

一种基于用户播放行为序列的个性化视频推荐策略

王娜 何晓明 刘志强 王文君 李霞

(深圳大学电子与信息工程学院 广东 深圳 518060)

(深圳大学智能信息处理广东省重点实验室 广东 深圳 518060)

摘要 本文针对在线视频服务网站的个性化推荐问题,提出了一种基于用户播放行为序列的个性化推荐策略.该策略通过深度神经网络词向量模型分析用户播放视频行为数据,将视频映射成等维度的特征向量,提取视频的语义特征.聚类用户播放历史视频的特征向量,建模用户兴趣分布矩阵.结合用户兴趣偏好和用户观看历史序列生成推荐列表.在大规模的视频服务系统中进行了离线实验,相比随机算法、基于物品的协同过滤和基于用户的协同过滤传统推荐策略,本方法在用户观看视频的 Top-N 推荐精确率方面平均分别获得 22.3%、30.7% 和 934% 的相对提升,在召回率指标上分别获得 52.8%、41% 和 1065% 的相对提升.进一步地与矩阵分解算法 SVD++、基于双向 LSTM 模型和注意力机制的 Bi-LSTM+Attention 算法和基于用户行为序列的深度兴趣网络 DIN 比较, Top-N 推荐精确率和召回率也得到了明显提升.该推荐策略不仅获得了较高的精确率和召回率,还尝试解决传统推荐面临大规模工业数据集时的数据要求严苛、数据稀疏和数据噪声等问题.

关键词 词向量;协同过滤;深度神经网络;个性化推荐

中图法分类号 TP393 **DOI号** 10.11897/SP.J.1016.2020.00123

Personalized Video Recommendation Strategy Based on User's Playback Behavior Sequence

WANG Na HE Xiao-Ming LIU Zhi-Qiang WANG Wen-Jun LI Xia

(College of Electronic and Information Engineering, Shenzhen University, Shenzhen, Guangdong 518060)

(Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen, Guangdong 518060)

Abstract Due to the spread of online web services about videos such as YouTube and Tencent, the number of videos posted to them is increasing. We can now watch a large amount of videos with ease, however, explosive videos available frequently overwhelmed users, leading them to make poor decisions. Recommendation systems are designed to predict the future preferences of users' based on their previous interactions with the items. For example, a user's previous viewing information can be used to make recommendations on his/her future viewing video. The vast amount of information produced by the users can be used by different methods for recommendations, with either neighborhood based methods, or machine-learning based methods or matrix-factorization based methods. These all use low-rank approximation of input data, however, may suffer from data sparsity or noise in data. Recently, word embedding methods in deep learning are successfully used to learn linguistic regularities and semantic information from large text datasets. It has made a big innovation in natural language processing and text mining fields. They can learn low-dimensional vector space representation of input elements with effect, which lays the foundation of the recommendation research in online video services. Actually, user's viewing history is one of the

important factors reflecting the user's attention and also it has good scalability, high accuracy and flexibility comparing to rating data and reviews written by users. In this work, we aim to recommend next videos by adopting word embedding techniques proposed in Word2Vec framework via analyzing user's playback behavior. Unlike the previous works that use Word2Vec for recommendation, a non-textual feature namely the past viewing videos of the users, is used to make recommendations. By representing each user (according to her/his history viewing behavior) with a high dimensional vector space, we can filter the target user potential interest in candidate video, thus a personalized recommendation strategy based on the user's playback behavior sequence for online video service websites is proposed. This strategy maps the video into feature vectors and extracts the semantic features of the video via user's video playing behavior. Then, the user's interest distribution matrix is modeled by clustering the feature vectors of the user's video playing history. A recommendation list is generated in combination with user interest preferences and user viewing history aging. Offline experiments were conducted in a large-scale video service system. Compared with random algorithms, item-based collaborative filtering recommendations, and user-based collaborative filtering recommendations, it improves the average accuracy of Top-N recommendation for users watching videos 22.3%, 30.7% and 93.4% respectively. The relative increase of 52.8%, 41% and 106.5% respectively in the recall rate indicator is achieved. Moreover, compared with the matrix decomposition algorithm SVD++, the model based on the bidirectional LSTM+Attention and the deep interest network DIN, the proposed method achieves different levels of improvement in both the Top-N recommendation accuracy rate and the recall rate. The recommendation strategy not only obtains better performance as a whole, but also attempts to solve the problems of data requirements, data sparsity and data noise faced by traditional recommendation algorithms.

Keywords word vector; collaborative filtering; deep neural network; personal recommendation

1 引 言

随着社会迈入移动互联时代,在线视频网站的用户规模和视频资源与日俱增,解决由此引发的用户信息过载问题势在必行,途径之一是个性化推荐系统.有效的推荐系统不仅能提高服务网站的用户忠诚度,增加用户流量和提升用户粘性,还可显著提高在线视频服务网站的收益.YouTube、NetFlix、腾讯视频和优酷视频等国内外在线视频网站每年花费大量资源来研究这一领域.根据视频服务网站中的视频资源不同可分为职业生产内容 OGC(Occupationally Generated Content)、专业生产内容 PGC(Professionally Generated Content)和用户生产内容 UGC(User Generated Content)三大类.根据其类型不同,在线视频网站采用的推荐模型也不相同.其中,PGC 拥有导演、介绍、主演和分类等属性信息,更适合采用基于内容属性信息的推荐模型,与之

相反,UGC 和 OGC 缺少属性信息,更多采用协同过滤模型^[1].当在线视频服务网站同时具有 PGC、OGC 和 UGC 等多种视频资源时,通常采用混合模型进行推荐,如 YouTube 采用基于内容推荐和基于协同过滤推荐的混合模型^[2].与 NetFlix 等 PGC 在线视频服务网站相比,含有 UGC 的视频服务网站除了视频本身外极少含有附加信息,其用户交互行为信息更加紊乱,因此含 UGC 在线服务网站的视频推荐也面临更大的噪声.

实际上,无论是具有哪种视频资源的在线视频服务网站,推荐策略可利用的数据信息主要有:(1) 用户行为数据,其中包括显式反馈和隐式反馈;(2) 视频标题、视频描述等视频元数据.由于用户显式反馈比隐式反馈更加容易获得并且简单易用,所以目前很多推荐算法都是基于用户显式反馈进行推荐.Koren 等人^[3]在用户-视频评分矩阵基础上,采用奇异值分解^[4]模型和交替最小二乘法将评分矩阵分解为一个用户-类别矩阵和一个类别-视频矩阵,

根据拟合每个用户对视频的预测评分进行推荐。这类基于矩阵分解的推荐算法必须维护物品相似度矩阵,但是计算相似度矩阵的复杂度随着物品数量的增多而增大,如果物品数量更新极快,维护相似度矩阵需要花费极大的代价。同时,由于用户评分数据极度稀疏和计算过程时间复杂度太高,其应用场合也大大受限。利用视频元数据的推荐算法,典型的如 Choi 等人^[5]尝试仅靠电影的分类信息进行个性化推荐,先计算得到所有电影类别间的相关性矩阵,然后根据用户的行为历史计算用户对所有分类的偏好程度,最后利用用户的类别偏好、目标电影的类别组合和平均评分预测用户对目标电影的评分,根据预测评分高低做出个性化推荐。Elkahky 等人^[6]通过分析用户的浏览记录和搜索记录来丰富用户的特征表示,并提出了一种多视角深度神经网络模型,该模型试图通过用户和物品两种信息实体的语义匹配来实现用户的物品推荐。Covington 等人^[7]通过利用用户信息、情景信息、历史行为数据和物品特征信息等多源异构数据,提出了一种深度神经网络模型用于 YouTube 视频推荐。Sedhain 等人^[8]提出了一种基于自编码器的协同过滤算法,该模型通过一个编码过程和一个解码过程产生输出。这类算法需要假设电影的分类信息是由若干类别组合而成的,并且算法难以捕捉对个性化推荐至关重要的用户兴趣变化。

