

输入数据：m×n用户app行为矩阵，数字表示用户使用强度

	app1	app2	app3		appN	
设备1	0.3	0	1		0.8	
设备2	0.1	2	0		1.2	
设备3	0.5	0	1		0	
⋮	⋮	⋮	⋮	⋮	⋮	⋮

目标函数：

$$\min_{b^*, q^*, p^*} \sum_{(u, a)} (r_{ua} - \mu - b_a - b_u - q_a^T p_u)^2 + \lambda(b_a^2 + b_u^2 + \|q_a\|^2 + \|p_u\|^2)$$

其中：

r_{ua} 为用户在app的平均使用强度

μ 为固定效应

λ 为正则项的权重

b_a 、 b_u 分别为应用的固定效应以及用户的固定偏好

q_a 、 p_u 分别为应用的隐含因子向量以及用户隐含因子向量

CHAPTER 03

用户特征在LFM中的使用

用户特征在LFM中的使用——性别的差异

- 女性用户对宝石消除游戏，美图拍照等比较关注。
- 男性用户对商务办公、汽车投资等比较感兴趣。



女性

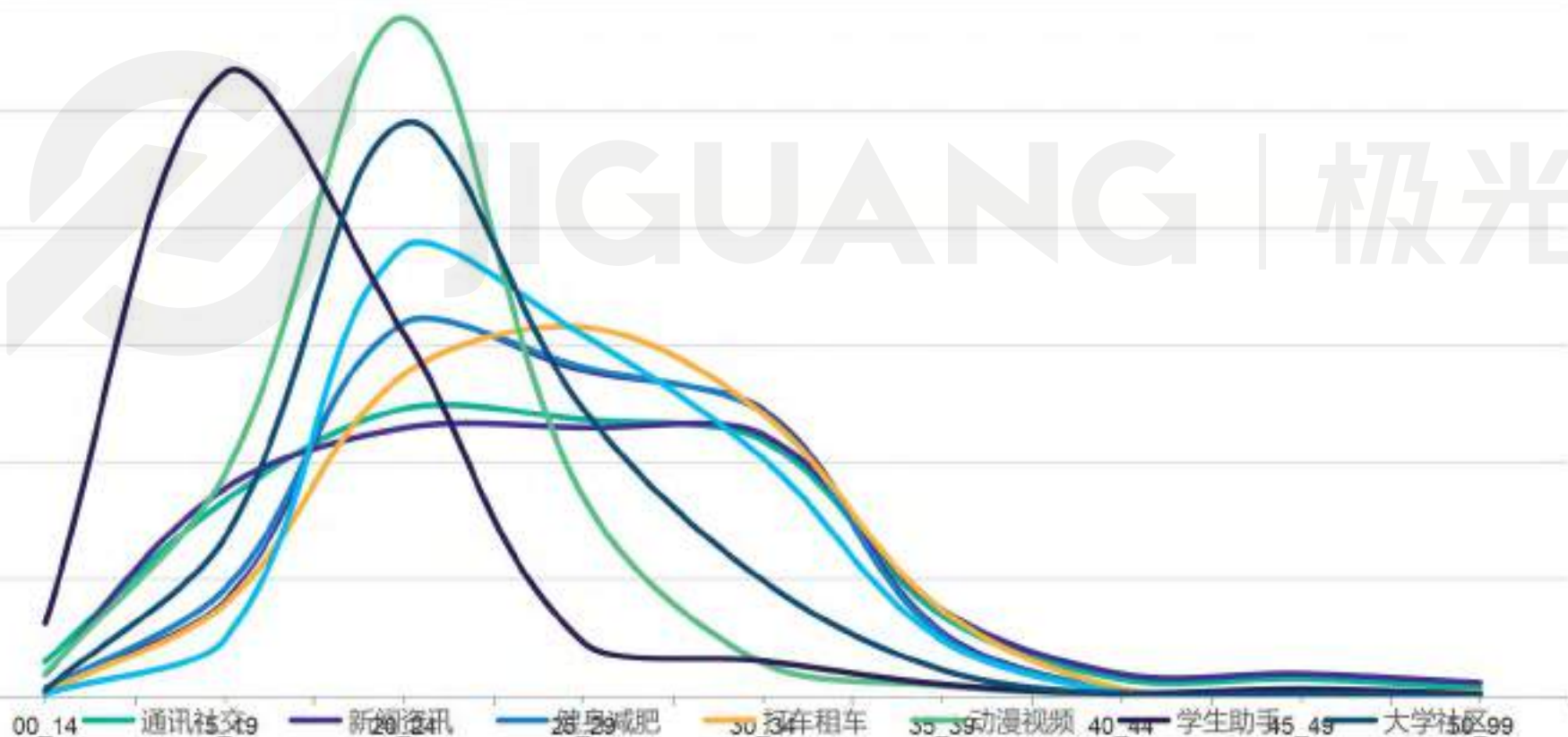


男性



用户特征在LFM中的使用——年龄的差异

- 不同的年龄在应用的选择上会有很大的差异
- 动漫视频、大学社区用户集中在20-24岁、打车租车类的集中在20-34岁。



目标函数：

