

UNIVERSIDADE DE PERNAMBUCO

ESCOLA POLITÉCNICA DE PERNAMBUCO

LINGUAGEM DE PROGRAMAÇÃO APLICADA A AUTOMAÇÃO-LPAA

SISTEMA HÍBRIDO DE COMBINAÇÕES PARA PREVISÃO
DE SÉRIES TEMPORAIS

Estudantes: Guilherme da Silva Barros
Professor: João Fausto Lorenzato de Oliveira
Turma: LP

Recife,
02 de julho de 2019

Sumário

1 - Resumo	3
2 - Introdução	4
3 - Método proposto	5
3.1 - Funcionamento dos modelos	5
3.1.1 - AR – Auto-regresivo	5
3.1.2 - MLP - Multi-layer Perceptron	5
3.1.3 - RF – Random Forest	5
3.2 – Analise de desempenho	6
4 – Experimento adicionais	7
4.1 – Analise de desempenho	7
Conclusão	8
Referencias	9

1. Resumo

No modelo híbrido proposto é de pegar previsões do modelo auto-regressivo (AR) e previsões do modelo MLP (RNA) e aplicar duas fases de previsões num novo modelo RF (RNA). A previsão de séries temporais é uma tarefa importante no campo do aprendizado de máquina, possuindo diversas aplicações em mercado de ações, hidrologia, meteorologia, entre outros. A análise da dependência existente nas observações adjacentes da série é necessária para que seja possível prever valores futuros com alguma precisão. Modelos dinâmicos são utilizados para realizar mapeamentos de uma série temporal, se aproximando do mecanismo gerador da série e sendo capazes de realizar previsões. No entanto, o mecanismo gerador de uma série temporal pode produzir padrões lineares e não-lineares que precisam ser devidamente mapeados. Modelos lineares como o auto-regressivo (AR) são capazes de mapear padrões lineares, porém não são indicados quando existem padrões não-lineares na série. Já os modelos não-lineares como as redes neurais artificiais (RNA) mapeiam padrões não-lineares, mas podem apresentar desempenho reduzido na presença de padrões lineares em relação aos modelos lineares. Fatores como a definição do número de elementos de entrada da RNA, número de amostras de treinamento podem afetar o desempenho. O modelo proposto possui potencial para melhorias, aplicando métodos que realizem mapeamento de tendência

2. Introdução

De maneira clara e objetiva, o modelo híbrido apresentado pega suas previsões do MLP e AR em seguida foi feita uma mediana com os parâmetros de previsão validados e treinados logo após no modelo RF essa mediana gerada pelas previsões foi feita um novo processamento dos dados da mediana passando pelo grid do Random Forest. O banco de dados utilizado foi o Saldo das operações de crédito por atividade econômica. A previsão de séries temporais consiste em analisar dados e realizar uma projeção para o futuro.

3. Método proposto

O método proposto é dividido em quatro fases: a normalização onde organiza o dataset para melhor processamento dos dados, a realização da primeira previsão dos modelos auto-regressivo e MLP(RNA), Em seguida é feita uma mesclagem das previsões utilizando o operador mediana e essa previsão mediana gerada é inserida no processamento de dados do Random Forest para gerar uma previsão final.

3.1 - Funcionamento dos modelos

3.1.1 - AR – Auto-regressivo

É uma série temporal do tipo: $X_t = \phi X_{t-1} + \varepsilon_t$ não linear. Onde o ε_t é o ruído e o ϕ suas constantes.

3.1.2 - MLP - Multi-layer Perceptron.

Este modelo otimiza a perda quadrática usando LBFGS ou descendente de gradiente estocástico.

Os parâmetros utilizados: de ativação, solucionadores e α .

Na ativação, é realizado o ativamente para a camada oculta da rede neural a implementa do código foi a 'relu' (a função de unidade linear retifica retornando o max entre 0 e x) e 'tanh' (a função bronzeada hiperbólica, retorna o tanh de x).

O solucionador é para otimização de peso, o utilizado foi na implementação do código foi 'lbfgs' ele é um otimizador da família dos métodos quase Newton.

3.1.3 - RF – Random Forest.

Este modelo é uma meta estimador que se ajusta a vários classificadores de árvore de decisão em várias sub amostras do conjunto de dados e usa a média para melhorar a precisão preditiva e controlar o ajuste excessivo. O tamanho da sub amostra é sempre igual ao tamanho da amostra de entrega original, mas as amostras são desenhadas com substituições. Algo notável é o seu alto custo de CPU visto que esse modelo surgiu por causa das novas tecnologias de potência de hardware.

