基于局部数据增强动态图的事件预测

潘磊, 刘欣, 陈君益, 程章桃, 刘乐源, 周帆

引用本文

潘磊, 刘欣, 陈君益, 程章桃, 刘乐源, 周帆. 基于局部数据增强动态图的事件预测[J]. 计算机科学, 2024, 51(3): 118-127.

PAN Lei, LIU Xin, CHEN Junyi, CHENG Zhangtao, LIU Leyuan, ZHOU Fan. Event Prediction Based on Dynamic Graph with Local Data Augmentation [J]. Computer Science, 2024, 51(3): 118-127.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.221200054

基于局部数据增强动态图的事件预测

潘 磊[1] 刘 欣[2] 陈君益[2] 程章桃[2] 刘乐源[2] 周 帆[2,3]
1 中国电子科技集团公司第十研究所 成都 610036
2 电子科技大学信息与软件工程学院 成都 610054
3 喀什地区电子信息产业技术研究院 新疆 喀什 844099
(mapan.lei@163.com)

摘 要 事件指在真实世界中特定的时间和地点发生的与特定主题相关的活动,例如,社会动乱、暴恐袭击、自然灾害和传染病流行等事件会对国家安全和人民群众的生活产生重大威胁。如果能对此类事件的发生进行有效预测,将最大程度地减少负面事件带来的影响或最大化正面事件带来的利益。关于事件的研究中,准确预测事件仍然是一个非常具有挑战性的任务。文中提出了一种基于图注意力网络的事件预测方法LAT-GAT(Local Augmented Temporal-GAT),该方法使用条件变分编码器,在所构建的事件图中对目标节点的邻居节点生成新的特征样本,与节点原有特征进行拼合,形成新的节点特征,实现了对事件的传播结构的利用;另外,LAT-GAT还考虑了历史事件发生的时间先后顺序,将网络在上一时间点的输出结果集成到当前时间的特征中,从而实现了对事件传播时间特性的利用。最后,在泰国、印度、埃及和俄罗斯这4个国家真实事件数据集上,与多种代表性基线方法进行了对比实验。实验结果表明,LAT-GAT在4个国家数据上的F1评分都优于基线方法;在泰国、俄罗斯和印度数据集上召回率优于基线方法;在泰国、埃及和印度数据集上也获得了最高的准确率。还通过消融实验考察了模型参数对最终结果的影响。

关键词:事件预测;图注意力网络;动态图;条件变分编码器;数据增强
中图分类号 TP391

Event Prediction Based on Dynamic Graph with Local Data Augmentation

PAN Lei[1],LIU Xin[2],CHEN Junyi[2],CHENG Zhangtao[2],LIU Leyuan[2] and ZHOU Fan[2,3]
1 No.10 Research Institute of China Electronics Technology Group Corporation,Chengdu 610036,China
2 School of Information and Software Engineering,University of Electronic Science and Technology of China,Chengdu 610054,China
3 Information Industry Technology Research Institute of Kashi Region,Kashi,Xinjiang 844099,China

Abstract Event refers to activities that occur in real world at specific time and places. For instance,unrest,violent terrorist attacks,natural disasters and the spread of infectious diseases,will bring great threats and losses to national security and human life. If the occurrence of such events could be predicted more precisely and effectively,the impact of negative events will be minimized,and it is possible to maximize the benefits of the positive events. It is still a very challenging task to predict events accurately. An event prediction method named local augmented temporal-GAT(LAT-GAT) based on graph attention network is proposed in this paper. It uses conditional variational encoders to generate new features,which will be concatenated with the original features to new one,based on neighbors of the current node. With this approach,our model can utilize the propagation structure of events. In addition,the chronological order of events occurrence is considered by our model. The feature of events in last time point is integrated into the output of the neural network in current time. The temporal property of event propagation is exploited through temporal data integration. And finally,the proposed method is compared with a number of representative baseline methods on the real-world datasets,including Thailand,India,Egypt and Russia. The results show that LAT-GAT has the best F1 scores in all datasets. The recall of our model exceeds that of any other baseline methods in the datasets of Thailand,Russia and India. In Thailand,Egypt and India,our model achieves the best precision. Ablation experiments are also conducted to investigate

到稿日期:2022-12-08 返修日期:2023-04-07
基金项目:国家自然科学基金(62176043,62072077);四川省自然科学基金(2022NSFC0505);四川省科技计划(2022YFSY0006);厅市共建智能终端四川省重点实验室开放课题(SCITLAB-20006)
This work was supported by the National Natural Science Foundation of China(62176043,62072077),Natural Science Foundation of Sichuan Province,China(2022NSFC0505),Sichuan Science and Technology Program(2022YFSY0006) and Open Project of Intelligent Terminal Key Laboratory of Sichuan Province(SCITLAB-20006).
通信作者:刘乐源(leyuanliu@uestc.edu.cn)

the influence of the model parameters on the final results.

Keywords　Event prediction，Graph attention network，Dynamic graph，Conditional variational auto-encoder，Data augmentation

1　引言

事件(Event)指在真实世界中特定的时间和地点发生的与特定主题相关的活动[1]。其指代范围较为宽泛，如社会动乱、暴恐袭击和传染病流行等社会事件；地震、水灾等自然灾害事件；系统故障、传感器失效等计算机系统故障事件；人类日常行为等个人事件。其中社会事件对人类发展、国家治理和居民日常生活都会产生重大影响，如果能够对未来社会事件进行准确的预测，将最大程度地减少负面事件带来的影响或最大化正面事件带来的利益。例如：缩短自然灾害发生时的应急响应的时间，最大程度地挽救人民的生命财产；精准防控传染性疾病，第一时间阻断疫情传播等。目前事件分析技术大体可分为3类：事件抽取、事件检测和事件预测。从其研究对象所在的时间线来说，前两者的研究对象是过去和现在已经发生的事件，是一种对过去事实进行回溯的分析；而事件预测的研究对象所处的时间线是未来可能会发生的事件，也是本文的研究目标。

虽然事件预测的重要性显而易见，但过去人类对事件发生的原因及其内在发展机制缺乏足够的认知，准确进行事件预测成为了一项极具挑战性的研究。近年来，随着信息技术的发展，一方面，新兴媒体和移动互联网的流行产生了与事件相关的多源海量数据；另一方面，人工智能、模式识别等技术的迅速发展，赋予了机器对大数据进行挖掘并从中发现事件内在规律的能力。上述情况给研究者提供了一种绕过“通过事件发展规律来预测未来事件”的可能性。

目前大量基于人工智能的技术被应用于事件预测任务中，这些工作从事件发生的时间、地点和语义这3个维度对未来事件的发生进行预测。例如使用逻辑回归(Logistic Regression)[2]、支持向量机(Support Vector Machine，SVM)[3]、决策树(Decision Tree)[4]、点过程(Point Processes)[5]、循环神经网络(Recurrent Neural Network，RNN)[6]、联合概率分布[7]、注意力机制(Attention Mechanism)[8]等方法对事件发生的时间进行预测；使用核密度估计(Kernel Density Estimation，KDE)[9-10]和卷积神经网络(Convolutional Neural Networks，CNNs)[11]等方法对事件发生的位置进行预测；使用基于关联[12]和基于因果推断[13-15]的方法对语义进行预测等。这些方法在事件预测中取得了一定的效果，但它们大多针对社交媒体内容，基于特征工程对事件进行预测，没有很好地利用事件数据的结构化特征。

