



第九届中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2018

O2O搜索的算法实践

仙云森@美团点评

DTCC
2018

2018.05.10 - 12 北京国际会议中心



IT168.com

ChinaUnix

ITPUB

大纲

- O2O搜索简介
- 默认词/热搜词算法与实时化
- 查询分析整体介绍
- 智能排序的模型演进
- 全场景优化与展望



生活服务@美团

- ✓ 一站式生活服务平台
- ✓ 几十个业务线



搜索@美团

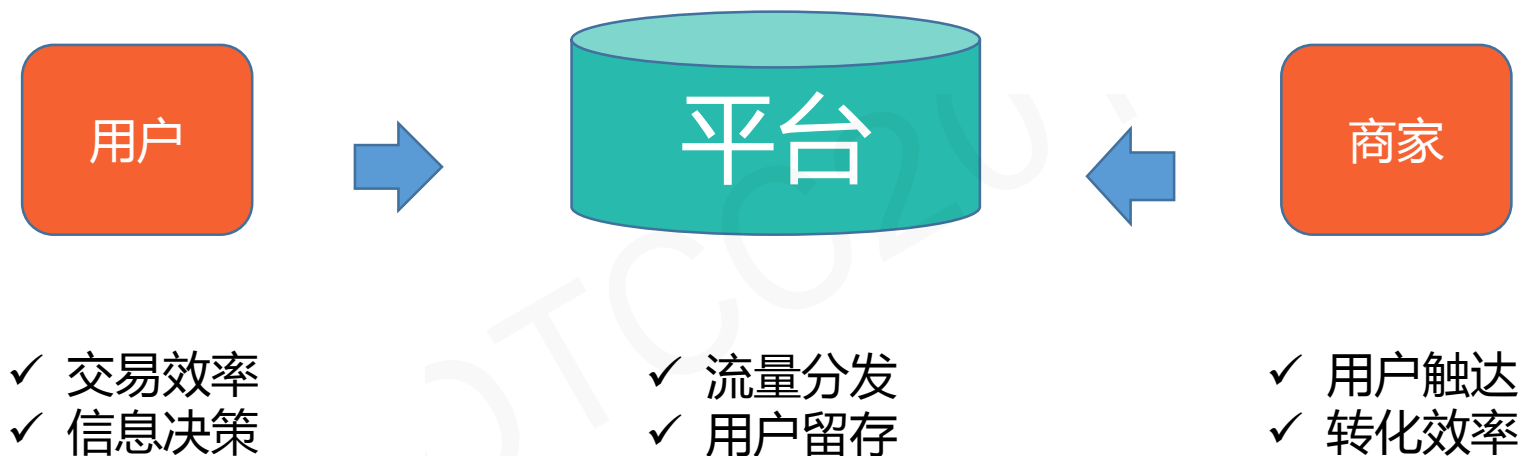


POI, 亿级SPU



搜索@美团

- Eat Better , Live Better
- 连接人与服务



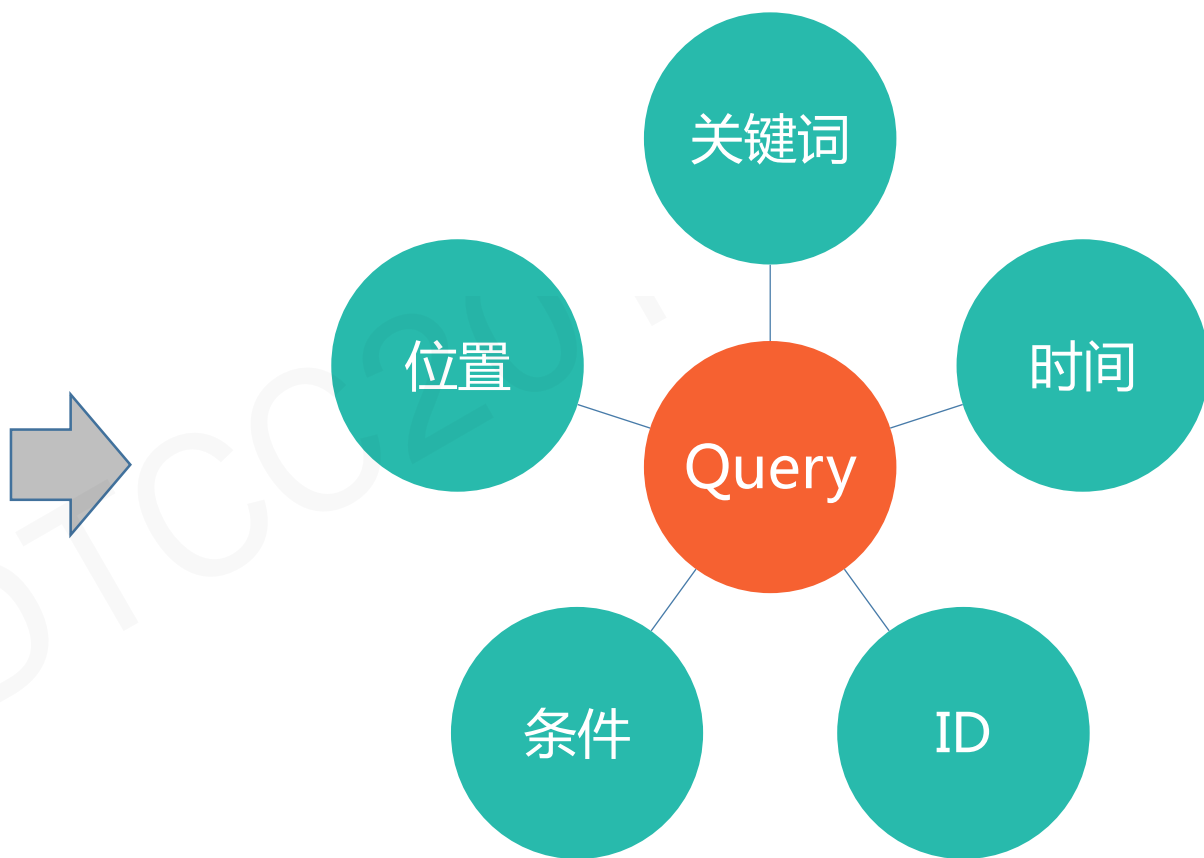
O2O搜索特点

	O2O	网页	购物电商
目标	交易	相关性	交易
