



Aprendizado Multitarefa na Predição de Produção em Usinas Eólicas

Prof. Dr. Fernando J. Von Zuben

Orientador

FEEC-UNICAMP

vonzuben@dca.fee.unicamp.br

Gabriel R. Lencione

Graduando em Engenharia de Computação

FEEC-UNICAMP

gabriellencione@gmail.com

Resumo — Com a implantação disseminada de usinas eólicas integradas a sistemas de geração de energia elétrica, em todo o mundo e particularmente no Brasil, torna-se relevante estimar a produção futura dessas usinas através da predição da velocidade do vento. Em face da existência de séries temporais históricas de intensidade de vento associadas a múltiplas usinas eólicas, devem ser resolvidas simultaneamente múltiplas tarefas de predição de séries temporais. Este trabalho propõe a adoção de técnicas consolidadas de aprendizado multitarefa, permitindo compartilhar informações entre tarefas de predição que se mostrarem relacionadas. Esse compartilhamento de informação se mostrou capaz de conduzir a ganhos de desempenho na predição, quando comparado ao tratamento independente de cada tarefa.

Palavras-chave — Aprendizado Multitarefa, Séries Temporais, Energias Renováveis

I. INTRODUÇÃO

Fontes renováveis têm ocupado um papel cada vez mais importante no cenário global de energia. Fatores como o impacto ambiental, a oscilação de preços e o caráter efêmero dos combustíveis fósseis têm motivado investimentos em produção de energia limpa. Nesse cenário, a energia eólica tem apresentado crescimento substancial nos últimos anos (Hossain, 2014).

Apesar da produção de energia eólica prover muitas vantagens, seu uso intensivo traz alguns desafios para a integração com o sistema de fornecimento de energia elétrica, uma vez que a natureza incerta e intermitente dos ventos torna difícil prever a quantidade de energia que será produzida em um determinado período. Jung & Broadwater (2014) apresentam diversas abordagens que já foram propostas na literatura e que tentam superar as dificuldades inerentes a esse tipo de problema. No entanto, devido ao comportamento complexo das variáveis envolvidas, os autores constatam que os modelos existentes ainda são muito dependentes das especificidades de cada usina.

Configura-se, assim, um cenário desafiador para técnicas de predição baseadas em aprendizado de máquina, inclusive pela existência de múltiplas tarefas de predição, sendo ao menos uma para cada usina eólica, quando não uma para cada turbina.

Buscando aproveitar o possível relacionamento entre as várias tarefas de predição, este trabalho explora o uso de técnicas consolidadas de aprendizado multitarefa linear nos parâmetros, empregando-as em um conjunto de dados reais de usinas eólicas da região de Miami, Flórida, Estados Unidos, e comparando suas performances com aquelas de modelos lineares independentes clássicos, incluindo modelos ARIMA.

II. FUNDAMENTAÇÃO TEÓRICA

A. Séries Temporais

Uma série temporal corresponde, de forma simplificada, a uma sequência de valores numéricos que refletem a evolução de uma dada variável no decorrer do tempo, geralmente em intervalos igualmente espaçados de tempo. Uma das metas centrais no estudo de séries temporais reside na predição de eventos referentes ao comportamento futuro da série temporal. Em outras palavras, monitorando-se o comportamento de um processo estocástico em instantes passados, deseja-se prever as características de sua evolução temporal futura.

Para que uma série temporal admita predição de seus valores futuros, é necessário que esses valores futuros estejam correlacionados com o passado da série temporal. Uma das principais motivações, então, para o estudo de séries temporais é analisar e modelar esta dependência entre passado e futuro.

Nos anos 70, foram propostas as técnicas e metodologias que obtiveram maior destaque a partir de então, reunidas no trabalho de Box & Jenkins (Box et al., 2015), cuja primeira edição data de 1976.

