

基于双向准循环神经网络和注意力机制的业务流程 剩余时间预测方法

徐兴荣¹, 刘 聪^{1,2}, 李 婷¹, 郭 娜¹, 任崇广¹, 曾庆田²

(1. 山东理工大学计算机科学与技术学院, 山东淄博 255000; 2. 山东科技大学计算机科学与工程学院, 山东青岛 266590)

摘 要: 业务流程预测可以有效帮助企业进行流程控制和传递高质量服务,因此作为此类场景中的核心任务之一,业务流程剩余时间预测得到国内外学者的广泛关注. 当前,在利用深度学习技术对业务流程剩余时间进行预测时,大都采用传统长短期记忆循环神经网络,然而,由于长短期记忆循环神经网络在处理序列数据的过程中缺乏并行性且建模能力有限,使得预测准确度还有进一步提升空间. 因此,本文提出一种基于双向准循环神经网络和注意力机制的业务流程剩余时间预测方法. 首先,该方法以双向准循环神经网络构建剩余时间预测模型,并在预测模型中融入注意力机制增强双向准循环神经网络输出的特征信息. 其次,设计了一种基于不同长度轨迹前缀训练迭代策略,解决流程实例中不同长度轨迹前缀数量存在差异性的问题. 最后,提出一种基于 Word2vec 的事件表示学习方法,实现对同一轨迹且经常出现事件的相似性向量表示,从而达到提高剩余时间预测准确度的目的. 经在5个公开事件日志数据集上实验,本文方法与已有方法相比在预测准确度上平均提高近15%,模型训练时间平均缩短约26%.

关键词: 深度学习; 准循环神经网络; 业务流程; 剩余时间预测; 事件表示学习

中图分类号: TP391;TP311.5

文献标识码: A

文章编号: 0372-2112(2022)08-1975-10

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20211477

Business Process Remaining Time Prediction: An Approach Based on Bidirectional Quasi Recurrent Neural Network with Attention

XU Xing-rong¹, LIU Cong^{1,2}, LI Ting¹, GUO Na¹, REN Chong-guang¹, ZENG Qing-tian²

(1. School of Computer Science and Technology, Shandong University of Technology, Zibo, Shandong 255000, China;

2. College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266590, China)

Abstract: Business process prediction can effectively facilitate enterprises to control processes and deliver high-quality services. As one of the core tasks of process prediction, remaining time prediction has been widely concerned by scholars. Currently, traditional long short-term memory(LSTM) neural networks have been used to predict the remaining time of business process instances. However, due to the lack of parallelism and limited modeling ability of LSTM in processing sequence data, the accuracy of prediction has further room to improve. In this paper, the remaining time prediction method based on bidirectional quasi-recurrent neural network with attention is proposed. Firstly, this method uses the bidirectional quasi-recurrent neural network to build the prediction model, and adds the attention mechanism to the model enhances the characteristic information of the bidirectional quasi-recurrent neural network output. Secondly, a training iteration strategy based on different length trace prefixes is designed, which solves the problem of the difference in the number of trace prefixes of different lengths. Finally, event representation learning method is proposed, to achieve vectors representation of similarity to the same traces and frequent events, improves the accuracy of the remaining time prediction. Experiments on five public event log datasets show this method has improved the accuracy of prediction by an average of nearly 15%, and the average training time is reduced by about 26%, compared with the existing methods.

Key words: deep learning; quasi-recurrent neural network; business process; remaining time prediction; event representation learning

收稿日期: 2021-11-02; 修回日期: 2022-04-28; 责任编辑: 李勇锋

基金项目: 国家自然科学基金(No.61902222); 山东省泰山学者工程专项基金(No.ts20190936, No.tsqn201909109); 山东省自然科学基金优秀青年基金(No.ZR2021YQ45); 山东省高等学校青创科技计划创新团队项目(No.2021KJ031)

1 引言

预测型流程监控^[1]作为流程挖掘领域中重要的研究方向,其目标是对正在执行的业务流程实例进行预测分析,进一步准确预测在未来一段时间内可能出现的执行状态,例如某个实例的演变、下一个将要执行的活动、实例剩余时间^[2]等。

业务流程剩余时间预测^[3]是预测型流程监控中的一项主要预测任务,旨在对当前业务流程实例剩余执行时间进行预测。对于传统剩余时间预测方法,其事先在业务流程历史日志中挖掘变迁系统^[4]、随机 Petri 网^[5]等形式化流程模型,然后利用挖掘的模型对剩余时间进行预测。近年来,深度学习技术被广泛应用于剩余时间预测任务中,如长短期记忆(Long Short-Term Memory, LSTM)循环神经网络^[6],并取得较好的预测效果。

然而,基于深度学习技术的业务流程剩余时间预测还处于起步阶段,如下问题尚未得到解决:

(1)应用深度学习技术的前提是需要建立高质量的训练数据,就目前已有方法中,主要是以时间戳或时间先后次序索引来提取历史流程实例轨迹前缀序列,并将其作为深度学习技术的训练数据。但是,在整个业务流程中,不同种类的流程实例轨迹前缀之间存在着较大的差异性^[7],例如对于一个刚开始不久的流程实例与一个即将结束的流程实例而言,在对它们剩余时间进行预测时所考虑的因素是各不相同的。此外,不同长度的流程实例轨迹前缀数量呈不均匀分布,且短流程实例轨迹前缀数量在一般情况下要多于长流程实例轨迹前缀,倘若单纯的将二者混合以训练剩余时间预测模型,会被数量较多的短流程实例轨迹前缀影响,从而在一定程度上降低剩余时间预测准确度。

(2)已有方法大都以传统 LSTM 构建剩余时间预测模型,但由于 LSTM 对序列数据建模能力有限且在处理大规模序列数据时存在训练速度缓慢的问题,因此如何在提升剩余时间预测效果的同时,提高预测模型的训练速度,需进行深入研究。

基于目前工作存在的不足,本文提出一种基于双向准循环神经网络和注意力机制的业务流程剩余时间预测方法,主要创新点体现在:首先,设计了一种基于注意力机制的双向准循环神经网络剩余时间预测模型。该预测模型在利用门函数和记忆单元保证序列信息上下文相关性的同时,通过应用矩阵卷积运算实现对输入序列信息及门函数的有效处理,进而提高了预测模型的并行计算能力,加快了预测模型的训练时间。引入注意力机制是为了在训练过程中通过对不同位置的序列信息权重的计算,进一步提高重要特征对预测结果的影响。然后,设计了一种基于不同长度轨迹前缀

训练迭代策略,解决并提升了流程实例中不同长度轨迹前缀间的数量差异性和关联性。最后,提出一种基于 Word2vec 的事件表示学习方法,通过对事件日志中事件的高质量向量化表示,达到增强事件辨识度和提升剩余时间预测效果的目的。