移动化	高	中	高
个性化	高	低	高
场景化	高	低	低
实时化	高	低	中



O2O搜索中的Query

- ✓ 品牌
- ✓ 类目
- ✓ 地点
- ✓ 项目
- ✓ 组合词



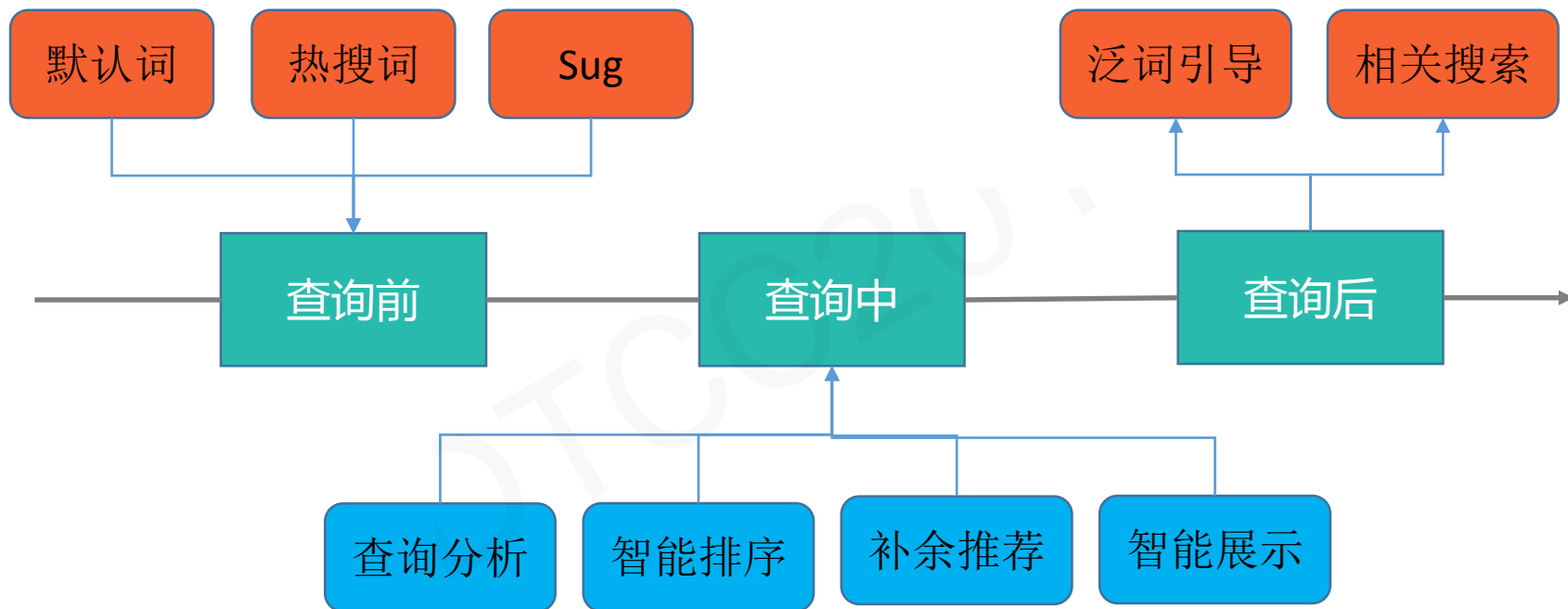
搜索技术全景图



搜索中的算法工作

降低成本，激发需求，搜索教练

细化需求，表达纠正，需求迁移

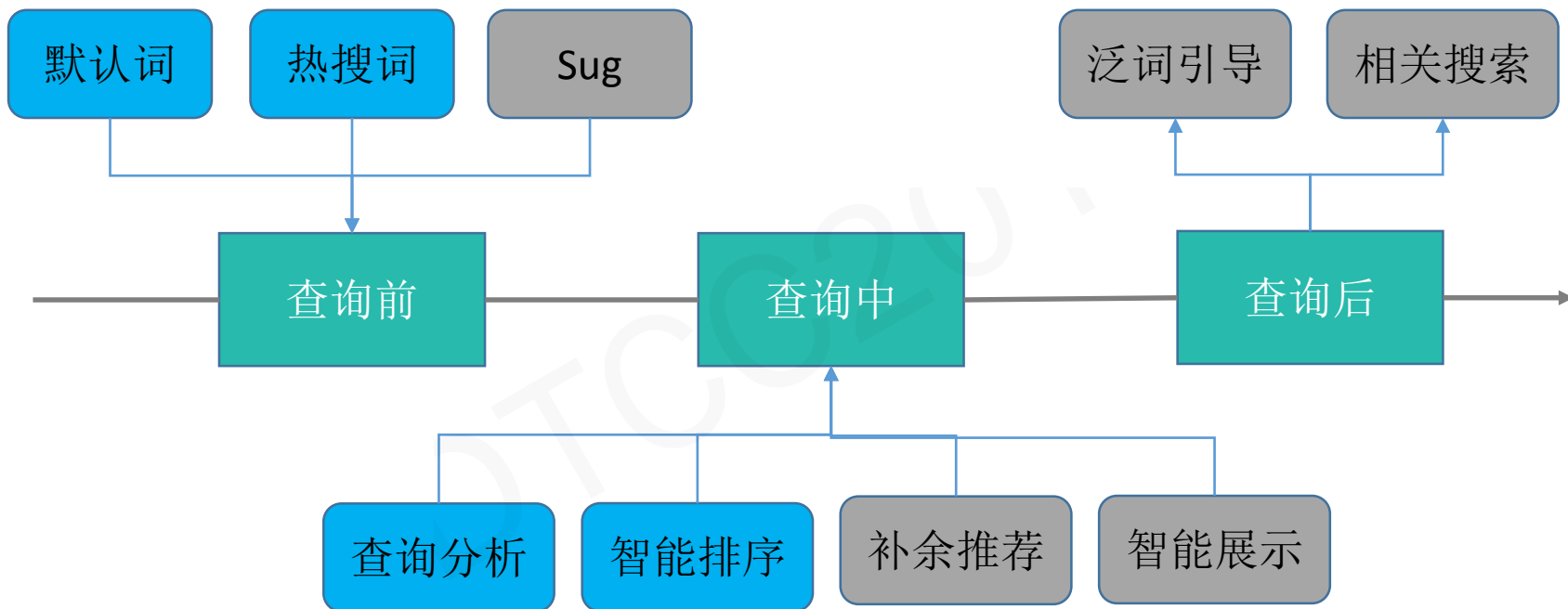


精准需求一键触达，提升决策效率

用户引导

降低成本，激发需求，搜索教练

细化需求，表达纠正，需求迁移

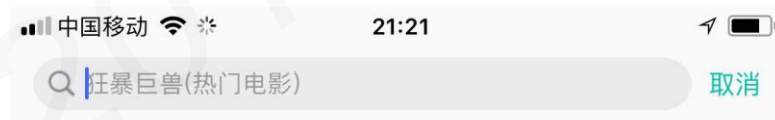


精准需求一键触达，提升决策效率

默认词/热搜词

- ✓ 产品定位：帮助用户快速发起搜索，激发用户需求
- ✓ 一套技术架构，两种排序策略

默认词



热门搜索

热门搜索

必胜客宅急送	嘉和一品粥	速8酒店	
