

刘元浩^{1,2} 曹婧¹ 沈华伟¹ 黄俊杰¹ 程学旗¹

(1. 中国科学院计算技术研究所 网络数据科学与技术重点实验室, 北京 100190; 2. 南开大学 软件学院, 天津 300071)

论文摘要

在COVID-19疫情的防控工作中,对疫情传播过程中确诊人数的预测工作具有重要意义。在现有疫情传播预测工作中,以SEIR(Susceptible-Exposed-Infected-Recovered)模型为代表的传染病模型能反映疫情相关人群人数变化,但由于其人群均匀接触的前提假设,模型的应用具有局限性。基于时间序列分析的模型可以通过简单建模历史确诊人数的时间序列对当前确诊人数进行预测,但缺乏对传染病传播的传染性、爆发性、衰减性等固有性质的认识,对疫情发展趋势变化的预测能力受到制约。为解决上述问题,该文采用基于自增强泊松过程(Reinforced Poisson Process, RPP)的模型对疫情确诊人数进行预测,考虑病毒传染性、级联传染的自增强效应和病毒传播的时效性等三个关键因子,对疫情传播的动态过程进行建模,从而对确诊人数做出预测。实验证明,相较SEIR模型,使用RPP模型进行疫情预测不依赖人群均匀混合假设,在各尺度的地理区域都有稳定且准确的预测结果,也解决了SEIR模型在后期预测值过高的问题;对比时间序列分析模型,RPP模型能够掌握疫情发展的内在规律,对疫情发展前、中、后期的发展趋势预测误差分别减小5.29%、5.04%、0.47%,并且能准确把握疫情发展的重要阶段性变化。该文方法已应用于线上平台实时疫情预测,平均误差率小于0.5%。

实践应用

本文方法也投入到了实际的防控应用中,自1月29日起先后对中国(全国、全国除湖北、湖北除武汉、武汉)、日本、韩国、意大利、伊朗、美国、法国、德国、西班牙等九个地区共12个地区的疫情发展趋势进行预测,累积确诊人数平均误差率小于0.4%。预测结果发布于中科天玑智疫通线上平台(<https://ncov.ictbda.com/#/>)。

国家	总预测天数 (截至7月26日)	MAPE
中国	162	0.05%
伊朗	137	0.54%
意大利	137	0.51%
韩国	137	0.15%
日本	90	0.80%
法国	112	0.72%
美国	112	0.42%
西班牙	112	0.25%
德国	112	0.19%



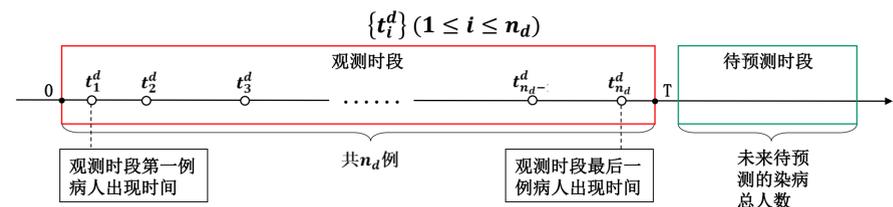
论文简介

自2019年底以来,新型冠状病毒肺炎疫情的持续蔓延对人们的生命安全造成巨大威胁,也对国家医疗建设、物资调配、隔离管控等方面带来挑战,而对疫情传播进行建模并对确诊病例数的增长进行及时准确地预测,对于医疗卫生资源的分配、防控重点的调整等具有重要的参考价值,并能够有效地对防控措施有效性进行合理评估。

目前主流的疫情预测方法包括基于微分方程的传染病模型和基于时间序列分析的疫情预测模型,然而上述模型在应用中存在各自的局限性。本文采用自增强泊松过程(RPP)模型(Shen et al., 2014)对疫情确诊人数变化趋势进行预测,将病毒感染人群的动态过程建模为不均匀泊松过程,对病毒传染性、级联传染的自增强效应和病毒传播的时效性等疫情传播中的三个关键因子进行建模,在本次COVID-19疫情传播数据上进行的实验,证明了该模型对于疫情传播预测的有效性。

模型方法

对于某传染病d,我们将其在时间段[0, T]内的疾病感染人群动态变化过程表示为个体染病事件发生的时间序列,并根据该观测序列预测传染病d在未来一段时间内的感染人群总数。

