



推荐系统实时化的实践和思考

Several Internals about Real-time RecSys

Tianjian Chen

Baidu Inc.

- 推荐系统实时化的需要
- 简单热身：实时更新的新热推荐
- 主流选择：基于反向索引的系统
- 延伸探索：协同过滤实时化

推荐时机 Timing

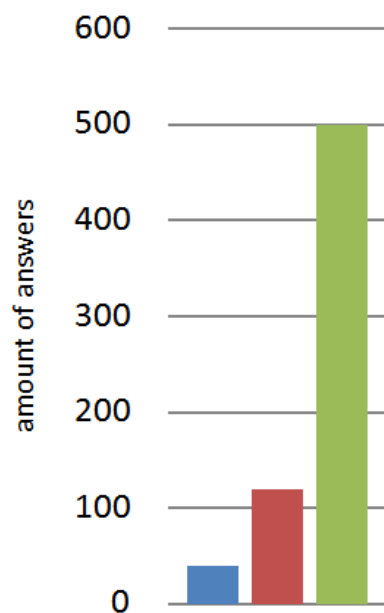
- 兴趣发现和收敛速度
- 对于智能程度的感知

推荐质量 Quality

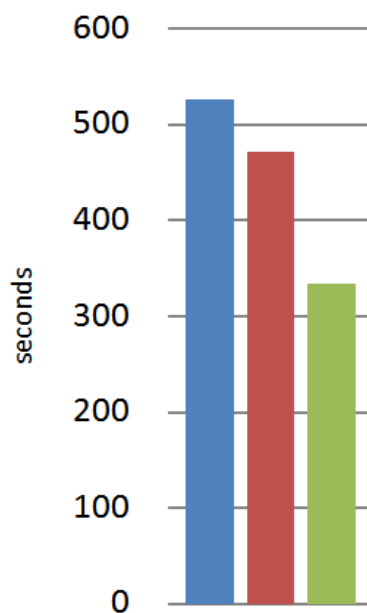
- 内容时效性
- 列表新颖性

推荐系统实时化的效果

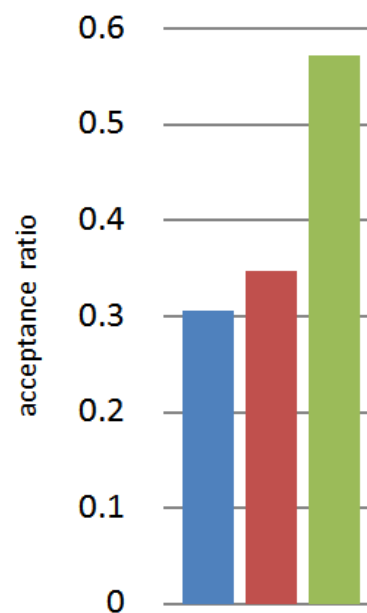
