

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA POLITÉCNICA
DEPARTAMENTO DE ENGENHARIA ELETRÔNICA E DE
COMPUTAÇÃO

PROPOSTA DE PROJETO DE GRADUAÇÃO

Aluno: João Victor Almeida Davim

Orientador: Alan Mathison Turing

1. TÍTULO

Alternativa para Previsão de Séries Temporais Utilizando Redes Neurais Sem
Peso

2. ÊNFASE

Projeto com ênfase em computação.

3. TEMA

O tema do trabalho é o estudo de uma alternativa para previsão de séries temporais. Neste sentido, o problema a ser resolvido é encontrar um modelo de aprendizado de máquina competitivo capaz de extrapolar o número de pessoas infectadas por Covid-19 no Rio de Janeiro com baixo consumo de recursos e pouco tempo de treinamento e inferência.

4. DELIMITAÇÃO

Os modelos a serem estudados são modelos de previsão de série temporal univariados, ou seja, utilizam dados históricos apenas da própria variável que será prevista. O modelo computacional apresentado como alternativa será uma Rede Neural sem Peso preparada para realizar tarefas de regressão supervisionada. O horizonte de previsão utilizado será de 7 dias.

5. JUSTIFICATIVA

Padrões podem ser observados em diversos fenômenos da natureza, fundamentais para a modelagem de fenômenos físicos, sociais e até mesmo biológicos. Para séries temporais não difere, ou seja, a utilização de padrões é essencial para a criação de modelos de previsão.

Previsão de séries temporais é um problema não tão abordado quanto problemas supervisionados de classificação e regressão, mas possui alguns fundamentos bem consolidados. Qualquer série temporal possui características que podem ser observadas como estacionaridade e autocorrelação. Além disso, também podem ser decompostas em tendência, sazonalidade e resíduo. Isso permite, poranto, modelagem do problema através do aprendizado de máquina com treinamento supervisionado. A partir de então, é possível inferir o comportamento esperado da série temporal para dias subsequentes.

Neste sentido, o presente projeto é uma alternativa a estudos anteriores, buscando otimizar a utilização de recursos e tempo de processamento sem comprometer a assertividade do modelo preditivo. Sua originalidade se dá pela inexistência de métodos para previsão de séries temporais utilizando RNSP. Até então, essas redes resolvem apenas problemas supervisionados de classificação e regressão. Assim, a importância desse trabalho está relacionada à expansão do conjunto de problemas resolvíveis por RNSP e viabilizar a previsão de séries temporais com utilização reduzida de recursos.

6. OBJETIVO

O objetivo é, então, propor uma adaptação dos problemas de previsão de séries temporais de forma que este seja resolvido utilizando uma RNSP de regressão. Desta forma, têm-se como específicos: (1) construir um modelo de RNSP ajustado a conjuntos dados de série temporal, e; (2) comparar a assertividade e os recursos utilizados com modelos pré-existentes de estudos anteriores ajustados com os mesmos conjuntos de dados.

7. METODOLOGIA

Este trabalho irá utilizar dados sintéticos, número de pessoas infectadas por Covid-19 no Rio de Janeiro e temperatura mínima diária em Melbourne para o ajuste dos modelos preditivos. Os dados sintéticos serão gerados de forma pseudo aleatória, já os dados do número de pessoas infectadas por Covid-19 e de temperatura mínima diária serão extraídos das fontes abertas eSUS e GitHub respectivamente.

Os modelos utilizados para ajuste serão Auto Regressive Integrated Moving Average (ARIMA)[1], Prophet [4] e Regression WiSARD (ReW) [3]. Cada modelo

será ajustado aos dados e terá seus hiperparâmetros otimizados através de uma busca em grade e/ou conhecimentos do domínio de cada problema individualmente. Os modelos então serão avaliados individualmente de acordo com os critérios: tempo de treinamento, tempo de inferência, uso de memória para treinamento, uso de memória para inferência e métricas de validação RMSE, MAPE, MAE e MPE.

A busca dos hiperparâmetros de cada modelo será realizada utilizando como critério de otimização cada métrica de validação apresentada. Para tal, o conjunto de dados será dividido em treino, validação e teste. O conjunto de treino será utilizado para o treinamento dos modelos enquanto o de validação para o cálculo das métricas no momento de busca dos hiperparâmetros. Então, para cada busca de hiperparâmetros será instanciado um modelo com a melhor configuração para ser treinado utilizando todo o conjunto completo (treino + validação) de dados. Em seguida, com uma configuração por métrica, todos os critérios serão avaliados para cada instância utilizando os dados separados para teste.

A partir dos resultados obtidos, será possível concluir se a utilização das RNSPs para previsão de séries temporais possui algum benefício prático. A partir desses modelos preditivos é possível, então, realizar extrapolações, que permitem a tomada de ações preventivas baseado no comportamento histórico dos dados, evidenciando informações como tendência e sazonalidade.

O êxito deste trabalho está centrado na apresentação de uma alternativa para modelos de previsão de séries temporais que exija menos recursos e tempo para ajuste, podendo assim abrir novas possibilidades de implantação desses modelos sem grandes perdas de assertividade.

8. MATERIAIS

Para este projeto será utilizado um computador de uso pessoal, no qual vai executar todos os experimentos em condições semelhantes.

A coleta de dados e os experimentos serão escritos na linguagem Python com auxílio, principalmente, das bibliotecas pandas, wisardpkg [2], statsmodels, fbprophet e memory_profiler.

Os dados serão coletados das fontes abertas eSUS e GitHub. A primeira é uma plataforma do governo de onde serão coletados os dados de número de casos confirmados de Covid-19 no Rio de Janeiro, enquanto a segunda é um repositório

aberto que será utilizado para baixar os dados de temperatura mínima diária em Melbourne.

9. CRONOGRAMA

Fase 1 (7 dias): Coleta dos dados abertos.

Fase 2 (21 dias): Experimentos com o modelo ARIMA.

Fase 3 (21 dias): Experimentos com o modelo ReW.

Fase 4 (21 dias): Experimentos com o modelo Prophet.

Fase 5 (14 dias): Compilação dos resultados e geração dos gráficos.

Fase 6 (30 dias): Escrita do texto.

Referências Bibliográficas

- [1] Saleh I. Alzahrani, Ibrahim A. Aljamaan, and Ebrahim A. Al-Fakih. Forecasting the spread of the covid-19 pandemic in saudi arabia using arima prediction model under current public health interventions. *Journal of Infection and Public Health*, 13(7):914–919, 2020.
- [2] Aluizio S. Lima Filho, Gabriel P. Guarisa, Leopoldo A. D. Lusquino Filho, Luiz F. R. Oliveira, Felipe M. G. Franca, and Priscila M. V. Lima. wisardpkg – a library for wisard-based models, 2020.
- [3] Leopoldo Lusquino Filho, Luiz Fernando de Oliveira, Aluizio Filho, Gabriel Guarisa, Lucca Felix, Priscila Lima, and Felipe França. Extending the weightless wisard classifier for regression. *Neurocomputing*, 03 2020.
- [4] Sean J Taylor and Benjamin Letham. Forecasting at scale. *PeerJ Preprints*, 5:e3190v2, September 2017.

Rio de Janeiro, 12 de julho de 2021

João Victor Almeida Davim - Aluno

Alan Mathison Turing - Orientador