



BEIJING 2018

# 爱奇艺信息流推荐 深度学习之路

演讲者 / 方非

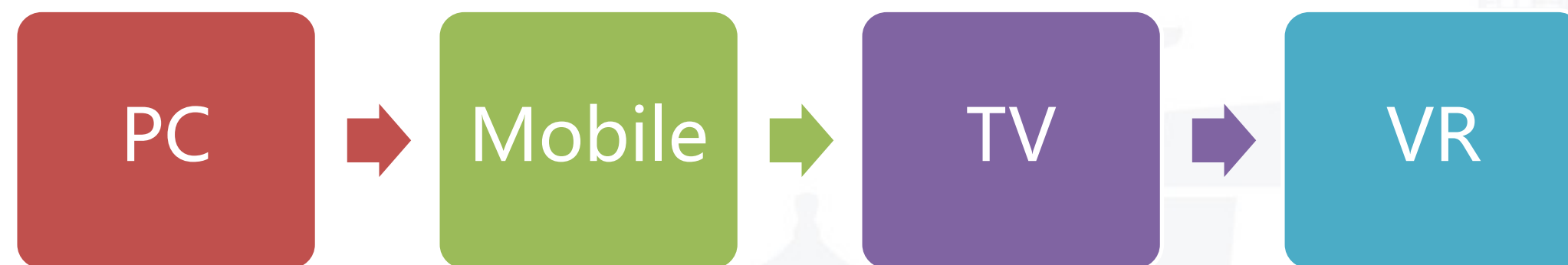
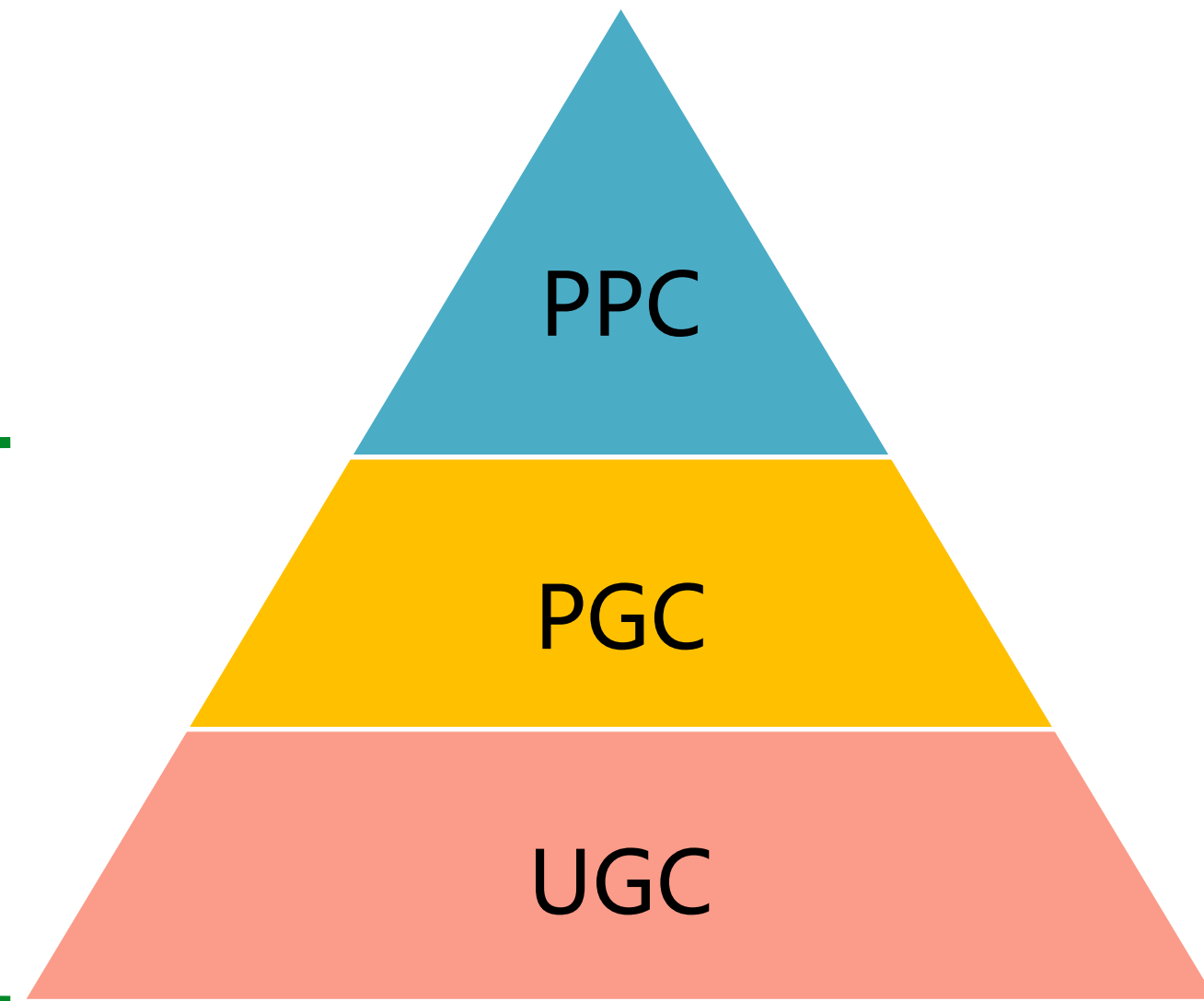
# Agenda

- 爱奇艺信息流推荐简介
- 深度学习设计思路和技术架构
  - 深度召回模型
  - 深度排序模型
- 深度学习平台实践及讨论
- 总结

# 爱奇艺

## 内容生态

## 短视频



**50.8 million** Subscribing Members

Data as of December 31, 2017. Subscribing members refers to the individuals who purchased our monthly, quarterly or annual membership packages, including individuals with trial membership, and excluding individuals who pay for video on-demand services.

**1.7 hours** Average Time Spent per Day per User

Data in December 2017. Refers to the average time spent per day per user watching video content on iQIYI mobile app.

**421.3 million** Mobile MAUs

Average data in 4Q2017. Mobile MAUs refers to the number of unique mobile devices that have accessed iQIYI platform through our iQIYI mobile app at least once during a calendar month.

