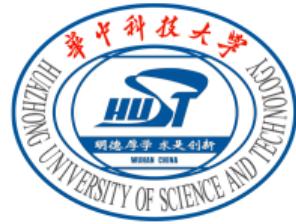


基于随机映射的时间序列深度学习预测建模技术及其应用研究

张心泽

School of Management,
Huazhong University of Science and Technology

April 26, 2023



基于随机映射的时间序列深度学习预测建模技术及其应用研究

预测建模问题是气象水文、公共卫生、电力系统等众多领域进行管理决策时的重要问题。

例如电力系统：

运行决策：

- 电力生产与配送计划
- 燃料购买与存用计划
- 电力价格定价策略

预测需求：

- 区域电力负荷
- 火电燃煤价格
- 风能光伏输出力

影响因素：

- 经济的发展
- 天气的变化
- 疫情的波动

受很多未知因素的存在影响，难以建立完备的因果关系来给出现象发生的定律，不可能建立出一个确定性的模型来精确计算现象的未来表现

时间序列预测定义

尽管有诸多因素干扰，依然有可能基于对现实现象的历史观测推导出一个模型，用来计算一定提前期的未来值。时间序列预测正是通过对时间序列观测值之间相互依赖性的分析，发展出动态模型，进而对时间序列的未来状态进行预测。

时间序列预测的过程本质上是一个以历史时间序列为自变量，以未来时间序列为因变量的函数逼近过程。

即，给定一个历史 T 步的时间序列 $\mathbf{x} = (x_1, \dots, x_T)$, $\mathbf{x} \in \mathbb{R}^T$, 以及未来提前 H 步的时间序列 $\mathbf{y} = (x_{T+1}, \dots, x_{T+H})$, $\mathbf{y} \in \mathbb{R}^H$, 时间序列预测问题可被公式化定义为:

$$\mathcal{F} : f(\mathbf{x}) + \mathbf{e} = \mathbf{y},$$

\mathcal{F} 是指时间序列预测模型所逼近的函数。 f 是指时间序列预测模型所学习出的预测函数。

传统时间序列预测建模技术

以统计学为基础的回归预测模型，通过构造历史观测值与相关因素的经验方程，对时间序列进行拟合和预测。

- 自回归（Auto-regressive, AR）模型
- 移动平均（Moving average, MA）模型
- 自回归移动平均混合（Auto-regressive moving average, ARMA）模型
- 整合移动平均自回归（Auto-regressive integrated moving average, ARIMA）模型
- 指数平滑（Exponential smoothing, ES）模型

归纳：

- 将时间序列数据的输入输出映射视为预定义的函数
- 仅适用于某一特定类型的平稳过程或非平稳过程
- 不能有效处理复杂非线性非平稳的时间序列预测问题

将预测问题考虑为一种以历史 T 步的时间序列 $\mathbf{x} = (x_1, \dots, x_T)$ 为输入，以未来提前 H 步的时间序列 $\mathbf{y} = (x_{T+1}, \dots, x_{T+H})$ 为输出的回归问题，构造监督机制下的回归模型予以求解。

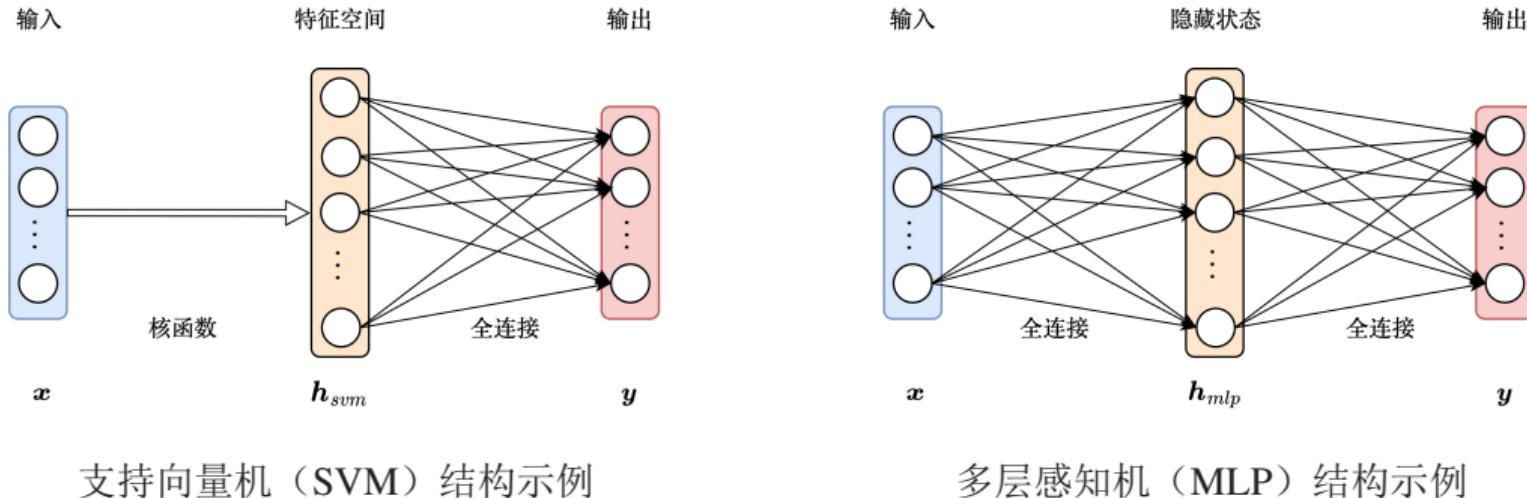
$$f(\mathbf{x}) + \mathbf{e} = \mathbf{y}, \quad f \in \{\vartheta, \theta\}.$$

θ 表示机器学习预测模型的权重参数集合， ϑ 表示定义机器学习预测模型与权重学习机制的超参数集合。

- ϑ 定义了机器学习预测模型的结构与权重参数的训练机制
- θ 表示机器学习模型中从输入变换到输出间的权重参数集合
- $f \in \{\vartheta, \theta\}$

如何选择合适的 ϑ 与 θ 从而建立一个精准的预测模型 f

ML 代表模型



支持向量机（SVM）结构示例

多层感知机（MLP）结构示例

以一组给定包含 N 个样本的时间序列数据集 $\mathbb{D} = \{(\mathbf{x}_n, \mathbf{y}_n) \in (\mathbb{R}^T \times \mathbb{R}^H)\}_{n=1}^N$ 为例，对基于 SVM 与 NN 的预测建模技术与应用展开概述。

梯度下降

常有学者将基于梯度下降方法的神经网络模型称为反向传播神经网络（Back propagation neural network, BPNN）模型。

- 梯度下降：梯度下降训练方法中各层权重参数的优化方法
- 反向传播：梯度下降训练方法中计算各层权重参数梯度时所使用的链式法则梯度传导过程

以 MLP 为例：

$$\ell(\mathbf{w}) = \frac{1}{N \times H} \sum_{n=1}^N \|\mathbf{e}_n\|,$$
$$\mathbf{e}_n = \mathbf{y}_n - f(\mathbf{x}_n).$$

