基于时空图卷积的网络漏洞态势预测

张迎春[1],李金[2],阿布都热依木·热西丁[3],张珣[2,3*],郝蒙蒙[4],江东[4]

1. 北京工商大学人工智能学院,北京 100048
2. 北京工商大学计算机学院,北京 100048
3. 和田师范专科学校数学与信息学院,和田 848099
4. 中国科学院地理科学与资源研究所,北京 100101

摘要 在网络空间要素预测过程中加入地理空间特征,可实现时空预测网络空间要素。针对网络安全要素预测过程中少有结合网络数据地理空间特征的研究现状,选择有地理空间特征的网络漏洞检测数据,构造网络漏洞时空数据集,通过构建结合图卷积和门控时间卷积的时空图卷积模型,实现网络漏洞态势发展的预测。选取ARIMA和LSTM时序预测模型进行对比实验,提出的网络漏洞时空图卷积预测模型在MAE、RMSE和MAPE的评价标准下显示有着更好的预测效果。

关键词 网络空间数据;地理空间;时空数据;时空图卷积;预测模型

大规模网络安全态势感知可以帮助网络管理者了解目标网络的安全状态,并在一定程度上提供决策参考依据,是网络安全领域的一个重要研究方向,对于网络信息安全防护具有重要意义[1-4]。态势感知的概念被定义为态势的认识、理解和预测的3层概念模型[5]。态势预测是网络安全感知在特定网络环境中,基于提取和理解到的安全要素信息,预测可能态势趋势的过程[6],是依照过去和当前搜集到的信息对一段时间内的未来网络安全状况发展做出预测。网络安全态势预测是对网络空间安全要素信息进行分析和学习,建立特定的预测模型来预测未来网络态势的过程。

网络安全威胁的随机性和不确定性决定了网络环境的态势变化是非线性的,统计模型也就无法应用到当前网络环境中,机器学习、深度学习等人工智能方法是解决此类问题的常用技术[3]。当前很多研究者使用机器学习方法,建立模型预测未来网络态势,例如马尔可夫链Markov[7-10]、支持向量机

收稿日期:2022-12-12;修回日期:2023-04-23
基金项目:国家重点研发计划项目(2020YFB1806500)
作者简介:张迎春,实验师,研究方向为网络安全、地理人工智能,电子信箱:zhangyingchun@btbu.edu.cn;张珣(通信作者),教授,研究方向为网络安全、地理人工智能,电子信箱:zhangxun@btbu.edu.cn
引用格式:张迎春,李金,阿布都热依木·热西丁,等. 基于时空图卷积的网络漏洞态势预测[J]. 科技导报, 2023, 41(13): 60-66; doi: 10.3981/j.issn.1000-7857.2023.13.006

SVM[11-13]、BP神经网络[14-15]。由于神经网络对于大规模数据的特征挖掘能力,现今已广泛应用于网络空间态势感知方面的研究来解决网络安全问题。胡昕[16]提出一种基于循环神经网络(recurrent neural network,RNN)的网络安全态势预测方法,考虑网络态势要素的时序特点,利用其时间序列的关联性进行网络安全态势的预测。为了获得更好的预测性,Feng等[17]提出带有门控循环单元(gated recurrent unit,GRU)的RNN态势预测模型,充分考虑原始网络时间序列数据对未来态势的影响。Fang等[18]考虑网络攻击序列的长期依赖性和高度非线性,基于具有长短期记忆的双向循环神经网络(BRNN-LSTM)开发了深度学习框架。为了提高长短期记忆网络(long short-term memory,LSTM)的抗噪能力和序列关联分析能力,Fan等[19]将注意力机制和事件嵌入引入了LSTM网络,提出ALEAP模型,将网络安全事件进行划分,分为不同的执行阶段,通过词嵌入将安全事件序列,更好地集中于历史安全事件中更相关的部分,理解攻击漏洞。Kishioka等[20]提出了一种使用卷积神经网络(convolutional neural networks,CNN)的动态预测方法,通过使用主机日志的邻接矩阵作为CNN的输入数据,来预测自演化僵尸网络的传播程度。何春蓉等[21]提出了一种新的网络安全态势预测方法,该方法主要考察网络安全态势要素之间的关联性以及历史数据的重要性,借助注意力机制为安全态势元素分配相应的权重,提取未来态势与历史态势之间的依赖关系。