总结以往的研究可知，使用深度学习(Deep Learning)的方法虽然取得了较传统方法更好的效果，但事件预测仍然存在诸多挑战，包括异构多输出预测、预测输出之间的复杂依赖性和实时流预测任务[1]。近年来，研究者发现，相比传统深度学习方法，图神经网络(Graph Neural Networks，GNNs)具有能够处理非欧几里得空间的图数据、高效利用样本实例间的结构性特征和能够进行关系推理等特点，更加适合处理事件预测任务(如事件传播依赖的网络结构特性、事件关联的实体间关系推理等)，一系列基于GNNs进行事件预测的研究由此展开。但使用GNNs解决事件预测任务，仍然面临许多难题。首先，节点邻居节点数量不足，影响了最终的预测效果。GNNs从节点的局部邻居节点学习节点表示，如果其邻居节点有限，就无法得到足够的邻居信息，GNNs的表示能力和性能都将受到限制[16]。虽然通过叠加图层来扩大感受野(Receptive Field)的方法，可以将节点的多跳邻居信息纳入模型，但会导致过平滑问题(Over-smooth)[17]，从而影响模型的预测性能。其次，现有的方法大多从语义相关性的角度考虑，忽略了事件的传播特征(Propagation Character)，没有完全利用现有数据带来的信息。最后，大多数方法依赖于有监督的训练，需要专家对大量的数据进行注释，分析上下文和各种权威报告，费时费力[18]。

为了应对上述挑战，本文提出了一种使用了局部数据增强的动态图神经网络模型LAT-GAT。本文的主要贡献如下：

(1)参考了现有研究[19]，在其基础上改进训练过程，使用条件变分编码器进行预训练，从事件图的邻居节点获得了更多的信息，且能避免过平滑问题；同时也使模型获得了自监督预训练的能力。

(2)通过条件变分编码器的预训练，从而利用了事件传播的结构特征，提升了事件预测的准确率。

(3)使用真实世界的事件数据集进行了相关实验，通过与现有的事件预测算法的对比，验证了模型的有效性，并针对部分参数对模型的性能影响进行了研究。

本文第2章对事件预测、基于GNN的事件预测以及数据增强等研究进展进行了回顾；第3章对本文要解决的问题给出了形式化定义并对文中的符号进行了介绍；第4章介绍了本文提出的基于局部数据增强动态图的事件预测模型；第5章展示了实验结果并进行分析；最后总结全文并展望未来。

2　相关工作

2.1　事件预测

现有的事件预测技术，根据其预测对象的不同可以分为3类：对于时间的预测、对于位置的预测和对于语义的预测。近年来，还有研究对以上3个因素中的2个或3个进行联合预测[20-23]。

对于时间的预测，一类研究是预测在特定的时间间隔内某个事件是否发生；另一类是预测某个事件可能发生的时间段或具体的事件点。文献[2-4]将事件预测问题归结为一个二分类问题，分别使用逻辑回归、支持向量机和决策树等方法，对特定时间事件发生与否进行二分类。此类方法仅能对连续时间事件发生的时间进行粗粒度预测，并不适用于需要细粒度时间预测的场景。以文献[5]为代表的基于点过程的方法适用于预测真实事件时间分布，并在预测连续时间事件

预测任务中取得了较好的效果。但此类方法缺乏对事件语义的挖掘且需要事件的先验概率，实际上大多数时候并没有足够的事件信息可以被用来计算先验概率。

对于地点的预测，可以按照其预测粒度划分为栅格预测和点预测。前者预测事件发生的大致区域，常用的做法是把物理空间按照固定的间隔划分为栅格，预测事件发生在哪个栅格所标定的物理空间。后者预测事件发生的细粒度点，其预测输出是如经纬度坐标这类具体的空间位置。文献[9]采用基于 KDE 的方法对事件位置进行栅格预测，以所有的地图上的格为输入，由一个分类器给出事件可能发生的格。近年来，以文献[11]为代表的使用 CNNs 预测事件发生位置的方法取得了巨大的成功，此类方法可以从图像和空间数据中学习复杂的空间模式，从而实现对位置的预测。

对于语义的预测，其目的是预测将发生事件的主题、事件描述或除了时间地点以外的其他事件元属性(Meta-attributes)。与上述两种预测不同，对于语义的预测，由于其数据来源和输出预测结果的特点，其使用的方法与上述两种预测相比也更具多样性，按照数据的组织方式可以将其分为 3 类：基于规则的方法(利用关联规则)、基于序列的方法和基于事件图的方法。文献[12]通过从候选规则中学习事件之间的所有显著关联，从而利用关联对事件进行预测。文献[13]使用情感语义动态图卷积模型，捕捉事件相关的图像特征与高阶语义之间的语义相关性。文献[14-16]通过发现历史事件之间的因果关系，从而使用因果推断对未来事件进行预测。其步骤包括首先对事件语义进行表示，其次对事件进行因果推断，最后对未来事件进行推断。

事件预测技术的应用场景较为宽泛，可以应用在社会事件、网络安全、犯罪调查、灾害应急救援、医疗、媒体、交通、商业和工程系统等领域中。在社会事件的预测中，主要是针对线下事件和线上事件进行预测。前者预测事件发生的时间、地点、主题、参与人群和事件范围等；后者预测在线社交平台上事件的走向和影响，并研究如何通过主动干预避免事件产生不良结果。在网络安全领域中，现有代表性研究是通过系统的各种脆弱性指标来预测未来系统的故障或者出现网络攻击的可能性。例如，有研究通过对社交媒体文本流进行分析，来确定未来发生 DDoS 攻击的目标；还有研究针对计算机上运行的二进制代码执行情况进行观察，基于推测分支、无序执行和共享末级缓存的观察结果来预测是否会产生系统攻击。在其他应用领域的应用情况，可参考文献[1]。

2.2 基于 GNNs 的事件预测

现有基于图神经网络的方法，其思路是将预测问题归结为一个分类推荐问题进行处理。首先使用不同的图结构对事件进行表示，再通过不同的 GNNs 对图结构数据进行处理，对可能的事件类型评分并将其作为输出。文献[20]使用了基于动态 GCN 的方法，并且考虑了时间和上下文因素，从而实现了对社会事件的预测。文献[24]使用动态知识图谱(Dynamic Knowledge Graph，DKG)构建事件图，并对事件参与的实体和多个事件都进行了预测。文献[25]将使用自然语言表示的事件构建为事件图，在 BERT 模型中集成了额外的结构化参数，减轻了事件图稀疏性的影响，在训练阶段学习事件之间的联系，实现了事件预测。上述方法虽然取得了一定的效果，但是没有考虑事件的传播结构。

2.3 GNNs 的数据增强

现有的基于 GNNs 的模型无法应对节点邻居节点数量偏少的情况，但在构建的事件图中这种情况又较为普遍。为了应对这种挑战，数据增强(Data Augmentation)技术被引入以产生更多的样本，提供更多的邻居节点信息以获得有效的节点表示。数据增强技术最早出现在机器视觉和自然语言处理领域，针对图的数据增强技术总体上可以分为拓扑级(Topology-level)和特征级(Feature-level)两种[20]。前者的主要思想是通过扰动原始图数据的邻接矩阵，从而生成新的图结构；后者通过对抗训练或生成模型，影响节点的原始特征，增强其泛化能力。

2.4 自监督学习

机器学习，特别是基于深度学习的技术取得的巨大成功，使得机器可以对信息社会运行中产生的海量数据进行挖掘，使用模型学习事件发展的内在规律，对新到来的事件相关数据进行准确的预测(分类)。在现有的方法中，有监督学习(Supervised Learning)占据主要地位。其训练和测试都需要有较强的监督信息，即标注数据。在实际应用中，往往难以获得高质量的标注信息，或者需要具备一定知识的专家花费大量的时间和人力成本进行数据标注。因此，学术界开始研究如何在缺乏标注数据的情况下，利用有限的监督信息进行学习，即自监督学习(Self-Supervised Learning，SSL)。SSL 通过构造辅助任务，从大规模的无标注数据中挖掘其自身的信息，从而对网络进行训练，学习到对下游任务有用的表征。目前自监督学习主要分为两大类：对比学习(Contrastive Learning)和生成学习(Generative Learning)[26]。

