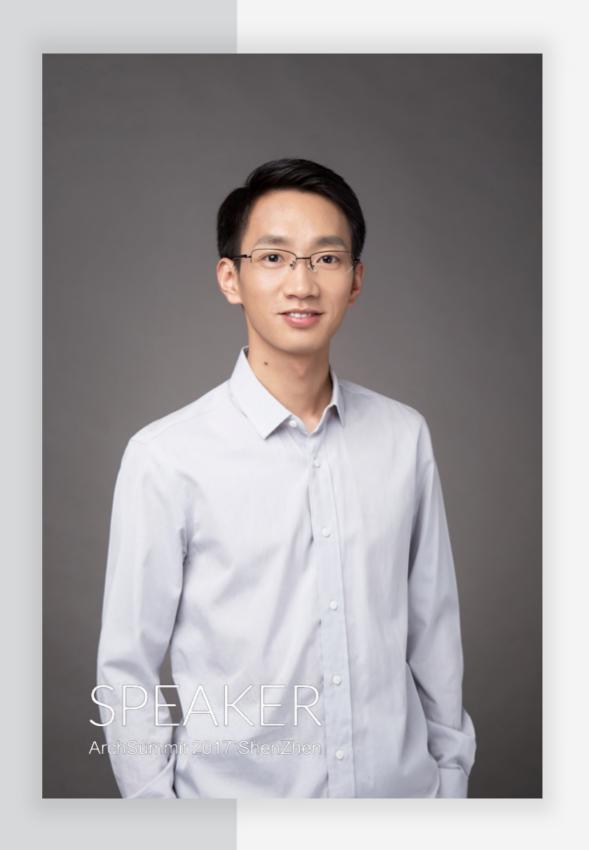


TABLE OF

CONTENTS 大纲

- 搜索广告业务简介
- CTR预估的前世今生
- 生成式广告触发
- 机器学习在内容生态的创新
- 未来的探索



搜索广告业务简介





关键词购买

按点击计费

GSP二价计费

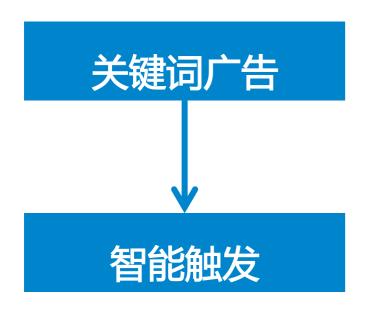


搜索广告业务简介

技术围绕业务场景

CTR预估、触发、文案生成、样式选择、弹性计算、风控....... 针对业务优化点,机器学习在凤巢几乎无处不在

例如:



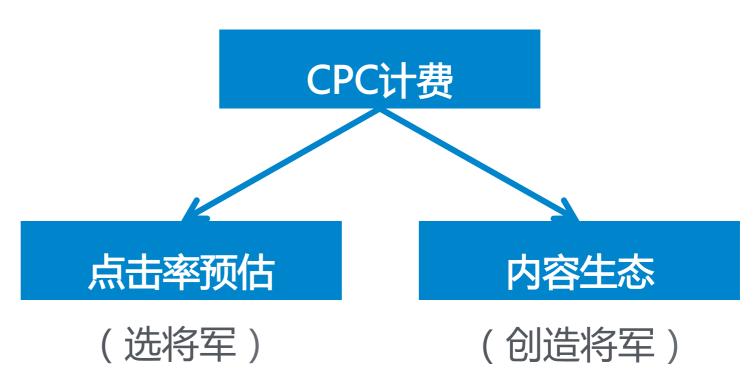






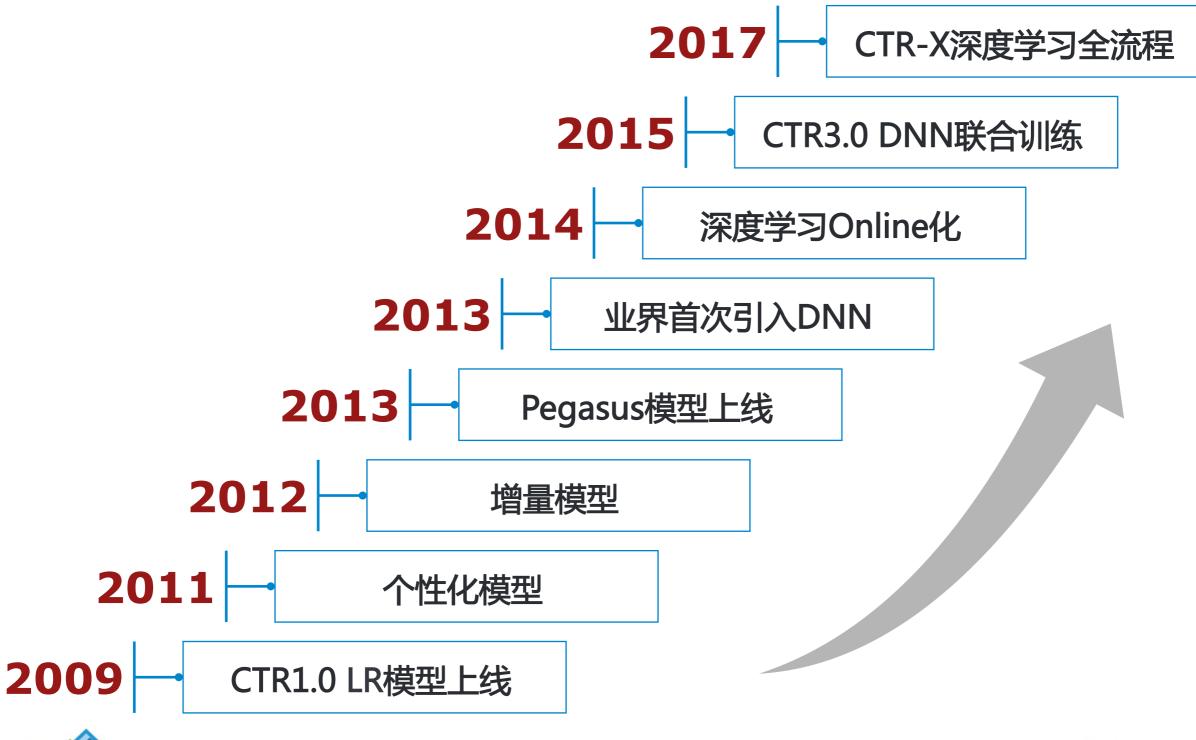
TABLE OF

CONTENTS 大纲

- 搜索广告业务简介
- CTR预估的前世今生
- 生成式广告触发
- 机器学习在内容生态的创新
- 未来的探索



CTR预估发展回顾(2009~2017)