$$\min_{b_*, q_*, p_*} \sum_{(u,a)} (r_{ua} - \mu - b_a - b_u - q_a(p_u + \sum_{a \in A(u)} y_a))^2 + \lambda(b_a^2 + b_u^2 + \|q_a^2\| + \|p_u^2\|)$$

其中：

集合A是性别、年龄、城市、收入、机型等一系列属性信息的集合

y_a 是这些属性的效应。

分值预测公式：

$$\hat{r}_{ua} = \mu + b_a + b_u + q_a(p_u + \sum_{a \in A(u)} y_a)$$



情景一：

随着时间的推移，app的产品会迭代，体验、定位、市场影响力都可能发生变化。



情景二：

因为工作学习环境改变、对事物的认识的改变会对应用服务的需求产生不同的需求。

分值预测做出以下调整：

$$\hat{r}_{ua}(t) = \mu + b_a(t) + b_u(t) + q_a(p_u(t) + \sum_{a \in A(u)} y_a)$$

其中 q_a 是应用本身的一些特质，并不会随着时间而变化。

CHAPTER 04

空间效应的使用

程序员的轨迹画像

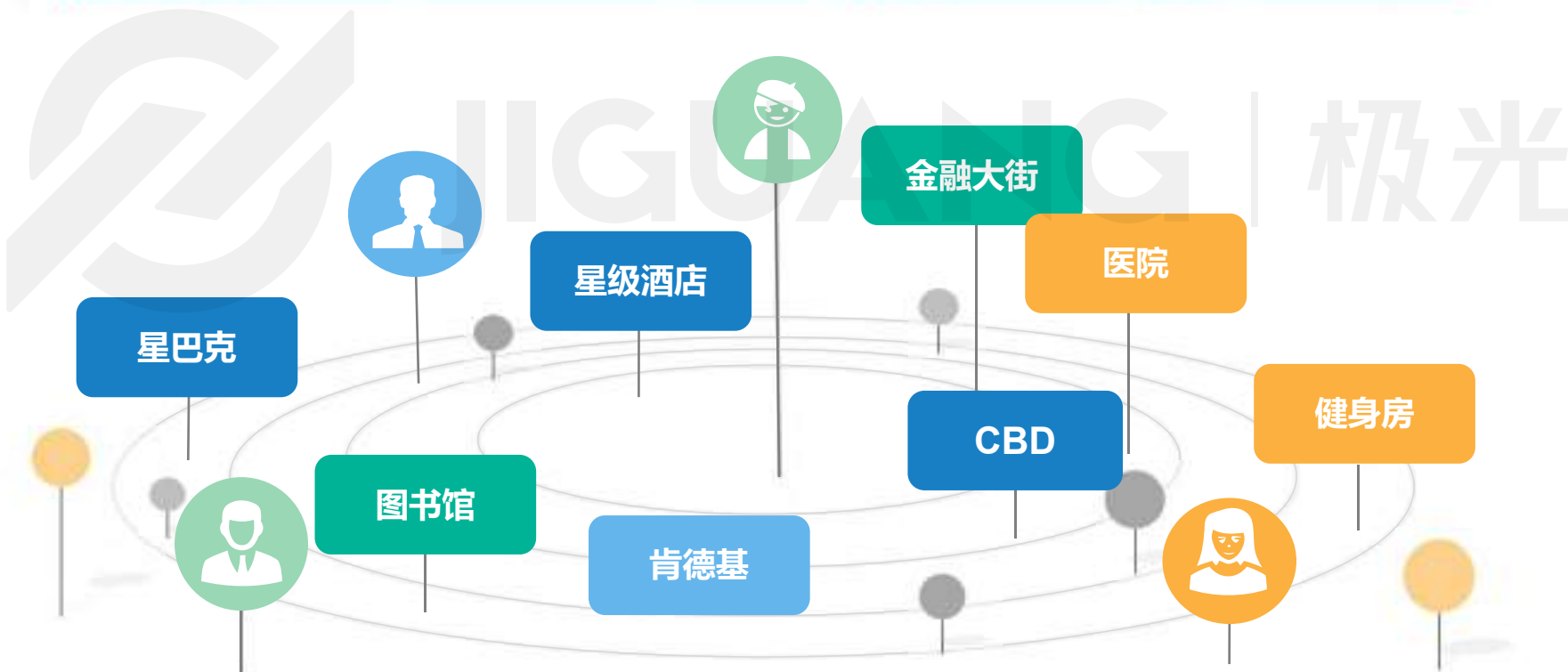
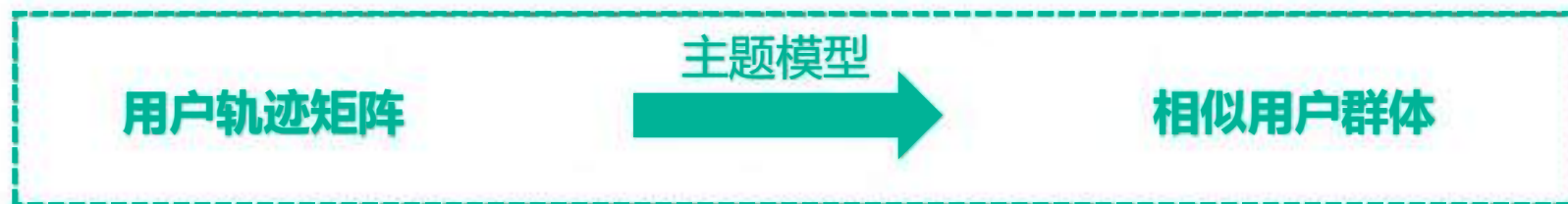