2 相关工作

目前,业务流程剩余时间预测在流程挖掘领域中拥有较强的研究价值,它不仅可以起到优化业务系统性能的作用,而且还可以为管理者提供可靠的决策性依据。起初,Will van der Aalst 等^[3]利用构建的多层次抽象变迁系统记录事件日志中所有流程实例可能出现的状态,并为每个状态标注所属时间信息,以之进行剩余时间预测; Rogge-Solti 等^[5]通过对事件日志中随机 Petri 网的挖掘以模拟正在执行的流程实例,并根据相应结果对剩余时间进行预测。由于传统流程挖掘方法在面时剩余时间预测任务时存在较多局限性,导致这些方法在处理海量事件日志的过程中不能获得较优的预测效果。

随着机器学习技术的广泛应用,研究者将其融入到剩余时间预测任务中。Folino 等^[8]通过对不同类型数据的分析,提出一种基于聚类的剩余时间预测方法,该方法利用逻辑规则对聚类模型进行表示,从而帮助隐形上下文更好地适应事后分析^[9]; Polato 等^[10]在文献[3]的基础上提出数据感知变迁系统。它通过朴素贝叶斯分类方式为存在的状态节点建立模型,然后利用支持向量机进行剩余时间预测; Verenich 等^[11]将业务流程抽象为流程树,并在每个流程树节点上训练回归模型以预测剩余时间。

近几年,研究者逐渐开始利用深度学习技术进行业务流程剩余时间的预测工作。Evermann 等^[12]通过利用简单的 LSTM 和嵌入方式对变量进行分类训练,实现对业务流程剩余时间预测的目标; Nguyen 等^[13]在现有的基础上提出新特征表示方法,以提高目前深度学习模型的预测能力; 倪维健等^[14]在文献[2]的基础上通过引入注意力机制,进一步提高了业务流程剩余时间的准确度。随着时间推移,部分研究者通过将深度神经网络与相关分析技术融合研究出新结构变体。Khan 等^[15]利用记忆增强神经网络来解决业务流程剩余时间预测任务中复杂问题的过程分析; Pasquadibisceglie 等^[16]提出一种数据工程方法。该方法事先将业务流程中事件时间数据转换为抽象的空间图像结构,然后利用卷积神经网络进行处理,从而获得剩余时间预测结果; Taymouri 等^[17]采用生成式对抗网络在生成事件后缀的同时,预测剩余时间; Bukhsh 等^[18]提出一种高级表示方法。该方法通过使用 Transformer 模型克服传统预测模

型不能有效捕捉输入序列长期依赖关系的问题,实现对剩余时间的准确预测。

3 任务描述

本节将对业务流程剩余时间预测任务的相关知识进行详细介绍,并给出形式化描述。

3.1 基本知识

定义 1 事件是业务系统中一个活动的执行实例,可以用元组 $e=(a, id, T_{start}, T_{end}, P_1, \dots, P_m)$ 进行表示. 其中, a 为事件执行活动, id 为事件所属流程实例 ID, T_{start} 和 T_{end} 分别为事件执行的开始时间和结束时间, P_1, \dots, P_m 为事件含有的属性,包括:事件相关资源、执行代价等信息。

定义 2 轨迹是非空且有限的事件序列,可以用 $\sigma=(e_1, \dots, e_{|\sigma|})$ 进行表示. 其中,对于 $\forall 1 \leq i \leq |\sigma|, e_i$ 代表执行的第 i 个事件, $|\sigma|$ 代表轨迹中事件数量。

定义 3 轨迹前缀是轨迹 σ 中的前 k 个事件,可以用 $\sigma^{(k)}=(e_1, \dots, e_k)$ 进行表示,其中 $1 \leq k \leq |\sigma|$. 轨迹前缀剩余时间 $\text{RemainTime}(\sigma, k)=e_{|\sigma|} \cdot T_{end} - e_k \cdot T_{end}$.

定义 4 流程实例是整个业务流程的执行实例,可以用元组 $c=(Cid, \sigma, Z_1, \dots, Z_n)$ 进行表示. 其中, Cid 为流程实例的 ID, σ 为该流程实例中包含的轨迹, Z_1, \dots, Z_n 为流程实例含有的属性。

定义 5 事件日志是对业务流程历史执行情况的记录,是全部流程实例的集合,可以用 $L=\{c_1, \dots, c_m\}$ 进行表示. 其中, $|L|$ 表示事件日志 L 包含的流程实例 c 数目。

定义 6 对于业务系统,假若系统给定事件空间为 X ,则流程实例中轨迹及轨迹前缀空间可表示为 X^* ,剩余时间预测目标是通过建立的预测模型 $f=X^* \rightarrow R^+$ 预测流程实例轨迹前缀的剩余时间。

需说明,本文只利用流程实例轨迹信息开展剩余时间预测任务,故在无特殊情况下通过轨迹表示其所属的流程实例,而如何融入流程实例中其它属性信息将作为下一阶段主要的研究工作。

3.2 业务流程剩余时间预测任务描述

基于深度学习技术的业务流程剩余时间预测可分为训练和应用两个阶段. 其中,训练阶段是通过抽取事件日志中流程实例集合训练预测模型 f ; 应用阶段,是利用预测模型 f 对正在执行的流程实例(即轨迹前缀)剩余时间进行预测。

训练阶段,训练数据集的构建是在流程实例轨迹前缀长度的基础上进行,详细步骤如算法 1 所示,基本思想是:遍历事件日志 L 中每条轨迹 σ ,并在设定的轨迹前缀长度范围 $[k_{min}, k_{max}]$ 内进行截取,随之得到长度不一的轨迹前缀 $\sigma^{(k)}$ 以及对应的剩余时间 RemainTime

(σ, k) , 并作为训练集 D_k 中的训练样本,最后经组合获得全部长度的训练数据集 D 。

算法 1 构建训练数据集

输入:事件日志 L ,最小轨迹前缀长度 k_{min} ,最大轨迹前缀长度 k_{max}

输出:训练数据集 $D=\{D_{k_{min}} \cup \dots \cup D_{k_{max}}\}$

```

1.  $D_k \leftarrow \emptyset$ ;
2. FOR  $\sigma \in L$  DO
3. FOR  $k \in [k_{min}, k_{max}]$  DO
4. IF  $k < \text{len}(\sigma)$  THEN
5.  $D_k \leftarrow \{(\sigma^{(k)}, \text{RemainTime}(\sigma, k))\}$ 
6. END IF
7. RETURN  $D=\{D_{k_{min}} \cup \dots \cup D_{k_{max}}\}$ 
8. END DO
9. END DO

```

应用阶段,是利用生成的训练数据集学习剩余时间预测模型 f ,为防止学习过程中过拟合情况的出现,引入正则项 $\Omega(f)$,即:

$$f^* = \underset{f \in F}{\text{argmin}} \sum_{(\sigma, t) \in D} (f(\sigma) - t)^2 + \Omega(f) \quad (1)$$