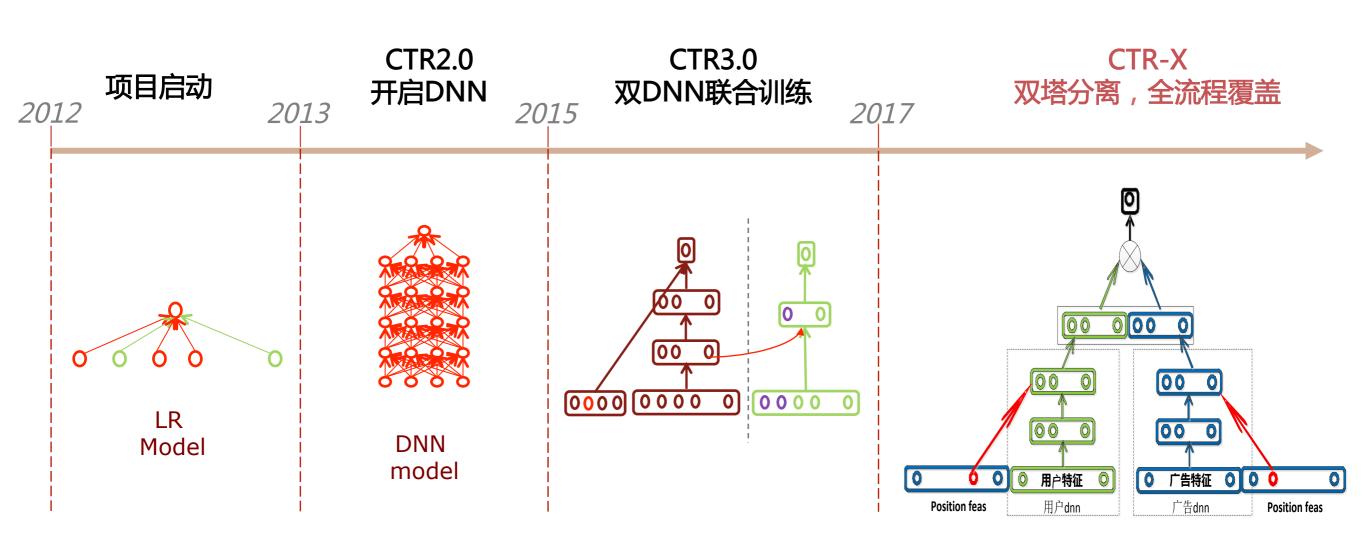
>5000

experiments





深度学习引入CTR进展回顾



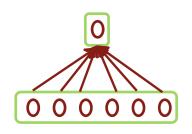


CTR-X:深度学习覆盖基础排序全流程

- 广告初筛阶段,计算量巨大,使用DNN有挑战。
- > 双塔分离DNN模型架构,优化计算逻辑和模型泛化能力,提升变现能力。

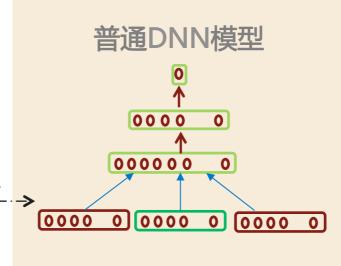
Before: LR模型

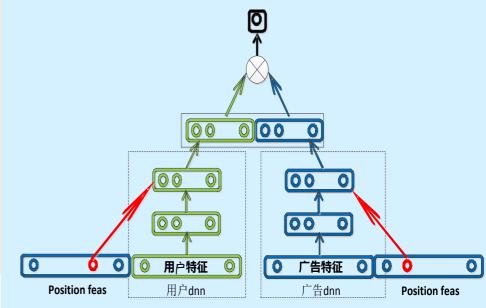
After:双塔分离DNN模型





长耗时/资源损耗,无法由LR直接升 级到DNN





特征抽取:用户侧、AD侧分别抽取

 \bigcirc

用户&广告两个维度的模型分开训练

由LR升级为双塔大规模离散DNN

耗时 优化 分布式字典:网络查询 → 内存查询

计算方面: 10万级别的float相乘 → 32个float相乘



自有大规模机器学习平台 Abacus | 超大规模

复杂性

可支持千亿级别特征和万亿模型参数

- ✓ 允许多机多线程异步实时访问和更新
- ✓ 用于支持大规模离散模型(如离散DNN)的异步sgd训练

扩展性

自主的Anti Over-Fitting机制,可随意使用扩展

- ✓ 应用于凤巢CTR模型
- ✓ 应用于手百Feed推荐模型



TABLE OF

CONTENTS 大纲

- 搜索广告业务简介
- CTR预估的前世今生
- 生成式广告触发
- 机器学习在内容生态的创新
- 未来的探索



Keyword Targeting—传统匹配模式

广泛匹配: 福克斯改造 匹配: 福特福克斯专卖店, 福克斯论坛

茶取流量从多到少人

短语匹配-核心包含:福克斯改造 匹配:福特福克斯改造,白色福克斯改装

短语匹配-同义包含:福克斯改造 匹配:福克斯改装,改造福克斯

短语匹配-精确包含:深圳福特福克斯改造 匹配:福克斯改造, 福特福克斯改造

插入、颠倒、 同义、省略 插入、颠倒、 同义

字符串包含

精确匹配: 福克斯改造

字面精确:福克斯改造

地域精确:深圳福克斯改造

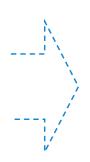
高级精确:福克斯改装





从传统触发模式到生成式触发,构建高度智能化的触发引擎

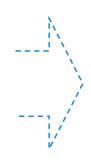














用户Query

数据检索

判别式触发

采用深度学习增强校验

支持多粒度检索

大幅提升召回率

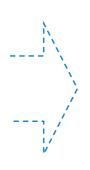
传统触发系统

检索和校验两步式架构 召回率低

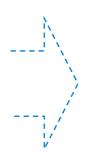


从传统触发模式到生成式触发,构建高度智能化的触发引擎











土八八代

生成式触发

充分利用深度学习能力 突破性地将检索和校验合二为一 显著提升用户体验

判别式触发

采用深度学习增强校验 支持多粒度检索 大幅提升召回率

传统触发系统

检索和校验两步式架构 召回率低





深刻理解用户意图,提升用户体验

Before

After

Bai A 百度 网页 地图

精油什么牌子的好

精油那个牌子的好?奇 一招解决秀发难题

欧莱雅奇焕润发精油,只需 蜕变, 百搭使用, 令秀发养 全网热销单品,解决各..

精油那个牌子的好?奇焕护发精油 一招解决秀发难题

欧莱雅奇焕润发精油,只需1瓶见证秀发新生 蜕变, 百搭使用, 令秀发养+护+亮+柔+顺。 全网热销单品,解决各..

巴黎欧莱雅 Va 119条评价 商业推广

巴黎欧莱雅 Vi 119条评价 高业推广

什么牌子精油最好_百度经验

- 1. 看价格 真正的精油价格不会太便宜普通精 油如香橙、茶树油大概10m/400-1300元,
- 2. 闻味道 多数纯精油味道柔和醇正, 因为是 100%从植物中提取的, 所以纯正的精油...
- 3. 看用法 绝大多数纯精油是不可直接用于皮 肤的, 否则会造成灼伤、脱皮, 除了玫瑰...

百度经验

查看全部内容

精油什么牌子的好? 百度知道

所调榜单、所调名牌、所调名人推荐,其实都 根本不是鉴定精油品质的关键。 迷信广告吹... 结果局限于字面 优质结果少



精油那个牌子的好? 一招解决秀发难题! 欧莱雅奇焕润发精油, 只

蜕变, 百搭使用, 令秀发 全网热销单品、解决各... 巴黎欧莱雅 Vi 119条评例

中国十大精油品牌,真实口碑_效果 测评 申领试用

兰蔻黄金油-全效精油, 蕴含8大臻稀植物精 华,深透肌...

