

Listas de conteúdo disponíveis em [ScienceDirect](#)

# Neurocomputação

página inicial da revista: [www.elsevier.com/locate/neucom](http://www.elsevier.com/locate/neucom)

## Um novo modelo de previsão de séries temporais com aprendizado profundo

Zhipeng Shen, Yuanming Zhang<sup>✉</sup>, Jiawei Lu, Jun Xu, Gang Xiao

Faculdade de Ciência da Computação e Tecnologia, Universidade de Tecnologia de Zhejiang, Hangzhou 310023, China

### informações do artigo

#### Historia do artigo:

Recebido em 30 de março de 2018

Revisado em 10 de novembro de 2018

Aceito em 6 de dezembro de 2018

Disponível on-line em 24 de abril de 2019

#### Palavras-chave:

Aprendizado profundo

Previsão de séries temporais

Memória de curto prazo longa

Convolução causal dilatada

### resumo

A previsão de séries temporais está emergindo como um dos ramos mais importantes da análise de big data. No entanto, os modelos tradicionais de previsão de séries temporais não podem efetivamente extrair recursos de dados de sequência suficientemente bons e, muitas vezes, resultam em baixa precisão de previsão. Neste artigo, um novo modelo de previsão de séries temporais, chamado SeriesNet, que pode aprender completamente os recursos de dados de séries temporais em diferentes comprimentos de intervalo. A SeriesNet consiste em duas redes. A rede LSTM visa aprender recursos holísticos e reduzir a dimensionalidade de dados multicondicionais, e a rede de convolução causal dilatada visa aprender diferentes intervalos de tempo. Esse modelo pode aprender recursos de vários níveis e de vários intervalos a partir de dados de séries temporais e tem maior precisão preditiva em comparação com os modelos que usam intervalos de tempo fixos. Além disso, esse modelo adota aprendizado residual e normalização em lote para melhorar a generalização. Os resultados experimentais mostram que nosso modelo tem maior precisão de previsão e maior estabilidade em vários conjuntos de dados de séries temporais típicos.

© 2019 Elsevier BV Todos os direitos reservados.

### 1. Introdução

Os recentes avanços nas tecnologias de sensores, computação e comunicação são as principais fontes ricas no fornecimento de dados para séries temporais. Esses avanços transformam a maneira como os sistemas complexos do mundo real são monitorados e controlados [1]. Por exemplo, os economistas usam dados de séries de flutuação de preços de ações para prever tendências econômicas, cientistas médicos usam dados de séries temporais biológicas para prever doenças e ambientalistas usam dados de séries temporais atmosféricas para prever mudanças climáticas ambientais. Além disso, a fabricação industrial também usa dados de séries temporais para monitorar a qualidade do produto. Particularmente, as ricas informações causais e dinâmicas contidas nas séries temporais permitem que elas sejam usadas para prever a evolução da complexa dinâmica de sistemas biológicos e de engenharia.

Consequentemente, a previsão de séries temporais está emergindo como um dos ramos importantes da análise de big data, e seu objetivo é prever com precisão as tendências futuras com base em dados de séries temporais passadas e presentes.

As séries temporais são geradas cronologicamente e possuem características de altas dimensionalidades e dependências temporais. As altas dimensionalidades são que cada ponto de tempo é uma dimensão, e as dependências temporais significam que mesmo que dois pontos numericamente idênticos pertençam a classes diferentes ou prevejam comportamentos diferentes.

De acordo com o número de variáveis de amostragem no ponto de tempo, as séries temporais podem ser divididas em séries temporais de variável única e séries temporais múltiplas.

Essas características complexas tornam muito difícil a previsão precisa de séries temporais.

No passado, a previsão de séries temporais tornou-se uma importante área de pesquisa. Várias abordagens diferentes de previsão de séries temporais foram propostas, como a abordagem autorregressiva [2,3], abordagem de média móvel integrada autorregressiva [4,5], abordagem de máquina de vetor de suporte [6,7] e abordagens baseadas em redes neurais [8–10]. Com base nessas abordagens básicas, várias abordagens híbridas [11–14] também foram propostas. No entanto, como os dados de séries temporais, como dados financeiros, dados de vento, dados sísmicos, etc., geralmente são não lineares e não estacionários, e essas abordagens existentes não podem efetivamente extrair recursos de dados de sequência suficientes para obter uma previsão precisa de séries temporais resultados.

Recentemente, deep learning [15,16] é uma nova abordagem que pode obter uma representação em vários níveis da entrada original combinando módulos simples, mas não lineares. Cada módulo da rede neural profunda transforma um nível de representação em um nível mais abstrato. Com essas transformações suficientes, o aprendizado profundo pode aprender todos os recursos dos dados de entrada. O aprendizado profundo tem sido usado para Visão Computacional [17–21], Reconhecimento de Fala [22–24], Processamento de Linguagem Natural [25,26], etc.

Como os métodos tradicionais geralmente são difíceis de extrair futuros de séries temporais suficientes, suas precisões de previsão são limitadas, especialmente em conjuntos de dados não lineares e não estacionários. Portanto, a fim de melhorar a precisão da previsão de séries temporais. Propomos um novo modelo de previsão de séries temporais baseado em aprendizado profundo, chamado SeriesNet, que pode aprender completamente os recursos de dados de sequência em diferentes comprimentos de intervalo. Nossas principais contribuições incluem:

#### <sup>✉</sup> Autor correspondente.

Endereços de e-mail: [zym@zjut.edu.cn](mailto:zym@zjut.edu.cn) (Y. Zhang), [viivan@zjut.edu.cn](mailto:viivan@zjut.edu.cn) (J. Lu),

[xujun@zjut.edu.cn](mailto:xujun@zjut.edu.cn) (J. Xu), [xg@zjut.edu.cn](mailto:xg@zjut.edu.cn) (G. Xiao).

<https://doi.org/10.1016/j.neucom.2018.12.084>

0925-2312/© 2019 Elsevier BV Todos os direitos reservados.

- Propomos um novo modelo de previsão de séries temporais com profundidade aprender e projetar uma nova estrutura de rede para extrair totalmente recursos de dados em série.
- Adotamos aprendizado residual e normalização de lote no SeriesNet para melhorar sua precisão de previsão e usar um ativar portão que faz com que o SeriesNet possa se adaptar a séries temporais dados.
- Avaliamos o SeriesNet em várias séries temporais típicas conjuntos de dados e mostrar que este modelo tem maior precisão de previsão e maior estabilidade do que os métodos tradicionais.

Na Seção 2, analisamos o trabalho relacionado e mostramos o tempo pré-processamento de dados em série na Seção 3. Em seguida, mostramos a estrutura do modelo SeriesNet na Seção 4, e na Seção 5, avaliamos isso modelo com vários conjuntos de dados típicos de séries temporais abertas e concluir este artigo na Seção 6.

## 2. Trabalho relacionado

O modelo clássico de previsão de séries temporais é autorregressivo (AR) proposto pelo estatístico britânico U. Yule. O modelo AR é uma representação do processo estocástico, e suas variáveis de saída são muito dependentes de seus valores anteriores e condições aleatórias.

Com base nesse modelo clássico, outros modelos de RA aprimorados, como a média móvel (MA) e a média móvel autorregressiva (ARMA) são propostas. No entanto, o modelo ARMA não é adequado para dados de séries temporais não estacionárias. Por esta razão, o modelo Autoregressive Integrated Moving Average (ARIMA) [27] é proposto posteriormente. A etapa de uma ou mais vezes diferenciando no ARIMA torna os dados da série temporal estacionários. A operação diferencial geralmente amplifica o ruído de alta frequência em dados de séries temporais, e então afetará a precisão da previsão. Baseado no ARIMA, Peter e Silvia [28] propõem o modelo Autoregressive Integrated Moving Average com modelo de entradas exógenas (ARIMAX) para processar outras séries temporais.

Thissen et al. [29] e Gui et al. [30] mostram outro novo modelo de previsão baseado em máquina de vetores de suporte (SVM). A previsão de séries temporais pertence ao problema de regressão. O SVM é adequado para lidar com este problema, que é chamado de vetor de suporte regressão (SVR). Ele mapeia amostras do espaço original para o espaço de alta dimensão e, em seguida, constrói uma função de decisão linear para alcançar a regressão linear. O SVR pode alcançar melhores resultados experimentais do que o modelo ARIMA anterior quando as séries temporais dados são muito complexos. Máquinas de vetor de suporte de mínimos quadrados (LS SVM) altera as restrições da função objetivo SVM [31]. Wang e Hu [32] compara o SVM e o LS-SVM para resolver o problema de regressão, e descobre que o LS-SVM é mais adequado para problema de grande escala, e a eficiência computacional também é muito superior.

O terceiro modelo de previsão é baseado em rede neural. Esse modelo tem grandes vantagens na resolução de problemas não lineares. De acordo com o teorema da aproximação universal, a rede neural pode aproximar qualquer função mensurável de Borel de um finito espaço dimensional para outro espaço dimensional finito com qualquer precisão [33]. Zhang e Berardi [8] apresentam um modelo de conjunto de redes neurais que combina diferentes estruturas de redes neurais.

Ele pode superar consistentemente a previsão de uma única rede neural, e também pode melhorar a capacidade de previsão evitando problemas de ajuste excessivo. No entanto, o desempenho da rede neural depende do tamanho da amostra e do nível de ruído para problemas lineares. Zhang [11] propõe um modelo híbrido que combina o ARIMA modelo e rede neural. Este modelo híbrido pode obter melhores precisão de previsão em dados lineares e não lineares com melhor desempenho. Além disso, Chow e Leung [34] propõem uma rede neural preditiva não linear para prever séries temporais de curto prazo.

Esta rede é baseada em uma rede de conexão realimentada com modelo autorregressivo não linear.