**424.1 million** PC MAUs

Average data in 4Q2017. PC MAUs refers to the sum of (i) the number of unique PC devices that have accessed our platform through our PC client terminal, and (ii) the number of PC devices that have accessed our platform through our website, in each case at least once during a calendar month.

**126.0 million** Mobile DAUs

Average data in 4Q2017. Mobile DAUs refers to the number of unique mobile devices that have accessed iQIYI platform through our iQIYI mobile app at least once during a day.

**53.7 million** PC DAUs

Average data in 4Q2017. PC DAUs refers to the sum of (i) the number of unique PC devices that have accessed iQIYI platform through our PC client terminal, and (ii) the number of PC devices that have accessed our platform through our website, in each case at least once during a day.



# 信息流

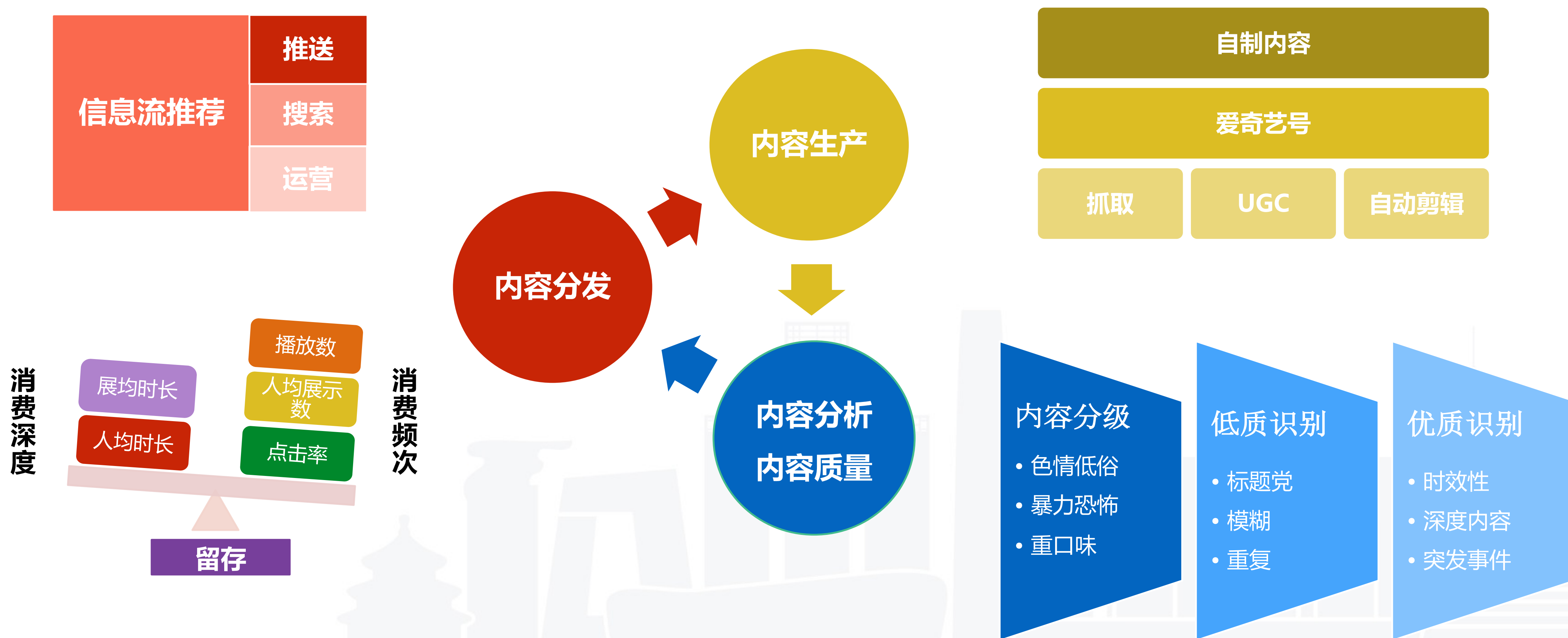
一种通过个性化推荐来实现  
高效内容分发的信息组织方式

## 特点

- 适合碎片化消费
- 培养用户习惯和粘性
- 充分挖掘长尾内容
- 更强的变现能力



# 内容平台

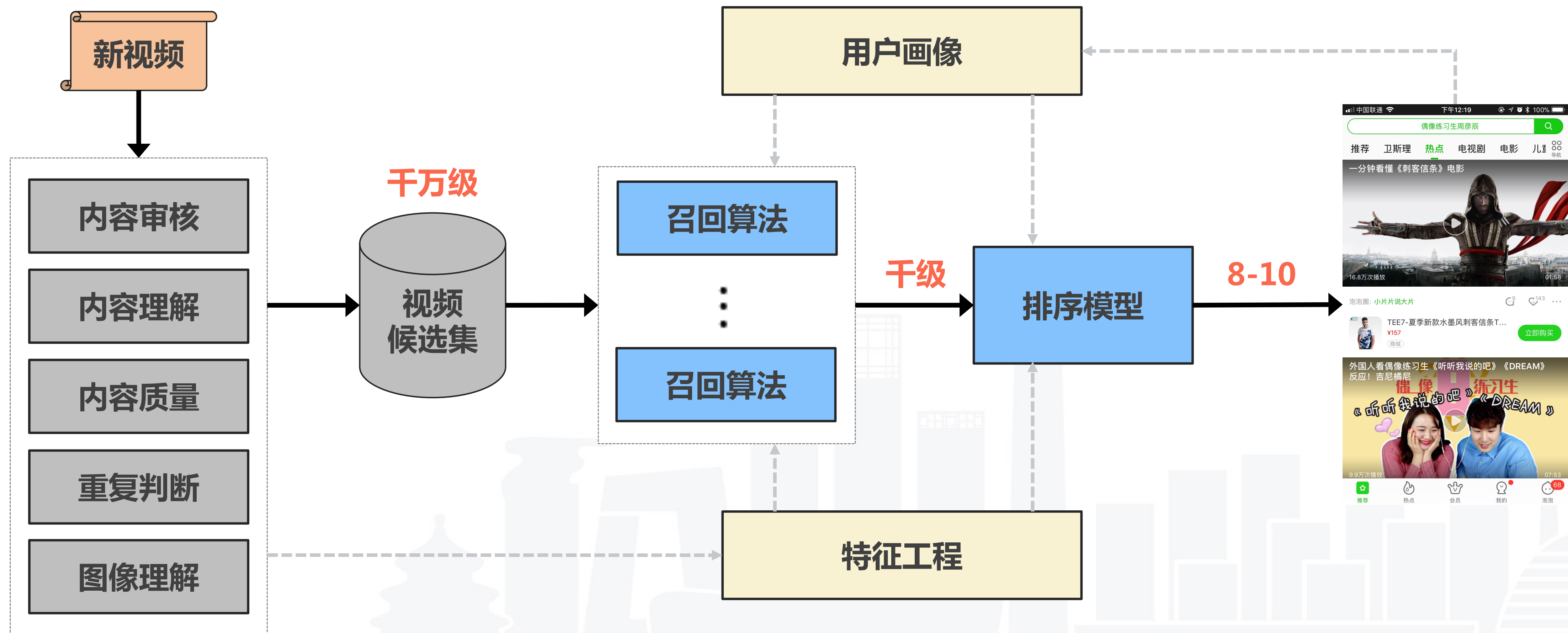


# Agenda

- 爱奇艺信息流推荐简介
- **深度学习设计思路和技术架构**
  - 深度召回模型
  - 深度排序模型
- 深度学习平台实践及讨论
- 总结



# 信息流推荐



# Agenda

- 爱奇艺信息流推荐简介