对比学习方式的核心思想，是让正样本和负样本在特征空间中学习样本的特征表示，目标是使样本与正样本的特征表示尽可能接近，与负样本的表示尽可能不同。文献[27]指出对比学习的模型框架总体上可以分为 3 类：基于负例的对比学习、基于非对称网络的对比学习和基于特征区相关的对比学习。虽然不同模型的增强方式不同，但增强后都需要通过损失函数计算，同时使正例损失和负例损失都达到最小，代表模型有 MoCo[28]，SimCLR[29]，BYOL[30]和 SimSiam[31]等。

而生成式方法的核心思想是从原始样本出发，通过技术手段生成新的数据，目标是使新生成的数据尽可能“还原”原始数据。其以自编码器为代表，包括生成式对抗网络(Generative Adversarial Networks，GAN)[32]、变分自编码器(Variational Auto-Encoder，VAE)[33]等。生成式 SSL 方法可以分为两类：自回归(Autoregressive)和自动编码器(Autoencoding)。前者的代表性模型有 GraphRNN[34]，GCPN[35]和 GPT-GNN[36]等；后者的代表性模型有 MGAE[37]，GATE[38]，NWR-GAE[39]和 GraphMAE[40]等。

3 问题定义

令 $\boldsymbol{C}=[\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_{|\boldsymbol{C}|}]$ 为事件数据集，其中 $\boldsymbol{C}_i$ 表示第 i

天的事件，$|\boldsymbol{C}|$表示事件的总天数。$\boldsymbol{G}=[\boldsymbol{G}_1,\boldsymbol{G}_2,\cdots,\boldsymbol{G}_n]$是根据原始输入构建的事件图。$\boldsymbol{X}$表示历史事件的特征矩阵，$\boldsymbol{X}'$表示数据增强后的事件特征矩阵，$\hat{y}$表示模型输出的事件预测结果。

事件$\boldsymbol{C}_i$，$\boldsymbol{C}_i=\{t_i,city_i,doc_i\}$，$t_i$表示事件发生的时间，$city_i$表示事件发生的城市，$doc_i$表示事件相关的文本内容。本文中的事件预测任务定义为给定t_c-k到t_c-1天的历史事件来预测在t_c天特定城市是否会发生对应的事件$\hat{\boldsymbol{y}}_{t_c,\text{city}}$ $\hat{\boldsymbol{y}}_{t_c,\text{city}}=1$表示事件发生，反之，$\hat{y}_{t_c,\text{city}}=0$。

本文使用$\boldsymbol{G}=(V,E)$表示事件图，V是顶点集合，E是边集合。$\boldsymbol{A}\in\{0,1\}^{N\times N}$为对应图的邻接矩阵。$N_i\in\{V_i\mid A_{i,j}=1\}^{N\times N}$为节点$v_i$的所有邻居节点。图对应的特征矩阵为$\boldsymbol{X}\in\mathbb{R}^{N\times F}$，其中$F$为特征向量的维度，$\boldsymbol{X}_i$为节点$i$的特征向量。在图上定义操作“$\|$”用来将原始特征与数据增强得到的特征进行拼接。则事件预测过程可以描述为：

$$\boldsymbol{C}\xrightarrow{\text{encode}}\boldsymbol{G}\xrightarrow{\text{Augmentation}}\boldsymbol{X}\,\|\,\boldsymbol{X}'\xrightarrow{\text{model}}y_{t,\text{city}}$$

即学习一个函数$f:\boldsymbol{C}\rightarrow y_{t,\text{city}}$。

4　基于自监督数据增强的事件预测模型

本章将先介绍整个LAT-GAT模型的架构以及模型的工作流程，之后对每个模块分别予以介绍。所构建的事件图来源于数据集中的原始文档，并基于事件图生成其邻接矩阵序列和特征矩阵；随后经过局部数据增强，生成新的特征；最终依次对时间信息进行提取。当模型训练完成后，就可以对t_c时刻的原始输入进行预测。基于局部数据增强的事件预测模型如图1所示，事件预测分为动态图构建、自监督数据增强、时间感知图注意力网络和事件预测网络4个部分。

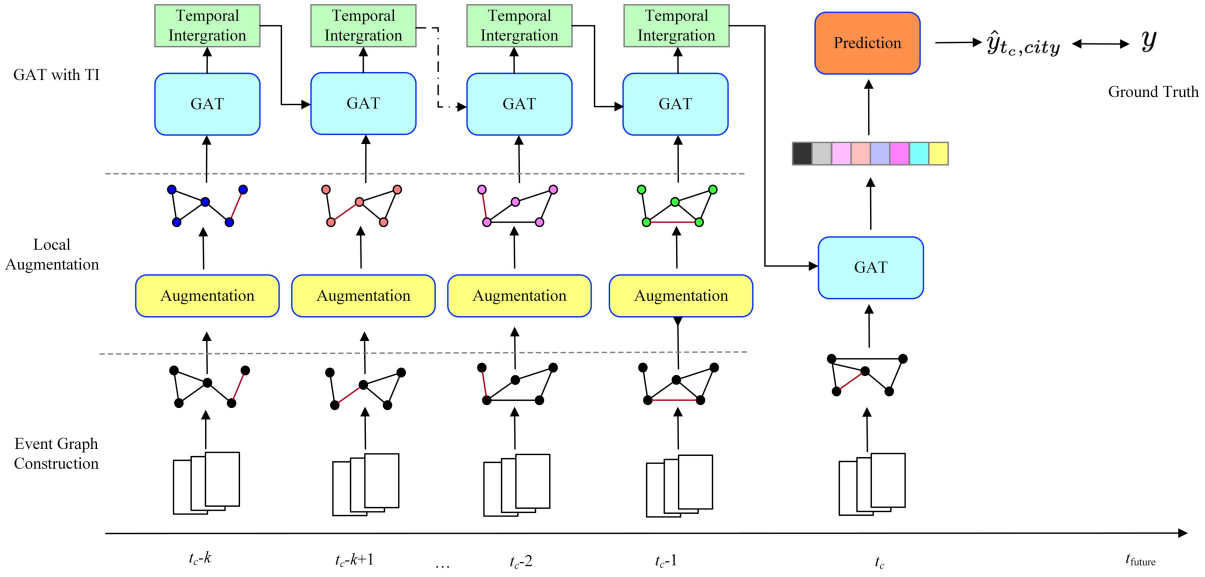

图1　模型框架示意图

Fig. 1　Illustration of model framework

4.1　动态图的构造

事件图构建是针对从社交媒体获取的事件文本，通过数据清洗，去掉无意义的词汇和罕见词汇，抽取与事件相关的词汇，然后挖掘词与词之间的关联关系，构建成事件图。在事件图中，每个节点表示一个词，每个节点的初始特征向量用词的嵌入向量表示（来源于Wikipedia数据库上的预训练[41]）。

事件图中的边表示事件文本中词与词之间的关系。以天为单位建立一个邻接矩阵序列$[\boldsymbol{A}_{t_c-k},\cdots,\boldsymbol{A}_{t_c-1}]$，其中一天的邻接矩阵为$\boldsymbol{A}_k\in\mathbb{R}^{n\times n}$。邻接矩阵的维度为每天经过筛选的词的总数量。本文使用文献[42]中提出的PMI（Point-wise Mutual Information）方法计算两个词之间的关联（边的权重）。对于事件$\boldsymbol{C}_i$，节点i与节点j之间的边权重计算式如式(1)所示：

$$\boldsymbol{A}_t[i,j]=\begin{cases}\log\dfrac{d(i,j)}{d(i)d(j)/D}, & PMIt(i,j)>0\\ 0, & \text{其他}\end{cases}\tag{1}$$

