基于图神经网络的视频推荐系统

Video Recommender System with Graph Neural Networks

高宸/GAO Chen，李勇/LI Yong，金德鹏/JIN Depeng
（清华大学，中国 北京 100084）
(Tsinghua University, Beijing 100084, China)

摘要： 提出了一种基于图神经网络的视频推荐模型，将用户的视频观看序列型行为建模为图结构，用结点代表用户与视频，用边代表行为，引入两种类型的向量传播方法分别对用户的长期兴趣与短时兴趣进行建模。其中，通过用户结点与视频结点的双向传播刻画长期兴趣，借助视频结点切换关系的单向传播刻画短时兴趣，并通过多层向量传播实现对图上高阶邻接信息的捕捉。在一个真实世界的视频网站观看数据集上的实验表明，提出的方法与现有最佳方法相比，其推荐精准度得到了有效提升。进一步的实验表明，该方法能够有效缓解数据稀疏性的问题。

关键词： 视频推荐系统；用户兴趣建模；图神经网络；深度学习

Abstract: A novel recommendation model with graph neural networks is proposed. Users' sequential video-watching behaviors are first constructed as a graph, which represents users and videos as nodes, and behaviors as edges. Then two kinds of embedding propagation methods are introduced for capturing users' long-term and short-term preferences, respectively. Specifically, a user-item bi-directional embedding propagation layer is used for capturing long-term preferences while an item-item embedding propagation layer for capturing short-term preferences. Moreover, the multi-layer propagation is proposed to extract high-order connectivity. Experiments on a real-world video-watching dataset verify that the proposed method can outperform the state-of-the-art methods. Further experiments demonstrate that the proposed method can effectively alleviate the data sparsity issue.

Keywords: video recommender system; user preference modeling; graph neural network; deep learning

DOI：10.12142/ZTETJ.202101007
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20210120.1010.002.html

网络出版日期：2021-01-20
收稿日期：2020-12-10

在信息超载时代，个性化推荐系统[1-2]成为用户获取信息的主要方式。推荐系统通过收集用户的历史行为来推断用户兴趣，进而生成推荐列表。与常见的电商网站推荐系统[3]不同，视频网站上的用户行为具有两个重要特性。首先，用户的视频观看行为呈现出高度的序列性。一段时间内浏览的视频表现出极高的相关性，且浏览的前后顺序十分重要，因此需要对用户行为进行序列化建模。其次，用户可能存在短期观看某一类/系列多个视频的“短时”兴趣，呈现出突发、多样的特点。因此，我们需要从长期兴趣与短时兴趣两方面对用户的兴趣进行细粒度化的建模。

针对序列化行为的推荐问题，现有的方法[4-6]仍然存在两部分缺陷。

基金项目：国家重点研发计划（2018YFB1800804）；国家自然科学基金（U1936217、61971267、61972223、61941117、61861136003）；北京自然科学基金（L182038）；北京国家信息科学与技术研究中心基金（20031887521）；清华大学-腾讯联合实验室项目

首先，仅仅使用权重或者卷积/循环神经网络对不同历史行为进行隐式建模的方法，缺乏对序列化行为中视频切换关系的显式建模；其次，目前的推荐方法没有考虑针对用户长期与短时兴趣的细粒度建模。本文中，我们设计了一种基于图神经网络的推荐模型，通过两种向量传播方式来分别对用户的长期兴趣和短时兴趣进行建模。此外，我们还引入了多层向量传播以捕捉图上高阶邻接信息。

1 问题定义

视频推荐系统的目标是尽可能地满足用户的需求，即为用户推荐最符合其兴趣的视频。在视频推荐系统中，相关输入数据为用户历史视频观看的行为序列，其中，序列中的前后关系代表用户观看视频的先后顺序关系。输出数据则为可计算给定用户下一次观看给定视频的概率模型。在得到该模型后，我们可对所有候选视频进行概率计算，并按照概率预估值从大到小排序，得到推荐列表。