由于词向量模型^[9]可以利用文本文档中的单词顺序,捕捉大型语料库中单词之间的语义关系,在自然语言处理任务中取得了良好的效果,近年来常被用于推荐系统当中。Barkan 等人^[10]根据用户网上听音乐和购买微软应用商店的商品的记录,使用基于负采样的 Skip-gram 算法(Skip-Gram with Negative Sampling, SGSN)建模用户行为序列,验证了 SGSN 算法对用户的行为序列建模的有效性。Ozsoy^[11]利用词向量模型将用户的历史签到行为嵌入到低维空间中,采用基于内容的推荐方法和基于用户的协同过滤方法,产生推荐列表。Grbovic 等人^[12]根据用户是否点击了电子邮件客户端中的广告,利用词向量算法建立 Prod2Vec 模型,依此可计算余弦相似度产生前 K 个最相似的商品,也可将商品聚类,向用户推荐与其先前购买的产品集群相关的产品。段琦^[13]利用词向量模型,获得餐厅特征向量,再通过预测用户打分,获得用户特征向量。以此计算餐厅特征向量相似度和用户特征向量相似度,选择与当前用户去过的餐厅相似度高的餐厅与当前用户相似的用户喜欢的餐厅作为推荐结果。以上四种算法本质

上都是基于相似度推荐,其推荐准确率较大地依赖于用户的历史记录,无法刻画用户的个性化兴趣。Vasile 等人^[14]提出了 Meta-Prod2Vec 模型,作为 Prod2Vec 模型的拓展,在 Prod2Vec 中融入了项目元数据作为辅助信息。同样,汪景福等人^[15]利用深度神经网络(Deep Neural Network, DNN)分析视频的媒资信息(视频简介、用户评论和标题等),通过词向量模型^[9]获取视频的隐话题特征,结合用户对视频的点击和不点击行为,构造用户的正负行为特征向量,筛选过滤对所推荐视频感兴趣的用户群。但这两种模型都使用了大量的物品元数据信息,因此仅限于元数据丰富的场景。

在线视频网站可利用的信息具有多模态、数据异构、大规模、数据稀疏和分布不均匀等复杂特征,对推荐策略而言,看似数据信息丰富,实则存在着很强的局限性:(1)传统的协同过滤推荐算法是浅层模型,难以提取用户和物品的深层次特征;(2)基于内容的推荐算法,主要利用人工设计的用户特征和物品特征,这种人工设计特征需要大量的领域知识和专家经验;(3)用户提供的显式反馈远远小于隐式反馈,利用用户的显式反馈进行推荐,应用场景受限。实际上,对任何一个在线视频服务网站,最具利用价值和最容易获得的就是用户历史播放数据,但需要有效的特征提取模型。本文基于用户历史播放序列和自然语言序列之间的相似性,通过自然语言处理中新颖的词向量模型,提出了一种基于深度语义模型的视频个性化推荐算法。该算法只需用户视频播放行为数据,无需使用用户和物品特征,省去了人工特征设计的过程。该算法不再局限于用户-物品评分矩阵和物品属性信息,不存在数据稀疏影响推荐效果的问题,对数据噪声也具有天然的鲁棒性。该算法仅仅依靠用户的行为记录历史序列即可完成视频向量特征提取,并且通过聚类算法,刻画用户的个性化兴趣,建模用户的兴趣矩阵,算法效率明显高于目前居于主流的神经网络模型,并且更适合大规模工业数据应用场景。

2 词向量模型

Mikolov 等人^[9]于 2013 提出了基于深度网络的词向量模型 Word2Vec 成功应用于大规模文本语料词向量训练。Word2Vec 包括 Continuous Bag-Of-Words(CBOW)和 Skip-gram 两种网络结构,每种结构又有 Hierarchical Softmax(HS)和 Negative

Sampling (NEG) 两种训练方法. 由于负采样 NEG 的计算量比 HS 要小得多且没有复杂的哈夫曼树环节, 所以基于负采样的词向量模型在计算速度上远胜于基于 HS 的词向量模型. 本文利用基于负采样的词向量模型提取用户观看的视频序列特征. NEG 词向量模型如下:

已知中心词 w_i 和上下文 $context(w_i)$, 其中 $context(w_i) = \{w_{i-k}, \dots, w_{i-1}, w_{i+1}, \dots, w_{i+k}\}$, k 为窗口大小. 采用负采样技术^[16] 构建负样本, 即对于给定的 $context(w_i)$, 词 w_i 本身是唯一一个正样本, 那么其它词就是负样本, 如图 1 所示.

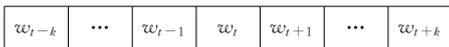


图 1 语言样本序列

根据词语在语料库中的词频对给定的 $context(w_i)$ 进行负采样, 负采样样本记为 $Neg(w)$, $\{w\}$ 为正样本集合, 样本 L^w 为中心词 w 对应的样本标签, 定义如下:

$$L^w(u) = \begin{cases} 1, & u \in \{w\} \\ 0, & u \in Neg(w) \end{cases} \quad (1)$$

对于给定的训练样本 ($context(w), w, Neg(w)$), 利用极大似然估计, 最大化词向量模型目标 $g(w)$

$$g(w) = \prod_{u \in \{w\} \cup Neg(w)} p(u | context(w)) \quad (2)$$

其中, $p(u | context(w))$ 表示当上下文为 $context(w)$ 时, 中心词为 u 的概率. $g(w)$ 表示当中心词为正样本且不为负样本的似然概率. 在训练开始的时候, 将语料库中每个词 u 表示为一个 m 维的向量 θ^u , 为上下文 $context(w)$ 中词向量的累加, 令

$$p(u | context(w)) = \begin{cases} \sigma(\mathbf{X}_w^T \theta^u), & L^w(u) = 1 \\ 1 - \sigma(\mathbf{X}_w^T \theta^u), & L^w(u) = 0 \end{cases} \quad (3)$$

其中, $\sigma(x)$ 为 Sigmoid 函数. 当 $L^w(u) = 1$ 时, $\sigma(\mathbf{X}_w^T \theta^u)$ 表示中心词是 w 的概率, 当 $L^w(u) = 0$ 时, $1 - \sigma(\mathbf{X}_w^T \theta^u)$ 表示中心词不是 w 的概率. 将 $p(u | context(w))$ 代入 $g(w)$ 可得

$$g(w) = \prod_{u \in \{w\} \cup Neg(w)} [\sigma(\mathbf{X}_w^T \theta^u)]^{L^w(u)} \cdot [1 - \sigma(\mathbf{X}_w^T \theta^u)]^{1 - L^w(u)} \quad (4)$$

那么, 对于给定的语料库 C , 基于负采样的连续词袋模型的优化目标 G 为

$$G = \prod_{w \in C} g(w) \quad (5)$$

为方便计算, 对 G 取对数, 最终的目标函数可表示为

$$\begin{aligned} l = \log G &= \log \prod_{w \in C} g(w) = \sum_{w \in C} \log g(w) \\ &= \sum_{w \in C} \sum_{u \in \{w\} \cup Neg(w)} \{L^w(u) \log[\sigma(\mathbf{X}_w^T \theta^u)] + \\ &\quad [1 - L^w(u)] \cdot \log[1 - \sigma(\mathbf{X}_w^T \theta^u)]\} \quad (6) \end{aligned}$$

在模型中利用随机梯度上升法优化目标函数, 可以选择合适的词向量维度、上下文窗口大小和学习速率等参数进行模型调优^[17].

