



## 地理位置和时间感知的表示学习框架

周俊, 张志强, 曹月恬, 郑小林

引用本文:

周俊, 张志强, 曹月恬, 等. 地理位置和时间感知的表示学习框架[J]. 智能系统学报, 2021, 16(5): 909–917.

ZHOU Jun, ZHANG Zhiqiang, CAO Yuetian, et al. A geography and time aware representation learning framework[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(5): 909–917.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202104011>

## 您可能感兴趣的其他文章

### 基于知识图谱和用户长短期偏好的个性化景点推荐

Personalized attraction recommendation based on the knowledge graph and users' long-term and short-term preferences  
智能系统学报. 2020, 15(5): 990–997 <https://dx.doi.org/10.11992/tis.201904064>

### 基于注意力机制的显著性目标检测方法

Salient object detection method based on the attention mechanism  
智能系统学报. 2020, 15(5): 956–963 <https://dx.doi.org/10.11992/tis.201903001>

### 基于注意力融合的图像描述生成方法

An image caption generation method based on attention fusion  
智能系统学报. 2020, 15(4): 740–749 <https://dx.doi.org/10.11992/tis.201910039>

### 层次化双注意力神经网络模型的情感分析研究

Hierarchical double-attention neural networks for sentiment classification  
智能系统学报. 2020, 15(3): 460–467 <https://dx.doi.org/10.11992/tis.201812017>

### 引入外部词向量的文本信息网络表示学习

Representation learning using network embedding based on external word vectors  
智能系统学报. 2019, 14(5): 1056–1063 <https://dx.doi.org/10.11992/tis.201809037>

### 多特征融合的兴趣点推荐算法

A point of interest recommendation algorithm based on multi-feature fusion  
智能系统学报. 2019, 14(4): 779–786 <https://dx.doi.org/10.11992/tis.201801048>

微信公众平台



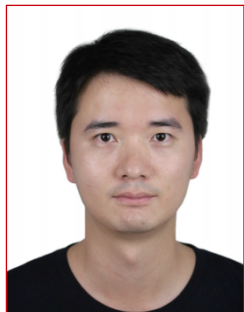
关注微信公众号, 获取更多资讯信息

## 吴文俊人工智能科技进步奖一等奖

成果名称：金融智能图机器学习系统研发与规模化应用

获 奖 人：周俊、朱军、石川、张志强、褚崴、何昌华、顾进杰

完成单位：蚂蚁科技集团股份有限公司、清华大学、北京邮电大学



### 周俊

蚂蚁集团&阿里巴巴达摩院-金融智能实验室资深专家，金融智能技术的创新研发先行者。现任蚂蚁集团-机器学习技术负责人，研发完成了图机器学习、数据安全与隐私保护、互联网金融智能风控系统等核心技术，致力于推动金融行业智能化。在NeurIPS、KDD等人工智能顶级刊物上发表论文60余篇，获得授权专利50余项，获得浙江省科技进步奖、CCF科学技术奖等多个科技奖项，为人工智能在行业的落地树立了良好的典范。带领团队在多项人工智能竞赛中取得世界排名第一的成绩，牵头和参与多项国际/行业标准的制定。周俊在推动人工智能与多个实际场景的结合过程中，与

清华大学、北京大学、浙江大学等高校建立了紧密的合作合作，在推动产学研结合方面发挥了积极作用，为人工智能行业培养了一批青年科技创新人才和工程技术人才。

## 团队简介

项目团队组建于2016年，由蚂蚁集团联合清华大学、北京邮电大学三家单位的资深专家组成，团队成员30余人。团队核心成员包括：蚂蚁科技集团有限公司资深技术专家周俊，清华大学教授朱军，北京邮电大学教授石川，蚂蚁科技集团有限公司高级技术专家张志强。团队研发完成了面向金融智能的图机器学习系统，包括：高精度的工业级图表示学习算法库，超大规模图机器学习架构，工业级全链路图学习服务平台，实现了海量图结构数据智能建模在金融领域中的大规模应用，取得了显著的经济价值和社会效益。基于该研究成果已获授权发明专利55项，近三年在NeurIPS、KDD、VLDB、SIGIR等人工智能重要会议和期刊发表相关论文超过40篇。

DOI: 10.11992/tis.202104011

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20210714.0944.002.html>

# 地理位置和时间感知的表示学习框架

周俊<sup>1,2</sup>, 张志强<sup>2</sup>, 曹月恬<sup>2</sup>, 郑小林<sup>1</sup>

(1. 浙江大学 计算机科学与技术学院, 浙江 杭州 310007; 2. 蚂蚁集团, 浙江 杭州 310013)

**摘要:** 现有时空感知的表示学习框架无法对强时空语义的实际场景存在的“When”、“Where”和“What”3 个问题给出一个统一的解决方案。同时, 现有的时间和空间建模上的研究方案也存在着一定的缺陷, 无法在复杂的实际场景中取得最优的性能。为了解决这些问题, 本文提出了一个统一的用户表示框架——GTRL (geography and time aware representation learning), 可以同时时间和空间的维度上对用户的历史行为轨迹进行联合建模。在时间建模上, GTRL 采用函数式的时间编码以及连续时间和上下文感知的图注意力网络, 在动态的用户行为图上灵活地捕获高阶的结构化时序信息。在空间建模上, GTRL 采用了层级化的地理编码和深度历史轨迹建模模块高效地刻画了用户的地理位置偏好。GTRL 设计了统一的联合优化方案, 同时在交互预测、交互时间预测以及交互位置 3 个任务上进行模型学习。最后, 本文在公开数据集和工业数据集上设计了大量的实验, 分别验证了 GTRL 相较学术界基线模型的优势, 以及在实际业务场景中的有效性。

**关键词:** 时空语义; 时间建模; 空间建模; 注意力机制; 图学习; 图神经网络; 用户行为建模; 用户行为表征  
**中图分类号:** TP183    **文献标志码:** A    **文章编号:** 1673-4785(2021)05-0909-09

中文引用格式: 周俊, 张志强, 曹月恬, 等. 地理位置和时间感知的表示学习框架 [J]. 智能系统学报, 2021, 16(5): 909-917.

英文引用格式: ZHOU Jun, ZHANG Zhiqiang, CAO Yuetian, et al. A geography and time aware representation learning framework[J]. CAAI transactions on intelligent systems, 2021, 16(5): 909-917.