三个渔夫灶台鱼	全时便利店	眉州东坡	金凤成祥

历史搜索

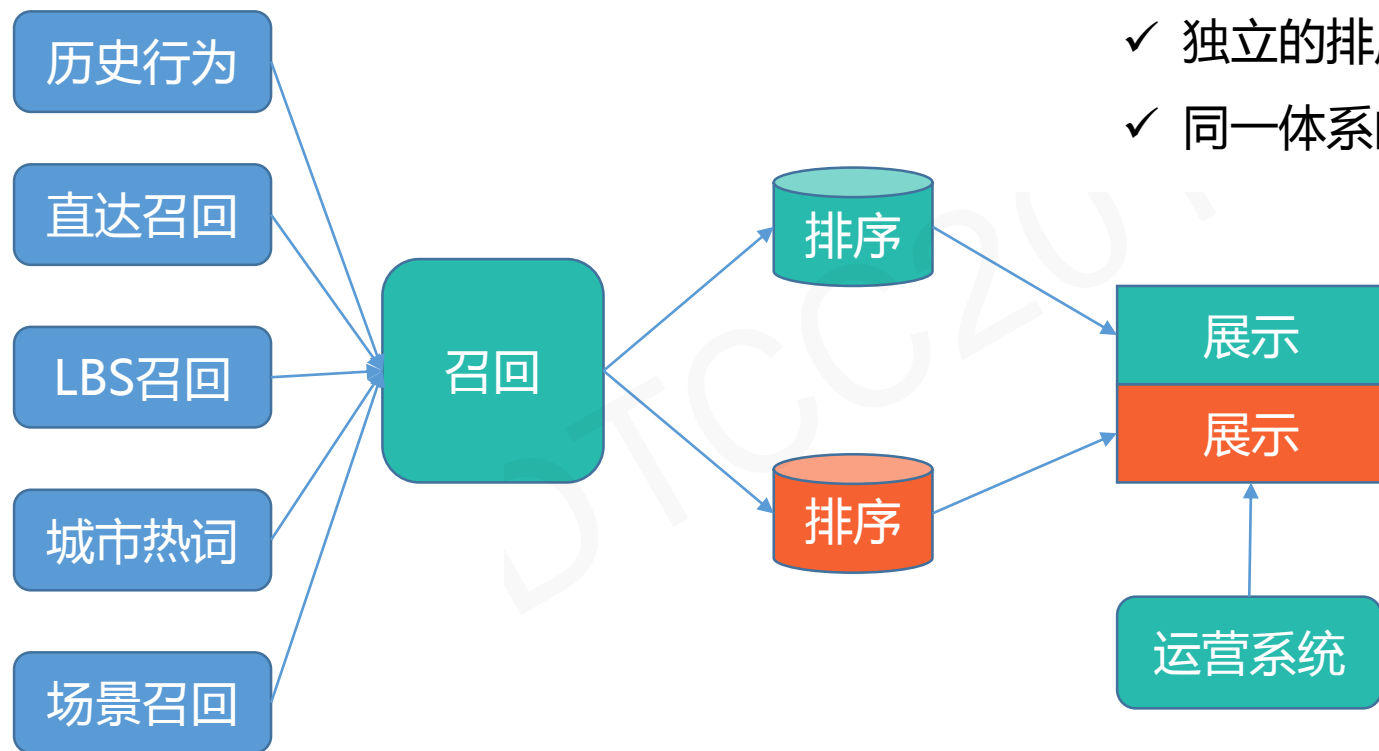
历史搜索

按摩	鲜花	火锅	spa	DIY	DIY手工坊
汗蒸	ktv	美睫	美甲		

热搜词

历史词

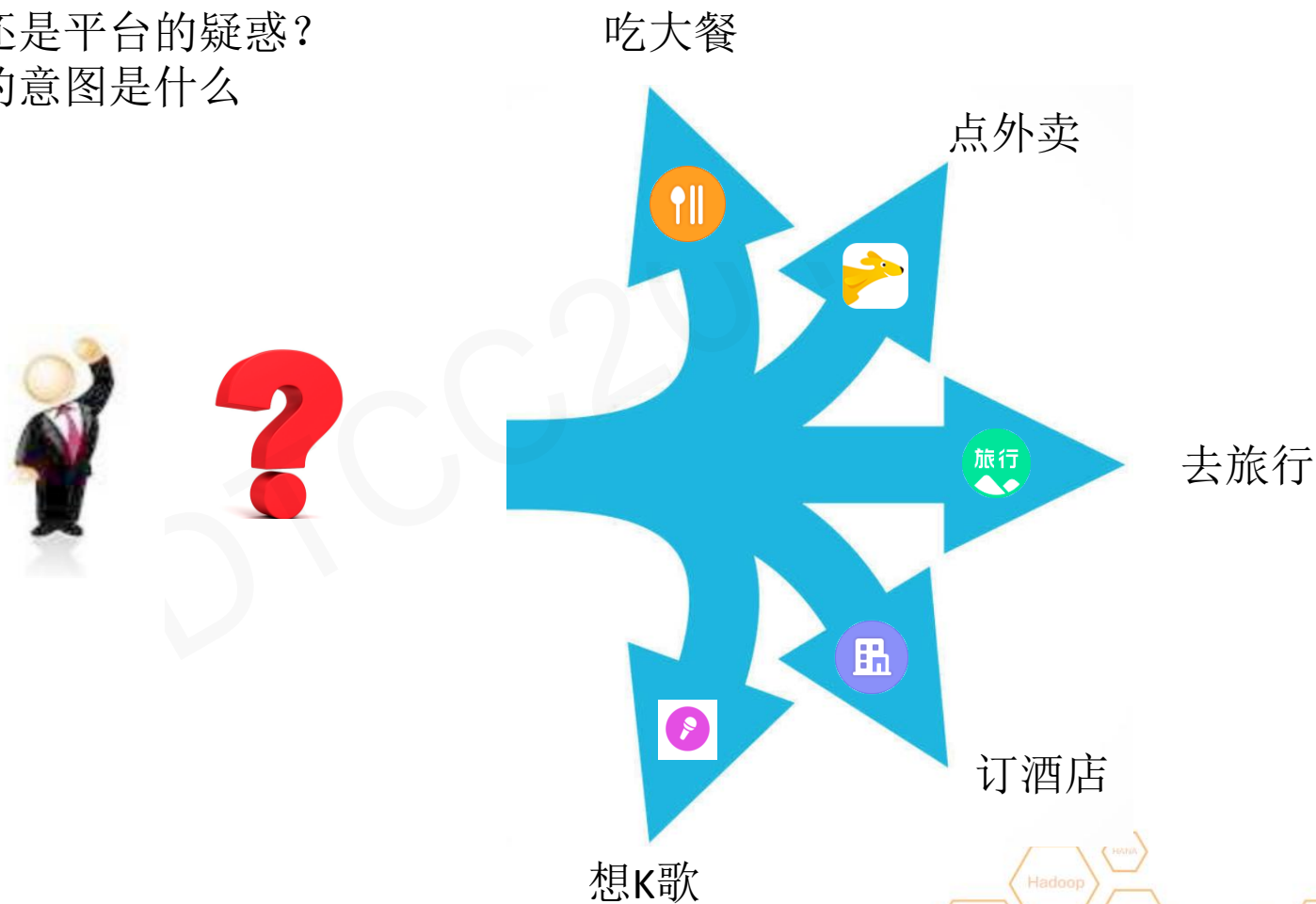
策略架构



- ✓ 统一的召回架构
- ✓ 独立的排序模型
- ✓ 同一体系的展示策略

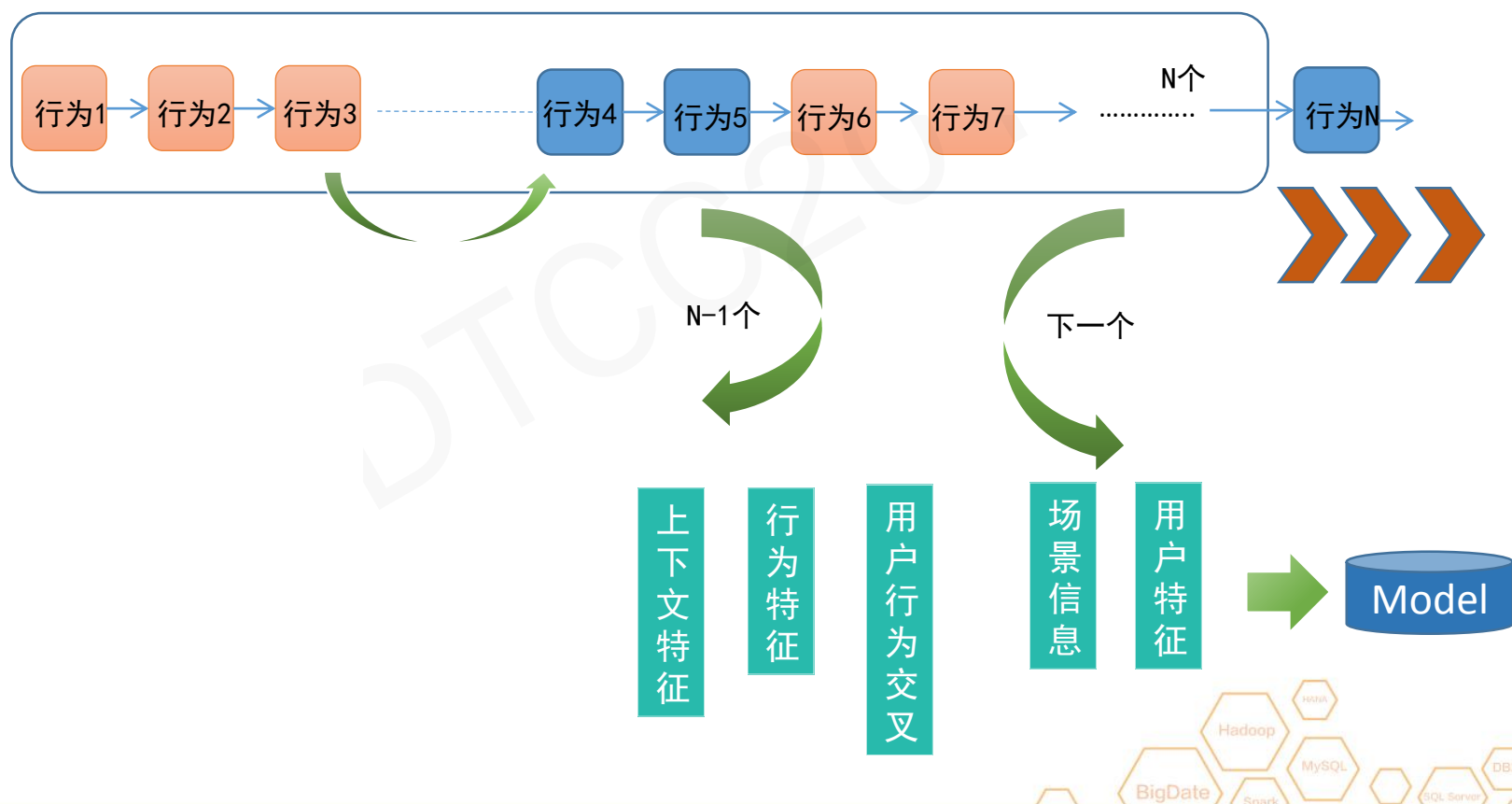
实时化-why

- ✓ 用户疑惑还是平台的疑惑?
- ✓ 用户当下的意图是什么

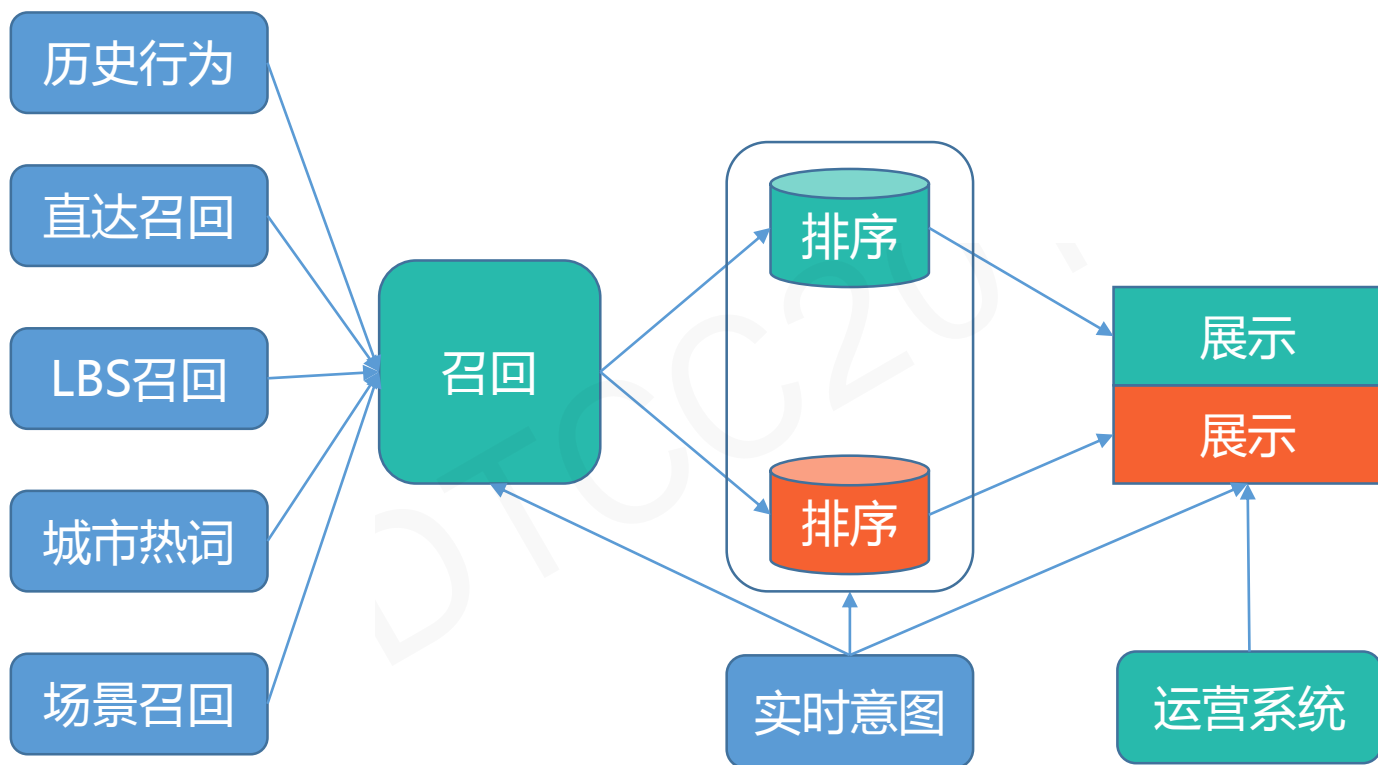


实时意图识别

- ✓ 以一级类目为意图体系
- ✓ 用过去的前N个行为预测用户下一个消费



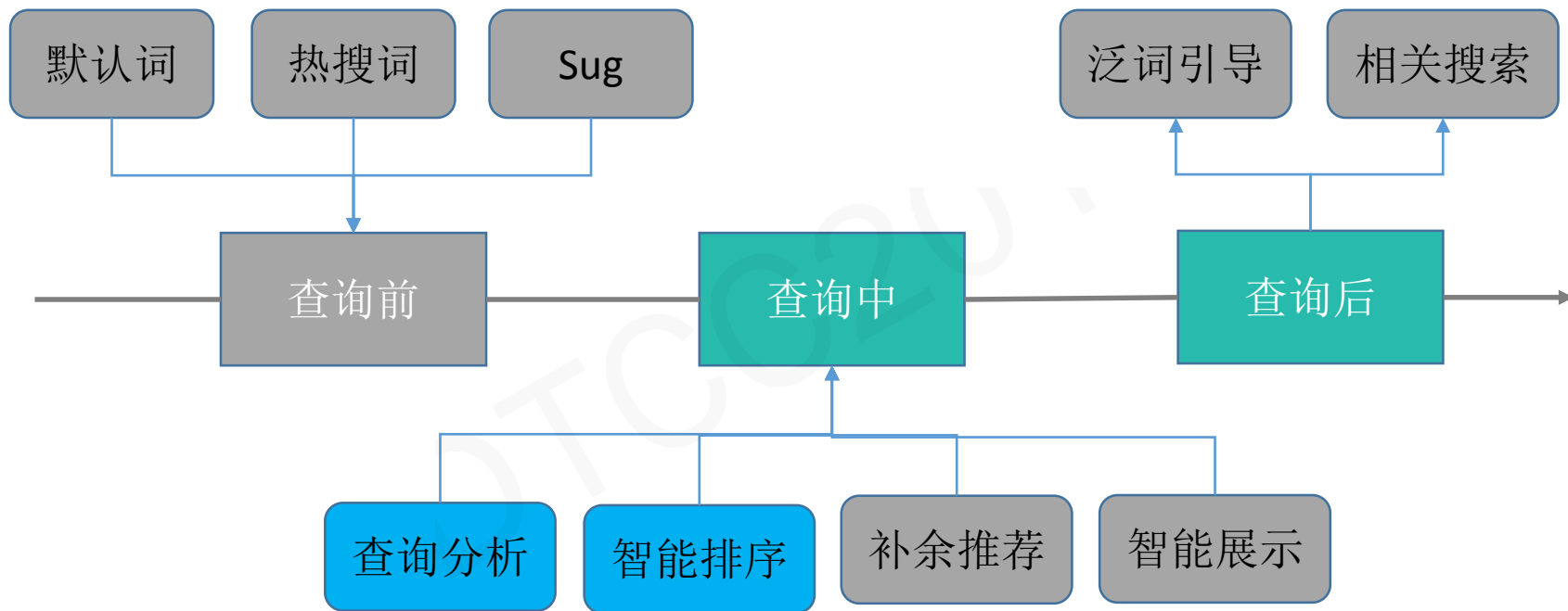