Nos últimos anos, o aprendizado profundo foi proposto para previsão de séries temporais. Connor e Martin [35] dão rede (RNN) que pode usar informações históricas de séries temporais para prever resultados futuros. Mais tarde, um RNN melhorado chamado Long Short Term Memory (LSTM) é proposto para previsão de séries temporais [36]. Mittelman [37] propõe o método totalmente convolucional não dizimado rede neural (UFCNN) para lidar com problemas de séries temporais. o UFCNN não levará ao desaparecimento ou explosão do gradiente, e é mais fácil de ser treinado. Dasgupta e Osogami [38] propõem a máquina de Boltzmann dinâmica gaussiana (Gaussian DyBM) combinando o RNN. O método de atualização de parâmetros do Gaussian DyBM não depende do algoritmo de retropropagação e, portanto, a atualização dos parâmetros é independente da dimensão dos dados ou atraso máximo. Essas características fazem com que o gaussiano DyBM mais robusto e escalável. Além disso, seu desempenho também é superior ao do popular modelo LSTM. Naduvil-Vadukootu et al. [39] mostra um modelo em pipeline que combina alguns métodos de processamento de dados de séries temporais de última geração com rede neural profunda (DNN). Pode reduzir a dimensionalidade dos dados de entrada. Nisso Desta forma, este modelo melhora a eficiência computacional da DNN.

Os outros modelos de previsão de séries temporais incluem espaço de estados vizinhança [40], processos gaussianos (GP) [41], modelo de decomposição funcional [42]. Esses modelos pertencem a abordagens não paramétricas e fornecem uma representação completa da dinâmica de acordo com os dados observados [1].

Comparado com os métodos acima, não usamos intervalos de tempo fixos para modelar séries temporais. Em vez disso, usamos LSTM e casual dilatado convolução para extrair recursos com diferentes intervalos de tempo de séries temporais e, finalmente, combinar esses recursos. Isso pode se inclinar totalmente características de séries temporais e pode ajudar a melhorar a precisão da previsão.

## 3. Pré-processamento de dados de séries temporais

Geralmente, como os dados de séries temporais são gerados pelo sistema natural, uma variedade de motivos inesperados levará à redundância de dados, dados ausentes, anormais e outros problemas no processo de dados aquisição e transmissão. O pré-processamento de dados é necessário para tornar os dados da série adequados para SeriesNet. A Fig. 1 mostra o procedimento de pré-processamento que inclui várias etapas.

A etapa de limpeza de dados é lidar com a redundância, NULL e outliers existentes nos dados de séries temporais originais. Dados redundantes fará com que as séries temporais não sejam alinhadas com a dimensão temporal, e levará a grandes erros no SeriesNet. Por outro lado, é também não é adequado simplesmente descartar valores ausentes, já que fazer isso perderá uma grande quantidade de informações e não ficará completo dados de séries temporais.

O estágio de normalização é dimensionar todas as dimensões do recurso para mesma escala, porque os dados a serem previstos muitas vezes apresentam inconsistências numéricas.

O estágio da janela deslizante é dividir o conjunto de dados de acordo com comprimento da janela deslizante. O comprimento da janela deslizante indica a comprimento de um único tamanho de passo de tempo de entrada, o que também significa que o modelo irá prever o valor da próxima vez com base nas informações históricas do comprimento da janela. Por exemplo, há um tempo sequência de série {1,2,3,4,5,6,7}. Quando o tamanho da janela deslizante é 3 e o tamanho do passo deslizante é 1, a saída das sequências de séries temporais é {{1,2,3},{2,3,4},{3,4,5},{4,5,6},{5,6,7}}. Nesta etapa, os dados da série original são construídos como a entrada formal de acordo com o deslizamento tamanho da janela.

O estágio de divisão é dividir os dados originais em três partes: os dados de trem que representam 60%, os dados de teste que representam 20% e os dados de validação que representam 20% do total

conjunto de dados.

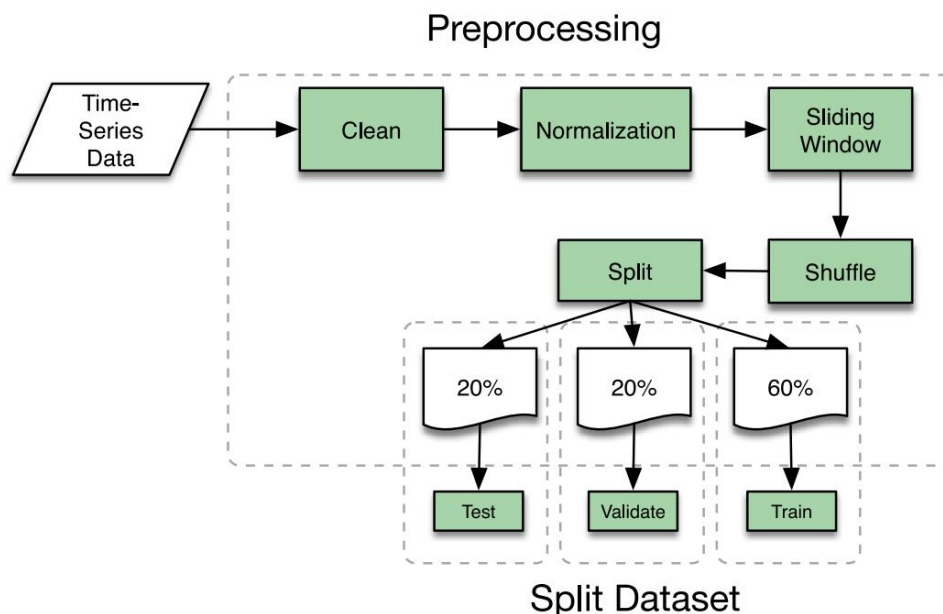


Fig. 1. Etapas de um pré-processamento.

#### 4. Principal da SeriesNet

##### 4.1. Rede de memória de longo prazo

Comparado com a rede neural artificial tradicional, o Recurrent A rede neural pode usar informações históricas de dados de séries temporais para estrutura de rede especial. O RNN é capaz de definir os dados de saída de acordo com as entradas atuais e passadas. No entanto, o RNN usa o Algoritmo BPTT para treinar a rede, quando processa muito tempo intervalos em sequência, o gradiente desaparecerá [43]. E quanto mais tempo os intervalos se tornam, mais sério o gradiente se tornará. Isso torna difícil treinar efetivamente o RNN por longos intervalos de tempo.

Para lidar com este problema, outras variantes de redes neurais são propostas para capturar dependências de longo prazo mais facilmente, como LSTM [44] e GRU [45] possuem uma célula especial estrutura, GF-GRU [46] usa uma forma especial de conexão entre cada duas camadas. O LSTM é uma variante básica, ele usa três estruturas de gate: gate de entrada e portão de saída, para controlar as informações do histórico de memória da camada oculta. As equações no tempo  $t$  são os seguintes:

$$it = \text{sigmoid}(W_{ix}xt + W_{hi}ht_{t-1} + W_{ci}ct) \quad (1)$$

$$ft = \text{sigmoid}(W_{fx}xt + W_{fh}ht_{t-1} + W_{fc}ct) \quad (2)$$

$$ct = ftct_{t-1} + it \tanh(W_{cx}xt + W_{ch}ht_{t-1}) \quad (3)$$

$$ot = \text{sigmoid}(W_{ox}xt + W_{oh}ht_{t-1} + W_{oc}ct) \quad (4)$$

$$ht = \text{otanh}(ct) \quad (5)$$

Onde  $i$  é a porta de entrada;  $f$  é a porta de esquecimento;  $o$  é a saída portão;  $c$  é a célula de memória;  $h$  é a saída da camada oculta;  $W$  é o matriz de pesos, e o subscrito é o item de associação de pesos. Por exemplo,  $W_{xi}$  representa o peso da conexão da camada de entrada para porta de entrada. Os valores de  $i, f, o$  dentro do intervalo de  $[0, 1]$  controlarão a proporção em que os dados históricos passam pelo portão.

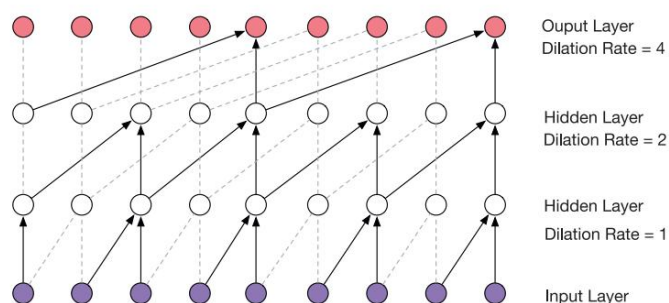


Fig. 2. Visualização de uma pilha de convolução causal dilatada.

##### 4.2. Convolução causal dilatada

A convolução dilatada é proposta para lidar com a perda de resolução ou cobertura devido à operação de downsampling na imagem segmentação semântica [47]. Ele usa convoluções dilatadas para agregar sistematicamente informações contextuais em várias escalas e melhorar a precisão do reconhecimento de imagem. A convolução causal é garantir que o núcleo de convolução da rede neural de convolução possa realizar operações de convolução exatamente na sequência de tempo [48], e que o kernel de convolução só pode ler as informações atuais e informações históricas.

O Google DeepMind combina a convolução dilatada e causal convolução no modelo WaveNet, e propõe uma relação causal dilatada convolução para gerar um som mais natural do que o tradicional método [48]. O WaveNet usa convolução causal dilatada e skip-connection para utilizar informações de longo prazo. A Fig. 2 mostra um pilha de convolução causal dilatada [48]. A dilatação é dobrada para cada camada, como 1, 2, 4, ...,  $2n$ , e tem um valor máximo. Após o valor máximo, ele começa a se repetir. O valor máximo em WaveNet é definido como 29. Como resultado, o campo receptivo aumenta com o aumento do número de camadas.

