

1. 选题的目的和意义

选题目的:

本选题来源于一个实际问题:通过全世界各地近 70 年观测到的温度预测将来世界各地温度可能的走向与趋势并对未来短时间内温度进行精确预测。为了解决该问题,本课题提出可解性的复杂网络方法,该方法具有一定的创新性,可以实现精度和可解释度的均衡。研究结果可以对于今后的全球气温变化提供参考。本选题的目的主要如下:

1. 改进传统稀疏辨识方法不能处理大规模数据、泛化性能不好的问题。通过设计一个“两高一低”(计算复杂度低、稀疏度高、可解释性高)的稀疏辨识的方法拟合从 1950 年到 2020 年全球各地气温随时间的分布规律,并根据拟合结果对不同地区今后 20 年的气温变化趋势做出预测。

2. 改进传统卷积神经网络方法过拟合、解释性差等问题。通过设计符合温度分布的时空卷积神经网络算法,结合 1950 年到 2020 年全球温度数据,给出全球今后 1 年温度的精确预测。

3. 通过辨识结果,计算真实温度与之的残差。并将残差放入卷积神经网络计算,实现稀疏辨识方法与卷积神经网络预测结果耦合,在模型上实现精度和可解释度的均衡。在实际预测中实现长期温度趋势预测和短期温度精确预测的兼顾。

4. 根据模型得出的结果给出一些指导性意见,通过全球各地温度变化趋势,对某些气候变暖的国家发出可持续化倡议;通过分近一年来可能发生的气温突变做出预判,并倡议予以防范和警告。

选题意义:

由于该选题来源于实际问题且应用的方法与模型在理论上也有创新,因此该选题兼具理论意义和现实意义。

本课题的理论意义主要体现在复杂网络和稀疏辨识两个方面。在复杂网络方面,自从 2012 年 Alex 教授利用其自己改进的卷积神经网络——AlexNet[1]获得 ImageNet 比赛的冠军后,复杂卷积神经网络在学界和业界均获得了长足的发展:翻译[2]、棋局[3]、图像处理[4]、自然语言处理[5]在此之后均取得了突破性的进展。可惜的是,鉴于网络结构的复杂性,网络中诸多参数难以得到解释[6]。虽然精度远高于一些可解释性模型,但是泛化性却不足,经常遇到长时间的调参才能给出一个较好的学习结果这样的问题。这使得该网络难以直接用于气象数据的长期趋势预测。为了解决这一问题,本课题提出可解释的复杂网络模型。其意义在于在预测温度的同时,还能从参数中洞悉未来很长一段时间气温可能的发展趋势,大大提高了复杂网络在时空数据下的实用性。而在稀疏辨识方面,稀疏辨识一般通过在大量的基函数中寻找影响事件的主要基

函数并赋予权重的方法找到控制一件事的“主要因素”[7]。该方法具有较好的泛化性，很少陷入过拟合问题，在信号处理、压缩[8]等领域有着长足的发展。但是，稀疏辨识的结果并不具备较高的精度，不适合进行短期精确温度的预测。在本课题中，引入复杂网络与稀疏辨识的结果相耦合。在洞悉长期预测温度趋势的同时，也能给予短期温度精确预测，实现了精度与可解释度的均衡，在预测方法中具有丰富的理论意义。

在现实意义方面，大规模时空数据的预测十分在很多场景中都有必要。例如本文选取的温度数据：温度趋势的长期预测，和温度精确的短期预测一直被各国科学家所关注，能较为准确的预测这两点内容可以有效的避免灾难的发生。温度的长期预测可以洞悉温度变化趋势，通过温度变化趋势可以制定相关政策防微杜渐。例如，Root 在[9]中表示，近 100 年来，全球气温已经上升了 0.6°C ，而在局部地区（例如极地），这种现象更加严重。如果能够较为精确的预测每个地区温度变化趋势，制定合适的碳排放策略（因为导致全球变暖的主要源于是大量使用如煤、石油等矿物燃料引起的）将会大大减缓全球变暖的趋势。同时，我们还可以根据预测情况对于全球变暖严重的低海平面国家发出警告，防止国家周边被淹没。而在温度短期精确预测在很多极端天气预测中也都有举足轻重的地位。例如，我们熟知的 ENSO 是热带太平洋异常冷暖条件之间的年际波动，是地球上最具影响力的海洋——大气气候现象之一[10]。ENSO 会影响海洋生态系统，商业渔业，农业和公共安全，故而短时间精确预测 ENSO 的发生是十分重要的；又如，在去年 7·20 河南暴雨中，郑州的降水量超过了 500mm，正是因为有了提前的红色降雨精确预警，才避免了更大的损失[11]。因此，本课题研究的结果对于在现实中进行气候趋势预测、短期极端天气预测均有较强的现实意义，有必要深入研究。