2 方法设计

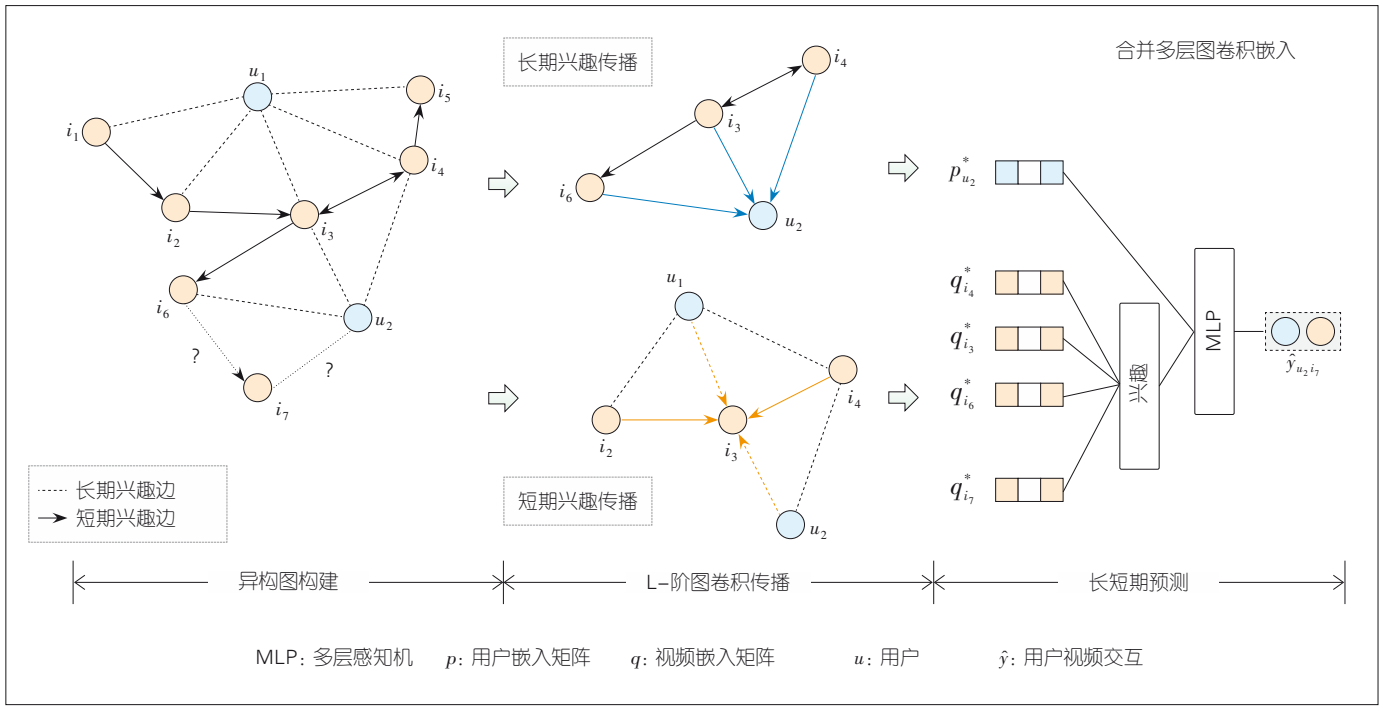
这里我们提出一种基于图卷积网络的视频（VGCN）推荐模型，具体如图1所示。该推荐模型主要由4个流程部分构成：构建包含用户视频结点与行为边的异构图、构建嵌入层以得到用户与视频的表征向量、设计向量以刻画用户的长期兴趣与短时兴趣、引入预测层得到用户观看视频的概率。

