面向火电控制系统辨识的循环神经网络可视分析

纪连恩^{1,2)}, 杨云^{1,2)}, 邱诗荣^{1,2)}, 王艺^{1,2)}, 田彬^{3,4)}

¹⁾ (中国石油大学(北京)石油数据挖掘北京市重点实验室 北京 102249)

²⁾ (中国石油大学(北京)计算机科学与技术系 北京 102249)

³⁾ (北京市电站自动化工程技术研究中心 北京 102211)

⁴⁾ (国能智深控制技术有限公司 北京 102211)

(jilianen@cup.edu.cn)

摘 要: 针对火电控制过程产生的数据连续性强、复杂度高, 循环神经网络模型行为与实际控制过程难以建立语义关联, 不能直观地进行模型的调试、优化和语义上的分析等问题, 将可视分析技术引入面向系统辨识的循环神经网络建模过程中, 提出可视分析系统 iaRNN. 首先, 通过可视化隐藏单元激活值分布与覆盖范围设计模型评估组合视图, 支持内外结合多方面评价模型性能; 然后, 从时序关系演变和敏感性分析等角度设计可视化视图, 以支持探索模型对控制参数的响应行为; 最后, 基于序列符号化和聚类分析提出了一种用于探索强时序依赖的实值时间序列与隐藏单元关联模式的可视化方法. 使用电厂真实数据进行案例分析, 验证了 iaRNN 在辅助用户理解模型工作机理和诊断模型缺陷方面的有效性.

关键词: 可视分析; 循环神经网络; 系统辨识; 火电控制

中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2021.19268

Visual Analytics of RNN for Thermal Power Control System Identification

Ji Lian'en^{1,2)}, Yang Yun^{1,2)}, Qiu Shirong^{1,2)}, Wang Yi^{1,2)}, and Tian Bin^{3,4)}

¹⁾ (Beijing Key Laboratory of Petroleum Data Mining, China University of Petroleum, Beijing 102249)

²⁾ (Department of Computer Science and Technology, China University of Petroleum, Beijing 102249)

³⁾ (Beijing Engineering Research Center of Power Station Automation, Beijing 102211)

⁴⁾ (China Energy Zhishen Control Technology Company Limited, Beijing 102211)

Abstract: Due to the problems such as strong continuity and high complexity of the data generated by the thermal power control process, and the difficulty of establishing a semantic correlation between the model behavior of recurrent neural networks and the actual control process, the model debugging, optimization and semantic analysis cannot be carried out intuitively. Visual analytics in recurrent neural networks modeling for system identification is applied and a visual analysis system called iaRNN is proposed. Firstly, by visualizing the activation value distribution and coverage of hidden units, the combined view for model evaluation is designed to support the evaluation of model performance in many aspects including inside and outside. Then, visual views are designed from the perspectives of temporal evolution, sensitivity analysis, etc, to support the exploration of model response behavior to thermal power control parameters. Finally, based on the symbolic representation of time sequences and clustering analysis, a method for exploring association

收稿日期: 2021-07-08; 修回日期: 2021-09-29. 基金项目: 国家自然科学基金(60873093); 国家能源集团 2030 重大项目(GJNY2030XDXM-19-08.2). 纪连恩(1972—), 男, 博士, 副教授, 硕士生导师, CCF 会员, 主要研究方向为可视化技术、人机交互、计算机图形学; 杨云(1995—), 女, 硕士研究生, 主要研究方向为可视化技术; 邱诗荣(1999—), 男, 硕士研究生, 主要研究方向为可视化技术; 王艺(1996—), 女, 硕士研究生, 主要研究方向为可视化技术; 田彬(1983—), 男, 硕士, 高级工程师, 主要研究方向为工业过程系统辨识与智能控制.

patterns between strong time-dependent real-valued time series and hidden units is proposed. A case study using real power plant data is conducted to verify the effectiveness of iaRNN in assisting users to understand the working mechanism of the model and diagnose model defects.

Key words: visual analysis; recurrent neural networks; system identification; thermal power control

目前, 循环神经网络(recurrent neural networks, RNN)及其变体已经广泛应用于各种与时间序列相关的任务中, 在语言识别、机器翻译和天气预测等领域取得显著性成果. 近年来, RNN 因具有强大的时序建模和非线性拟合能力, 在火力发电等工业控制过程的系统辨识中也取得了一定进展, 比传统的控制系统建模方法展现出精度更高和适用更广泛的潜力^[1].

火力发电厂主要由集散控制系统(distributed control system, DCS)对机组设备和系统的大量运行参数进行监测与控制, DCS 通常以秒为单位对这些参数进行不间断的采样和记录, 这些秒级数据中蕴含着丰富的控制信息, 为机组的运行过程建模和控制参数辨识提供了数据支撑^[2]. 在应用 RNN 对某个火电控制环节进行系统辨识的过程中, 其输入特征是由上述若干控制参数的大规模高密度采样值序列构成, 输出是某个受控参数的预测值序列, 输入/输出之间具有强时序依赖的特性, 即当前时刻受控参数的取值高度依赖于过去一段时间内控制参数的变化. 由于这些机组参数在实际的控制过程中瞬间波动较为频繁, 同时受机理规律和随机性影响, 难以直接发现和总结它们的变化与 RNN 模型表现之间的关联规律, 再加上 RNN 本身也因具有复杂的工作机理难以被控制领域人员理解, 因此, 针对这类工业系统辨识应用如何帮助领域人员快速、有效地构建和优化模型面临一系列特殊问题.