Receiving Answers



Avg Answer Time



Acceptance Ratio

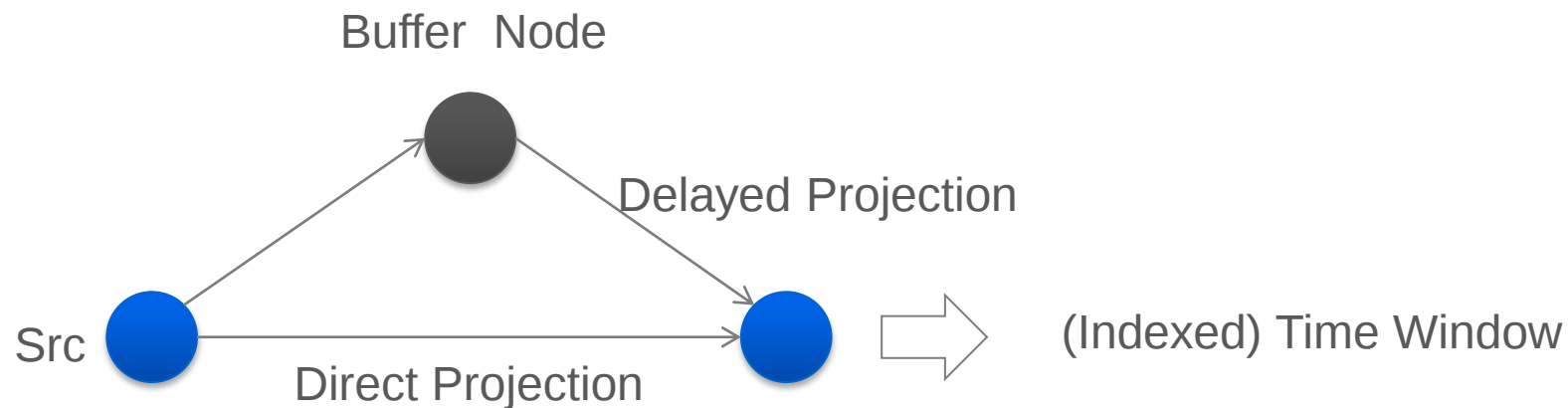


■ TRQ ■ UQRS ■ CRS

$$f(x) = Eval\{\sum_{t=T_0}^{T_0+\Delta T} X(t), \Delta T\}$$

- 时间窗口
- 观测值在时间窗口内的加和
- 新热度评估策略

简单热身：实时更新的新热推荐



一种基于流式计算的可扩展实现

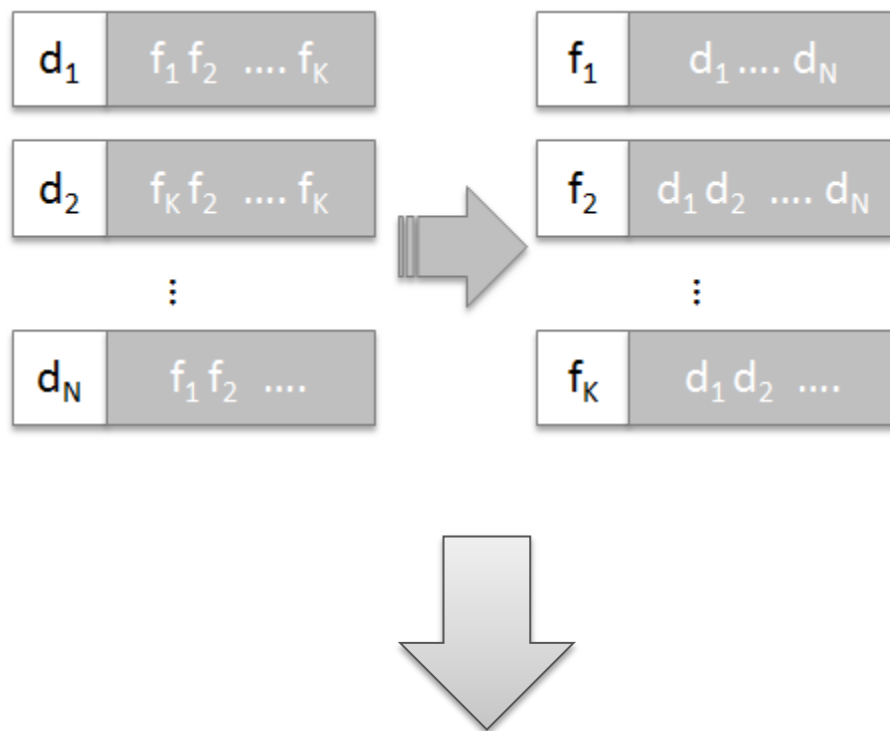
优点 Pros

- 算法设计和工程实现成本低

缺点 Cons

- 新颖性比较差
- 个性化拓展空间小