综上所述，该选题内容兼具理论意义和现实意义，适合进行深层次的研究。

2. 国内外研究现状

大规模时空数据预测一直是国内外学者研究的重点内容。一般来说，传统的时空预测方法分为两大块，一种基于物理模型（model-based）的预测方法，另一种基于纯数据（data-driven）的预测方法。这两种方法在国内外研究中均有长足的发展。

基于物理模型（model-based）的时空数据预测方法非常多。本文主要介绍通过辨识构造物理模型，并进行预测时空数据的方法，国内外辨识方法的发展主要分为三个阶段。

(1) 最早的辨识方法是在 19 世纪末 Galton 提出的[12]。他发现父母身高可以预测孩子的身高，于是将孩子的身高与父母的身高这一现象拟合成线性关系，形成一元线性回归。这个方法也开启了通过辨识方法预测数据的先河，人们在进行时间上的预测时仅需要将因变量 y 与时间 t 进行拟合，就可以较为简单的构建模型进行预测。但随后，人们发现因变量 y 可能给会和很多时间变量（例如 t_1, t_2, \dots, t_n ）有关，故而在 20 世纪初进一步地提出多元线性回归，它描述了因变量与多个自变量的关系，可以更加精确的预测随时间发展的数据。该方法即为原始的辨识方法，在当时医疗、生物、社

会数据分析[13]中均有较好的效果。

(2) 20 世纪末, 随着大数据时代到来, Robert 指出, 传统辨识方法自变量过多, 会产生以下缺陷: 模型结果中大部分自变量前系数过小, 很多自变量可能与因变量并无直接联系。他提出是否能舍去某些自变量以达到简化问题的效果(即让某些自变量前系数为 0)。这就是稀疏辨识问题, 旨在给定误差范数小于 ε 的情况下, 最小化系数的 l_0 范数, 这是一个 NP 问题, 当时人们只能寻找其近似解。Mallt 提出了基于贪心思想的正交匹配追踪(OMP)算法[14], 该方法将一个 NP 问题转化为一个多项式时间内可以求解的问题, 并给出了该问题的近似解。但是由于其用的是贪心的思想, 实际的稀疏效果泛化性能较差, 即若真实数据有较小的扰动, 结果将会发生较大的变化。Robert 尝试对原问题条件修改, 将原问题中最小化系数的 l_0 范数分别改为 l_1 范数与 l_2 范数, 分别提出 LASSO 和岭回归方法给出了稀疏解[15], 该方法可以在较短时间内寻到 l_1 范数与 l_2 范数意义下的全局最优解。但是 Robert 并没有解决原问题, 而是解决了原问题的近似问题, 虽然结果符合稀疏辨识的期望, 但是结果并非原问题的最优解。