快速发展的可视分析技术在解决上述问题方面发挥着关键作用, 许多可视分析人员尝试将领域知识融入建模分析过程^[3], 结合人的视觉感知能力和推理能力, 帮助领域专家对模型进行更好的理解、诊断和优化. 已有 RNN 模型的可解释性工作大多围绕着以自然语言处理为代表的符号化序列建模问题进行, 如 RNNVis^[4]和 LSTMVis^[5]是通过离散化的单词或其他符号本身自带的语义信息解释隐藏单元的状态和变化. 但使用 RNN 对火电控制系统进行辨识时, 作为模型输入/输出的机组参数主要是在一定范围内频繁变化的实值时间序

列, 每个时刻的数据不具有上述类似的高度离散化特征和明确的语义信息, 导致难以直接对模型输入/输出进行有效理解和解释, 进而无法直观地分析隐藏单元与这些控制参数之间的关联模式.

在非符号序列建模分析中, 通过将实数类型的天气特征按照其数值范围进行类别划分, MultiRNN-Explorer^[6]实现输入/输出序列的离散化分组, 能够帮助用户探索和解释 RNN 在多维时间序列上的表现. 但由于大气监测数据采样稀疏, 时序依赖性较弱, 因此可以单个时间步提取模型输入特征, 并以此为基础分析隐藏单元对不同输入特征的响应. 而在火电控制系统应用中, RNN 的模型输入序列是由高密度、连续采样的机组参数值构成的, 这样的时间序列不仅规模庞大, 而且由于时序依赖的影响, 大多数情况下尽管相邻时刻间的参数值存在持续波动, 但相对于参数取值的上下界而言变化幅度并不大, 直接按照数值范围进行特征分类将导致序列内部出现大量重复和信息冗余, 无法有效地反映控制系统状态, 因此难以按照单个时间步进行特征提取和分析. 此外, 火电机组运行状态复杂、控制环节多样、参数间相关性高且耦合性强, 导致对模型结构选择、模型性能评估和建模结果分析更为困难. 这些都是可视分析技术应用于该领域中模型行为理解和性能优化面临的新挑战.

为了应对这些挑战, 本文从以下方面开展研究. 首先, 在损失值等模型外部评价指标的基础上, 从模型内部表现角度引入隐藏单元激活值分布与覆盖率指标, 实现对模型的综合评估; 其次, 通过建立隐藏单元激活值与控制过程的输入/输出参数在时序上的对应关系以及在统计上的变化敏感性, 支持探索模型对控制参数的响应行为; 然后, 通过对连续变化的控制参数进行离散化和符号化处理, 并结合协同聚类分析提出了一种探索强时序依赖的实值时间序列与隐藏单元关联模式的新方法; 最后提出了可视分析系统 iaRNN, 通过组合多个可视化视图支持领域专家对模型行为进行交互式探索. 基于真实数据的应用案例分析和评估表明了该系统具有较好的有效性和可用性.

1 相关工作

1.1 RNN 及其在工业领域的应用

RNN 是一类用于处理序列数据的神经网络, 它引入状态变量存储过去的信息, 与当前输入共同决定当前输出. 由于普通 RNN 在训练过程中容易出现梯度爆炸和梯度消失等问题^[7], 导致其无法适应长跨度依赖的问题, 长短期记忆(long short-term memory, LSTM)神经网络应运而生^[8]. LSTM 通过“门”结构将短期记忆与长期记忆相结合, 以产生梯度长时间持续流动的路径. 但 LSTM 的结构过于复杂, 门控循环单元(gated recurrent unit, GRU)^[9]作为 LSTM 的一种变体, 在简化其结构的同时拥有着与之相似的效果. 本文针对火电控制参数变化特点, 在这些模型的基础上添加了能够将历史状态信息积累的初始化过程, 以提升模型在预测前期的精度.

目前 RNN 不仅在天气预报^[10]、语音识别^[11]和自然语言处理(natural language processing, NLP)等任务^[12]中被证明是有效的, 也逐渐应用于流程工业数据建模中^[13]. 例如, 杨国田等^[14]验证了 LSTM 神经网络模型在火电厂 NO_x 排放预测中相较于传统模型预测效果更好, 稳定性更高. 火电控制系统辨识是一个从输入/输出中确定出一个与所测系统等价的最佳模型的反复迭代过程^[15]. 随着电力行业的发展, 火电厂控制设备的结构和动态特性越来越复杂, 导致生产控制的难度大大增加, 但对控制品质的要求却越来越高. 传统的机理建模方法需要构建复杂的数学模型, 并进行大量的现场试验, 建模成本高, 进行简化假设后又容易导致控制精度不足^[13]. RNN 模型借助强大的非线性映射能力和自学习适应能力, 在解决火电控制系统辨识问题上提供了一种新的思路. 但由于火电控制过程和 RNN 内部本身的复杂性, 应用人员在模型理解和优化上存在很大困难, 模型的调试代价巨大且过度依赖控制人员自身的经验, 限制了模型更进一步的应用. 本文致力于借助可视分析技术建立模型行为与火电控制过程间的语义关联, 帮助用户诊断分析 RNN 的模型表现, 提升系统辨识的效果和迭代速度.

1.2 深度学习的可视分析现状

近年来, 可视分析技术已逐渐成为解释深度学习模型的一种重要手段^[16-17]. 随着深度学习框架的发展, 出现了支持用户分析和调试模型的可视化工具包. 如 TensorFlow 的 TensorBoard^[18]能够

展示模型的网络计算图, 将模型训练期间产生的各项定量指标和参数绘制成图像, 帮助模型开发人员理解和调试模型. Visdom^[19]可以与 PyTorch 等深度学习库配合使用, 在模型训练过程中针对权重、偏置等模型参数和性能指标提供实时绘图. 虽然这些可视化工具能够直观地表示深度学习模型本身直接提供的低层次信息, 但对深度学习模型的解释能力并不能满足一些特定领域的深层次分析需求.