##### 4.3. Estrutura da SeriesNet

Esta seção descreve a estrutura do SeriesNet. Comparar com o modelo LSTM tradicional que não pode extrair

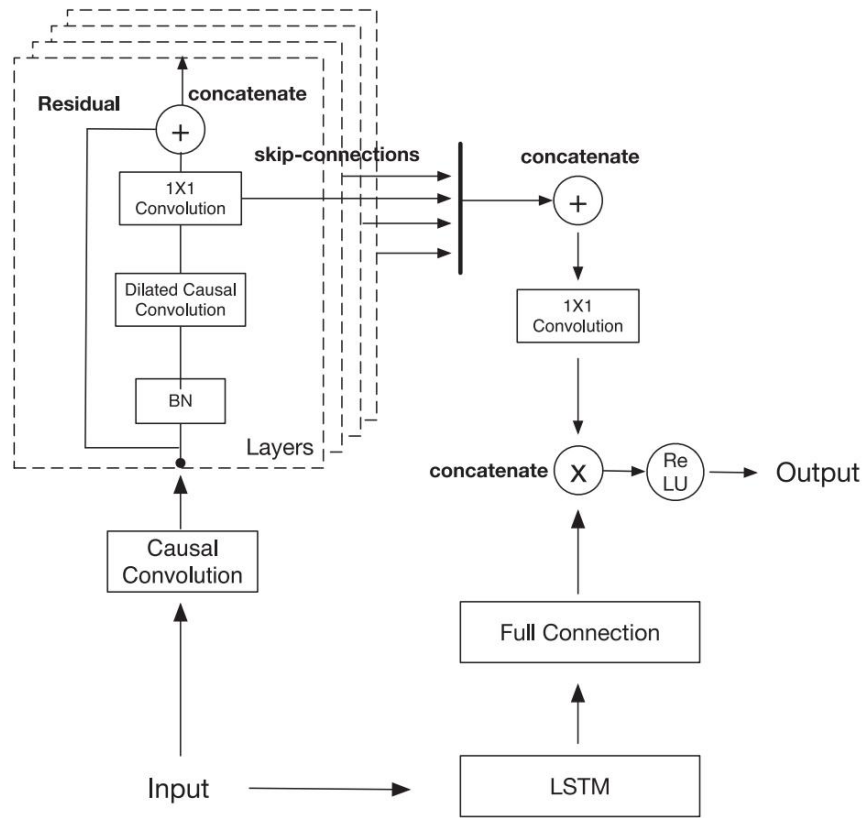


Fig. 3. A estrutura do SeriesNet.

recursos de dados de séries temporais, o SeriesNet pode extrair totalmente recursos de dados temporais.

Assumindo que a sequência de entrada é  $X = (x_1, x_2, \dots, x_{t-1}, x_t)$ , o máxima verossimilhança  $p(X)$  é a seguinte:

$$p(X) = \prod_{t=1}^T p(x_{t+1} | x_1, x_2, \dots, x_t) \quad (6)$$

O  $x_{t+1}$  terá todas as vantagens de todas as informações de passo de tempo da série.

A Fig. 3 mostra a estrutura do SeriesNet. usamos a camada LSTM para aprender os recursos holísticos dos dados de entrada e, em seguida, usa o camada de conexão completa para definir a dimensionalidade de saída do LSTM. Outra parte do seriesNet, convolução causal é adicionada a garantir que as séries temporais estejam em conformidade com a relação causal e, em seguida, ele se conecta ao bloco de conexão residual. A ligação residual bloco inclui operação de convolução causal dilatada. Empilhando o bloco de conexão residual e aumentando a taxa de dilatação convolução causal fará com que o SeriesNet tenha um campo receptivo maior Tamanho. Diferentes taxas de dilatação terão diferentes tamanhos de campo receptivo. Cada bloco de conexão residual extrairá recursos de séries temporais em Níveis diferentes.

O SeriesNet usa a operação de conexão com salto longo para mesclar o recursos de série temporal extraídos de cada bloco de conexão residual. Isso pode garantir que os recursos de série temporal não sejam perdidos nas redes. Além disso, não só a operação pode melhorar o desempenho, mas também pode lidar com diferentes níveis de ruído [49].

Por fim, concatenamos as duas partes da saída, as vantagens da SeriesNet são que suas camadas rasas de blocos residuais aprenderão recursos de intervalo curto, suas camadas profundas de blocos residuais aprenderão recursos de intervalo longo e a camada LSTM aprenderão recursos holísticos. Isso significa que o SeriesNet pode usar informações abundantes de recursos hierárquicos na previsão de regressão para melhorar a precisão da previsão. No SeriesNet, usamos a operação de adição e multiplicação

operação para conectar diferentes partes do modelo. A adição operação é usada para conectar a saída de diferentes camadas no conexão de salto. No bloco de conexão residual, faz parte do função de identidade como em ResNet [50]. A operação de multiplicação é usado para conectar dois tipos de rede. Pode misturar as características, que são extraídas da rede de convolução causal dilatada e rede LSTM.

Por fim, o algoritmo RMSProp é usado para treinar os parâmetros, o pseudocódigo do SeriesNet é descrito no Algoritmo 1 a função de perda. Nós usamos de erro quadrático médio como objetivo do SeriesNet:

$$perdamin = \frac{1}{T} \sum_{t=1}^T (Y_{t+1} - \hat{y}_{t+1})^2 \quad (7)$$

#### 4.4. Série temporal multicondicional

Séries temporais multicondicionais são dados adicionais que podem ter grande impacto na série temporal alvo a ser prevista. Os dados de séries multicondicionais podem ajudar a melhorar a precisão da previsão.

Para séries temporais multicondicionais, o LSTM e a conexão completa camadas podem reduzir a dimensionalidade de dados multicondicionais. Variáveis de entrada adicionais serão introduzidas. Então, a Eq. (8) será alterado da seguinte forma:

$$p(X|y) = \prod_{t=1}^T p(x_{t+1} | x_1, x_2, \dots, x_t, y_1, \dots, y_t) \quad (8)$$

onde  $y_i$  é a variável de entrada da condição adicional  $i$ . O LSTM a rede mapeia os dados da série temporal multicondicional para uma série temporal dimensional inferior, que aprenderá a relação interna entre dados de séries temporais de múltiplas condições.

Além disso, o aprendizado residual garante que a precisão não diminua em redes muito profundas e usa normalização de lote (BN)

[51] para evitar o desaparecimento do gradiente.

**Algoritmo 1:** SeriesNet.

**Requer:** Treinar dataSet  $D_{train} = \{x_i\}_{i=1}^n$   $D_{test} = \{x_i\}_{i=n+1}^m$ , conjunto de dados de teste  $i=1, \dots, m$

$\{x_i\}_k$  **Requer:** Passo de tempo  $s$ , parâmetro do modelo  $\tilde{y}$ , taxa de aprendizado  $\tilde{\gamma}$ , pequena constante  $\tilde{\epsilon}$ , taxa de decaimento  $\tilde{\alpha}$  e tamanho do lote  $m$  **Requer:** Janela deslizante de previsão  $j$  **Entrada:**  $D_{in} = \{x_i\}_{i=1}^n, \{x_i\}_{i=n+1}^m, \dots, \{x_i\}_{i=n\tilde{y}+1}^m$

$i=2$ ,  $\{x_i\}_{i=2}^n, \{x_i\}_{i=n+1}^m, \dots, \{x_i\}_{i=n\tilde{y}+1}^m$  **Saída:**  $\{y_i\}_{i=1}^m$

Resultados da previsão Rota 1 Inicializar os parâmetros *SeriesNet*  $\tilde{y}$ ; **2** Inicialize a variável de acumulação de gradiente  $r = 0$ ; **3** Faça uma amostra de um minilote de  $m$  exemplos da entrada; **4 enquanto** o critério de parada não é satisfeito **do** **5** Calcular gradiente (minilote):  $g = \frac{\partial L}{\partial \tilde{y}}$   $\tilde{y} = \tilde{y} - \tilde{\gamma} g$ ; **6** Gradiente acumulado:  $r = \tilde{\alpha} r + (1 - \tilde{\alpha}) g$ ; **7** Novos parâmetros:  $\tilde{y} = \tilde{y} - \tilde{\gamma} r$ ; **8** Parâmetros de atualização:  $\tilde{y} = \tilde{y} - \tilde{\gamma} r$ ; **9**  $\tilde{y} = \tilde{y} - \tilde{\gamma} r$ ; **10**  $\tilde{y} = \tilde{y} - \tilde{\gamma} r$ ; **11**  $\tilde{y} = \tilde{y} - \tilde{\gamma} r$ ; **12**  $\tilde{y} = \tilde{y} - \tilde{\gamma} r$ ; **13** Calcular o valor da previsão do próximo passo:  $\tilde{y}$ ; **14** Anexe  $\tilde{y}$  ao final de  $\{x_i\}_{i=1}^n, \{x_i\}_{i=n+1}^m, \dots, \{x_i\}_{i=n\tilde{y}+1}^m$ ; **15** **final**

**16** Gere o resultado da previsão:  $R_{out} = \{v_1, v_2, \dots, v_j\}$

4.5. Residual profundo

Uma rede muito profunda levará ao desaparecimento do gradiente e isso diminuirá a precisão da previsão. Para lidar com este problema chave, a aprendizagem residual [52] é proposta.

Suponha que o mapeamento original da camada oculta seja  $H(x)$ , e deixe que cada bloco residual aprenda o mapeamento residual entre entrada e saída:

$$F(x) = H(x) \oplus x$$

Ou seja, a saída da camada oculta original é representada como:

$$H(x) = F(x) \oplus x$$

Essa modificação principal permite que a rede apenas precise aprender o mapeamento residual de cada camada em vez de aprender o mapeamento de entrada e saída. Quando o mapeamento entre  $H(x)$  e  $F(x)$  é identidade, a rede pode detectar pequenas perturbações. Ao empilhar os blocos residuais, a rede residual profunda garante que a precisão da previsão não será diminuída. Quando o modelo aprende recursos de alto nível, o aprendizado residual torna possível treinar uma rede muito profunda.

Em comparação com os dados do sinal de áudio, os dados de séries temporais serão muito mais simples. A unidade de ativação de portas no WaveNet é adequada para interações complexas [53], como sinais de áudio. Ele divide os mapas de recursos que são saídas de cada operação de convolução causal e, em seguida, usa uma função de ativação tanh e sigmoid para processar a saída anterior. Sua não linearidade funciona muito melhor do que a função de ativação linear retificada [48]. No entanto, em comparação com os dados do sinal de áudio do som, os dados de séries temporais são muito mais simples. Ou seja, o rec

a ativação linear especificada é adequada para dados de séries temporais. Então, a ativação da convolução causal dilatada é alterada para:

$$z = \text{ReLU}(W_k \oplus x)$$

Onde  $k$  é a sequência dos blocos residuais;  $W_k$  é o peso do filtro de convolução no  $k$ -ésimo bloco residual.