对网络空间的科学刻画是网络事件分析、网络空间治理、网络安全保障的重要基石,也是信息化时代地理学研究拓展的新领域[22]。从网络空间安全的角度看,物理域和信息域主要关注以信息技术为核心的网络基础设施安全及网络信息通信安全,而认知域和社会域更关注以人为核心的认知文化等精神层面以及个人与集体相互作用的社会层面。网络空间作为人类信息活动的产物是与地理空间交织融合的,地理空间是网络空间所依附的客观载体,网络空间中物理和信息2部分都不能离开地理空间单独存在[23]。网络态势感知研究少有综合考虑网络空间与地理空间的特点和相互作用的,网络态势感知的预测模型也并没有将网络要素的地理信息特征进行数据发掘,实现将网络空间要素和地理空间关联研究的预测模型[24]。

为在预测问题中融入网络漏洞数据的地理空间特征,对网络漏洞数据进行了时空结构化处理,并提出一个新的预测城市网络安全漏洞的时空图卷积模型,以综合考虑网络监测数据中的地理空间和时间序列的特征,进行网络漏洞态势预测。

1　研究方法

1.1　问题形式化定义

网络数据时空预测问题,是根据网络监测数据发掘其中的时空特征,通过数据分析等手段,用已有的历史数据来对未知的将来数据进行预测,旨在挖掘数据中蕴含的时空模式,从而在当前网络监测数据上进行推断,以根据已知的特征值来预测目标特征的未来值。

$$\left[X^{T_0},X^{T_1},X^{T_2},\cdots,X^{T_i};G\right]\rightarrow\left[X^{T_{(i+1)}},X^{T_{(i+2)}},X^{T_{(i+3)}},\cdots,X^{T_{(i+j)}}\right] \tag{1}$$

如式(1)所示网络漏洞数据预测任务的目标是根据给定前T_i个时间点的历史网络数据以及网络空间结构图,预测后T_j个时间点的网络数据。

输入特征:网络—地理空间网络漏洞历史时间序列数据和地理空间关系邻接矩阵。

输出结果:空间图网络各点的漏洞数据预测值。

模型技术框架如图1所示,数据预处理得到历史网络漏洞数据的时序特征数据和空间图网络,分别为特征数据集和图网络邻接矩阵。在进行网络漏洞时序特征数据的预测时,同时聚合网络漏洞空间特征,提高网络漏洞态势预测的效果。

1.2　模型介绍

本文提出的时空图卷积网络模型包括空间图卷积模块和时间门控卷积模块2部分。图卷积模块主要用于捕获网络漏洞时空数据中的空间特征,将所构造的图结构的邻接矩阵和时空数据矩阵作

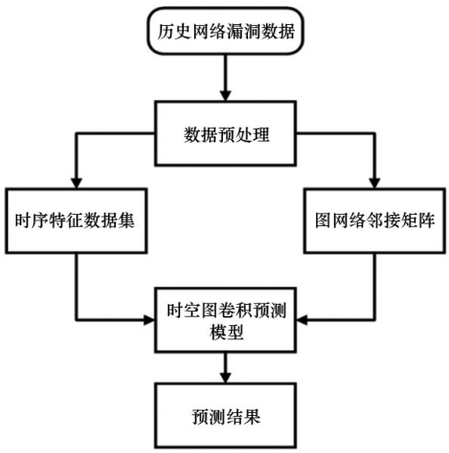

图1 网络漏洞态势预测流程框架

为图卷积网络的输入,通过图卷积网络强大的网络结构学习能力来达到直接获得空间相关性的目的。时间门控卷积模块以图卷积模块的输出结果为输入,使用多层堆叠的门控卷积单元来间接捕获时空数据的时空相关性,并使用全连接层来得到模型的最终预测结果。

如图2所示,图卷积模块为2层图卷积网络(GCN),输出维度依次为16、32;门控时间卷积模块为2层一维CNN,输出维度依次为32、64;2层全连接层,输出维度为64。