针对时序建模分析需求, 已经有一些专门的可视分析工具来帮助用户更好地理解 RNN. 例如, 在 NLP 领域, RNNVis^[4]用于理解和比较不同的 RNN 模型, 主要将单词与隐藏单元进行协同聚类, 揭示了隐藏单元与单词间的关联模式; LSTMVis^[5]是针对隐藏单元时序变化的动态过程进行可视化的工具, 能够让用户将隐藏状态与词组进行关联并形成假说, 然后自动匹配出具有相似激活特征的其他文本, 进而理解模型. 这些工作都是基于自然语言本身所蕴含的明确语义, 通过比较隐藏单元对不同单词或类别响应的相似度, 建立模型行为与模型输入输出之间的语义信息关联. 在医疗领域中, RetainVis^[20]帮助领域专家使用 RNN 模型分析时间间隔不均匀的电子医学数据, 通过加入时间信息, 显著提升了预测性能. 此外, 文献[6]针对应用于空气污染物预测中以非符号化多维时序数据为输入的 RNN 模型, 设计了可视分析系统 MultiRNNE Explorer, 从模型机制和输入特征在不同时间步上的重要性 2 个方面理解高维时序预测中的 RNN. 虽然这些工作都能够在一定程度上展现出模型内部的工作机理, 但与强时序依赖的火电控制数据不同, 它们的时间序列都是高度离散化的, 分析时默认不同时刻的数据是离散的, 可以直接提取应用语义信息. 而在火电控制系统辨识中相邻时刻的控制参数变化幅度不大, 如果对 RNN 输入序列以单个时间步进行离散分析和特征提取, 将产生大量的连续重复状态, 无法观察到有用的变化规律, 也难以建立模型输入特征与隐藏单元在响应上的关联, 因此并不能直接应用上述方法进行分析. 本文以多个时间步为单元对模型输入数据进行粗粒度提取, 并使用符号化特征近似代替原始时间序列, 以揭示隐藏单元与不同控制参数间的潜在关联.

此外, 火电控制系统辨识的最终目的是更好地支持调控运行参数以优化系统的运行, 所以还需要探索不同控制参数对被预测的受控参数的影响强弱, 因此, 本文结合 RNN 模型内部隐藏单元

响应特点, 提出了一种量化分析控制参数敏感性的方法.

2 数据描述及模型

2.1 数据描述

本文数据来源于某火力发电厂 DCS 数据库, 记录了机组在稳定工况下某个运行时段中不同控制环节的各项控制参数变化. 在机组运行过程中需要对整个发电流程各个环节中的设备进行监控和调优, 并以秒为单位记录产生的所有参数, 利用这些历史数据建模和预测可以很大程度上提升控制系统运行的质量.

火电参数在实际的控制过程中变化复杂, 采集到的原始数据集规模大、时间跨度长、密度高. 在模型训练前, 需根据所分析的控制环节确定作为预测输出的受控参数, 再选择与受控参数相关性最强的 n 个控制参数作为输入. 然后通过预处理校正或删除原始数据中存在的一些坏点或异常值, 以消除噪声对模型的影响. 另外, 由于这些控制参数间的度量单位不尽相同, 取值范围存在较大差别. 为了使模型结构得到更好的训练效果, 在输入模型前还需要对所有控制参数进行 min-max 处理, 统一归一化到 $[0,1]$. 这样, RNN 的输入变量 $\mathbf{X}$ 是由 n 个控制参数组成的多维时间序列, 模型的预测输出 $\mathbf{Y}$ 为受控的单参数时间序列, 即

$$\mathbf{X} = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_t\}, \mathbf{x}_i \in \mathbb{R}^n;$$

$$\mathbf{Y} = \{y_0, y_1, \dots, y_t\}, y_i \in \mathbb{R}.$$

2.2 模型结构

火电控制过程中各个参数的变化往往具有明显的惯性特点, 使传统 RNN 模型在前期的时间步因为历史信息积累不足而导致预测值与真实值的误差偏大. 为了解决这个问题, 本文结合火电控制系统机理, 采用了如图 1 所示的模型结构^[21].

在实际预测前, 先利用一段时间步长的控制参数和受控参数进行初始状态估计, 可以向模型

传递控制系统中各个参数内部和相互间关系的知识, 更快速地确定火电控制系统当前的运行状态. 模型以此为基础再进行预测, 能够有效地提升预测的精度. 预测过程分为以下 2 个步骤.

Step1. 模型预测前首先进行历史控制信息的积累, 即状态初始化过程为

$$\mathbf{h}_s^{\text{init}} = \sigma(\mathbf{W}\mathbf{h}_{s-1}^{\text{init}} + \mathbf{V}_1\mathbf{x}_s + \mathbf{V}_2\mathbf{z}_s + \mathbf{b}) \quad (1)$$

其中,

$$\mathbf{z}_s = [y_{s+1}, y_{s+2}, \dots, y_{s+m}] \quad (2)$$

Step2. 当模型记忆的历史信息足以让其比较准确地预测输出时, 进入输出参数的趋势预测过程为

$$\mathbf{h}_t^{\text{pred}} = \sigma(\mathbf{W}\mathbf{h}_{t-1}^{\text{pred}} + \mathbf{V}_1\mathbf{x}_{s+t} + \mathbf{b}) \quad (3)$$

式(1)~式(3)中, $\mathbf{W}$, $\mathbf{V}_1$, $\mathbf{V}_2$ 为权重矩阵; $\mathbf{b}$ 为偏置矩阵; σ 为 $\tanh$ 激活函数. 输出的隐藏单元激活值 $\mathbf{h}_t \in (-1,1)$.

3 任务分析及模型解释

3.1 任务分析

通过与火电控制领域人员的反复交流和讨论, 针对复杂的火电控制系统辨识过程提出如下分析目标.