其中，$d(i,j)$是第t天词i和j同时出现的文章的总数，$d(i)$表示第t天词i出现的文章的数量，$d(j)$同理；D为数据集中的所有文章数。

4.2　自监督的数据增强

基于GNNs的方法是通过聚合邻居节点信息来更新当前节点的表示。但在实际情况中，对于构建的事件图，常常会出现节点邻居节点稀疏的情况，无法有效地建模事件图结构信息，从而影响模型的预测性能。因此，本文基于自监督的思想，考虑从事件图本身捕捉额外的自监督信息来增强事件图结构学习，无需额外的标签数据，适用于事件图上的数据增强。本文采用条件变分自编码器（Conditional Variational Auto Encoder，CVAE）[43-44]以图结构生成的方式进行数据增强，通过捕捉额外的自监督信息来辅助事件预测任务。条件变分自编码器以当前节点u为中心，学习其所有一阶邻居节点v特征的条件分布，并利用条件分布生成新的节点u的特征。一旦生成模型训练完成，便以当前节点u的一阶邻居节点为条件，生成新的节点u特征。其原理示意图如图2所示，对节点u进行数据增强，使用算法对u的邻居节点分布进行采样，并生成u的生成特征，通过CONCAT(·)操作进行

特征拼接。对其他节点也进行类似操作，最终生成原始特征和生成特征组成的特征矩阵。

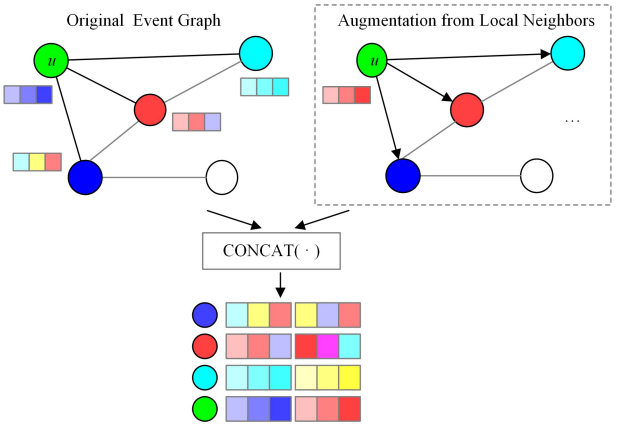

图 2 局部数据增强示意图

Fig. 2 Illustration of local augmentation

考虑节点 u，CVAE 的基本思想是以 $\boldsymbol{X}_v$ 为条件，所有一阶邻居节点 v 的特征 $\boldsymbol{X}_v$ 与 $\boldsymbol{X}_u$ 分布相关，其中，$v\in \boldsymbol{N}_u$，$\boldsymbol{N}_u$ 表示节点 $\boldsymbol{u}$ 的一阶邻域，通过 $\boldsymbol{q}_\phi(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u)$ 生成隐变量 $\boldsymbol{z}$，再通过对隐变量 $\boldsymbol{z}$ 进行采样，利用相关的分布 $\boldsymbol{p}_\theta(\boldsymbol{X}_u|\boldsymbol{z},\boldsymbol{X}_v)$ 生成节点 u 特征 $\bar{\boldsymbol{X}}_u$，其中 θ 表示解码模块中神经网络的参数，ϕ 表示编码模块中神经网络的参数，条件变分自编码器推断过程如式(2)所示：

$$\begin{aligned}\log \boldsymbol{p}_\theta(\boldsymbol{X}_u|\boldsymbol{X}_v)=&\int_z \boldsymbol{q}_\phi(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u)\log\frac{\boldsymbol{q}_\phi(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u)}{\boldsymbol{p}_\theta(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u)}\mathrm{d}\boldsymbol{z}+\\&\int_z \boldsymbol{q}_\phi(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u)\log\frac{\boldsymbol{p}_\theta(\boldsymbol{z},|\boldsymbol{X}_v,\boldsymbol{X}_u)}{\boldsymbol{q}_\phi(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u)\boldsymbol{p}_\theta(\boldsymbol{X}_v)}\\&\mathrm{d}\boldsymbol{z}\\=&\boldsymbol{D}_{\mathrm{KL}}(\boldsymbol{q}_\phi(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u),\boldsymbol{p}_\theta(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u))+\\&\boldsymbol{L}(\boldsymbol{X}_u,\boldsymbol{X}_v;\theta,\phi)\end{aligned}\tag{2}$$

其证据下界(Evidence Lower Bound，ELBO)如式(3)所示：

$$\begin{aligned}\boldsymbol{L}(\boldsymbol{X}_u,\boldsymbol{X}_v;\theta,\phi)=&\int_z \boldsymbol{q}_\phi(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u)\log\frac{\boldsymbol{q}_\phi(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u)}{\boldsymbol{p}_\theta(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u)}\mathrm{d}\boldsymbol{z}\\=&\int_z \boldsymbol{q}_\phi(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u)\log\frac{\boldsymbol{p}_\theta(\boldsymbol{z}|\boldsymbol{X}_v)}{\boldsymbol{q}_\phi(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u)}\mathrm{d}\boldsymbol{z}+\\&\int_z \boldsymbol{q}_\phi(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u)\log\boldsymbol{p}_\theta(\boldsymbol{X}_u|\boldsymbol{X}_v,\boldsymbol{z})\mathrm{d}\boldsymbol{z}\\=&-\boldsymbol{D}_{\mathrm{KL}}(\boldsymbol{q}_\phi(\boldsymbol{z}|\boldsymbol{X}_v,\boldsymbol{X}_u)|\boldsymbol{p}_\theta(\boldsymbol{z}|\boldsymbol{X}_v))+\\&\mathbb{E}_{\boldsymbol{q}_\phi}[\log\boldsymbol{p}_\theta(\boldsymbol{X}_u|\boldsymbol{X}_v,\boldsymbol{z})]\end{aligned}\tag{3}$$

其中，$\boldsymbol{L}(\boldsymbol{X}_u,\boldsymbol{X}_v;\theta,\phi)$ 作为损失函数，在训练阶段，训练的目的是使用节点 u 和其邻居 v 作为节点对输入，使得 ELBO 最大化。在生成阶段，使用节点 u 的一阶邻域节点的特征作为“条件”(Condition)，采样隐变量 z，输入解码器得到与节点 u 相关的生成特征向量。

4.3 时间感知图注意力网络

GAT 提出了使用注意力机制对邻域节点特征加权求和，其特征的权重取决于节点特征，独立于图结构，无须对整张图进行计算，可以减少训练时间。当 4.2 节中生成数据增强的样本后，本文选择使用 GAT 作为下游模型进行训练。为了使 GAT 可以利用 CVAE 的生成样本，本文重新对 GAT 的第一层进行定义，如式(4)所示：

$$\begin{aligned}H(1)=&(\Big\|_{k=1}^{K/2}\sigma(\sum_{v\in N_{u-t}}\alpha_{uv}^k W_k^{(1)}X_v))\ \|\\&(\Big\|_{k=K/2+1}^{K}\sigma(\sum_{v\in N_{u-t}}\alpha_{uv}^k W_k^{(1)}\overline{X_v}))\end{aligned}\tag{4}$$

需要注意的是，在最后一层的计算过程中，不采用“‖”拼接操作，而是对多个注意力机制的输出求平均值。其计算式如式(5)所示：

$$H^{(l+1)}=\Big\|_{k=1}^{K/2}\sigma\left(\frac{1}{K}\sum_{v\in N_{u-t}}^{K}\alpha_{uv}^k W^k H_v^{(l)}\right)\tag{5}$$

其中，α_{uv}^k 为在 $\boldsymbol{X}$ 或 $\bar{\boldsymbol{X}}$ 上计算得到的 k 头注意力系数，$\boldsymbol{W}_k^{(l)}$ 为权重矩阵。在 LAT-GAT 网络中，由于原始事件文档是按照时间进行组织的，因此，在每天的数据上，注意力层需要重新计算当天数据的注意力系数，如式(6)所示：

$$\alpha_{uv}=\frac{\exp(LeakyRelu(\vec{\alpha}^{\mathrm{T}}[W\vec{h_u}\ \|\ W\vec{h_v}]))}{\sum_{i\in N_{u,t}}\exp(LeakyRelu(\vec{\alpha}^{\mathrm{T}}[W\vec{h_u}\ \|\ W\vec{h_i}]))}\tag{6}$$