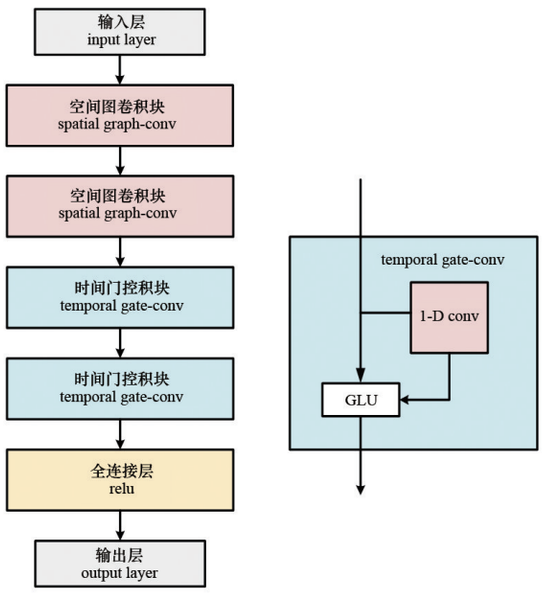

图2 时空图卷积模型结构与门控时间卷积模块

图卷积网络GCN是捕获网络结构信息的有效方法[25],它可以将节点邻域的信息聚集到节点自身,从而有效地提取网络中的空间结构信息,图卷积网络GCN在基于图结构的任务上取得了超常的性能,被广泛运用于各种网络数据挖掘任务中。通过叠堆多个GCN来构建空间图卷积模块,单层GCN的公式为

$$Z = \sigma f(X,A) = \sigma(AXW) \tag{2}$$

其中,X表示特征矩阵;A是邻接矩阵;W是需要学习的权重矩阵;$\sigma(\cdot)$是非线性激活函数。

时间卷积模块由多个门控时间卷积单元(GLU)叠堆而成,而门控时间卷积单元则由一维卷积神经网络层CNN和门控线性单元GLU组合而成,一维CNN用来捕获时空数据中的时间信息,门控线性单元GLU用来选择哪些时间信息会被保留[26]。相比RNN等时间序列预测神经网络耗时的迭代以及对动态变化的响应缓慢等问题,GLU有训练速度快且不受前几步依赖约束的优势。门控时间卷积单元表达式为

$$C \times_{\circ} R = P \odot \delta(M) \tag{3}$$

其中,C表示一维CNN的卷积核;R表示输入的特征向量;P、M分别表示经过CNN后的输出矩阵,将P、M作为GLU的输入来得到最终的输出结果;$\delta(M)$表示GLU中的门控单元,用来对输入P的时间信息进行筛选,从而保留其中与时间密切相关的信息,其中,$\delta(\cdot)$表示sigmoid函数,$\odot$表示哈达玛积。

此外,本研究在时间卷积模块后接入多个全连接层来生成未来若干个时刻的预测值,通过增加2层全连接层网络来增强模型的能力。

2 数据预处理

本研究数据来自国内某市网安支队提供的某市近6个月内的网络漏洞检测数据,如式(4)所示,单条网络漏洞数据E包括漏洞点ip、地理空间位置、漏洞描述、漏洞等级以及时间戳。

$$E = [ip, 经纬度, 描述, 等级, 时间戳] \tag{4}$$

数据包括网络特征、地理空间特征和时间序列,支持实现网络漏洞的时空预测。图3为某时间

段城市网络漏洞点分布示意图,网络漏洞分布明显受所在城市空间分布和城市功能划分的影响。

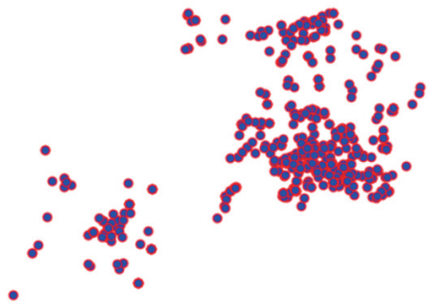

图3　网络漏洞点分布示意

如图4所示,网络漏洞数据时空结构化过程中,首先将网络漏洞数据去噪,去除缺乏经纬度信息及位于城市外范围的网络漏洞数据。依照网络漏洞数据的地理位置信息,进行地理网格划分,将网络漏洞数据所处空间范围划分为经度a、纬度b的空间网格。网络漏洞数据按时间段划分,以相近的网络漏洞检测时间段为范围,将各时间段的网络漏洞数据划分开,形成时间序列。网络漏洞等级数值加和,在各个时间段内,将各空间网格内的网络漏洞点的漏洞等级以不同数值表示,并加和获得时间段内网格内的整体网络漏洞等级,形成时间序列网络漏洞特征矩阵。构造地理关系邻接矩阵,将网格中网络漏洞时间序列不全为0的网格块作为图数据的G的顶点V_i。根据地理位置相关性构成图网络,将距离在$dist$内的点V_i间的关系$edge$设为1,构造邻接矩阵$\boldsymbol{A}$。