$\ell(\mathbf{w})$ 为神经网络预测模型在训练集 \mathbb{D} 上的 MSE 损失

核心思想：

- 基于神经网络模型中激活函数的连续性与可微性
- 向 \mathbf{w} 添加一个很小的动量 $\Delta_{\mathbf{w}}$, 即 $\|\Delta_{\mathbf{w}}\|$ 很小, 亦等价于 $\mathbf{w} + \Delta_{\mathbf{w}}$ 近似 \mathbf{w}
- 利用泰勒近似将复杂非凸的 $\ell(\mathbf{w})$ 函数优化问题当作一个简单的函数极小值问题

$$\ell(\mathbf{w} + \Delta_{\mathbf{w}}) \approx \ell(\mathbf{w}) + \nabla_{\mathbf{w}}^T \Delta_{\mathbf{w}}.$$

$\nabla_{\mathbf{w}}$ 表示 \mathbf{w} 在误差函数 \mathbf{w} 中的梯度。在最陡下降中, 定义学习速率 (Learning rate) 为 η , 且 $\eta > 0$, 因此:

$$\Delta_{\mathbf{w}} = -\eta \nabla_{\mathbf{w}}.$$

当 η 足够小, 可证:

$$\ell(\mathbf{w} - \eta \nabla_{\mathbf{w}}) \approx \ell(\mathbf{w}) - \eta \nabla_{\mathbf{w}}^T \nabla_{\mathbf{w}} < \ell(\mathbf{w}).$$

Algorithm 神经网络模型梯度下降方法训练过程

- 1: **while** 未满足 ϑ 界定的收敛条件 **do**
- 2: $i \leftarrow i + 1$
- 3: 基于 θ^i , 确定模型 $f \leftarrow f \in \{\vartheta, \theta^i\}$
- 4: 基于数据集 \mathbb{D} 和模型 f , 完成前馈过程 $f(\mathbf{x})$, 计算模型损失函数值 $\ell(\mathbf{w})$
- 5: **for** w in θ^i **do**
- 6: 选取计算梯度所需的样本
- 7: 根据反向传播链式法则计算梯度信息 ∇_w
- 8: 根据梯度下降更新公式更新权重参数 w
- 9: **end for**
- 10: **end while**
- 11: **return** 已完成更新过程的权重参数集合 $\theta \leftarrow \theta^i$

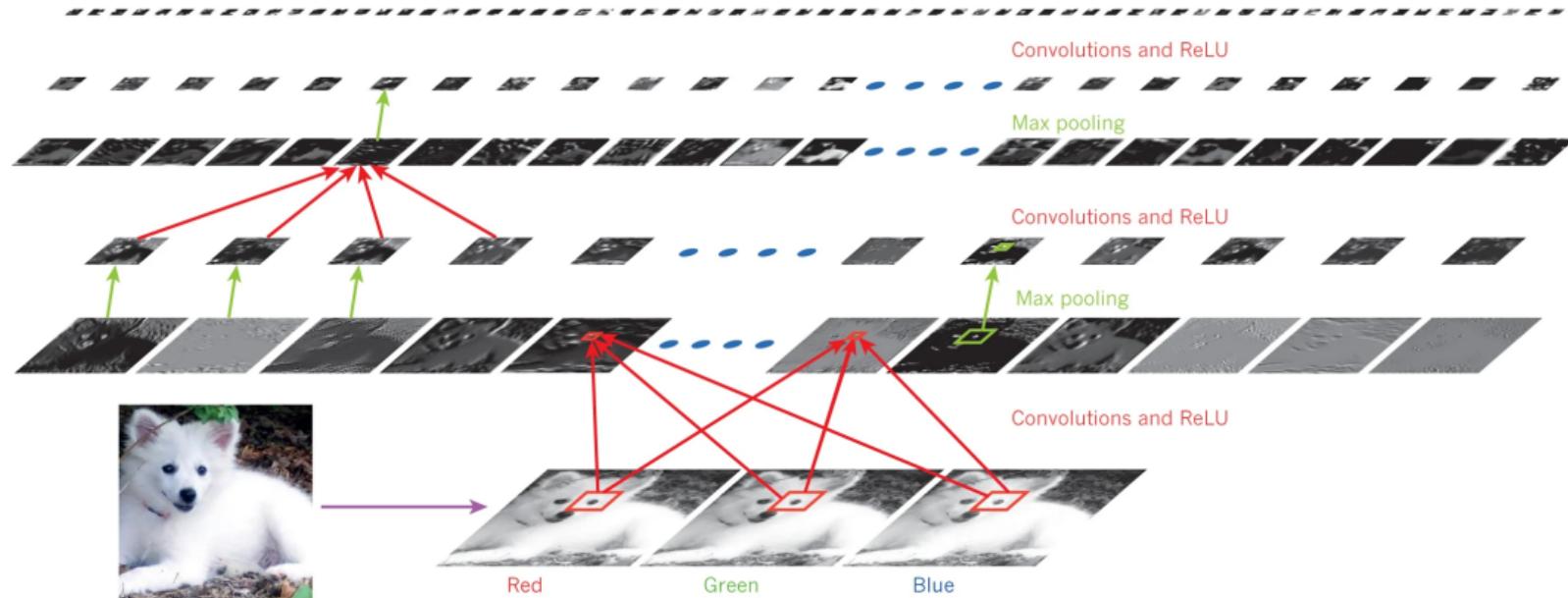
深度学习建模技术是机器学习建模技术中一类基于深度神经网络的建模方法。

- 海量数据的积累: CIFAR10, CIFAR100, ImageNet, WMT
- 计算能力的提升: GPU 加速, 分布式计算, 并行计算
- 梯度下降的改进: SGD, Moment, Adam
- 通用框架的提出: TensorFlow, Pytorch

- 传统机器学习: 人工经验建立特定的特征提取方法
 - 基于图像像素数值高斯分布描述图像特征的 Fisher Vector 方法
 - 基于语言文档内词频信息描述语句特征的词频-逆文档 (TF-IDF) 方法
- 深度学习: 依靠神经网络结构学习数据的抽象特征表示
 - CNN 中的卷积结构
 - RNN 中的循环结构

深度学习示例

Samoyed (16); Papillon (5.7); Pomeranian (2.7); Arctic fox (1.0); Eskimo dog (0.6); white wolf (0.4); Siberian husky (0.4)



卷积神经网络图像分类模型示例（引用于 Lecun et al. “Deep Learning.” Nature, 2015, 521.7553: 436-444）

模型选择问题

对于由超参数 ϑ 和权重参数 θ 所定义的深度学习预测模型 f , 其模型选择问题便是如何选择合适的 ϑ 与 θ 从而提升模型 f 预测性能的问题。

- 权重参数 θ 的优化研究 → 梯度下降方法的改进
- 超参数 ϑ 的优化研究 → 神经网络结构的优化

神经网络结构的分解:

- 表征结构
 - 隐藏结构 → 神经网络结构搜索 (NAS)
 - ◇ 启发式优化、强化学习 → 计算开销大, 选择耗时长
 - 输入结构 → 预测领域的多输入特征选择
 - ◇ 封装法更优 → 多时步维度的输入特征结构
- 输出结构 → 预测领域的多输出策略优化
 - MIMO 策略, Direct 策略, Iterative 策略
 - ◇ MIMO 更优 → MIMO 策略下依然存在多种输出结构

θ 依赖于超参数 ϑ :

- 模型神经网络结构参数（如 CNN 模型中卷积层层数、卷积核宽度和卷积核数量等等）
- 梯度下降方法参数（如 SGD 方法中的学习速率、批次样本选取数量和动量惯性系数等等）

不同的 ϑ 必然会导致 θ 的差异

模型选择过程

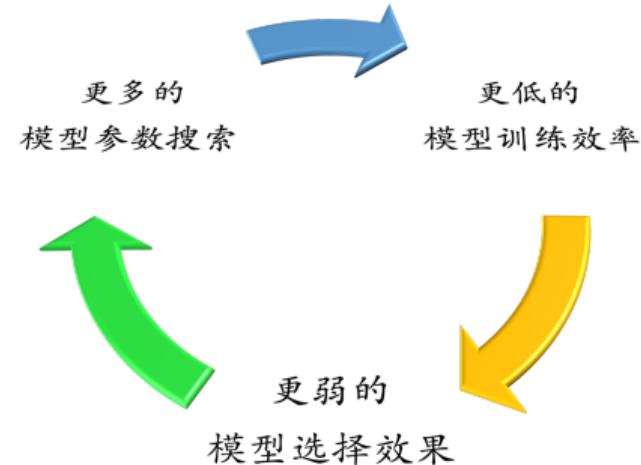
Algorithm 基于梯度下降方法的神经网络预测建模技术模型选择过程

- 1: **while** 未满足模型选择方法界定的收敛条件 **do**
 - 2: $j \leftarrow j + 1$
 - 3: 从超参数集合 ϑ 的搜索空间 Ω 中选择或更新出当前的超参数集合 ϑ^j
 - 4: 基于 ϑ^j :
 - 执行梯度下降算法所示步骤
 - 得到当前 ϑ^j 试验下的权重参数集合 $\theta|\vartheta^j$
 - 确定模型 $f \leftarrow f \in \{\vartheta^j, \theta|\vartheta^j\}$
 - 5: 基于数据集 \mathbb{D} 和模型 f , 计算模型误差
 - 6: 基于模型误差与超参数更新方法:
 - 更新最优的神经网络超参数集合 ϑ^*
 - 更新最优的神经网络超参数集合 $\theta^* \leftarrow \theta|\vartheta^*$
 - 7: **end while**
 - 8: **return** 已完成更新过程的超参数集合 $\vartheta \leftarrow \vartheta^*$ 与权重参数集合 $\theta \leftarrow \theta^*$
-

模型选择挑战

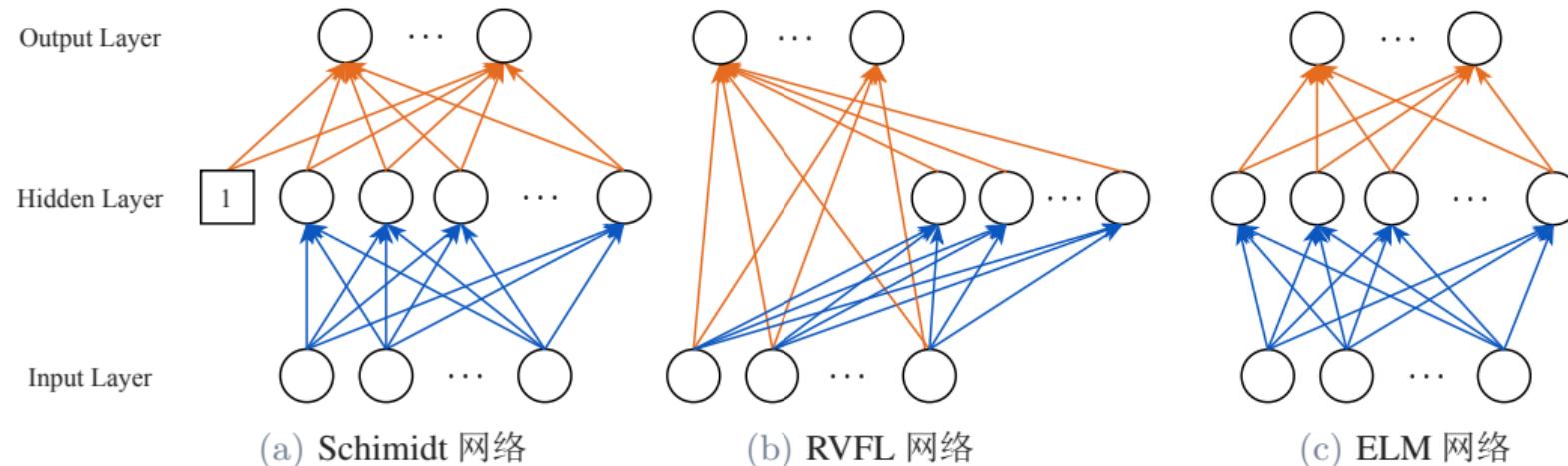
深度学习预测建模技术所面临的模型选择挑战尤为突出。

- 更多的权重参数以及更加复杂的梯度计算方式，导致模型选择效率低
- 复杂的神经网络结构加深了超参数搜索空间的复杂程度，导致模型选择效果弱



随机映射方法研究概况

随机映射是一种采用非迭代学习机制的 MLP 建模方法。



蓝色线表示固定的随机初始化权重，黄色线表示闭式求解得输出权重。

随机映射建模过程

以 MLP 为例：随机映射多层感知机（Stochastic multiple percentage, SMLP）模型的预测过程为：

$$\begin{aligned}f_{smlp}(\mathbf{x}) &= \mathbf{w}_{out}\mathbf{h}_{smlp}, \\ \mathbf{h}_{smlp} &= \sigma(\mathbf{w}_{hid}\mathbf{x}).\end{aligned}$$

其模型损失函数依然保持 MSE 损失形式

$$\begin{aligned}\mathbf{w}_{out} &= \operatorname{argmin}_{\mathbf{w}_{out}} \|\mathbf{y} - \mathbf{w}_{out}\mathbf{h}_{smlp}\| \\ &= \mathbf{y}\mathbf{h}_{smlp}^T(\mathbf{h}_{smlp}\mathbf{h}_{smlp}^T)^{-1}.\end{aligned}$$

支持向量机模型是一种 SMLP 模型的特例

SMLP 的收敛性质

传统的 SMLP 模型避免了梯度下降训练计算开销高的问题，但这种权重随机机制使其模型性能的收敛性与稳定性受到质疑。

针对于此：

- Huang 等¹基于递归添加神经元的方式构造了一种增长极限感知机（Incremental extreme learning machine, IELM），并证明了 IELM 对于任意连续有界目标函数的学习收敛能力。
- Wang 等²通过迭代增加隐藏层神经元，基于监督机制闭式挑选神经元隐藏层权重，全局更新输出层权重的方式，构造出增长的 SMLP 模型，并证明了 SCN 的普适逼近性质（Universal approximation property）。

¹Huang et al. "Universal approximation using incremental constructive feedforward networks with random hidden nodes." IEEE Trans. Neural Networks 17.4 (2006): 879-892.