其中，“‖”为拼接操作，$N_{u,t}$ 为第 t 天节点 u 的所有邻居节点集合，$\vec{\alpha}^{\mathrm{T}}$ 为单层前馈神经网络参数，$LeakyRelu(\ )$ 为激活函数。

在 RNN 网络中，网络通过循环核实现了对事件序列信息的提取。本模型借鉴了这种思想，将上一个时间的输出，集成到当前层 GAT 的输出中作为下一时刻的输入，从而实现对时序信息的利用。需要注意的是，这里是将 LAT-GAT 中上一时刻的输出直接和当前时刻 GAT 网络的输出进行拼接，并非作为 GAT 网络中的一个节点，因此无须计算其注意力系数。令 $H_{\mathrm{out}}^{(t)}$ 代表 t 时间 LAT-GAT 的输出节点特征，则 $H^{(t)}$ 表示这一时刻 GAT 网络的输出，本文通过一个可学习的线性变换集成每个时刻的特征以及初始的词嵌入向量，如式(7)所示：

$$H_{\mathrm{out}}^{(t)}=\tanh[H^{(t)}W_p^{(t)}\ \|\ H_0^{(t)}W_e^{(t)}]\tag{7}$$

模型最后对比预测值和真实值，并优化交叉熵损失，计算方法如式(8)所示：

$$\boldsymbol{L}=-\sum y\ln\hat{y}\tag{8}$$

其中，y 为真实值，$\hat{y}$ 为预测值。为训练事件预测模型，本文将交叉熵损失和条件自编码器的证据下界结合，通过反向传播来优化模型参数，计算方法如式(9)所示：

$$\boldsymbol{L}=-(\sum \boldsymbol{y}_{t_c,\mathrm{city}}\log\hat{\boldsymbol{y}}_{t_c,\mathrm{city}}+\boldsymbol{L}(\boldsymbol{X}_u,\boldsymbol{X}_v;\theta,\phi))\tag{9}$$

5 实验

5.1 数据集构造

实验所使用的数据来自于“综合危机早期警系统项目”(Integrated Conflict Early Warning System，ICEWS)[45]，该系统可以用于监测、评估和预测国家的内部危机，指导战略资源的分配以减轻危机。作为系统的一部分，其事件库包含了从 1991 年 1 月至今的多种语言(英语、西班牙语、葡萄牙语和阿拉伯语等)的未分类事件；且经过深度和浅解析技术，构建了超过 250 万件独立事件，每个事件是由(源参与者，事件类型，目标参与者)组成的三元组，并且与事件的发生地理位置、时间等元数据以及新闻文章中的事件描述等内容相关联。与文献[16]一致，我们选取了来自印度、埃及、泰国和俄罗斯 4 个

国家的事件数据。其中印度、埃及的数据为 2012－2016 年的，泰国和俄罗斯的数据为 2010－2016 年的。印度包含 15 个城市，埃及、泰国分别包含 1 个城市（为其首都）；俄罗斯的数据选取了莫斯科和其他两个相关城市。它们关注的事件类型为抗议（Protest）。其他统计信息如表 1 所列。

表 1　数据集统计信息

Talel 1　Dataset statistics

文档	词	样本	正例	负例
247457	172731	21472	7941	13531

文档列表示数据集中作为原始输入的文章数量，词列表示移除低频词汇和无效词汇以后的数量。值得注意的是，所选国家不同，其相关的词数据量有较大不同，整体的正反例比约为 5∶3，使用 70％的实例作为训练集，15％作为验证集，15％作为测试集。

5.2　实验设置

本文选取了其他能够进行事件预测的算法作为基线方法，在同样的数据集上与本文方法进行对比。

（1）GCN：只使用基本 GCN 网络对事件进行预测，不考虑利用事件传播的时间特征，即所有的历史数据用一张图表示。

（2）DynamicGCN：即文献[20]中提到的方法，在每个 GCN 层嵌入上一时刻输出的特征，从而达到利用事件传播时间信息的目的。

（3）EvolveGCN[46]：其沿时间维度调整 GCN 模型。它使用 RNN 来进化 GCN 参数从而捕捉图序列的动态。输入是一系列只包含单词节点的动态同源图。

（4）HGT[47]：其包括节点和边类型依赖参数，以表征每条边上的注意力。HGT 引入了相对时间编码技术来处理节点时间戳可能不同的动态异构图，它需要动态异构上下文图作为输入。考虑到数据中的边是不断变化的，我们对边应用时间编码来调整该模型。

（5）GWet[48]：一种前沿的交通流量预测模型。我们对该模型稍加改动，使之可以应用于本文研究的问题中。

为了对本文所提模型和基线方法的效果进行评估，本文使用了准确率、F1 和召回率 3 种评价指标，还考察了 LAT-GAT 模型参数对预测准确率（Accuracy，Acc）的影响。本文使用 Adam Optimizer 对所有模型进行训练。所有模型的初始重要参数分别设置为：$Dropout\ Rate=0.2$，$Batch\ Size=5$，$Learning\ Rate=1\times10^{-3}$，$Weight\ Decay=5\times10^{-4}$。对于所有模型，我们均设置 $Eary\ Stopping=10$，即模型的表现效果如果在 10 个训练轮回里没有明显增长，则停止训练以防止过拟合。所有模型默认的训练轮次为 $Iteration=50$。最后，所有的实验结果均取 5 次随机实验的平均值，以减少偶然性。

5.3　性能比较和结果分析

本文使用 4 个国家的事件数据集，在相同配置的单 GPU 的 PC 上对 3 种方法进行了测试，分别计算其 F1 评分（F1-score）、召回率（Recall，Rec）和精确率（Precision，Prec）。表 2 列出了两种具有代表性的基线方法与本文所提方法在所选数据集上的实验结果，预测类型为“抗议”（Protest）事件。其中基于 GCN 的方法对以历史数据构建的事件图提取特征并进行预测，没有使用时间传播的时间信息；DynamicGCN 使用一个两层的 GCN 网络提取每一天的文章信息构建的事件图的特征，并通过一个类似于 RNN 结构的网络，将上一时刻 GCN 提取到的特征作为当前网络的输入，从而在一定程度上利用了事件传播的时间信息。

表 2　数据集上的性能对比

Table 2　Performance comparison on datasets

	Thailand			Egypt			Russia			India		
	F1	*Rec*	*Prec*	*F1*	*Rec*	*Prec*	*F1*	*Rec*	*Prec*	*F1*	*Rec*	*Prec*
GCN	0.7304	0.7281	0.7363	0.8291	0.7910	0.8683	0.8040	0.8506	0.6843	0.6232	0.6981	0.6296
DynamicGCN	0.7644	0.7766	0.7526	0.8477	0.8170	0.8807	0.8120	0.8232	0.8011	0.6803	0.6987	0.6629
EvolveGCN	0.7712	0.7604	0.7621	0.8501	0.8030	0.8765	0.7650	0.7905	0.7003	0.6103	0.7002	0.6501
HGT	0.7701	0.7502	0.7754	0.8498	0.7960	0.8801	0.7720	0.8006	0.7256	0.5962	0.7156	0.6431
GWet	0.7521	0.7601	0.7831	0.8510	0.8040	0.8902	0.7830	0.8123	0.7385	0.6901	0.7103	0.6699
LAT-GAT	0.7849	0.7767	0.7935	0.8514	0.8043	0.9043	0.7885	0.8343	0.7475	0.6946	0.7196	0.6713