十大精油品牌 什么精油好

rosebeauty.com.cn V3 114条评价 商业推广

中国十大精油品牌,兵大口呼一次木 测评 申领试用

兰鼓黄金油-全效精油,蕴含8大臻稀植物精 华,深透肌...

十大精油品牌 什么精油好

rosebeauty.com.cn vi 114条评价 形型学厂

什么牌子精油最好 百度经验

- 1. 看价格 真正的精油价格不会太便宜普通精 油如香橙、茶树油大概10m/400-1300元...
- 2. 闻味道 多数纯精油味道柔和醇正, 因为是 100%从植物中提取的 所以纯正的精油

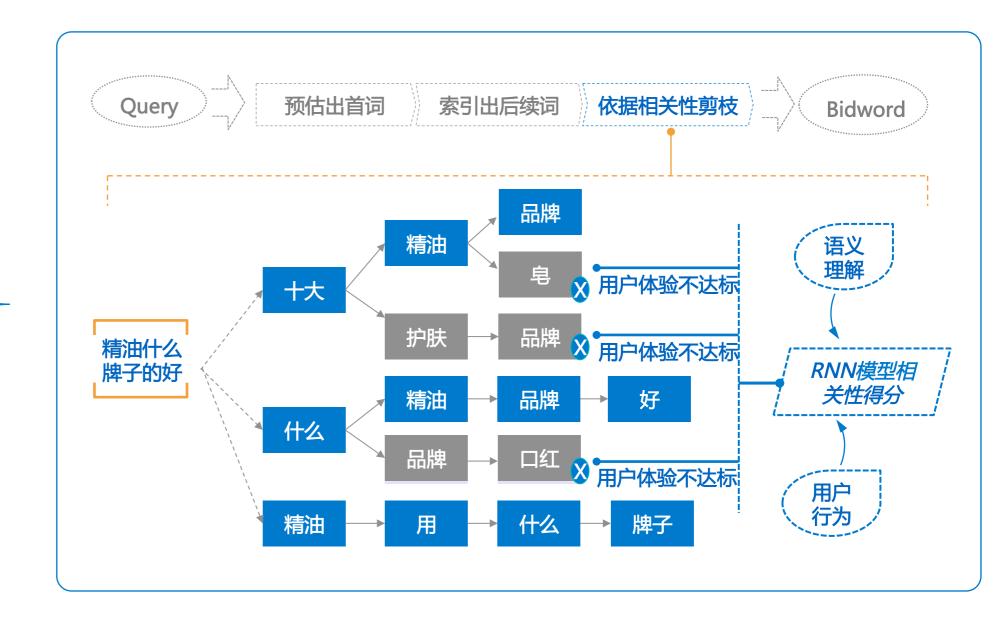
具备AI "思考" 能力 深度理解用户意图 呈现更多优质结果

Geekbang>. InfoQ



生成式触发模型, 以AI技术深度理解用户需求









生成式触发模型, 以AI技术深度理解用户需求

编码器: Bi-RNN

注意力模型

解码器: RNN

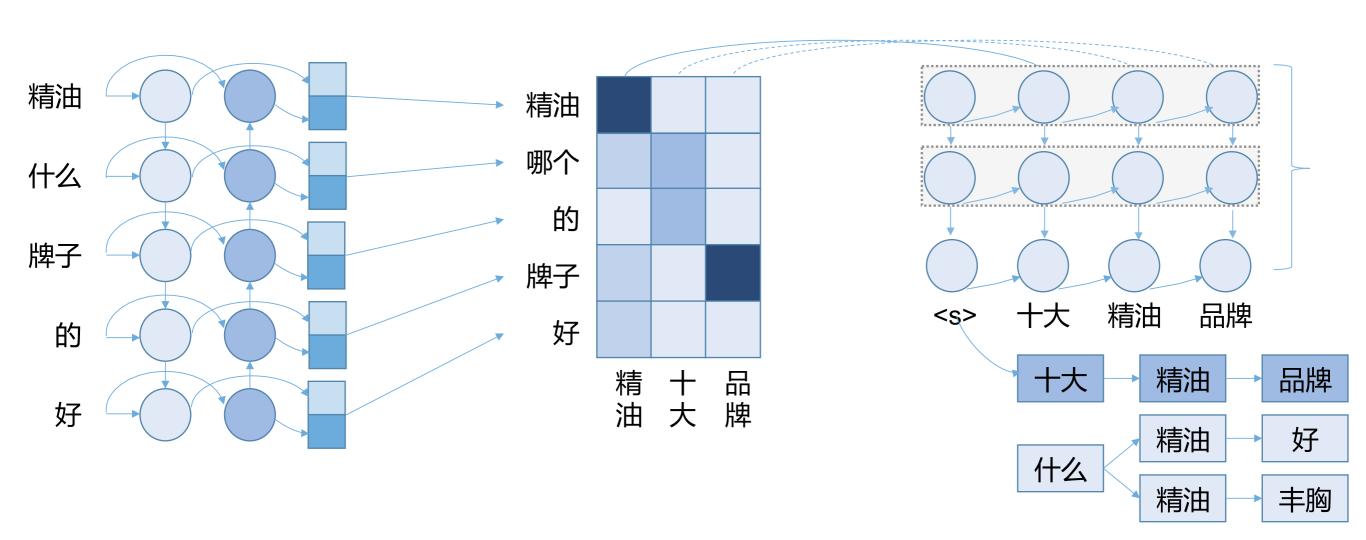






TABLE OF

CONTENTS 大纲

- 搜索广告业务简介
- CTR预估的前世今生
- 生成式广告触发
- 机器学习在内容生态的创新
- 未来的探索



自动生成、校验商业文案,提升可信度

去八达岭长城 只需59元起 不满意退款

去八达岭长城-北京国际旅行总社官方网,市区免费接 送, 给您完美旅行体验, 多款京城游玩线路, 供您选 择,价格优惠,欢迎咨询:010-59422266 m.bjglwj.com Vi 评价 商业推广

去八达岭长城 仅售100元 无购物 多景点

去八达岭长城-北京国际旅行总社官方网,市区免费接

松你中羊花仁体心 多数古地游玩线路, 供您选

H

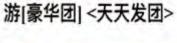
咨询

长城一日游[豪华团]

2266