²Wang et al. "Stochastic configuration networks: Fundamentals and algorithms." IEEE transactions on cybernetics 47.10 (2017): 3466-3479.

随机映射 RNN 的研究:

- Jaeger 等³在 2001 年提出随机映射 RNN，称其为状态回声网络（Echo state network, ESN）。
- 技术：Leaky ESN, Growing ESN, Deep ESN, Regularize ESN
- 应用：电力负荷预测，原油价格预测

随机映射 CNN 的研究:

- He 等⁴发现随机映射 CNN 在纹理生成和风格迁移等任务上具有不亚于梯度下降训练 CNN 的表现。
- 预测研究较少

³Jaeger et al. "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication." science 304.5667 (2004): 78-80.

⁴He et al. "A powerful generative model using random weights for the deep image representation." Advances in Neural Information Processing Systems 29 (2016).

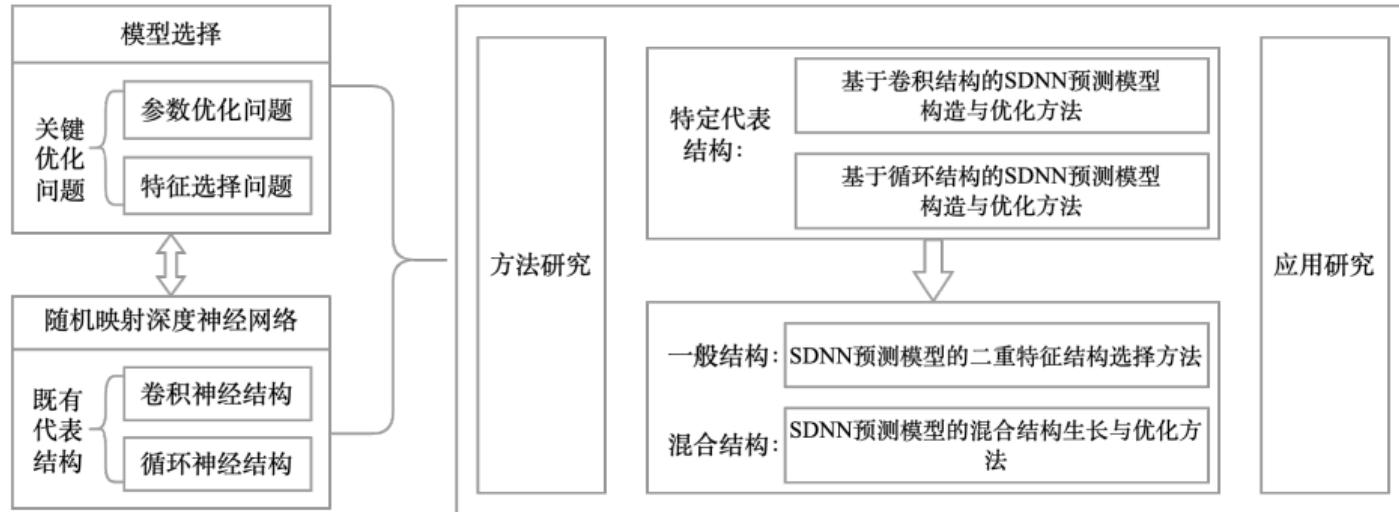
由此，针对时间序列深度学习预测技术模型选择挑战中的低效问题，构造基于随机映射的深度学习预测建模技术是一种有效的解决途径

- 深度学习建模技术优异的预测潜能
- 随机映射建模技术高效的建模效率

随机映射深度学习预测建模技术的挑战：

- 建模效率的提高建立在预测性能的牺牲上
- 更高的模型选择要求

技术研究体系



- 在理论层面，本论文将深入研究不同神经网络结构下的 SDNN 预测模型和新的理论分析框架，如构建基于卷积结构和混合结构的 SDNN 预测模型的预测误差收敛性分析框架，直观理解 SDNN 预测模型的构造与决策逻辑
- 在方法层面，本论文将构造多套预测模型构造与优化方法，包括适用于卷积结构、循环结构、一般结构至混合结构 SDNN 模型的优化技术，形成从特定到一般再到混合的综合技术体系与框架，为 SDNN 预测建模技术的复杂模型选择问题提供新颖的方法与思路
- 在现实层面，本论文将立足于现实时间序列预测任务的多个场景，包括公共卫生领域中的流感阳性样本率预测任务、能源市场中的原油价格预测任务、金融市场中的股票指数预测任务、大气污染中的 PM2.5 预测任务、电力系统中的电力负荷与电力价格预测任务等