Os parâmetros utilizados: número de arvores, características, folhas e as divisões.

No número de arvores, são feitas a escolha de quanto arvores vai gerar para rodar o determinado tipo de características da arvore.

Nas características, a profundidade máxima da árvore. Os utilizados foi o 'auto', 'sqrt', 'log2' e none

Nas folhas é o número mínimo de amostras necessárias para dividir um nó interno

Na divisão é o número mínimo de amostrar necessárias para estar em um nó da folha.

3.5 – Análise de desempenho

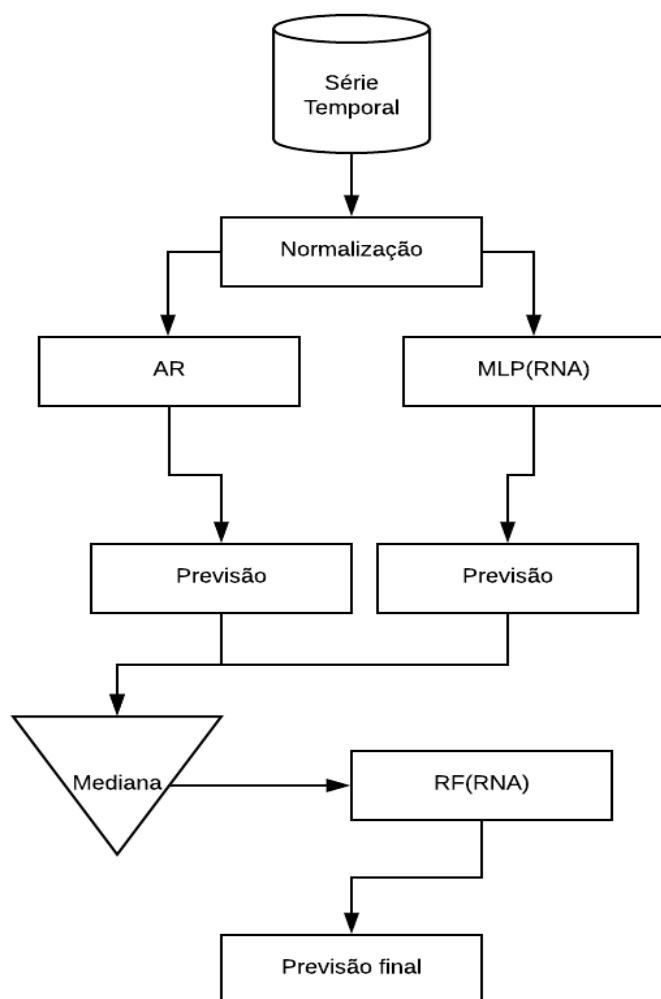


Figura 1 – Método Proposto, Fonte: do autor.

Tabela de erro	Crédito	Downjones
MLP	22008,15	571
AR	21919,77	543,85
Mediana	21919,7	555,8738
RF-Híbrido	356751,37	1288,2563

Tabela 1 – Tabela de erro

Fonte: do autor

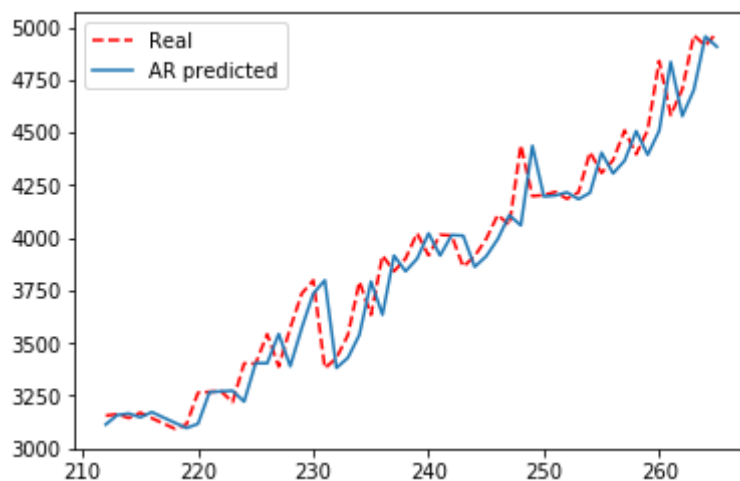


Figura 2 – Gráfico AR – Dataset: Credito.