4 基于双向准循环神经网络和注意力机制的业务流程剩余时间预测方法

基于双向准循环神经网络和注意力机制的业务流程剩余时间预测方法框架如图 1 所示。

首先,根据算法 1 将事件日志划分为不同长度轨迹前缀训练数据集;然后,抽取训练数据集中全部事件,利用基于 Word2vec 的事件表示学习方法对事件进行向量表示,并将其作为预测模型的输入. 同时,为进一步克服和提升不同长度轨迹前缀间的数量差异性和关联性,利用基于不同长度轨迹前缀训练迭代策略对预测模型进行迭代训练;最后,获得业务流程剩余时间预测结果. 经实验结果表明,将该方法应用于业务流程剩余时间预测任务后,在预测效果等方面展现出一定优势。

4.1 基于 Word2vec 的事件表示学习方法

目前,基于深度学习技术的业务流程剩余时间预测方法大都以独热(one-hot)编码的方式对事件进行表示,但因 one-hot 编码存在语义鸿沟等问题,使得剩余时间预测准确度还有待提升. 基于此,提出一种基于 Word2vec 的事件表示学习方法,以之对事件日志中事件进行表示学习,并将训练得到的事件向量作为预测模型的输入,从而实现提升事件表示质量和预测效果的目标。

由定义 1 可知,执行活动与执行时间是事件的关键组成部分,因此将其当作事件的主要标识,并进行向量化表示. 然而,由于事件执行时间是连续值,故需对其离散化处理,旨在将事件日志中事件对应活动可能出

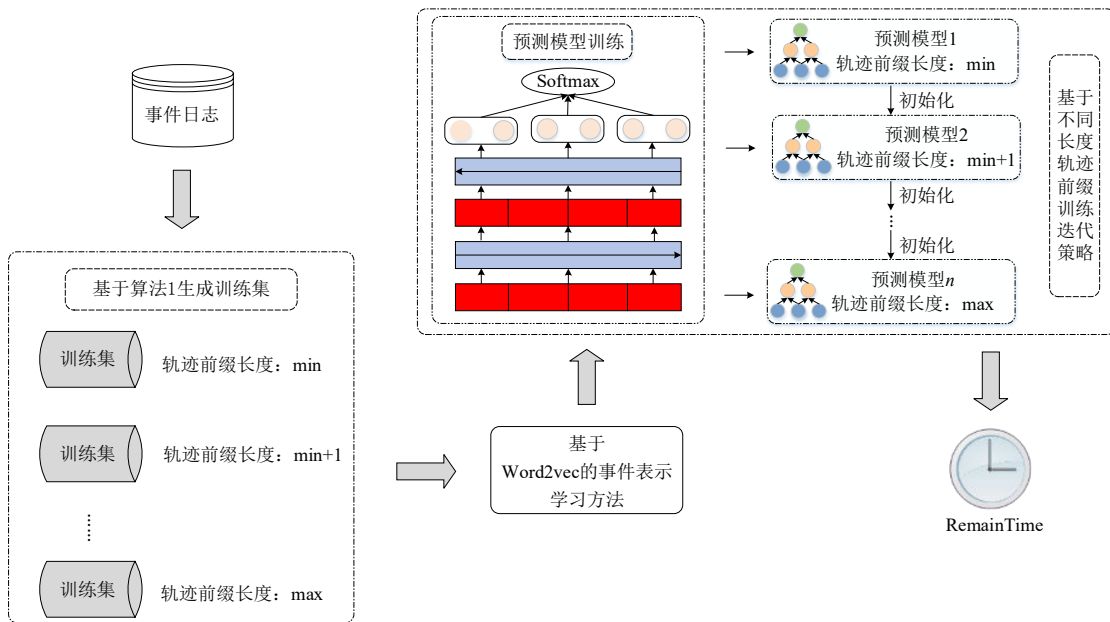


图1 基于双向准循环神经网络和注意力机制的业务流程剩余时间预测方法框架图

现的执行时间进行细化分类. 经离散化后事件 e 的执行时间为:

$$\tilde{t}(e) = \left\lfloor \frac{t(e) - t_{\text{MIN}}(\text{EndTime} - \text{StartTime})(e.a)}{t_{\text{MAX}}(\text{EndTime} - \text{StartTime})(e.a) - t_{\text{MIN}}(\text{EndTime} - \text{StartTime})(e.a)} \cdot N \right\rfloor \quad (2)$$

式中, $t(e)$ 代表事件 e 的连续执行时间, $t_{\text{MIN}}(\text{EndTime} - \text{StartTime})(e.a)$ 和 $t_{\text{MAX}}(\text{EndTime} - \text{StartTime})(e.a)$ 代表经计算整个事件日志所有活动 a 执行时间后选择的最短与最长执行时间, N 代表离散化后事件 e 执行时间所划分的数量 (本文 $N=10$). 由此, 事件表示学习的目标是将业务系统中全部活动与离散化执行时间进行拼接, 通过训练得到事件向量 $\theta^{<a, \tilde{t}(e)>} \in \mathbf{R}^d$.

对事件进行表示学习时, 主要利用自然语言处理领域中基于 Word2vec 中的 CBOW (continuous bag-of-word model) 词向量学习方法^[19], 基本思想为: 首先, 取第 t 时刻事件 e_t 的上下文 c 个事件 $\{e_{t-c}, \dots, e_{t-1}, e_{t+1}, \dots, e_{t+c}\}$ 进行 one-hot 编码 (实验环节 c 设置为 2), 并作为训练模型的输入, 此时输入向量通过与权重矩阵相乘及加权平均后获得隐藏层向量; 然后, 隐藏层向量随之与输出层之间权重矩阵相乘, 经 Softmax 操作后得到事件 e_t 的概率分布. 最终, 经训练获得的权重矩阵为事件 e_t 的表示向量.

4.2 基于注意力机制的双向准循环神经网络

由于 LSTM 在计算每个时间步时都依赖于前一个时间步的输出, 使得在处理大规模序列数据时无法进

行有效地并行计算, 进一步降低了对序列数据的建模能力. 而准循环神经网络 (Quasi-Recurrent Neural Network, QRNN)^[20] 通过一种交替卷积层的神经序列建模方法, 在输出时充分利用输入序列的顺序信息^[21], 实现了对序列数据跨时间步长的并行处理. 除此之外, QRNN 在 LSTM 结构的基础上进行简化, 只需计算遗忘门和输出门, 减少了网络计算量^[22]. 卷积层和池化层是 QRNN 主要组成部分, 其网络模型结构如图 2 所示. 其中, 卷积层的作用在于提取输入序列数据的特征信息并对门函数进行卷积处理, 池化层则是提取卷积层输出的特征信息, 以达到减少特征数目的目的.