- 深度学习设计思路和技术架构
  - **深度召回模型**
  - 深度排序模型
- 深度学习平台实践及讨论
- 总结



# 召回

- 给定**上下文**，从所有推荐集合中过滤出**具备推荐价值**的内容
  - 上下文 = { User , Context }
  - Output = { Item List }
  - 推荐价值：**兴趣相关性**、热门、好友都在看、关注、LBS...

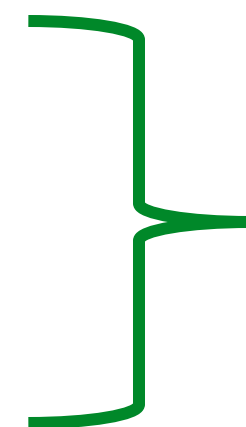
# 召回模型

Content Filtering

Collaborative Filtering

Popularity

SNS , LBS , etc.



基于兴趣相关性

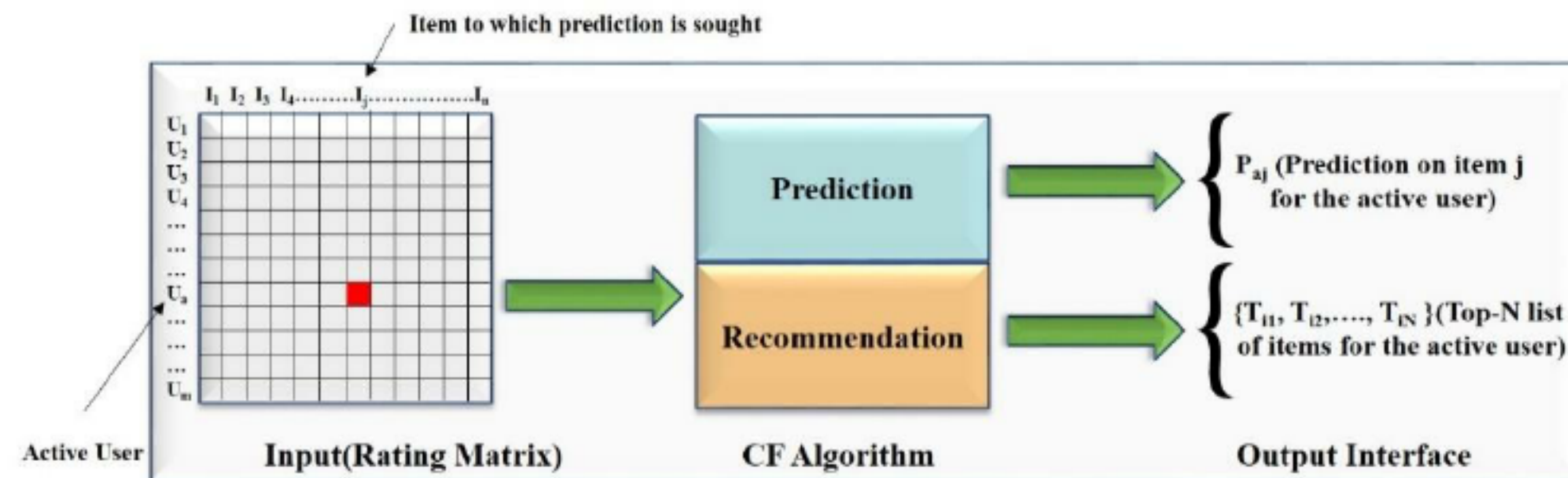
基于热度和质量

基于社交和场景

# 协同过滤

基于用户行为来 预测/推荐 的一类算法

1. 相似的视频更大概率会被同一个用户（相似的用户）看过
2. 相似兴趣的用户会更大概率看同一个视频（相似的视频）
3. 历史会重复发生





# 协同过滤

- Neighborhood based method
  - Item-based CF
  - User-based CF

## 优势



- ☐ 实现简单，不依赖内容本身
- ☐ 基本稳定可靠的结果
- ☐ 可解释性强

## 劣势



- ☐ 对稀疏性非常敏感
- ☐ 没有考虑行为顺序和上下文变化
- ☐ 难以加入其它特征
- ☐ 高维空间建模，难以和其他模型结合

# Item-based CF

目标：计算物品和物品的相似度

- 1. 构建物品**向量**
- 2. 计算相似度

亿维空间

	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	.....
	1	0	1	0	1	0	1	1	0	
	1	0	1	1	1	0	1	1	1	
	0	1	0	1	0	0	0	1	1	

# 协同过滤

- Model based method
  - Matrix Factorization
    - SVD & ALS
  - Bayesian Model
  - Classification & Clustering
  - Embedding
    - **Item2Vec**
    - **Neural CF**
  - Deep Learning
    - RBM
    - **MLP**
    - **LSTM**
    - **Attention**