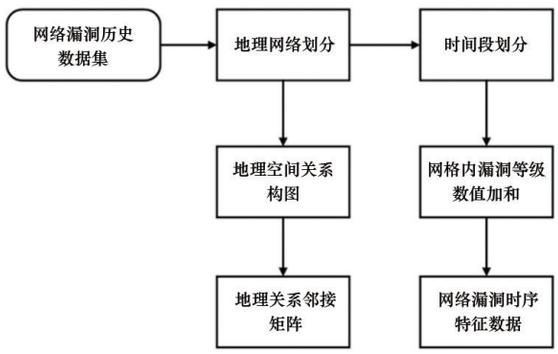

图4　数据处理流程图

图5为网络漏洞数据时空化构造过程。通过地理网格划分,取特征时序数据中非全0节点,构建地理关系图,生成对应的空间关系邻接矩阵。

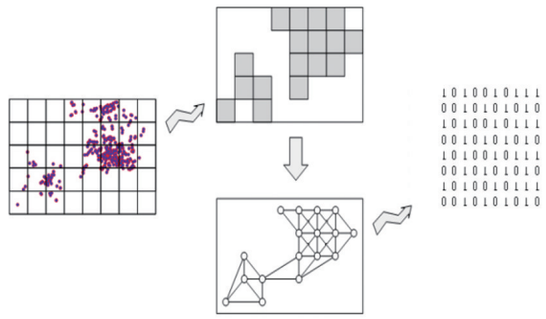

图5　时空数据构造过程

3　模型实验

数据预处理中,设置地理节点相关距离为1.5,地理网格经纬度间距为0.1,进行时空数据构造。网络漏洞监测数据预处理后,时空数据中节点44个,节点关系边264条,时间节点88个,特征数为1。

实验设置中,时间序列预测的模型分别对各点位的时间序列数据进行训练验证,将模型训练评价指标求平均值。时空预测模型实验中,按照时间顺序对所有数据集进行划分,将前80%用于训练,后20%用于验证。实验使用PyTorch深度学习框架搭建所提图网络模型,模型按照前期数据处理。在训练过程中,使用均方误差MSE作为损失函数,并采用随机梯度下降法进行迭代反向传播学习,优化器使用Adam,固定迭代次数为200次,选取验证集误差最小的输出作为预测结果。

对实验选取了差分整合移动平均自回归模型(autoregressive integrated moving average model,ARIMA)和LSTM作为对比模型。差分整合移动平均自回归模型结合自回归和移动平均2种方法,可对平稳时间序列或经过差分处理后转换为平稳时间序列的非平稳时间序列进行预测。LSTM解决了RNN无法处理的长期依赖问题,被广泛应用于时间序列预测任务中。由于对比模型进行时间序列上的预测,故实验中将各点位的时间序列分别进行模型训练验证,最后对训练预测结果求平均值。

模型目标是使预测数据尽可能地接近真实数

据,选择平均绝对误差(MAE)、均方根误差(RMSE)、平均绝对百分比误差(MAPE)3种预测评价指标。

实验设置为12个时间点预测之后4个时间点的短时预测,表1为对比实验结果,可以看出本研究提出的时空图卷积预测模型在MAE、RMSE和MAPE 3种评价标准下都优于对比模型ARIMA和LSTM,有更高的预测精度,在对网络漏洞数据的时空预测上具有较好的性能。

表1 对比实验结果

模型	MAE	RMSE	MAPE
ARIMA	48.46	56.31	68.69
LSTM	59.76	65.95	56.70
时空图卷积	9.37	17.29	1.42

对比仅对时间序列进行预测的ARIMA和LSTM模型,本模型对于加入地理空间特征的网络漏洞时空数据的时空序列预测依然有着较为明显的优势,可见卷积神经网络对于时空预测特征提取的效果。同时,网络漏洞时序数据具有很强的非线性,对比考虑时间序列数据长期依赖的另外2个模型,时空图卷积模型使用一维卷积还可以保留模型的非线性,大幅提高预测精度。