## A geography and time aware representation learning framework

ZHOU Jun<sup>1,2</sup>, ZHANG Zhiqiang<sup>2</sup>, CAO Yuetian<sup>2</sup>, ZHENG Xiaoling<sup>1</sup>

(1. College of Computer Science and Technology, Zhejiang University, Hangzhou 310007, China; 2. Ant Group, Hangzhou 310013, China)

**Abstract:** The existing geography and time aware representation learning framework fails to give a unified solution to three problems of “When”, “Where” and “What” in real-world scenarios with strong geography and time aware semantics, and the existing research schemes for time and spatial modeling also have certain defects and thus cannot achieve optimal performance in complex actual scenarios. To fill in these gaps, this paper proposes a unified user representation framework —GTRL (geography and time aware representation learning), which is capable of jointly modeling users’ historical behaviors in both time and spatial dimensions. In terms of time modeling, GTRL adopts a functional time encoding and a continuous time and context aware graph attention network to flexibly exploit high-order and structured temporal information from dynamic user behaviour graphs. In terms of spatial modeling, GTRL uses a hierarchical geography encoding and a deep historical trajectory modelling module to effectively depict users’ preference for geographical location. Moreover, a unified optimization strategy is carefully designed to carry out model learning on three tasks of interaction prediction task, interaction time stamp prediction task and interaction geography prediction. At last, extensive experiments on public and industrial data sets are carried out, demonstrating the advantages of GTUL over the academic baseline model and its effectiveness in actual business scenarios, respectively.

**Keywords:** semantic-space-time mode; time modelling; spatial modelling; attention mechanism; graph learning; graph neural network; user behavior modeling; user behavior representation

收稿日期: 2021-04-07. 网络出版日期: 2021-07-14.

基金项目: 国家重点研发计划项目 (2018YFB1403001).

通信作者: 郑小林. E-mail: xlzheng@zju.edu.cn.

近年来, 许多研究者致力于使用表示学习的手段来学习用户表征, 刻画用户的历史行为, 并

将其用于各种各样的在线服务中,比如商品推荐、优惠券投放、在线广告等。但现有的方法大多聚焦于对用户历史的点击/购买商品序列建模,无法直接应用于具有强时空语义的实际场景中(如美团、饿了么等本地生活平台)。在这样的场景中,要同时解决“*When*”、“*Where*”和“*What*”的问题,即在特定的时间、特定的地点为用户推荐合适的商品。为此,需要学习一个好的用户表征,它不仅仅能捕获用户在点击/购买序列上表现的兴趣偏好,更要刻画这些兴趣偏好随着时间和位置的演化特征。本文的核心研究任务是如何在时间和空间的维度上学习用户的统一表征,刻画用户行为的时空语义。

在时间建模上,考虑到行为稀疏性和用户关联性,本文主要在用户-商品交互(如购买)的时序图上进行用户表示学习。国外针对时序比较成熟的表示学习方法主要有3类:1)基于随机游走的方法<sup>[1-2]</sup>直接将时序信息融合到随机游走序列中,并通过经典的 skip-gram 方法学习动态表征。但这类方法忽略了节点和边上的丰富信息,无法进行归纳学习。2)基于时间切片的方法<sup>[3-4]</sup>先在不同的时间切片上进行表示学习,然后通过神经网络将不同切片上的表示融合为最终的节点表示。这类方法忽略了每个时间切片内的动态性,一定程度上损害了节点关系随时间演化的特性;3)基于连续时间的方法<sup>[5-6]</sup>尝试在连续时间上进行时间序列或时序图的表示学习。但是此类方法大多仍然无法进行归纳学习。更重要的是,它们在进行时间编码的时候忽略了时序图上节点间的交互和时间维度上的上下文信息及其影响。

在空间建模上,如何对地理位置进行编码至关重要。当前的表征方法包括使用 geohash 网格信息、兴趣点(point of interest, POI)或兴趣面(area of interest, AOI)的 one-hot 编码,以及使用经纬度信息等。但是这些方式都有自己的缺点:1) geohash 网格信息或者 POI 信息虽然可以精确区分不同的地点,但是由于数量太大,做 one-hot 编码特征会导致模型参数爆炸。另外 geohash 网格或 POI 忽视了地理位置的远近关系信息,不利于模型使用。2) AOI/POI 的数量在几百万左右,规模不算很大,但是很多 AOI/POI 覆盖的面积很大,这样就不能区分其中包含的不同位置实体,另外目前 AOI/POI 的覆盖率也不是很高;3) 直接经纬度信息虽然避免了 one-hot 编码特征维度过大的问题,也保持了地理位置的远近关系,但是由于经度和纬度值的微小变化,实际位置就会偏离很

远,模型很难使用这类特征。

针对以上不足,本文的目标是开发一个统一的用户表示框架——GTRL (geography and time aware representation learning),可以同时时间和空间维度上对用户的历史行为轨迹进行联合建模。

在时间信息建模上,首先, GTRL 通过 Mercer 理论<sup>[7]</sup>来实现函数式的时间编码,将时间戳(时间差)映射到高维空间中,并保存节点的时序性。其次,设计了连续时间和上下文感知的图注意力网络(C<sup>2</sup>GAT),一个既可以利用图上高阶结构信息也可以在连续时间上刻画节点局部上下文的、全新的、归纳式的图表示学习模型。

在空间信息建模上,为 GTRL 设计了一种新颖的地理位置编码器——GeoEncoder,以更高效简洁的方式将每个地理位置映射为固定长度的字符串。基于经典的 *n*-gram 技术, GTRL 可以在更加细粒度的视角上对位置信息进行刻画。最后,依赖新兴的注意力机制对用户的历史位置轨迹进行了深度建模,最终生成了用户的地理偏好表征。

在模型优化方面, GTRL 致力于同时解决上述的“*When*”、“*Where*”和“*What*”3个问题,因此为 GTRL 设计了统一的联合优化方案。值得注意的是,在优化时间上,该方案引入了用户行为中的时间间隔信息,用以刻画行为的周期性、偏好的衰减性等一系列细粒度特征。

最后,本文在公开数据集和工业数据集上设计了大量的实验,一方面验证了 GTRL 的性能可以超过当前学术界最优的基线模型,另一方面证明了 GTRL 在实际业务场景中的优势。

## 1 整体框架

本文提出的 GTRL 整体算法框架如图1所示,框架会利用用户的历史行为进行建模,包括历史交互的商品(如用户购买过的商品),交互时间以及交互发生的地理位置(购买的商铺所在的地理位置)。框架的两个核心模块共同支持时空表示学习:1)时序信息表示模块包含函数式的时间编码及连续时间和上下文感知的图注意力网络,主要用于学习时序图上的用户表示;2)空间信息表示模块包含层次化的地理位置编码和深度用户轨迹建模子模块,主要对用户历史位置偏好表征的生成。为了统一解决时空语义场景的通用问题(即“*When*”、“*Where*”和“*What*”3个问题),设计了统一的且可解耦的目标函数来同时优化交互预测、时间间隔预测以及地理位置预测。