实际上, 在在线视频网站中的用户历史视频播放序列和自然语言之间有很强的类比性, 用户视频播放行为具有连续性的特点, 用户相继观看的视频之间具有很强的关联性, 如用户某个时间点观看了时事新闻, 那么这个用户接下来更有可能倾向于播放时事新闻类的视频; 用户在看了电视某集剧集, 接下来很大概率会播放同一剧的下一集. 因此本文将用户播放序列比作自然语言处理中的词序列, 利用词向量模型对用户视频播放序列建模, 提取视频深层次的语义特征, 从而生成视频特征向量, 并提出了一种基于词向量视频特征的推荐策略.

3 基于词向量的视频推荐策略

在在线视频网站中, 用户和视频繁杂, 数据庞大, 但通常存在噪声大、计算复杂、数据不完整、数据稀疏和数据隐私等问题, 对形成有效的推荐策略是一大挑战. 本文试图只利用在线视频网站中最普遍、噪声小也最易获取的视频用户播放行为数据, 利用词向量模型分析并处理用户播放视频的历史记录, 构建视频特征向量, 提取视频的语义特征, 并对用户播放历史中的视频特征向量聚类, 建模用户兴趣分布矩阵, 结合用户的兴趣模型和用户观看历史时效性生成推荐列表, 如下:

(1) 分析并处理用户的播放日志数据, 构建视频词向量训练模型, 生成视频特征向量, 计算视频相似度;

(2) 聚类视频特征向量, 结合用户的播放历史构建用户的兴趣模型, 得到用户的兴趣分布矩阵;

(3) 结合用户的播放历史和兴趣分布矩阵, 产生高效的个性化推荐.

3.1 视频特征向量生成

如图 2 所示, 每一条用户的播放历史日志包括视频的唯一标识(video_id)、用户的唯一标识(user_id)、播放时间(timestamp)和播放时长(duration), 视频播放序列按照用户播放的时间先后顺序从日志中抽取 video_id 生成.

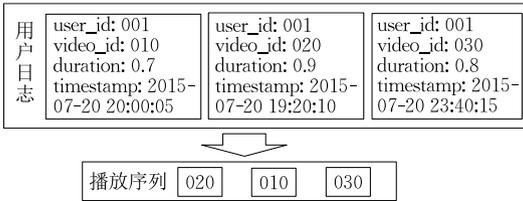


图 2 用户 001 日志对应的播放序列

每一个用户的播放序列都会被类比成自然语言中的词序列, 类比 Word2Vec 训练词序列生成词特征向量, 本文利用 Word2Vec 训练所有用户的播放序列得到视频特征向量. 训练数据集包含腾讯视频提供的约 3000 万用户在 22 天里产生的 100 多亿次播放日志. 经过上述数据清洗后, 训练数据仅仅由视频 ID 组成, 且过滤部分播放量低的冷门视频后约包括 1500 万部视频. 训练数据中的视频类型包括电影、电视剧、ACG、综艺、新闻、音乐、体育、生活等. 视频向量训练过程如图 3 所示.

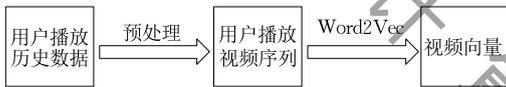


图 3 视频向量模型的训练过程

一般来说, 视频向量的表征效果与视频向量特征维度是正相关关系, 但是若想获得高维度的视频向量, 必须增加更多的用户播放历史数据, 并且训练时间也会明显增加. 本文在 12 核 3.3 GHz CPU、32 GB 内存和 Windows10 操作系统下完成 Word2Vec 训练, 生成视频向量特征. 整个训练过程耗时与视频向量维度和上下文窗口成正比. 当视频向量维度设置为 200, 上下文窗口设置为 5 时, 训练耗时约为 6 小时. 模型训练采用自适应学习速率.

3.2 视频相似度计算

生成视频特征向量后, 就可以计算视频之间的相似度. 相似度计算主要有欧式距离和余弦距离方法, 余弦距离是用向量空间中两个向量夹角的余弦作为个体差异的度量, 主要体现方向上的差异性. 欧式距离衡量的是空间中各个点的绝对距离, 与点的位置密切相关. 考虑到视频向量具有高维的特点, 本文使用余弦相似度来衡量视频之间的相似度. 计算方法如下:

$$\omega_{ij} = \frac{\sum_{x=1}^m (\mathbf{V}_i^x \times \mathbf{V}_j^x)}{\sqrt{\sum_{x=1}^m (\mathbf{V}_i^x)^2} \times \sqrt{\sum_{x=1}^m (\mathbf{V}_j^x)^2}} \quad (7)$$

其中, ω_{ij} 为视频 i 和视频 j 的相似度, \mathbf{V}_i 和 \mathbf{V}_j 分别

是视频 i 和视频 j 对应的视频向量, 视频 i 对应的视频向量 $\mathbf{V}_i = \{V_i^1, V_i^2, \dots, V_i^m\}$, m 为视频向量维度.

3.3 用户兴趣建模

视频推荐系统的核心思想是向用户推荐其感兴趣的视频. 面对规模庞大的工业数据应用场景, 直接从庞大的视频库中找到用户大概率感兴趣的视频是很耗费资源和时间的, 因此需先产生候选集. 如美团在推荐商品时就是先利用协同过滤的方法获得向量的相似性产生候选集, 但只利用相似度推荐会忽略用户兴趣偏好. 在 YouTube 推荐系统的研究上, Davidson 等人^[2]和 Covington 等人^[7]采用了对候选集排序的方法建模用户兴趣. 本文的推荐策略是首先利用视频特征向量计算相似度产生视频候选集, 称之为粗排序. 再利用 K-Means 算法聚类用户观看历史中所有视频向量特征, 建模用户的兴趣偏好, 产生用户兴趣分布矩阵, 进而对视频候选集进行精排序, 产生推荐.

3.3.1 视频聚类

用户的兴趣偏好可由其历史播放记录体现, 且用户播放历史行为越丰富, 表现出的视频类别偏好信息越明显. 如某用户历史播放记录中 A 类视频占比显著大于其它类, 那么用户很大概率对 A 类视频具有较大兴趣. 原本训练数据集中的视频有十五个大类的原始分类信息(电影、新闻、娱乐、体育等), 但这种分类信息颗粒度过大且不准确. 如一个视频归类为新闻类, 但该视频内容又具有娱乐属性, 此时分类信息就不具有太高的参考价值. 因此使用原始视频分类信息直接建模用户兴趣显然不可靠. 本文在获取的视频特征向量基础上利用 K-means 聚类, 通过指定簇的数量得到视频分类信息的更细的粒度划分. 实验仿真(详见 4.3.5 节实验)也证实了视频分类的颗粒度大小对推荐算法具有明显的影响, 且视频类属划分提高了推荐算法的准确性.

K-means 是机器学习中一种简单有效、使用非常广泛的聚类算法, K-means 其时间复杂度接近于线性, 尤其适合大数据集. 本文用局部密度和距离来确定聚类最佳簇数^[18], 即把每个样本点 i 用密度 ρ_i 和距离 δ_i 来表示

$$\rho_i = \sum_{j \in I_S} e^{-\left(\frac{d_{ij}}{d_c}\right)} \quad (8)$$

I_S 为样本集合, d_c 为截断距离, d_{ij} 为样本 i 和样本 j 的距离. 样本 i 的局部密度 ρ_i 的大小与其截断距离 d_c 内样本的数量有关.

$$\delta_i = \begin{cases} \min_{j \in I_S^i} \{d_{ij}\}, & I_S^i \neq \emptyset \\ \max_{j \in I_S^i} \{d_{ij}\}, & I_S^i = \emptyset \end{cases} \quad (9)$$

$$I_S^i = \{k \in I_S : \rho_k > \rho_i\} \quad (10)$$

I_S^i 为局部密度 ρ 比样本 i 的局部密度更大的所有样本集合, I_S 表示所有点的集合. 对于非局部密度最大点, δ_i 为 I_S^i 中和样本 i 距离最小的样本之间的欧氏距离; 对于局部密度最大点, δ_i 为 I_S 中和样本 i 距离最大的样本之间的欧氏距离.