由表 2 可知，仅使用 GCN 的方法，在对 Russia 事件预测时达到了最高的准确率（0.8506）。但其他两种使用了时间信息的方法在剩余指标上都高于仅使用 GCN 的方法，说明使用时间数据能有效提升事件预测的性能。

与 DynamicGCN 方法相比，在 Thailand 和 India 数据集上，LAT-GAT 所有的预测性能指标均高于 DynamicGCN，尤其在 Egypt 事件数据集上，达到了 0.9043 的准确率；在 Thailand，Egypt 和 India 数据集上，LAT-GAT 准确率平均提升约 3.7％；在 Thailand 数据集上，准确率最高提升约为 5％；在 Egypt 数据集上，仅召回率落后约 1％。说明在大多数情况下，本文所提模型能有效提升预测性能。

与 EvolveGCN 方法相比，LAT-GAT 在所有数据集上的表现都优于 EvolveGCN。一部分原因是 EvolveGCN 所利用的时间维度方法过于宽泛，因而无法在数据集上的特定事件上做出准确预测。另一部分原因是有关 RNN 的方法缺乏相关的配套组件来帮助其不断更新所获得的动态图信息，导致预测结果出现了偏差。

与 HGT 方法相比，本文方法在所有数据集上的表现都更优。这是因为 HGT 虽然利用的是时间戳特征，并不断更新所获得的动态图信息，但是由于其整体模型中缺乏对数据的预处理和生成新样本，从而未能在数据的预处理环节获得最为有效的数据的空间特征和时间特征，导致最后的预测出现了偏差。

最后，本文方法在所有的数据集上的表现都优于基线方法 GWet。部分原因是 GWet 的设计初衷是进行交通流量预测，这与本文方法的研究目标不同，导致其在本实验中的性能

下降。另外，同样地，该方法没有利用预处理模型在训练之初抓住节点与节点之间的空间特征与时间特征以及节点间的边的特征，从而导致后续训练效果不佳。

Russia 数据集具有一定的特殊性，因为在俄罗斯的相关报道中，关于“抗议”的新闻数量较少，测试数据集还引入了除莫斯科以外其他两个有一定数量“抗议”事件发生的城市新闻数据[16]，这也是本文所提方法在此类数据集上表现不佳的原因，我们没有足够的事件传播时间信息和传播结构信息可以被利用。

5.4 消融实验

为了研究 Temporal-GAT 模型的每个组成部分的贡献，我们进行了一项消融研究，测试了 4 种实验设置下的模型性能。

(1)有 Temporal Integration 但没有 CVAE 的模型(-C)。

(2)有 CVAE 但没有 Temporal Integration 的模型(-T)。

(3)有 CVAE 和 Temporal Integration 的模型，但只使用 CVAE 进行一次训练，后续不使用(CT-C)。

(4)有 CVAE 和 Temporal Integration 的模型，但只使用 Temporal Integration 进行一次训练，后续不使用(CT-T)。

(5)有 CVAE 和 Temporal Integration 的完整模型(Temporal-GAT)。

在 Thailand，Egypt，Russia 和 India 这 4 个不同国家的测试集上，使用 F1、召回率和精确率指标对每个实验设置的性能进行了评估，并与完整模型进行了比较，其结果如表 3 所列。

表 3 数据集上的消融实验

Table 3 Ablation study on datasets

	Thailand			Egypt			Russia			India		
	F1	*Rec*	*Prec*	*F1*	*Rec*	*Prec*	*F1*	*Rec*	*Prec*	*F1*	*Rec*	*Prec*
-C	0.6904	0.6881	0.6903	0.8091	0.7810	0.8083	0.6704	0.6906	0.6243	0.6122	0.6801	0.5896
-T	0.7514	0.7266	0.7226	0.8357	0.7770	0.8607	0.7020	0.7932	0.7211	0.6403	0.6587	0.6129
CT-C	0.5004	0.6001	0.6501	0.7096	0.7020	0.6823	0.5903	0.6801	0.6024	0.6544	0.6905	0.5402
CT-T	0.5231	0.6705	0.7014	0.8254	0.7540	0.7804	0.6732	0.7541	0.7001	0.6143	0.6501	0.5921
LAT-GAT	0.7849	0.7767	0.7935	0.8514	0.8043	0.9043	0.7885	0.8343	0.7475	0.6946	0.7196	0.6713

在第一个实验中，我们去除了 CVAE，但保留了 Temporal Integration。结果表明，此模型在所有国家数据集上，尤其是在 Russia 数据集上表现都大幅下降。这是由于 Russia 数据集内样本较少，因此使用 CVAE 让模型对数据进行预处理并生成新样本尤为重要。这表明 CVAE 在增强局部节点特征方面起着至关重要的作用，它的缺失极大地影响了模型的性能。

在第二个实验中，我们去除了 Temporal Integration，但保留了 CVAE。结果再次表明，与完整模型相比，更改后的模型在所有国家的数据集上的表现都显著下降。这表明，虽然 CVAE 对于增强局部节点特征很重要，但仅使用它还不够，必须与 Temporal Integration 配合才能实现最佳性能。

在第三个实验中，我们保留了 CVAE 和 Temporal Integration，但只使用 CVAE 进行一次训练，后续不使用。结果表明，模型在所有国家的数据集上的表现都显著下降，且下降程度在所有分类实验中最为明显。这表明，尽管 CVAE 对于增强局部节点特征很重要，但仅将其用于一轮训练是不够的，因为 CVAE 对数据的预处理和生成新样本需要通过不断的训练才能获得适合模型的最佳选项，因此需要连续使用 CVAE 以获得最佳性能。

在第四个实验中，我们保留了 CVAE 和 Temporal Integration，但只使用 Temporal Integration 进行一次训练，后续不使用。结果表明，模型在所有国家的数据集上的性能都大幅下降，下降程度仅次于第三个实验。这也从另一侧面说明，不论是 CVAE 还是 Time Integration，组件的有效使用都必须通过多次、完整的训练才可以达到最佳效果，不当的使用反而会对核心下游分类模型(GAT)产生较为明显的不利影响。

总之，这些实验强调了 LAT-GAT 模型中 CVAE 和 Temporal Integration 组件的重要性。仅在一轮中使用任一组件或移除任一组件都会极大程度地影响模型的性能，不当的训练和使用组件更会对模型产生较大的不利影响。因此，在模型中同时包含 CVAE 和 Temporal Integration 并合理地训练和使用该两组件对于优化性能至关重要。

5.5 模型不同损失函数的变化

本文提出了两种损失函数，分别是利用条件自分编码器(CVAE)使模型进行自监督学习和对数据进行预处理的损失函数以及利用模型的预测值和真实值之间的差异得到的交叉熵损失。将两种损失函数的值相加，即可得到总体模型的损失函数值。受消融实验结果的启发，我们想更全面地探究两部分损失函数的变化对模型的总体影响。为方便阅读，我们将第一部分的损失函数简洁表示为 $L|CVAE$，将第二部分的交叉熵损失表示为 $L|Cross\ Entropy$，将模型的总损失表示为 $L|All$。

从图 3 的各部分损失函数的变化中可以看出，$L|CVAE$ 值的下降速率明显低于 $L|Cross\ Entropy$ 与 $L|All$。这从另一个侧面表明，本文提出模型中的预处理组件对模型的整体表现产生了较为重要的积极影响。

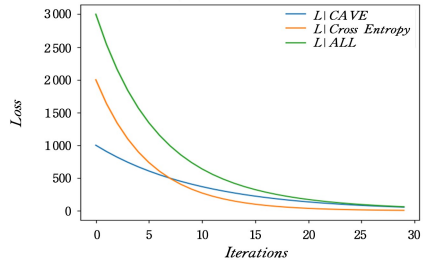

图 3 损失函数变化图

Fig. 3 Variation of loss function

为了更好验证我们的设想，我们移除了 Time Integration 组件，仅保留条件自分编码器，其结果如图 4 所示。从图中函数值的变化趋势可以得出，即使单独使用模型中的预处理