G1. 综合对比不同 RNN 模型在特定控制环节中的性能表现, 确定候选模型.

G2. 理解 RNN 模型应用于系统辨识的工作机理.

G3. 探索模型行为与火电控制参数的关联.

G4. 支持模型参数及训练样本量的评估, 为优化模型提供指导性原则.

基于上述领域分析目标, 提炼了以下可视分析任务.

T1. 多指标评估模型预测精度(G1). 使用单一的损失指标评价模型具有局限性, 多种模型损失值评价指标可以综合对比预测值与真实值之间的误差, 更全面地反映模型的表现性能.

T2. 解释隐藏层状态整体分布规律(G2, G4). 隐藏单元状态蕴含着丰富的模型行为信息, 结合激活函数和隐藏层状态分布可以判断模型内部梯度的大小, 而梯度可间接地反映模型内部权重更新的速度, 即模型的学习速度.

T3. 展示隐藏单元对火电控制参数的时序响应行为(G2, G3). 本文使用的 RNN 是多对多的序列模型, 沿时间步展开分析隐藏单元激活值与控制参数、受控参数的变化规律有助于理解模型的响应行为.

图 1 模型结构

T4. 探索模型对控制参数的敏感性(G2, G3). 火电控制过程中有多个可能的控制参数(模型输入)共同影响最终的受控参数(模型输出), 通过定量化分析模型内部响应与不同控制参数之间的关联程度, 可以帮助用户调整控制参数的选择和权重设置, 以优化目标值预测.

T5. 构建输入序列中多维火电控制参数的特征表示, 并探索与隐藏单元之间的语义关联模式(G2, G4). 隐藏单元中存储着与控制参数特征相关的信息, 分析和解释两者之间的关联程度, 能够帮助用户根据不同的控制环节选择合适的模型参数和数据集.

T6. 基于案例探索不同模型的行为(G2, G4). 不同模型的对比分析可以帮助用户理解和探索不同模型结构之间的区别, 进而根据控制过程特点有针对性地选择合适的模型.

3.2 隐藏单元激活值分布

隐藏单元激活值是模型中间计算结果的直接反馈, 对于理解模型的行为至关重要. 文献[22]在观察非线性激活函数对模型性能的影响中指出, 对称激活函数 $\tanh$ 的激活值分布在 0 左右是最合适的, 因为这种情况下梯度能够很好地在神经网络中流动; 相反, 分布在极值附近的激活值会阻塞梯度的倒流, 阻止下层学习到有用的信息.

此外, 实验中发现 RNN 模型在训练开始时, 在 Xavier^[22]权重初始化的作用下, 隐藏单元的激活值分布随着训练次数的增加逐渐从 0 附近的小范围区间向两边扩散, 模型的性能也会逐渐提升. 在此现象的启发下, 本文提出了一种估计模型性能潜力的新指标, 称为激活值覆盖率 c , 即

$$c = \left(\frac{1}{m} \sum_{i=0}^{m-1} n_i \right) \times 100\% \quad (4)$$

首先将隐藏单元激活值的值域(a, b)等比例分为 m 个子区域. 其中, n_i 表示是否有隐藏单元的激活值落在第 i 子区域中, 即

$$n_i = \begin{cases} 1, & \left(\exists h_t^k, h_t^k \in \left[a + \frac{i \times m}{b-a}, a + \frac{(i+1) \times m}{b-a} \right) \right) \\ 0, & \text{其他} \end{cases}$$

用 $\mathbf{H} = \{\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_{T-1}\}$ 表示不同时间步的激活值序列的集合, 其中, $\mathbf{h}_t \in \mathbf{H}$ 且 $\mathbf{h}_t = \{h_t^0, h_t^1, \dots, h_t^{D-1}\}$ 表示时间步为 t 时 D 个隐藏单元的激活值序列, $i = 1, 2, \dots, m-1$.

隐藏单元激活值的丰富组合能够在各层中传递多样性数据, 有利于发挥出模型的拟合性能. 激

活值覆盖率越低, 代表模型性能提升的潜力越大. 例如, 在训练结束并满足所需的精度时, 如果模型的激活值覆盖率未能达到 100%, 表明模型的性能未能完全发挥便完成了拟合任务, 此时可考虑缩小模型的规模, 避免大材小用. 与评价隐藏层内部表现的其他指标对比, 如判别指标和密度指标^[23], 激活值覆盖率取值范围为 0~100%, 能够简单、清晰地表示激活值分布的广度, 且具有较好的通用性, 也能够应用于其他深度学习模型中, 但要求激活值的取值是在有限区间内.

3.3 控制参数敏感性分析

在实际火电控制过程中, 不同控制参数对受控参数的影响是不同的. 评估模型对不同控制参数的敏感度有助于验证和发现控制参数在实际控制系统中的重要性程度, 进而发掘出控制系统潜在的工作机理. 本文利用 Spearman 相关系数^[24]度量隐藏单元与控制参数的相关性.

首先求出每组隐藏单元激活值序列 $\mathbf{h}^k$ 与控制参数序列 $\mathbf{x}^w$ 的相关性大小 $\rho_{k,w}$, 记录每个隐藏单元的最大相关系数 $\rho_k = \max(\rho_{k,0}, \rho_{k,1}, \dots, \rho_{k,m-1})$ 及其对应的控制参数 w , 再分别统计各个控制参数出现的次数, 统计结果代表了不同控制参数的强相关性隐藏单元的数量. 以隐藏单元与控制参数之间的相关性系数代表隐藏单元对控制参数的敏感度, 这样, 对比不同控制参数的强相关性隐藏单元的数量, 可以近似估计模型对不同控制参数的敏感性强弱.