默认词/热搜词实时化



查询分析

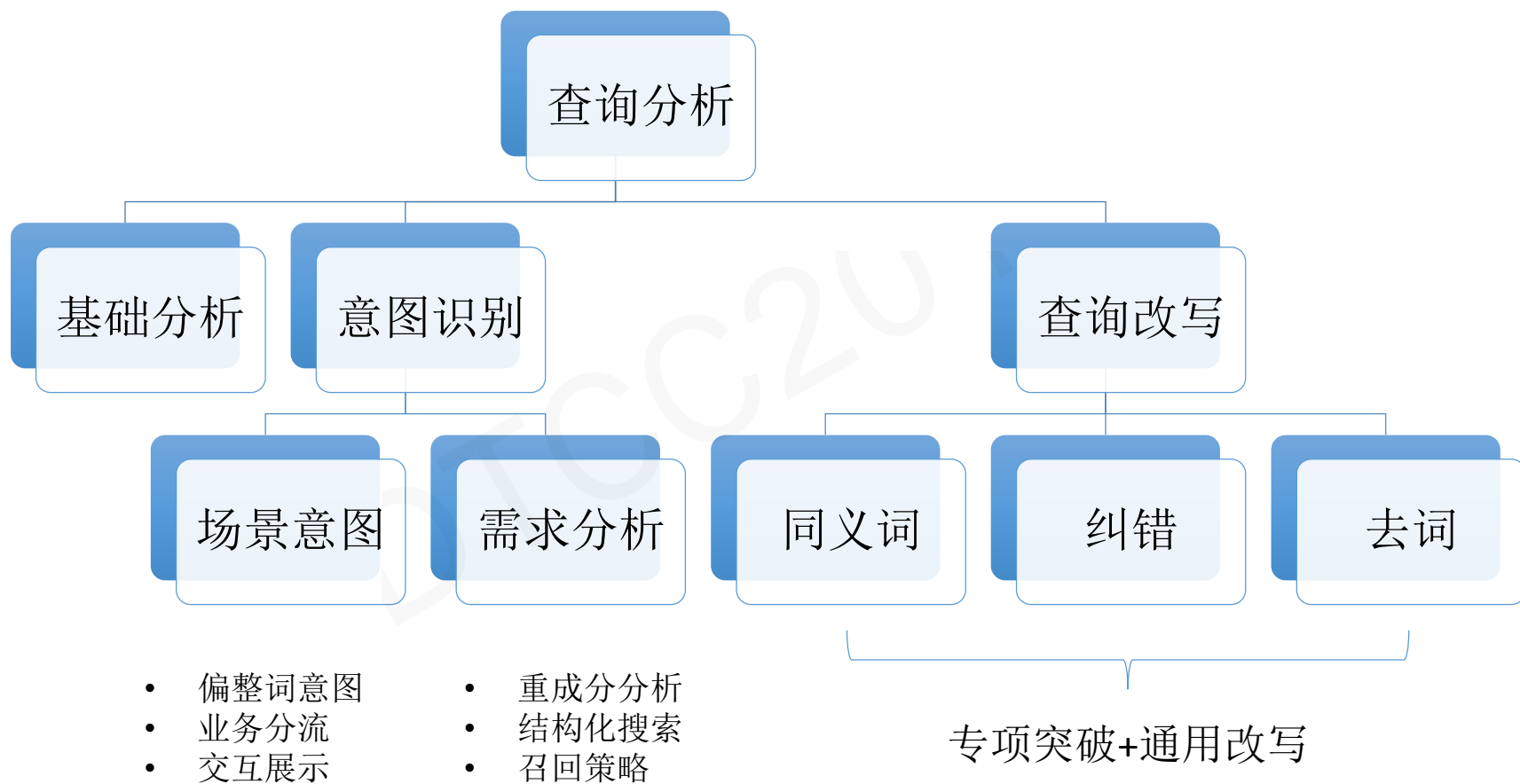
降低成本，激发需求，搜索教练

细化需求，表达纠正，需求迁移

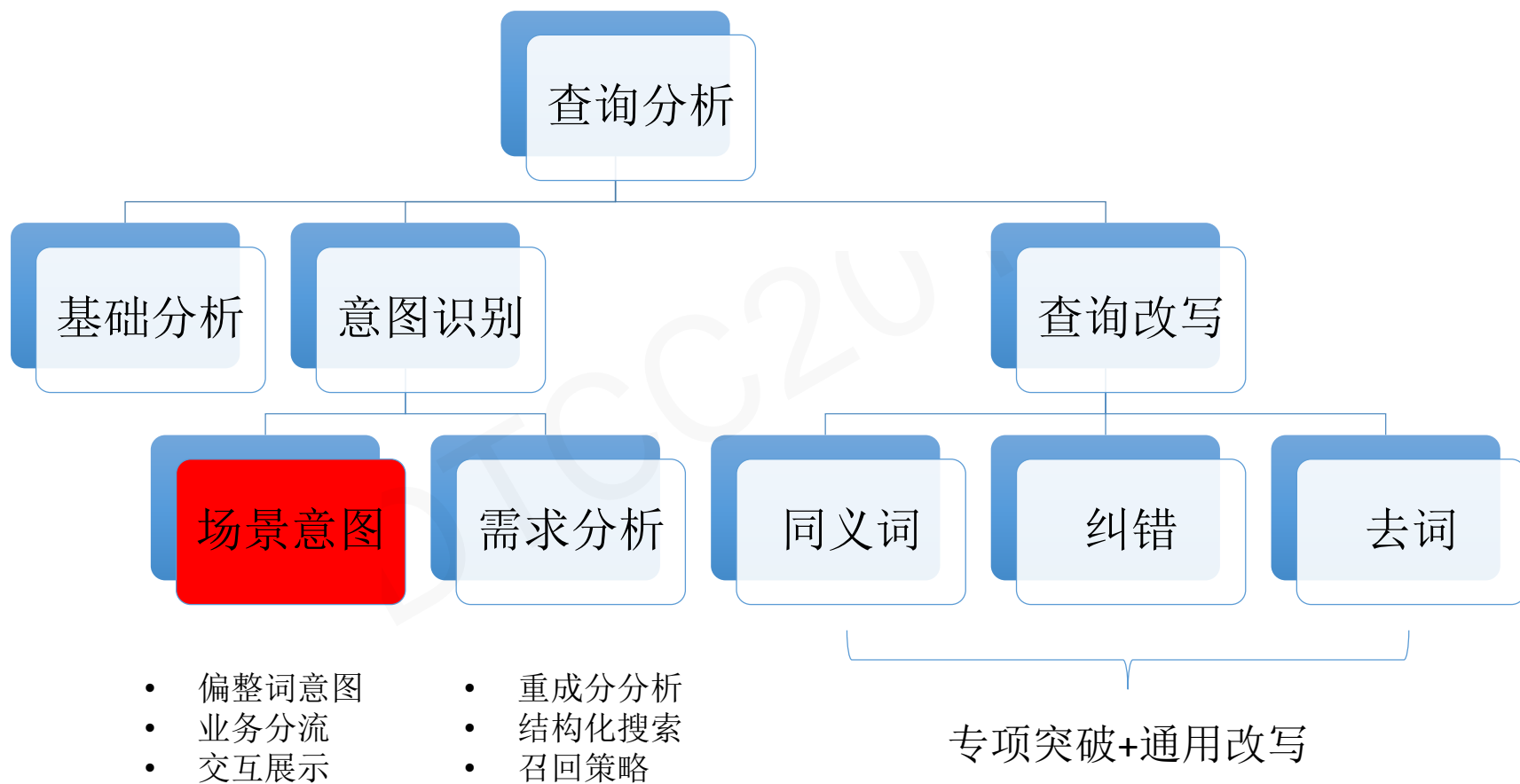


精准需求一键触达，提升决策效率

查询分析全景



查询分析全景



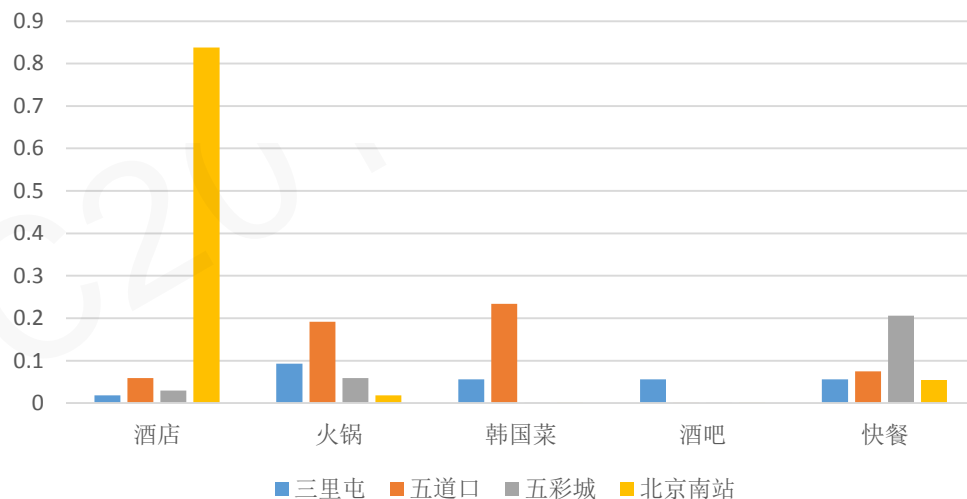
查询分析-场景意图 1/3



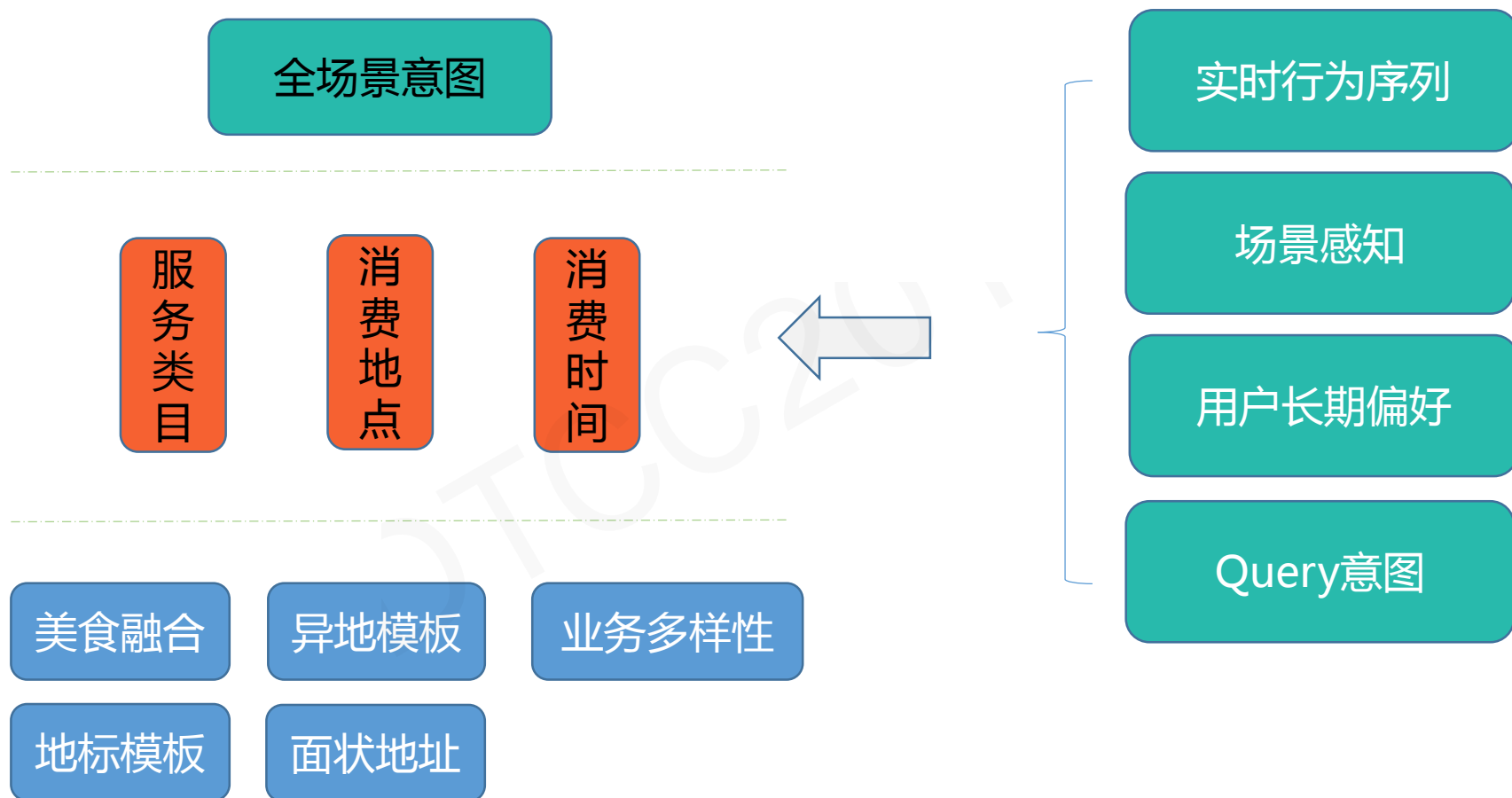
查询分析-场景意图图 2/3

- 本地特色
- 消费的及时性
- 跨类目多意图查询

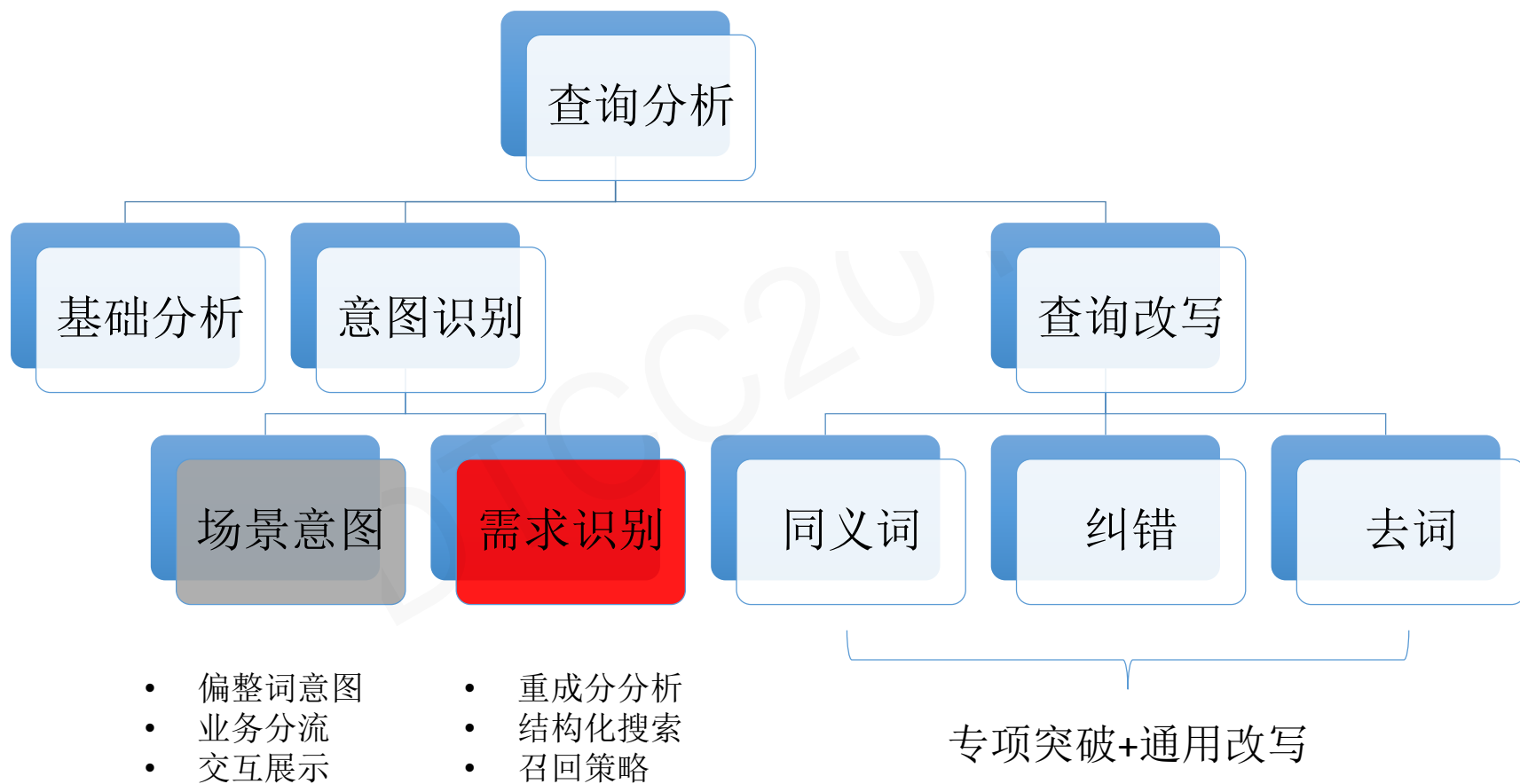
地址类查询消费分布



查询分析-场景意图 3/3



查询分析全景

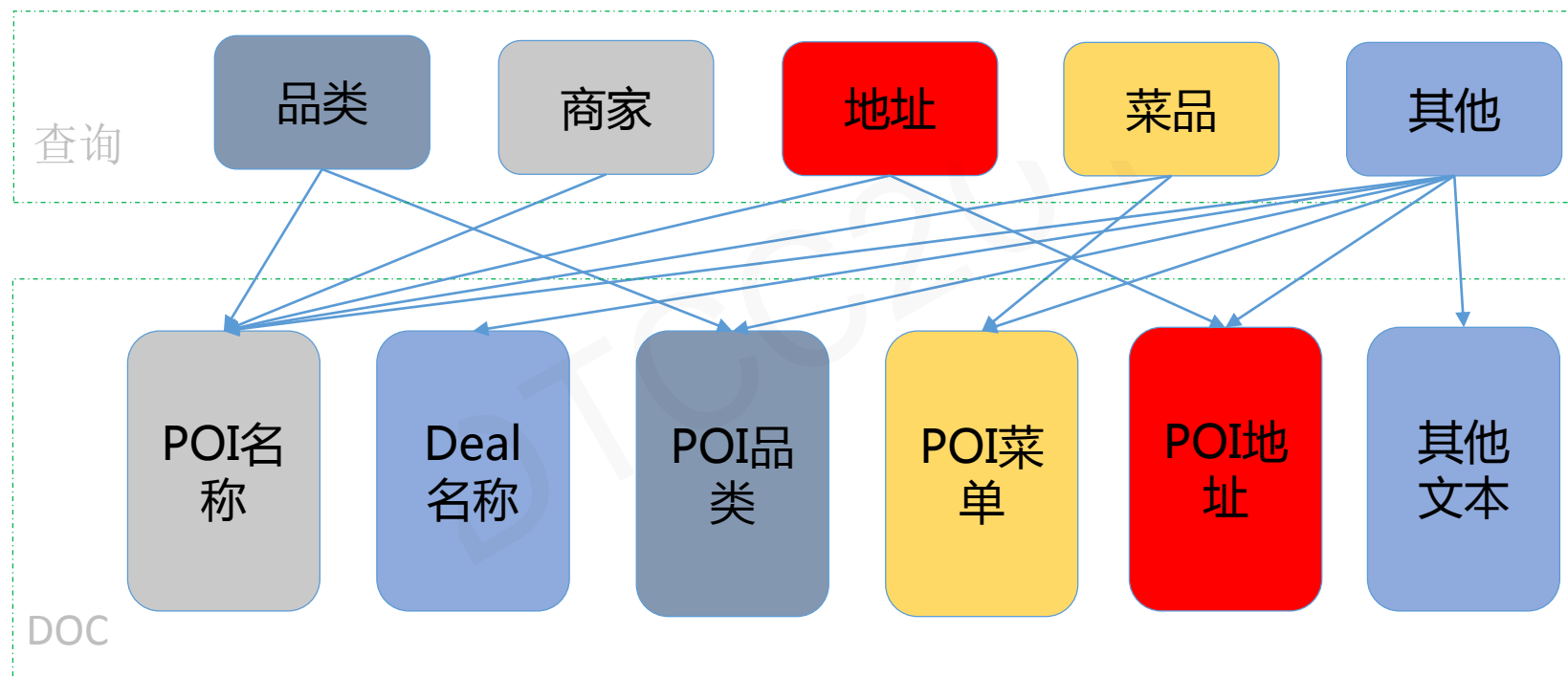