网络漏洞态势预测结果为未来几个时间片的地理空间网络漏洞态势,图6为一个时间片的预测结果示意图,网络漏洞数据的预测结果受地理空间的时间序列数据的影响,需考虑时空预测中时间与空间特征共作用。对比模型ARIMA和LSTM,时空图卷积模型可以通过获取邻居网格站点的特征信息,提升模型的预测精度,同时也减少了对历史时间数据的依赖,故在得到信息较少的短时预测中有更好的非线性,可以实现较少数据下的准确预测。

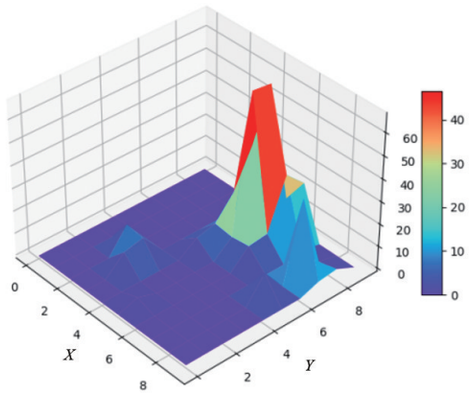

图6 预测结果示意

4 结论

针对现有网络空间预测研究少有将地理空间关系作为特征加入的现状,考虑网络空间数据与地理空间关系性,选取拥有地理空间特征的网络漏洞监测数据,根据其地理空间位置信息关系构造了时空特征数据。提出一种结合图卷积和门控卷积的时空图卷积预测模型,对构造的时空特征数据进行训练预测。选取ARIMA和LSTM时序预测模型进行对比实验,本研究提出的网络漏洞时空图卷积预测模型在MAE、RMSE和MAPE的评价标准下显示出更好的预测效果,表明了使用时空图卷积对网络—地理数据进行时空预测的有效性。

网络空间要素是依附于地理空间存在的,网络要素结合地理空间的研究方法应更普遍地应用于网络空间态势感知和网络空间数据预测的相关研究中。时空图卷积模型在时序预测中能够有效地从邻居节点获取历史特征信息进行预测,减少了对历史数据的依赖,在数据较少的短时预测中会有较好的效果,对于要求短时、快速、准确的网络空间预警的研究场景会有较好的应用。例如网络空间监管的推荐系统,网络监管者可以自主加入地理空间或网络空间关系信息构建关系网络要素图网络。通过增加网络要素的关联信息,提高网络空间态势预测的准确性。推荐系统可对高密度网络威胁位置发出警报,也可推荐最可能的网络攻击行为。

参考文献(References)

[1] 张焕国, 韩文报, 来学嘉, 等. 网络空间安全综述[J]. 中国科学: 信息科学, 2016, 46(2): 125-164.

[2] 龚俭, 臧小东, 苏琪, 等. 网络安全态势感知综述[J]. 软件学报, 2017, 28(4): 1010-1026.

[3] 肖喜生, 龙春, 彭凯飞, 等. 基于人工智能的安全态势预测技术研究综述[J]. 信息安全研究, 2020, 6(6): 506-513.

[4] 席荣荣, 云晓春, 金舒原, 等. 网络安全态势感知研究综述[J]. 计算机应用, 2012, 32(1): 1-4.

[5] Endsley M R. Design and evaluation for situation awareness enhancement[C]//Proceedings of the Human Factors Society annual Meeting. Los Angeles, CA: Sage Publications, 1988, 32(2): 97-101.

[6] Bass T. Multisensor data fusion for next generation distributed intrusion detection systems[C]//Proceedings of the IRIS National Symposium on Sensor and Data Fusion. Laurel, MD: Citeseer, 1999, 24(28): 24-27.

[7] Man D, Wang Y, Yang W, et al. A combined prediction method for network security situation[C]//2010 International Conference on Computational Intelligence and Software Engineering. Wuhan, China: IEEE, 2010: 1-4.

[8] Wang Y, Li W, Liu Y. A forecast method for network security situation based on fuzzy Markov chain[J]. Lecture Notes in Electrical Engineering, 2014(260): 953-962.

[9] Liang W, Long J, Chen Z, et al. A security situation prediction algorithm based on HMM in mobile network[J]. Wireless Communications and Mobile Computing, 2018, 2018: 1-11.