决策图中标记所有样本 (ρ_i, δ_i) , 最优聚类中心数量为决策图中的离群点样本个数, 记为 K . 再利用 K -means 算法把视频库中所有对应的视频向量 V 聚为 K 类, 类属中心为 $C = \{c_1, c_2, c_3, \dots, c_K\}$.

3.3.2 用户兴趣分布

设用户 u 在时间窗 T 内的播放历史为 $H_u^T = \{V_1, V_2, V_3, \dots, V_m\}$, 用户 u 对类属 c 的兴趣偏好记为 $I(u, c)$, 那么用户视频类属偏好计算方法如下:

$$I(u, c_i) = \frac{1}{n} \sum_{V_x \in H_u^T} e^{\frac{-\|V_x - c_i\|^2}{\sigma}}, \quad c_i \in C \quad (11)$$

其中, $C = \{c_1, c_2, c_3, \dots, c_K\}$ 为所有视频的 K 个聚类中心, n 为用户 u 在时间窗 T 播放历史中视频总数, 目的是为了平衡用户视频播放量对类别偏好的差异, σ 为待定超参数. 根据上式分别计算, 就可得到用户 u 在时间窗 T 内对所有类属 C 的兴趣分布为 $I_u = \{I(u, c_1), I(u, c_2), \dots, I(u, c_K)\}$.

设用户集合 $U = \{u_1, u_2, u_3, \dots, u_i\}$, 则所有用户的兴趣分布矩阵建立如表 1.

表 1 用户兴趣分布矩阵

| U | C | | | |
|----------|---------------|---------------|----------|---------------|
| | c_1 | c_2 | \dots | c_K |
| u_1 | $I(u_1, c_1)$ | $I(u_1, c_2)$ | \dots | $I(u_1, c_K)$ |
| u_2 | $I(u_2, c_1)$ | $I(u_2, c_2)$ | \dots | $I(u_2, c_K)$ |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| u_i | $I(u_i, c_1)$ | $I(u_i, c_2)$ | \dots | $I(u_i, c_K)$ |

3.4 生成推荐列表

视频推荐根据其和用户播放历史中的视频相似度以及用户对视频类属的兴趣偏好进行计算, 公式如下:

$$p_{uj} = \frac{I(u, c_x)}{\sum_{j \in c_x} \tau_{ij}} \quad (12)$$

其中, τ_{ij} 为视频 i 和视频 j 的相似度, $I(u, c_x)$ 表示用户 u 对 c_x 类视频的兴趣度, $N(u)$ 是用户距离推荐时间最近观看的视频集合, $S(j, K)$ 是和视频 j 最相似的 K 个视频. 按照用户对视频的兴趣度 p_{uj} 降序

排列, 取前 Top- N 生成对用户的视频推荐列表. 由式(12)可以看出, 与用户观看历史中视频越相似, 且与用户历史兴趣越吻合的视频越有可能在推荐列表中获得比较高的排名. 该推荐策略在利用用户观看历史视频相似性进行粗排序的基础上, 综合用户的兴趣, 进一步对候选视频精排序, 产生高效推荐.

图 4 给出了本文推荐策略在腾讯视频数据集上的一个完整示例. 首先, 根据用户的播放视频日志, 通过词向量模型获得视频特征向量, 它是后续视频相似度计算、用户兴趣建模的关键; 其次, 聚类视频特征向量, 实现视频细粒度分类, 计算用户兴趣分布矩阵(式(11)), 如表 2 所示. 然后, 根据用户历史视频播放记录中的 Downtown Abbey 和 Game of Thrones 两个视频作为种子视频, 通过视频相似度(式(7))找出与这两个视频最相似的视频, 其中视频 Missing 属于 c_2 类, 视频 Rome 属于 c_3 类等; 最后, 生成视频推荐列表(式(12)) $\{Rome, Spartacus, Missing, Atlantic City, Prejudice\}$. 显而易见, 视频 Rome 和视频 Downtown Abbey 的相似度为 0.6, 与视频 Game of Thrones 的相似度为 0.8, 综合起来 Rome 视频得分最高, 同时也符合用户视频类别兴趣偏好.

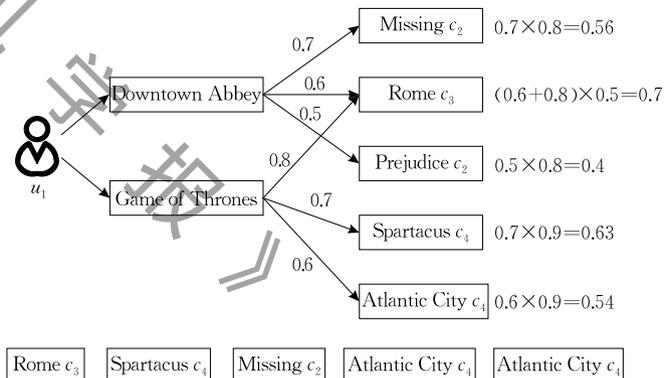


图 4 推荐策略简单示例

表 2 用户 u_1 的兴趣分布矩阵

| U | C | | | |
|-------|-------|-------|-------|-------|
| | c_1 | c_2 | c_3 | c_4 |
| u_1 | 0.4 | 0.8 | 0.5 | 0.9 |

4 实验结果分析

本文对腾讯视频的用户历史播放数据进行了离线实验, 在离线环境下不仅和传统协同过滤推荐算法 ItemCF^[19]、UserCF^[20] 等热门推荐的推荐效果进行比较, 还与矩阵分解算法 SVD++^[21]、基于双向 LSTM 模型和注意力机制的 Bi-LSTM+Attention

算法^[22]和基于用户行为序列的深度兴趣网络 DIN^[23]的推荐效果进行比较,并从推荐的准确率和多样性等多个指标对实验结果进行分析。

4.1 实验数据和实验环境

腾讯视频是聚合热播影视、综艺娱乐、体育赛事、新闻资讯等为一体的综合视频平台.截止 18 年 7 月,腾讯月均活跃用户超过 5 亿,移动端日均活跃用户达到 1.37 亿,每天可以产生 10 亿以上的播放量.腾讯视频海量的用户播放记录为本文提供了可靠的数据支撑,具有较强工程应用意义.本文使用 3000 万用户在腾讯视频里 22 天的播放记录,约有 1000 万部视频,大约 65 亿次播放记录。

4.2 评价指标

精确率(Precision)原本是用户信息检索和统计分类领域的评价指标,在推荐系统中用来评价对用户推荐的准确程度,即算法生成的推荐列表和用户实际观看历史的重合率.精确率计算公式如下:

$$Precision = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (13)$$

其中, U 表示用户全集, $R(u)$ 为算法向用户 u 的推荐列表, $T(u)$ 为用户实际观看的视频集合,精确率是用来衡量推荐算法好坏的主要标准之一。

召回率是指用户实际播放视频中有多少出现在推荐列表中,是评价推荐算法挖掘用户整体实际播放量的评价指标,其计算公式如下:

$$Recall = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|} \quad (14)$$

其中, U 表示用户全集, $R(u)$ 为算法向用户 u 作出的推荐列表, $T(u)$ 为用户实际观看的视频集合。

一般来说,精确率和召回率反映了推荐算法性能的两个方面,精确率反映了推荐的精度,召回率衡量推荐系统的查全率,即用户感兴趣的内容出现在推荐列表中的概率.精确率和召回率有可能出现矛盾的情况,即推荐系统可具有较高精确率而召回率却很低.单一的某个指标不能较为全面地评价一个推荐算法的好坏,为了平衡精确率和召回率的影响,我们引入了综合评价指标 $F1$,计算如式(15). $F1$ 值越高时说明实验结果越好。

