



推荐系统

Recommender System

陈震

清华大学基础工业训练中心

机器学习

推荐系统

推荐系统

推荐系统通过分析大量用户的消费和购物行为，找出相似性的用户和物品，通过预测技术向用户推荐新的用户和物品。

定义

➤分析用户的历史行为，挖掘出兴趣爱好，预测其未来感兴趣的内容

优势

➤不需要提供搜索的关键词



推荐方法

- 定义（维基百科）

- 利用某兴趣相投、拥有共同经验之群体的喜好来推荐使用者感兴趣的资讯

- 商业系统



电子商务推荐

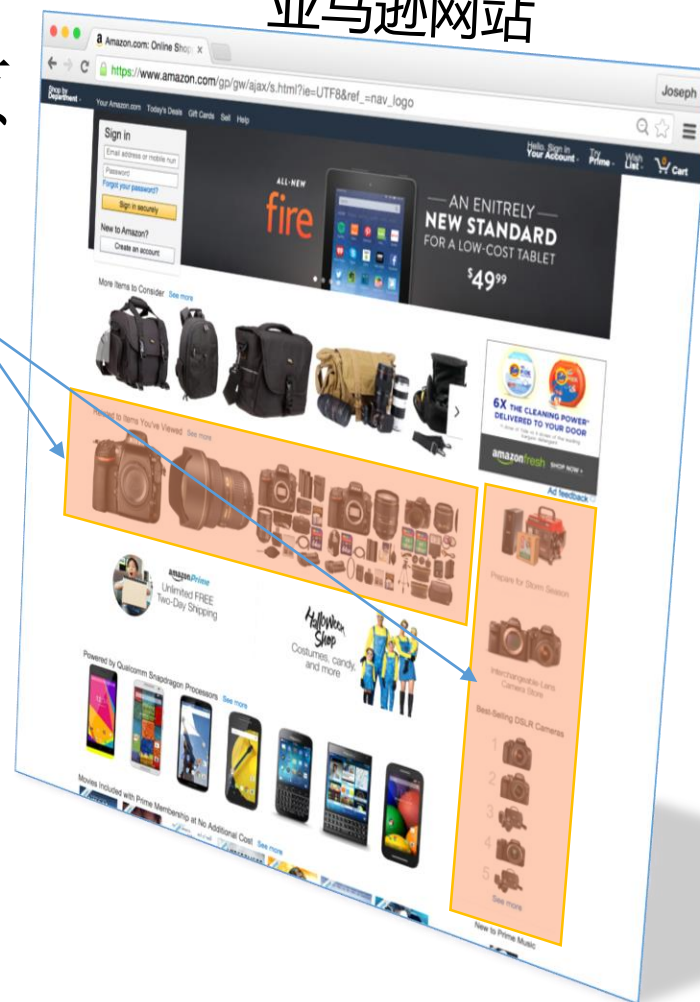
(1) 当用户浏览照相机网页

...