查询分析-需求识别

中关村火锅

如家上地

水煮鱼



- ✓ 知识+模型
- ✓ 实体识别

实体识别 1/3

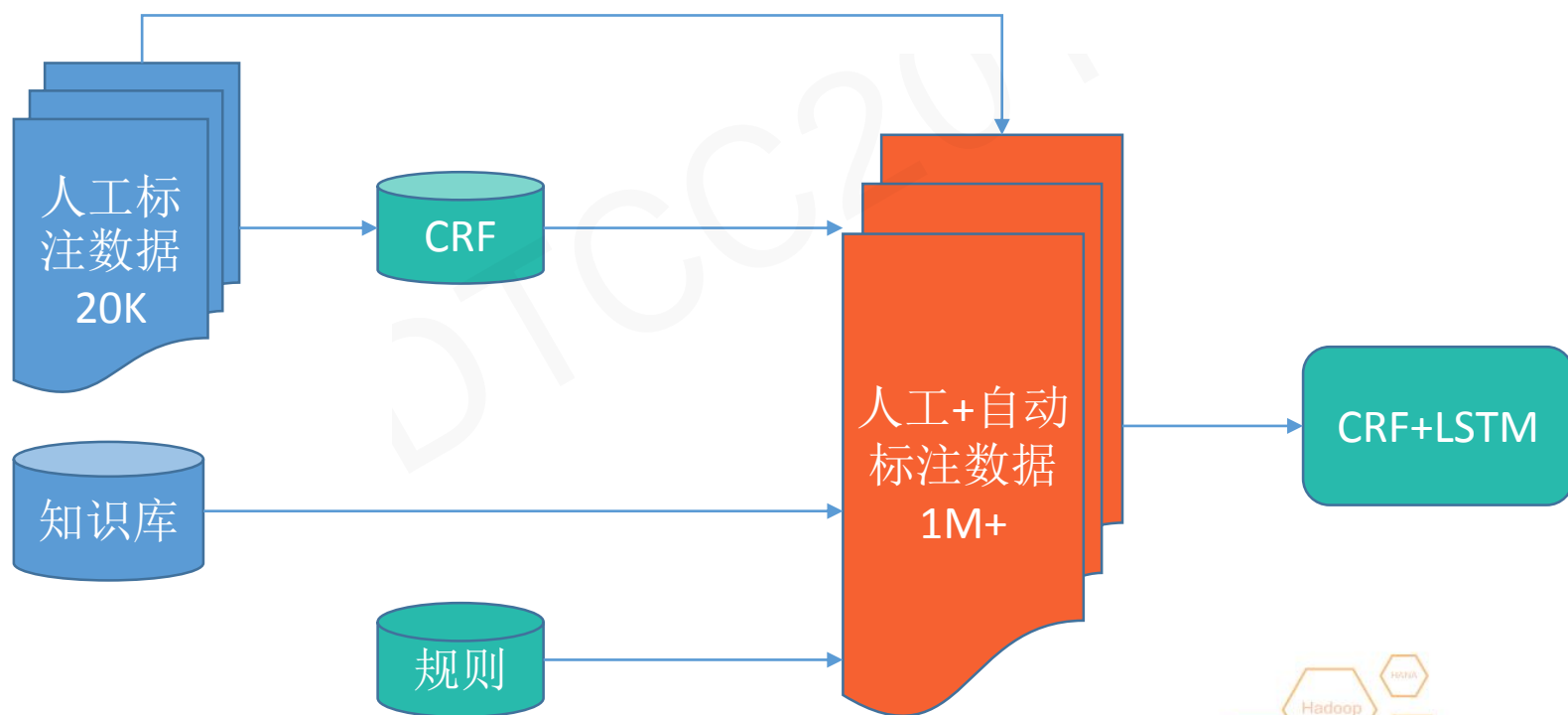
- 定义

- ✓ 商家，如：金百万，海底捞
- ✓ 类目，如：火锅，休闲娱乐
- ✓ 地址
 - 点状地标，如：朝阳大悦城，北京西站
 - 街道，如：知春路，吉庆街
 - 商圈区划，如：望京，五道口
- ✓ 产品
 - 菜品，如：宫保鸡丁
 - 服务项目，如：光子嫩肤
 - 旅游项目，如：线路游
- ✓ SPU
- ✓ 未知

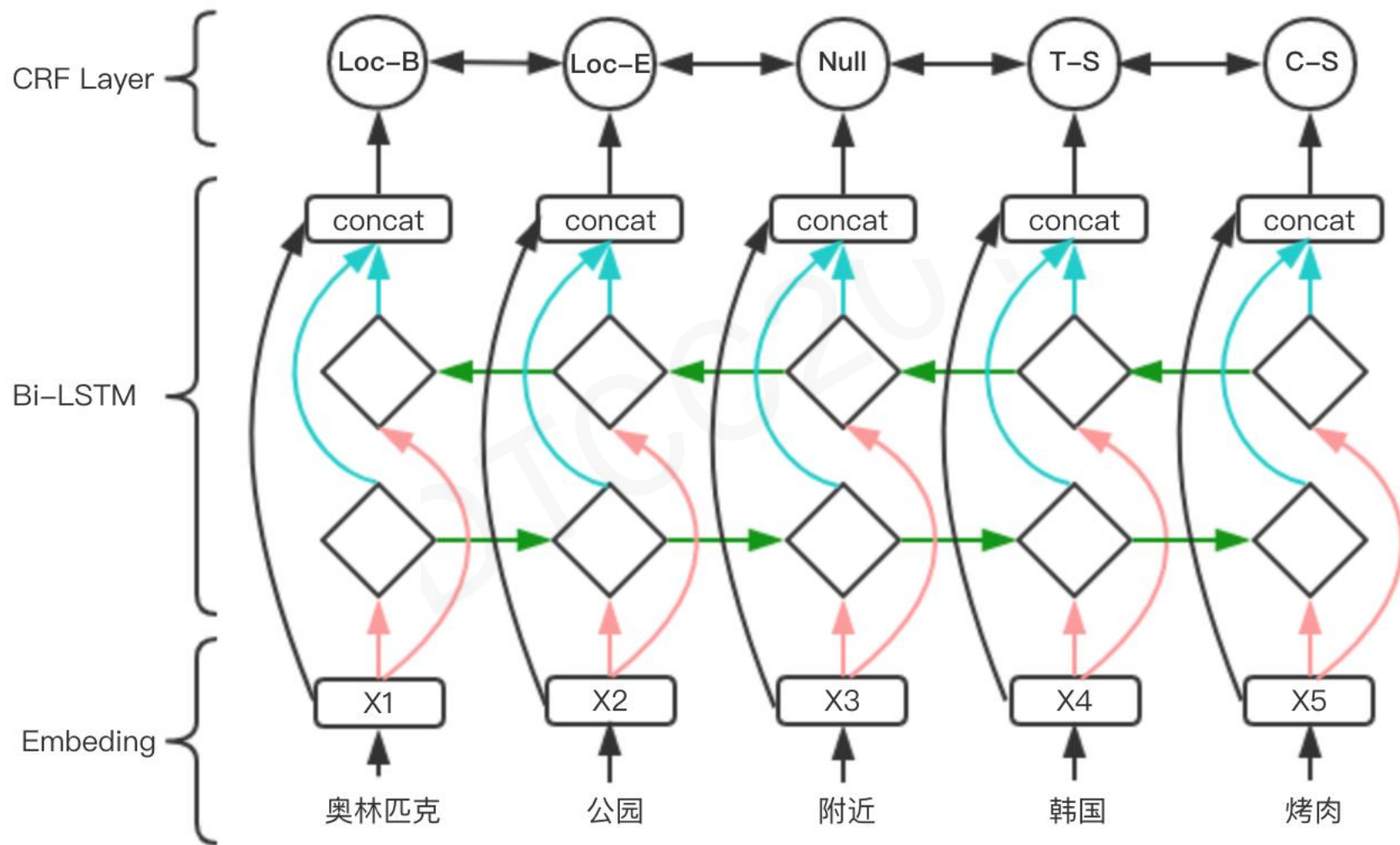


实体识别 2/3

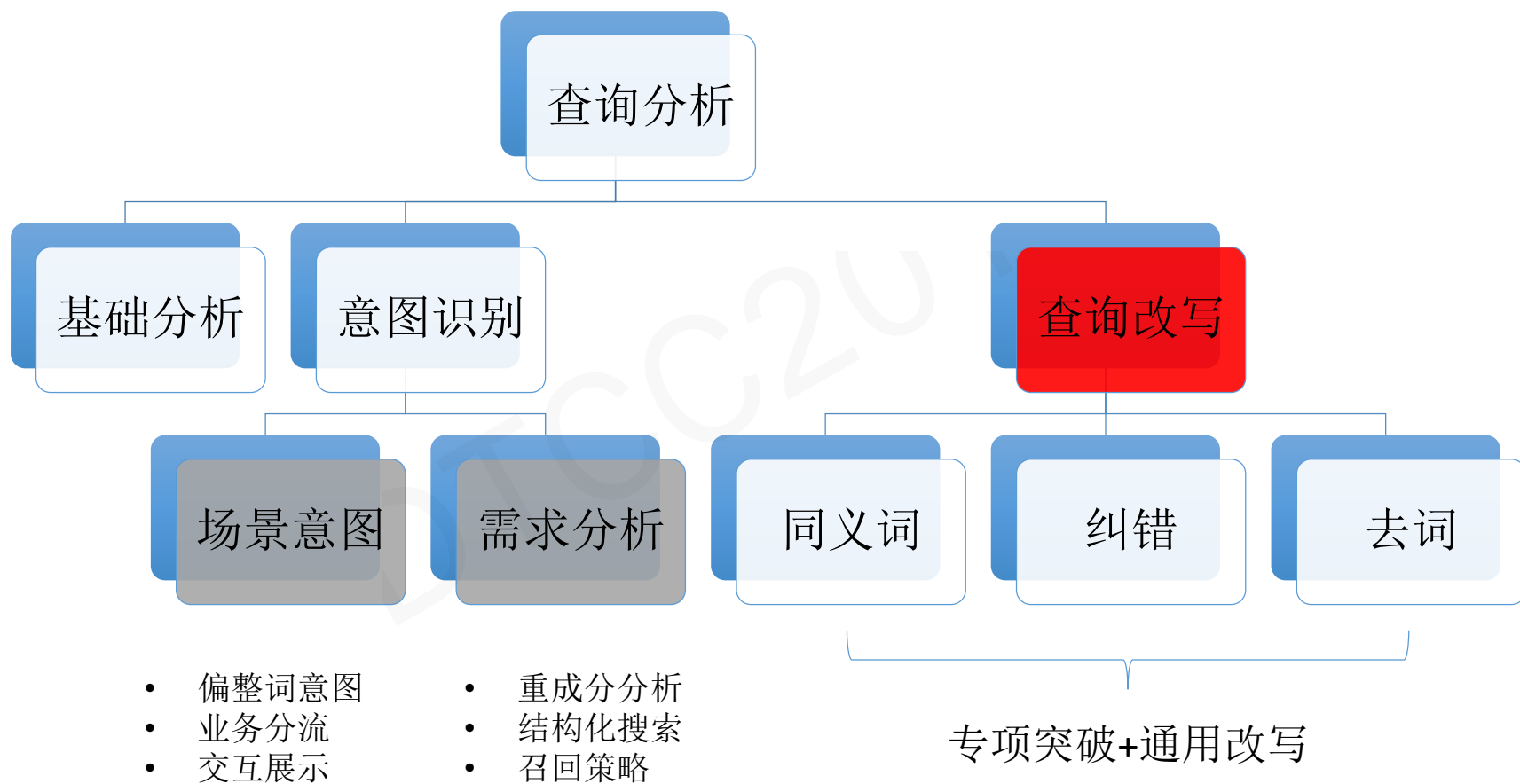
- ✓ 人工标注+CRF做基础模型，产出高精度结果补充标注
- ✓ 知识库+规则方式产出大量样本
- ✓ CRF+LSTM作为最终模型，准确率90%，召回75%