Conv-SDNN 预测模型——引言

基于卷积结构的 SDNN 预测模型：随机映射 CNN 模型

- 随机映射 CNN 在图像生成任务（纹理生成、风格迁移）上具有不亚于梯度下降训练 CNN 的表现

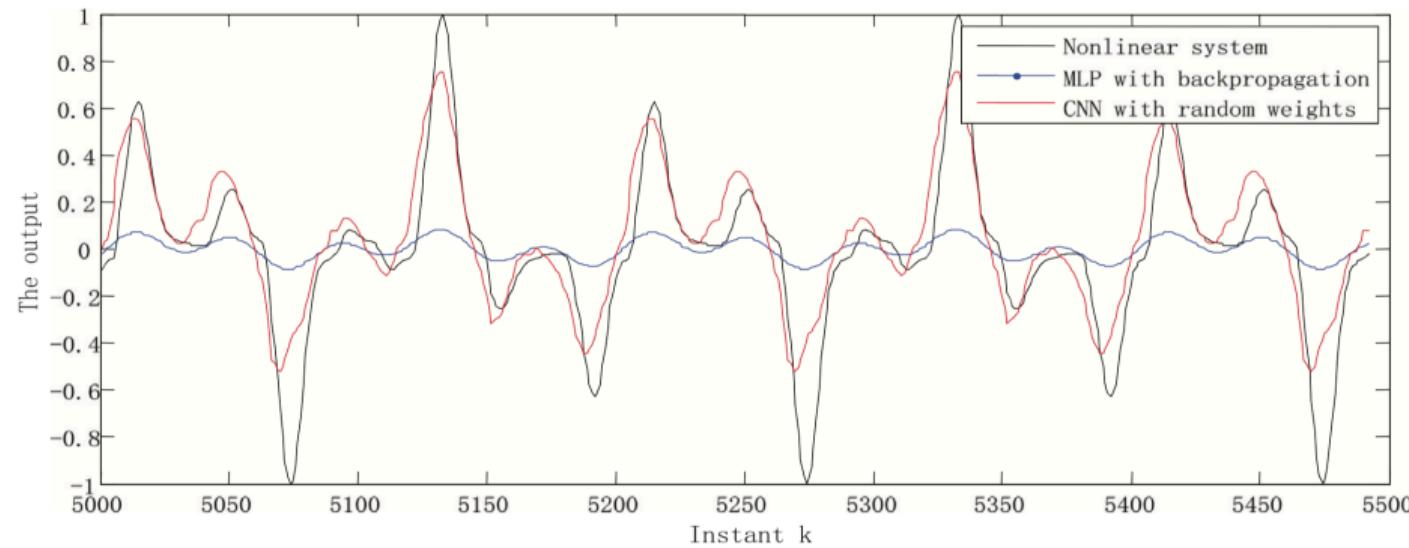


Conv-SDNN 预测模型——引言

随机映射 CNN 模型研究：

- 一维随机映射 CNN 在一些人工时间序列预测数据集上展现出比 MLP 模型更好的拟合性能（Yu et al. 2019）

W. Yu, M. Pacheco / Information Sciences 477 (2019) 1–14



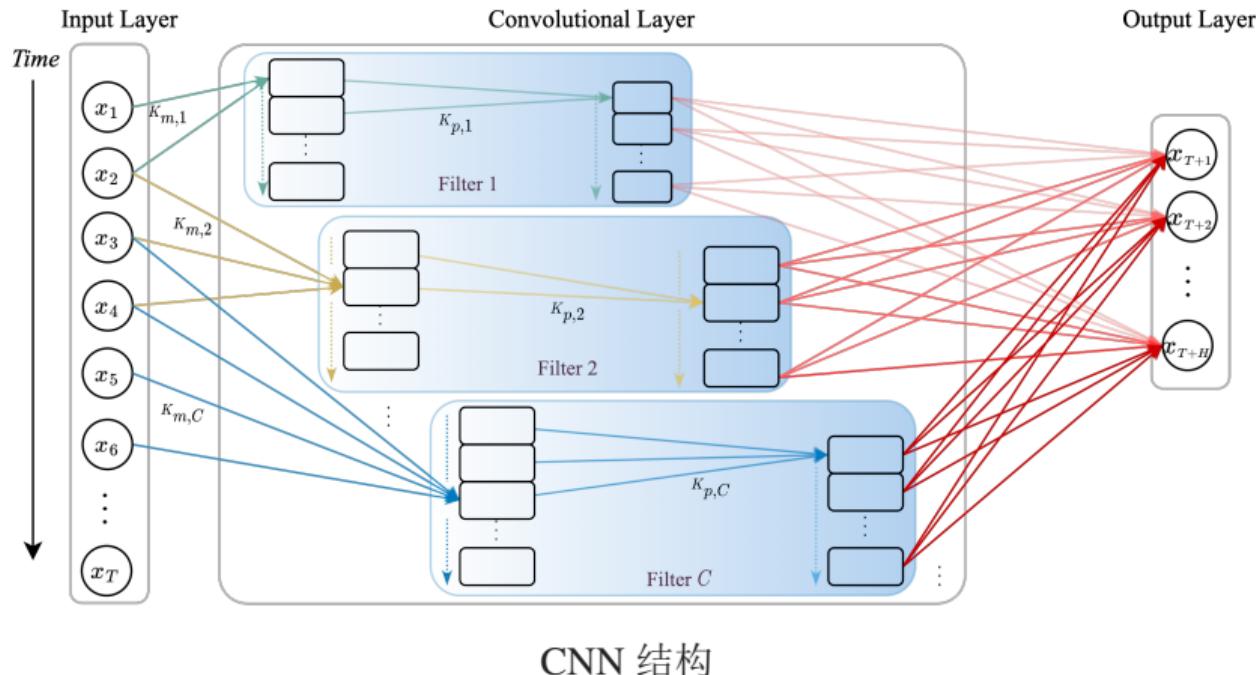
随机映射 CNN 模型研究归纳:

- 既有研究较少，已有随机映射 CNN 模型性能有限，亟待新的随机映射 CNN 预测模型及其构造技术

本章贡献:

- 基于误差反馈随机映射的随机映射 CNN 预测模型构造方法
 - 通过递归生成随机映射卷积核的方式提高了模型构造效率
 - 借助基于误差反馈策略保证了所构造模型的理论收敛性
- 引入贪心算法解决模型构造中随机映射卷积核的选择问题
 - 自适应的在单卷积层内具备不同卷积宽度的卷积核
- 与梯度下降深度学习模型和传统随机映射模型相比，ESM-CNN 具有优秀的预测性能与建模效率

Conv-SDNN 预测模型——CNN 预测模型



Conv-SDNN 预测模型——误差反馈随机映射构造方法

- 过在单卷积层内递归增长添加固定随机初始化权重卷积核的方式构造 CNN 隐藏结构
- 利用误差反馈闭式计算新增卷积核所对应的输出权重参数

对于一个已具有 C 个卷积核的单卷积层 ESM-CNN

$$f_C = \sum_{j=1}^C \sum_{i=0}^{T-K+2} \beta_j^i p_j^i$$

当前预测误差为:

$$e_C = y - f_C = [e_C^1, \dots, e_C^H].$$

新增卷积核所对应的全连接输出层权重将基于当前模型的预测误差 e_C 反馈，通过 OLM 加以求解:

$$[\beta_{C+1}^0, \dots, \beta_{C+1}^{T-K+2}] = \underset{\beta}{\operatorname{argmin}} \|e_C - \sum_{i=0}^{T-K+2} \beta_{C+1}^i p_{C+1}^i\|.$$

Conv-SDNN 预测模型——拟合收敛性

定义 ESM-CNN 的预测误差中间项为: $\tilde{e}_{C+1}^0, \dots, \tilde{e}_{C+1}^{T-K+2}$, 新增卷积核所对应的全连接输出层权重中间项为: $\tilde{\beta}_{C+1}^0, \dots, \tilde{\beta}_{C+1}^{T-K+2}$, 中间项之间的计算关系为:

$$\tilde{e}_{C+1}^{i+1} = \tilde{e}_{C+1}^i - \tilde{\beta}_{C+1}^{i+1} p_{C+1}^{i+1}, \quad i = 0, \dots, T - K + 1,$$

其中,

$$\tilde{\beta}_{C+1}^{i+1} = [\tilde{\beta}_{C+1}^{i+1,1}, \dots, \tilde{\beta}_{C+1}^{i+1,h}, \dots, \tilde{\beta}_{C+1}^{i+1,H}],$$

$$\tilde{\beta}_{C+1}^{i+1,h} = \langle \tilde{e}_{C+1}^{i,h}, p_{C+1}^{i+1} \rangle / \|p_{C+1}^{i+1}\|^2, \quad h = 1, \dots, H.$$