组件而不使用时间特征，其对模型产生的积极影响也是非常明显的。

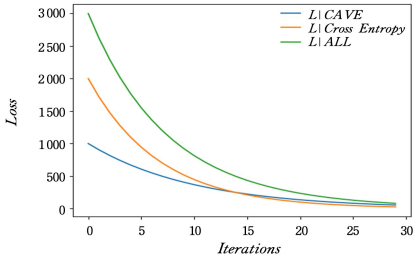

图 4　损失函数变化图(去除 Time Integration)

Fig. 4　Variation of loss function(without Time Integration)

5.6　模型参数影响

本文还对 LAT-GAT 模型参数对模型性能的影响进行了研究。我们通过改变图注意力层的数量来考察其对模型准确率的影响。其中 Parameters 为模型参数规模，K 为图注意力层的数量，实验在 Thailand 数据集上进行。其结果如表 4 所列。可以看到，准确率并非随着层数的堆叠一直保持上升趋势，但是模型的参数规模一直在增长。例如，在层数从 2 增加到 7 时，准确率提升约 1%，但参数量却增加了 3 倍之多。

表 4　模型层数对准确率的影响

Table 4　Influences of layers on Acc of model

层数	*Acc*	Parameters
$K=1$	0.851 6	27 385
$K=2$	0.858 7	47 785
$K=3$	0.841 0	68 185
$K=4$	0.841 0	88 585
$K=5$	0.865 7	108 985
$K=6$	0.855 1	129 385
$K=7$	0.869 3	149 785

其次，本文还了 Batch 参数对模型 Acc 的影响，其结果如表 5 所列。Batch 指在每一个训练轮回中，模型获得的图的个数。在本研究中，Batch 可以理解为一次训练轮回中，模型获得的样本事件数。实验结果表明，模型在 $Batch=10$ 时，可以获得很高的准确率(Acc)，这表明该模型获得的事件个数越多，就越能够发现事件之间的传播特征，得出准确的预测结果。

表 5　图数量对模型准确率的影响

Table 5　Influences of graph batch on Acc of model

Batch	*Acc*
1	0.855 1
2	0.844 5
5	0.844 5
10	0.858 7

最后，本文研究了 Dropout 参数对预测准确率的影响，其结果如表 6 所列。Dropout 指在每一个训练轮回中，随机抽取图中的节点的边并移除，以增强从图中提取的特征信息，提高模型的训练效率并验证分类模型的鲁棒性。该值一般设定为 0.5 以下。本文中，由于已经使用了预处理模型进行特征信息的增强，因此可以使用 Dropout 参数模拟外部因素对图内链接节点边的扰动，以验证所提模型的鲁棒性。实验结果表明，本文提出的 LAT-GAT 在 Dropout 取值高或低时准确率都较高，这在一定程度上反映 LAT-GAT 模型具有很强的抗扰动能力和鲁棒性。

表 6　Dropout 值对模型准确率的影响

Table 6　Influences of dropout on Acc of model

Dropout	Acc
0.1	0.855 1
0.2	0.851 6
0.4	0.855 1

结束语　本文提出了一种基于局部数据增强动态图的事件预测模型，该模型利用事件的传播结构，为节点生成额外的特征向量，最后使用基于 GAT 的方法对未来事件进行预测。在真实事件数据集上对本文提出的模型进行验证，并与其他未做数据增强的基线方法和基于 GNNs 的方法进行对比，从整体的性能表现上验证了本文模型的有效性。本文模型可以应用于事件预测任务中，也可以应用于流言、假新闻传播预测任务中。在未来的工作中，我们将研究如何挖掘多种事件之间的内在联系，利用这种联系对多种事件进行预测；此外也将研究如何利用丰富的社交媒体信息(如图片)，与文本进行联合事件预测。

参 考 文 献

[1] ZHAO L. Event prediction in the big data era: A systematic survey[J]. ACM Computing Surveys(CSUR), 2021, 54(5): 1-37.

[2] ZHAO L, YE J, CHEN F, et al. Hierarchical incomplete multi-source feature learning for spatiotemporal event forecasting [C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 2085-2094.

[3] INCEOGLU F, JEPPESEN J H, KONGSTAD P, et al. Using machine learning methods to forecast if solar flares will be associated with CMEs and SEPs[J]. The Astrophysical Journal, 2018, 861(2): 128.

[4] DE CAIGNY A, COUSSEMENT K, DE BOCK K W. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees[J]. European Journal of Operational Research, 2018, 269(2): 760-772.

[5] QIAO Z, ZHAO S, XIAO C, et al. Pairwise-ranking based collaborative recurrent neural networks for clinical event prediction [C]// Proceedings of the Twenty-seventh International Joint Conference on Artificial Intelligence. 2018.

[6] SIMMA A, JORDAN M I. Modeling events with cascades of Poisson processes[J]. arXiv: 1203. 3516, 2012.

[7] HOU X L, ZHOU P P, ZHAO J B. An automatic exposure model of image sequence acquisition for HDR scenes[J]. Journal of Chongqing University of Technology: Natural Science, 2022, 36(4): 153-161.

[8] BERHICH A, BELOUADHA F Z, KABBAJ M I. An attention-based LSTM network for large earthquake prediction[J]. Soil Dynamics and Earthquake Engineering, 2023, 165: 107663.

[9] RAMA-MANEIRO E, VIDAL J C, LAMA M. Embedding graph convolutional networks in recurrent neural networks for predictive monitoring[J]. IEEE Transactions on Knowledge and Data