2.1 异构图构建

图是一种具备强大数据表示能力的结构。在视频推荐系统中，一种直观并且有效的做法是，把用户与视频分别表示为图中两种类型的结点，并将用户的观看行为建模为图上的边。具体而言，我们构建异构图 $G=<V,E>$，其中，V 表示所有结点的集合，E 表示所有边的集合。同时集合 V 中有两类结点：用户结点 $u\in U$ 与视频结点 $i\in I$。集合 E 中有两类边：一类是用户与视频的交互边 $r\in R$，其中 r_{ui} 代表用户 u 与视频 i 存在交互行为（观看）；另一类是视频切换边 $r\in T$，其中 t_{ij} 代表视频 i 到视频 j 的切换行为，并且该边的权重由数据集中所有用户的切换次数决定。如果该权重为0，即边不存在，则代表没有用户产生从观看 i 切换到观看 j 的行为。总的来说，我们得到了包含两类结点与两类边的异构图。

2.2 嵌入层

针对图的表征学习可定义为：通过机器学习的方法，为点、边或图学习其在低维空间的表征。该方法可以将高维的图数据转换为低维特征向量，实现预测、分类等任务[7]。在通用视频推荐任务中，由于用户画像、视频属性等数据较难收集，用户与视频一般而言仅有身份标识（ID）



▲图1 视频图卷积网络推荐模型示意图

特征，即仅有用户与视频的原始编号。因此，我们针对ID特征设计适用于独热编码的嵌入层，为用户与视频分别建立嵌入矩阵$\boldsymbol{P}$与$\boldsymbol{Q}$。$\boldsymbol{P}$的维度为N乘以D，$\boldsymbol{Q}$的维度为M乘以D。其中，N为用户的数目，M为视频的数目。D为低维空间的维度，是一个可以调整的超参数，其过大的维度会带来过拟合问题，而过小的维度则存在欠拟合问题。

我们通过用户与视频的嵌入矩阵与独热编码可得到每个用户与视频的低维表征向量，如公式(1)所示：

$$p_u = \boldsymbol{P}^{\mathrm{T}} v_u^U, q_i = \boldsymbol{Q}^{\mathrm{T}} v_i^I, \tag{1}$$

其中，v_u^U与v_i^I分别为用户与视频的独热编码。

独热编码是一种仅有一个位置为1且其余位置为0的高维向量(值为1的位置即为原始编号)。用户独热编码的长度为N，视频独热编码的长度为M。在嵌入矩阵完成随机初始化后，模型的后续部分将从嵌入矩阵得到最终的预测结果。当基于预测损失的随机梯度下降时，嵌入矩阵即可从初始化的随机向量逐渐调整至可刻画用户与视频特征的高质量表征向量。

2.3 向量传播层

我们首先建立了上述用户与视频的嵌入矩阵。该嵌入矩阵可以被视为第0层用户/视频向量。接着，我们设计向量传播层以利用图上的高阶邻接关系，以捕捉用户的长期与短时兴趣。图卷积网络是一类最典型的图神经网络[8]。向量传播是图卷积网络的核心模块，其核心思想是将向量传播给图上的邻居结点，以实现图结构邻接性到向量相似性的转化，并可通过多层向量传播实现对高阶邻接关系的建模。借助向量传播方法，图卷积网络在诸多任务上取得了当前最佳性能[8-9]。

在视频推荐系统中，需要对用户进行两方面的兴趣建模：长期兴趣与短时兴趣。其中长期兴趣侧重于用户较为固定的、不随时间变化的兴趣，短时兴趣则与之相反。具体而言，我们通过用户结点的表征向量对其长期兴趣进行建模，通过用户上一时刻交互的结点的表征向量对其短期兴趣进行建模。这种做法与用户长期兴趣与短时兴趣的物理意义相契合。

2.3.1 长期兴趣向量传播层

用户观看视频的行为被表示为图上用户结点与视频结点之间的边。如果将用户观看过的所有视频结点的向量表征往其传播，聚合之后自然就代表了与具体时间无关的长期兴趣。因此，我们可以通过用户结点-视频结点向量传播对长期兴趣进行建模。具体而言，某个用户观看过的所有视频往其传播，向量传播公式如下：

$$p_u^{(l+1)} = \sigma(W_1^{(l+1)}(p_u^{(l)} + aggregate(q_i^{(l)} | i \in R_u)) + b_1{}^{(l+1)}), \tag{2}$$

$$q_{i,1}^{(l+1)} = \sigma(W_1^{(l+1)}(q_{i,1}^{(l)} + aggregate(p_u^{(l)} | u \in R_i)) + b_1{}^{(l+1)}), \tag{3}$$