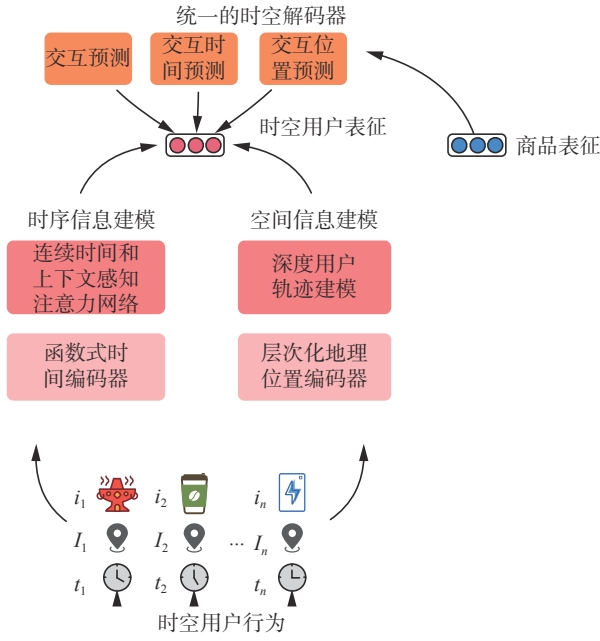


图 1 GTRL 框架的整体结构

Fig. 1 Overall architecture of GTRL framework

## 2 时序信息建模

为了更好地克服行为稀疏性和捕获用户行为的高阶关联性, 本文从图的角度出发进行时序信息的建模, 并设计了连续时间和上下文感知的时序图学习方法, 主要分为两个部分: 1) 为了避免离散时间建模带来的信息损失, GTRL 尝试用函数式的时间编码技术将时间戳映射到连续可微的向量空间上更好地保存时序图上的动态性; 2) 在最近新兴的图神经网络结构上, 融合时间编码, 设计了连续时间和上下文感知的注意力机制, 可在图上按时间顺序选取重要的邻居, 并同时考虑不同邻居在时间上的相互影响。

### 2.1 函数式的时间编码

函数式时间编码的目的是找到一个从时间域到  $d$  维向量空间的映射  $\Phi(\cdot)$ 。具体地, 考虑任意两个时间戳  $t_1, t_2 \in \mathcal{T}$ , 需要关心的是如何学习这两个时间间隔之间的关系模式, 可以表示成它们对应的时间编码的内积, 即  $\langle \Phi(t_1), \Phi(t_2) \rangle$ 。由此, 可以将上述的时序模式进一步形式化成具有平移不变性的核函数  $\mathcal{K}$ , 且其映射函数为  $\Phi(\cdot)$ 。

为了学习上述的时序核函数及其映射函数, 在 Mercer 理论的启发下, 可以将该映射函数定义为

$$t \mapsto \Phi^M(t) := [\sqrt{c_1} \phi_1(t), \sqrt{c_2} \phi_2(t), \dots]$$

经验上来说, 时序模式可以被一系列周期核函数刻画, 参照文献 [8] 介绍的定理, 以  $\omega$  为频率的映射函数  $\Phi(\cdot)$  可以进一步被形式化为

$$t \mapsto \Phi_\omega^M(t) := \left[ \sqrt{c_1} \cos\left(\frac{j\pi t}{\omega}\right), \sqrt{c_{2j+1}} \sin\left(\frac{j\pi t}{\omega}\right), \dots \right]$$

这样一个傅里叶级数形式的表示有比较好的截断特性, 因为可以用截断的  $d$  维的映射函数  $\Phi_{\omega,d}^M(t)$  来近似原始的无穷维的映射。随后, 将  $k$  个周期 (即  $\{\omega_1, \omega_2, \dots, \omega_k\}$ ) 的映射函数拼接, 形成最终的时间编码为

$$t \mapsto \Phi_d^M(t) := [\Phi_{\omega_1,d}^M(t) \parallel \dots \parallel \Phi_{\omega_k,d}^M(t)]$$

值得注意的是, 这样的时间编码是节点无关的。换句话说, 由于时间编码和节点不相关, 会导致相同时间间隔下的任意两个节点的时间编码都是一样的。但在具体的应用场景中, 如个性化推荐, 用户在过去同一时间上购买的两个不同的商品, 对当下用户兴趣的影响是不同的。如一天前喝过咖啡今天很可能会继续喝, 但一天前吃过火锅今天再吃的可能性就比较低了。因此, 这样的节点无关的时间编码不适合用于后续图神经网络来进行时序图表示。给定一个特定的节点  $v$ , 可以进一步将该节点的时间编码定义为

$$v, t \mapsto \Phi_\omega^M(v, t) := \left[ \sqrt{c_1(v)} \cos\left(\frac{j\pi t}{\omega}\right), \sqrt{c_{2j+1}(v)} \sin\left(\frac{j\pi t}{\omega}\right), \dots \right]$$

$$v, t \mapsto \Phi_d^M(v, t) := [\Phi_{\omega_1,d}^M(v, t) \parallel \dots \parallel \Phi_{\omega_k,d}^M(v, t)]$$

式中  $c_i(u): \mathbf{R}^d \rightarrow \mathbf{R}, i=1, 2, \dots$  是一系列节点相关的映射函数, 用来计算傅里叶系数。由于多层感知机对复杂交互的优秀建模能力, 在实验中将其用于实现  $c_i(u)$ , 并强制感知机的最后一层输出为正数, 以此来满足 Mercer 理论的内在特性。