Engineering, 2024, 36(1): 137-151.
[10] HAO M, JIANG D, DING F, et al. Simulating spatio-temporal patterns of terrorism incidents on the Indochina Peninsula with GIS and the random forest method[J]. ISPRS International Journal of Geo-Information, 2019, 8(3): 133.
[11] PIRAJAN F, FAJARDO A, MELGAREJO M. Towards a deep learning approach for urban crime forecasting[C]// Workshop on Engineering Applications. Cham: Springer, 2019: 179-189.
[12] HAN J, PEI J, TONG H. Data mining: concepts and techniques [M]. Morgan kaufmann, 2022.
[13] SU Y T, WANG J, ZHAO W, et al. Dynamic graph convolutional neural network for image sentiment distribution prediction[J]. Journal of Jilin University(Engineering and Technology Edition), 2023, 53(9): 2601-2610.
[14] RADINSKY K, DAVIDOVICH S, MARKOVITCH S. Learning causality for news events prediction[C]// Proceedings of the 21st International Conference on World Wide Web. 2012: 909-918.
[15] LEI L, REN X, FRANCISCUS N, et al. Event prediction based on causality reasoning[C]// Asian Conference on Intelligent Information and Database Systems. Cham: Springer, 2019: 165-176.
[16] YANG Y, WEI Z, CHEN Q, et al. Using external knowledge for financial event prediction based on graph neural networks[C]// Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019: 2161-2164.
[17] LIU S, YING R, DONG H, et al. Local augmentation for graph neural networks[C]// International Conference on Machine Learning. PMLR, 2022: 14054-14072.
[18] LI Q, HAN Z, WU X M. Deeper insights into graph convolutional networks for semi-supervised learning[C]// Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
[19] SHU K, SLIVA A, WANG S, et al. Fake news detection on social media: A data mining perspective[J]. ACM SIGKDD Explorations Newsletter, 2017, 19(1): 22-36.
[20] DENG S, RANGWALA H, NING Y. Learning dynamic context graphs for predicting social events[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 1007-1016.
[21] BILOŠ M, CHARPENTIER B, GÜNNEMANN S. Uncertainty on asynchronous time event prediction[J]. arXiv: 1911.05503, 2019.
[22] WANG Q, JIN G, ZHAO X, et al. CSAN: A neural network benchmark model for crime forecasting in spatio-temporal scale [J]. Knowledge-Based Systems, 2020, 189: 105120.
[23] YUAN Z, ZHOU X, YANG T. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018: 984-992.
[24] DENG S, RANGWALA H, NING Y. Dynamic knowledge graph based multi-event forecasting[C]// Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020: 1585-1595.
[25] DU L, DING X, ZHANG Y, et al. A Graph Enhanced BERT Model for Event Prediction[C]// Findings of the Association for Computational Linguistics: ACL 2022. 2022: 2628-2638.
[26] LIU X, ZHANG F, HOU Z, et al. Self-supervised learning: Generative or contrastive[J]. arXiv: 2006.08218, 2021.
[27] LI X, LIU X P, LI W C, et al. Survey on Contrastive Learning Research[J]. Journal of Chinese Computer Systems, 2023, 44(4): 787-797.
[28] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 9729-9738.
[29] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations [C]// International Conference on Machine Learning. PMLR, 2020: 1597-1607.
[30] TIAN Y, CHEN X, GANGULI S. Understanding self-supervised learning dynamics without contrastive pairs[C]// International Conference on Machine Learning. PMLR, 2021: 10268-10278.
[31] CHEN X, HE K. Exploring simple siamese representation learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 15750-15758.
[32] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
[33] BOWMAN S R, VILNIS L, VINYALS O, et al. Generating sentences from a continuous space[C]// 20th SIGNLL Conference on Computational Natural Language Learning, CoNLL 2016. Association for Computational Linguistics(ACL), 2016: 10-21.
[34] YOU J, YING R, REN X, et al. Graphrnn: Generating realistic graphs with deep auto-regressive models[C]// International Conference on Machine Learning. PMLR, 2018: 5708-5717.
[35] YOU J, LIU B, YING Z, et al. Graph convolutional policy network for goal-directed molecular graph generation[J]. arXiv: 1806.02473, 2018.
[36] HU Z, DONG Y, WANG K, et al. Gpt-gnn: Generative pre-training of graph neural networks[C]// Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020: 1857-1867.
[37] WANG C, PAN S, LONG G, et al. Mgae: Marginalized graph autoencoder for graph clustering[C]// Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017: 889-898.
[38] SALEHI A, DAVULCU H. Graph Attention Auto-Encoders [C]// 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence(ICTAI). IEEE Computer Society, 2020: 989-996.
[39] TANG M, YANG C, LI P. Graph Auto-Encoder via Neighborhood Wasserstein Reconstruction[J]. arXiv: 2202.09025, 2022.
[40] HOU Z Y, LIU X, CEN Y K, et al. GraphMAE: Self-Supervised Masked Graph Autoencoders[C]// The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'22). 2022: 594-604.

[41] MIKOLOV T，SUTSKEVER I，CHEN K，et al. Distributed representations of words and phrases and their compositionality [C]// Proceedings of the 26th International Conference on Neural Information Processing Systems. 2013：3111-3119.

[42] CHURCH K，HANKS P. Word association norms，mutual information，and lexicography[J]. Computational Linguistics，1990，16(1)：22-29.

[43] KINGMA D P，WELLING M. Auto-Encoding Variational Bayes [J]. arXiv：1312.6114，2014.

[44] SOHN K，LEE H，YAN X. Learning structured output representation using deep conditional generative models[C]// Proceedings of the 28th International Conference on Neural Information Processing Systems. 2015：3483-3491.

[45] ELIZABETH B，JENNIFER L，SEAN O，et al. ICEWS Coded Event Data. Harvard Dataverse[EB/OL]. [2023-12-03]. https：//dataverse. harvard. edu/dataset. xhtml?persistentId=doi：10.7910/DVN/28075.

[46] PAREJA A，DOMENICONI G，CHEN J，et al. Evolvegcn：Evolving graph convolutional networks for dynamic graphs[J]. arXiv：1902.10191，2020.

[47] HU Z N，DONG Y X，WANG K S，et al. Heterogeneous graph transformer[C]// Proceedings of The Web Conference 2020. 2020：2704-2710.

[48] WU Z H，PAN S R，LONG G D，et al. Graph wavenet for deep spatial-temporal graph modeling[J]. arXiv：1906.00121，2019.

PAN Lei，born in 1986，Ph.D，senior engineer. His main research interests include NLP，multimodal data feature extraction，crisis event analysis and intelligent text generation.

LIU Leyuan，born in 1982，Ph.D，research associate. His main research interests include graph learning，social network data mining and event prediction.

（责任编辑：何杨）

一文带你回顾 CNCC 的发展

2003 年，首届中国计算机大会 CNCC2003 在北京召开，举行了 14 个领域的分会场学术论文交流。参会人数达 400 余人。

2007 年，CNCC2007 在苏州举行，共举办了 2 个产业发展论坛，在 24 个学术领域宣读论文交流，来自全国计算机研究、教育、应用、产业、政府等各界专业人士近 600 人参加了大会。

2010 年，CNCC2010 在杭州举行，本届大会创下了 CNCC 历史上的许多记录，首次举办科技新成果展览、CNCC 期间首次举办 YOCSEF 论坛、CNCC 首次进行网上直播、首次资助经费困难者参加 CNCC 等。共有 1 200 余人出席大会，CNCC 的会议规模首次突破千人。

2012 年，CNCC2012 在大连举行，会议英文名更名为 China National Computer Congress，CNCC 简称不变并一直沿用至今。本届大会吸引了来自全国 30 个省市区、港澳及海外的 2 000 名计算领域的专家学者参会。

以“计算改变未来（Computing Changes the Future）”为主题的 CNCC2016 在太原举行，来自国内学术界、产业界、政府部门、媒体界的相关人士参加了大会，参会人数超过 5 000 人。

2019 年，CNCC2019 在苏州举行，主题为“智能＋引领社会发展（AI＋ Leading the Development of Society）”，共有 8 000 余人参加，本次大会还特别为会员举办了会员之夜（CNCC Night）。

2020 年，受疫情影响，CNCC2020 在北京主会场、沈阳、杭州、济南分会场、重庆、厦门专场及线上同时举行，本届大会主题为“信息技术助力社会治理（Information Technology Empowering Social Governance）”。

2021 年，CNCC2021 于深圳主会场、北京分会场及线上同时举行，主题为“计算赋能加速数字化转型（Expediting Digital Transformation with Computing Empowerment）”。3 场大会主题论坛，111 场前沿技术论坛，复杂疫情下全新的办会模式被参会者誉为“CNCC 历史上最难忘的一届”。

2022 年，CCF 创建 60 周年，会员数首次突破 10 万。CNCC 历史上首次采用全线上方式举行，主题为“算力 数据 生态”。大会设置 15 场特邀报告，3 场大会论坛，120 余场技术论坛，线上注册参会人数突破 1.3 万，全网直播人气超过 570 万。

2023 年，CNCC 迎来第二十届，也是经历了 3 年线上活动后第十二届 CCF 理事会任期内的首次全线下大会。CNCC2023 在沈阳举办，大会主题为“发展数字基础设施，支撑数字中国建设”，包括 19 个特邀报告、3 场大会论坛、130 场技术论坛和丰富的活动及展览展示，ACM、IEEE CS、IPSJ、KIISE 等国际合作学会的代表出席。本届 CNCC 参会人数突破 1.3 万，是 CNCC 历史上真正实现了万人规模的盛会。

回顾过去二十年的历程，CNCC 历经了从首届仅有 14 个领域的学术论文交流，逐渐发展到涵盖数十个方向 130 场技术论坛，从最初的 400 余人参加，成长为如今 700 余位国内外讲者踊跃参与，超 13 000 人注册的年度盛会。二十载不忘初心，CNCC 旨在为所有参会者呈上一场精彩宏大的专业盛宴，也期待所有人都能够在这场盛宴中获得助益，提升自身的专业价值，获得前行的动能。

CNCC2024 将于浙江省东阳市横店召开。期待与您再次相聚！