可证:

$$\begin{aligned} & \|\tilde{e}_{C+1}^{i+1}\|^2 - \|\tilde{e}_{C+1}^i\|^2 \\ &= \sum_{h=1}^H \left(\langle \tilde{e}_{C+1}^{i,h} - \tilde{\beta}_{C+1}^{i+1,h} p_{C+1}^{i+1}, \tilde{e}_{C+1}^{i,h} - \tilde{\beta}_{C+1}^{i+1,h} p_{C+1}^{i+1} \rangle - \langle \tilde{e}_{C+1}^{i,h}, \tilde{e}_{C+1}^{i,h} \rangle \right) \\ &= \sum_{h=1}^H \left(-\langle \tilde{e}_{C+1}^{i,h}, p_{C+1}^{i+1} \rangle^2 / \|p_{C+1}^{i+1}\|^2 \right) \leq 0 \end{aligned}$$

Conv-SDNN 预测模型——拟合收敛性

该性质在 $\|\tilde{e}_{C+1}^0\|^2$ 与 $\|e_C\|^2$ 间仍然保持，因此，ESM-CNN 的预测误差收敛性如下：

$$\|e_{C+1}\|^2 \leq \|\tilde{e}_{C+1}^{T-K+2}\|^2 \leq \|\tilde{e}_{C+1}^0\|^2 \leq \|e_C\|^2.$$

方法优势：

- 迭代局部更新输出层权重的方式具有更小的计算开销，这种低开销优势会随隐藏层结构的增大同时扩大
- 随着卷积层中递归新增卷积核的不断加入，链接至输出层的池化特征图向量维度会同时倍数增大，与预测目标维度相比过大的特征维度会导致闭式求解算法的病态问题（Ill-posed problem）
- 基于误差反馈递归更新输出权重构造的预测模型能在神经网络隐藏构造的过程中不断弥补上步预测误差，同时，这种历史输出权重保持固定的方式在一定程度上承担了正则化的作用，从而增强了所构造模型的性能稳定性

Conv-SDNN 预测模型——卷积核选择方法

针对 ESM-CNN 预测模型构造中的卷积核选择问题，本章节提出了一种基于贪心算法的卷积核选择方法，使得 ESM-CNN 具有在单卷积层中同时具备不同宽度的卷积核结构，以此增强模型对于不同尺度时间序列特征的学习建模能力。

本节提出了一种卷积核评分 $\Delta_{C+1,s}$ ，计算如下：

$$\begin{aligned}\Delta_{C+1,s} &= \|e_{C+1,s}\|^2 - \|e_C\|^2 \\ &= \|e_C - \sum_{i=0}^{T-K'+2} \beta_{C+1,s}^i p_{C+1,s}^i\|^2 - \|e_C\|^2\end{aligned}$$

基于所提的卷积核评分，实现最好预测效果提升的备选卷积核 p_{C+1}^* 将被选出作为新增卷积核正式加入当前神经网络结构，该过程可被表述为：

$$p_{C+1}^* = \operatorname{argmax}_{p_{C+1,s}} \{\Delta_{C+1,s}, s = 1, \dots, S\}.$$

Conv-SDNN 预测模型——实验设计

数据集信息

数据集名称	平稳性	趋势性	季节性	起始截止日期	数据集大小
AR1	✗	0.97	0.09	-	500
BTC	✗	0.99	0.66	05/25/2020 ~ 03/20/2021	2181
ILI	✓	0.51	0.61	01/15/2010 ~ 04/15/2020	535
BRENT-weekly	✗	0.97	0.07	05/15/1987 ~ 04/30/2021	1773
BRENT-daily	✗	0.97	0.06	05/20/1987 ~ 05/03/2021	8620
WTI-weekly	✗	0.96	0.08	01/03/1986 ~ 04/30/2021	1844
WTI-daily	✗	0.96	0.07	01/02/1986 ~ 05/03/2021	8904
S&P 500	✗	0.99	0.40	12/31/2009 ~ 11/15/2017	1984
NASDAQ	✗	0.99	0.27	12/31/2009 ~ 11/15/2017	1984
DJI	✗	0.99	0.40	12/31/2009 ~ 11/15/2017	1984
NYSE	✗	0.98	0.45	12/31/2009 ~ 11/15/2017	1984

Conv-SDNN 预测模型——实验设计

对比模型：

- 统计预测模型
 - Naive、ARIMA、Holt's Winters
- 随机映射模型
 - RVFL、IELM、SCN
- 深度学习模型
 - GS-CNN、DeepAR、CLSTM
- 消融实验模型
 - 移除卷积核选择
→ ES-CNN
 - 移除误差反馈与卷积核选择
→ Stoc-CNN

评价指标：

- 百分比误差
 - MAPE、SMAPE
- 绝对值误差
 - RMSE

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|.$$

$$SMAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i + \hat{y}_i} \right|.$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}.$$

Conv-SDNN 预测模型——实验步骤

- 数据集预处理：基于 Z-score 方法⁵归一化至标准正太分布
- 数据集切分：按照 0.64, 0.16 和 0.2 的比例切分为训练集、验证集和测试集
- 测试集评价：预测生成结果均先被还原至原有数值范围，再进行预测准确度计算

均对每一预测模型建模 20 次，对 20 次预测建模结果结合 MAPE、SMAPE 和 RMSE 指标计算预测准确度平均值，以综合表现预测模型预测准确性与稳定性

本章节所有实验均基于 CUDA 10.1 版本的 GPU 加速 Pytorch 框架、Ubuntu 20.04 系统环境、Intel 8700K CPU 和 Nvidia GTX 1070 GPU 环境进行，以保证计算环境的一致性

本章节的实验设置与算法代码已开源在 Github 平台⁶

⁵<https://scikit-learn.org/stable/modules/preprocessing.html#preprocessing-scaler>.