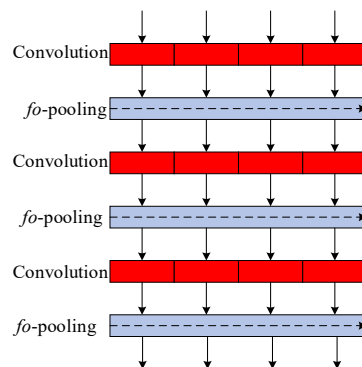


图2 QRNN 网络模型结构

假设长度为 T 的序列 $X=(x_1, x_2, \dots, x_T)$ 是 QRNN 卷积层输入序列, 首先以卷积方式经数量为 w 、宽度为 r 的滤波器在时间维度上对输入序列特征信息加以卷积处理, 使当前及过去时刻信息输入至 QRNN 单元, QRNN

单元结构如图 3 所示. 倘若在 t 时刻, 输入序列为 $X=(x_{(t-r+1)}, \dots, x_t)$, 则输出序列 z_t 、遗忘门 f_t 和输出门 o_t 计算过程为:

$$z_t = \tanh(W_z^1 x_{t-r+1} + W_z^2 x_{t-r+2} + \dots + W_z^r x_t) \quad (3)$$

$$f_t = \text{Sigmoid}(W_f^1 x_{t-r+1} + W_f^2 x_{t-r+2} + \dots + W_f^r x_t) \quad (4)$$

$$o_t = \text{Sigmoid}(W_o^1 x_{t-r+1} + W_o^2 x_{t-r+2} + \dots + W_o^r x_t) \quad (5)$$

式中, W_z 、 W_f 、 W_o 代表权重矩阵, Sigmoid 和 tanh 代表激活函数. 当滤波器宽度 r 增大时, 模型可以计算更多 N-Garm 特征, 实验环节滤波器宽度为 2.

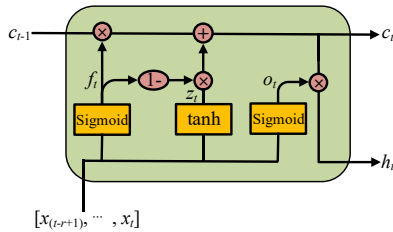


图 3 QRNN 单元结构图

池化层采用 fo -pooling 池化方式对第 t 时刻记忆单元状态 c_t 进行更新:

$$c_t = f_t \odot c_{t-1} + (1 - f_t) \odot z_t \quad (6)$$

最后, QRNN 输出 h_t 由第 t 时刻输出门 o_t 和记忆单元状态 c_t 计算而得:

$$h_t = o_t \odot c_t \quad (7)$$

式中, \odot 代表向量按位相乘运算, c_{t-1} 表示上一时刻记忆单元状态.

针对单向 QRNN 在处理序列数据时不能较好地顾及上下文信息之间的联系以及有效区分不同位置的重要程度, 本文使用带有注意力机制的双向准循环神经网络 (Bi-directional Quasi-Recurrent Neural Network with Attention, Att-Bi-QRNN) 构建剩余时间预测模型. 该预测模型包含如下关键模块:

(1) 基于双向 QRNN 的上下文编码

双向 QRNN 由一条正向 QRNN 和一条反向 QRNN 组成, 相较于单向 QRNN, 双向 QRNN 可以充分考虑轨迹前缀中不同事件间的相关性. 假设正向 QRNN 和反向 QRNN 获得的输出隐向量分别是 \vec{h}_t 和 \overleftarrow{h}_t , 则第 t 时刻上下文编码为 $H_t = [\vec{h}_t, \overleftarrow{h}_t]$.

(2) 基于注意力机制的轨迹前缀编码

获得第 t 时刻轨迹前缀上下文编码 H_t 后, 利用注意力机制计算整个轨迹前缀编码, 计算过程为:

$$v = \sum_t a_t \cdot H_t \quad (8)$$

式中, a_t 代表第 t 时刻上下文编码权重, 同时也进一步反映出轨迹前缀中不同事件对剩余时间预测任务的重要程度.

4.3 基于不同长度轨迹前缀训练迭代策略

针对训练单一剩余时间预测模型而言, 利用多个预测模型可以在一定程度上提高剩余时间预测的有效性和针对性, 但也面临着众多困难与挑战. 第一, 利用算法 1 将事件日志划分为多个训练集后, 其内部的轨迹前缀数量会进一步减少, 尤其是在现实事件日志, 轨迹前缀数量会根据其长度出现逐渐下降的现象, 导致长度较长的轨迹前缀数据量变得较少, 再加上深度学习技术需要大量的训练数据作为支撑, 倘若只在各个训练集中进行训练, 将会很难得到准确的预测模型; 第二, 由于长度较长的轨迹前缀涵盖了所有长度小于其本身的轨迹前缀, 因此长度较短的轨迹前缀预测模型可以为长度较长的轨迹前缀预测模型提供有用的预测信息, 而仅在各长度不一的轨迹前缀训练集上独立训练预测模型, 忽视了不同长度轨迹前缀之间的关联性.

基于此, 设计了一种基于不同长度轨迹前缀训练迭代策略, 详细步骤如算法 2 所示. 首先, 设定轨迹前缀长度范围 $[\min, \max]$, 其中 \min 表示事件日志中最短轨迹前缀长度, \max 表示过滤掉不频繁长轨迹前缀后的最大轨迹前缀长度. 而过滤掉不频繁长轨迹前缀主要原因在于不频繁出现的长轨迹前缀在整个事件日志中数据量极少, 因此在对其进行训练时很容易造成过拟合的情况, 进而影响剩余时间预测效果. 然后, 利用算法 1 生成多个训练集 $D_{k, \min}, \dots, D_{k, \max}$, 并在长度短 (\min) 且数量多的轨迹前缀训练集 $D_{k, \min}$ 上训练剩余时间预测模型 $F_{k, \min}$, 当预测模型 $F_{k, \min}$ 训练完毕, 将其训练参数共享于长轨迹前缀预测模型, 并进行迭代训练, 直至得到最长 (\max) 轨迹前缀的剩余时间预测模型 $F_{k, \max}$. 利用此策略不仅可以充分使用不同长度轨迹前缀内在关联性, 而且还可以解决长轨迹前缀数据量较少的问题, 从而提升剩余时间预测准确性.

算法 2 基于不同长度轨迹前缀训练迭代策略

输入: 不同长度的轨迹前缀训练集 $D_{k, \min} \cup \dots \cup D_{k, \max}$, 迭代次数 n

输出: 不同长度的轨迹前缀剩余时间预测模型 $\{F_{k, \min}, F_{k, \min+1}, \dots, F_{k, \max}\}$

```

1. FOR  $i \in [1, n]$  DO
2.  $F_{k, \min} \leftarrow \text{Train}(D_{k, \min})$ 
3. FOR  $k \in [k_{\min+1}, k_{\max}]$  DO
4.  $F_{(k, \min+1, \dots, k, \dots, \max)} \leftarrow \text{Train}(D_{(k, \min+1, \dots, k, \dots, \max)}, F_{(k, \min, \dots, k-1, \dots, \max-1)})$ 
5. RETURN  $\{F_{k, \min}, F_{k, \min+1}, \dots, F_{k, \max}\}$ 
6. END DO
7. END DO

```

5 实验结果与分析

本节对提出的基于双向准循环神经网络和注意力机制的业务流程剩余时间预测方法进行实验探究与分析.