其中，σ是非线性激活函数，$W_1^{(l+1)}$与$b_1^{(l+1)}$分别是从第l层往l+1层传播时的网络参数与偏置参数(此处我们添加下标1以与后文的短时兴趣相关参数加以区分)，R_u是用户u交互过的所有视频，R_i是观看过视频i的所有用户，$aggregate(\cdot)$是向量聚合操作(常见做法有求平均等)，$p_u^{(l)}$为用户u第l层的表征向量，$p_u^{(l+1)}$为用户u第l+1层的表征向量，$q_i^{(l)}$为视频i第l层的表征向量，$q_i^{(l+1)}$为视频i第l层的表征向量。

总的来说，前文所述的向量操作实现了长时兴趣侧从低层向量到高层向量的计算方式。随着层数的逐渐提升，更高阶的邻接关系将会被提取至表征向量中。但值得一提的是，层数不能过高，这是因为向量传播可以被理解为一种局部图的近邻平滑作用，如果层数过深，则相当于实现了全局平滑，反而会使学习到的表征向量无效。

2.3.2 短时兴趣向量传播层

上述长期兴趣向量传播层通过忽略序列关系的历史行为边传播，来刻画用户的长期兴趣。接着，我们进一步设计用于对用户的短时兴趣进行建模的向量传播方法。考虑到用户的短时兴趣与视频观看的切换行为需要相契合，我们采用基于视频切换行为的有向边来设计向量传播方法。换而言之，向量传播的路径就是上一个视频到下一个视频的有向边。由于此处不涉及用户结点的表征向量，因此，我们可以实现长期兴趣与短时兴趣的解耦建模。

具体而言，传播公式如公式(4)所示：

$$q_{i,2}^{(l+1)} = \sigma(W_2^{(l+1)}(q_{i,2}^{(l)} + aggregate(q_j^{(l)} | j \in T_i)) + b_2{}^{(l+1)}), \tag{4}$$

其中，σ是非线性激活函数，$W_2^{(l+1)}$与$b_2^{(l+1)}$是从第l层往l+1层传播时的网络参数与偏置参数(此处我们添加下标2，以与前文所述的长期兴趣相关参数加以区分)，T_i是存在往i有向边的所有视频的集合。此处的$aggregate(\cdot)$是带权重的聚合操作，该权重为往i的有向边的权重，即存在该切换行为的用户的个数，同时该权重可从一定程度上对切换关系的强弱进行建模。

2.4 预测层

在以上向量传播层的基础上，我们进一步设计得到最终预测结果的预测层。通过长期兴趣向量传播与短时兴趣向量传播，我们得到了两部分视频向量。此处我们使用求和操作，将两部分合并为一部分，即 $q_i^{(l+1)} = q_{i,1}^{(l+1)} + q_{i,2}^{(l+1)}$。