⁶<https://github.com/XinzeZhang/TimeSeriesForecasting-torch>

Conv-SDNN 预测模型——实验结果

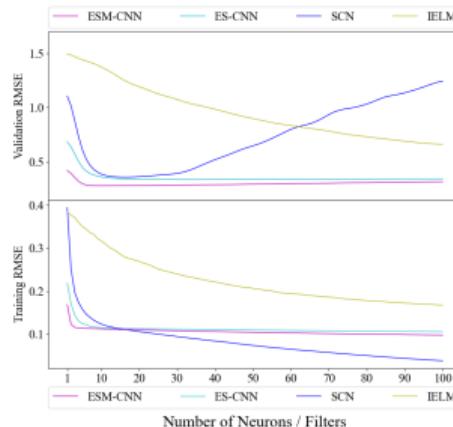
ESM-CNN 及对照组模型的 MAPE 结果对比

数据集	H	统计模型			梯度下降模型			随机映射模型			消融模型		所提模型 ESM-CNN
		Naive	ARIMA	Holt	GS-CNN	DeepAR	CLSTM	RVFL	IELM	SCN	Stoc-CNN	ES-CNN	
AR1	1	2.54e-01	5.95e-01	5.94e-01	1.06e-01	1.08e-01	8.88e-02	7.61e-01	9.84e-02	6.50e-02	6.10e+01	4.53e-02	3.49e-02*
	3	2.62e-01	6.00e-01	5.98e-01	1.23e-01	1.64e-01	1.20e-01	1.17e+00	1.07e-01	9.53e-02	6.34e+01	6.41e-02	5.28e-02*
	6	2.68e-01	6.08e-01	6.05e-01	1.48e-01	2.17e-01	1.24e-01	1.23e+00	1.26e-01	1.16e-01	1.54e+02	8.92e-02	7.70e-02*
BTC	1	2.55e-01	3.17e-01	3.17e-01	6.27e-02	2.26e-01	4.09e-01	1.20e+00	1.85e-01	1.11e-01	2.14e+02	4.62e-02	2.70e-02*
	3	2.54e-01	3.18e-01	3.18e-01	6.43e-02	1.98e-01	4.41e-01	1.13e+00	1.84e-01	1.29e-01	2.86e+02	4.73e-02	2.96e-02*
	6	2.53e-01	3.21e-01	3.20e-01	6.49e-02	2.35e-01	4.61e-01	1.38e+00	1.84e-01	1.40e-01	3.32e+02	4.84e-02	3.26e-02*
ILI	1	1.94e-01	5.70e-01	5.94e-01	1.28e-01	1.14e-01	9.50e-02	3.03e-01	1.38e-01	1.15e-01	2.48e+00	1.00e-01	8.93e-02*
	4	2.51e-01	5.39e-01	5.90e-01	1.52e-01	1.59e-01	1.26e-01	3.53e-01	1.58e-01	1.53e-01	2.33e+00	1.33e-01	1.25e-01*
	8	3.07e-01	5.34e-01	5.98e-01	1.70e-01	1.82e-01	1.45e-01*	3.67e-01	1.74e-01	1.86e-01	2.38e+00	1.62e-01	1.53e-01

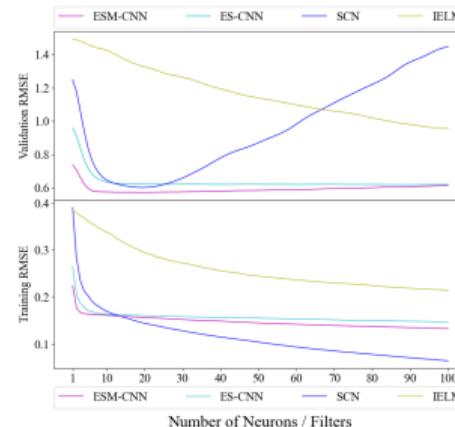
- 1) 作为随机映射神经网络, ESM-CNN 在本章节选取的人工合成时间序列数据集与真实时间序列数据集上都保持了优秀的预测准确度, 展现出 ESM-CNN 预测模型的良好应用性。
- 2) 与 Naive、ARIMA 和 Holt 所代表的统计模型相比, ESM-CNN 在本章节设置的所有预测任务和所有评价指标上都表现出更优的预测准确度, 展现出 ESM-CNN 预测模型的良好预测性能。
- 3) 与 GS-CNN、DeepAR 和 CLSTM 所代表的梯度下降训练深度学习预测模型相比, ESM-CNN 在本章节设置的绝大部分预测任务和评价指标上取得了更优的预测准确度; 在梯度下降训练模型取得最优结果的预测任务中, ESM-CNN 预测模型也取得了差距很小的次优预测结果, 展示出 ESM-CNN 预测模型与梯度下降训练深度学习预测模型匹敌的预测能力。
- 4) 与 RVFL、IELM 和 SCN 所代表的随机映射 MLP 预测模型相比, ESM-CNN 同样在本章节设置的所有预测任务和所有评价指标上都表现出更优的预测准确度, 展示出 ESM-CNN 预测模型在随机映射预测模型中的预测优势。

- 1) 仅单纯引入随机映射方法的 Stoc-CNN 预测模型在本章节选取的所有预测任务和所有评价指标上都表现出显著弱于 ES-CNN 与 ESM-CNN 的预测性能，并在多项预测任务中表现出最差的预测准确度，验证了病态问题在全局更新输出权重方式下所导致的预测性能问题。
- 2) 引入误差反馈随机映射构造策略的 ES-CNN 预测模型与 Stoc-CNN 相比，取得了显著的准确度提升，并在大部分预测任务和评价指标中表现出次优的水平，展现出基于误差反馈随机映射策略构造 CNN 预测模型的有效性与必要性。
- 3) 与 ES-CNN 和 Stoc-CNN 预测模型相比，ESM-CNN 在本章节设置的所有预测任务和所有评价指标上都表现出更优的预测准确度，证明了所提卷积核选择方法在 ESM-CNN 预测模型构造过程中的有效性与必要性。

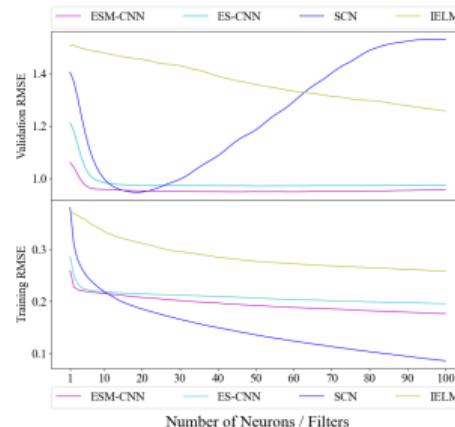
Conv-SDNN 预测模型——收敛性分析



(a) ILI, $H = 1$



(b) ILI, $H = 4$



(c) ILI, $H = 8$

ILI 数据集上递归增长随机映射神经网络的 RMSE 曲线

- 1) 同一预测模型在不同预测时长任务上的收敛情况是一致的。
- 2) 尽管 SCN 模型在不同预测任务的训练集 RMSE 上都能做出最快收敛，但 SCN 模型在进行真实预测建模中出现了明显的过拟合现象，而采用误差反馈策略的 IELM 模型并没有出现过拟合，但表现出最慢的收敛速度。这一对比结果证明了误差反馈策略解决过拟合问题的有效性，同时也指出了误差反馈策略下 MLP 结构的性能局限性。
- 3) 基于卷积结构的 ESM-CNN 和 ES-CNN 在所有的验证集 RMSE 中都表现出比 IELM 和 SCN 更快且更低的收敛效果，证明了基于误差反馈随机映射构造 CNN 预测模型的收敛性与鲁棒性。
- 4) 与 ES-CNN 相比，ESM-CNN 具备更低的训练集 RMSE 曲线和测试集 RMSE 曲线，从收敛性角度再次验证了所提卷积核选择方法的有效性与必要性。
- 5) 在 ESM-CNN 与 ES-CNN 的收敛过程中，预测性能的最大增益明显来自于构造过程的前 10 步，表现出 ESM-CNN 具备通过牺牲微小性能换来巨大效率提升的平衡策略，从而使得 ESM-CNN 具有更加广泛和高效的现实应用性。