Fonte: do autor

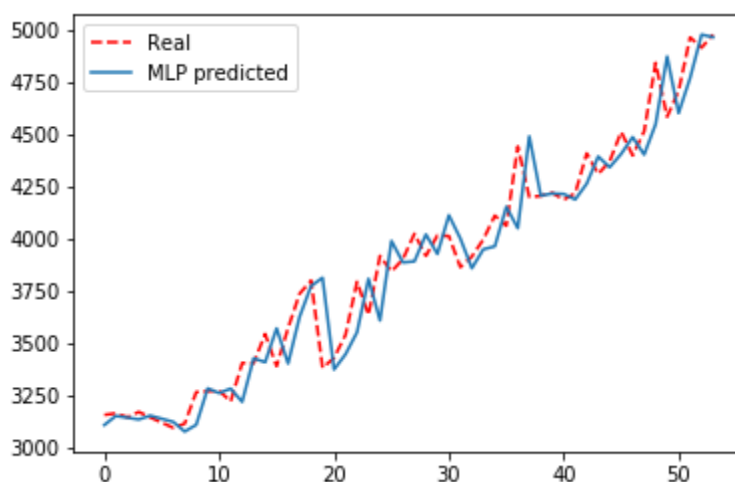


Figura 3 – Gráfico MLP – Dataset: Credito.

Fonte: do autor

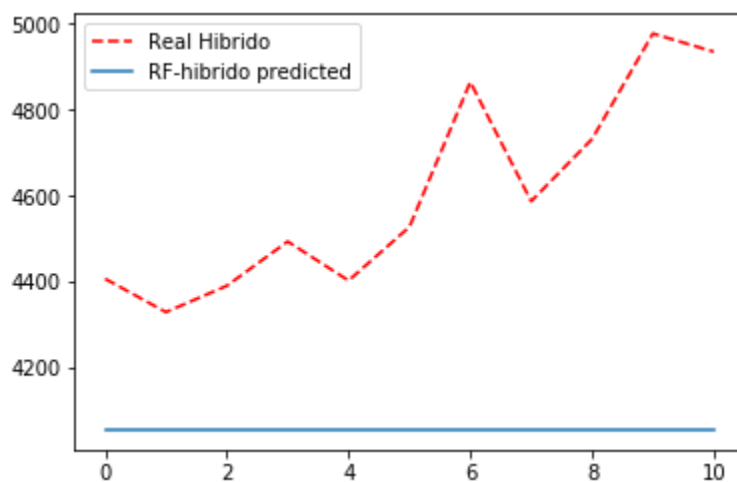


Figura 4 – Gráfico Híbrido – RF – Dataset: Credito

Fonte: do autor

4. Experimentos Adicionais

Os experimentos adicionais foram realizados com propósito de analisar outro tipo de dataset utilizando o mesmo modelo híbrido. Como visto pelo fato do RF(RNA) trabalhar com várias camadas de arvores de busca a série temporal de operações de crédito é pequena por utiliza dados mensais logo obteve um erro final maior visto que quando tirado a mediana e passado pelo um novo processamento de dados deixou ela menor ainda com isso não se obteve o melhor aproveitamento do grid do Random Forest

Logo, utilizando uma nova base de dados (Dow-Jones Dow-Jones Industrial Stock Price Index) com um maior número de dados se observa um ganho na utilização do híbrido visto que no final do processo de dados sobra mais previsões do processamento do RF e as arvores consegue ter uma melhor análise.