实验基于 PC Intel Core i5-10400F 2.90 GHz、NVIDIA GeForce RTX 2070 SUPER 环境,使用 Pytorch 框架实现.

5.1 实验数据集

实验使用 5 个来自 4TU Center for Research 平台公开的事件日志数据集,分别为:BPIC_2012_A、BPIC_2012_O、BPIC_2012_W、Helpdesk、Hospital_Billing,其中:

BPIC_2012_A:是记录荷兰财政机构贷款申请审批流程的事件日志;

BPIC_2012_O:同 BPIC_2012_A,且仅记录了贷款审批状态的变更信息;

BPIC_2012_W:同 BPIC_2012_A,且仅记录了贷款审批资源状态的变更信息;

Helpdesk:是记录意大利软件公司票务管理系统后台流程的事件日志;

Hospital_Billing:是记录某地区医院 EPR 系统中出院结算流程的事件日志;

上述 5 个事件日志数据集的基本信息统计情况如表 1 所示.

表 1 事件日志数据集基本信息统计

数据集	轨迹数量	事件数量	活动数量	轨迹最大长度	轨迹最小长度
BPIC_2012_A ^[23]	13 087	73 022	10	10	3
BPIC_2012_O ^[23]	5 015	41 728	7	39	4
BPIC_2012_W ^[23]	9 658	147 450	6	153	1
Helpdesk ^[24]	3 804	13 710	9	14	1
Hospital_Billing ^[25]	100 000	451 359	18	217	1

表 2 业务流程剩余时间预测方法对比实验结果

方法	BPIC_2012_A	BPIC_2012_O	BPIC_2012_W	Helpdesk	Hospital_Billing
LSTM ^[14]	3.588	7.993	8.021	3.542	42.050
Bi-LSTM ^[14]	4.600	8.407	7.221	2.952	41.515
Att-Bi-LSTM ^[14]	3.895	7.324	6.153	2.677	36.691
Trans-Att-Bi-LSTM ^[14]	3.489	5.858	5.826	3.357	33.201
GRU ^[14]	4.888	7.394	7.753	3.425	47.400
Bi-GRU ^[14]	4.461	8.144	7.334	3.222	37.157
Att-Bi-GRU ^[14]	3.512	7.306	6.338	3.303	33.201
Trans-Att-Bi-GRU ^[14]	3.438	5.863	5.821	3.299	32.187
QRNN	3.543	6.779	7.516	3.199	36.611
Bi-QRNN	3.992	6.757	7.128	2.884	35.248
Att-Bi-QRNN	3.200	6.276	6.007	2.626	32.101
Trans-Att-Bi-QRNN	2.373	5.158	5.275	2.423	31.436

对比 QRNN、LSTM、GRU 三种预测模型(QRNN vs LSTM vs GRU、Bi-QRNN vs Bi-LSTM vs Bi-GRU、Att-Bi-QRNN vs Att-Bi-LSTM vs Att-Bi-GRU),QRNN 整体预测效果在 5 个事件日志数据集中优于 LSTM 和 GRU,因此

5.2 评价指标

利用平均绝对误差(Mean Absolute Error, MAE)作为剩余时间预测任务的总体评价指标,通过计算轨迹前缀剩余时间真实值和预测值之间差值的绝对值评判剩余时间预测准确度. MAE 值越低,表明预测的剩余时间越准确.

$$MAE(f) = \sum_{(\sigma^{(k)}, \text{RemainTime}(\sigma, k)) \in D} |f(\sigma^{(k)}) - \text{RemainTime}(\sigma, k)| \quad (9)$$

式中, $\sigma^{(k)}$ 代表轨迹前缀, $f(\sigma^{(k)})$ 代表轨迹前缀剩余时间的预测值, $\text{RemainTime}(\sigma, k)$ 代表轨迹前缀剩余时间的真实值.

实验通过 5 折交叉验证的方式进行有效评估,故将数据集随机划分为 5 等份,每次将 4 份数据选作训练集,1 份选作测试集,重复 5 次实验,最后把每次实验 MAE 值的平均值作为最终评估结果.

5.3 实验结果

5.3.1 业务流程剩余时间预测准确度对比分析

本文在 5 个事件日志数据集上进行了对比实验,实验结果如表 2 所示. 为验证以 QRNN 为基准的预测模型(即 QRNN、引入双向机制的 Bi-QRNN、引入双向机制和注意力机制的 Att-Bi-QRNN)在剩余时间预测任务上的有效性,将与文献[14]提出的基于注意力双向循环神经网络剩余时间预测方法进行比较,同时在预测模型顺序结构和隐层节点数的设置上与对比方法^[14]保持一致. 此外, Batch_size 设置为 512, 迭代 150 轮,使用 Adam 优化算法.

可认为 QRNN 比 LSTM、GRU 更加适合剩余时间预测任务.

引入双向机制后, Bi-QRNN 相比于 QRNN 在 5 个事件日志数据集中获得较低 MAE 值,说明双向机制可以

有效提升事件之间的关联性,达到提高预测效果的目的.而在 Bi-QRNN 基础上引入注意力机制后,Att-Bi-QRNN 在各事件日志数据集上 MAE 值进一步降低,表明注意力机制经对不同位置事件信息权重的计算,有效区分出不同事件对剩余时间预测任务的影响,进而获取关键事件信息.此外,在同级相比的条件下,以 QRNN 为基准的预测模型 MAE 值在 5 个事件日志数据集中比以 LSTM 和 GRU 为基准的预测模型平均下降近 10% (Bi-QRNN vs Bi-LSTM、Bi-GRU) 和 4% (Att-Bi-QRNN vs Att-Bi-LSTM、Att-Bi-GRU),说明将 Att-Bi-QRNN 预测模型应用于业务流程剩余时间预测任务具有一定的可行性.

相比于文献[14]以 LSTM 和 GRU 为基准的最佳预测方法,本文方法(即 Trans-Att-Bi-QRNN,以 Att-Bi-QRNN 作为基础预测模型,融入基于不同长度轨迹前缀训练迭代策略和基于 Word2vec 的事件表示学习方法)

在 5 个不同的事件日志数据集中 MAE 值平均下降近 15%,表明本文提出的剩余时间预测方法在剩余时间预测任务中具备优越性.