实体识别 3/3



查询分析全景



查询分析-改写 1/2

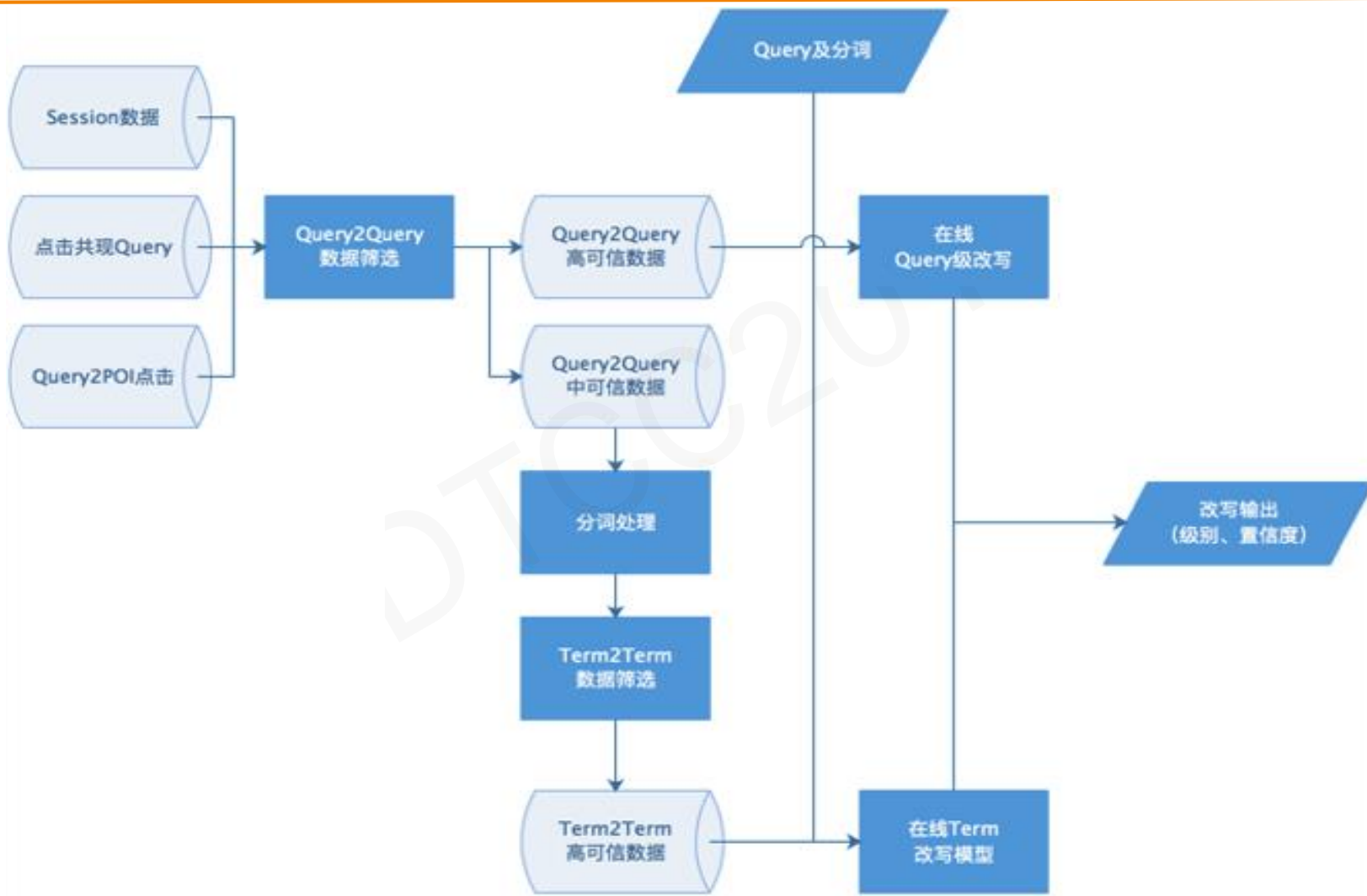
$$P(A/B) = \frac{P(B/A)P(A)}{P(B)}$$



$$P(A/B,C) = \frac{P(B/A,C)P(A,C)}{P(B,C)}$$

✓ C代表场景，包含地域，所属类目等

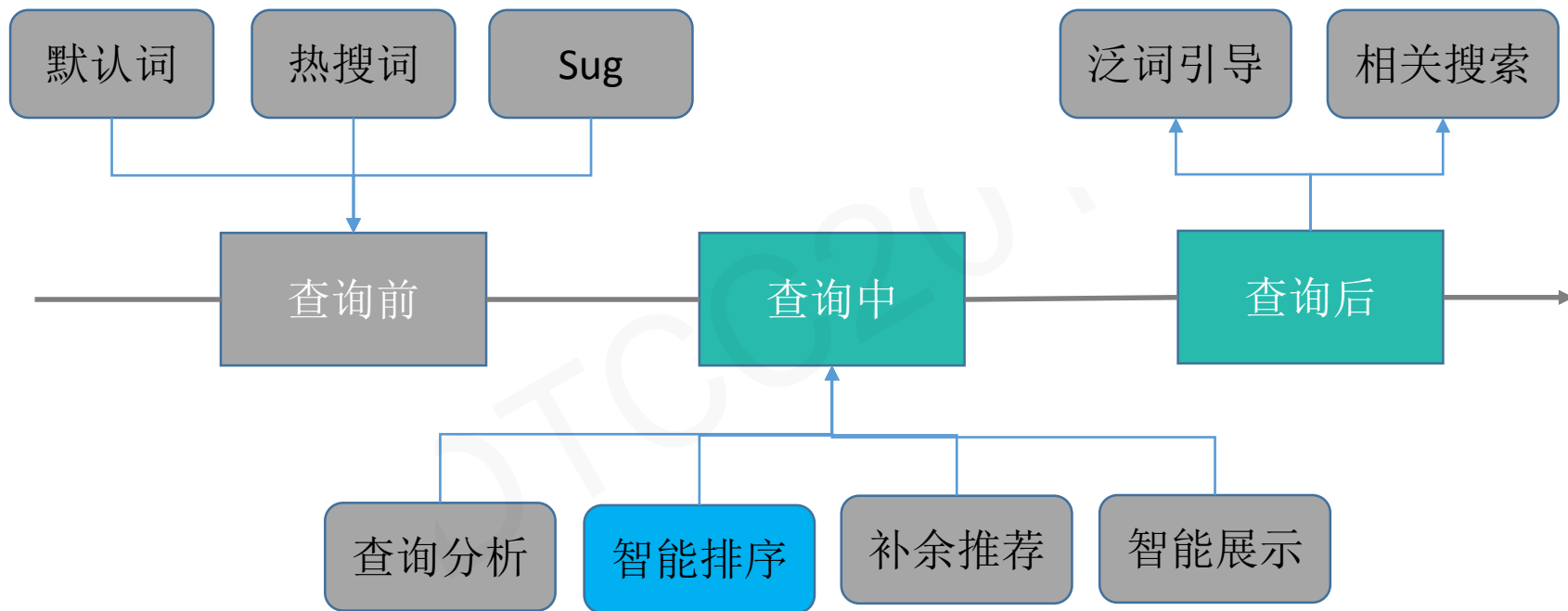
查询分析-改写 2/2



智能排序

降低成本，激发需求，搜索教练

细化需求，表达纠正，需求迁移



精准需求一键触达，提升决策效率

智能排序

召回层



模型层



业务层

- ✓ 解决有无问题
- ✓ 优质结果进精排

- ✓ 转化率目标

- ✓ 可解释性
- ✓ 业务支持
- ✓ 干预运营

智能排序-模型演进

线性模型

GBDT

- ✓ GBDT->XGB
- ✓ 详见技术博客文章《[美团O2O排序解决方案](#)》

PairWise

- ✓ 定义session
- ✓ Pair标注
- ✓ 修改XGB源码提交GitHub被陈天奇审核通过

实时模型

- ✓ 特征实时
- ✓ XGB+LR

深度学习

- ✓ 在线深度学习

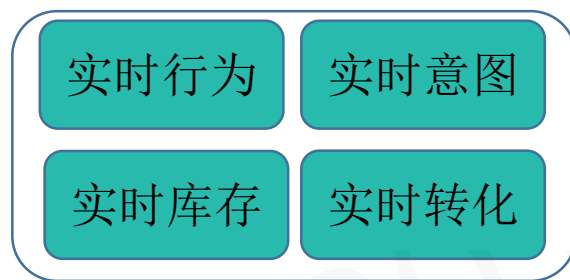


实时模型

实时模型-why



实时化-演进



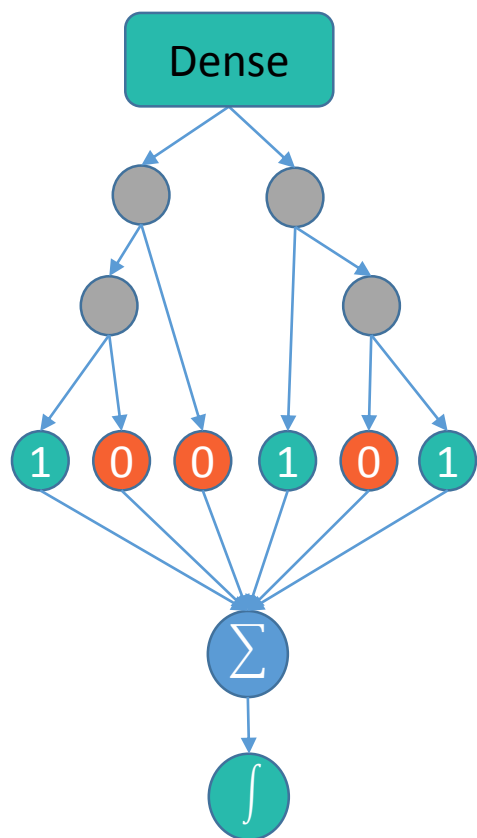
$$y = f\left(\sum_i w_i x_i\right) \rightarrow y = f\left(\sum_i w_i \tilde{x}_i\right) \rightarrow y = f\left(\sum_i \tilde{w}_i \tilde{x}_i\right)$$