考虑到不同层数的用户或视频向量包含了不同阶数的图上邻接关系，我们将不同层数的向量使用拼接操作进行聚合，如公式(5)和公式(6)所示：

$$p_u^* = p_u^{(0)} \| p_u^{(1)} \| \ldots \| p_u^{(L)}, \tag{5}$$

$$q_i^* = q_i^{(0)} \| q_i^{(1)} \| \ldots \| q_i^{(L)}, \tag{6}$$

其中，L为一个可以调整的超参数，$\|$代表向量的拼接操作。

随后，我们通过一个基于注意力网络的预测函数，对给定的用户与视频预测观看概率，具体如公式(7)所示：

$$\hat{y}_{ui} = MLP(p_u^* \| ATN(q_i^*, R_u)), \tag{7}$$

其中，MLP代表一个多层感知机，ATN代表一个注意力网络。最终输出的$\hat{y}_{ui}$为一个从0到1之间的概率值，该值越大，用户u越有可能观看视频i。

2.5 训练方法

在获得对于任意给定用户与视频的观看概率预估后，我们基于对数损失函数进行优化。由于数据中仅记录了用户观看过的视频，即正样本，我们需要从未观看的视频中随机采集一些样本作为负样本。对于正样本而言，模型的预测结果要尽可能接近1；对于负样本，模型的预测结果要尽可能接近0。损失函数具体计算方式如公式(8)所示：

$$Loss = -\sum_{(u,i) \in Y^+ \cup Y^-} y_{ui} \log \hat{y}_{ui} + (1 - y_{ui}) \log (1 - \hat{y}_{ui}), \tag{8}$$

其中，Y^+与Y^-分别为正样本集合与负样本集合，y_{ui}与$\hat{y}_{ui}$分别为数据真实标签与模型预测输出。

基于以上损失函数，我们可通过随机梯度下降进行更新。在我们提出的VGCN模型中，需要学习的参数主要为用户与视频的嵌入矩阵$\boldsymbol{P}$与$\boldsymbol{Q}$。向量传播操作中的W与b等参数则占据着较少的参数量，而每加深一次图卷积网络，只会引入额外的W与b等参数。因此，不同深度的VGCN模型的参数量基本相当，均约等于嵌入矩阵的参数量大小。而嵌入矩阵参数量大小不仅与用户数目和视频数目的和呈线性关系，也与嵌入矩阵向量维度D呈线性关系。

3 实验验证

为了验证提出的VGCN方法的有效性，我们对真实视频观看数据集进行了推荐性能的验证。

3.1 实验设置

3.1.1 数据集

我们在一个视频网站上收集了2020年10月的用户视频观看行为数据。由于完整数据规模过大，我们随机选取了一部分用户。经过预处理后的数据，包括了60 813个用户与292 286个视频产生的14 952 659条观看记录。对于每一个用户而言，其观看记录为一条包含了若干个视频的序列。

3.1.2 性能指标

视频推荐乃至通用推荐系统最常使用的指标为排序指标，其中最具代表意义的指标为特征曲线下方的面积(AUC)、平均倒数排名值(MRR)与归一化折损累计增益(NDCG)[10]。AUC可衡量模型对于所有正样本与负样本相对关系的区分能力，MRR衡量模型将正样本排在列表靠前位置的能力，NDCG则衡量模型排序结果与理想排序结果的距离。

3.1.3 基线模型

我们选取两个极具竞争力的模型作为基线模型：卷积序列嵌入推荐模型(CASER)[5]与深度兴趣网络(DIN)[6]。其中，CASER通过卷积网络建模用户的行为序列，DIN通过注意力网络建模用户行为序列。

3.2 推荐性能比较

我们首先对整体的推荐结果进行比较，如表1所示。

由表1可知，与现有模型相比，我们提出的VGCN模型在AUC、MRR、NDCG@1、NDCG@2等指标上，可以有效且稳定地提升推荐性能，且平均相对提升值约为1.7%。对于推荐系统模型而言，该提升值是显著的。

图2则展示了不同方法训练时的模型损失曲线。由图2可以看出，我

▼表1 视频推荐精准度性能比较

模型	AUC	MRR	NDCG@1	NDCG@2
CASER	0.7471	0.8788	0.7576	0.9105
DIN	0.7561	0.8856	0.7712	0.9156
VGCN	0.7781	0.8943	0.7886	0.9220

AUC：特征曲线下方的面积　MRR：平均倒数排名值
CASER：卷积序列嵌入推荐模型　NDCG：归一化折损累计增益
DIN：深度兴趣网络　VGCN：基于图卷积网络的视频推荐算法