Conv-SDNN 预测模型——建模效率分析

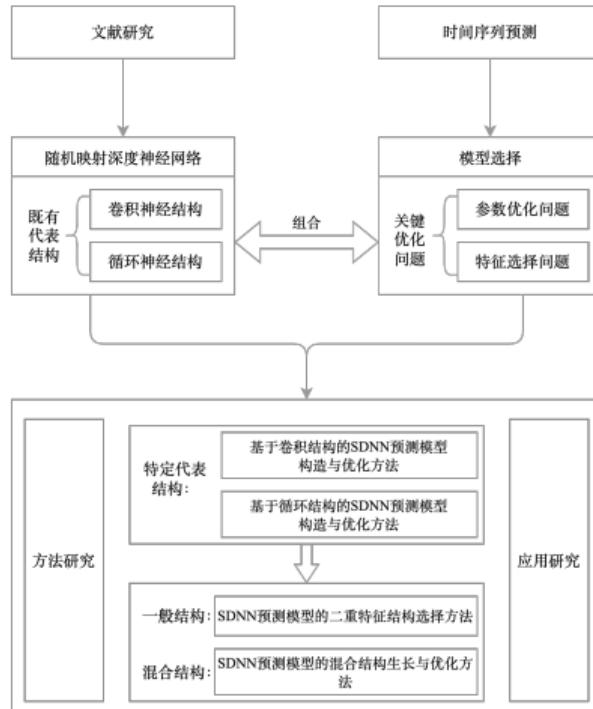
神经网络模型在 AR1、BTC 和 ILI 数据集上的单次平均建模时间 (s)

建模方法	模型	AR1			BTC			ILI		
		1	3	6	1	3	6	1	4	8
梯度下降	GS-CNN	74.65	76.95	78.55	93.70	96.45	99.40	67.05	70.25	73.00
	DeepAR	225.25	238.95	257.70	223.75	251.10	268.65	256.80	279.10	275.30
	CLSTM	150.45	154.80	160.25	143.00	141.50	142.25	167.40	170.80	175.30
随机映射	RVFL	-	-	-	-	-	-	-	-	-
	IELM	2.65	2.85	2.80	3.40	3.30	3.20	2.55	2.55	2.40
	SCN	22.85	57.40	107.75	21.10	58.00	111.40	16.10	49.60	90.95
消融方法	Stoc-CNN	-	-	-	-	-	-	-	-	-
	ES-CNN	6.50	6.40	6.50	6.35	6.50	6.45	6.45	6.50	6.50
所提方法	ESM-CNN	9.65	9.65	9.70	7.10	7.25	7.25	7.15	7.15	7.20

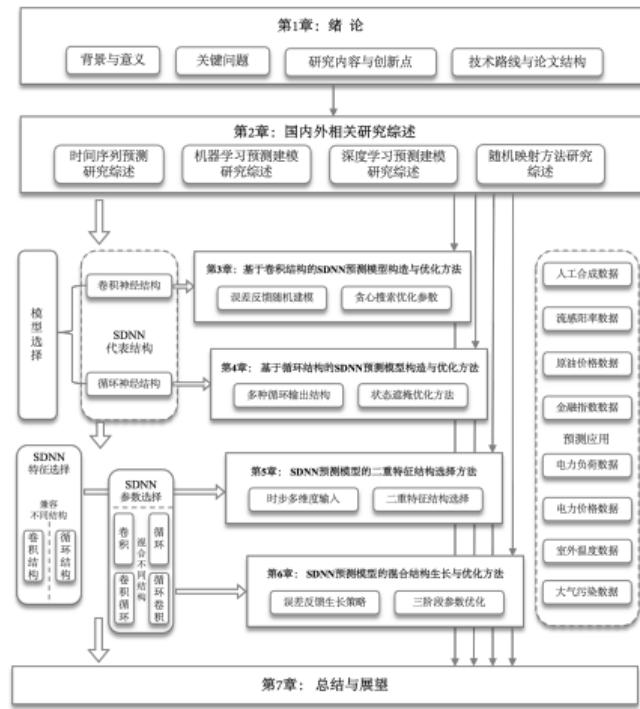
(1) 基于卷积结构的 SDNN 预测模型构造与优化方法

- 通过代数推导证明了基于该方法所构造的预测模型具有随着卷积核的增加而单调下降的预测误差，以此保证了所提方法的收敛性，且其确定的输出权重与既有方法相比具有更小的 L_2 范数，以此提升了模型的预测稳定性；
- 区别于既有方法仅能用相同宽度的卷积核构造单卷积层，通过贪心选择由不同卷积核宽度组成的备选集，使得模型能够在迭代构造模型的过程中自适应确定卷积参数，其构造的单卷积层具备不同宽度的卷积核，使模型借助单卷积层具备不同尺度局部特征的学习能力；
- 在人工合成数据、流感阳率数据、原油价格数据和金融指数数据上的实验表明，与梯度下降训练的深度神经网络预测模型相比，基于本方法所构造的模型在具备相近甚至更优的预测准确度同时具备极高的建模效率，验证了所提方法的有效性。

总结与展望



本研究技术路线



本论文主要结构

卷积结构的 SDNN 预测模型构造与优化方法：

- 可以通过计算不同方法的时间复杂度与空间复杂度，从而精确比较各方法的建模效率优劣
- 基于历史数据训练和构造的模型不一定适应于时变的时间序列数据，因此可以通过考虑剪枝策略，设计模型结构对时变时间序列数据的自适应方法

循环结构的 SDNN 预测模型构造与优化方法研究

- 未有考虑状态遮掩机制对于不同随机循环隐藏结构的影响，因此可以进行状态遮掩输出结构在其他随机循环隐藏结构（如 GESN 或 DESN 结构）上的迁移实验
- 在梯度下降的 RNN 预测模型训练中同样存在类似的问题，因此可以通过设计基于状态遮掩的新颖梯度下降损失函数，提升 RNN 预测模型的预测效果。

SDNN 预测模型的二重特征结构选择方法研究：

- 二维的时间序列输入特征结构同样适用于梯度下降训练的 CNN 与 RNN 模型，因此可以考虑进行二重特征结构选择方法在训练的深度神经网络模型上的迁移实验
- 可以考虑联立循环结构 SDNN 预测模型的输入特征时步掩码和隐藏特征时步掩码，构建多阶段整体优化的循环结构 SDNN 预测模型，进一步提升其预测效果

SDNN 预测模型的混合结构生长与优化方法研究

- 限于 SDNN 预测模型的输入与隐藏结构，未有考虑输出结构的生长与优化问题，因此可以构造混合多种不同输出结构的自适应端到端预测模型
- 随机映射方法下的误差反馈生长策略并不适用于梯度下降方法，因此可以设计基于误差反馈的子网络新颖权重训练方法，构造具有收敛保证的混合结构 DNN 预测模型

围绕上述不足之处与改进方法，未来的研究工作将持续探索时间序列预测建模技术的更多可能与思路。

希望借助随机映射方法，通过进一步的学习与交流，完善深度学习预测模型的解释性研究，探寻预测模型与其优化方法在构造选择效率与预测性能提升间的更佳平衡，使其在更为广泛的应用领域发挥可信且有效的支撑作用。

- 衷心感谢博士生导师鲍玉昆教授、计算机科技与技术学院的何琨教授、硕士研究生导师蔡淑琴教授对我的指导。
- 感谢预答辩委员会的胡斌教授、王林教授、杨彦武教授和吴庆华教授。
- 感谢我的爱人、父母、朋友、师弟师妹，以及所有的支持者！