降维



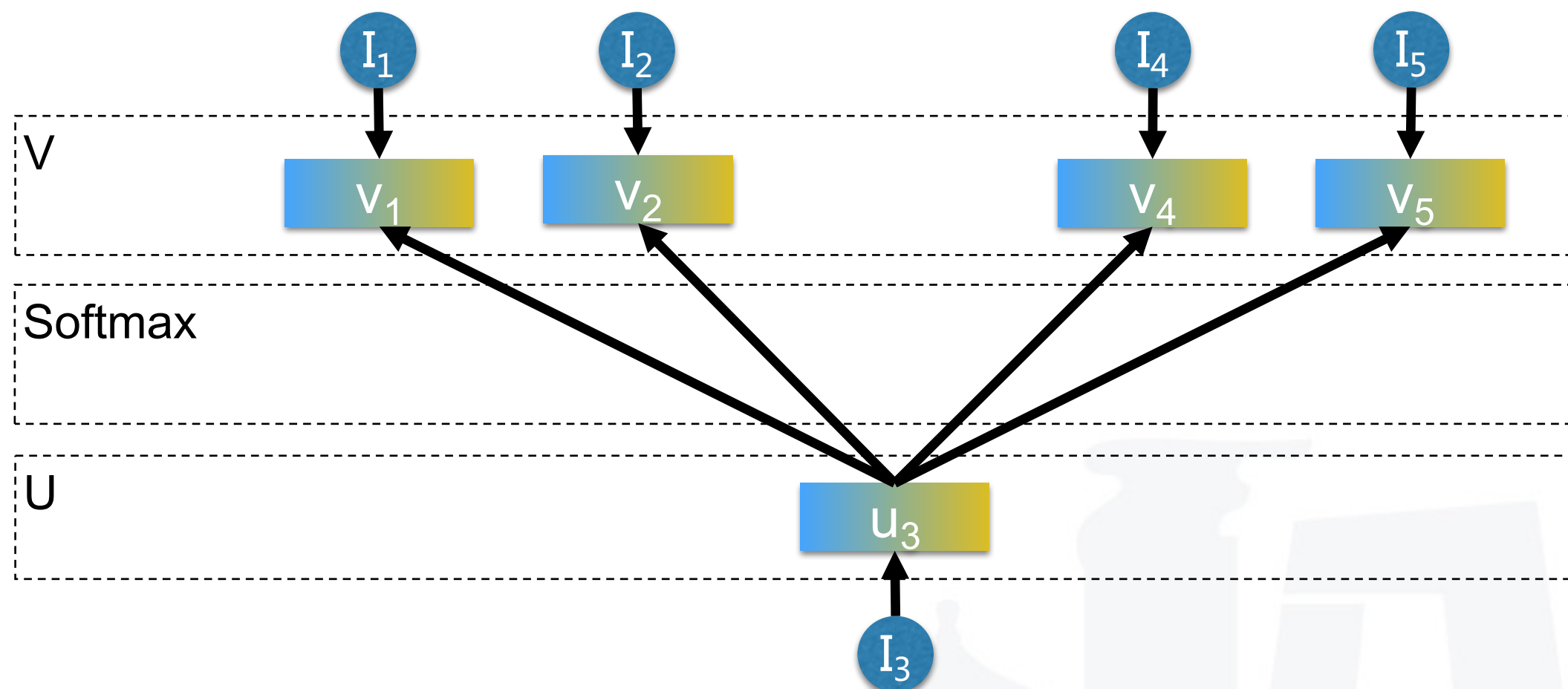
计算相似度



# Item2Vec

Barkan O, Koenigstein N. Item2Vec: Neural Item Embedding for Collaborative Filtering[J]. 2016:1-6.

$$P(I_j|I_i) = \sigma(u_i^T v_j) \prod_{k=i}^N \sigma(-u_i^T v_k) \quad \longrightarrow \quad \begin{array}{l} \mathbf{U} : \text{表征向量} \\ \mathbf{V} : \text{预测向量} \end{array}$$