们的VGCN方法可以取得更小的训练损失。

3.3 稀疏度影响研究

在推荐系统尤其是视频推荐系统中，数据稀疏十分重要。具体而言，对于不同稀疏性的用户，能否均取得较好的效果，是衡量一个推荐模型好坏的重要指标。因此，我们将用户的历史交互行为数目分3组进行研究：0～50、50～200、200以上。每组均有足够的用户数目，以消除随机性。对于每组的用户，为计算其平均推荐精准度，我们选取了AUC与NDCG@2两个排序指标，具体结果如图3所示。

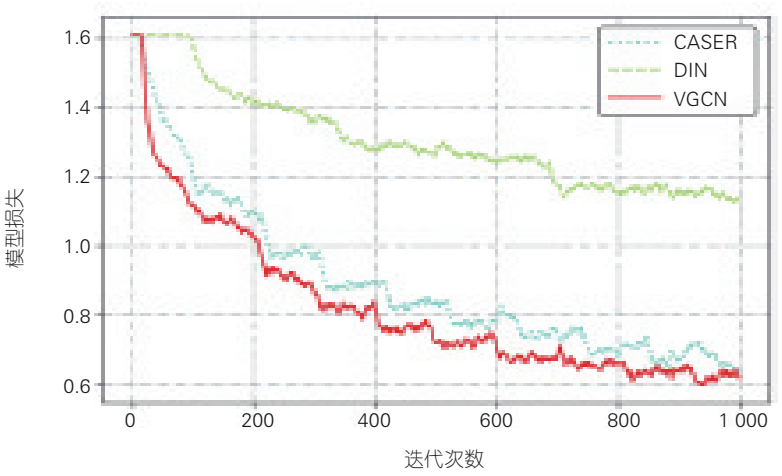
由图3可知，我们提出的VGCN方法在不同稀疏度的用户组里，均可取得有效且稳定的性能提升。这一结果进一步验证了VGCN方法的有效性。

3.4 超参数影响研究

在模型训练的过程中，L2正则系数是一项重要的超参数，图4展示了不同L2正则系数对视频推荐性能的影响。根据图4可以看出，不论选择何种L2正则系数，我们提出的VGCN方法均可以取得最佳推的荐性能。此外，L2正则系数对于模型视频推荐精准度性能的影响较小，即模型对于该超参数的敏感度较低，这意味着模型不需要花费太多的调参时间与算力。

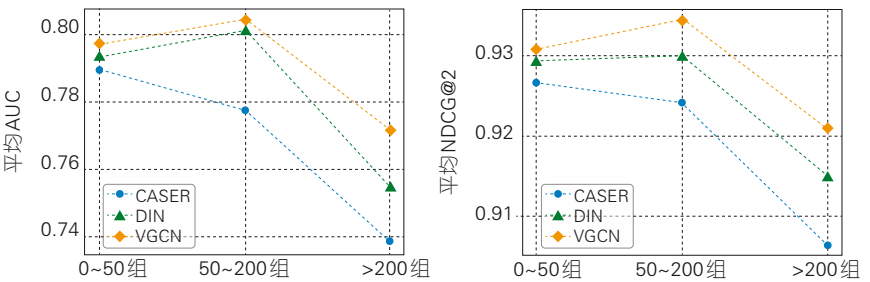
4 结束语

视频推荐系统是提升用户视频观看体验的重要技术。本文设计了一种基于图神经网络的推荐模型，并通过两种向量传播方式对用户长期兴趣与短时兴趣建模。基于真实数据集的实验有效验证了整体推荐精