### 2.2 连续时间和上下文感知的图注意力网络

在 Mercer 理论的帮助下, 可以对时序图上的每个节点相关的时间间隔进行时间编码, 以此在连续时间上建模节点的时序模式。在此基础上, 本文设计了连续时间和上下文感知的图注意力网络 (C<sup>2</sup>GAT) 来自适应地学习图上的时序信息, 一方面在时间上衡量不同邻居节点的重要性/相关性, 另一方面考虑邻居节点间在时间维度上的相互影响。具体地, 在 C<sup>2</sup>GAT 的第  $l$  层, 给定一个  $t$  时刻的目标节点  $v_t$ , 需要对该节点的邻居节点集合  $\mathcal{N}_{v_t}(t) = \{v | t_v < t\}$  产生一个注意力分布用来融合每个邻居节点的表示。由于之前定义的时间核函数具有平移不变性, 可以使用时间差  $\{t - t_v\}_{v \in \mathcal{N}_{v_t}(t)}$  来替代绝对时间。在  $t$  时刻, 对于目标节点  $v_t$  及其任意一个邻居  $v \in \mathcal{N}_{v_t}(t)$ , 该节点对之间的连续时间和上下文感知的注意力值可以计算为

$$\alpha_{v,v_t}(t) = \frac{\mathbf{Q}_{v_t}(t) \mathbf{K}_v^T(t)}{\sqrt{d}}$$

$$\mathbf{Q}_{v_i}(t) = [h_{v_i}^{(l-1)}(t) \| 0 \| \Phi_d^M(0)] \mathbf{W}_Q$$

$$\mathbf{K}_v(t) = [\mathcal{F}^{(l-1)}(v, t, t_v) \| e_{v_i, v}(t) \| \Phi_d^M(t - t_v)] \mathbf{W}_K$$

式中:  $\mathbf{W}_Q$  和  $\mathbf{W}_K$  是两个映射矩阵, 分别用来获取“Query”和“Key”矩阵;  $e_{v_i, v}(t)$  为  $t$  时刻节点  $v_i$  和节点  $v$  的边特征向量;  $h_v^{(l-1)}$  为节点在  $C^2GAT$  第  $l-1$  层的输出。显然, 这里设计的注意力机制可以灵活地扩展为多头注意力机制, 以此保证更加稳定的训练。如上所述, 节点邻居间在时间上的上下文信息及其相互影响是刻画目标节点和邻居节点注意力值的一个很重要的因素。为此, 设计了一个上下文聚合函数  $\mathcal{F}^{(l-1)}(v, t, t_v)$  来显式地刻画这样的影响。给定目标节点  $v_i$ , 对任意一个邻居节点  $v$  及其对于的时间  $t_v$ , 可以定义如上的上文邻居集合  $S_{v_i, v}(t_v) = \{v' \in \mathcal{N}_{v_i}(t) | t'_v \leq t_v\}$ , 进一步地, 这里可以使用两种聚合函数来实现  $\mathcal{F}(\cdot)$ 。

循环聚合器在序列上下文上应用复杂的 LSTM(long short-term memory) 结构来赋予  $C^2GAT$  强大的表达能力。

$$\mathcal{F}_R = \text{LSTM}(h_v^{(l-1)}(t); v' \in S_{u, v}(t_v))$$

卷积聚合器使用了卷积操作来保证  $C^2GAT$  可以扩展到大规模数据集上。

$$\mathcal{F}_C = \sum_{i=1}^{d-1} \sum_{j=0}^{|S_{u, v}(t_v)|} h_j^{(l-1)}(t) W_{ji}$$

式中  $W \in \mathbf{R}^{|S_{u, v}(t_v)| \times d \times d}$  是卷积核。

通过叠加  $L$  层上述的  $C^2GAT$  层, 该方法可以从更广更深的维度在时序图中利用高阶的结构信息。由此, 可以将每个节点在  $t$  时刻的节点表示记为  $\{h_v^T(t) = h_v^{(L)}(t)\}_{v \in V}$ 。本文更关心用户的表示, 因此, 对于每个用户  $u$ , 将其在  $t$  时刻的表示重写为  $h_u^T(t)$ 。

### 3 空间信息建模

空间信息建模, 即建模用户在空间上的轨迹信息, 是对用户时空行为刻画的另外一个主要模块。具体来说, 空间轨迹一方面可以在空间上区别人与人之间的关系, 刻画用户的空间迁移性, 从而实现更好的用户表示; 另一方面, 它也可以描述人与物(服务、权益和店铺等)的关系, 增强模型的预测能力。为此, 本文设计了层级化的地理编码和深度的历史轨迹建模来实现这一目标, 如图2所示。

#### 3.1 层级化的地理编码

为了避免传统地理位置编码(geohash 网格、POI 或者 AOI 的 one-hot 编码以及经纬度编码等)

的不足, 设计的地理位置编码器(GeoEncoder)参考了谷歌地图和必应地图中广泛使用的 Tile Map System, 将全地图等分为4块, 分别标为0、1、2、3。然后将标号为“0”的地块(Tile)继续等分为4小块, 分别标为00、01、02、03, 这4小块有共同的前缀:“0”, 也就是它们上层那个地块的编号。其他标号为“1”、“2”、“3”的地块进行同样的划分操作。这样递归地将地图不断划分下去, 直到最小地块满足要求的进度为止。采样这种地理编码方式, 中国地图, 考虑最小地块长度为10米, 只需要长度为19的序列编码就可以覆盖全国所有位置。这样产出的编码具有很好的性质: 给定任意两个地块, 它们的共同前缀越短, 它们距离越远; 反之则越近。图3具体展示了对位置“杭州市-西湖区-天目里”的层级化地理编码。

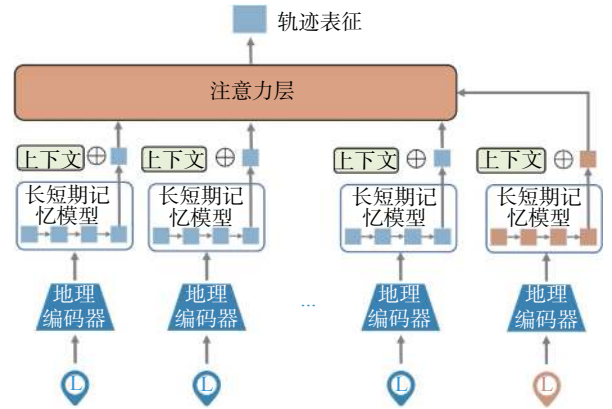


图2 空间信息建模示意

Fig. 2 Illustration of geography aware modelling

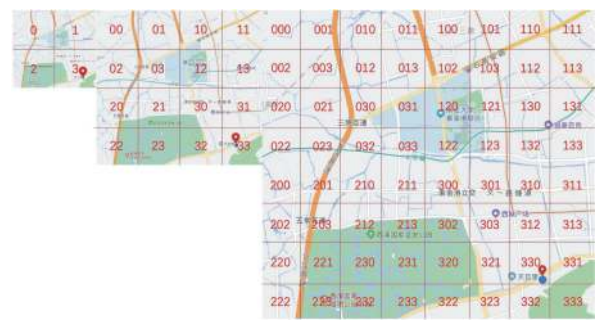


图3 层级化地理位置编码示意(以“杭州市-西湖区-天目里”为例)