(3) 进入 21 世纪, 更多学者致力于加强该问题的条件将该问题彻底解决。2001 年, Tipping 等人提出相关向量机(RVM)[16], 从贝叶斯学习的角度进行回归, 开创了稀疏贝叶斯学习的先河。该方法给了原问题一个先验假设: 即系数与误差均满足正态分布先验, 并利用贝叶斯方法根据数据计算其后验分布。该方法在其假设条件下又快又准的解决了原问题, 但是该方法比较依赖误差项是正态分布的假设, 若已经有先验信息知道不满足, 则 RVM 失效。随后十年, 诸多学者对于不同的先验假设进行研究, 并利用稀疏贝叶斯方法给出结果。至 2012 年, 稀疏贝叶斯方法在稀疏辨识领域已经相当成熟, Liu 等人[17]为总结常用的稀疏贝叶斯方法。2016 年, Brunton 等人[18]提出阈值最小二乘法用于解决动力学辨识问题, 成功通过稀疏辨识方法拟合了复杂的常微分方程组。2018 年, 在北京智源大会的“人工智能的数理基础专题论坛”上, 北京大学教授董斌将稀疏辨识的方法正式用于偏微分方程反问题的求解他构造的 PDE-Net 可用于通过少量数据迅速求解偏微分方程, 这开创了人工智能求解偏微分方程的先河。自此, 稀疏辨识不再仅仅构造简单的代数关系, 在常微分关系、偏微分关系等动力系统的构造中也有亮眼的表现。2019 年, Yuan 等人[19]提出了 HIYDE 模型, 完成了分段动力学辨识问题。该方法可以辨识多个系统之间的切换, 这将动力学系统的稀疏辨识的方法运用的更加贴合实际。2020-2021 年, 很多学者将已发展的动力学系统辨识与时空数据预测方法用于实际, 并解决了很多现实生活的难题[20]。

基于大数据(data-driven)的时空数据预测方式是近年来新兴起的一种预测方式, 它并不需要太多的机理做支撑, 仅需要大量的数据放入计算机, 就可以让计算机自行的学习, 形成一种端到端的预测模式。data-driven 的预测方式是随着人工神经网络的发展而兴起的。它的发展可以分为两个阶段:

(1) 可用于较为精确预测的多层神经网络课追溯到 1974 年, Paul 在哈佛大学攻读博士学位期间, 在其博士论文中发表了关于误差反向传播神经网络(即我们常说的 BP 神经网络学习算法), 可是当时并未引起重视。1982 年, John 提出了连续和离散的 Hopfield 神经网络模型[21], 他利用该网络给当时的 NP-hard 问题——商旅问题(即

TSP 问题)提出了一个可行解,好于目前的绝大多数可行解。这引起了世界范围内的极大关注。此后一段时间,对 BP 神经网络的研究络绎不绝,很多关于 BP 神经网络著名的文章均出于此时期。该网络用于预测、拟合、分类的效果均好于传统的方法,因此,在传统模型训练大规模时空数据的预测无能为力时,往往 BP 网络可以显得得心应手。但是,该方法也有一定的问题,其中比较突出的几个问题是:层数少、梯度爆炸、梯度消失、过拟合、需要频繁调参。在 20 世纪 90 年代至 21 世纪初,随着统计学习的迅速发展,神经网络逐渐被 Vapnik 发明的支持向量机(SVM)所取代[22]。此后十年, data-driven 的预测方式逐渐被遗忘。

(2) data-driven 的预测方式再一次兴起是因为在 2012 年 ImageNet 的比赛中, Alex 教授利用其自己改进的卷积神经网络——AlexNet 以极大的优势获得冠军(高于第二名约 10 个百分点)[1]。他通过卷积的预处理方式使得训练时间降低的同时大大提高了准确率和泛化性。在 2013 至 2017 年期间,很多研究聚焦于卷积神经网络(CNN), CNN 不仅在 ImageNet 中大放异彩(例如谷歌与 2014 年提出的 VGG, 2015 年 ImageNet 冠军 ResNet 等),而且在各种预测、翻译、对弈中均展现出了远高于传统水平的表现,神经网络学习进入了深度学习时代。深度学习中的“深度”体现在对于复杂网络的训练时间长、网络层数多上,这解决了很多 BP 神经网络中无法解决的问题。2016-2017 年,利用深度学习构造的围棋机器人 AlphaGo 先后击败九段围棋大师李世石、世界围棋第一人柯杰[3],再一次震惊了世界;同一时间,谷歌提出了一种新的翻译方式 Transformer[2],仅用了 attention 机制就完全颠覆了传统的翻译方法,将翻译精度提升了好几个量级。正是因为复杂网络在各个领域中有着极大的优势,很多学者利用深度学习的方法将时空预测很好的应用在社会各个方面。Jing 等人[23]设计了一个基于自动编码器的深度模型预测蜂窝网络的时空数据; Jin 等人[24]设计了地理语义集成神经网络(GSEN)进行城市热点时空预测;而 Elsaadani[25]等人则通过 CNN 和长短时记忆网络(LSTM)的结合 ConvLSTM 预测了土壤水分的时空变化量。这体现了复杂网络在时空预测中的优势。