仅售100元!享受中国国旅提供价

值200元 无购物: 无自费长城一日





微信

落地页

Query + Title + 创意 + 落地页.... Input IN X1' X2' ··· Χ'n X1 X2 X3 Encoder 双向 **RNN** context Decoder Y3., OUT 高牌校验 品牌校验 验 业务校验 Quality **Control** All Yes Output 商业文案 Sentences

±ma Geekbang>. infoQ



商业文案从人工撰写迈向AI自动生成

Before

After:智能创意

雅思口语考试 雅思7分通过率90%

雅思口语考试签约直达,不达标免费重读雅思名师教学,致力于全程一对一个性化课时辅导,致力于打造 高分之路!

m.beijing.gedu.org V1 评价 高业推广

- 标题描述夸大不一致
- 创意空洞冗长, 无重点



雅思口语考试 环球英语培训学校

[常见问题] 雅思考试口语水平如何提高?

[推荐答案] 1.模仿磁带或广播。2.大声朗读。3.用自己的话复述。4.选择语言地道、内容新颖的教材。详情...

m.beijing.gedu.org Vi 评价 商业推广

- & 电话咨询
- AI知识和内容理解
- 结构化核心内容前置
- 知识直达最相关页面

如何提高自己的口语水平?

2016-07-02 (SW : III III III III

如何提高自己的口语水平?看一些口语书有用吗?这是在学习英语当中很多人都会遇到的困惑。

众所周知,对话是提高口语能力最有效的方法,所以应该抓住一切机会和讲英语的人士对话。如果找不到和你进行对话的合适人选,你可以采取以下几种方法:

第一、模仿。提高口语能力的第一 步就是模仿磁带或广播,模仿以英语为 母语的人的语音和语调。如配有磁带的 《新概念英语》,你可以根据自己的喜 好,选择英音版或美音版的磁带加以模 仿。