Fig. 3 Illustration of hierarchical location encoding (take “Hangzhou-Xihu-Tianmuli” as an example)

但是并不能直接使用这个编码作为ID特征。进一步地, 参考2020 KDD的工作<sup>[8]</sup>, 可以对这个编码做  $n$ -gram 处理(一般取  $n=6$ ) 转换成序列的形式。具体来说, 给定编码“120220011012000332”做 6-gram 处理之后会变为 [‘120220’, ‘202200’, ‘022001’, ‘220011’, ‘200110’, ‘001101’, ‘011012’,

‘110 120’, ‘101 200’, ‘012000’, ‘120003’, ‘200033’, ‘000332’] 的序列, 其中单词量为 4 096(4<sup>6</sup>)。

一个地理位置基于 GeoEncoder +  $n$ -gram 做序列编码之后, 该序列从左到右逐步刻画了更细粒度的位置信息。可以类比于从国家粒度, 逐步精细到地区、省、市、县等。不同的是, GeoEncoder +  $n$ -gram 的序列编码方式还可以描述相对位置信息。考虑  $n=1$  的最简单情况, 当一个序列编码的第  $k$  个位置为 0、1、2 和 3 时, 表示它在当前区域的西北、东北、西南和东南区域。当  $n$  取值变大, 序列编码则可以表示更复杂的相对位置信息。为刻画上述序列编码层级递进的性质, GTRL 的地理编码模块使用 LSTM 模型来建模序列编码得到每个地理位置的低维表征。

### 3.2 深度的历史轨迹建模

对每个用户  $u$ , 将其历史的轨迹 (即地理位置) 集合表示为  $\mathcal{H}_u^G$ 。通过层次化的地理编码, 每个位置  $g$  都被表示为低维向量  $h_g$ 。显然, 对于一个目标位置  $g_t$  来说, 用户历史地理位置对其有着不同的影响。因此, 本文设计了一个注意力层来刻画这些信息。具体地, 该注意力层为每个历史位置学习一个自适应的注意力值:

$$\alpha_{g_t, g} = \frac{\exp(\mathbf{v}^T f(\mathbf{W}_1^G [h_{g_t} || h_g]))}{\sum_{g' \in \mathcal{H}_u^G} \mathbf{v}^T f(\mathbf{W}_1^G [h_{g_t} || h_{g'}])}$$

式中:  $\mathbf{W}_1^G$  和  $\mathbf{v}$  分别为映射矩阵和映射向量;  $f(\cdot)$  为激活函数, 一般定义为  $\tanh(\cdot)$ 。紧接着, 利用学到的自适应注意力值, 通过加权的方式整合历史地理位置信息, 从而生成最终的用户地理位置偏好表示。

$$h_u^G = \sum_{g \in \mathcal{H}_u^G} \alpha_{g_t, g} h_g$$

## 4 模型学习

如上, GTRL 已经从空间和时序的角度来充分刻画用户的行为偏好, 分别为  $h_u^G$  和  $h_u^T(t)$ 。现在, 本文将讨论如何将学习到的用户时空表示用于实际的场景中。以具有强时空语义的推荐场景为例, 该场景主要关心用户在当前时刻和当前地点是否会点击/购买一件商品。其中,  $h_u^G$  和  $h_u^T(t)$  分别刻画了用户在时间和空间上的历史交互行为, 另一方面, 用户行为中的时间间隔也往往蕴涵着丰富的信息。比如在 O2O 业务的优惠券核销率预估场景中, 用户从拿到优惠券到核销的时间间隔越短则说明用户有较强的使用券的心智, 反之则说明用户对券的使用心智较弱。

为此, GTRL 使用时序点过程的方式来建模用户行为的时间间隔, 该方法的核心为定义强度

函数, 在本文中 can 表示为  $\lambda_{v_i}(\tau|h_u(t))$ , 其中  $v_i$  为目标商品,  $\tau = t - t_i$  为当前时刻距上一时刻的时间间隔,  $h_u(t) = h_u^G || h_u^T(t)$  表示融合用户历史时空语义的最终表示。参照 FullyNet 的思想<sup>[9]</sup>, GTRL 直接建模累计强度函数为

$$\psi_{v_i}(\tau|h_u(t)) = \int_0^\tau \lambda_{v_i}(s|h_u(t)) ds$$

相比于传统的时序点过程, GTRL 使用了一种强度自由的形式来建模用户的时间间隔, 可以更好地适应复杂的真实应用场景。由于神经网络强大的拟合能力, GTRL 将其作为累积强度函数是实现基础, 整体的公式为

$$\psi_{v_i}(\tau|h_u(t)) = f(\mathbf{W}_L \cdots f(\mathbf{W}[h_u(t)||\tau] + \mathbf{b}_1) + \mathbf{b}_L)$$

s.t.  $\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_L \geq 0, \mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_L \geq 0$

式中:  $\mathbf{W}_*$  和  $\mathbf{b}_*$  分别为权重矩阵和偏置向量;  $f(\cdot)$  为激活函数, 在这里被定义为  $\text{ReLU}(\cdot)$ 。最终, 强度函数可以通过微分得到, 即  $\lambda_{v_i}(\tau|h_u(t)) = \frac{\partial \psi_{v_i}(\tau|h_u(t))}{\partial \tau}$ 。

为了解决时间  $t$ 、位置  $g$ 、商品  $v$  的联合预测, GTRL 通过如下的联合分布来估计:

$$p(v, g, t|u) = p(v|u) \cdot p(g|u) \cdot p(t, |u, v)$$

式中: 下一次交互预测  $p(v|u)$  与下一次位置预测  $p(g|u)$  分别由以  $h_u(t)$  为输入的多层感知机器得到; 而下一次交互间隔预测  $p(t, |u, v)$  由时序点过程支持。由于商品和位置量级很大, GTRL 使用了随机负采样的方式来进行近似计算。最后, 根据极大似然估计的策略, 最终的损失函数为

$$\mathcal{L} = \sum_{(u, v, g, t, \tau) \in S_v} C(y, p(v|u)) + \sum_{(u, v, g, \tau, y) \in S_g} C(y, p(g|u)) - \sum_{(u, v, g, t, \tau) \in S^+} \left( \log \lambda_v(\tau) + \sum_{v' \in P_{\text{neg}}} \int_0^\tau \lambda_{v'}(s) ds \right)$$