近年来,有些学者尝试将两种传统的预测方法结合,形成一种复杂的时空数据预测方式。这种预测方式兼具传统预测方法的优点,既能取得较好的精度,又能兼具一定的可解释性。例如,在 model-based 方面,在 Brunton 提出 SINDy 模型完成了常微分关系的动力学辨识方法后, Kaiser[26]等人于 2017 年提出将 SINDy 和 MPC(模型预测控制)结合,形成了一种与 data-driven 的复合模型,可以用于少量数据情形下的精确预测;类似地,在 data-driven 方面,在提出 Transformer 后, Beltagy 等人[27]于 2020 年提出了一种随序列长度线性扩展的注意机制结合入 Transformer,这种结合 model-based 方法具有更好的翻译效果;同年, Choromanski 等人[28]基于正交随机特征(FAVOR)的快速注意力机制,这种机制可以给出某些情况下的稀疏解,进一步拓展了 transformer 的适用性,有实用意义价值。基于此,本文主要针对全球气温数据进行辨识与预测,该数据集是一个大规模的时空数据,我们旨在克服辨识和神经网络各自的缺点,形成一个一体化的预测模型,得到比使用单独模型更好的效果。

3. 研究内容
<p>1. 当下研究时空数据的背景介绍，阐明当下时空预测的问题和瓶颈所在；</p> <p>2. 稀疏辨识方法的阐述与应用（包括正交匹配追踪、稀疏贝叶斯方法、LASSO、阈值最小二乘等方法）；</p> <p>3. 复杂卷积神经网络算法的基本介绍和应用（包括空间卷积网络 CNN，VGG，ResNet 以及时间循环网络 RNN，LSTM，GRU 等）；</p> <p>4. 阐明将辨识方法与复杂网络方法相耦合的过程；</p> <p>5. 基于理论，用于实际数据，基于全球基站往年温度记录，对于今后温度进行评估与预测。</p>
4. 实施方案、进度安排及预期效果
<p>2021.11~2022.01：学术调研，查阅文献，熟悉数据并进行简单的数据预处理，熟练掌握稀疏辨识和各种神经网络中的算法原理，思考数据处理方式，做好论文准备阶段的规划为下一阶段的研究做准备；</p> <p>2021.02：大量阅读与稀疏辨识和复杂网络相关的相关文献，从发展历程和算法核心两个角度了解、学习知识。确定论文标题，在导师的指导下提交论文标题、任务书和开题报告。</p> <p>2021.03：确定论文写作模块，并逐步开始论文写作。与此同时开始编程，对于自己的每一个想法进行编程并给出预测结果。力求在三月底基本完成初稿内容，向老师提交初稿并检查。</p> <p>2021.04：在导师的指导下，检查论文结果是否严谨，数据是否有误，并多次对初稿做进一步的检查、修改，并最终完善定稿。</p> <p>2022.05：在确定无误后准备 PPT 进行答辩。</p>
5、已查阅参考文献：
<p>[1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C] International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.</p> <p>[2] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[J]. Computing Research Repository, 2017: 1706.03762.</p> <p>[3] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge[J]. Nature, 2017, 550(7676):354-359.</p> <p>[4] Zhang K, Zuo W, Chen Y, et al. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising[J]. IEEE Transactions on Image Processing, 2017. 26:3142-3155.</p> <p>[5] Hassan A, Mahmood A. Efficient Deep Learning Model for Text Classification Based</p>