特征实时化

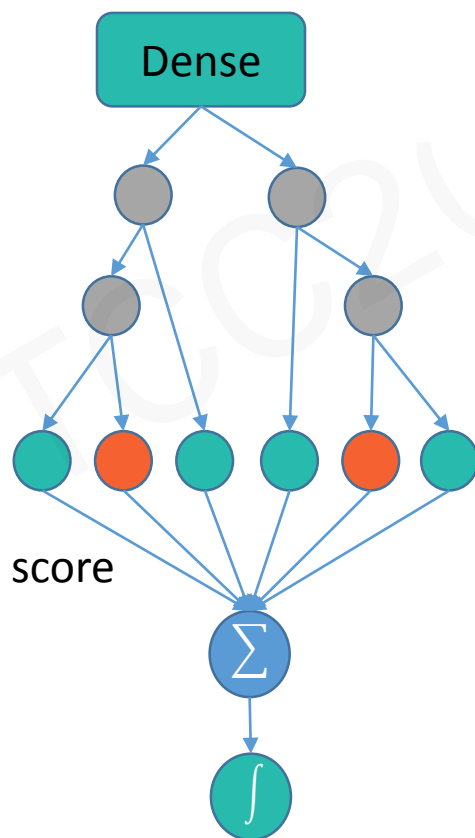
在线学习

实时化-模型优化

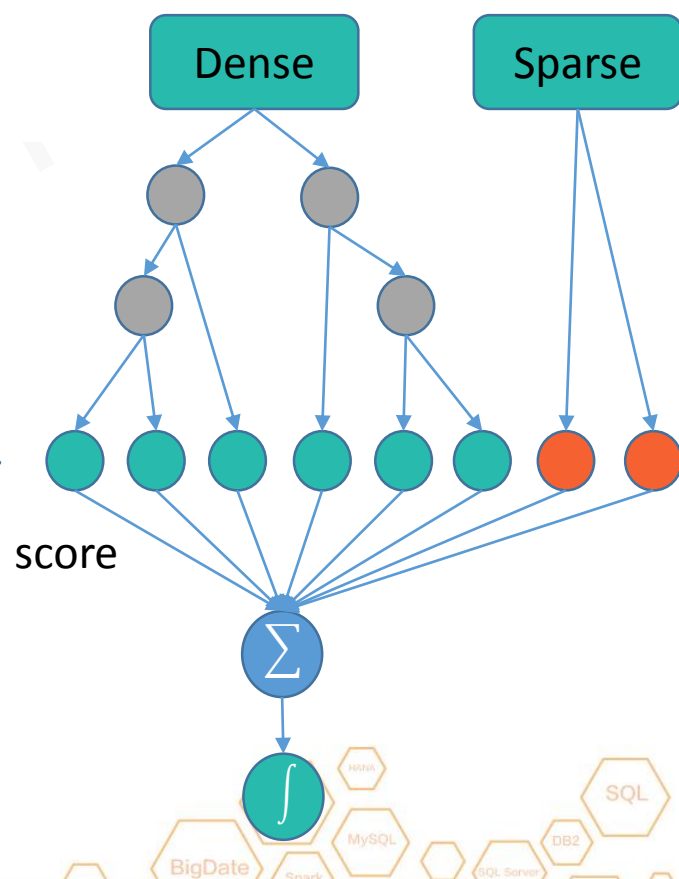
GBDT+LR, LR部分在线更新
树模型特征选择和组合



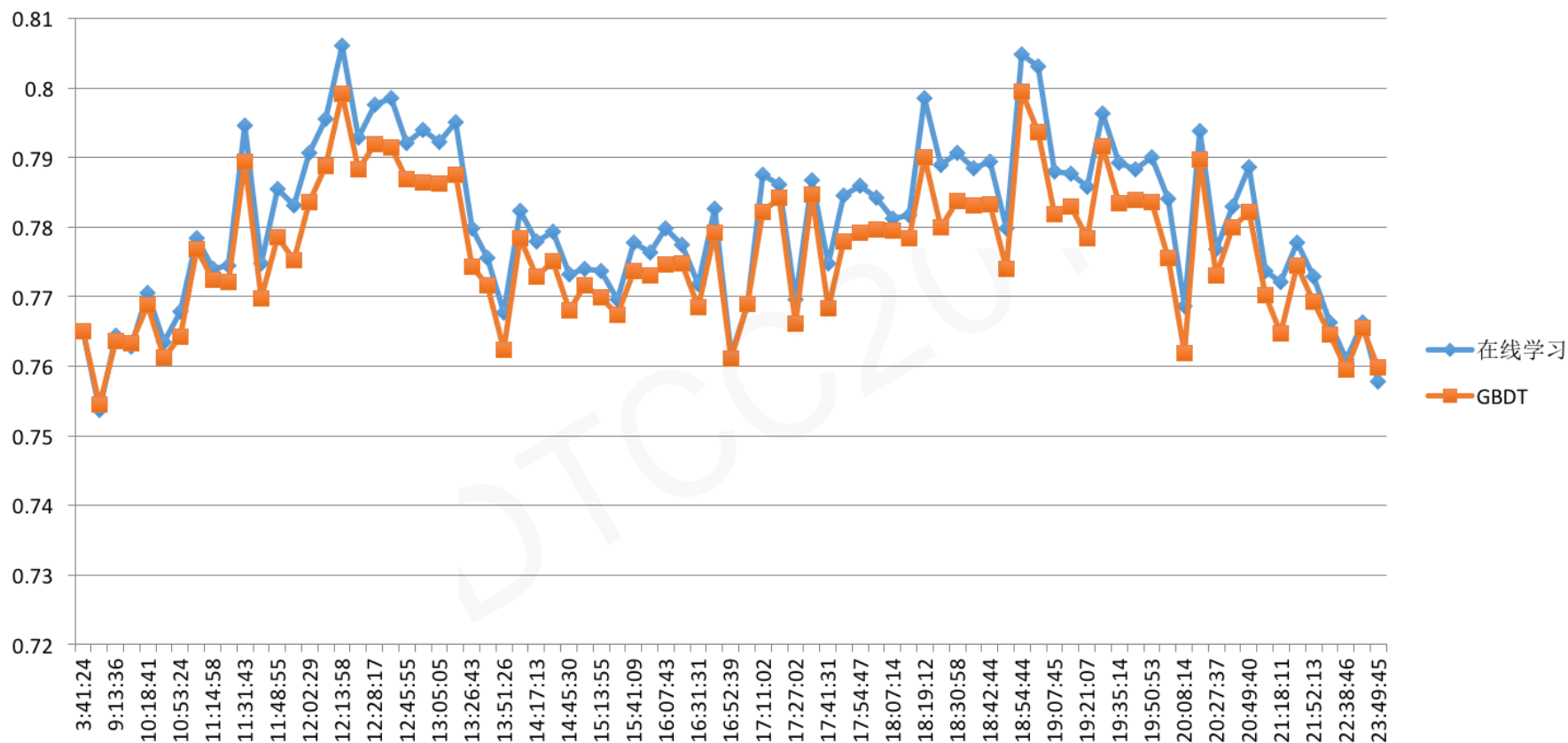
Score信息价值



加入ID特征，提升模型能力



实时化-模型效果



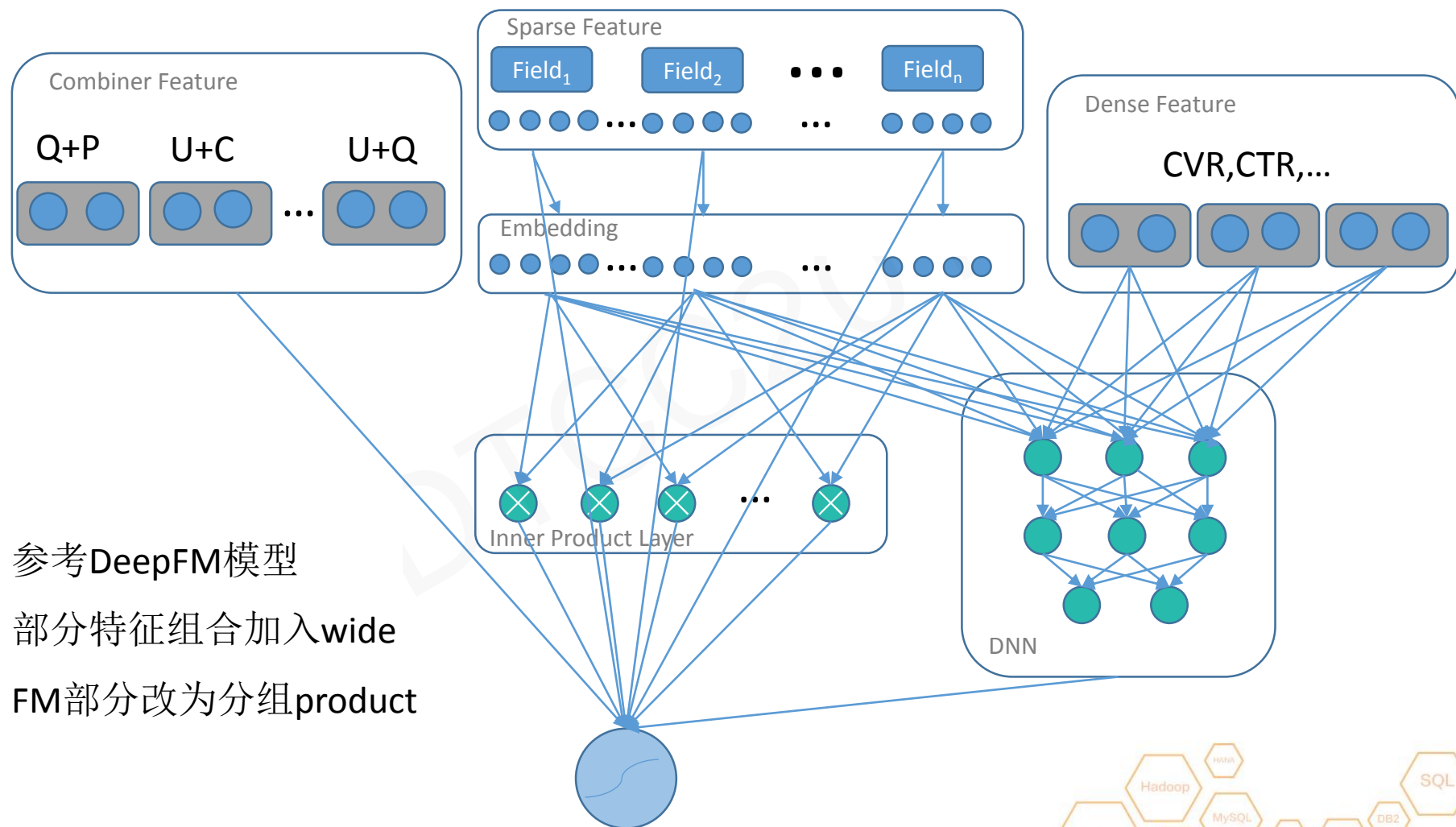
深度学习

深度学习

- ✓ 场景复杂
- ✓ 业务多而差异大
- ✓ 领域专家不够用



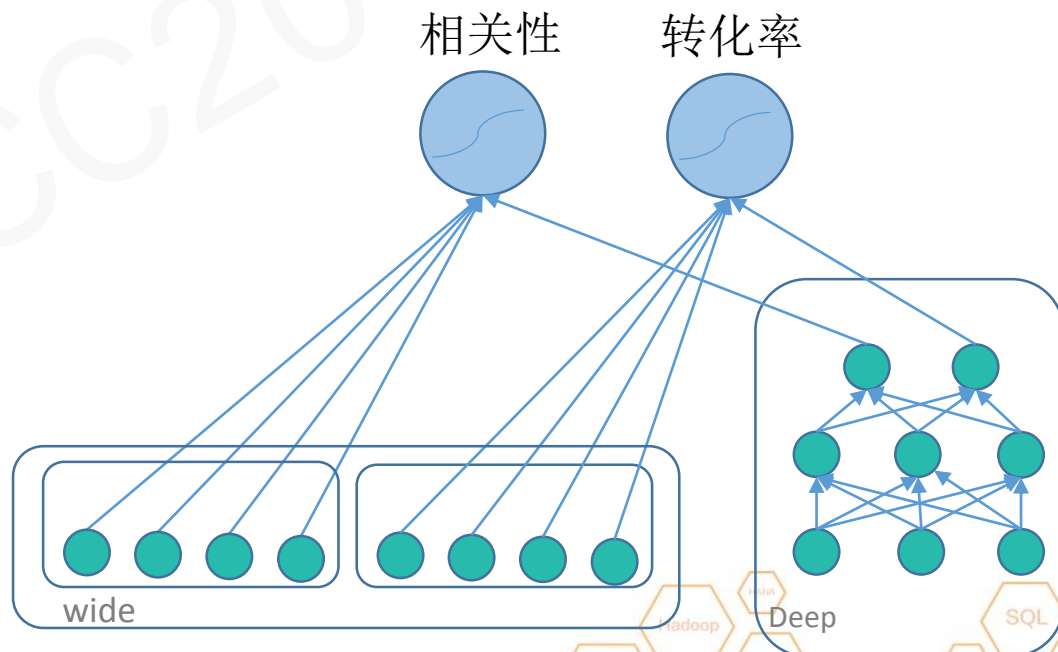
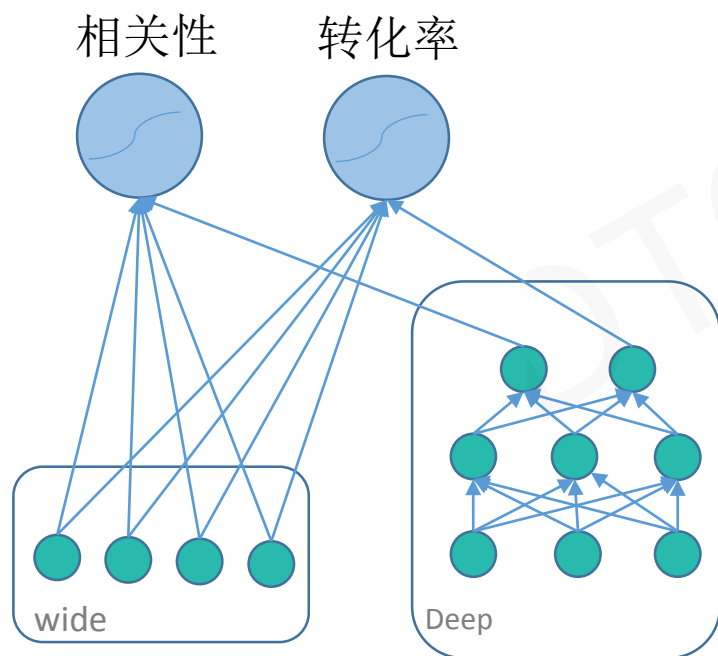
深度学习-基础网络



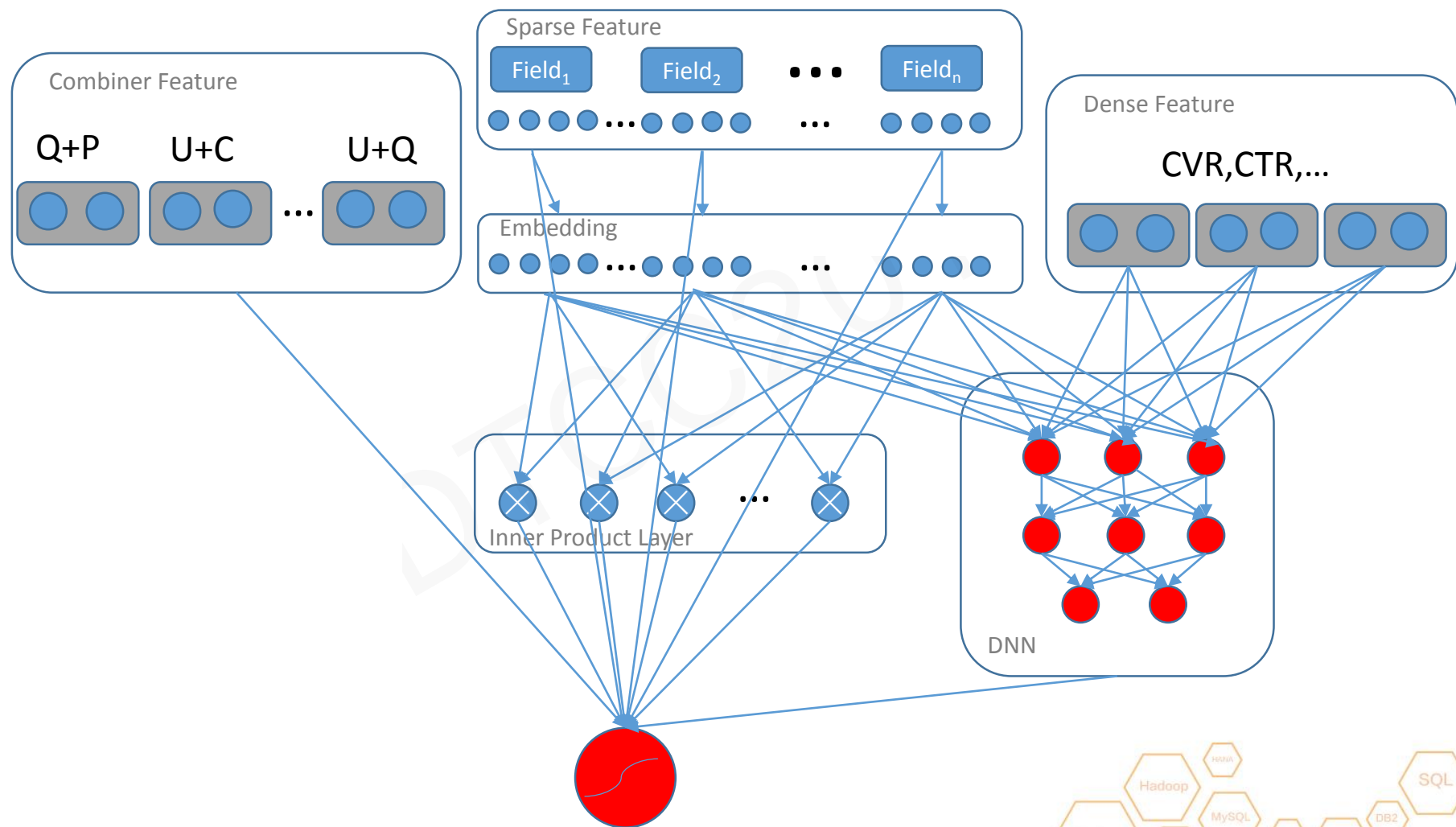
- ✓ 参考DeepFM模型
- ✓ 部分特征组合加入wide
- ✓ FM部分改为分组product