式中:  $S_v = S_v^+ \cup S_v^-$  表示对商品进行负采样后的所有训练样本集合; 同理,  $S_g = S_g^+ \cup S_g^-$  表示对地理位置进行负采样后的所有训练样本集合;  $S^+ = S_v^+ \cup S_g^+$  为训练集合中所有正样本集合;  $C(\cdot, \cdot)$  表示交叉熵损失函数。值得注意的是, GTRL 框架与具体的任务是完全解耦的, 使用者可以选取框架中的任意子模块进行任意预测任务 (如下一个商品预测, 下一个时间间隔预测或下一个地理位置预测)。

## 5 实验分析

### 5.1 公开数据集上的有效性实验

首先, 本文分别在 4 个公开数据集 (Reddit、Wikipedia、MOOC 和 LastFM) 上进行 (推导式/归纳式) 链路预测和节点分类任务, 来验证 GTRL 中的时序信息建模模块 (即 C<sup>2</sup>GAT) 的有效性。4 个数据集的详细统计信息如表 1 所示。

表1 4个公开数据集的统计信息  
Table 1 Statistics of four public datasets

数据集	节点数	边数	特征数
Reddit	11 000	672 477	172
Wikipedia	9 227	157 474	172
MOOC	7 145	411 749	4
LastFM	2 000	1 293 103	0

在实验中,选取了9个经典的基准方法,大致

表2 公开数据集上(推导式/归纳式)链路预测任务的性能比较

Table 2 Performance comparison (AUC) for (transductive / inductive) link prediction task on public datasets

方法		Reddit		Wikipedia		MOOC		LastFM	
		Accuracy	AP	Accuracy	AP	Accuracy	AP	Accuracy	AP
推导式	Time-LSTM	0.702 5	0.715 7	0.562 5	0.564 8	0.560 1	0.567 3	0.510 3	0.521 6
	Jodie	0.908 8	0.974 2	0.835 4	0.929 3	0.782 2	0.774 6	0.621 1	0.650 5
	GraphSAGE	0.932 3	0.983 0	0.888 9	0.959 9	0.700 4	0.745 9	0.644 1	0.692 2
	GAT	0.931 7	0.983 3	0.880 7	0.953 9	0.673 2	0.721 7	0.654 8	0.680 0
	CTDNE	0.781 0	0.859 4	0.552 1	0.568 9	0.580 2	0.591 9	0.392 0	0.439 9
	M <sup>2</sup> DNE	0.862 2	0.942 9	0.816 7	0.909 1	0.685 8	0.694 5	0.592 6	0.620 1
	GCRN	0.933 8	0.982 9	0.885 5	0.955 2	0.710 6	0.746 2	0.654 1	0.721 3
	GraphSAGE-T	0.930 3	0.982 3	0.896 9	0.964 8	0.756 6	0.786 8	0.679 1	0.776 5
	GAT-T	0.932 3	0.983 4	0.898 4	0.964 7	0.755 3	0.790 1	0.678 5	0.757 6
	TGAT	0.934 2	0.983 7	0.874 3	0.950 2	0.686 9	0.715 7	0.676 5	0.673 2
	GTRL-T	<b>0.935 5</b>	<b>0.984 5</b>	<b>0.906 1</b>	<b>0.971 4</b>	<b>0.790 8</b>	<b>0.861 4</b>	<b>0.717 1</b>	<b>0.812 9</b>
归纳式	GraphSAGE	0.900 1	0.965 0	0.862 7	0.944 2	0.697 3	0.736 5	—	—
	GAT	0.901 8	0.966 9	0.854 3	0.937 2	0.661 0	0.699 7	—	—
	GCRN	0.900 2	0.963 6	0.853 3	0.932 8	0.694 2	0.743 8	—	—
	GraphSAGE-T	0.899 1	0.965 0	0.876 1	0.956 6	0.764 1	0.797 6	—	—
	GAT-T	0.903 1	0.968 1	0.880 3	0.956 2	0.767 7	0.804 3	—	—
	TGAT	0.905 6	0.967 9	0.853 6	0.935 3	0.678 9	0.703 6	—	—
	GTRL-T	<b>0.906 9</b>	<b>0.968 2</b>	<b>0.883 6</b>	<b>0.960 5</b>	<b>0.796 2</b>	<b>0.844 2</b>	—	—

表3 公开数据集上节点分类任务的性能比较(AUC)  
Table 3 Performance comparison (AUC) for node classification task on public datasets

方法	Reddit	Wikipedia	MOOC
Time-LSTM	0.630 5	0.777 3	0.693 5
Jodie	0.610 6	0.762 9	0.665 5
GraphSAGE	0.650 2	0.797 4	0.674 2
GAT	0.661 7	0.848 0	0.645 9
GCRN	0.674 3	0.857 5	0.669 5
GraphSAGE-T	0.659 4	0.851 5	0.677 9
GAT-T	0.674 3	0.850 8	0.669 4
TGAT	0.641 1	0.860 6	0.678 6
GTRL-T	<b>0.684 6</b>	<b>0.875 3</b>	<b>0.688 1</b>

可以分为3类:1) 基于深度递归网络的方法: Time-LSTM<sup>[10]</sup>和Jodie<sup>[11]</sup>;2) 静态的图神经网络方法: GraphSAGE<sup>[12]</sup>和GAT<sup>[13]</sup>;3) 时序图神经网络方法: CTDNE<sup>[14]</sup>, M<sup>2</sup>DNE<sup>[15]</sup>、GCRN<sup>[16]</sup>、GraphSAGE-T和GAT-T和TGAT<sup>[5]</sup>。

本文将C<sup>2</sup>GAT与上述的9个基准方法在(推导式/归纳式)链路预测和节点分类任务上的性能对比实验结果分别展示在表2和表3中。从表2、3中可以得出如下两点结论:

1) GTRL中集成的C<sup>2</sup>GAT模型同时在链路预测和节点分类任务上都一致且明显优于其余基准模型,验证了C<sup>2</sup>GAT的有效性。这得益于C<sup>2</sup>GAT使用了一种更加有效的方式来捕获图上的动态性,并同时利用高阶结构信息和局部序列式的上下信息。

2) 在这些对比方法中,多数情况下,时序图模型的性能优于静态图模型优于深度序列模型。这证明了动态性、高阶结构以及节点/边特征对于时序图表示学习的重要性。

## 5.2 工业数据集上的实验

首先在支付宝首页的本地生活场景进行了一系列关于时间维度的实验,选取了外卖(“饿了么”服务)、旅游(“飞猪”服务)和栏目排序这3个

具有强时间语义的数据集,具体的数据集统计信息见表4。

表4 3个工业数据集的统计信息  
Table 4 Statistics of four industrial datasets

指标	外卖	旅游	栏目
用户数	1 388 754	1 323 131	8 396 541
商品数	1 937 950	17 859	17
训练数	4 351 640	3 839 993	31 673 510
验证数	200 000	200 000	200 000
测试数	3 385 8181	3 186 217	1 682 579