3.4 控制参数与隐藏单元关联分析

3.4.1 控制参数特征提取

火电控制参数的时间序列包含着很多有价值的信息, 如控制参数的数值大小及其时序变化在一定程度上能够反映机组的运行状态, 但难以直观地分析原始的高密度实值时间序列, 这为探索控制参数与 RNN 模型隐藏单元之间的关联带来了挑战. 为了提取控制参数序列的时序特征, 本文根据火电控制过程的特点, 借助符号累计近似技术(symbolic aggregate approximation, SAX)^[25]对原始数据进行粗粒度化处理, 即以多个时刻为单元提取各个控制参数的均值, 再根据参数值的变化范围将均值映射为不同的类别符号, 原始序列被转换为符号序列, 从而保留表达序列特征的关键信息, 而忽略过于微观的控制参数变化细节. 利用类别符号便于直观地分析原始序列中隐含的特征. 原始的 n 维控制参数经过符号化后用 n 个字母组成

的字符串表示, 每个字母对应着一种控制参数。

本文通过多组实验进行比较, 并结合火电领域的参数值划分标准及控制数据的分布特性, 发现将控制参数序列划分成 4 类能够比较准确地对其进行分类。按照控制参数值域从小到大划分, 依次用 4 个字母“A, B, C 和 D”的组合替代原始的一段控制参数序列。图 2 所示为二维控制参数序列的符号化过程, 在 0~5s 时间段的序列用字符串“DC”表示, 5~10s 时间段的序列用字符串“BA”表示。

图 2 二维控制参数符号化

3.4.2 聚类分析

火电控制参数的时间序列规模庞大且时序特征复杂, 多维参数符号化后对应的字符串有上百种, RNN 模型每个隐藏层通常包含着成百上千个隐藏单元, 因此, 难以直接建立符号化控制参数特征与隐藏单元之间的关联模式。由于每个隐藏单元对所有的符号化控制参数特征均有一定的响应, 而每个符号化控制参数特征也同时对应着所有隐藏单元的响应值, 因此本文利用谱协同聚类算法 (spectral co-clustering, SCoC)^[26]同时对隐藏单元和符号化控制参数特征进行聚类, 并使用轮廓系数评估聚类效果, 采用二分图结构建立了符号化控制参数特征簇与隐藏单元簇之间多对多的关联模式。

上述每种字符串对应的各个隐藏单元激活值构成一个向量 $(\delta_f^0, \delta_f^1, \dots, \delta_f^{D-1})$, 表示模型对该字符串的响应。其中, δ_f^i 表示第 i 个隐藏单元对字符串

串 f 的响应激活值。为了直观地展示模型对不同符号化控制参数的响应特点, 本文进一步使用 t -SNE 降维方法将这些高维向量投影到二维空间, 借助散点图展示这些对应控制参数特征的字符串在空间上的分布, 结合聚类关系研究不同簇的差异程度和各个簇的内部相似性, 从而推断模型对不同簇的辨识能力。

4 可视化设计

基于可视分析任务设计了可视化系统 iaRNN, 其主界面如图 3 所示, 由 5 个部分组成。从图 3a 开始, 区域配置面板开始用户可以选择数据集和感兴趣的模型; 之后, 通过图 3b 从外到内评估模型性能; 为探索模型的响应行为, 用户可在图 3c 整体观察控制参数与隐藏单元状态在时序关系上的对应变化, 再通过图 3d 观察模型对不同控制参数敏感度的量化评估结果; 最后, 为了揭示模型隐藏单元与符号化控制参数特征的关联模式, 图 3e 展示了两类聚类后的结果和相关性强弱, 并支持交互式选择符号化控制参数特征或隐藏单元, 将其同步到细节视图进一步探索局部特征。

iaRNN 系统分为前端和后端 2 个部分。前端可视化与交互设计使用 ECharts.js, D3.js 等工具实现; 后端使用 TensorFlow 框架搭建神经网络模型, 使用 Python 语言进行数据处理和分析。前端和后端的数据传输通过微型 Web 框架 Flask 实现。

4.1 模型评估组合视图

在图 3b 中, 区域①的多精度雷达图映射了 5 个模型评价指标, 包括均方误差 (mean square error, MSE)、确定系数 (R-squared)、平均绝对误差 (mean absolute error, MAE)、平均绝对百分比误差 (mean absolute percentage error, MAPE) 和对称平均绝对百分比误差 (symmetric mean absolute percentage error, SMAPE), 用于综合展示 RNN 模型预测结果与真实值之间的偏差 (T1), 此外, 为支持从内部评估模型性能, 设计隐藏单元激活值可视化视图, 用于展示不同训练次数 (epochs) 下指定隐藏层的激活值总体变化规律 (T2); 区域②利用柱形图比较不同训练次数下的隐藏单元激活值覆盖率, 从左到右每个柱形对应着下方数轴上标记的各个训练次数节点, 即每个柱形高度映射相应训练次数完成时的激活值覆盖率大小; 区域③展示了选中训练次数下的隐藏层激活值概率分布图, x 轴表

图 3 iaRNN 主界面

示激活值取值范围 $[-1,1]$, y 轴表示对应的概率密度;区域④利用盒须图展示每个隐藏单元的激活值分布, x 轴为隐藏单元标号, 每个盒子代表一个隐藏单元激活值的分布状态。

4.2 响应行为组合视图

4.2.1 时序关系视图

如图 3c 所示, 区域①为模型输入对应的火电控制参数序列折线图, 将不同控制参数序列用不同颜色标记, 支持交互式选择折线图中需展示的控制参数; 区域②为模型各个中间层的隐藏单元激活值序列视图; 区域③为模型输出序列折线图(T3). 为支持上述大规模、长时间跨度序列的有效展示, 借助“概览+细节”思想设计了 2 种交互方法. 一种方法是滑动窗口技术, 图 3c 中区域⑤是带有刷选功能的条带, 交互式选取时间轴上的展示范围, 此方法利于观察序列的整体趋势和细节特点; 另一种方法是时间盒技术, 通过区域④的框选工具在折线图区域选定 x 轴与 y 轴的展示范围, 更加方便用户观察数据细节. 无论哪种交互方式都会引起整个视图的更新, 控制参数、隐藏单元激活值和受控参数序列始终在时序上同步展示。