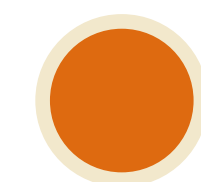
降维

任意Item都能  
得到U和V两个  
向量，维度空  
间64



分布式

多样本并行训  
练，独立更新，  
训练速度快



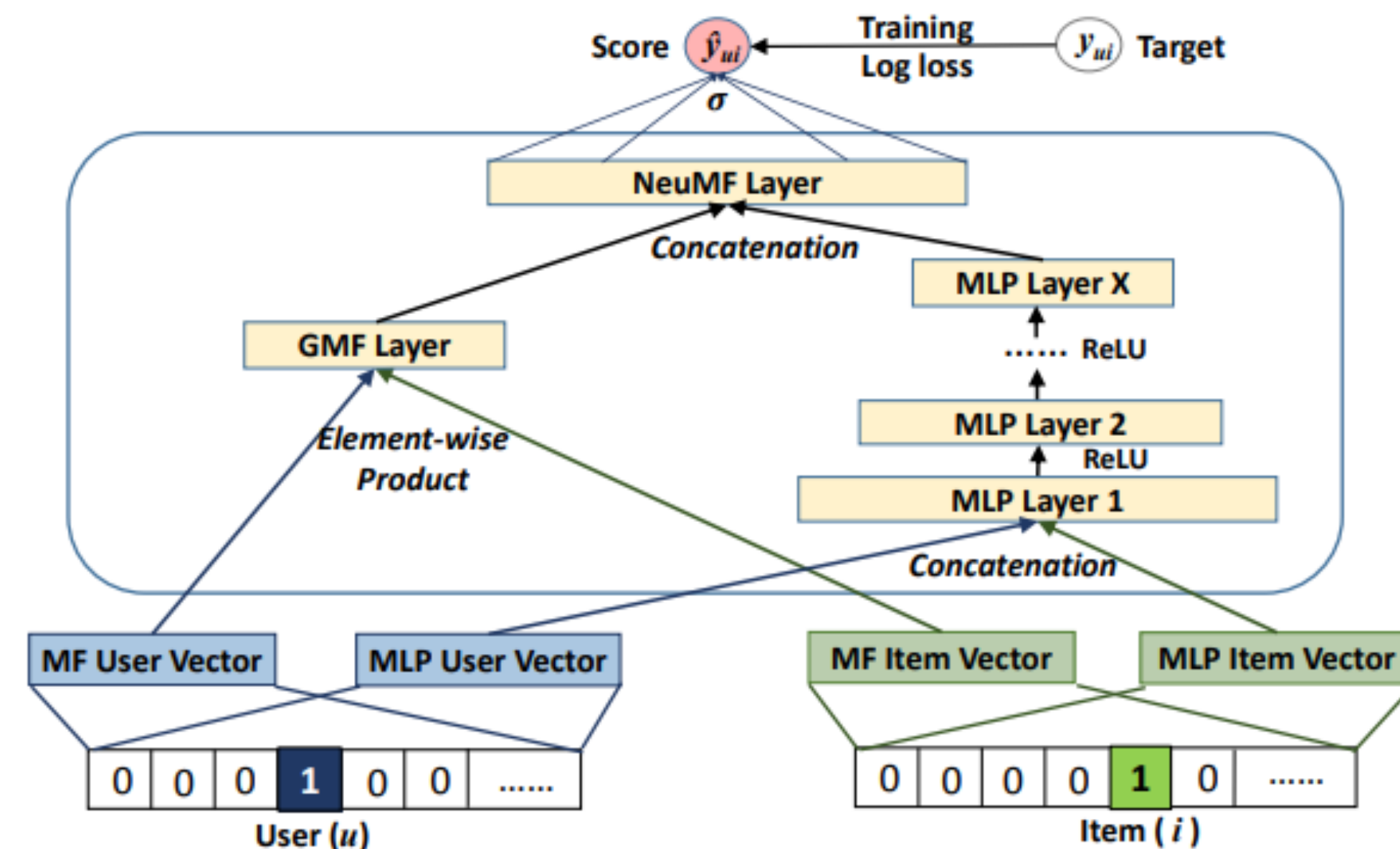
负采样

随机抽取负样  
本

# Neural CF

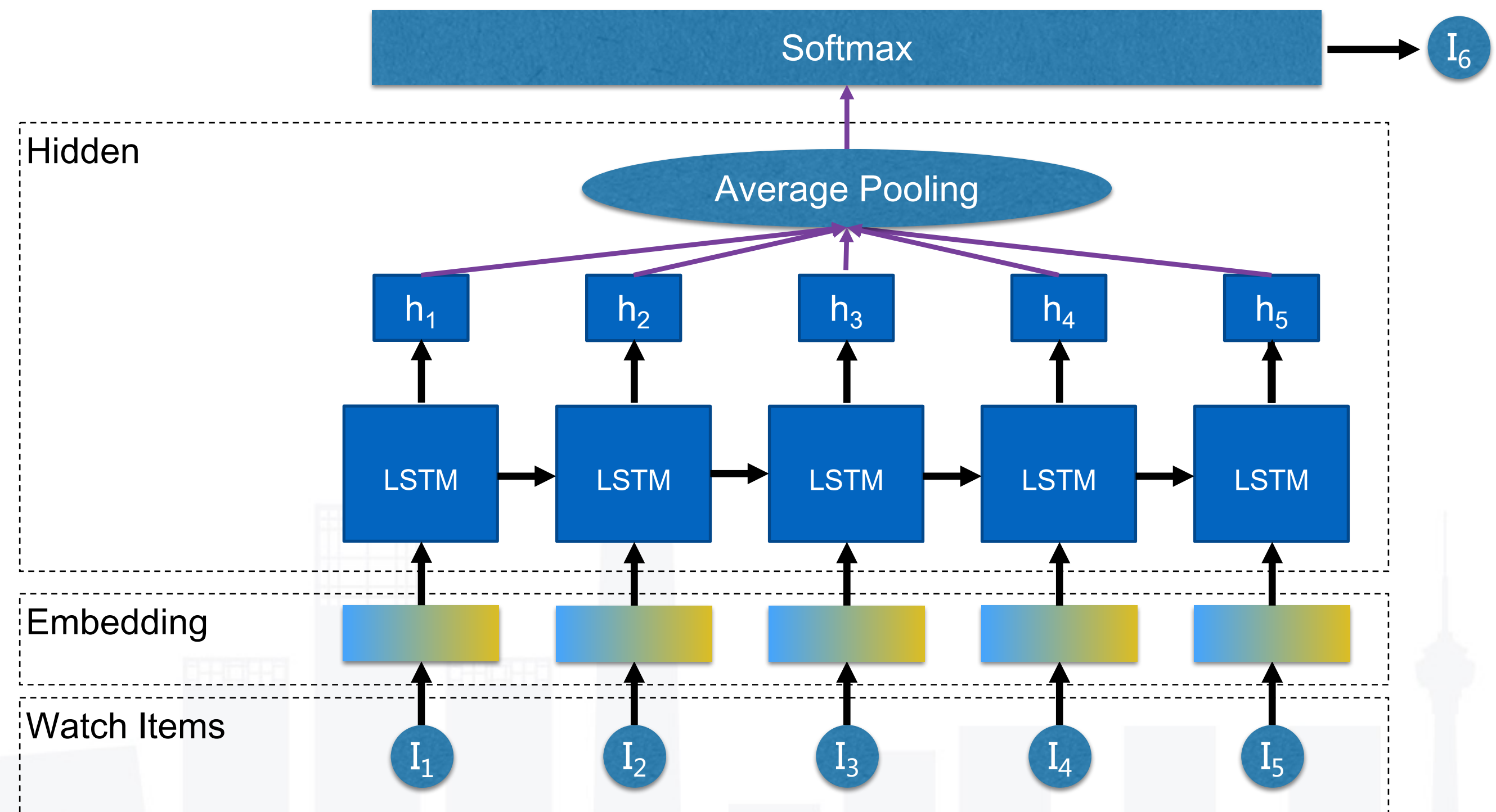
- Xiangnan He , Lizi Liao , Hanwang Zhang , Liqiang Nie , Xia Hu , Tat-Seng Chua, **Neural Collaborative Filtering**, Proceedings of the 26th International Conference on World Wide Web, April 03-07, 2017, Perth, Australia