考虑到数据规模的量级,本文针对性地选取了一系列代表性的基准方法,如基于特征的DeepFM<sup>[17]</sup>模型、基于序列的GRU4Rec<sup>[18]</sup>、DIN<sup>[11]</sup>、SASRec<sup>[20]</sup>,以及基于兴趣演化的DIEN<sup>[21]</sup>和Time-LSTM<sup>[10]</sup>模型。详细的性能比较如表5所示。可以发现:在3个实际的场景数据集上,本文提出的方法对比最优的基准方法可以带来0.12%~1.71%的提升,验证了该方法在实际场景中捕获时间语义的有效性。

表5 工业数据集的性能比较

Table 5 Performance comparison on industrial datasets %

方法	外卖		旅游		栏目	
	GAUC	RI	GAUC	RI	GAUC	RI
DeepFM	75.34	3.31	71.98	4.86	88.37	0.26
GRU4Rec	76.82	1.31	73.73	2.37	88.39	0.15
DIN	76.82	1.31	73.71	2.40	88.46	0.16
DIEN	77.01	1.08	74.21	1.71	88.49	0.12
SASRec	76.88	1.25	73.62	2.52	88.43	0.19
Time-LSTM	76.98	1.12	73.99	2.01	88.42	0.20
GTRL-T	77.84	—	75.48	—	88.60	—

进一步,将更加显式地展示GTRL在时间间隔预测上的有效性。因此,本文在支付宝/本地生活的券核销预估中应用了GTRL。首先对比线上最优的基准方法(MMoE<sup>[22]</sup>),GTRL在核销率上可以带来4.23%的提升。其次,展示了这个任务中三类券的真实核销量-时间分布(蓝色柱状图)以及GTRL对券核销时间分布的估计(红色实线),如图4所示。可以发现,GTRL的预测与真实的核销分布基本吻合,也反映出了不同类型券核销的时间周期性:1)餐饮券:领券后一般都是尽快核销,核销量随时间呈明显衰减趋势;2)电影券:

领券后尽快核销(现领现买),领券第2天核销(1天后看电影);3)出行券:时间衰减不明显,根据实际出行需求进行核销,呈现天周期。

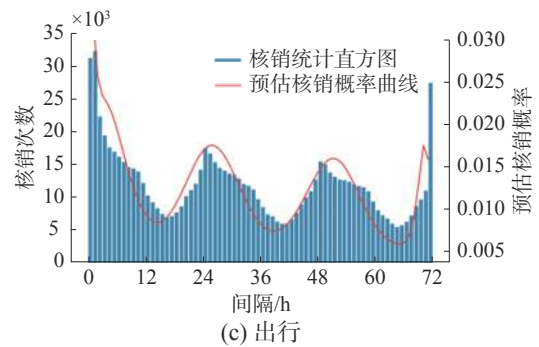
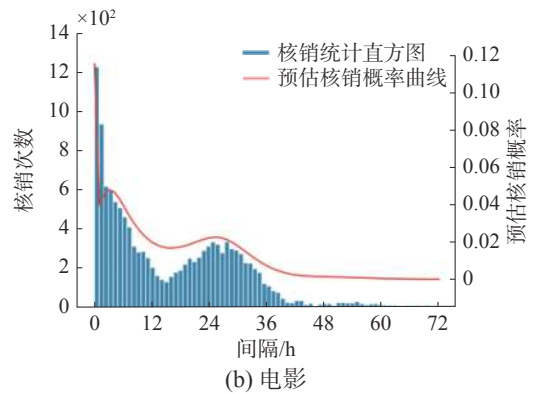
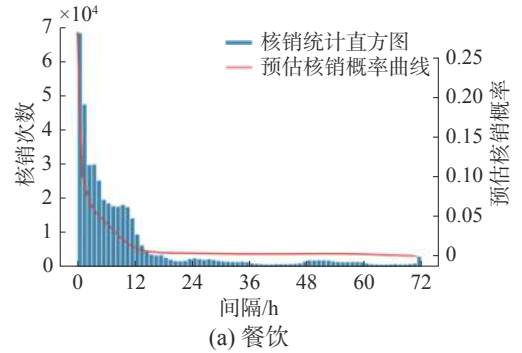
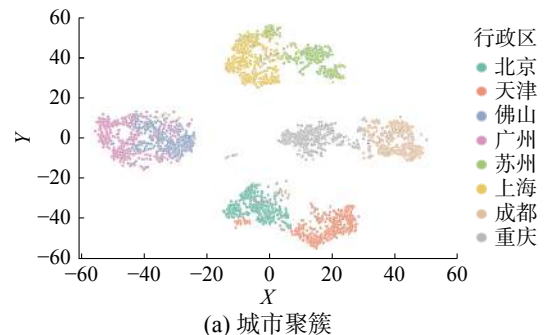


图4 时间间隔预测的案例学习

Fig. 4 Case study for time interval prediction

除了预测时间任务外,GTRL也可以进行空间地理信息的建模<sup>[23-25]</sup>。将GTRL应用于建模用户历史线下交易的空间轨迹。首先,本文将学习到的用户空间表征进行可视化,如图5所示。



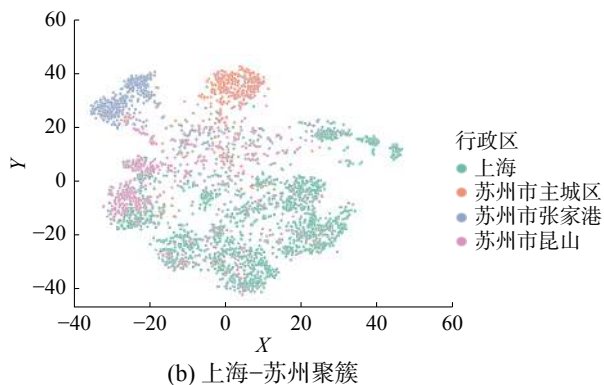


图5 GTRL-G学习到的用户表示示意

Fig. 5 Illustration of user embeddings learned by GTRL-G

从图5中,可以发现:1)每个城市聚簇明显,北京-天津、佛山-广州、苏州-上海、成都-重庆4组城市对也呈现聚簇状态,其中成都-重庆呈现较为明显的分离(可能是成都-重庆距离较远)。2)图(b)中:上海与苏州昆山的聚簇overlap较大(住昆山,在上海工作),苏州主城区和张家港则明显呈现独立聚簇。进一步地,在双十二的券抽奖场景上线了GTRL-G模型以帮助预测用户未来支付的位置,对比线上基线模型(MMoE),GTRL-G在整体核销率上可以带来3.31%的提升,在非消费券核销率上可以带来13.94%的提升。