4.2.2 参数敏感性视图

按照第 3.3 节中计算方法得到的模型, 通过图

3d 展示控制参数的敏感性量化结果(T4). 百分比面积柱状图展示了每个隐藏单元与不同控制参数相关性的大小, 横坐标为隐藏单元标号, 纵坐标为各个控制参数与当前隐藏单元相关性大小占比. 圆环图的外环为各控制参数对应的强相关性隐藏单元个数所占比例, 内环展示了百分比面积柱状图中每个控制参数对应的柱形面积总和. 这 2 类视图中的相同控制参数映射为同一颜色。

4.3 聚类视图

为进一步探索火电控制参数与隐藏单元之间的关联关系, 本文对符号化控制参数与隐藏单元激活值的聚类结果进行了可视化(T5). 如图 3e 所示, 区域①将多维符号化控制参数簇映射为图符集合, 区域③将隐藏单元簇映射为热力图, 两者间的相关性通过连线表示, 如区域②所示。

4.3.1 符号化控制参数特征图符

为了直观地展示多维控制参数序列的符号化结果, 设计了控制参数特征图符用于可视化对应的字符串. 如图 4 右侧所示, 该图符由同一圆心的多个圆环组成, 不同颜色的扇形区域代表不同的控制参数; 环的个数代表不同字母, 如 1 个环代表字母 A, 2 个环代表字母 B 等; 以圆环上半部分垂直过圆心的直线为轴, 轴右侧的第 1 个扇形为起始

控制参数, 顺时针依次描述不同控制参数, 图 4 右侧图符代表控制参数字符串“ABCD”, 左侧为一个控制参数簇. 为便于总结簇内的整体分布, 同一簇中符号化控制参数特征通过层次化排序依次进行展示.

图 4 符号化控制参数特征图符

4.3.2 聚类关联视图

在图 3e 的区域③中, 每个矩形框中由多个色块组成的热力图对应一个隐藏单元簇. 每个色块对应一个隐藏单元, 以颜色为视觉通道映射隐藏单元激活值的均值, 红色为正值, 蓝色为负值, 绝对值越大, 颜色越深. 符号化控制参数特征簇 F_i 与隐藏单元簇 H_j 之间连线的宽度 W 代表了它们之间聚合相关性. ∂_f^h 表示隐藏单元 h 对符号化控制参数特征 f 的响应值, 颜色编码的方式与热力图相同, 正值表示正相关, 负值表示负相关. W 的计算公式为

$$W(F_i, H_j) = \frac{\sum_{f \in F_i, h \in H_j} \partial_f^h}{\sum_{k=1}^n \left| \sum_{f \in F_i, h \in H_k} \partial_f^h \right|}.$$

除了能够概览控制参数簇与隐藏单元簇的联系, 还可以通过交互手段探索细节信息. 在图 3e 中, 单击区域③内隐藏单元簇中的任一隐藏单元, 可在区域④观察选中隐藏单元对同簇符号化控制参数特征的具体响应; 此外, 用户也可以在区域①选择若干个符号化控制参数特征, 通过区域⑤展示每个隐藏单元对它们响应的激活值, y 轴为隐藏单元响应值, x 轴为根据响应值大小排序后的隐藏单元标号.

为了进一步分析模型对不同簇中符号化控制参数特征的响应特点, 图 3e 的区域⑥用散点图展示了所有符号化控制参数特征对应的隐藏单元激活值向量经过 t -SNE 降维后的结果, 用不同颜色编码不同簇, 右侧标记簇与颜色的编码关系, 同时展示对应簇的轮廓系数.

5 案例分析与用户评估

本文采用火电厂的实际运行数据设计了应用

案例. 所用数据来源于电厂锅炉主汽温控制环节在 2019 年 1 月的实际运行记录, 并在其中选取 10 天正常工作的数据作为训练样本, 共 16000 组; 另外选择一天的数据共 1600 组作为测试样本. 模型的输入包括减温水流量、主蒸汽流量、燃料量和主蒸汽压力 4 种控制参数, 模型的预测输出为受控参数主汽温.

表 1 列出了案例中使用的不同模型的超参数及其预测误差. 邀请火电控制领域的专家使用 iaRNN 进行案例的探索和分析, 然后在此基础上进行了系统的评估.

表 1 模型配置和性能

模型	状态 初始化	隐藏层 大小	隐藏层 数量	均方误差
RNN	是	100	2	15.07±2.62
	否			19.81±0.53
GRU ^[9]	是	100	2	2.13±0.24
	否			6.29±0.18
LSTM ^[8]	是	100	2	1.99±0.21
	否			5.14±0.10
RNN	是	100	3	109.78±20.14
		500	2	471.52±131.27
GRU ^[9]	是	500	2	0.50±0.03

5.1 模型综合评估

首先从图 3b 的多精度雷达图中观察模型在各个评价指标上的表现, 对比 2 层 100 个隐藏单元的 GRU 模型与普通 RNN 模型后发现, 前者的各项指标值都远优于后者.

进一步观察隐藏单元分布, 如表 2 所示. GRU 在训练过程中隐藏单元激活值始终在 0 两侧紧密分布, 这有利于梯度在训练过程中传播, 且适当扩大模型规模有助于提升模型性能. 但是普通 RNN 在训练过程中激活值覆盖率很快就达到了 100%, 隐藏层激活值概率分布也逐渐呈现饱和状态. 根据上述这些现象可以诊断出模型发生了梯度消失^[22]. 通过多组实验对比发现, GRU 与 LSTM 的模型结构比 RNN 更加适用于针对主汽温控制环节这类具有明显延迟特征的系统辨识任务. 观察模型训练过程中的上述变化规律, 可以帮助用户根据控制系统中不同辨识对象的特点确定合适的候选模型.