5.3.2 基于不同长度轨迹前缀训练迭代策略分析

在实验中,以带有注意力机制的双向 QRNN 为基准模型(即 Att-Bi-QRNN),并在此基础上训练不同长度轨迹前缀预测模型,具体训练方式如下:

(1)在不同长度轨迹前缀数据集上单独训练剩余时间预测模型,以 Diff-Att-Bi-QRNN 表示;

(2)在不同长度轨迹前缀数据集上利用第 4.3 节基于不同长度轨迹前缀训练迭代策略训练多个剩余时间预测模型,以 Trans-Att-Bi-QRNN 表示;

两种不同方法在 5 个事件日志数据集上所获得的业务流程剩余时间预测准确度(MAE)如图 4(a)~(e)所示.

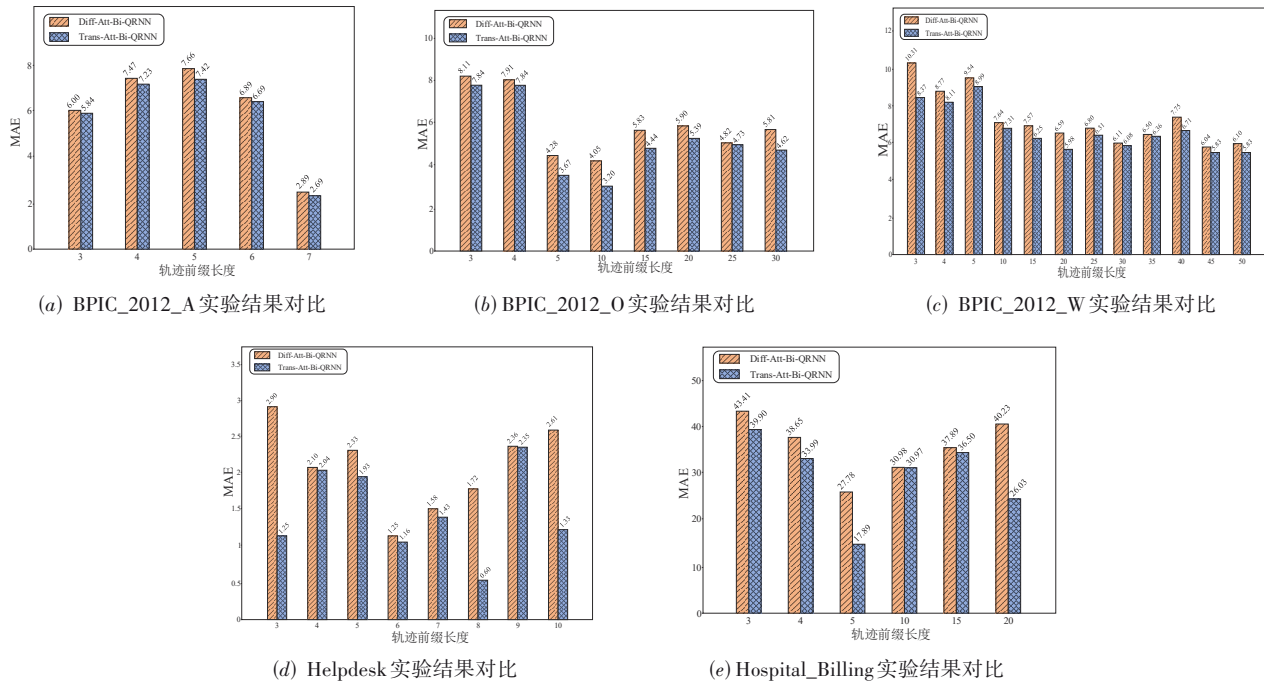


图 4 基于不同长度轨迹前缀训练迭代策略效果对比图

由图 4 实验结果可知,Trans-Att-Bi-QRNN 在各事件日志数据集中获得的 MAE 值优于 Diff-Att-Bi-QRNN,同时,对于 BPIC_2012_O、BPIC_2012_W、Helpdesk、Hospital_Billing 事件日志数据集,Trans-Att-Bi-QRNN 在长轨迹前缀上展现出一定优势,说明基于不同长度轨迹前缀训练迭代策略可解决不同长度轨迹前缀之间数量存在差异性的问题,提升了不同长度轨迹前缀的关联性.

为进一步说明 Trans-Att-Bi-QRNN 在剩余时间预测任务中的有效性,将 Trans-Att-Bi-QRNN 及 Diff-Att-Bi-QRNN 在不同长度轨迹前缀数据集下取得的 MAE 值进

行加权平均,结果如表 3 所示.

通过表 3 结果可知,Trans-Att-Bi-QRNN 在各事件日

表 3 基于不同长度轨迹前缀训练迭代策略实验结果

	Diff-Att-Bi-QRNN	Trans-Att-Bi-QRNN
BPIC_2012_A	3.880	3.739
BPIC_2012_O	5.838	5.219
BPIC_2012_W	7.475	6.844
Helpdesk	2.108	1.511
Hospital_Billing	36.488	30.878

志数据集中 MAE 值相比于 Diff-Att-Bi-QRNN 降低近 13%,表明在引入基于不同长度轨迹前缀训练迭代策略后,其对剩余时间预测效果呈现出正影响状态,印证了基于不同长度轨迹前缀训练迭代策略确有助于提升剩余时间预测准确度。

5.3.3 基于 Word2vec 的事件表示学习方法分析

在实验中,以 Trans-Att-Bi-QRNN 为基础模型,分别将基于 Word2vec 的事件表示方法得到的向量(表示为 Trans-Att-Bi-QRNN-Vector)及传统 one-hot 编码(表示为 Trans-Att-Bi-QRNN-NoVector)得到的向量作为预测模型的输入,以验证事件表示学习方法对剩余时间预测效果的影响。