## 6 结束语

本文提出了一个统一的用户表示学习框架——GTRL,可以有效地捕获用户行为在时空维度的语义。具体地,GTRL在时间建模上使用了函数式的时间编码和连续时间和上下文感知的图注意力网络用于在图上灵活地捕获高阶的结构化时序信息。与此同时,GTRL在空间建模上引入了层级化的地理编码和深度历史轨迹建模模块高效地刻画了用户的地理位置偏好。最后GTRL给出了一个统一且可解耦的优化方案来联合预测下一次交互,及相应的时间和地理位置。大量的公开数据集和工业数据集上的实验分别验证了GTRL相较学术界基线模型的优势,以及在实际业务场景中的有效性。在未来的工作中,将进一步致力于研究用户行为的稀疏性问题,将更加复杂的结构信息(如用户侧的社交网络信息和商品侧的知识图谱信息)集成到GTRL框架中。

## 参考文献:

- [1] DU Lun, WANG Yun, SONG Guojie, et al. Dynamic network embedding: an extended approach for skip-gram based network embedding[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18). Stockholm, Sweden, 2018: 2086–2092.
- [2] SINGER U, GUY I, RADINSKY K. Node embedding over temporal graphs[C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao, China, 2019: 4605–4612.
- [3] GUO Shengnan, LIN Youfang, FENG Ning, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[J]. Proceedings of the AAAI conference on artificial intelligence, 2019, 33(1): 922–929.
- [4] YANG Shuo, ZHANG Zhiqiang, ZHOU Jun, et al. Financial risk analysis for SMEs with graph-based supply chain mining[C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20). Yokohama, Japan, 2020: 4661–4667.
- [5] XU Da, RUAN Chuanwei, KORPEOGLU E, et al. Inductive representation learning on temporal graphs[C]//2020 International Conference on Learning Representations. Addis Ababa, Ethiopia, 2020: 1–19.
- [6] HUANG Hong, FANG Zixuan, WANG Xiao, et al. Motif-preserving temporal network embedding[C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20). Yokohama, Japan, 2020: 1237–1243.
- [7] MINH H Q, NIYOGI P, YAO Yuan. Mercer's theorem, feature maps, and smoothing[C]//Proceedings of the 19th Annual Conference on Learning Theory (COLT 2006). Pittsburgh, USA, 2006: 154–168.
- [8] LIAN Defu, WU Yongji, GE Yong, et al. Geography-aware sequential location recommendation[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Virtual Event, USA, 2020: 2009–2019.
- [9] OMI T, UEDA N, AIHARA K. Fully neural network based model for general temporal point processes[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada, 2019: 2120–2129.
- [10] ZHU Yu, LI Hao, LIAO Yikang, et al. What to do next: modeling user behaviors by time-LSTM[C]//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17). Melbourne, Australia, 2017: 3602–3608.
- [11] KUMAR S, ZHANG Xikun, LESKOVEC J. Predicting dynamic embedding trajectory in temporal interaction networks[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, USA, 2019: 1269–1278.
- [12] HAMILTON W L, YING R, LESKOVEC J. Inductive

- representation learning on large graphs[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA, 2017: 1025–1035.
- [13] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]//The 6th International Conference on Learning Representations. Vancouver, BC, Canada, 2018: 1–12.
- [14] NGUYEN G H, LEE J B, ROSSI R A, et al. Continuous-time dynamic network embeddings[C]//Companion Proceedings of the the Web Conference 2018. Lyon, France, 2018: 969–976.
- [15] LU Yuanfu, WANG Xiao, SHI Chuan, et al. Temporal network embedding with micro-and macro-dynamics [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing, China, 2019: 469–478.
- [16] SEO Y, DEFFERRARD M, VANDERGHEYNST P, et al. Structured sequence modeling with graph convolutional recurrent networks[C]//Processing of the 25th International Conference (ICONIP 2018). Siem Reap, Cambodia, 2018: 362–373.
- [17] GUO Huifeng, TANG Ruiming, YE Yunming, et al. DeepFM: a factorization-machine based neural network for CTR prediction[C]//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17). Melbourne, Australia, 2017: 1725–1731.
- [18] HIDASI B, KARATZOGLOU A, BALTRUNAS L, et al. Session-based recommendations with recurrent neural networks[C]//International Conference on Learning Representations. San Juan, Puerto Rico, 2016: 1–10.
- [19] ZHOU Guorui, ZHU Xiaoqiang, SONG Chenru, et al. Deep interest network for click-through rate prediction [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, United Kingdom, 2018: 1059–1068.
- [20] KANG Wangcheng, MCAULEY J. Self-attentive sequential recommendation[C]//2018 IEEE International Conference on Data Mining (ICDM). Singapore, Singapore, 2018: 197–206.
- [21] ZHOU Guorui, MOU Na, FAN Ying, et al. Deep interest evolution network for click-through rate prediction[J]. Proceedings of the AAAI conference on artificial intelligence, 2019, 33(1): 5941–5948.
- [22] MA Jiaqi, ZHAO Zhe, YI Xinyang, et al. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, United Kingdom, 2018: 1930–1939.
- [23] HU Binbin, ZHANG Zhiqiang, ZHOU Jun, et al. Loan default analysis with multiplex graph learning[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020: 2525–2532.
- [24] WANG D, ZHANG Z, ZHOU J, et al. Temporal-aware graph neural network for credit risk prediction[C]//Proceedings of the 2021 SIAM International Conference on Data Mining (SDM). 2021: 702–710.
- [25] HU B, ZHANG Z, SHI C, et al. Cash-out user detection based on attributed heterogeneous information network with a hierarchical attention mechanism[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Hawaii, USA, 2019: 946–953.

#### 作者简介:



周俊, 蚂蚁集团研究员, 主要研究方向为机器学习, 获得了第十届吴文俊人工智能科技进步奖一等奖(第一完成人)。发表学术论文 40 余篇。



张志强, 蚂蚁集团高级算法专家, 主要研究方向为图机器学习, 发表学术论文 20 余篇。



郑小林, 教授, 博士生导师, 主要研究方向为人工智能、智能电商、金融智能、服务计算。主持国家自然科学基金重点等多项。发表学术论文 50 余篇。