4.6. Normalização em lote

Para aprender representações abstratas de alto nível de dados de entrada, a profundidade da rede profunda pode ser adicionada. No entanto, isso fará com que a distribuição dos parâmetros de cada camada mude ao treinar a rede neural.

Como a distribuição de peso de cada camada é diferente, a taxa de aprendizado unificada tem efeito diferente na atualização de peso de cada camada. É difícil coordenar a atualização de peso entre diferentes camadas. Devido a esta razão, é difícil selecionar uma taxa de aprendizado adequada para treinar a rede. Como resultado, a rede terá um gradiente nulo, chamado de deslocamento covariável interno [51]. À medida que o aprofundamento da rede, os efeitos negativos tornam-se mais graves e levam a uma formação eficaz.

O SeriesNet adiciona a camada de normalização de lote (BN) após a entrada de cada camada no bloco de conexão residual e redistribui a entrada. Ele irá normalizar cada característica escalar dos dados de entrada independentemente para fazer com que o valor da média seja 0 e o valor da variância seja 1. Os parâmetros da BN também podem ser aprendidos durante o treinamento, o que faz com que o modelo recupere facilmente para a entrada de origem .

5. Avaliação

Usamos três conjuntos de dados típicos de séries temporais abertas: o S&P 500 Index, o Shanghai Composite Index e a temperatura de Hangzhou para avaliar o SeriesNet. Os dados do S&P 500 Index e do Shanghai Composite Index são do Yahoo Finance e os dados meteorológicos são da Administração Meteorológica da China. A Tabela 1 mostra as informações de perfil desses três conjuntos de dados. Ele mostra o nome da série temporal, o intervalo de tempo, o número do conjunto de dados do trem, o número do conjunto de dados de validação e o número do conjunto de dados de teste. Os dados econômicos são não estacionários e não lineares, os dados de temperatura são relativamente estacionários.

A Tabela 2 mostra os melhores hiperparâmetros do SeriesNet no conjunto de dados S&P 500. Nesta tabela, a coluna Tipo significa o funcionamento de cada camada. Usamos a adição de matriz para conectar duas camadas. Se a adição estiver marcada com Skip-Connection, ela usará a operação skip-connection para conectar todas as camadas de saída. A coluna de Unidades é o número de neurônios. Quando seus tipos são convolucionais, os valores da coluna são o número de filtros. A coluna de Tamanho é o tamanho do kernel dos filtros. O SeriesNet usa diferentes taxas de dilatação em diferentes camadas quando prendemos os blocos de conexão residuais. Além disso, a taxa de dilatação da operação convolucional normal é definida como 1. A coluna de Padding é a forma de preenchimento. A coluna de Saída são as dimensões de saída de cada camada. Digno de nota é que os hiperparâmetros do modelo e o número de blocos residuais são ligeiramente ajustados quando o SeriesNet se aplica a diferentes conjuntos de dados.

Comparamos o SeriesNet com o LSTM, o UFCNN, o ANN e o SVM, o LSTM é o modelo de aprendizado profundo mais popular e o UFCNN é um modelo avançado de rede neural de convolução, o

Tabela 1  
Conjunto de dados de séries temporais.

Série temporal	Período	Treinar dados	Dados de validação	Dados de teste
Índice S&P 500	2001,01–2017,05	2750	2005,01–	878
Shanghai Composite Index A	2017,06	1530		510
temperatura de Hangzhou	2011,01–2017,01	1880	470	470

mesa 2

Hiperparâmetros do SeriesNet.

Tipo	Unidades/Filtros	Tamanho	Saída de preenchimento de taxa de dilatação		
Convolutacional (entrada)	8	30	1	causal	90 x 8
Normalização em lote					
Convolutacional	8	7		causal	90 x 8
Convolutacional	1	1	1 1	mesmo	90 x 1
Adicionar					90 x 8
Normalização em lote					
Convolutacional	8	7	2	causal	90 x 8
Convolutacional	1	1	1	mesmo	90 x 1
Adicionar					90 x 8
Normalização em lote					
Convolutacional	8	7	4	causal	90 x 8
Convolutacional	1	1	1	mesmo	90 x 1
Adicionar					90 x 8
Normalização em lote					
Convolutacional	8	7	8	causal	90 x 8
Convolutacional	1	1	1	mesmo	90 x 1
Adicionar					90 x 8
Normalização em lote					
Convolutacional	8	7	16	causal	90 x 8
Convolutacional	1	1	1	mesmo	90 x 1
Adicionar(Saltar conexão)					90 x 1
Convolutacional	1	1	1	mesmo	90 x 1
LSTM (Entrada)	10				90 x 10
LSTM	10				90 x 10
FC	1				90 x 1
Multiplicar					90 x 1

Tabela 3

O resultado da janela deslizante de previsão é 1.

Modelo de série temporal S&P 500 Index		Índice Composto de Xangai A temperatura de Hangzhou		
	RMSE MAE R2	RMSE MAE R2	RMSE MAE R2	
SeriesNet10	<b>17,32 13,15 0,982</b>	63,94 38,37 0,976	2,82 2,06 0,903	
SeriesNet50	19.97 15.56 0,976	<b>63.59</b> 38,41 <b>0,977</b> 64,54	2,92 2,16 0,896	
SeriesNet100	33.54 26,72 0,933 18.55	39.37 0,976 80,60 50.56	2,83 2,09 0,903	
SeriesNetNoRes	13.94 0,979 19.04 14.42	0,962 63.84 <b>38,05</b> 0,976	2,91 2,13 0,897	
LSTM2 <sub>20</sub>	0,978 20.79 15.57	64.72 40.55 0,975 64,85	2,86 2,09 0,901	
LSTM2 <sub>50</sub>	0,974 18.59 14.02 0,979	41,21 0,975 66,18 39.66	2,87 2,07 0,900	
LSTM2 <sub>100</sub>	26.31 20.13 0,959	0,975 93.06 57.77 0,950	2,89 2,09 0,899	
LSTM4 <sub>100</sub>	24.36 19,84 0,965 24.22	66,25 39.35 0,975	2,87 2,10 0,901	
UFCNN	20.84 0,965		<b>2,64 1,97</b> 0,907	
ANN			2,95 2,14 0,895	
SVM	31,55 22,92 0,941	69,02 40,43 0,972	2,74 2,04 <b>0,909</b>	

Tabela 4

O resultado da janela deslizante de previsão é 5.

Índice S&P 500 do Modelo de Série Temporal		Índice Composto de Xangai A temperatura de Hangzhou		
	RMSE MAE R2	RMSE MAE R2	RMSE MAE R2	
SeriesNet10	<b>31,47 23,76 0,941</b>	121,53 71,18 0,915	3,95 3,02 0,811	
SeriesNet50	47,34 35,60 0,866 51,40	<b>116,59 67,43 0,921</b>	4,19 3,20 0,788	
SeriesNet100	40,72 0,843	121,32 72,24 0,915	4,04 3,14 0,802	
SeriesNetNoRes	36,94 27,22 0,918	144,55 88,67 0,879	4,10 3,13 0,796	
LSTM2 <sub>20</sub>	34,98 25,85 0,927 40,36	122,25 69,81 0,914	4,18 3,25 0,788	
LSTM2 <sub>50</sub>	30,11 0,903 38,55 28,81	121,65 77,13 0,915	4,07 3,16 0,799	
LSTM2 <sub>100</sub>	0,912 61,19 43,46 0,778	123,64 81,06 0,912	4,11 3,18 0,796	
LSTM4 <sub>100</sub>	36,26 29,13 0,922 67,05	130,87 78,35 0,901	4,08 3,18 0,799	
UFCNN	55,68 0,733 50,34 36,14	194,81 126,23 0,781	4,80 3,85 0,692	
ANN	0,849	129,98 76,43 0,902 138,64	4,84 3,69 0,716	
SVM		78,13	<b>3,73 2,88 0,832</b>	

ANN é o modelo de rede neural mais básico, e o SVM é o modelo de previsão de séries temporais mais clássico.

Este artigo usa o erro quadrático médio (RMSE), a média erro absoluto (MAE) e o coeficiente de determinação (R2) como critérios de avaliação, a fórmula de R2, RMSE e MAE são tão segue:

$$R2 = 1 - \frac{\sum_{i=1}^n (f_i - y_i)^2}{\sum_{i=1}^n (f_i - \bar{y})^2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2}$$