- 基于MLP的协同过滤
  - 增加网络深度
- 问题
  - 缺乏时序性考虑
  - 没有其他的特征信息



# LSTM

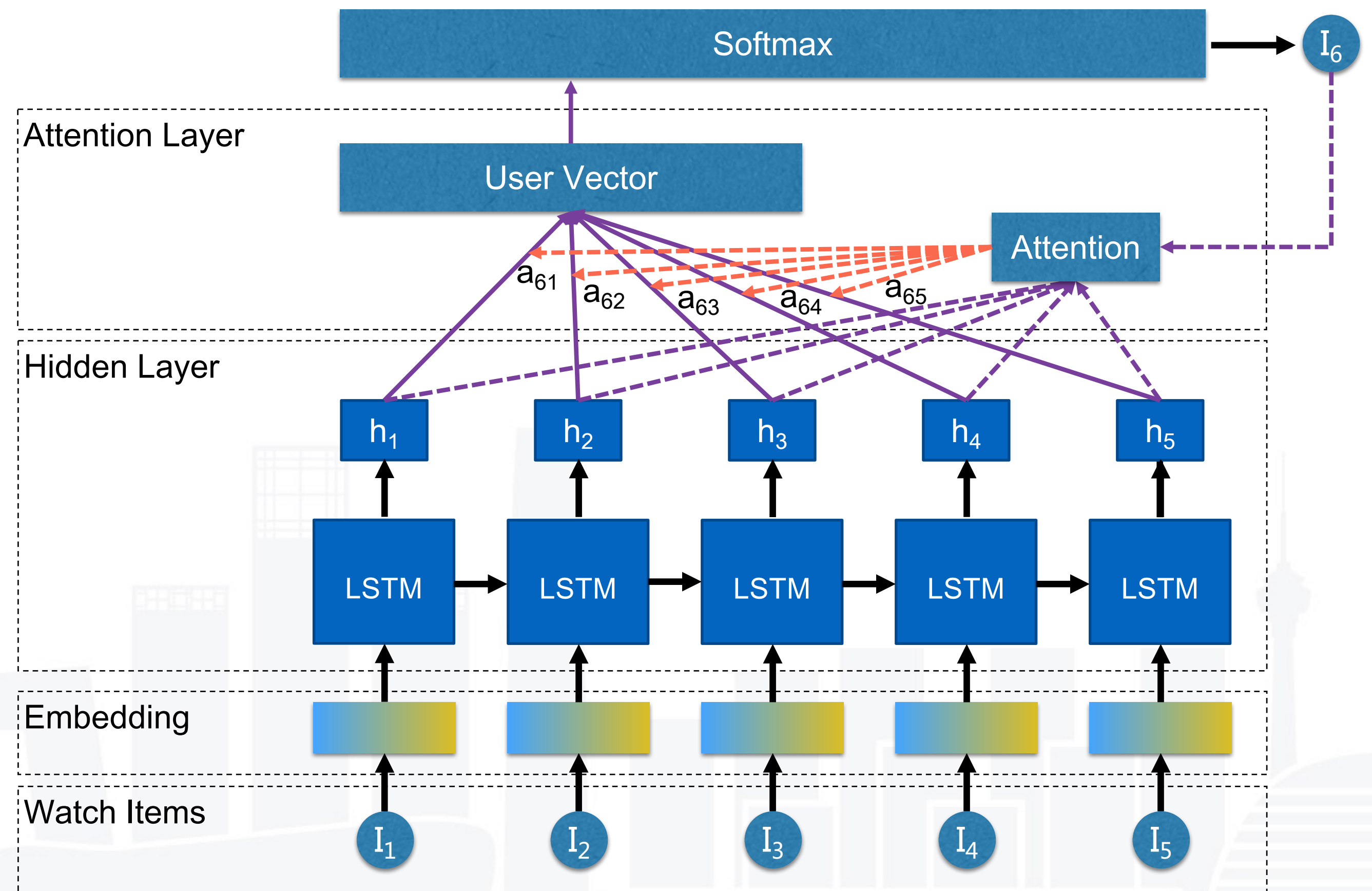
- 基于LSTM的协同过滤
  - 加入时序性考虑
- 问题
  - 缺乏对历史项重要性的学习
  - 没有其他的特征信息



