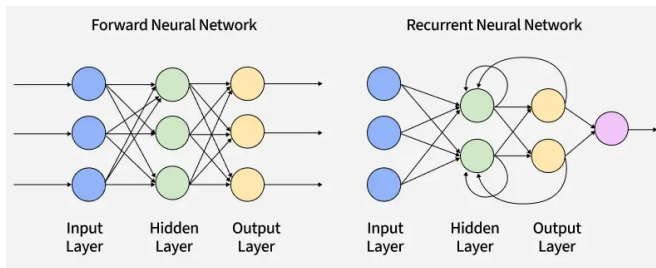


Redes Neurais Recorrentes

Cesar de Oliveira F. Silva

Setembro / 2026

Redes Neurais Recorrentes



Ideia central

RNNs são arquiteturas projetadas para dados sequenciais. Elas processam a entrada ao longo do tempo, mantendo um **estado interno** que resume a informação observada até o instante atual.

Modelagem de séries temporais ambientais — visão geral

- Objetivo: prever variáveis ambientais como **vazão, nível, qualidade da água, nutrientes e temperatura**.
- A principal dificuldade está em modelar:
 - dependências temporais;
 - sazonalidade e tendências;
 - eventos extremos e não linearidades.
- Problemas típicos:
 - previsão de um passo à frente;
 - previsão multi-step;
 - previsão probabilística;
 - detecção de anomalias e eventos raros.

O que é uma RNN?

- Uma **Recurrent Neural Network** modela sequências processando observações uma a uma.
- Em cada instante t , a rede recebe a entrada atual x_t e combina essa informação com o estado anterior h_{t-1} .

Notação básica

$$h_t = f(h_{t-1}, x_t; \theta), \quad y_t = g(h_t)$$

- x_t : entrada no tempo t ;
- h_t : estado oculto;
- y_t : saída da rede.

Intuição

O estado oculto funciona como uma memória compacta da sequência passada.

RNNs ao longo do tempo

- Uma RNN pode ser vista como a mesma célula reaplicada em vários instantes temporais.
- Os parâmetros são compartilhados ao longo do tempo.
- Isso permite modelar dependências temporais sem precisar de um conjunto distinto de pesos para cada instante.

Propriedade importante

Compartilhamento temporal de parâmetros:

$$\theta_1 = \theta_2 = \dots = \theta_T$$

- A rede aprende padrões como persistência, atraso de resposta, ciclos e relações entre variáveis exógenas e a série alvo.

Limitações das RNNs simples

- RNNs básicas têm dificuldade em capturar dependências de longo prazo.
- O principal problema ocorre durante o treinamento por backpropagation through time.

Problemas clássicos

- **Vanishing gradients:** o gradiente tende a zero.
- **Exploding gradients:** o gradiente cresce excessivamente.
- Isso dificulta aprender relações entre eventos distantes no tempo, como:
 - efeitos acumulados de chuva;
 - respostas lentas da bacia;
 - padrões sazonais de longo alcance.

Explosão e desvanecimento do gradiente

- Durante o treinamento, o gradiente é propagado ao longo de muitos passos temporais.
- Repetidas multiplicações podem fazer a informação:
 - desaparecer (*vanishing*);
 - crescer de forma instável (*exploding*).

Mitigações comuns

- LSTM e GRU;
- inicialização cuidadosa;
- normalização;
- *gradient clipping*;
- otimizadores adaptativos, como Adam.

LSTM: memória com gates

- A **Long Short-Term Memory** introduz uma célula de memória explícita.
- O fluxo de informação é controlado por portas (*gates*):
 - **forget gate**: descarta informação irrelevante;
 - **input gate**: incorpora nova informação;
 - **output gate**: controla o que é exposto como estado oculto.

Equações simplificadas

$$\begin{aligned}f_t &= \sigma(W_f[x_t, h_{t-1}] + b_f), & i_t &= \sigma(W_i[x_t, h_{t-1}] + b_i) \\ \tilde{c}_t &= \tanh(W_c[x_t, h_{t-1}] + b_c), & c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ h_t &= o_t \odot \tanh(c_t)\end{aligned}$$

GRU: simplificação eficiente da LSTM

- A **Gated Recurrent Unit** simplifica a LSTM.
- Em vez de três portas principais, usa dois mecanismos:
 - **update gate**;
 - **reset gate**.
- Frequentemente possui:
 - menos parâmetros;
 - treinamento mais rápido;
 - desempenho competitivo com LSTM.