图 5 展示了两种不同方法的 MAE 值。可以看出,Trans-Att-Bi-QRNN-Vector 在 5 个事件日志数据集上的

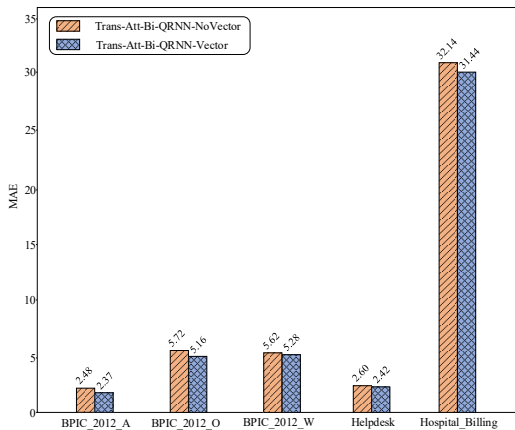


图 5 基于 Word2vec 的事件表示学习方法效果对比图

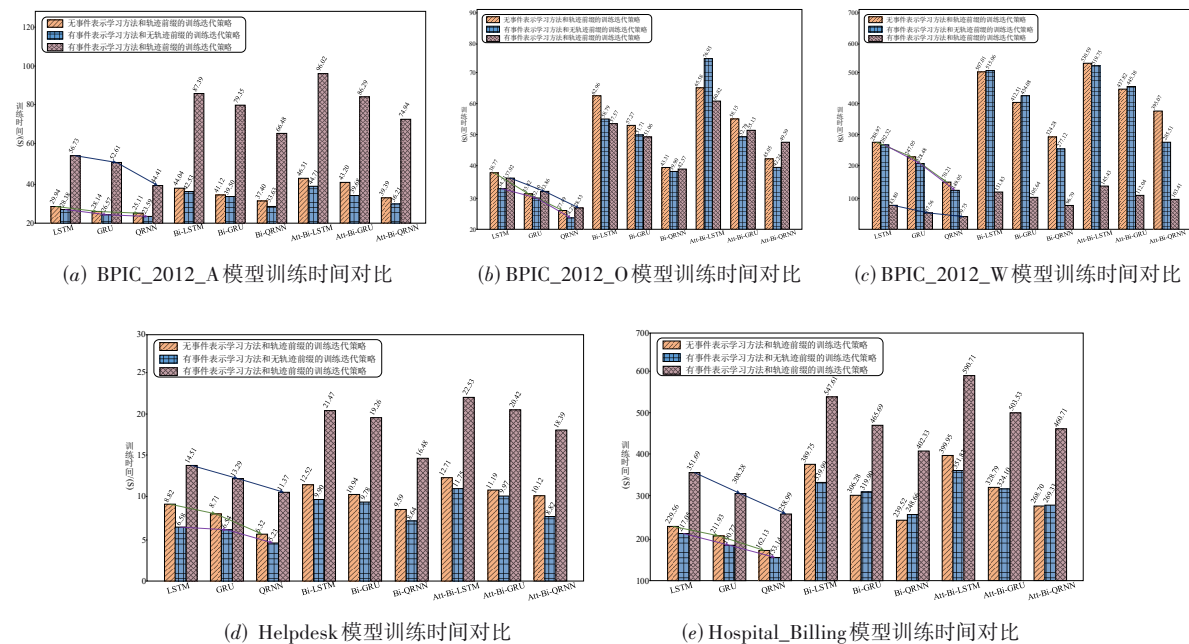
MAE 值比 Trans-Att-Bi-QRNN-NoVector 的 MAE 值低 3%,因此可以认为基于 Word2vec 的事件表示学习方法对提升业务流程剩余时间预测效果具有积极的作用。

5.3.4 时间性能对比分析

本节在 5 个事件日志数据集上对各预测模型的训练时间(单位:秒)进行了对比实验,实验结果如图 6(a)~(e)所示。在实验过程中,预测模型在不同方法下采用相同模型参数:迭代次数为 150、学习率为 0.01、Batch_size 为 512、输入事件向量维度为 3、神经元隐向量维度为 5。

由实验结果可知,在有无基于 Word2vec 的事件表示学习方法和基于不同长度轨迹前缀训练迭代策略的条件下,对于 BPIC_2012_A,以 QRNN 为基准的预测模型训练时间比以 LSTM 和 GRU 为基准的预测模型训练时间平均缩短 11%~18%;对于 BPIC_2012_O,以 QRNN 为基准的预测模型训练时间比以 LSTM 和 GRU 为基准的预测模型训练时间平均缩短 19%~29%;对于 BPIC_2012_W,以 QRNN 为基准的预测模型训练时间比以 LSTM 和 GRU 为基准的预测模型训练时间平均缩短 20%~35%;对于 Helpdesk,以 QRNN 为基准的预测模型训练时间比以 LSTM 和 GRU 为基准的预测模型训练时间平均缩短 13%~20%;对于 Hospital_Billing,以 QRNN 为基准的预测模型训练时间比以 LSTM 和 GRU 为基准的预测模型训练时间平均缩短了约 26%,说明本文使用的预

综上所述,以 QRNN 为基准的预测模型相比于以 LSTM 和 GRU 为基准的预测模型在 5 个事件日志数据集中训练速度平均缩短了约 26%,说明本文使用的预



(a) BPIC_2012_A 模型训练时间对比

(b) BPIC_2012_O 模型训练时间对比

(c) BPIC_2012_W 模型训练时间对比

(d) Helpdesk 模型训练时间对比

(e) Hospital_Billing 模型训练时间对比

图 6 预测模型训练时间对比图

测模型具备更快的系统运行速度,从而使预测模型的整体性能得到有效提升.

综合分析 5 个事件日志数据集在各项实验中的结果可知,相比于 LSTM、Bi-LSTM、Att-Bi-LSTM、GRU、Bi-GRU、Att-Bi-GRU,所提方法无论是在剩余时间预测效果,还是在预测模型训练速度,都取得较优的实验结果,说明该方法在剩余时间预测任务中存在一定优势.

另外,表 4 给出 9 种剩余时间预测模型的参数规模,可见在同级相比的条件下,QRNN 预测模型参数数量明显少于 LSTM 和 GRU 预测模型,说明 QRNN 预测模型在没有明显增加网络参数的条件下进一步提升了业务流程剩余时间预测性能.

表 4 各预测模型参数规模

预测模型	参数量/M
LSTM	0.007 73
Bi-LSTM	0.013 83
Att-Bi-LSTM	0.014 03
GRU	0.006 23
Bi-GRU	0.010 83
Att-Bi-GRU	0.011 03
QRNN	0.003 83
Bi-QRNN	0.004 23
Att-Bi-QRNN	0.004 43

6 结论

本文提出一种基于双向准循环神经网络和注意力机制的业务流程剩余时间预测方法.该方法采用处理序列数据更佳的准循环神经网络,有效克服传统 LSTM 因缺少并行计算能力而导致训练缓慢的问题.同时,通过引入双向机制和注意力机制进一步提升了上下文信息之间的联系以及重要特征对预测结果的影响.此外,经使用基于不同长度轨迹前缀训练迭代策略和基于 Word2vec 的事件表示学习方法,不仅加强了业务流程剩余时间预测的针对性,而且还为预测模型提供了高质量的输入事件向量.由实验结果表明,本文方法在预测效果以及训练时间上明显优于已有研究.

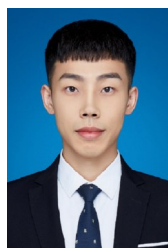
然而,所提方法只是针对流程实例中轨迹信息、事件信息进行剩余时间预测,而流程实例中还包含执行者、资源等额外属性,因此可在当前工作的基础上增加其他属性,以验证多属性对剩余时间预测效果的影响.此外,作为一种深度学习方法,该方法虽然获得较优的预测效果,但可解释性较差,即使部分研究者开始着手于此方面的研究,但这些方法对深度学习技术应用于剩余时间预测任务的解释效果并不是很理想,因此如何提高剩余时间预测任务的可解释性,是今后一项重要的研究工作.