$$FÁCIL = \frac{1}{n} \sum_{i=1}^n |f_i - y_i|$$



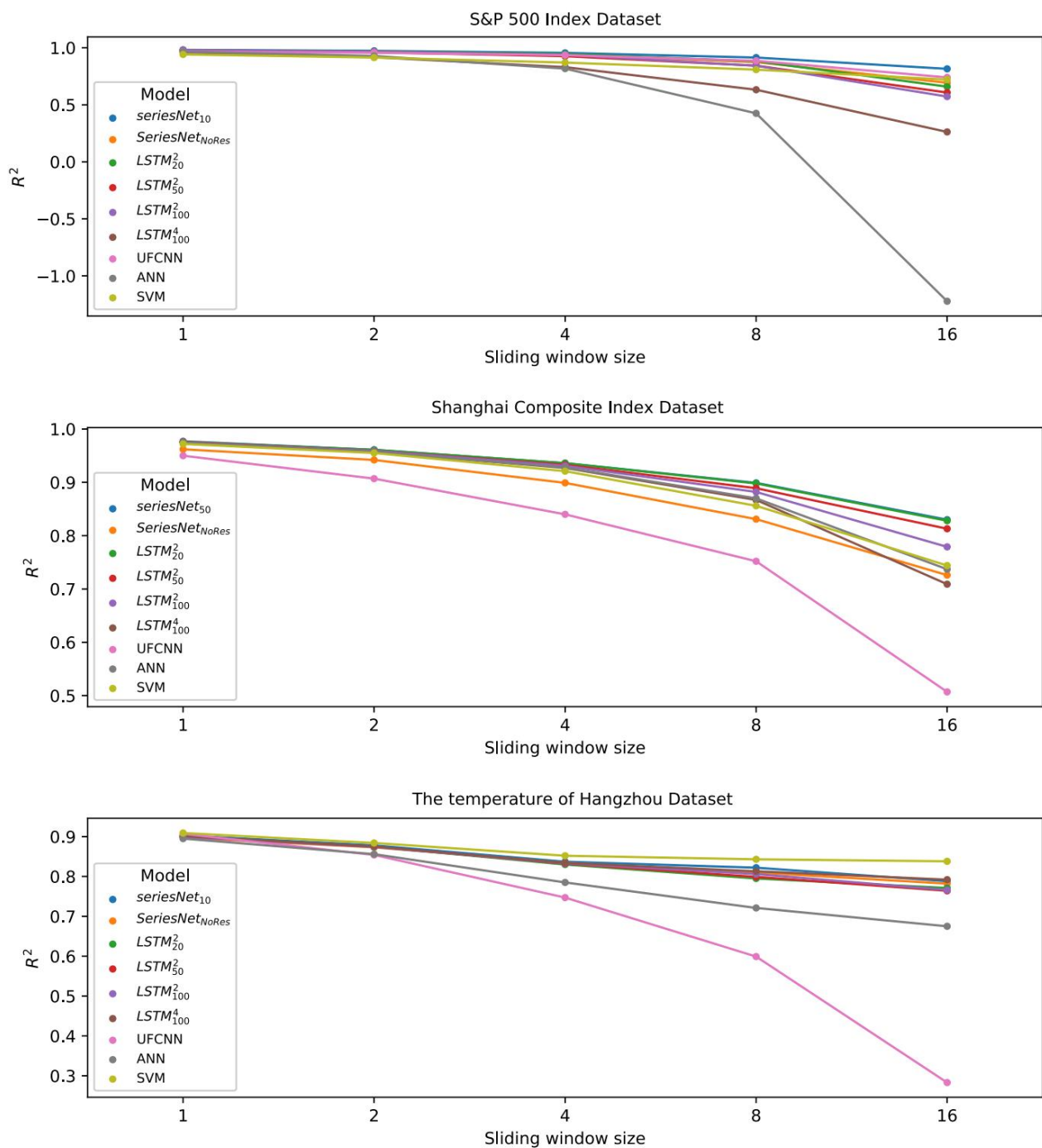


Fig. 4. Comparação de diferentes tamanhos de janelas deslizantes em três conjuntos de dados.

Quanto mais próximo de 1 o  $R^2$  estiver, maior será a precisão da previsão. E quanto menor o RMSE e o MAE, menor o erro estão.

A Tabela 3 mostra os resultados experimentais com três conjuntos de dados, quando a janela deslizante de previsão que representa o intervalo de tempo futuro é 1.

Usamos quatro diferentes superparâmetros do LSTM como linha de base para comparar com o SeriesNet, o UFCNN, o ANN e o SVM em três conjuntos de dados. LSTM2 tem 2 camadas de células LSTM e cada camada tem 2 camadas de célula contêm 20 neurônios, a camada 50 LSTM contém 50 neurônios e assim por diante. Além disso, também usamos três modelos SeriesNet diferentes que possuem tamanhos de LSTM diferentes. Além disso, um modelo SeriesNet sem resíduo também é usado para verificar a eficácia do aprendizado residual.

Pode-se observar na Tabela 3 que os modelos SeriesNet têm a melhor precisão de previsão em conjuntos de dados não lineares e não estacionários em comparação com outros modelos. Para os modelos SeriesNet, diferentes conjuntos de dados têm diferentes precisões. Por exemplo, o SeriesNet10 é o melhor no conjunto de dados do índice S&P 500 e o SeriesNet50 é o melhor no conjunto de dados do Shanghai Composite Index. A razão desse fenômeno é que as características das séries desses conjuntos de dados são bastante diferentes. Flutuações de dados complexos exigem um número maior de neurônios ou modelo complexo para ajustá-los. Pelo contrário, flutuações de dados simples adotando estruturas complexas levarão a uma baixa capacidade de generalização. Ou seja, existe overfitting no modelo. Como resultado, os hiperparâmetros do SeriesNet precisam ser cuidadosamente ajustados.

Além disso, o resultado mostra que o SeriesNet não é o melhor para

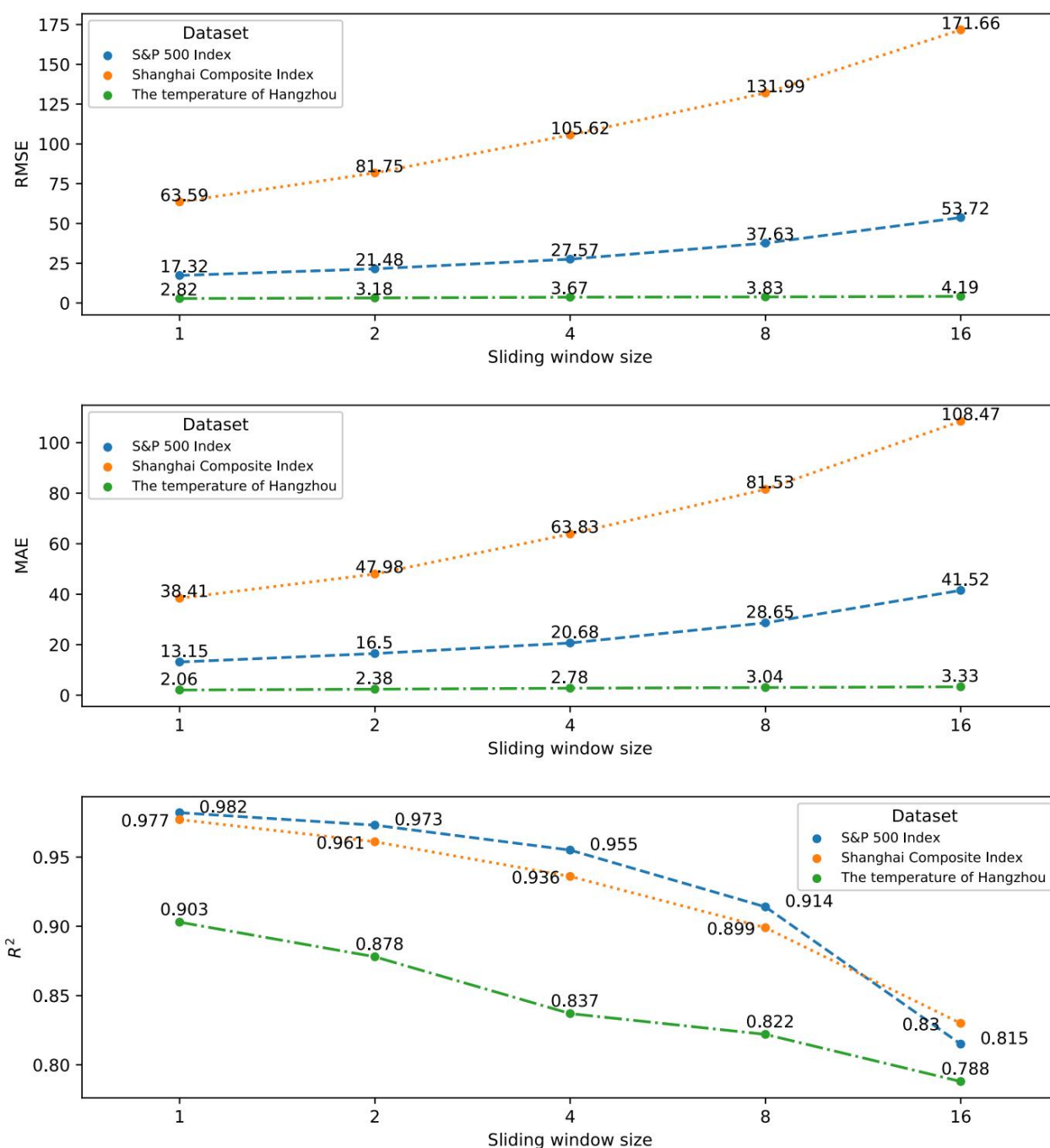


Fig. 5. Comparação de diferentes tamanhos de janela deslizante.

dados de séries temporais estacionárias, como dados de temperatura, porque parâmetros excessivos levam ao overfitting facilmente, enquanto o modelo SVM adota o princípio de minimizar o risco estrutural para encontrar a solução ideal. Isso faz com que o SVM se ajuste bem aos dados de séries temporais de regularidade forte.

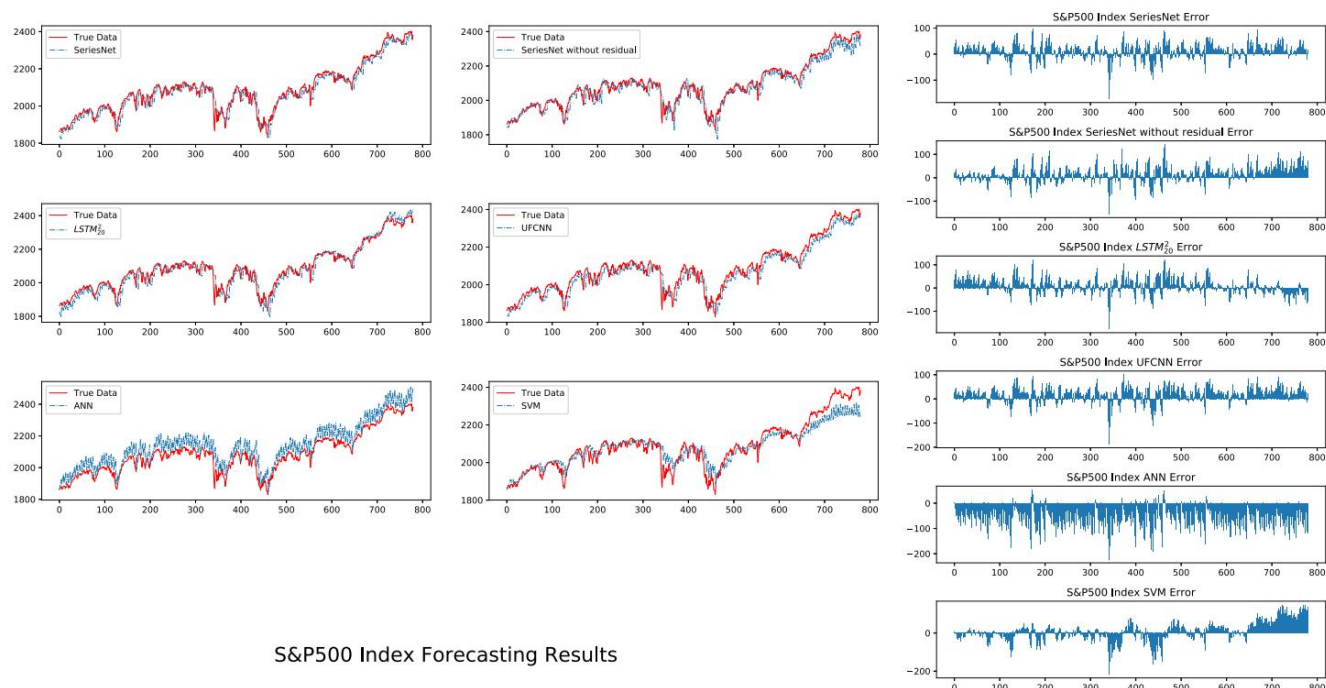
Tirando os melhores resultados do SeriesNet, SeriesNet sem resíduo, LSTM, UFCNN, ANN e SVM em três datasets. O valor médio de  $R^2$  desses seis modelos é 0,954, 0,946, 0,952, 0,941, 0,945 e 0,941, respectivamente, para os três conjuntos de dados. Pode-se observar que o SeriesNet é geralmente superior a outros modelos.