•基于大量实时的lbs信息我们可以很好地对用户进行生活、工作各个角度的地理位置画像。

•生活轨迹会反应着用户的应用使用习惯。

空间效应的使用——基于轨迹画像寻找相似群体

基于同城用户的轨迹矩阵通过主题模型（LDA）训练，我们可以很好地得到不同轨迹人群相似的群体。



- 紧密的社交网络圈子是最优秀的内容传播渠道
- 在wifi连接信息的基础上，借助图挖掘中的频繁子图识别算法我们可以很好地识别社交网络中关系紧密的圈子

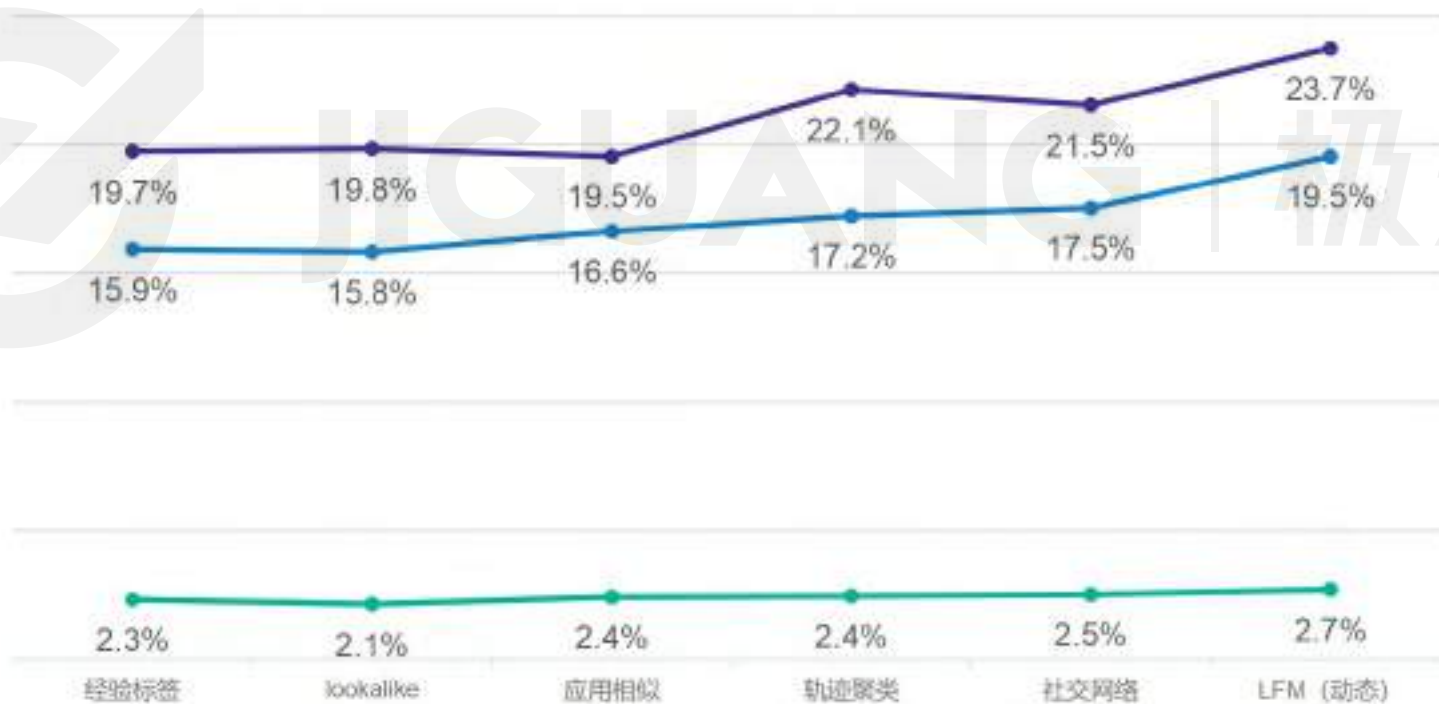
社交网络中的频繁子图



空间效应——结果分析

- Lookalike以及应用相似方法基本好经验标签持平
- 空间效应和LFM会优于经验结果。其中LFM会表现的最好。

各方法效果对比