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (15)$$

推荐多样性指的是推荐系统推荐不同类别物品的能力.由推荐列表中物品两两间的相似性来描述,

相似性越高,多样性越低.用户的推荐列表的所有物品的平均相似度计算如式(16).

$$ListSimilarity(L_u) = \frac{2 \times \sum_{i,j \in L_u, i \neq j} Similarity(i,j)}{|L_u| \times (|L_u| - 1)} \quad (16)$$

其中, $Similarity(i,j)$ 可以采用皮尔逊相关系数、向量余弦相似度、Jaccard 系数等相似度指标来表示.求推荐系统中所有用户的推荐列表的平均相似度的平均值得到总体相似度,计算公式如下:

$$Similarity = \frac{1}{n} \sum_{u \in U} ListSimilarity(L_u) \quad (17)$$

$Similarity$ 值越大,说明用户的推荐列表内的物品之间的相似度越高,多样性越低。

4.3 实验分析

基于词向量模型的视频特征提取是本文推荐策略的核心,本文分别从视频相似度和同类视频距离两方面验证了词向量模型生成视频向量特征的有效性。

4.3.1 视频特征向量验证

表 3 给出了基于词向量模型得到的 200 维视频特征向量的最相似 Top-5 视频结果图,越靠前表示越相似.从表 3 中,我们知道,无论是联系非常紧密的电视剧,还是用户上传的 UGC 类视频,对种子视频的最邻近查询的结果内容都是非常相似的,说明了本文采用的词向量模型生成的视频特征向量能够很好地对视频进行描述,能够反映视频的特征。

表 3 与种子视频最相似的 Top-5

| Seed Video | Top-5 Most Similar Videos | Seed Video | Top-5 Most Similar Videos |
|---------------|---------------------------|---------------|---------------------------|
| 新洛神_01 电视剧 | 新洛神_02 电视剧 | 矮仔多情 电影 | 临时同居 电影 |
| | 新洛神_03 电视剧 | | 吉星高照 2015 电影 |
| | 新洛神_65 电视剧 | | 全城热恋 电影 |
| | 新洛神_04 电视剧 | | 百变爱人 电影 |
| | 新洛神_05 电视剧 | | 有种你爱我 电影 |
| 恐龙先生弄丢了 动漫 | 最好的朋友 动漫 | 王者组无限 火力游戏 | 1V5 5 杀集锦 游戏 |
| | 泥坑 动漫 | | DOPA 集锦 游戏 |
| | 鸚鵡波莉 动漫 | | 炫酷五杀精彩集锦 游戏 |
| | 捉迷藏 动漫 | | lol 亡灵暴力五杀 游戏 |
| | 猪妈妈在工作 动漫 | | 男枪酷炫走位中路 游戏 |

为了更直观地展示视频特征向量的有效性,选取如表 4 所示已知标题的视频,图 5 为表 4 的视频降维后的结果,视频 b0303p52105 和 f030313oq9e 均为娱乐类的视频,q00201bqsfm 和 b0020clme80 属于同一电视剧视频,可以看出本文方法能够深度挖掘视频深层信息并完整表达,同一类的视频空间距离也相近。

表 4 视频 ID 对应的标题

| 视频 ID | 标题 |
|-------------|------------------------|
| b0303p52105 | 刘亦菲穿私服长这样 朴素靓丽接地气 娱乐 |
| f030313oq9e | 文章马伊琍携手拍戏 全程无交流貌合神离 娱乐 |
| q00201bqsfm | 金水桥边_23 电视剧 |
| b0020clme80 | 金水桥边_24 电视剧 |

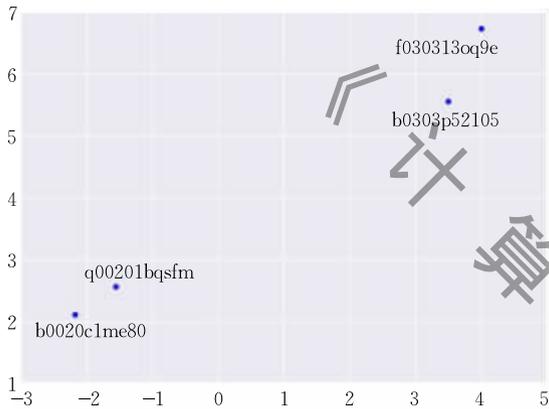


图 5 视频向量示例

进一步地,对基于词向量模型生成的视频特征向量通过 t-SNE 降维^[24]至二维空间,如图 6 所示.其中,不同颜色代表视频不同的原始分类信息.SERIES 表示电视剧类、ACG 表示动漫类、ENT 表示娱乐类、NEWS 为新闻类、FUNNY 为搞笑类等.从图 6 可以很清晰地看出,本文所提的基于词向量算法的视频特征向量建模具有较为理想的表征效果,同类视频分布得比较集中,不同类的视频距离较远,说明计算得到的视频特征向量能够体现视频的本质信息。

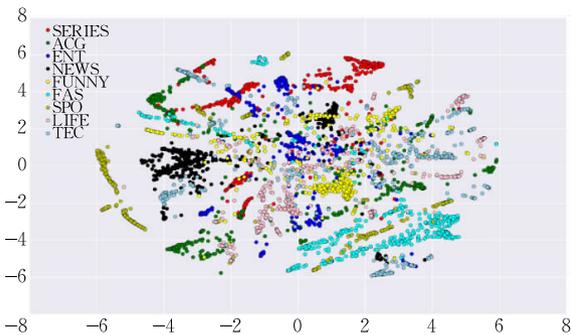


图 6 所有视频向量可视化效果图

此外,对同属一类的视频特征向量表示进行了可视化验证,以连续剧类为例,图 7 描述了 8 部不同的连续剧视频特征向量表示图.可以看出,相同连续剧的特征向量紧聚,不同连续剧的特征向量相距较远,进一步说明了算法提取的特征向量是有效的,利用此视频特征向量计算视频与视频的相似度是合理的。

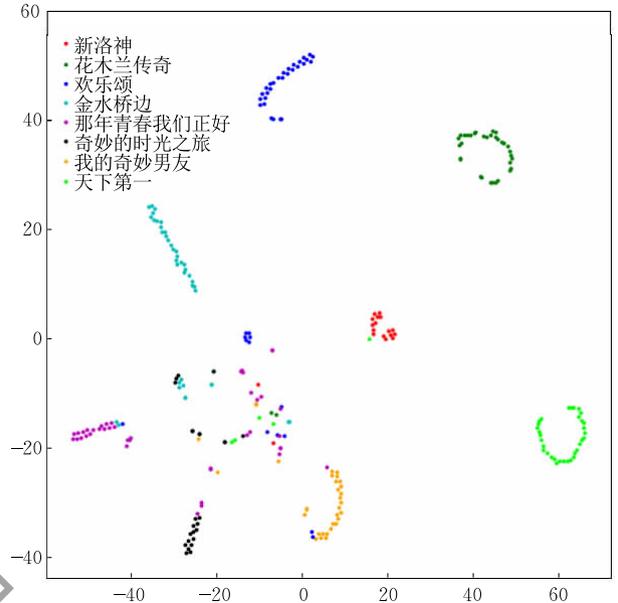


图 7 连续剧类视频特征向量可视化

4.3.2 最佳聚类簇区间和聚类簇数量 K

在建模用户观看视频的兴趣偏好中,采用了 K-means 聚类处理,对视频进行更细种类划分.本文首先对聚类簇数区间进行了实验仿真,如图 8 所示.通过 γ 值($\gamma_i = \rho_i \delta_i$)的大小对视频特征向量进行降序排列,这些点在区间[30, 60]时明显地趋于平缓,因此确定最佳的聚类中心数量应在[30, 60]。