A Tabela 4 mostra os valores de RMSE, MAE e  $R^2$  com os três conjuntos de dados, quando o tamanho da janela deslizante é 5. O valor médio de  $R^2$  de seis modos é 0,891, 0,864, 0,880, 0,798, 0,784 e 0,857 respectivamente para os três conjuntos de dados.

Geralmente, o tamanho grande da janela deslizante resultará em baixa precisão de previsão. O próximo resultado da previsão será realizado com base no resultado da previsão anterior, pois aumentar o tamanho da janela deslizante introduzirá mais erros cumulativos. A Fig. 4 mostra o declínio de  $R^2$  à medida que o tamanho da janela deslizante aumenta nos três conjuntos de dados e a Fig. 5 mostra três métodos de avaliação no modelo SeriesNet. Esses números mostram que a precisão diminui rapidamente com o aumento do tamanho da janela deslizante. Embora seja difícil prever com precisão séries temporais de longo período, é mais valioso obter tendências futuras de séries temporais. O erro do SeriesNet diminui lentamente com o aumento do tamanho da janela deslizante em comparação com outros modelos nos conjuntos de dados não estacionários.

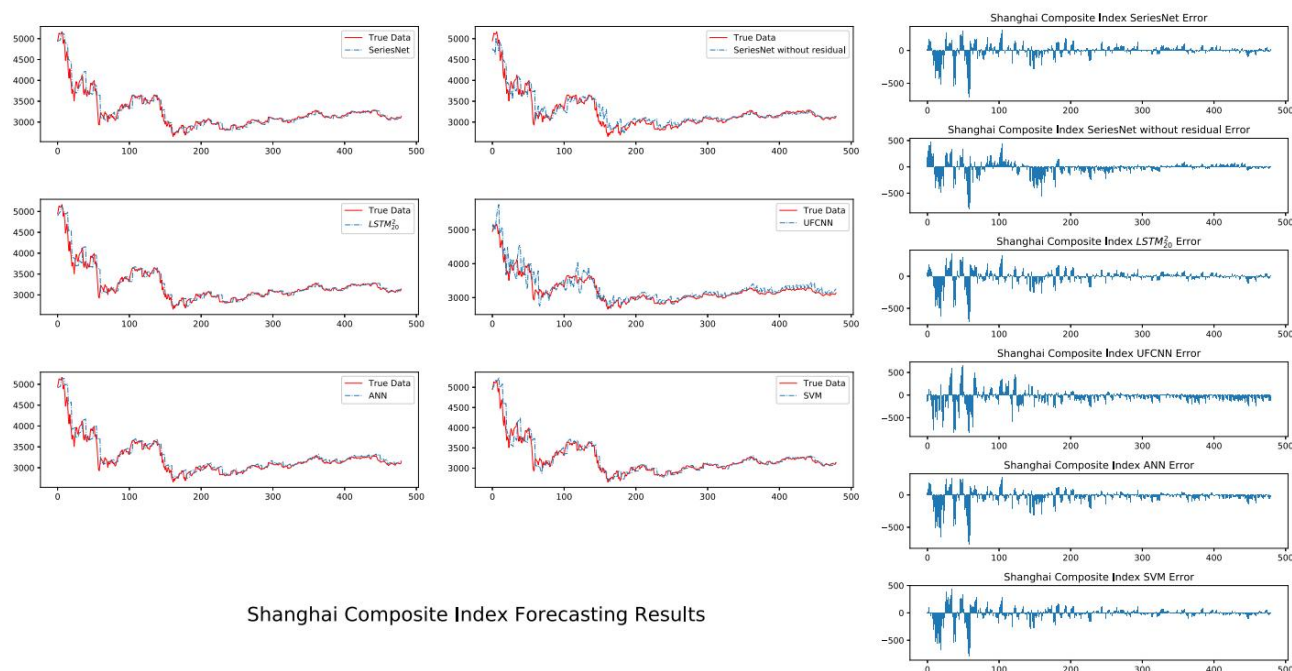
A fim de demonstrar ainda mais os resultados experimentais, usamos as figuras de previsão e erro para ilustrar. A Fig. 6 compara





S&amp;P500 Index Forecasting Results

Fig. 6. Comparação dos valores de previsão e valores reais para o índice S&amp;P 500.



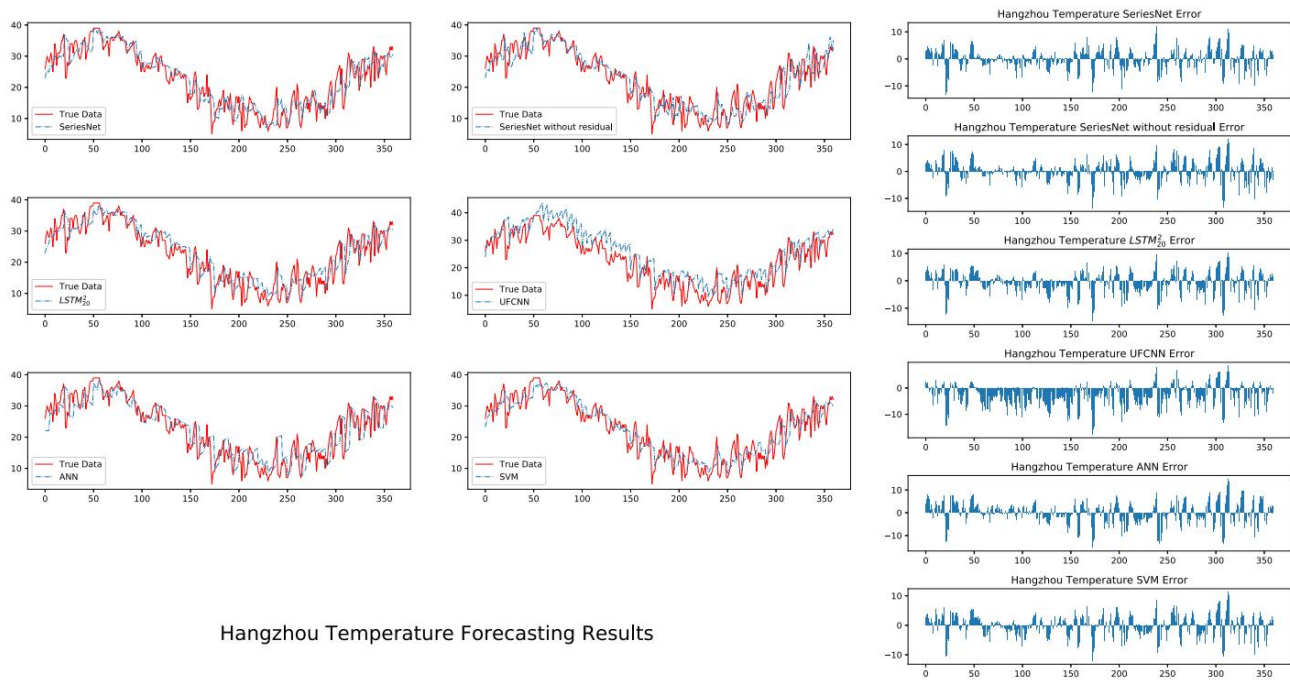
Shanghai Composite Index Forecasting Results

Fig. 7. Comparação dos valores de previsão e valores reais para o Shanghai Composite Index.

os valores de previsão do SeriesNet, o SeriesNet sem resíduo, o LSTM, o UFCNN, o ANN e o SVM com o verdadeiro valor no conjunto de dados do índice S&P 500. Na figura da esquerda, o eixo x indica a unidade de tempo e o eixo y é o valor de correlação. Esse A figura tem sete curvas de previsão e valores verdadeiros. A figura de erro é obtida calculando os valores verdadeiros e a previsão valor na figura à esquerda. E na figura do erro, o eixo x indica a unidade de tempo e o eixo y indica o erro entre o valor de previsão e o valor real. De acordo com a previsão e erro acima figura, pode-se ver que o erro do modelo SeriesNet é menor e o erro é relativamente estável.

Além disso, a Fig. 7 compara os valores de previsão do SeriesNet, do SeriesNet sem resíduo, do LSTM, do UFCNN, do ANN e o SVM com o valor real no conjunto de dados do Shanghai Composite Index. A Fig. 8 compara os valores de previsão do SeriesNet, do SeriesNet sem resíduo, do LSTM, do UFCNN, do ANN e o SVM com o valor real no conjunto de dados da temperatura diária de Hangzhou.

Além disso, dividimos cada conjunto de dados em 5 partes e calculamos o erro de previsão de cada parte. As Tabelas 5-7 mostram os erros de previsão de cada parte quando o tamanho da janela deslizante é definido como 5. Combinado com as Figs. 6-8 e Tabelas 5-7, pode-se ver que o



Hangzhou Temperature Forecasting Results

Fig. 8. Comparação dos valores de previsão e valores reais para a temperatura de Hangzhou.