第二、朗读。提高口语能力的第二 步是大声朗读。朗读也可以说是第二模 仿,所以也需要有意识地以磁带里的语









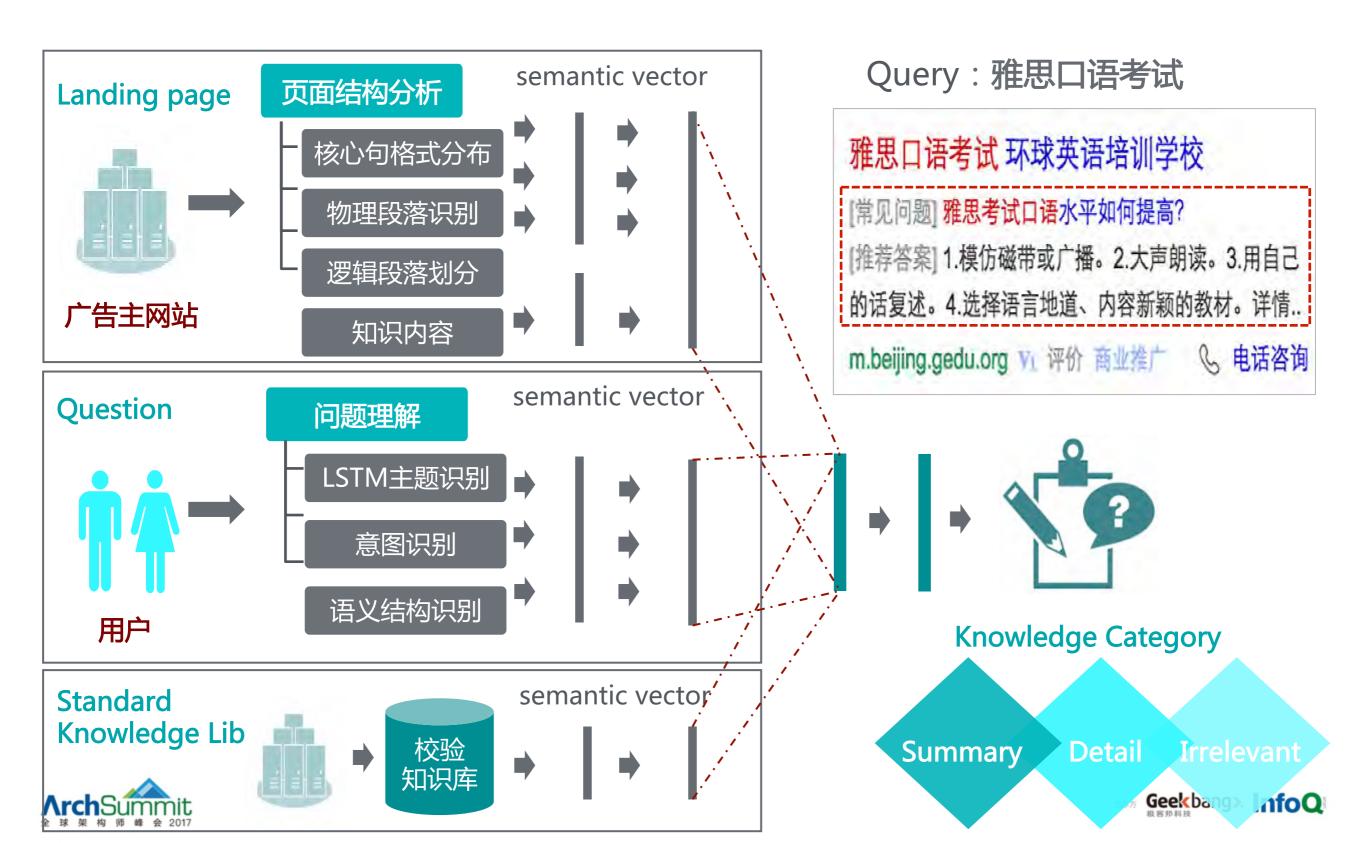


您想了解都方面问题呢?



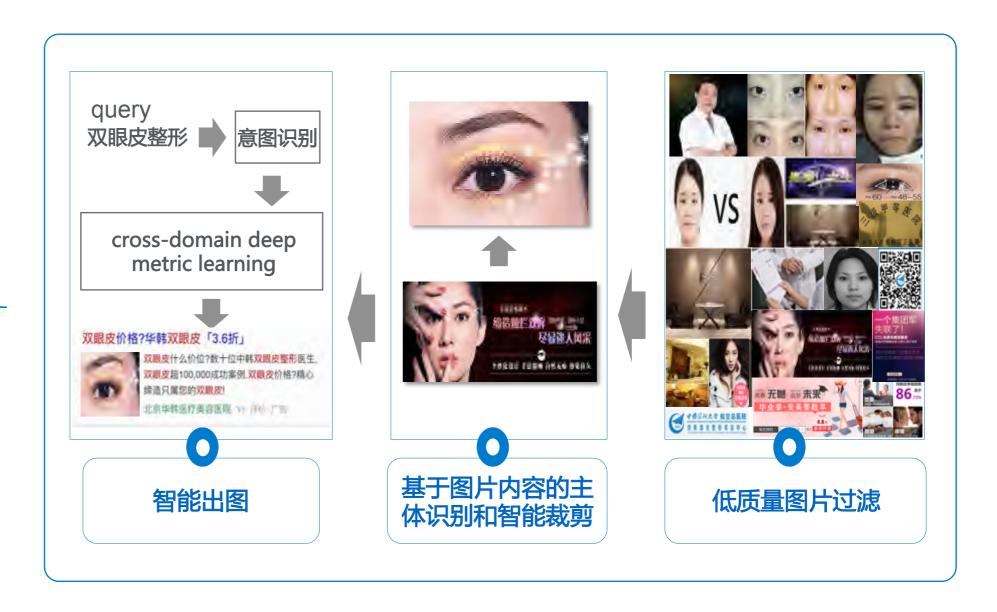


隐式知识理解与凝练,提高用户获取信息效率



基于用户意图智能出图,快速Catch用户









机器学习打造智能落地页

Before

全新园具设备,油烟净化器产品,好品牌就选东莞市 科怡环保油烟净化器

餐饮油烟净化器 油烟净化设备 全新厨...