5.2 响应模式分析

5.2.1 时序关系

根据图 3c 所展示的控制参数、受控参数与隐藏单元激活值间的时序演变规律, 可以观察发现

表 2 GRU 与 RNN 高隐藏层的激活值变化对比

模型	激活值覆盖率/%	激活值概率分布图		
		训练 10 次	训练 100 次	训练 500 次
GRU ^[9]
RNN

多维控制数据在时序上的波动幅度与隐藏单元的变化幅度相似,模型中的隐藏单元在整体上能对不同变化趋势的序列做出不同程度的类似响应。当控制参数序列从剧烈波动趋于平坦时,第1层的激活值的波动也很快变得平缓,但第2层的激活值却需要经过较长的缓冲时间使其变化幅度降低。基于上述现象并结合领域知识,发现高层隐藏单元已经学习到了主蒸汽温度控制环节具有较长惯性延迟的特点。

5.2.2 参数敏感性验证

接下来通过查看敏感性分析视图,定量估计模型对不同控制参数的敏感性,如图 3d 所示。在环形图中相比其他 3 种控制参数,减温水流量所占的面积最大,说明模型对减温水流量的敏感性更强。对此,领域专家指出,“在机组的实际运行中,减温水确实是调节主蒸汽温度的最直接手段,比其他 3 种控制参数的影响都要快”。这表明 RNN 学习到了这种机理特性,能够区分受控参数对不同控制参数的敏感性。此外,主汽温对其他 3 种控制参数的敏感性强弱在火电控制领域并没有明确的排序,敏感性分析视图展示的结果为此提供了参考。

5.3 关联模式分析

聚类视图清晰地呈现了符号化控制参数特征簇与隐藏单元簇之间的关联关系,如图 3e 所示。观察发现同一簇内大多数的隐藏单元对符号化控制参数特征的响应是相似的,例如,图中与左侧第 2 个符号化控制参数特征簇高相关的第 1 个隐藏单元簇的响应值整体上为负值,第 2 个隐藏单元簇的响应值都为正值。

领域专家还选择不同的符号化控制参数特征进行分析,隐藏单元对同一簇中的符号化控制参数特征(如“AAAA”和“AAAB”)的响应分布高度相似,但对不同簇的符号化控制参数特征(如“AAAD”与“AAAA”)的响应分布则有较大差异,

如图 5 所示。上述现象体现出符号化方法有效地表达了控制参数的特征,协同聚类方法也较好地构建了符号化控制参数特征与隐藏单元之间的关联模式。

图 5 隐藏单元对符号化控制参数特征的响应

此外,领域专家使用聚类视图也发现了一些异常之处。如图 3e 所示,与其他聚类簇相比,第 2 个符号化控制参数特征簇所对应的隐藏单元簇中包含更多隐藏单元,且通过区域⑥可以看出,第 2 个符号化控制参数特征簇的投影点较为分散,体现出模型对此簇的辨识能力不足。上述现象表明,模型在辨别这类控制参数时分配了大量隐藏单元且得不偿失。在检查了训练用的数据集后发现,第 2 簇中的符号化控制参数特征对应的样本量普遍偏少,对此,领域专家猜测模型预测性能的不均衡是因训练集样本不均衡导致的。在补充了更多此类样本后重新训练,模型预测精度确实得到了一定程度的提升,进一步验证了此前的猜想。

5.4 用户评估与反馈

在 iaRNN 设计与开发过程中, 始终与领域专家保持紧密合作, 经过反复迭代历时一年多最终形成目前的原型系统. 在这个过程中, 收集了领域专家的反馈和建议. 总体上, 领域专家认为 iaRNN 是一种新颖、有效的可视分析工具, 能够支持多方面评估不同模型性能, 辅助用户基于不同控制环节选择适合的模型; 通过揭示模型对不同控制参数响应的特点, 领域内的机理规律得到验证和拓展; 从多个角度探索控制参数与模型的关联模式, 可以有效地理解和解释模型行为.

领域专家也对系统的进一步优化提出以下建议.

(1) 模型可解释性方面还可以继续扩展, 如增加对模型结构和门控信息等的可视化展示.

(2) 针对不同控制环节的时序特征提取还需进一步细化研究, 如符号化时子序列长度和特征量的选择.

(3) 系统可扩展性还存在不足, 为了适应分析更大规模的神经网络模型, iaRNN 在可视化设计方面还需进一步优化.

6 结 语

本文以面向火电控制系统辨识的 RNN 为研究对象, 从理解模型的运行机制、诊断和优化模型等角度设计并实现了可视分析系统 iaRNN. 首先, 通过综合预测精度和隐藏单元状态信息对模型性能进行多方面评估, 帮助用户选择模型结构及规模; 然后, 从模型对控制参数的响应行为、隐藏单元与输入特征的关联模式等方面探究模型行为与实际控制过程的语义关联; 最后, 通过真实数据进行案例设计与分析, 验证了 iaRNN 系统的可用性和有效性. 虽然本文是以火电控制系统辨识为例, 但从设计上, iaRNN 可推广到其他类似的系统辨识与预测过程中, 能为工业控制领域的 RNN 建模提供有力的支持.