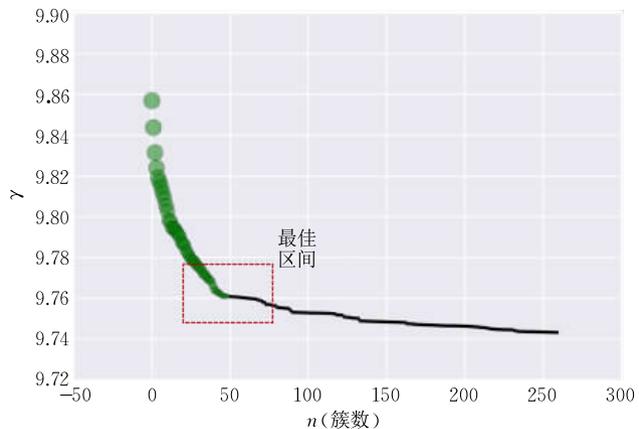


图 8 聚类簇数选取最佳区间

为了与原始分类信息效果对比,在这个区间下本文对聚类中心数量 $K \in \{0, 30, 40, 50, 60\}$ 进行了实

验,以推荐列表长度 $\text{Top-N} \in \{5, 10, 15, 20, 25, 30\}$ 为自变量分别检验实验结果的精确率,召回率和 F1 值,如图 9 所示。

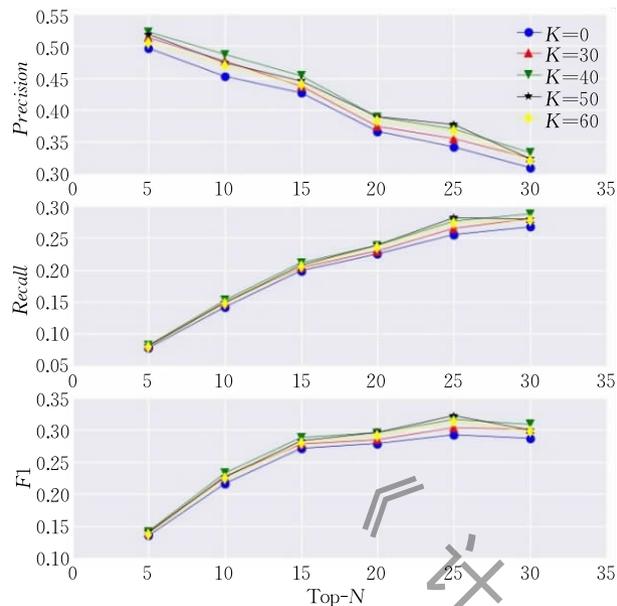


图 9 K 在不同 Top-N 下对推荐效果的影响

可以看出:(1)精确率随着 Top-N 的上升而下降.当 $\text{Top-N}=5$ 时,各精确率高于 50%.即使当 $\text{Top-N}=30$ 时,最低精确率也高于 30%.本文推荐算法选取用户近期播放视频作为推荐的种子视频,更偏重于用户的近期行为;(2)不同 K 值在精确率指标上具有较明显的差异.当 $K=0$ 时,不对视频进行聚类处理,只利用视频原始分类信息,精确率指标最低. $K=30$ 次之. $K=0$ 与 $K=40$ 对应的精确率始终相差约 4 个百分点,其余 K 值对应的精确率均处于 $K=0$ 和 $K=40$ 之间;(3)聚类簇数 $K=40$ 时,精确率在不同的 Top-N 取值下都是最佳,可选择聚类簇数为 40;(4)从图 9 可看出,在相同的推荐列表长度 Top-N 下,不同的聚类簇数量 K 对应的召回率差异较小,并且随着 Top-N 的增加,召回率指标在增加的同时,不同 K 值对应的召回率指标差异在逐渐增大.由于用户的视频播放列表是不变的,所以随着 Top-N 的增加,有更多包含在视频播放列表的视频被推荐给用户,所以召回率会增加,而精确率会下降;(5)从图 9 中可以看出,F1 指标随 Top-N 先上升后下降,并于 $\text{Top-N}=25$ 时取得最佳值。

4.3.3 视频特征向量维度影响

本文还实验验证了不同视频向量维度 $|\mathbf{V}_i|$ 对个性化推荐效果的影响.固定兴趣缩放因子 $\sigma=1$,聚类中心数 $K=45$,对 $|\mathbf{V}_i| \in \{50, 100, 150, 200, 250\}$

进行了多次实验,对比推荐 $|\text{Top-N}|=20$ 下的精确率、召回率和 F1 值,如图 10 所示。

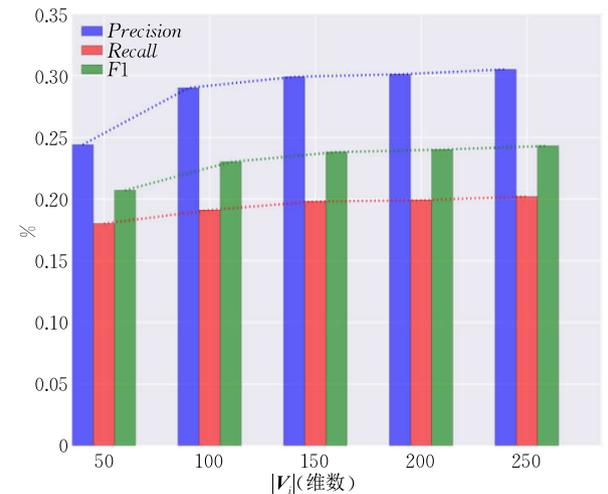


图 10 视频向量维度对推荐效果影响

可以看出,推荐精确率随视频向量维度 $|\mathbf{V}_i|$ 的增大而增长,其中维度由 50 维上升至 100 维时精确率有明显增长,当维度 $|\mathbf{V}_i| > 100$ 时,精确率略微提升,召回率和 F1 值的结果类似.这也不难理解,如果向量维度过小,那么视频特征向量并不能完全反映视频的本质信息.当维度增加到一定程度时,视频特征向量已完整地表达了视频信息,推荐效果不再有明显提升.另外,视频特征向量建模训练时间和产生推荐列表的时间都会随着维度 $|\mathbf{V}_i|$ 的增加而增加,所以应该选择合适 $|\mathbf{V}_i|$ 来权衡计算速度以及推荐效果.本文后续实验选择特征向量维度值为 200。

4.3.4 兴趣缩放因子 σ

兴趣缩放因子 σ 是用户兴趣偏好 $I(u, c_i)$ 的一个参数,不同的参数取值也会对推荐效果有不同影响.设 $I(u, c_i)$ 中

$$g(\sigma) = \frac{-\|\mathbf{V}_x - c_i\|^2}{\sigma} \quad (18)$$

那么 $g(\sigma)$ 的导数如下:

$$g'(\sigma) = \frac{\|\mathbf{V}_x - c_i\|}{\sigma^2} \quad (19)$$

当参数 σ 取值增大时, $g'(\sigma)$ 减小,用户兴趣偏好 $I(u, c_i)$ 梯度减小,此时用户对每一类视频的兴趣程度近似呈现为均匀分布; $g'(\sigma)$ 随着参数 σ 的减小而逐渐增大,如果 σ 减小至 0,表明用户对所有类别视频的兴趣均为零.因此 σ 需选择合适的参数值.本文在 $|\mathbf{V}_i|=200$, $|\text{Top-N}|=20$ 条件下对不同的 $\sigma \in \{1, 5, 10, 15, 20\}$ 进行仿真实验,如图 11 所示。

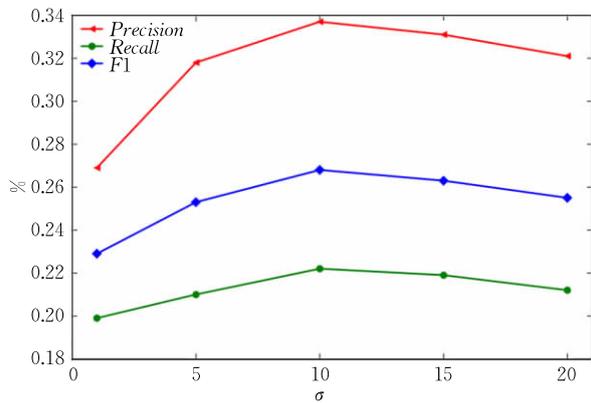


图 11 兴趣缩放因子 σ 对推荐效果影响

从图 11 可以看出,随着参数 σ 的增大,精确率、召回率和 F1 值先增大后减小,并于 $\sigma=10$ 达到峰值.在后续的推荐实验中,取经验值 $\sigma=10$.