www.keyisb.com v Fff FEAT

- 内容空洞千篇一律
- 专业化、结构化不足
- 移动用户体验质量堪忧



After:智能落地页

餐饮油烟净化器 节能耐用 超强静音

相关产品: 约11款 优势:环保|节能|可定做|质量领先 专业油烟净化器生产商、提供新研制的低空排放油烟。

www.keyisb.com vy Williams

- 重构高质量落地页
- 智能化、专业结构化
- 实现内容体验闭环







机器学习打造智能落地页

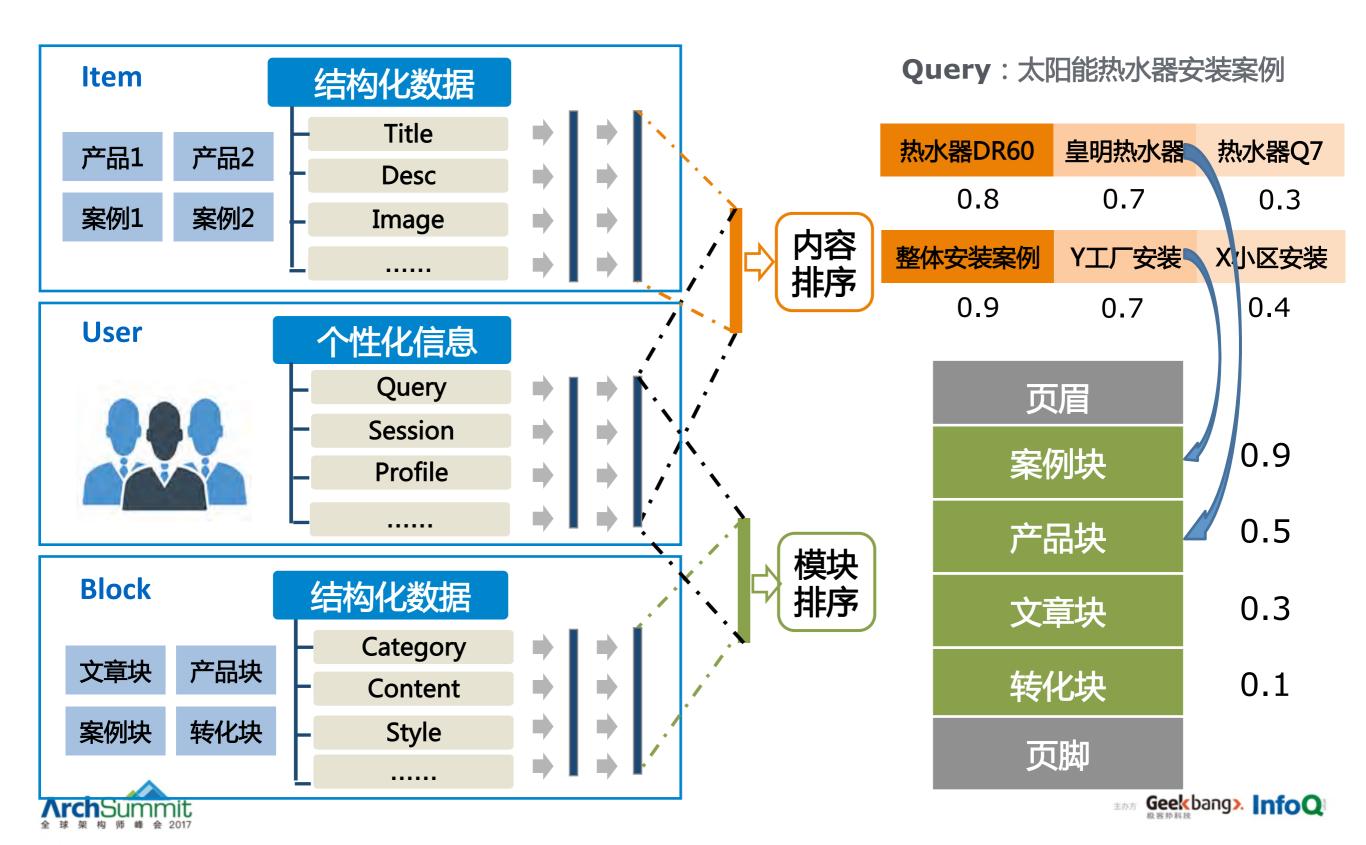


TABLE OF

CONTENTS 大纲

- 搜索广告业务简介
- CTR预估的前世今生
- 生成式广告触发
- 机器学习在内容生态的创新
- 未来的探索







以图片化信息集成快速吸引用户并满足需求

实现图片创意的"一键美颜"

2017马尔代夫旅游报价特惠 Before 旅游报价全包套餐!