(2) 相近的照相机和配件

反馈回路：基于算法自动生成推荐物品

亚马逊网站



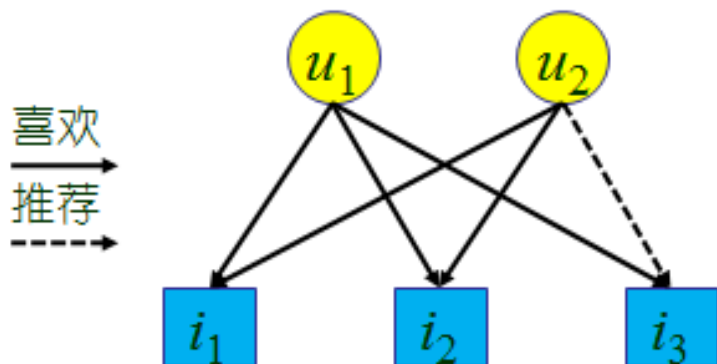
协同过滤推荐方法

推荐算法研究是电子商务重点。典型推荐算法有：

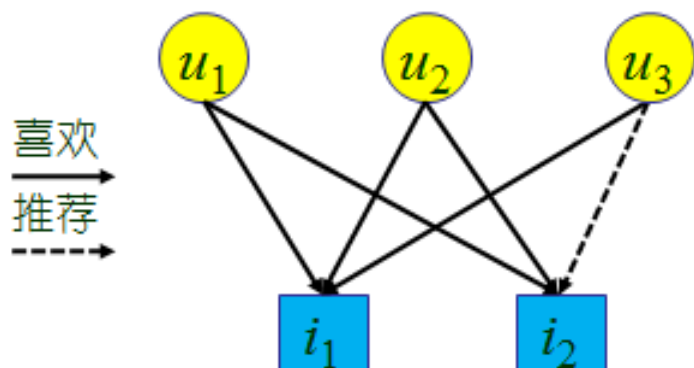
- (1) 基于协同过滤的推荐。
- (2) 基于内容的推荐；
- (3) 基于人口统计学推荐；

推荐算法分为非个性化和个性化推荐的方法。
协同过滤是其中重要的一种方法。

基于用户的协同过滤

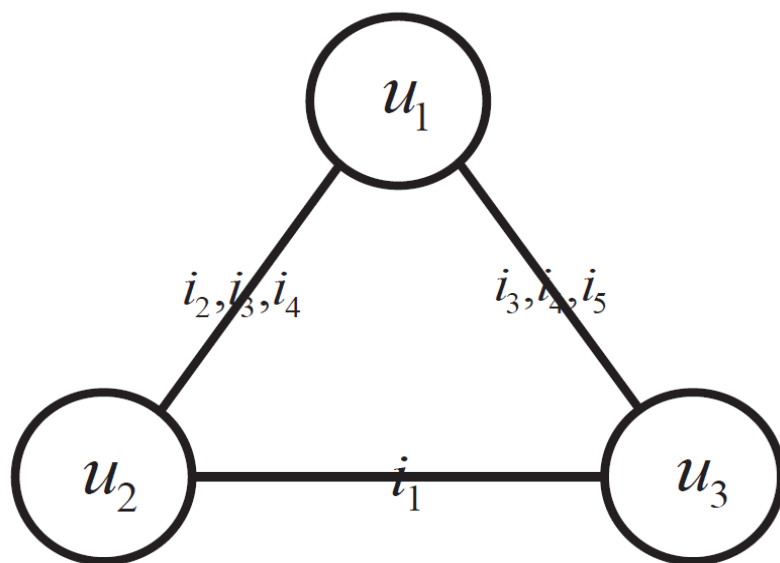


基于项目的协同过滤



基于相似性传递的协同过滤方法

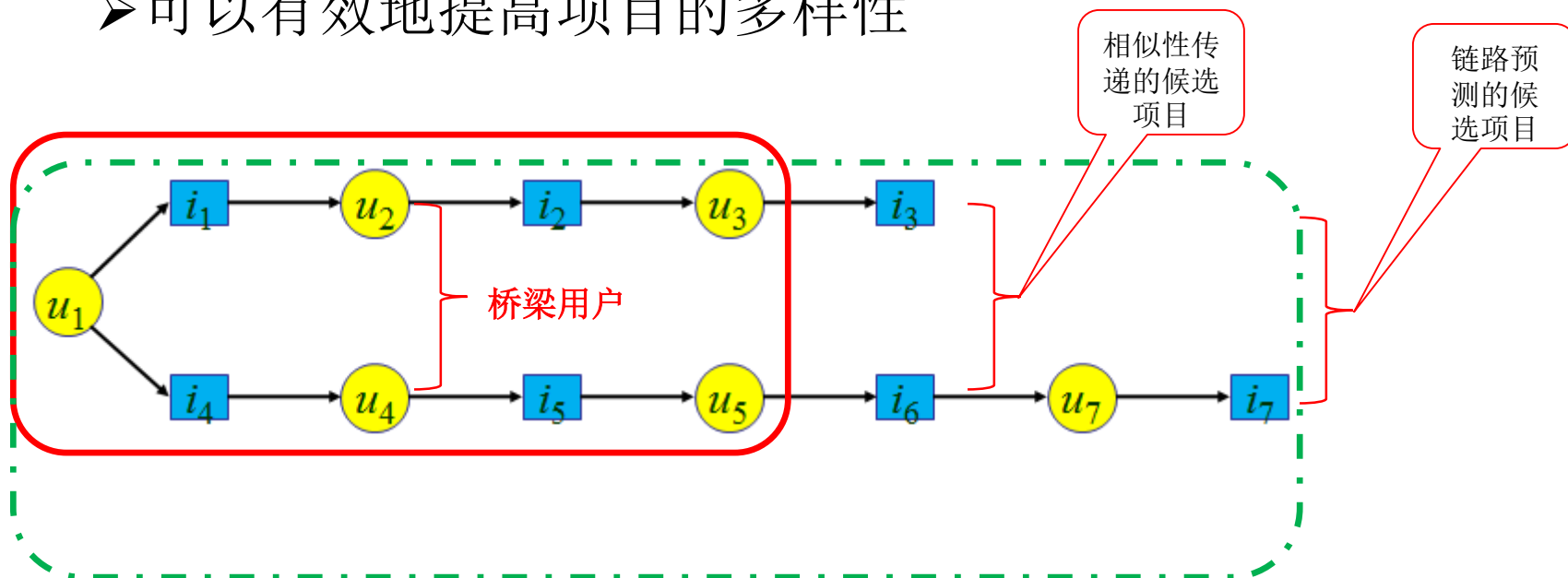
- 相似性传递（Threshold based Similarity Transitivity, TST）
 - 设定阈值避免不准确的相似性度量
 - 利用相似性传递性增加相似性数量
 - 实现相似性质量和数量的双赢