- on Recurrent and Convolutional Layers[C] 16th IEEE International Conference On Machine Learning And Applications (ICMLA2017). IEEE, 2017.
- [6] Kim Y. Convolutional Neural Networks for Sentence Classification[J]. Computing Research Repository, 2014: 1408.5882.
 - [7] Brunton S L, Proctor J L, Kutz J N. Discovering governing equations from data: Sparse identification of nonlinear dynamical systems[J]. Proceedings of the National Academy of Sciences of the United States of America, 2015, 113(15):3932.
 - [8] Kopsinis Y, Slavakis K, Theodoridis S. Online Sparse System Identification and Signal Reconstruction Using Projections Onto Weighted Balls[J]. IEEE Transactions on Signal Processing, 2011, 59(3):936-952.
 - [9] Root T L, Price J T, Hall K R, et al. Fingerprints of global warming on wild animals and plants. [J]. Nature, 2003, 421(6918):57-60.
 - [10] Meng J, Fan J, Ludescher J, et al. Complexity-based approach for El Nio magnitude forecasting before the spring predictability barrier[J]. Proceedings of the National Academy of Science, 2020, 117.
 - [11] Zhang P, Wang L, Zhang N. Study on the Early Warning System for Rainstorm Water logging in Wuhan City and its Simulated Results[J]. Torrential Rain and Disasters, 2010.01:29.
 - [12] Galton F. The geometric mean, in vital and social statistics[J]. Proceedings of the Royal Society, 1897:365-7.
 - [13] Hocking R R. A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression[J]. Biometrics, 1976, 32(1):1-49.
 - [14] Pati Y C, Rezaiifar R, and Krishnaprasad P S, et al. Orthogonal Matching Pursuit-Recursive Function Approximation with Applications to wavelet decomposition, Proc. 27thAnnu. Asilomar Conf. Signals, Systems, and Computers, Pacific Grove, CA, Nov.1993,1:40-44.
 - [15] Tibshirani R. Regression Shrinkage and Selection Via the Lasso[J]. Journal Of The Royal Statistical Society: Series B (Methodological), 1996,58(1), 267-288.
 - [16] Michael T. Relevance vector machine: WO0118667A9[P], 2002.
 - [17] Liu B, Fan H, Lu Z, et al. The Annealing Sparse Bayesian Learning Algorithm[J]. Computer Science, 2012, 1209.1033.
 - [18] Brunton S, et al. Discovering governing equations from data by sparse identification of nonlinear dynamics. PNAS, 2016, 113(15):3932-3937.
 - [19] Yuan Y, Tang X, Zhou W, et al. Data driven discovery of cyber physical systems[J]. Nature Communications, 2019, 10(1).
 - [20] BMD Silva, Champion K, Quade M, et al. PySINDy: A Python package for the Sparse Identification of Nonlinear Dynamics from Data[J]. The Journal of Open Source Software, 2020, 5(49):2104.
 - [21] Hopfield J. Neural Computation of Decisions in Optimization Problem[J]. Biological Cybernetics, 1985, 52.
 - [22] Guyon I, Weston J, Barnhill S, et al. Gene Selection for Cancer Classification using Support Vector Machines[J]. Machine Learning, 2002, 46(1-3):389-422.
 - [23] Jing W, Jian T, Xu Z, et al. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach[C] IEEE INFOCOM 2017 -

- IEEE Conference on Computer Communications. IEEE, 2017, 10.1109.
- [24] Jin G, Sha H, Feng Y, et al. GSEN: An Ensemble Deep Learning Benchmark Model for Urban Hotspots Spatiotemporal Prediction[J]. Neurocomputing, 2021, 455(4).
- [25] Elsaadani M, Habib E, Abdelhameed A M, et al. Assessment of a Spatiotemporal Deep Learning Approach for Soil Moisture Prediction and Filling the Gaps in Between Soil Moisture Observations[J]. Frontiers in Artificial Intelligence, 2021, 4:636234.
- [26] Kaiser E, Kutz J N, Brunton S L. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit[J]. Proceedings of The Royal Society A Mathematical Physical and Engineering Sciences, 2017, 474(2219).
- [27] Beltagy Lz, Matthew E Peters, and Arman C. Longformer: The long-document transformer[J]. Computation and Language, 2020, 2004.05150.
- [28] Choromanski K et.al. Masked language modeling for proteins via linearly scalable long-context transformers. Computing Research Repository, 2020, 2006.03555.