Regra prática

- **LSTM**: maior flexibilidade de memória;
- **GRU**: arquitetura mais enxuta e eficiente.

RNN, LSTM e GRU — comparação conceitual

- **RNN simples:**

- estrutura mais básica;
- menor custo;
- dificuldade com dependências longas.

- **LSTM:**

- memória explícita;
- melhor retenção temporal;
- maior número de parâmetros.

- **GRU:**

- compromisso entre simplicidade e desempenho;
- útil quando há restrição computacional.

Estratégias de previsão multi-step

- **Recursive / autoregressiva:**
 - prevê um passo de cada vez;
 - usa previsões anteriores como entrada;
 - simples, mas acumula erro.
- **Direct:**
 - um modelo por horizonte;
 - evita propagação de erro;
 - custo computacional maior.
- **Multi-output / seq2seq:**
 - prevê vários passos em um único passe;
 - equilíbrio entre flexibilidade e estabilidade.

Treinamento seq2seq: teacher forcing e scheduled sampling

- Em modelos multi-step, a forma de alimentar a sequência futura afeta o treinamento.

Teacher forcing

Durante o treino, a rede recebe o valor real anterior como entrada para o próximo passo.

- Vantagem: acelera a convergência.
- Limitação: gera **exposure bias**, pois na inferência a rede não terá acesso à verdade futura.

Scheduled sampling

Mistura gradualmente valores reais e previsões do próprio modelo durante o treinamento.

Arquiteturas modernas e alternativas

- **Seq2Seq encoder-decoder:** útil para previsão multi-step.
- **Attention:** permite focar em janelas temporais mais relevantes.
- **Transformers temporais:** capturam dependências longas por self-attention.
- **Temporal Convolutional Networks (TCN):**
 - convoluções causais e dilatadas;
 - receptive field controlado;
 - alternativa eficiente às RNNs.
- **Modelos híbridos:** CNN temporal + RNN/Transformer.

Modelagem multivariada e variáveis exógenas

- Em problemas ambientais, a série alvo raramente é suficiente sozinha.
- É comum incorporar covariáveis exógenas como:
 - precipitação;
 - temperatura;
 - uso do solo;
 - operação de barragens;
 - sazonalidade e calendário.
- Estratégias de entrada:
 - concatenação direta das features em cada tempo t ;
 - encoders separados por fonte/sensor;
 - fusão tardia entre representações.

Pré-processamento e qualidade dos dados

- Tratamento de *missing values*:
 - interpolação;
 - imputação baseada em modelo;
 - máscaras explícitas de ausência.
- Tratamento de outliers e inconsistências.
- Harmonização entre sensores, estações e frequências temporais.
- Construção de janelas temporais (*sliding windows*).
- Engenharia de atributos:
 - lags;
 - médias móveis;
 - diferenças;
 - agregações sazonais.

Formatos de entrada e saída em RNNs

- **Many-to-one:** sequência de entrada, uma saída final.
 - Ex.: classificar regime hidrológico de uma janela temporal.
- **One-to-many:** uma entrada inicial, sequência de saídas.
 - Ex.: gerar múltiplos horizontes a partir do estado atual.
- **Many-to-many:** sequência de entrada, sequência de saída.
 - Ex.: previsão multi-step e modelagem seq2seq.

Funções de perda e objetivos

- **Losses pontuais:**

- MSE;
- RMSE;
- MAE.

- **Previsão probabilística:**

- NLL;
- perdas quantílicas (*pinball loss*);
- CRPS;
- Mixture Density Networks.

- **Multi-horizon forecasting:**

- perdas ponderadas por lead time;
- combinação de objetivos por horizonte.

Escolha prática

A função de perda deve refletir o objetivo operacional: erro médio, robustez a extremos ou qualidade da incerteza.