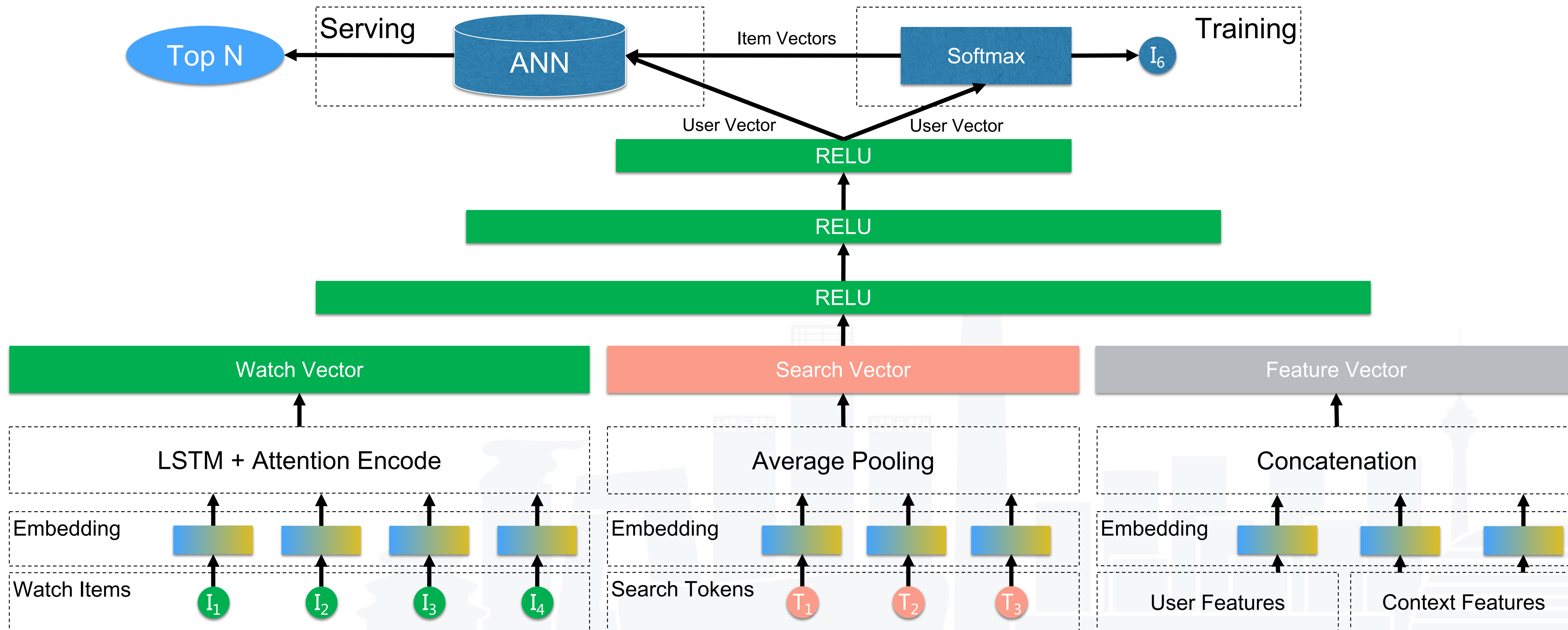


# LSTM + Attention

- 基于LSTM+Attention的协同过滤
  - 加入时序性考虑
  - 加入对历史项Attention的学习
- 问题
  - 没有其他的特征信息



# 深度召回模型



# Agenda

- 爱奇艺信息流推荐简介
- 深度学习设计思路和技术架构
  - 深度召回模型
  - **深度排序模型**
- 深度学习平台实践及讨论
- 总结

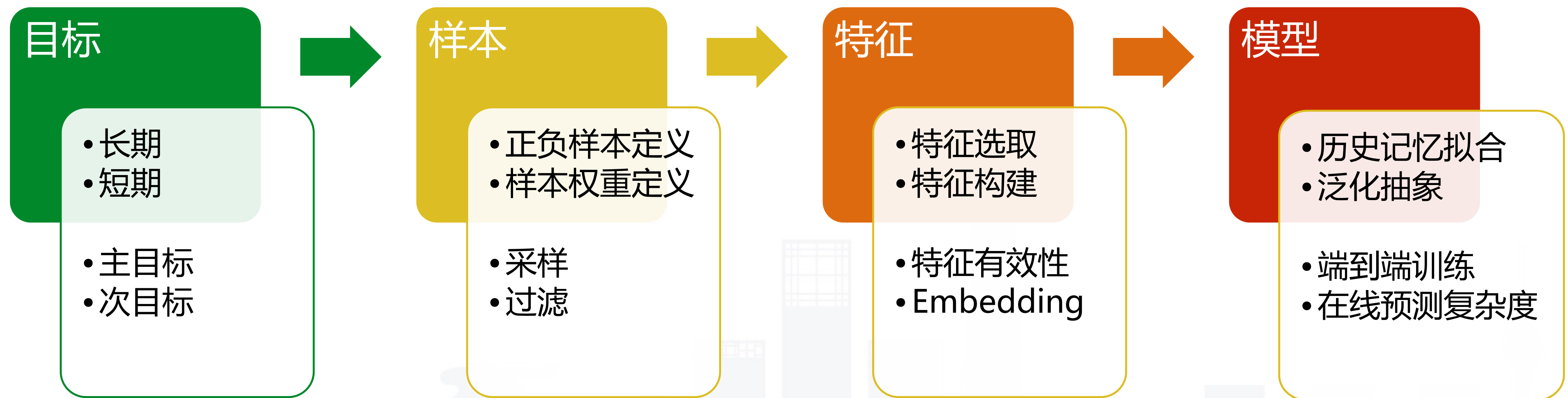


# 排序模型

- 从召回算法选取的内容集合中，找到最符合**推荐预定目标**的K个结果
  - Pointwise
  - Pairwise
  - Listwise

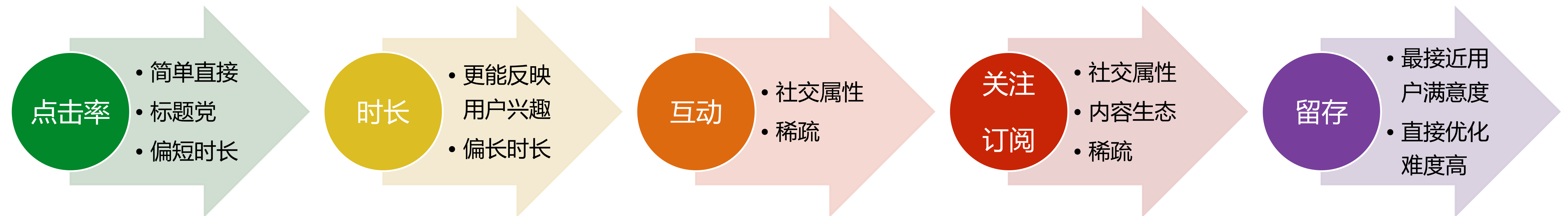
# 挑战

- 排序模型四要素



# 目标

最大化：



约束：



# 样本

根据目标仔细设计样本：

- 正样本：点击、点赞、评论、转发、关注、留存...
- 负样本：展示不点击、不喜欢、负面评论...
- 样本权重：播放时长、反馈类型、样本有效性...