基于链路预测的协同过滤方法

- 推荐问题是链路预测的子问题

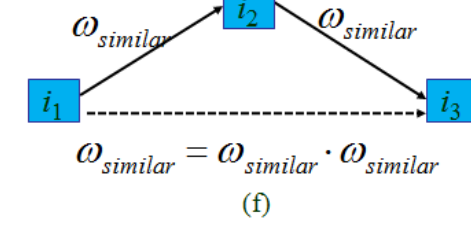
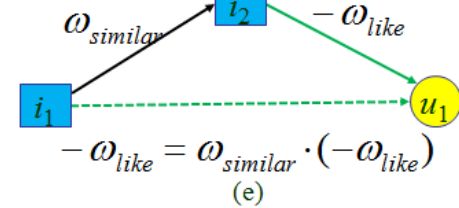
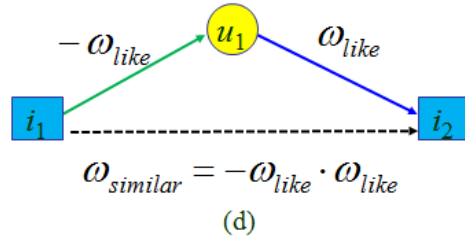
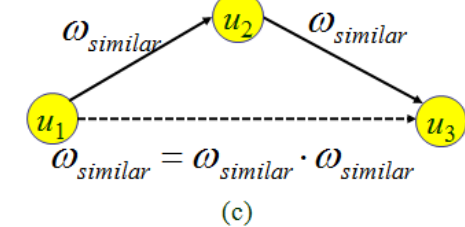
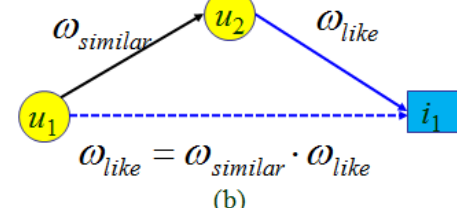
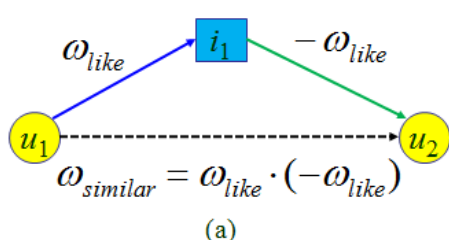
- 相似性传递方法的进一步拓展
- 本质上仍属于协同过滤的范畴
- 可以有效地提高项目的多样性



基于链路预测的协同过滤方法

• 三角关系

- 用复数表示用户与项目之间边的权重
- 用实数表示用户与用户、项目与项目之间边的权重



$$\begin{aligned}
 \omega_{\text{similar}} &= -\omega_{\text{like}}^2 \\
 \omega_{\text{like}} &= \omega_{\text{similar}} \times \omega_{\text{like}} \\
 \omega_{\text{similar}} &= \omega_{\text{similar}}^2
 \end{aligned}
 \xrightarrow[\omega_{\text{like}} = j]{\omega_{\text{similar}} = 1}
 \begin{aligned}
 1 &= -j^2 \\
 j &= 1 \times j \\
 1 &= 1^2
 \end{aligned}$$