Em linhas gerais, os métodos de Box & Jenkins (tais como AR, MA, ARMA etc), baseiam-se na proposição de que o valor atual da série temporal é a combinação de p valores precedentes e q impactos aleatórios anteriores, mais o impacto atual. Os p valores antecedentes formam o componente autorregressivo e os q impactos prévios formam o componente de média móvel da série. A modelagem de uma série temporal busca, então, a determinação dos valores de p e q , seguida da estimação dos respectivos coeficientes da combinação linear. Esta análise pode ser feita empregando



a autocorrelação entre passado e futuro da série temporal. Dentre os algoritmos de ajuste dos $p + q$ coeficientes temos Maximização da Verossimilhança (*Maximum Likelihood* - ML) e Quadrados Mínimos (Brockwell & Davis, 1991).

B. Aprendizado Multitarefa

O tratamento conjunto de múltiplas tarefas em aprendizado de máquina é denominado aprendizado multi-tarefa (MTL, do inglês *multi-task learning*) (Caruana, 1997; Zhang & Yang, 2017). Há ganho de desempenho em MTL, em comparação com o aprendizado independente das tarefas, quando o compartilhamento de informação entre as tarefas é benéfico ao aprendizado (Baxter, 2000). Explorar a relação entre as tarefas pode conduzir a uma melhor capacidade de generalização, particularmente quando o conjunto de amostras disponíveis para treinamento é reduzido e quando se consegue modelar adequadamente e utilizar de forma efetiva a estrutura de relacionamento entre as tarefas.

Há várias aplicações bem-sucedidas de MTL, sendo que o grupo de pesquisa se destaca na previsão de variáveis climáticas (Gonçalves et al., 2015; Aguiar, 2019), como temperatura, precipitação e umidade.

Motivada por essas várias frentes de aplicação, já foram propostas muitas formulações alternativas para modelar e empregar adequadamente a estrutura de relacionamento entre as tarefas, mecanismo denominado de regularização estrutural (Argyriou et al., 2009). O MALSAR (Zhou et al., 2012) é um pacote de software voltado para o emprego de técnicas de otimização na solução de muitas dessas formulações alternativas, as quais geralmente contêm um termo de perda, associado à composição do erro de treinamento das múltiplas tarefas, e um termo de regularização, que procura implementar o compartilhamento de informação entre as múltiplas tarefas.

É muito comum, em MTL, a adoção de modelos de aprendizado lineares nos parâmetros, pois isso torna os problemas de otimização associados convexos, ou ao menos biconvexos, além de permitir explorar soluções esparsas via o emprego de LASSO (Gonçalves et al., 2016). Assim, este trabalho lida somente com preditores lineares, seja nas abordagens MTL, ou nos modelos individuais a serem comparados.

III. METODOLOGIA

Os modelos adotados nesta pesquisa são divididos entre duas classes principais: modelos de aprendizado independente (ou monotarefa, também conhecido como STL do inglês *Single Task Learning*) e modelos de aprendizado multitarefa, sendo que parte da metodologia usada para estudar os STL vai nos servir de apoio para formalizar os MTL. Assim, começamos apresentando a formulação monotarefa.

A. Modelos de Aprendizado Independente (Monotarefa)

Em séries temporais e, de forma mais geral, em problemas de regressão, é comum o uso de técnicas que abordam cada tarefa separadamente, estimando seus parâmetros de forma independente. Neste trabalho, são considerados modelos Auto-regressivos Integrados de Médias Móveis (ARIMA na sigla inglesa), modelos de regressão linear via OLS (do inglês *Ordinary Least Squares*) e modelos de regressão linear via regularização Ridge e LASSO (Bishop, 2006).

Seja uma série temporal $Y_i, i \in \mathbb{N}$. Dizemos que ela segue um modelo $ARIMA(p, d, q)$ se ela puder ser descrita da seguinte forma:

$$Y_i(1 - z)^d(1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p) = (1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q)W_i \quad (1)$$

onde p e q são os ordens dos componentes AR e MA, d é a ordem de diferenciação, W_i é um ruído branco e z é o operador de atraso: $Y_i - Y_{i-1} = (1 - z)Y_i$. Um processo $ARMA(p, q)$ é um caso particular do modelo ARIMA em que $d = 0$.

Métodos como os de Box & Jenkins, embora gerem preditores lineares, o que pode ser limitante dependendo da natureza da série temporal, ainda são amplamente utilizados e apresentam performance muito competitiva em vários cenários (De Giorgi et al., 2011; Makridakis et al., 2018).