2017马尔代夫旅游报价特惠技

After

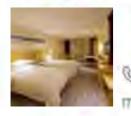
旅游报价全包套餐!



海洋国旅马尔代夫旅游报价选岛, 控房控票代理, 马尔代夫旅游报价专业...

m.ooccl.cn Vi -17 M VFI (II)

「携程网」上海酒店预定 - 2 Before 省money - 放心的价格...



「携程网」上海酒店预定 - 2

After

省money - 放心的价格..



国内酒店 客栈民宿

团购酒店 附近酒店

6 电话咨询 400-828-0067

APP下载

m ctrip.com * F =

海外酒店

实现图片创意信息的"一览无遗'

携程拉萨 订酒店 高额返现 住哪里都..

携程拉萨 订酒店保证低价,新用户注册成功可获1000积分+返现特权!!



m.ctrip.com v 168条评价 广告

上海旅游 北京出发 海量线路+报价 北...

适宜季节: 春季 景区类型: 城市 游玩时长: 3~7天 北青旅-上海旅游,海量线路,一价全含,纯玩品质团.



m.iutour.cn v 108条评价 广告



以图片化信息集成快速吸引用户并满足需求

实现图片创意的"一键美颜'

2017马尔代夫旅游报价特惠 Before 旅游报价全包套餐!



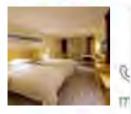
2017马尔代夫旅游报价特惠技 旅游报价全包套餐!



海洋国旅马尔代夫旅游报价选岛, 控房控票代理, 马尔代夫旅游报价专业...

m.ooccl.cn Vi -I/MVPI0 []

「携程网」上海酒店预定 - 2 Before 省money - 放心的价格...



「携程网」上海酒店预定 - 2

After

After

省money - 放心的价格..



国内酒店 客栈民宿 海外酒店 APP下载

团购酒店 附近酒店

电话咨询 400-828-0067

m ctrip.com * F =

实现图片创意信息的"一览无遗'











加强端到端的富媒体化程度...

Before





After







营销信息提炼成短视频呈现,让用户一秒读懂









视频合成

订婚宴 就选花嫁丽舍



Video-Out视频广告:视频处理、图像理解、检测技术



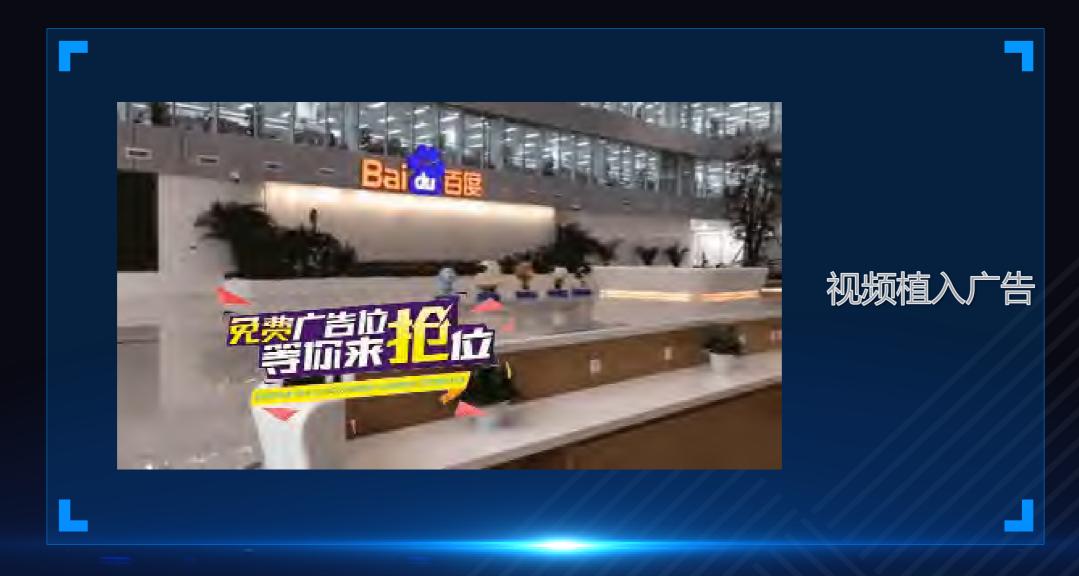


Video-In视频广告:图像前景背景检测与分割、AR虚拟植入





Video-In视频广告:图像前景背景检测与分割、AR虚拟植入





新媒体的产生渴求新技术, 在AI时代这波浪潮中,欢迎各位大咖加入

欢迎联系:fc-ai@baidu.com





THANKS!





让创新技术推动社会进步

HELP TO BUILD A BETTER SOCIETY WITH INNOVATIVE TECHNOLOGIES

Geekbang》。 极客邦技



专注中高端技术人员的技术媒体



高端技术人员学习型社交平台



实践驱动的IT教育平台