Tabela 5

O resultado da avaliação do índice S&P 500 em diferentes horários de início.

Modelo de série temporal 1	2	3	4	5	
	RMSE MAE R2	RMSE MAE R2	RMSE MAE R2	RMSE MAE R2	
SeriesNet10	27,20 21,27 0,721	30,27 22,69 0,197	<b>41,27 30,77 0,738</b>	<b>27,95</b> 22,22 <b>0,793</b>	28,52 21,88 0,899
SeriesNet50	29,41 0,491 54,25	46,34 31,96 0,882 46,48 35,84 0,668	47,18 37,01 0,411	57,67 43,83 0,586	
SeriesNet100	36,74 48,12 49,62 39,63 -0,110 -1,157 <b>24,00 0,783</b>	69,36 51,65 0,260	42,06 35,17 0,532	34,99 29,04 0,848	
SeriesNetNoRes	<b>16,80</b> 33,67 24,81 0,007 35,33 <b>27,49 0,052 0,136</b> 38,4,49	48,66 37,19 0,636	28,62 <b>22,00</b> 0,783	43,89 35,31 0,760	
LSTM20	0,336 37,72 28,32 56,12 41,15 <b>52,07 0,545 38,92</b> ,31	46,40 35,13 0,669	31,15 22,07 0,743	26,44 19,77 0,913	
LSTM250	51,69 39,22 -0,483 0,589 42,22 <b>34,97 0,005 28,98</b> ,90 0,647 30,73 24,89 0,644		28,32 21,00 0,788	31,75 25,60 0,875	
LSTM2100	31,94 26,05 0,106 46,57 36,58 <b>70,65 0,061 4,173</b> 0,189 70,76 59,83 y3,388		30,69 22,98 0,751	<b>25,32 18,97 0,920</b>	
LSTM4100			36,06 29,57 0,65,38	111,59 95,87 y0,550	
UFCNN			-0,073 53,38 63,073	35,87 30,25 0,840	
ANN				67,34 57,49 0,435	
SVM	28,56 18,96 0,692	<b>28,74 22,02 0,276</b>	62,83 45,75 0,393	30,15 25,41 0,759	78,55 68,55 0,232

Tabela 6

O resultado da avaliação do Shanghai Composite Index em diferentes horários de início.

Modelo de série temporal 1	2	3	4	5	
RMSE MAE R2	RMSE MAE R2	RMSE MAE R2	RMSE MAE R2	RMSE MAE R2	
SeriesNet10	233,79 169,81 0,852	112,51 84,35 0,897	<b>60,81 45,37 0,367</b>	37,65 27,99 0,747	37,55 28,37 0,687
SeriesNet50	<b>224,48 162,38 0,864</b>	104,82 76,56 0,910	63,77 45,81 0,304	38,52 28,85 0,736	32,19 23,58 0,770
SeriesNet100	234,98 174,15 0,850	<b>103,45 74,89 0,913</b>	61,23 45,80 0,358	50,74 39,11 0,541	36,80 27,25 0,699
SeriesNetNoRes	253,00 194,56 0,827	176,25 129,11 0,746	81,90 61,02 y0,148	37,72 28,70 0,747	35,66 29,89 0,718
LSTM20	237,66 172,44 0,847	110,42 82,36 0,900	61,29 <b>44,29</b> 0,357	<b>35,44 26,68 0,776</b>	<b>32,12 23,26 0,771</b>
LSTM250	228,75 175,60 0,858	107,71 79,60 0,905	71,91 54,28 0,115	53,62 41,42 0,488	44,92 34,75 0,552
LSTM2100	228,13 177,19 0,859	109,76 80,86 0,902	78,60 60,85 y0,057 60,30 47,06 0,352 66,32 47,65		50,35 39,35 0,437
LSTM4100	247,87 176,15 0,834	122,23 91,87 0,878	0,247	45,99 34,64 0,623	52,36 41,46 0,391
UFCNN	360,59 271,43 0,648	189,92 146,98 0,705 89,63 69,84 y0,375 76,43 59,59 -0,040 98,87 83,29 y1,172			
ANN	248,86 176,94 0,832	118,91 88,51 0,884	65,17 46,78 0,273	43,28 32,64 0,666	47,82 37,28 0,492
SVM	276,38 201,61 0,793	114,46 86,73 0,893	62,58 47,98 0,330	38,82 28,62 0,732	34,61 25,87 0,734

**Tabela 7**  
O resultado da avaliação da temperatura de Hangzhou em horário de início diferente.

Modelo de série temporal 1	2			3			4			5		
	RMSE	MAE	R2	RMSE	MAE	R2	RMSE	MAE	R2	RMSE	MAE	R2
SeriesNet10	3,02	0,000	2,39	3,95	1,82	0,720	4,00	3,26		4,50	3,69	0,048
SeriesNet50	<b>2,08</b>	<b>-0,025</b>	<b>1,59</b>	<b>0,788</b>	-0,027	4,00	3,21	2,51	0,153	4,40	3,54	0,004
SeriesNet100	1,91	0,691	3,73	2,88	0,107	2,18	1,62	0,768	4,01	0,269		
SeriesNetNoRes	3,67	2,87	0,138	2,83	2,24	<b>0,606</b>	<b>0,385</b>	<b>2,385</b>	4,40	3,44	0,004	3,96
LSTM <sub>20</sub>	0,649	3,51	<b>2,67</b>	0,210	2,54	<b>2,85</b>	<b>0,169</b>	<b>2,360</b>	4,87	3,77	0,228	4,11
LSTM <sub>50</sub>	1,82	0,735							4,74	3,67	0,165	4,58
LSTM <sub>100</sub>									4,60	3,57	0,098	4,34
LSTM <sub>400</sub>									4,64	3,57	0,113	4,34
UFCNN	3,51	2,88	0,708	4,71	4,16	0,161			6,79	5,32	0,087	
ANN	4,49	3,50	0,289	2,33	1,84	0,734	<b>3,41</b>	2,69	5,38	4,21	0,501	
SVM	<b>0,256</b>	2,45	1,87	0,707					<b>4,03</b>	<b>3,21</b>	<b>0,158</b>	
									<b>3,69</b>	2,97	<b>0,057</b>	
										4,68	3,68	0,356

SeriesNet é mais adaptável do que outros modelos e pode alcançar melhores precisões de previsão, pois as flutuações do conjunto de dados são dramáticas. Como as mudanças de dados não são óbvias, outros modelos também chances de obter melhores precisões de previsão do que nossa SeriesNet.

6. Conclusão

A fim de melhorar a precisão da previsão de séries temporais, este paper projeta um novo modelo de previsão de séries temporais chamado Series Net. Este modelo pode aprender completamente os recursos de dados de séries temporais em diferentes comprimentos de intervalo. Ele usa o LSTM para aprender recursos holísticos e reduzir a dimensionalidade de dados multicondicionais. Além disso, ele usa as convoluções causais dilatadas para aprender diferentes tempos característicos do intervalo. Além disso, adota o aprendizado residual e a normalização em lote para melhorar a generalização. Este modelo pode aprenda recursos multi-intervalo e multinível de dados de séries temporais. Os resultados experimentais mostram que este modelo tem uma previsão muito maior precisão e estabilidade muito maior.

Conflito de interesses

Nenhum.

Agradecimentos

Os autores agradecem o apoio do Zhejiang projeto especial especial de ciência e tecnologia (nº 2018C01064); Projetos Públicos da Província de Zhejiang (nº 2017C31014).

Referências

[1] C. Cheng, A. Sa-Ngasongsong, O. Beyca, T. Le, H. Yang, Z. Kong, ST Bukkap atam, Previsão de séries temporais para processos não lineares e não estacionários: a revisão e estudo comparativo, IIE Trans. 47 (10) (2015) 1053-1071.

[2] H. Akaike, Ajustando modelos autorregressivos para previsão, Ann. Inst. Estado. Matemática. 21 (1) (1969) 243-247.

[3] CM Hurvich, C.ŷL. Tsai, Regressão e seleção de modelos de séries temporais em pequenas amostras, Biometrics 76 (2) (1989) 297-307.

[4] GE Box, DA Pierce, Distribuição de autocorrelações residuais em modelos de séries temporais de média móvel autoregressiva integrada, J. Am. Estado. Associação 65 (332) (1970) 1509-1526.

[5] B. Williams, P. Durvasula, D. Brown, Previsão de fluxo de tráfego em rodovias urbanas: aplicação de modelos de média móvel integrada autoregressiva sazonal e suavização exponencial, Transp. Res. Gravando. J. Transp. Res. Conselho 1644 (1998) 132-141.

[6] L.-J. Cao, FEH Tay, Máquina vetorial de suporte com parâmetros adaptativos na previsão de séries temporais financeiras, IEEE Trans. Rede Neural. 14 (6) (2003) 1506-1518.

[7] K.-R. Müller, AJ Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, V. Vap nik, Previsão de séries temporais com máquinas de vetor de suporte, em: Proceedings of a Conferência Internacional sobre Redes Neurais Artificiais, Springer, 1997, págs. 999–1004.

[8] GP Zhang, V. Berardi, Previsão de séries temporais com conjuntos de redes neurais: um pedido de previsão da taxa de câmbio, J. Oper. Res. Soc. 52 (6) (2001) 652-664.

[9] MM Noel, BJ Pandian, Controle de um sistema de nível de líquido não linear usando um novo abordagem de aprendizado por reforço baseado em rede neural artificial, Appl. Suave Computar. 23 (2014) 444-451.

[10] Y. Chen, B. Yang, J. Dong, Previsão de séries temporais usando uma wavelet linear local rede neural, Neurocomputing 69 (4-6) (2006) 449-465.

[11] GP Zhang, Previsão de séries temporais usando um arima híbrido e rede neural modelo, Neurocomputing 50 (2003) 159-175.

[12] A. Jain, AM Kumar, modelos de rede neural híbrida para séries temporais hidrológicas previsão, Ap. Computação Suave. 7 (2) (2007) 585-592.