4.3.5 算法对比

本文与基于物品的协同过滤算法 ItemCF、基于用户的协同过滤算法 UserCF 和热门推荐 HotRec 进行了 Top-N=25 的对比实验,如表 5 所示. Seq4Rec 代表本文提出的个性化推荐算法.

表 5 传统算法对比

| Algorithm | Precision | Recall | F1 |
|-----------|-----------|--------|-------|
| Seq4Rec | 0.362 | 0.268 | 0.308 |
| ItemCF | 0.296 | 0.195 | 0.235 |
| UserCF | 0.277 | 0.183 | 0.221 |
| HotRec | 0.035 | 0.023 | 0.028 |

从表 5 可看出,本文提出的推荐算法在各个指标上都明显优于 ItemCF、UserCF 和 HotRec 算法.这是由于 ItemCF 和 UserCF 等传统的推荐算法无法建模用户的兴趣偏好,所以在同等数据的前提下,本文的推荐效果显著优于传统的推荐算法.

在性能方面,ItemCF 和 UserCF 算法由于需要消耗大量的资源去维护一个极度稀疏且高维的 User-Item 矩阵,且在计算 Item 之间和 User 之间相似度时相当耗时,而本文策略有效地避免了这个问题,所以在性能和效果上都优于对比算法.另外,本文中的视频特征向量更新是利用用户新的播放行为进行迭代即在线学习,视频特征向量模型只与视频库中视频数量呈正相关,算法效率高效.而随着用户行为历史的积累,UserCF 和 ItemCF 算法中 User-Item 矩阵却会变得越来越大,维护难度也会变得越来越高,推荐的性能和效果会大大受限.

进一步地,本文在不同 Top-N 的条件下分别与奇异值分解算法 SVD++、Bi-LSTM+Attention

模型、深度兴趣网络 DIN 模型和 Video2Vec 模型进行了比较.图 12 为不同算法的推荐准确率、召回率、F1 值. SVD++ 算法是在 SVD 模型中融入用户的隐式兴趣. Bi-LSTM+Attention 使用引入注意力机制的双向 LSTM 模型. DIN 模型是从用户的历史行为中自适应地学习用户的兴趣表征.此外为了说明用户兴趣建模的必要性,本文设计了 Video2Vec 模型作为对比,即直接利用用户最近观看的视频求其最相似的视频集(粗排序),不经过用户兴趣建模(精排序)产生推荐.

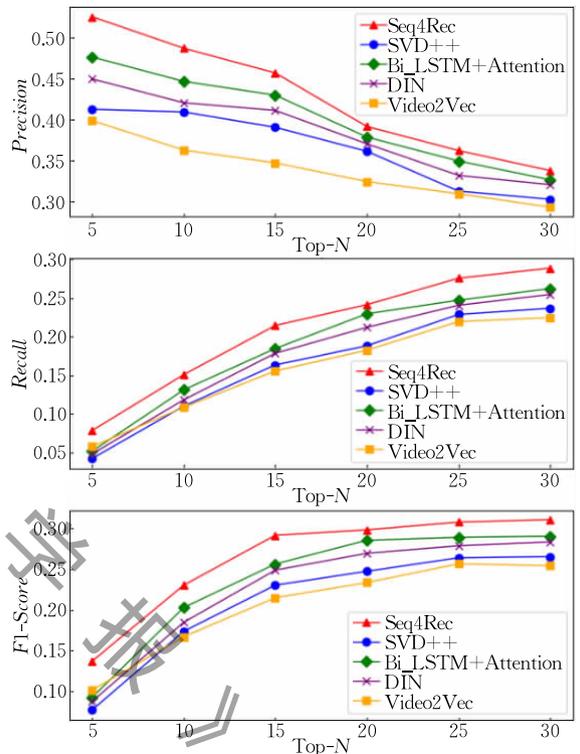


图 12 不同算法对比(准确率、召回率、F1 值)

从图 12 可以看出,本文提出的推荐算法在各项指标上都优于 SVD++、Bi-LSTM+Attention、DIN 模型和 Video2Vec 模型.这是由于 SVD++ 算法无法利用用户的视频观看序列,不能更好地建模视频向量. Bi-LSTM+Attention、DIN 和 Video2Vec 模型虽然可以利用用户的视频观看序列,但是无法针对每个用户建立个性化兴趣模型,因此本文提出的推荐算法具有更明显的优势.另外, Bi-LSTM+Attention 和 DIN 模型都需要消耗大量资源去维护模型参数,并且模型的计算复杂度与用户观看序列的长度呈正相关关系.换言之,随着用户历史行为的积累, Bi-LSTM+Attention 和 DIN 模型会变得更加不可取.与 Video2Vec 算法相比,本文的算法也具有更加明显的优势,充分说明了用户个性化兴趣

建模的必要性与有效性。

图 13 是不同推荐算法的多样性指标图. 本文算法略逊于 SVD++、Bi_LSTM+Attention 和 DIN 模型, 但比单纯的 Video2Vec 模型要好. 这是因为 SVD++、Bi_LSTM+Attention 和 DIN 模型不采用用户的视频相似性产生推荐, 因此其多样性高于本文算法和 Video2Vec 算法. Video2Vec 模型是仅仅依靠用户历史行为视频的相似性产生候选列表, 导致了推荐视频列表具有很高的相似性. 本文算法引入了对用户的个性化兴趣建模, 只有符合用户的个性化兴趣的视频才会推荐给用户, 通过对用户候选列表的精排序过程, 排除了与用户的历史行为视频拥有极高相似度但却不符合用户个性化兴趣的视频, 获得了推荐高准确率、高召回率和高 F1 值的同时, 还兼顾了推荐多样性.

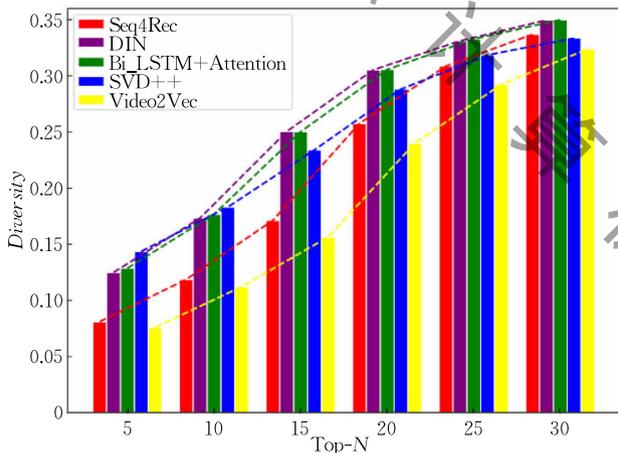


图 13 不同算法对比(多样性)

5 结 论

针对当前视频服务网站内容数量趋于爆炸的趋势, 且内容混杂 PGC、OGC 和 UGC 的情况下, 我们提出了一种通用的推荐算法, 该算法无需人工特征设计, 无需视频的媒资信息, 只利用用户播放行为通过“粗排序+精排序”结合的方法, 可对大规模工业应用数据产生准确稳定的推荐. 在粗排序阶段, 主要利用了基于深度学习的词向量模型建立视频向量, 根据用户的观看历史, 选取与用户观看历史最相似的若干视频作为候选推荐列表. 在精排序阶段, 利用 K-Means 算法聚类用户观看历史中所有视频特征向量, 利用聚类建模出用户的兴趣偏好. 本文和传统推荐算法 ItemCF、UserCF 算法和热门推荐的推荐

效果进行比较, 还与矩阵分解算法 SVD++、基于双向 LSTM 模型的 Bi-LSTM+Attention 算法和基于用户行为序列的深度兴趣网络 DIN 的推荐效果进行了比较. 实验结果表明, 本文提出的推荐算法在准确度、召回率等指标上具有明显效果.