Validação temporal adequada

- Em séries temporais, **random split** normalmente é inadequado.
- A divisão deve respeitar a ordem temporal dos dados.

Estratégias recomendadas

- **Holdout temporal;**
 - **rolling validation;**
 - **walk-forward validation.**
-
- Essas estratégias simulam o cenário operacional real.
 - Também é recomendável reportar desempenho:
 - por horizonte;
 - por estação do ano;
 - em eventos extremos.

Métricas para hidrologia e qualidade da água

- **RMSE / MAE:** magnitude média do erro.
- **NSE (Nash–Sutcliffe Efficiency):** compara o modelo à média observada.
- **KGE (Kling–Gupta Efficiency):** incorpora correlação, viés e variabilidade.
- **Skill scores:** comparação com baselines como persistência ou climatologia.

Boa prática

Avaliar não apenas erro médio, mas também:

- capacidade de antecipar picos;
- robustez em estiagens e enchentes;
- calibração de intervalos de previsão.

Comparação com modelos estatísticos

- Modelos como **ARIMA**, **SARIMA**, **state-space** e **Prophet** são fortes quando a estrutura temporal é relativamente estável e interpretável.
- Vantagens dos métodos estatísticos:
 - maior interpretabilidade;
 - menor exigência de dados;
 - incerteza frequentemente mais estruturada.
- Vantagens de DL:
 - captura não linearidades;
 - integra múltiplas entradas;
 - representa interações complexas entre variáveis.

Comparação com modelos físicos (hidrológicos)

- Modelos físicos como **HBV, SWAT e HEC-HMS** representam processos hidrológicos explicitamente.
- Têm parâmetros interpretáveis e permitem simular cenários de manejo.
- Modelos de ML/DL:
 - podem ter forte desempenho preditivo;
 - são especialmente úteis em previsão de curto e médio prazo;
 - não substituem a modelagem física.

Integração promissora

- correção de viés de modelos físicos;
- aprendizagem de resíduos;
- híbridos guiados por estrutura física.

Interpretação, confiabilidade e responsabilidade

- Interpretação temporal:
 - importância de lags;
 - pesos de atenção;
 - sensibilidade a covariáveis.
- Verificar se o modelo aprende relações fisicamente plausíveis.
- Identificar:
 - correlações espúrias;
 - mudança de regime;
 - *concept drift*.
- Em aplicações operacionais, é essencial quantificar incerteza e limites de uso.

Regularização e técnicas práticas

- **Regularização:**
 - dropout;
 - variational dropout para RNNs;
 - weight decay;
 - early stopping;
 - ensembles.
- **Data augmentation temporal:**
 - jittering;
 - time warping;
 - bootstrapped windows.
- **Tuning:**
 - tamanho da janela;
 - dimensão oculta;
 - número de camadas;
 - horizonte de previsão;
 - conjunto de covariáveis.

Implantação e manutenção operacional

- Inferência em produção pode ocorrer em:
 - streaming;
 - bateladas periódicas;
 - pipelines de atualização contínua.
- Questões operacionais importantes:
 - latência por horizonte;
 - atualização incremental;
 - retreinamento programado;
 - monitoramento de drift.
- Em previsão probabilística, recalibração pode ser necessária ao longo do tempo.

Casos de uso: previsão hidrológica e qualidade da água

- **Previsão de vazão:**

- precipitação observada e prevista;
- lags de vazão;
- umidade antecedente;
- variáveis fisiográficas e operacionais.

- **Qualidade da água:**

- turbidez;
- nutrientes;
- oxigênio dissolvido;
- resposta a eventos de escoamento e descargas.

- **Critérios de sucesso:**

- antecipação de picos;
- estabilidade sazonal;
- robustez em extremos.

Resumo prático

- 1 Escolha a arquitetura conforme horizonte, volume de dados e complexidade temporal.
- 2 Respeite a cronologia: use validação temporal.
- 3 Trate faltantes, outliers e normalize por variável.
- 4 Inclua covariáveis exógenas quando houver base física para isso.
- 5 Compare sempre com baselines estatísticos e físicos.
- 6 Reporte incerteza, skill por horizonte e desempenho em eventos extremos.