[13] CH Aladag, E. Egrioglu, C. Kadilar, Previsão de séries temporais não lineares com um metodologia híbrida, Appl. Matemática. Lett. 22 (9) (2009) 1467-1470.

[14] LP Maguire, B. Roche, TM McGinnity, L. McDaid, Prevendo um tempo caótico série usando uma rede neural difusa, Inf. Sci. 112 (1-4) (1998) 125-136.

[15] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436.

[16] J. Schmidhuber, Deep learning em redes neurais: uma visão geral, Neural Netw. 61 (2015) 85-117.

[17] A. Krizhevsky, I. Sutskever, GE Hinton, classificação Imagenet com redes neurais convolucionais profundas, em: Proceedings of the Advances in Neural Information Processsing Systems, 2012, pp. 1097–1105.

[18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Van houecke, A. Rabinovich, Aprofundando com convoluções, em: Proceedings of the Conferência IEEE sobre Visão Computacional e Reconhecimento de Padrões, 2015, pp. 1–9.

[19] R. Girshick, J. Donahue, T. Darrell, J. Malik, Hierarquias de recursos ricos para detecção precisa de objetos e segmentação semântica, em: Proceedings of the Conferência IEEE sobre Visão Computacional e Reconhecimento de Padrões, 2014, pp. 580–587.

[20] X. Lu, B. Wang, X. Zheng, X. Li, Explorando modelos e dados para geração de legendas de imagens de sensoriamento remoto, IEEE Trans. Geosci. Sensor Remoto 56 (4) (2018) 2183-2195.

[21] X. Lu, X. Zheng, Y. Yuan, classificação de cena de sensoriamento remoto por aprendizado de representação não supervisionado, IEEE Trans. Geosci. Sensor Remoto 55 (9) (2017) 5148-5157.

[22] A.-r. Mohamed, GE Dahl, G. Hinton, modelagem acústica usando redes de crenças profundas, IEEE Trans. Áudio, Fala, Lang. Processo. 20 (1) (2012) 14–22.

[23] GE Dahl, D. Yu, L. Deng, A. Acero, neuronal profundo pré-treinado dependente de contexto redes para reconhecimento de fala de grande vocabulário, IEEE Trans. Áudio, Fala, Lang. Processo. 20 (1) (2012) 30–42.

[24] I. Sutskever, O. Vinyals, QV Le, Sequência para sequenciar aprendizagem com neural redes, em: Proceedings of the Advances in Neural Information Processing Sistemas, 2014, pp. 3104-3112.

[25] Y. Bengio, R. Ducharme, P. Vincent, C. Jauvin, Uma linguagem neural probabilística modelo, J. Mach. Aprender. Res. 3 (2003) 1137-1155. fevereiro

[26] T. Mikolov, I. Sutskever, K. Chen, GS Corrado, J. Dean, Distributed representações de palavras e frases e sua composicionalidade, em: Proceedings of the Avanços em Sistemas de Processamento de Informação Neural, 2013, pp. 3111–3119.

[27] GEP Box, GM Jenkins, Análise de séries temporais: previsão e controle, J. Am. Estado. Associação 68 (342) (1970) 199-201.

[28] D. Peter, P. Silvia, Arima vs. Arimax – qual abordagem é melhor analisar e previsão de séries temporais macroeconômicas, em: Proceedings of the 30th International Conferência Métodos Matemáticos em Economia, 2012, pp. 136–140. Karviná, República Checa

[29] U. Thissen, R. Van Brakel, A. De Weijer, W. Melszen, L. Buydens, Usando suporte máquinas vetoriais para previsão de séries temporais, Chemom. Intel. Laboratório. Sistema 69 (1) (2003) 35-49.

[30] B. Gui, X. Wei, Q. Shen, J. Qi, L. Guo, Previsão de séries temporais financeiras usando máquina de vetores de suporte, em: Anais da Décima Conferência Internacional sobre Inteligência e Segurança Computacional (CIS), IEEE, 2014, pp. 39–43.

[31] JA Suykens, T. Van Gestel, J. De Brabanter, Least Squares Support Vector Machines, World Scientific, 2002.

[32] H. Wang, D. Hu, Comparação de svm e ls-svm para regressão, em: Proceedings da Conferência Internacional sobre Redes Neurais e Cérebro, ICNN&B'05., 1, IEEE, 2005, pp. 279-283.

[33] K. Hornik, M. Stinchcombe, H. White, redes feedforward multicamadas são aproximadores universais, Rede Neural. 2 (5) (1989) 359-366.

[34] T. Chow, C.-T. Leung, Rede neural integrada autorregressiva não linear modelo para previsão de carga de curto prazo, IEE Proc. Gerador. Transm. Distribuir 143 (5) (1996) 500-506.

[35] JT Connor, RD Martin, LE Atlas, Redes neurais recorrentes e tempo robusto previsão de série, IEEE Trans. Rede Neural. 5 (2) (1994) 240-254.

- [36] FA Gers, D. Eck, J. Schmidhuber, Aplicando LSTM a séries temporais previsíveis através de abordagens de janela de tempo, em: *Proceedings of the Neural Nets WIRN Vietri-01*, Springer, 2002, pp. 193–200.
- [37] R. Mittelman, Modelagem de séries temporais com neu totalmente convolucional não dizimado redes ral, *arXiv:1508.00317* (2015) 1–9.
- [38] S. Dasgupta, T. Osogami, Máquinas de Boltzmann dinâmicas não lineares para previsão de séries temporais, em: *Proceedings of the AAAI*, 2017, pp. 1833–1839.
- [39] S. Naduvil-Vadukootu, RA Angryk, P. Riley, Avaliando estratégias de pré-processamento para previsão de séries temporais usando arquiteturas de aprendizado profundo., em: *Proceedings of a Conferência FLAIRS*, 2017, pp. 520–525.
- [40] N. Hamid, M. Noorani, Modelagem de sistema de previsão: uma aplicação da abordagem do vizinho mais próximo para dados caóticos, *Appl. Matemática. Computar. Intel.* 2 (1) (2013) 137–148.
- [41] CE Rasmussen, processos gaussianos em aprendizado de máquina, em: *Palestras avançadas sobre aprendizado de máquina*, Springer, 2004, pp. 63–71.
- [42] MG Frei, I. Osorio, Decomposição em escala de tempo intrínseca: tempo-frequência-energia análise e filtragem em tempo real de sinais não estacionários, *Proc. R. Soc. Londres. UMA: Matemática. Física Eng. Sci.* 463 (2007) 321–342. A Sociedade Real
- [43] Y. Bengio, P. Simard, P. Frasconi, Aprender dependências de longo prazo com descida de gradiente é difícil, *IEEE Trans. Rede Neural.* 5 (2) (1994) 157–166.
- [44] S. Hochreiter, J. Schmidhuber, Long curto prazo memory, *Neural Comput.* 9 (8) (1997) 1735–1780.
- [45] K. Cho, B. van Merriënboer, Ç. Gülçehre, F. Bougares, H. Schwenk, Y. Bengio, Aprendendo representações de frases usando codificador-decodificador RNN para tradução de máquina estatística, *arXiv:1406.1078* (2014) 1–15.
- [46] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Redes neurais recorrentes de feedback fechado, em: *Proceedings of the International Conference on Machine Learning*, 2015, págs. 2067–2075.
- [47] F. Yu, V. Koltun, agregação de contexto multi-escala por convoluções dilatadas, *arXiv:1511.07122* (2015) 1–13.
- [48] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, AW Sênior, K. Kavukcuoglu, Wavenet: um modelo generativo para áudio bruto, *arXiv: 1609.03499* (2016) 1–15.
- [49] X. Mao, C. Shen, Y.-B. Yang, Restauração de imagem usando redes de codificador-decodificador convolucional muito profundas com conexões de salto simétricas, em: *Proceedings of the Advances in Neural Information Processing Systems*, 2016, págs. 2802–2810.
- [50] K. He, X. Zhang, S. Ren, J. Sun, mapeamentos de identidade em redes residuais profundas, em: *Anais da Conferência Europeia sobre Visão Computacional*, Springer, 2016, págs. 630–645.
- [51] C. Laurent, G. Pereyra, P. Brakel, Y. Zhang, Y. Bengio, Redes neurais recorrentes normalizadas em lote, em: *Proceedings of the IEEE International Conference on Acústica, Fala e Processamento de Sinais (ICASSP)*, IEEE, 2016, pp. 2657–2661.
- [52] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the Computer Vision and Pattern Recognition*, 2016, págs. 770–778.
- [53] A. van den Oord, N. Kalchbrenner, L. Espeholt, O. Vinyals, A. Graves, et al., geração de imagem condicional com decodificadores pixelCNN, em: *Proceedings of the Advances in Neural Information Processing Systems*, 2016, págs. 4790–4798.



**Zhipeng Shen** está atualmente cursando o mestrado da Faculdade de Ciência da Computação e Tecnologia da Universidade de Tecnologia de Zhe Jiang, China. Ele recebeu o grau de BE pela Fujian Normal University, China. Sua pesquisa interesses incluem aprendizado profundo e aprendizado de máquina.



**Yuanming Zhang** recebeu o Ph.D. grau do Departamento de Ciência da Informação da Universidade de Utsunomiya, Japão, em 2010. Atualmente é Professor Associado em a Faculdade de Ciência da Computação e Tecnologia da Universidade de Tecnologia de Zhe Jiang China. Seus interesses de pesquisa incluem análise de big data, computação paralela. Ele é um membro do CCF.



**Jiawei Lu** é professor na faculdade de Ciência da Computação e Tecnologia da Universidade de Tecnologia de Zhejiang China. Seus atuais interesses de pesquisa incluem big data e arquitetura de software. É membro do CCF.



**Xu Jun** é um Experimentalista Sênior na Universidade de Zhejiang de Tecnologia China. Seus interesses de pesquisa incluem serviço de software, tecnologia de interação. Ele é membro de CCF.



**Gang Xiao** é professor na Faculdade de Ciência da Computação e Tecnologia da Universidade de Tecnologia de Zhejiang China. Seus interesses de pesquisa incluem componentes de software, família de produtos de software e sistema de informação inteligente. É membro do CCF.