CASER：卷积序列嵌入推荐模型
DIN：深度兴趣网络
VGCN：基于图卷积网络的视频推荐算法

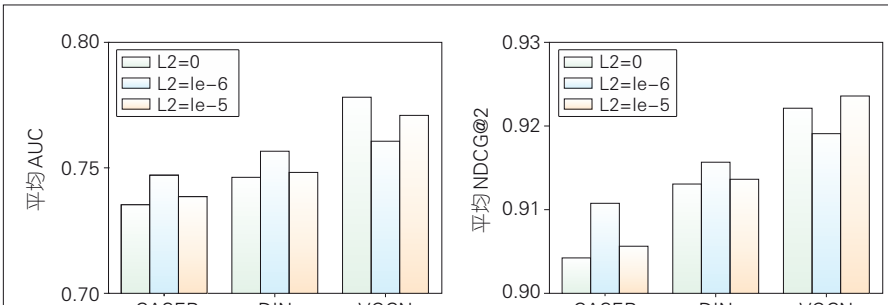
▲图2 训练时模型损失曲线比较



（a）平均AUC性能　（b）平均NDCG@2性能

AUC：特征曲线下方的面积
CASER：卷积序列嵌入推荐模型
DIN：深度兴趣网络
NDCG：归一化折损累计增益
VGCN：基于图卷积网络的视频推荐算法

▲图3 不同观看次数用户的视频推荐精准度性能比较



（a）平均AUC性能　（b）平均NDCG@2性能

AUC：特征曲线下方的面积
CASER：卷积序列嵌入推荐模型
DIN：深度兴趣网络
NDCG：归一化折损累计增益
VGCN：基于图卷积网络的视频推荐算法

▲图4 不同超参数设置（L2正则系数）的视频推荐精准度性能比较

准度与不同稀疏度用户推荐精准度的性能提升。同时，超参数影响的实验进一步验证了推荐精准度性能提升的稳定性。

致谢

本研究得到清华大学常健新同学的帮助，谨致谢意！

参考文献

[1] 许海玲, 吴潇, 李晓东, 等. 互联网推荐系统比较研究 [J]. 软件学报, 2009, 20(2): 350–362. DOI: 10.3724/SP.J.1001.2009.03388
[2] LU J, WU D S, MAO M S, et al. Recommender system application developments: a survey [J]. Decision support systems, 2015, 74: 12–32. DOI: 10.1016/j.dss.2015.03.008
[3] 朱岩, 林泽楠. 电子商务中的个性化推荐方法评述 [J]. 中国软科学, 2009(2): 183–192. DOI: 10.3969/j.issn.1002–9753.2009.02.022
[4] RENDLE S, FREUDENTHALER C, SCHMIDT–THIEME L. Factorizing personalized Markov chains for next–basket recommendation [C]// Proceedings of the 19th International Conference on World Wide Web. North CA, USA: ACM Press, 2010. DOI: 10.1145/1772690.1772773
[5] TANG J X, WANG K. Personalized top–N sequential recommendation via convolutional sequence embedding [C]//Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. Los Angeles, CA, USA: ACM, 2018. DOI: 10.1145/3159652.3159656
[6] ZHOU G R, ZHU X Q, SONG C R, et al. Deep interest network for click–through rate prediction [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. NY, USA: ACM, 2018. DOI: 10.1145/3219819.3219823
[7] GOYAL P, FERRARA E. Graph embedding techniques, applications, and performance: a survey [J]. Knowledge–based systems, 2018, 151: 78–94. DOI: 10.1016/j.knosys.2018.03.022
[8] 徐冰冰, 岑科廷, 黄俊杰, 等. 图卷积神经网络综述 [J]. 计算机学报, 2020, 43(5): 755–780. DOI: 10.11897/SP.J.1016.2020.00755
[9] KIPF T N, WELLING M. Semi–supervised classification with graph convolutional networks [C]//Proceedings of International Conference on Learning Representations. Toulon, France: University of Montreal, 2017
[10] MANNING D C, SCHÜTZE H, RAGHAVAN P. Introduction to information retrieval [M]. Cambridge: Cambridge University Press, 2008

作 者 简 介

高宸，清华大学电子工程系在读博士研究生；主要研究领域为用户行为建模与挖掘。

李勇，清华大学电子工程系副教授；主要研究领域为网络科学、城市计算、用户行为建模与挖掘。

金德鹏，清华大学电子工程系教授；主要研究领域为网络科学、城市计算、用户行为建模与挖掘。