4.1 – Análise de desempenho

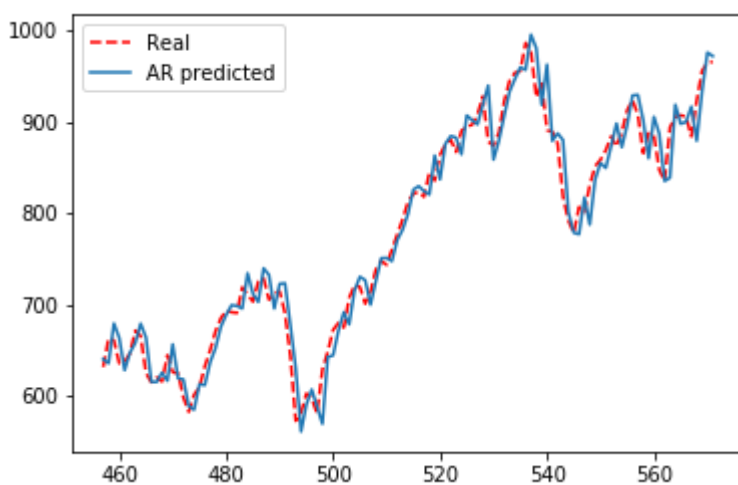


Figura 5 – Gráfico AR – Dataset: Downjones

Fonte: do autor

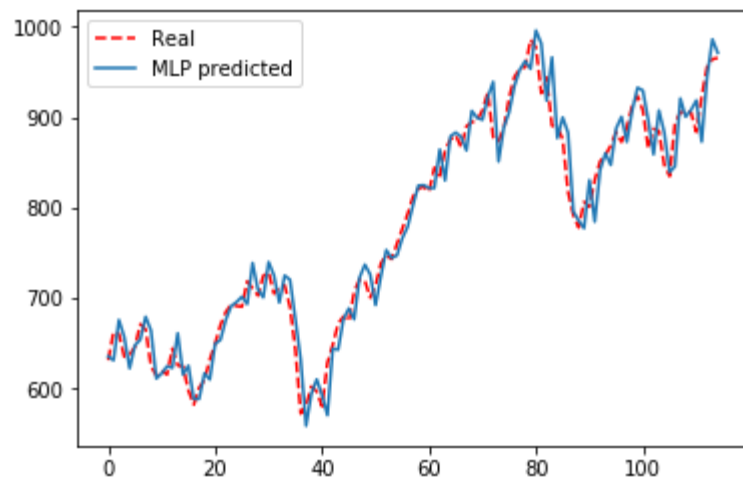


Figura 6 – Gráfico MLP – Dataset: Downjones.

Fonte: do autor

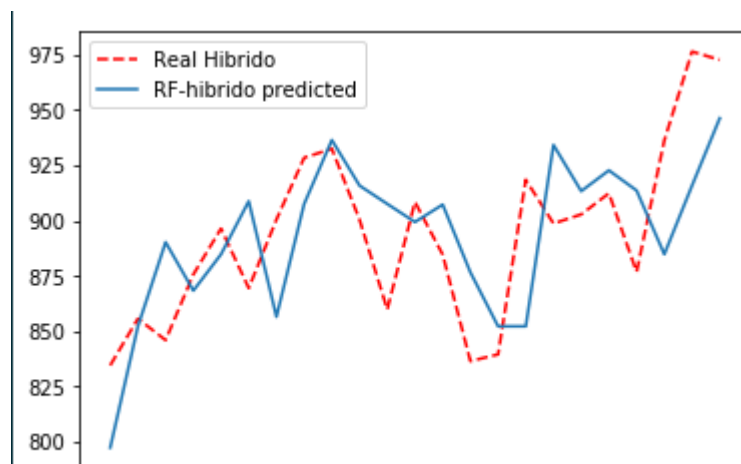


Figura 7 – Gráfico Híbrido – RF – Dataset: DownJones Fonte: do autor

5. Conclusões

Este trabalho apresenta uma abordagem de sistema híbrido por método de combinação aplicados à previsão de séries temporais. Os sistemas são compostos por modelos lineares e não-lineares para realizar o mapeamento dos padrões de diferentes séries temporais. Os experimentos foram realizados em bases de problemas reais (Banco Central do Brasil - BCB).

Os experimentos adicionais contemplam a utilização de diferentes parâmetros e dimensões diferente para análise da metodologia e também a análise residual. O sistema proposto apresentou melhoria nos resultados.

O modelo AR para o modelo MLP, o MLP obteve melhores previsões, esse comportamento foi observado em 5 bases, em relação ao erro obtido. No entanto sua utilização pode apresentar resultados variados em decorrência do tamanho da série e do tamanho mínimo da subsérie

O método de combinação, foi assumida uma relação linear entre a previsão do modelo linear e a previsão do modelo não linear, sendo adotado um modelo aditivo para as previsões. No entanto outras estratégias podem ser adotadas como o erro multiplicativo ou outras combinações utilizando rede neurais. Outra possibilidade é a utilização de meta-learning para o treinamento de um modelo específico para combinar as previsões.

6. Referencias

1. Slides, prof. JOÃO FAUSTO LORENZATO DE OLIVEIRA
2. <http://dados.gov.br/dataset/27734-saldo-das-operacoes-de-credito-por-atividade-economica-atacado-bens-duraveis-e-nao-durave>
3. . https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html
4. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
5. FIRMINO, P. R. A.; NETO, P. S. de M.; FERREIRA, T. A. Correcting and combining time series forecasters. *Neural Networks*, v. 50, p. 1 – 11, 2014.