基于灰色预测模型的协同过滤方法

- 基于灰色预测模型的方法

- 利用项目相似性生成原始序列
- 根据原始序列构建灰色预测模型
- 利用灰色预测模型得到预测结果

- 灰色预测模型的优势

- 有效缓解数据稀疏性
- 充分利用数据关联性

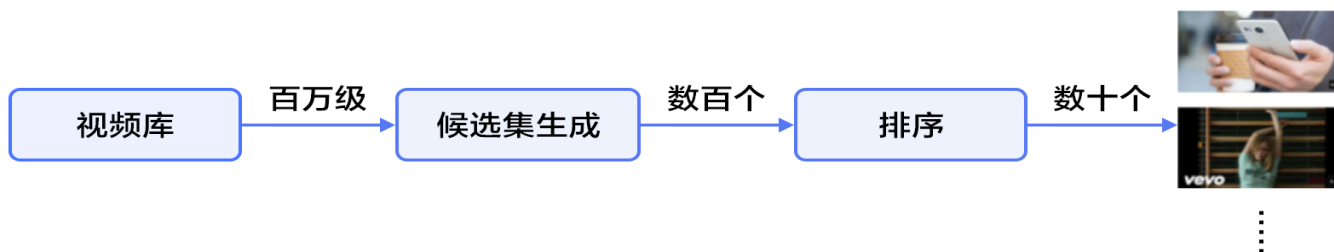
应用-内容推荐

- 研究思路：通过大数据技术，寻找用户历史与感兴趣内容之间的相关性
- 研究影响1：爱奇艺视频推荐系统，覆盖其近4千万用户
- 研究影响2：应用汇AppChina移动应用推荐，覆盖其上百万用户