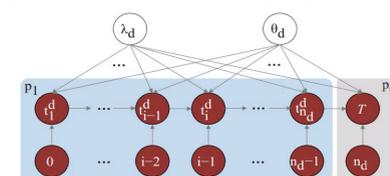


- 疾病传播过程中的三大现象:
 - 病毒传染性,即病毒自身的传染性对最终的感染人数起决定作用;
 - 级联传播的自增强效应,即病毒当前的感染人数越多越容易进行新的传播感染;
 - 病毒传播的时效性,即随时间推移,病毒感染人群继续感染他人可能性会下降。
- 将疾病d感染人群的动态过程建模为一个自增强泊松过程,速率为:

$$x_d(t) = \lambda_d f_d(t; \theta_d) i_d(t)$$

病毒自身传染性 病毒传播速率随时间的演变过程 病毒d在时刻t已感染的人数

整个疾病感染人群的动态过程可由下列概率图模型表示:



疾病感染人群动态过程的产生式概率图模型*

*Shen Huawei, Wang Dashun, Song Chaoming, et al. Modeling and Predicting Popularity Dynamics via Reinforced Poisson Processes. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI'14). Canada: 2014: 291-97.

实验结果

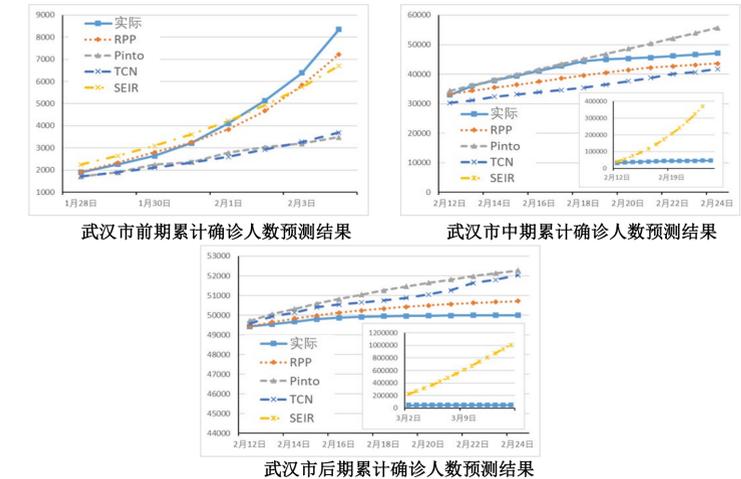
对比本文模型与SEIR模型、Pinto模型、TCN模型在COVID-19疫情预测中任务中的表现。

1) RPP模型在同一时期不同地区预测误差小,预测效果稳定

各模型2月1日至2月7日预测结果MAPE				
	RPP	Pinto	TCN	SEIR
全国	1.86%	8.44%	2.85%	108.52%
全国(除湖北)	3.62%	20.72%	4.27%	46.73%
湖北(除武汉)	3.54%	22.08%	8.25%	53.79%
武汉	6.82%	17.84%	8.22%	20.40%

各模型2月11日至2月17日预测结果MAPE				
	RPP	Pinto	TCN	SEIR
全国	2.10%	9.22%	4.16%	>1000%
全国(除湖北)	1.61%	4.09%	1.95%	474.77%
湖北(除武汉)	2.90%	5.52%	8.70%	256.27%
武汉	3.46%	3.13%	12.87%	90.33%

2) RPP模型在同一地区不同疫情发展阶段均能准确预测疫情发展趋势



论文结论

本文应用自增强泊松过程(RPP)模型来预测COVID-19的疫情确诊病例数。我们的实验结果表明,RPP模型在预测疫情确诊人数的任务中明显优于传统的传染病模型和时间序列分析模型。在空间上,RPP模型克服了SEIR模型基于人群均匀混合的局限,在各尺度的地理区域都有稳定且准确的预测结果。在时间上,一方面,RPP模型解决了SEIR模型在人口总数很大时累计确诊人数持续增长的问题;另一方面,RPP模型通过建模疫情发展过程中的关键因素,摆脱了时间序列分析模型仅对历史数据建模的局限性,从而对疫情发展各个阶段的疫情走势能够进行更精确的预测,并且能准确把握疫情发展的重要阶段性变化,其结果在实际应用更具有参考价值。