仔细设计样本有效性

- 无效样本判断过滤



# 特征

## 稠密特征

- 点击率
- 播放量
- 播放时长
- 年龄、性别

## 稀疏特征

- 视频 ID
- 视频标签
- 用户标签
- 订阅作者 ID
- LBS

## Embedding特征

- 标题 Embedding
- Item2Vec Embedding
- 图片 Embedding
- 社交 Embedding
- End2End Embedding

# 相关模型

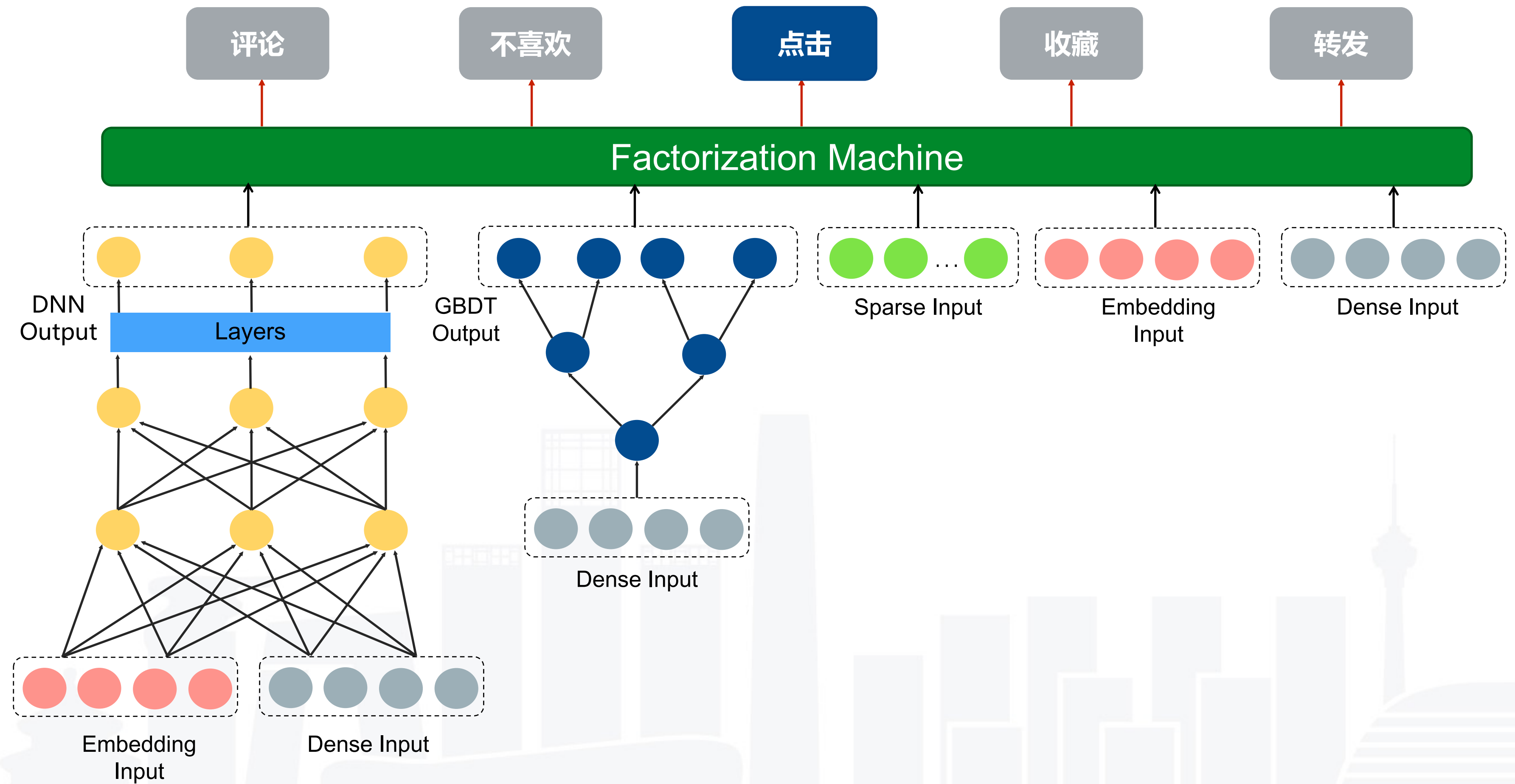
- LR
- FM
- GBDT
- GBDT + FM ( LR )
- DNN + FM
  - DeepFM: A Factorization-Machine based Neural Network for CTR Prediction

# 深度排序模型

## Ensemble DNN

1. DNN
2. GBDT
3. FM

- Multi Objective
- Wide And Deep
- Embedding



# Agenda

- 爱奇艺信息流推荐简介
- 深度学习设计思路和技术架构
  - 深度召回模型
  - 深度排序模型
- **深度学习平台实践及讨论**
- 总结



# 问题

## 背景：

- **目标、样本、特征**：业务导向
- **模型**：最优化问题求解

## 出发点：

- **业务优化**和**模型优化**解耦

## 目标：

- **服务化**：提高业务算法开发效率
- **模块化**：提高模型升级效率

# 排序服务平台

## 训练平台搭建

- 搭建训练集群
- 训练任务调度

## 模型训练服务

- 训练过程模块化
- 训练过程配置化
- 交叉验证工具

## 可视化工具

- 训练过程可视化
- 训练结果分析工具
- 特征可视化分析
- 优化建议

## 在线监控和报警

- 模型正确性校验
- 模型AUC监控报警

# 排序服务

- 业务算法：

- 目标，特征，样本
- 超参数调优

- 平台算法：

- 模型开发，可视化
- 模型模块通用开发

```
java -cp ${dnn_service} com.qiyi.qiyu.rankingservice.gear.DnnServiceClient \  
-c ${config_path} \  
-a ${dnn_train_path} \  
-b ${dnn_test_path} \  
-t ${cur_dt} \  
-n ${resource_num} \  
-o ${dnn_output_path}
```

[dnn]	
batch_size=100	每个batch的样本数量
max_steps=30000	最大迭代次数
log_frequency=100	日志打印频率
issync=0	更新参数方式，0代表异步，1代表同步
regularizer_type=2	正则化方式，2代表L2正则，1代表L1正则
regularization_rate=0.01	正则化系数
moving_average_decay=0.99	滑动平均衰减值
learning_rate_base=0.1	基础学习率
learning_rate_decay=0.1	学习率衰减值
feature_length=784	稠密特征长度
num_classes=10	label种类

# Agenda

- 爱奇艺信息流推荐简介
- 深度学习设计思路和技术架构
  - 深度召回模型
  - 深度排序模型
- 深度学习平台实践及讨论
- **总结**



# 总结

- 深度学习在推荐中的应用
  - 与传统模型自然融合
  - 具备更强的表达能力
    - MLP
    - LSTM
    - Attention
- 端到端训练和预测
  - Embedding



关注QCon微信公众号，  
获得更多干货！

# Thanks!



INTERNATIONAL SOFTWARE DEVELOPMENT CONFERENCE

主办方 **Geekbang**  **InfoQ**  
极客邦科技