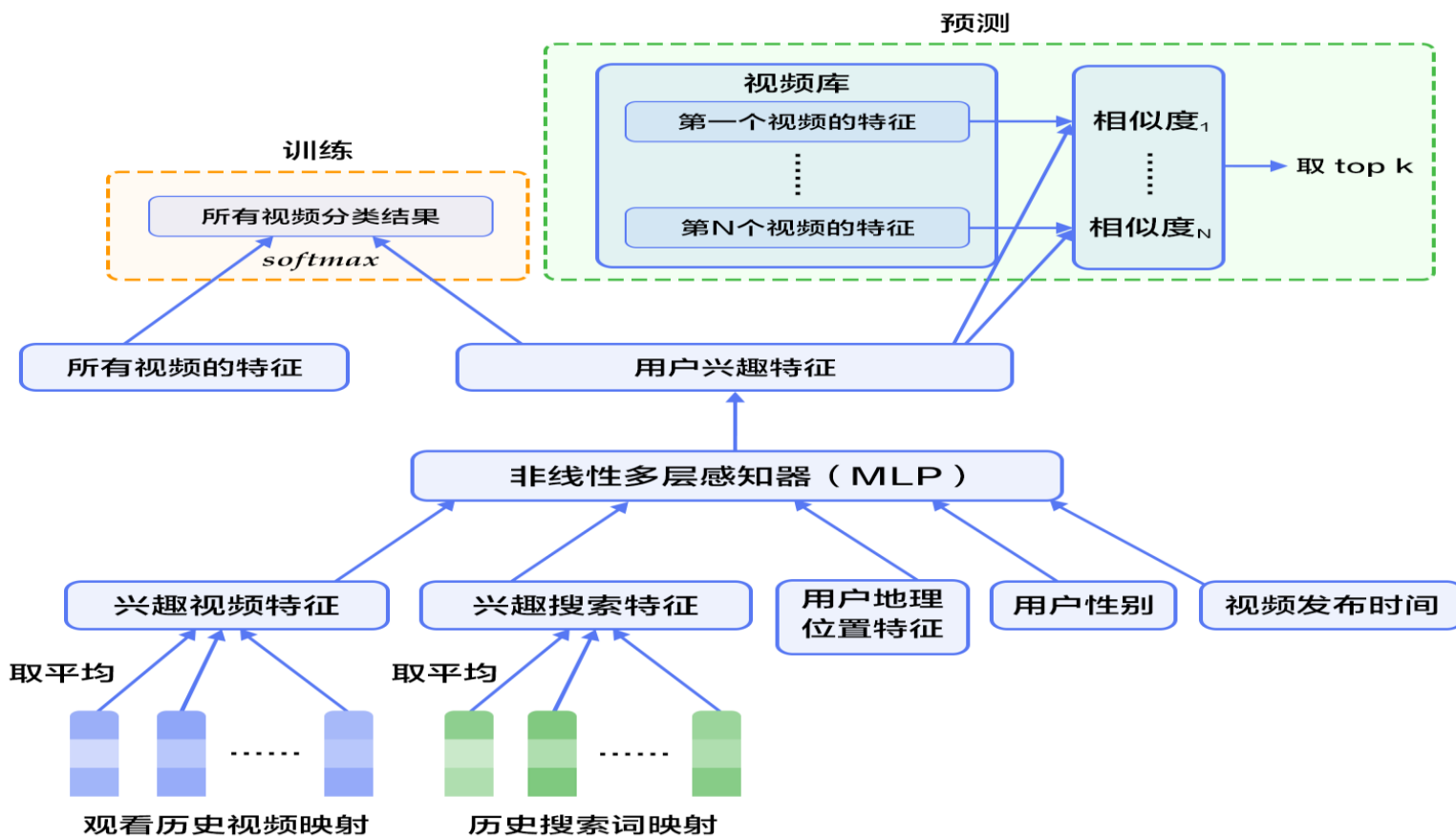
基于深度学习的推荐系统（YouTube）

- 整个系统由两个神经网络组成：候选生成网络和排序网络。
- 候选生成网络从百万量级的视频库中生成上百个候选。
- 排序网络对候选进行打分排序，输出排名最高的数十个结果。



[x] Covington P, Adams J, Sargin E. Deep neural networks for youtube recommendations, ACM Recsys, 2016.

候选生成网络（Candidate Generation Network）



候选生成网络（Candidate Generation Network）

- 候选生成网络将推荐问题建模为一个类别数极大的多类分类问题：
- 对于一个Youtube用户，使用其观看历史（视频ID）、搜索词记录（search tokens）、人口学信息（如地理位置、用户登录设备）、二值特征（如性别，是否登录）和连续特征（如用户年龄）等
- 对视频库中所有视频进行多分类，得到每一类别的分类结果（即每一个视频的推荐概率），最终输出概率较高的几百个视频

候选生成网络（Candidate Generation Network）

- 首先，将观看历史及搜索词记录这类历史信息，映射为向量后取平均值得到定长表示；
- 同时，输入人口学特征以优化新用户的推荐效果，并将二值特征和连续特征归一化处理到 $[0, 1]$ 范围。
- 接下来，将所有特征表示拼接为一个向量，并输入给非线性多层感知器（MLP）处理。
- 最后，训练时将MLP的输出给softmax做分类，预测时计算用户的综合特征（MLP的输出）与所有视频的相似度，取得分最高的 k 个作为候选生成网络的筛选结果。

候选生成网络（Candidate Generation Network）

- 对于一个用户 u ，预测此刻用户要观看的视频 ω 为视频 i 的概率公式为：

$$P(\omega = i|u) = \frac{e^{v_i u}}{\sum_{j \in V} e^{v_j u}}$$

- 其中 u 为用户 U 的特征表示， V 为视频库集合， v_i 为视频库中第 i 个视频的特征表示。 u 和 v_i 为长度相等的向量，两者点积可以通过全连接层实现。
- 考虑到softmax分类的类别数非常多，为了保证一定的计算效率：
 - 1) 训练阶段，使用负样本类别采样将实际计算的类别数缩小至数千；
 - 2) 推荐（预测）阶段，忽略softmax的归一化计算（不影响结果），将类别打分问题简化为点积（dot product）空间中的最近邻（nearest neighbor）搜索问题，取与 u 最近的 k 个视频作为生成的候选。

排序网络

- 排序网络的结构类似于候选生成网络，但是它的目标是对候选进行更细致的打分排序。
- 和传统广告排序中的特征抽取方法类似，这里也构造了大量的用于视频排序的相关特征（如视频 ID、上次观看时间等）。
- 这些特征的处理方式和候选生成网络类似，不同之处是排序网络的顶部是一个加权逻辑回归（**weighted logistic regression**），它对所有候选视频进行打分，从高到底排序后将分数较高的一些视频返回给用户。

谢谢！