本文提出的视频推荐算法主要面对当前移动互联网用户个性化推送场景, 然而在实际应用中仍然需要考虑较多的状况. 如用户观看视频时的行为可能隐式地表达出用户的喜好程度, 用户对正在观看的视频的喜好程度不应以是否点击来进行衡量. 比如用户虽然点击视频但是在观看过程中全程快进和正常速度播放所表达出的对视频喜好程度是不一致的. 所以如何利用用户的隐反馈信息来改善推荐效果是本文后续研究的重要工作之一.

参 考 文 献

- [1] Schafer J B, Dan F, Herlocker J, et al. Collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 2004, 22(1): 5-53
- [2] Davidson J, Liebald B, Liu J, et al. The YouTube video recommendation system//*Proceedings of the ACM Conference on Recommender Systems*. Barcelona, Spain, 2010: 293-296
- [3] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. *Computer*, 2009, 42(8): 30-37
- [4] Polat H, Du W. SVD-based collaborative filtering with privacy //*Proceedings of the 2005 ACM Symposium on Applied Computing*. Santa Fe, New Mexico, 2005: 791-795
- [5] Choi S M, Ko S K, Han Y S. A movie recommendation algorithm based on genre correlations. *Expert Systems with Applications*, 2012, 39(9): 8079-8085
- [6] Elkahky A M, Song Y, He X. A multi-view deep learning approach for cross domain user modeling in recommendation systems//*Proceedings of the International World Wide Web Conferences Steering Committee*. Florence, Italy, 2015: 278-288
- [7] Covington P, Adams J, Sargin E. Deep neural networks for YouTube recommendations//*Proceedings of the ACM Conference on Recommender Systems*. Boston, USA, 2016: 191-198
- [8] Sedhain S, Menon A K, Sanner S, et al. AutoRec: Autoencoders meet collaborative filtering//*Proceedings of the International Conference on World Wide Web*. Florence, Italy, 2015: 111-112
- [9] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. *arXiv preprint arXiv: 1301.3781*, 2013
- [10] Barkan O, Koenigstein N. Item2Vec: Neural item embedding for collaborative filtering//*Proceedings of the IEEE 26th*

- International Workshop on Machine Learning for Signal Processing (MLSP), Salerno, Italy, 2016; 1-6
- [11] Ozsoy M G. From word embeddings to item recommendation. arXiv preprint arXiv:1601.01356, 2016
- [12] Grbovic M, Radosavljevic V, Djuric N, et al. E-commerce in your inbox: Product recommendations at scale//Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney, Australia, 2015; 1809-1818
- [13] Duan Qi. Design and Implementation of Personalized Restaurant Recommendation System Based on Word2Vec [M. S. thesis]. Chongqing University, Chongqing, 2017(in Chinese)
(段琦. 基于 Word2Vec 的个性化餐饮推荐系统设计与实现 [硕士学位论文]. 重庆大学, 重庆, 2017)
- [14] Vasile F, Smirnova E, Conneau A. Meta-Prod2Vec: Product embeddings using side-information for recommendation//Proceedings of the 10th ACM Conference on Recommender Systems. Boston, USA, 2016; 225-232
- [15] Chen Liang, Wang Jing-Fu, Wang Na, et al. DNN based mobile online video recommendation. Chinese Journal of Computers, 2016, 39(8): 1626-1638(in Chinese)
(陈亮, 汪景福, 王娜等. 基于 DNN 算法的移动视频推荐策略. 计算机学报, 2016, 39(8): 1626-1638)
- [16] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality//Proceedings of the Advances in Neural Information Processing Systems (NIPS 2013). Lake Tahoe, USA, 2013; 3111-3119
- [17] Goldberg Y, Levy O. Word2Vec explained: Deriving Mikolov et al.'s negative-sampling word-embedding method. arXiv preprint arXiv:1402.3722, 2014
- [18] Rodriguez A, Laio A. Clustering by fast search and find of density peaks. Science, 2014, 344(6191): 1492
- [19] Linden G, Smith B, York J. Amazon.com recommendations; Item-to-Item collaborative filtering. IEEE Internet Computing, 2003, 7(1): 76-80
- [20] Wang J, Yin J. Combining user-based and item-based collaborative filtering techniques to improve recommendation diversity //Proceedings of the IEEE 6th International Conference on Biomedical Engineering and Informatics. Hangzhou, China, 2013; 661-665
- [21] Kumar R, Verma B K, Rastogi S S. Social popularity based SVD++ recommender system. International Journal of Computer Applications, 2014, 87(14): 33-37
- [22] Zhang Y, Dai H, Xu C, et al. Sequential click prediction for sponsored search with recurrent neural networks//Proceedings of the 28th AAAI Conference on Artificial Intelligence. Quebec, Canada, 2014; 1369-1375
- [23] Zhou G, Zhu X, Song C, et al. Deep interest network for click-through rate prediction//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK, 2018; 1059-1068
- [24] Laurens V D M, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research, 2017, 9(2605): 2579-2605



WANG Na, Ph. D., professor. Her main research interests include intelligent computing, machine learning, and pattern recognition.

HE Xiao-Ming, M. S. candidate. His main research interests include recommendation system and data analysis.

LIU Zhi-Qiang, M. S. candidate. His main research interests include recommendation system and data analysis.

WANG Wen-Jun, M. S. candidate. His main research interests include recommendation system and data analysis.

LI Xia, Ph. D., professor, Ph. D. supervisor. Her main research interests include machine learning, intelligent computing and signal processing.

Background

Along with the development of mobile Internet, massive video resources are pouring into people's modern life all the time, and the issue of personalized video recommendation has aroused interest of people in mobile online video services. The explosion valuable information to design a smart mobile video recommendation strategy become increasingly important. Currently, there exist many developed studies on the personalized recommendation system, including content-based

recommendation, item based collaborative filtering, user-based collaborative filtering and so on. Meanwhile, deep learning has been studied in depth in the area of natural language processing (NLP) and it is already proved to be more suitable for abundant descriptions in current big data environment. It is promising to explore the deep learning based recommendation strategy for mobile online videos.

This paper introduces a personalized recommendation

strategy based on the user's playback behavior sequence for online video service websites. There are three major steps involved. Firstly, the video feature vectors are obtained based on rich corpus of user's playing videos with Word2Vec tool. Secondly, the user's interest distribution matrix is modeled by clustering the feature vectors of the user's video playing history. Thirdly, we filter the candidate videos a target video list that a user has potential interest in based on his interest preferences and viewing history aging. Offline abundant experiments were conducted in a large-scale video service system, and we evaluate our proposal. We compared it with random algorithms, item-based collaborative filtering recommendations, and user-based collaborative filtering recommendations, the results show that our proposal gets significant improvement on the recommendation performance.

Moreover, compared with the matrix decomposition algorithm SVD++, the model based on the bidirectional LSTM + Attention and the deep interest network DIN, the proposed method achieves different levels of improvement in both the Top-N recommendation accuracy rate and the recall rate. The recommendation strategy not only obtains better performance as a whole, but also attempts to solve the problems of data requirements, data sparsity and data noise faced by traditional recommendation algorithms.

The project was supported by the National Natural Science Foundation of China (61871273), the Shenzhen Innovation Technology Program(JCYJ20160422112909302), and the Guangdong Provincial Science Technology Plan Project (20138021500017).

《计算机学报》