多任务学习

- 用户体验与转化率的平衡
- 转化率与相关性多任务

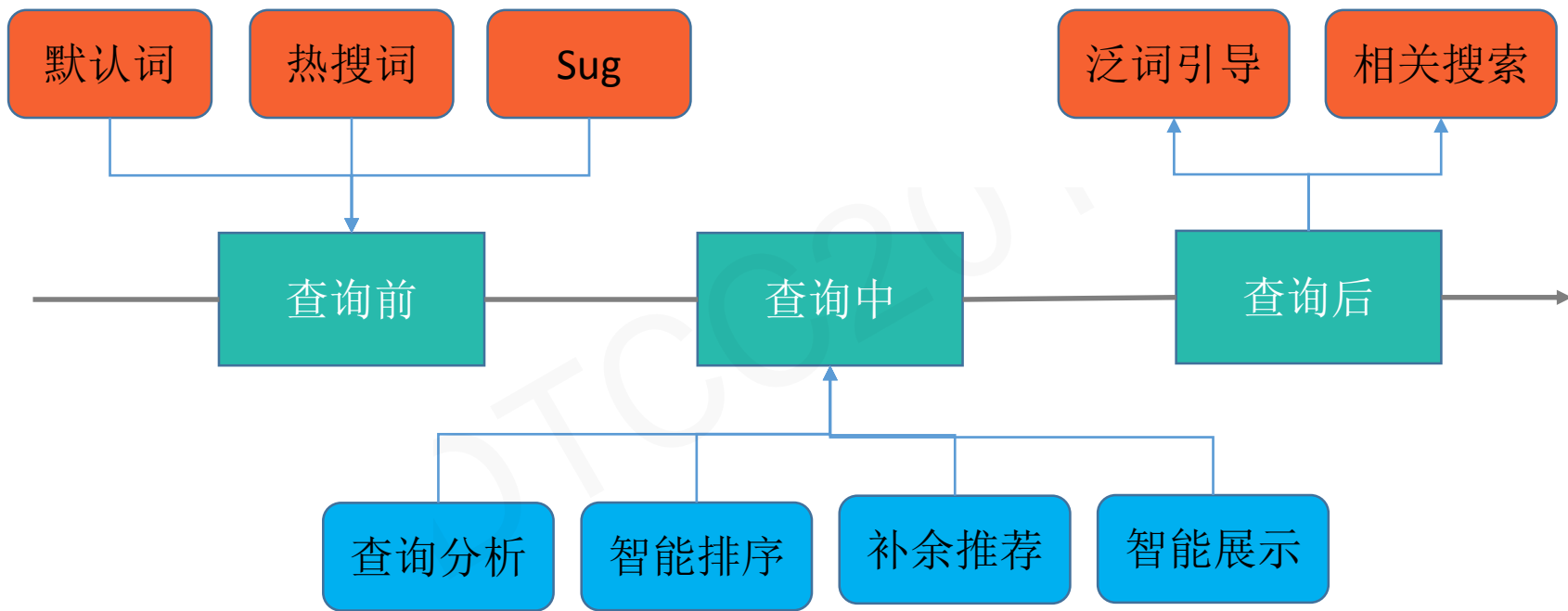


在线深度学习



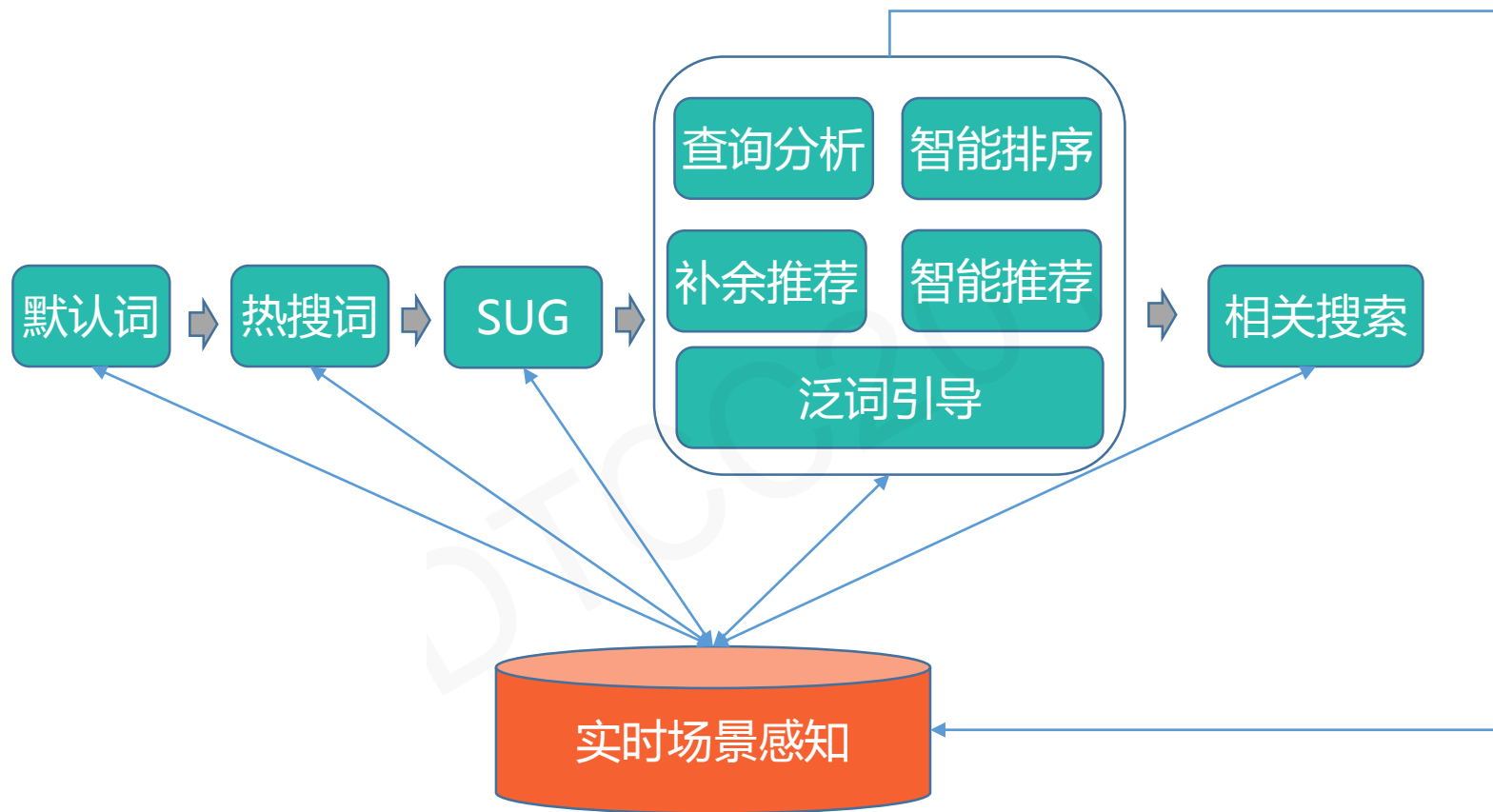
降低成本，激发需求，搜索教练

细化需求，表达纠正，需求迁移

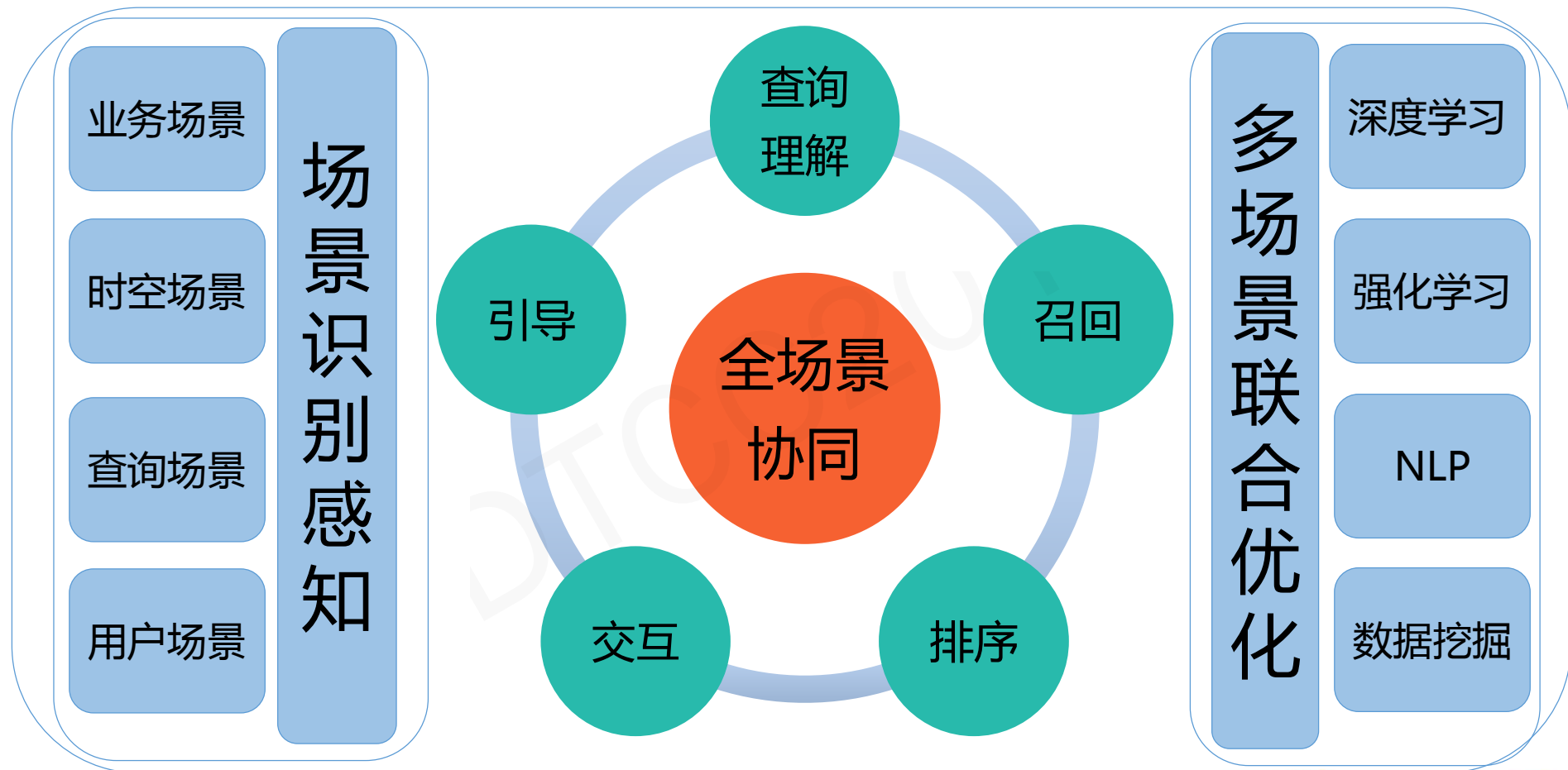


精准需求一键触达，提升决策效率

全场景协同 1/2



全场景协同 2/2





弄潮 

北京 海淀



扫一扫上面的二维码图案，加我微信

SQL

DB2

SQL Server

NewSQL

Oracle

NoSQL

MPP

OLAP



讲师申请

联系电话（微信号）：18612470168

关注“ITPUB”更多
技术干货等你来拿~

与百度外卖、京东、魅族等先后合作系列分享活动



让学习更简单

微学堂是以ChinaUnix、ITPUB所组建的微信群为载体，定期邀请嘉宾对热点话题、技术难题、新产品发布等进行移动端的在线直播活动。

截至目前，累计举办活动期数60+，参与人次40000+。

ITPUB学院

ITPUB学院是盛拓传媒IT168企业事业部（ITPUB）旗下
企业级在线学习咨询平台
历经18年技术社区平台发展
汇聚5000万技术用户
紧随企业一线IT技术需求
打造全方式技术培训与技术咨询服务
提供包括企业应用方案培训咨询（包括企业内训）
个人实战技能培训（包括认证培训）
在内的全方位IT技术培训咨询服务

ITPUB学院讲师均来自于企业
一些工程师、架构师、技术经理和CTO
大会演讲专家1800+
社区版主和博客专家500+

培训特色

无限次免费播放
随时随地在线观看
碎片化时间集中学习
聚焦知识点详细解读
讲师在线答疑
强大的技术人脉圈

八大课程体系

基础架构设计与建设
大数据平台
应用架构设计与开发
系统运维与数据库
传统企业数字化转型
人工智能
区块链
移动开发与SEO



联系我们

联系人：黄老师
电话：010-59127187
邮箱：edu@itpub.net
网址：edu.itpub.net
培训微信号：18500940168