参考文献

- [1] TEINEMAA I, DUMAS M, ROSA M L, et al. Outcome-oriented predictive process monitoring: Review and benchmark[J]. *ACM Transactions on Knowledge Discovery from Data*, 2019, 13(2): 17.
- [2] TAX N, VERENICH I, LA ROSA M, et al. Predictive business process monitoring with LSTM neural networks [C]//*International Conference on Advanced Information Systems Engineering*. Aachen: Springer, 2017: 477-492.
- [3] VAN DER AALST W P M, SCHONENBERG M H, SONG M. Time prediction based on process mining[J]. *Information Systems*, 2011, 36(2): 450-475.
- [4] 赵海燕, 李帅标, 陈庆奎, 等. 面向业务过程的时间预测方法[J]. *小型微型计算机系统*, 2019, 40(2): 42-48.
ZHAO Hai-yan, LI Shuai-biao, CHEN Qing-kui, et al. Method of time prediction for business process[J]. *Journal of Chinese Computer Systems*, 2019, 40(2): 42-48. (in Chinese)
- [5] ROGGE-SOLTI A, WESKE M. Prediction of business process durations using non-Markovian stochastic petri nets [J]. *Information Systems*, 2015, 54: 1-14.
- [6] NAVARIN N, VINCENZI B, POLATO M, et al. LSTM networks for data-aware remaining time prediction of business process instances[C]//*IEEE Symposium Series on Computational Intelligence*. Hawaii: IEEE, 2017: 1-7.
- [7] LEONTJEVA A, CONFORTI R, DI FRANCESCO MARIANO C, et al. Complex symbolic sequence encodings for predictive monitoring of business processes[C]//*International Conference on Business Process Management*. Dubai: Springer, 2016: 297-313.
- [8] FOLINO F, GUARASCIO M, PONTIERI L. Mining predictive process models out of low-level multidimensional logs[C]//*International Conference on Advanced Information Systems Engineering*. Luxembourg: Springer, 2014: 533-547.
- [9] 崔亮. 基于机器学习的业务流程系统的预测[D]. 北京: 北京邮电大学, 2019.
- [10] POLATO M, SPERDUTI A, BURATTIN A, et al. Time and activity sequence prediction of business process instances[J]. *Computing*, 2018, 100(9): 1005-1031.
- [11] VERENICH I, NGUYEN H, LA ROSA M, et al. White-box prediction of process performance indicators via flow analysis[C]//*Proceedings of the 2017 International Conference on Software and System Process*. Paris: ACM, 2017: 85-94.
- [12] EVERMANN J, REHSE J, FETTKE P. A deep learning

- approach for predicting process behaviour at runtime[C]// Business Process Management Workshops. Rio de Janeiro: Springer, 2016: 327-388.
- [13] NGUYEN A, CHATTERJEE S, WEINZIERL S, et al. Time matters: Time-aware LSTMs for predictive business process monitoring[C]//International Workshop on Leveraging Machine Learning in Process Mining. Padua: ICPM, 2020: 1-12.
- [14] 倪维健, 孙宇健, 刘彤, 等. 基于注意力双向循环神经网络的业务流程剩余时间预测方法[J]. 计算机集成制造系统, 2020, 26(6): 1564-1572.
NI Wei-jian, SUN Yu-jian, LIU Tong, et al. Business process remaining time prediction using bidirectional recurrent neural networks with attention[J]. Computer Integrated Manufacturing Systems, 2020, 26(6): 1564-1572. (in Chinese)
- [15] KHAN A, LE H, DO K, et al. DeepProcess: Supporting business process execution using a MANN-based recommender system[EB/OL]. (2018) [2021]. <https://arxiv.org/abs/1802.00938>.
- [16] PASQUADIBISCEGLIE V, APPICE A, CASTELLANO G, et al. Using convolutional neural networks for predictive process analytics[C]//International Conference on Process Mining, Aachen: ICPM, 2019: 29-136.
- [17] TAYMOURI F, ROSA M L. Encoder-decoder generative adversarial nets for suffix generation and remaining time predication of business process models[EB/OL]. (2020) [2021]. <https://arxiv.org/abs/2007.16030>.
- [18] BUKHSH Z A, SAEED A, DIJKMAN R M. ProcessTransformer: Predictive business process monitoring with transformer network[EB/OL]. (2021) [2021]. <https://arxiv.org/abs/2104.00721>.
- [19] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[EB/OL]. (2013)[2021]. <https://arxiv.org/abs/1301.3781>.
- [20] BRADBURY J, MERITY S. Quasi-recurrent neural networks[EB/OL]. (2019) [2021]. <https://arxiv.org/abs/1611.01576>.
- [21] 王栋, 李业刚, 张晓, 等. 基于准循环神经网络的中文命名实体识别[J]. 计算机工程与设计, 2020, 41(7): 2038-2043.
WANG Dong, LI Ye-gang, ZHANG Xiao, et al. Chinese name entity recognition based on quasi-recurrent neural networks[J]. Computer Engineering and Design, 2020, 41(7): 2038-2043. (in Chinese)
- [22] 时云龙, 袁文浩, 胡少东, 等. 一种用于实时语音增强的卷积准循环网络[J]. 西安电子科技大学学报, 2022, 49(3): 1-8.
Shi Yun-long, Yuan Wen-hao, Hu Shao-dong, et al. Convolutional quasi-recurrent network for real-time speech enhancement[J]. Journal of Xidian University, 2022, 49(3): 1-8. (in Chinese)
- [23] VAN DONGEN. Bpi2012[OL]. <https://doi.org/10.4121/uuid:3926db30-f712-4394-aebc-75976070e91f>.
- [24] IYA VERENICH. Helpdesk[OL]. <https://mendeley.figshare.com/articles/dataset/Helpdesk/8951084>.
- [25] FELIX MANNHARDT. Hospital Billing[OL]. https://data.4tu.nl/articles/dataset/Hospital_Billing_-_Event_Log/12705113.

作者简介



徐兴荣 男, 1995年5月出生于山东省济南市. 现为山东理工大学计算机科学与技术学院硕士研究生. 主要研究方向为流程挖掘.
E-mail: Nick_xu0522@163.com



刘聪(通讯作者) 男, 1990年8月出生于山东省淄博市. 现为山东理工大学计算机科学与技术学院教授、博士生导师. 主要研究方向为流程挖掘.
E-mail: liucongchina@sdust.edu.cn



李婷 女, 1996年11月出生于山东省济宁市. 现为山东理工大学计算机科学与技术学院硕士研究生. 主要研究方向为流程挖掘.
E-mail: Ltingoo@163.com

郭娜 女, 1996年7月出生于山东省淄博市. 现为山东理工大学计算机科学与技术学院博士研究生. 主要研究方向为流程挖掘.
E-mail: guona_7@163.com

任崇广 男, 1982年7月出生于山东省淄博市. 现为山东理工大学计算机科学与技术学院教授、硕士生导师. 主要研究方向为智能装备.
E-mail: renchg@sdut.edu.cn

曾庆田 男, 1976年1月出生于山东省潍坊市. 现为山东科技大学计算机科学与工程学院教授、博士生导师. 主要研究方向为流程挖掘.
E-mail: qtzeng@163.com