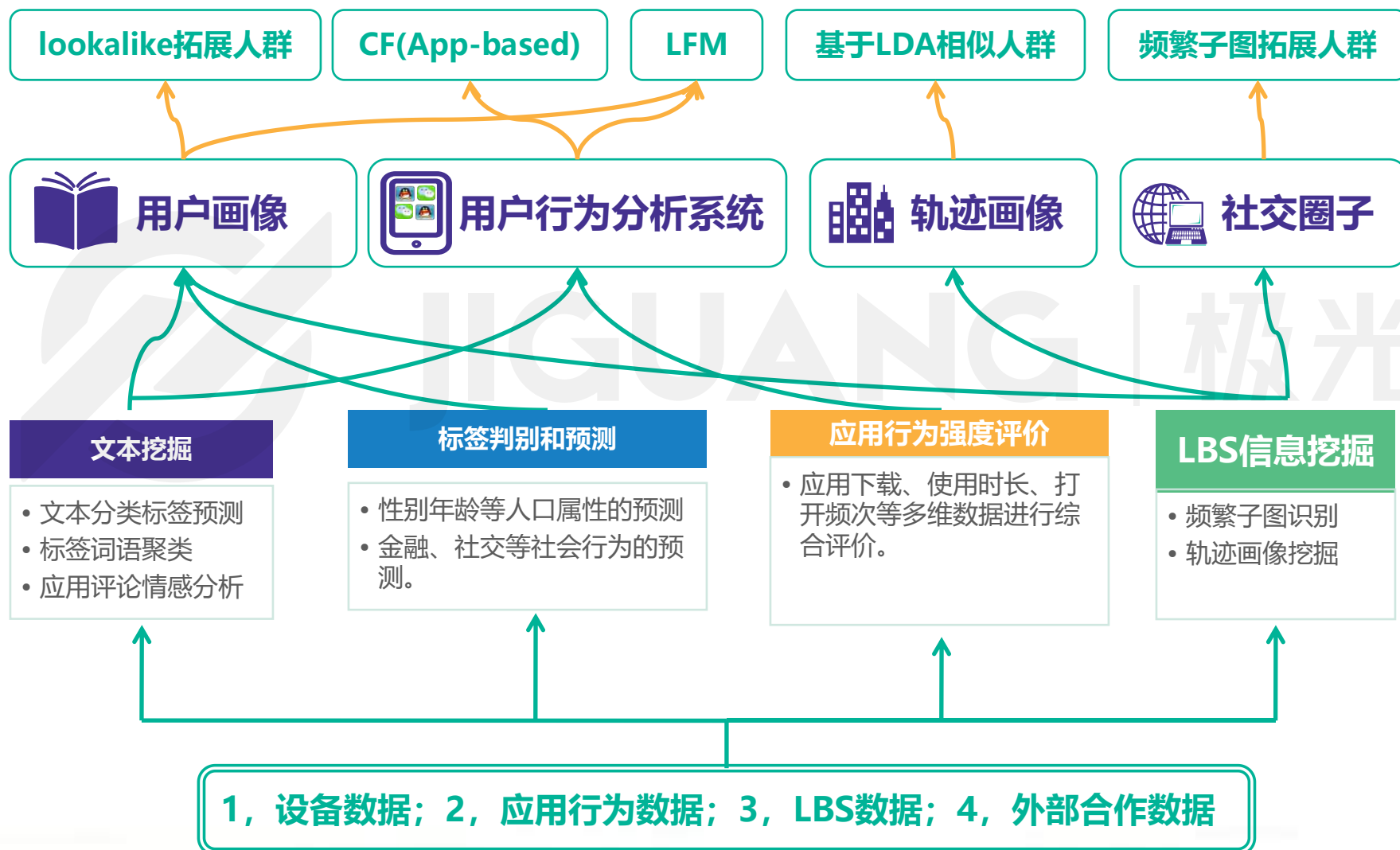


** 对比来至于12个大行业 百余款产品 在多类型渠道投放的监控数据的均值

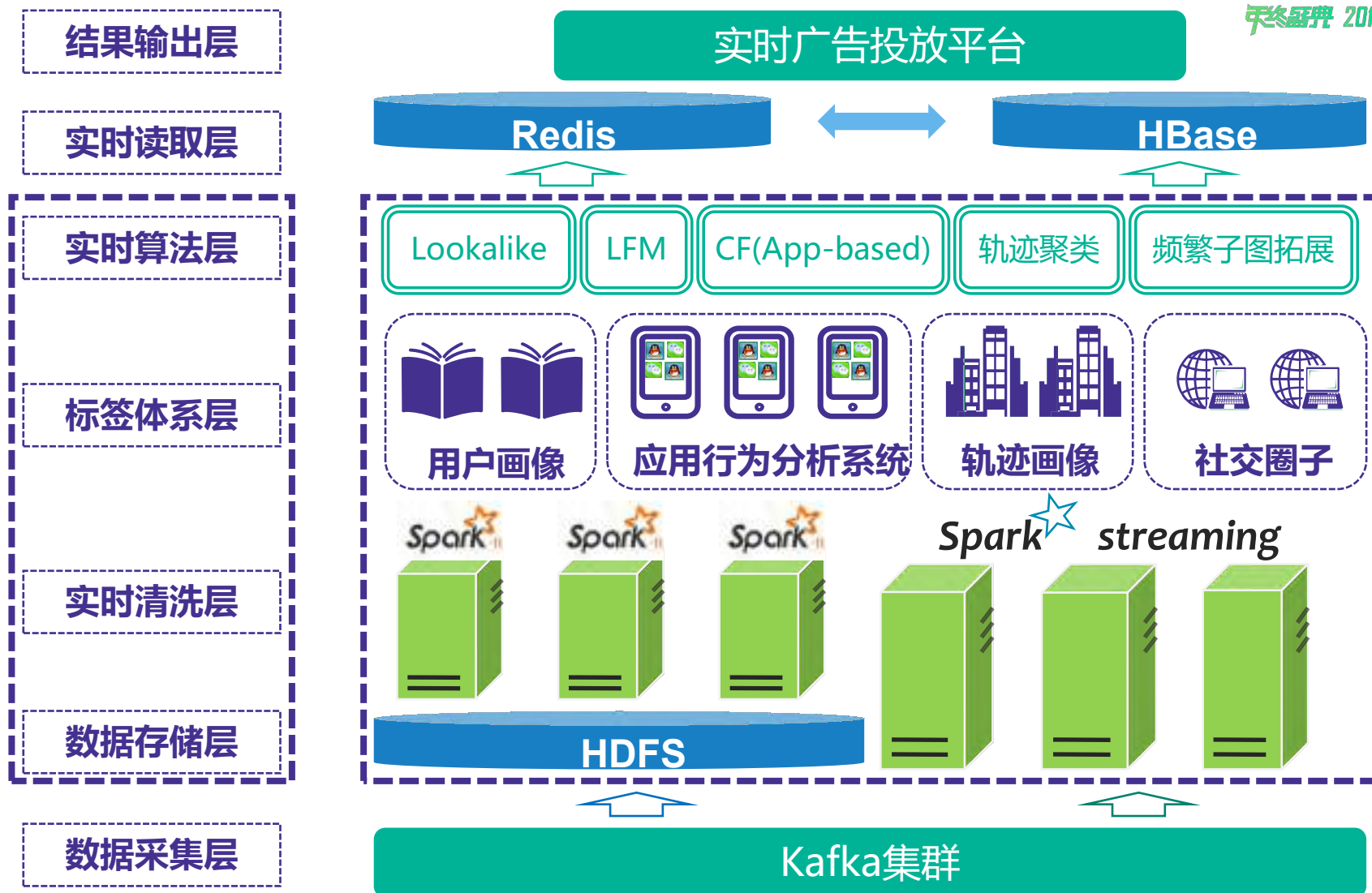
CHAPTER 05

实时算法架构体系

实时算法架构体系——算法逻辑结构



实时算法架构体系——算法实现物理架构



END

T H A N K S