参考文献(References):

- [1] Duan Yanjie, Lv Yisheng, Zhang Jie, *et al.* Deep learning for control: the state of the art and prospects[J]. *Acta Automatica Sinica*, 2016, 42(5): 643-654(in Chinese)
(段艳杰, 吕宜生, 张杰, 等. 深度学习在控制领域的研究现状与展望[J]. *自动化学报*, 2016, 42(5): 643-654)
- [2] Liu Jizhen, Wang Qinghua, Fang Fang, *et al.* Data-driven-based application architecture and technologies of smart power generation[J]. *Proceedings of the CSEE*, 2019, 39(12): 3578-3587(in Chinese)
(刘吉臻, 王庆华, 房方, 等. 数据驱动下的智能发电系统应用架构及关键技术[J]. *中国机电工程学报*, 2019, 39(12): 3578-3587)
- [3] Ji Lianen, Kong Yumeng, Wang Yanlin, *et al.* Visual analytics of time series data of thermal power control for system identification[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2019, 31(10): 1677-1686(in Chinese)
(纪连恩, 孔雨萌, 王炎林, 等. 面向系统辨识的火电控制时序数据可视分析[J]. *计算机辅助设计与图形学学报*, 2019, 31(10): 1677-1686)
- [4] Ming Y, Cao S Z, Zhang R X, *et al.* Understanding hidden memories of recurrent neural networks[C] // *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Los Alamitos: IEEE Computer Society Press, 2017: 13-24
- [5] Strobel H, Gehrmann S, Pfister H, *et al.* LSTMVis: a tool for visual analysis of hidden state dynamics in recurrent neural networks[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2018, 24(1): 667-676
- [6] Shen Q M, Wu Y H, Jiang Y Z, *et al.* Visual interpretation of recurrent neural network on multi-dimensional time-series forecast[C] // *Proceedings of the IEEE Pacific Visualization Symposium*. Los Alamitos: IEEE Computer Society Press, 2020: 61-70
- [7] Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks[C] // *Proceedings of the 30th International Conference on Machine Learning*. Lille: PMLR Press, 2013: 1310-1318
- [8] Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780
- [9] Cho K, van Merriënboer B, Gulcehre C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation[C] // *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: ACL Press, 2014: 1724-1734
- [10] Shi X J, Chen Z R, Wang H, *et al.* Convolutional LSTM network: a machine learning approach for precipitation nowcasting[C] // *Proceedings of the 28th Neural Information Processing Systems*. Cambridge: MIT Press, 2015: 802-810
- [11] Mikolov T, Karafiát M, Burget L, *et al.* Recurrent neural network based language model[C] // *Proceedings of Conference of the International Speech Communication Association*. New York: Curran Associates, 2010: 1045-1048
- [12] Liu S J, Nan Y, Mu L, *et al.* A recursive recurrent neural network for statistical machine translation[C] // *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: ACL Press, 2014: 1491-1500
- [13] Yuan Xiaofeng, Wang Yalin, Yang Chunhua, *et al.* The application of deep learning in data-driven modeling of process industries[J]. *Chinese Journal of Intelligent Science and Technology*, 2020, 2(2): 107-115(in Chinese)
(袁小锋, 王雅琳, 阳春华, 等. 深度学习在流程工业过程数据建模中的应用[J]. *智能科学与技术学报*, 2020, 2(2): 107-115)
- [14] Yang Guotian, Zhang Tao, Wang Yingnan, *et al.* Prediction

- model for NO_x emissions from thermal power plants based on long-short-term memory neural network[J]. *Thermal Power Generation*, 2018, 47(10): 12-17(in Chinese)
(杨国田, 张涛, 王英男, 等. 基于长短期记忆神经网络的火电厂 NO_x 排放预测模型[J]. *热力发电*, 2018, 47(10): 12-17)
- [15] Liu Chen, Han Pu, Sun Ming. Research on system identification methods for thermal power plant process[J]. *Computer Simulation*, 2016, 33(4): 283-286(in Chinese)
(刘琛, 韩璞, 孙明. 火电厂热工过程系统辨识方法的研究[J]. *计算机仿真*, 2016, 33(4): 283-286)
- [16] Zahavy T, Ben-Zrihem N, Mannor S. Graying the black box: understanding DQNs[C] // *Proceedings of the 33rd International Conference on Machine Learning*. Lille: PMLR Press, 2016: 1899-1908
- [17] Liu M C, Shi J X, Cao K L, *et al.* Analyzing the training processes of deep generative models[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2017, 24(1): 77-87
- [18] Wongsuphasawat K, Smilkov D, Wexler J, *et al.* Visualizing dataflow graphs of deep learning models in TensorFlow[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2017, 24(1): 1-12
- [19] Urbanek J, Maaten L, Malmaud J, *et al.* Facebook research[OL]. [2021-07-08]. <https://github.com/facebookresearch/visdom>
- [20] Kwon B C, Choi M J, Kim J T, *et al.* RetainVis: visual analytics with interpretable and interactive recurrent neural networks on electronic medical records[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2018, 25(1): 299-309
- [21] Savran A. Discrete state space modeling and control of nonlinear unknown systems[J]. *ISA Transactions*, 2013, 52(6): 795-806
- [22] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[C] // *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*. Lille: PMLR Press, 2010: 249-256
- [23] Zhong W, Xie C, Zhong Y, *et al.* Evolutionary visual analysis of deep neural networks[OL]. [2021-07-08]. <http://icmlviz.github.io/assets/papers/2.pdf>
- [24] Croux C, Dehon C. Influence functions of the Spearman and Kendall correlation measures[J]. *Statistical Methods & Applications*, 2010, 19(4): 497-515
- [25] Lin J, Keogh E, Wei L, *et al.* Experiencing SAX: a novel symbolic representation of time series[J]. *Data Mining and Knowledge Discovery*, 2007, 15(2): 107-144
- [26] Dhillon I S. Co-clustering documents and words using bipartite spectral graph partitioning[C] // *Proceedings of the 7th ACM Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2001: 269-274