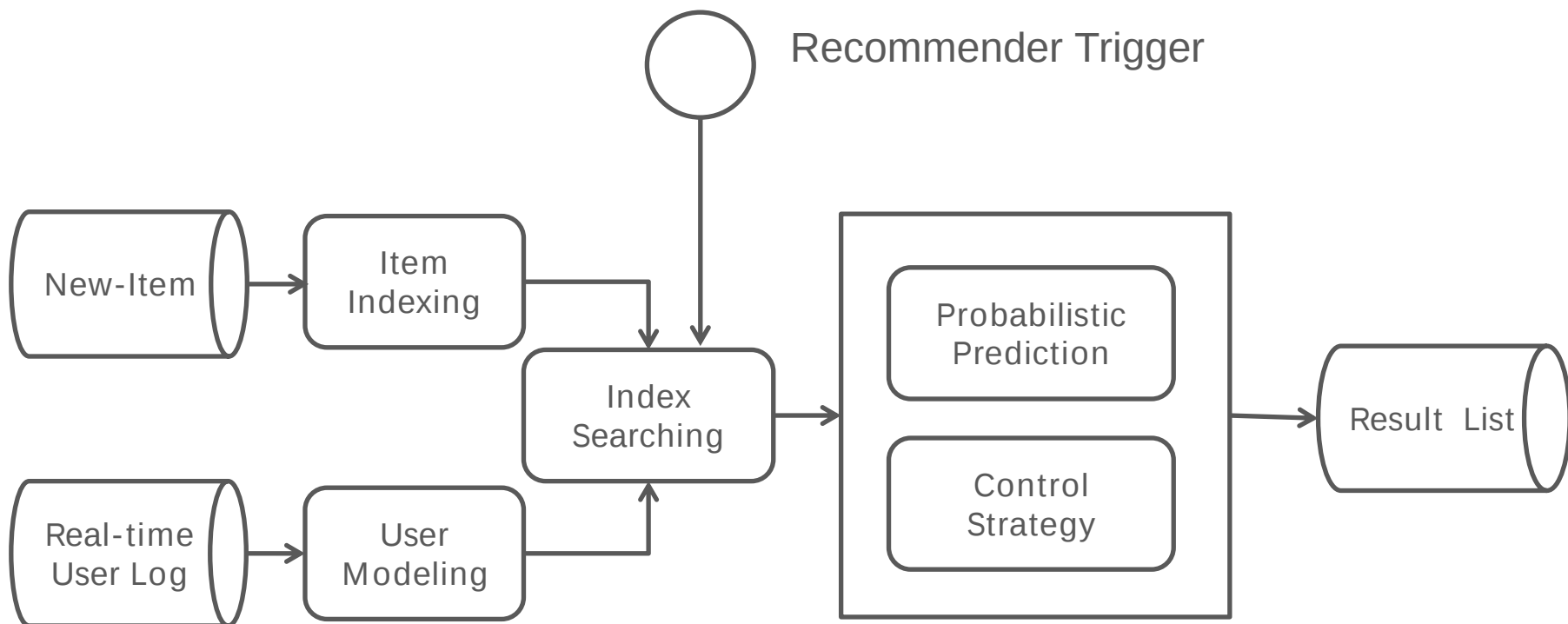
主流选择：基于反向索引的系统



对召回的结果进行排序

概率预估: Max Ent, RF, LR

主流选择：基于反向索引召回的系统



优点 Pros

- 准确性和新颖性平衡
- 时效性较好

缺点 Cons

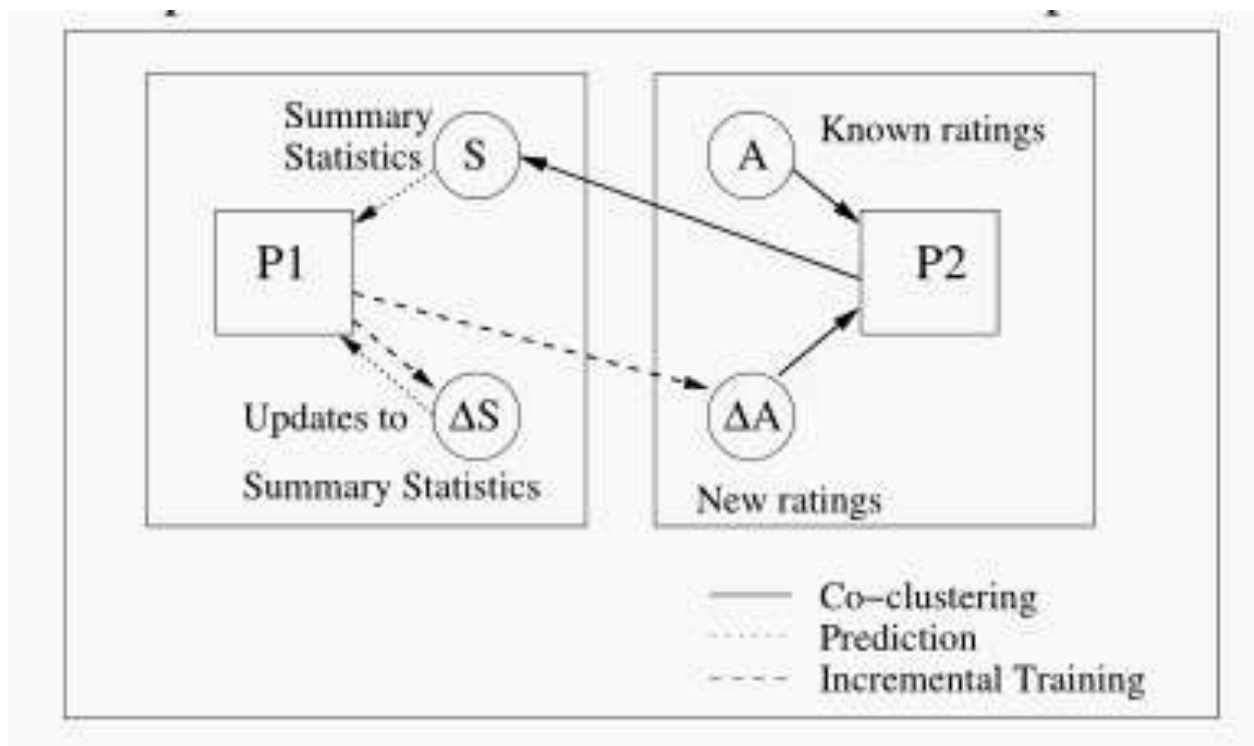
- 召回过程的性能问题

延伸探索：协同过滤实时化



- Cached CF
- Parallel Co-Clustering

延伸探索：协同过滤实时化



优点 Pros

- 时效性大为改进
- 准确性几乎相当

缺点 Cons

- 工程代价比较大

Q & A Time



Thanks😊

Email: chentianjian@baidu.com