[10] 李欣, 段詠程. 基于改进隐马尔可夫模型的网络安全态势评估方法[J]. 计算机科学, 2020, 47(7): 287-291.

[11] 孙卫喜. 用于网络安全态势预测的粒子群与支持向量机算法研究[J]. 计算机应用与软件, 2019, 36(6): 308-316.

[12] Lu H, Zhang G, Shen Y. Cyber security situation prediction model based on GWO-SVM[C]//13rd International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. Cham: IEEE, 2019: 162-171.

[13] Hu J, Ma D, Liu C, et al. Network security situation prediction based on MR-SVM[J]. IEEE Access, 2019, 7: 130937-130945.

[14] Xiao P, Xian M, Wang H. Network security situation prediction method based on MEA-BP[C]//2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT). Chengdu, China: IEEE, 2017: 1-5.

[15] Zhang X, Ye Z, Yan L, et al. Security situation prediction based on hybrid rice optimization algorithm and back propagation neural network[C]//2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS). Lviv, Ukraine: IEEE, 2018: 73-77.

[16] 胡昕. 基于RNN的网络安全态势预测方法[J]. 现代计算机, 2017(6): 14-16.

[17] Feng W, Wu Y, Fan Y. A new method for the prediction of network security situations based on recurrent neural network with gated recurrent unit[J]. International Journal of Intelligent Computing and Cybernetics, 2020, 13(1): 25-39.

[18] Fang X, Xu M, Xu S, et al. A deep learning framework for predicting cyber attacks rates[J]. EURASIP Journal on Information Security, 2019, 2019(1): 1-11.

[19] Fan S, Wu S, Wang Z, et al. Aleap: Attention-based LSTM with event embedding for attack projection[C]//2019 IEEE 38th International Performance Computing and Communications Conference (IPCCC). London, UK: IEEE, 2019: 1-8.

[20] Kishioka K, Hongyo K, Kimura T, et al. Prediction method of infection spreading with CNN for self-evolving botnets[C]//2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). Hawaii, USA: IEEE, 2018: 1810-1815.

[21] 何春蓉, 朱江, 张欣. 基于复杂样本的安全态势要素分类架构[J]. 重庆邮电大学学报(自然科学版), 2022, 34(4): 719-727.

[22] 高春东, 郭启全, 江东, 等. 网络空间地理学的理论基础与技术路径[J]. 地理学报(英文版), 2019, 29(12): 1949-1964.

[23] 郭启全, 高春东, 孙开锋, 等. 基于"人-地-网"关系的网络空间要素层次体系建设[J]. 地理研究, 2021, 40(1): 109-118.

[24] 王奕钧. 网络空间地理图谱在城市网络安全综合管控中的应用研究[J]. 信息安全研究, 2022, 8(8): 801-811.

[25] Henaff M, Bruna J, Lecun Y. Deep convolutional networks on graph-structured data[J]. arXiv preprint, 2015, arXiv: 1506.05163.

[26] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. arXiv preprint, 2018, arXiv: 1803.01271, 2018.

Network vulnerability situation prediction based on spatio-temporal graph convolution

ZHANG Yingchun[1], LI Jin[2], ABDUREYIM Raxidin[3], ZHANG Xun[2,3*], HAO Mengmeng[4], JIANG Dong[4]

1. School of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China
2. School of Computer Science and Engineering, Beijing Technology and Business University, Beijing 100048, China
3. School of Mathematics and Information, Hotan Normal College, Hotan 848099, China
4. Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

Abstract In view of the increasingly serious problem of network security, geographical space features are added into the prediction process to realize spatio-temporal prediction of network space elements in this study. Considering the research status that network data are often rarely combined with geospatial characteristics in the prediction process of network security elements, network vulnerability detection data with geospatial characteristics are also selected to construct the spatio-temporal data set of network vulnerabilities. By constructing a spatio-temporal graph convolution model combining graph convolution and gated time convolution, the development of network vulnerability situation can be predicted. ARIMA and LSTM temporal prediction models are selected for comparative experiments, and the proposed network vulnerability spatio-temporal graph convolution prediction model shows better prediction effect under MAE, RMSE and MAPE evaluation criteria.

Keywords cyberspace data; geography space; spatio-temporal data; spatio-temporal graph convolution; prediction model

（责任编辑　傅雪）