Neste trabalho, serão considerados modelos $ARMA(p, q)$ e modelos $ARIMA(p, 1, q)$, com $p \in [0, \dots, 8]$ e $q \in [0, \dots, 2]$ de forma a obter componentes auto-regressivos suficientemente representativos, o que será necessário nas próximas etapas, e, ao mesmo tempo, manter os modelos parcimoniosos. O papel da diferenciação nos modelos ARIMA é eliminar possíveis tendências que perturbem a estacionariedade das séries e também será estudada como técnica de pré-processamento nos demais modelos.

Para cada série temporal do nosso conjunto de dados, estimamos os modelos ARMA e ARIMA (suas ordens e seus parâmetros) utilizando o Critério de Informação de Akaike Corrigido (AICc em inglês), que é reportado na literatura (Brockwell & Davis, 2002) por evitar o sobreajuste a que pode incorrer o Critério de Informação de Akaike original (do inglês AIC). Desse procedimento retemos também as ordens p dos modelos $ARMA(p, 0)$ e $ARIMA(p, 1, 0)$ que minimizam seus respectivos critérios AICc's, de maneira que elas serão utilizadas para construir os conjuntos de regressão tanto para as séries originais como para as séries diferenciadas.

Procedemos, então, à montagem dos conjuntos de regressão linear. Seja Y_i^t a série temporal relativa à tarefa t , seja também S_i^t a série resultante da diferenciação de grau um de Y_i^t ($S_i^t = Y_i^t - Y_{i-1}^t$). Obtemos o conjunto de regressão para cada uma das séries selecionando como vetor de entrada L valores



anteriores ao elemento de saída, sendo L um único valor para todas as T séries originais e um outro único valor para as T séries diferenciadas, calculados como o valor máximo do parâmetro p , estimado na etapa anterior para os modelos ARMA e ARIMA respectivamente. A Figura 1 ilustra esse procedimento.

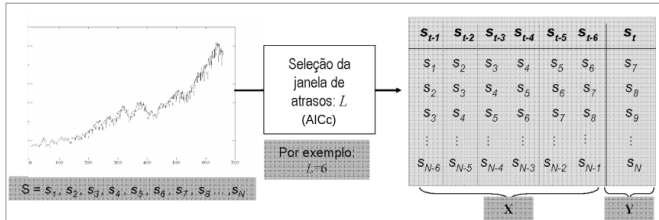


Fig. 1. Preparação dos dados para predição de um passo à frente com $L=6$ valores atrasados da série.

Uma vez que dispomos das T matrizes de entrada \mathbf{X}_t e dos T vetores de saída \mathbf{Y}_t para cada um dos dois pré-processamentos, adicionamos uma coluna de 1's para incorporar o viés, obtendo $M = L + 1$ parâmetros de regressão para cada tarefa, e aplicamos os métodos OLS, Ridge e LASSO, seguindo um procedimento de validação cruzada em cinco pastas sobre o conjunto de treino para determinação dos hiperparâmetros dos modelos.

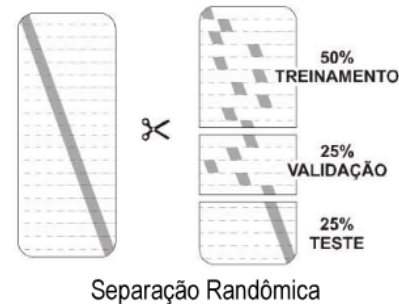
A utilização da validação cruzada no contexto de séries temporais é algo questionado por muitos autores, porém, desde que assumamos sua relativa estacionariedade, ou a obtenção da mesma através de diferenciação, seu uso é validado por alguns estudos (Bergmeir & Benítez, 2012). Fazemos, então, um recorte clássico do conjunto de dados, em que um percentual de pontos é separado como conjunto de treinamento e validação, a serem repartidos como tais a posteriori, e o restante sequencial dos pontos é mantido como conjunto de teste, a ser utilizado somente para fins de comparação de performance entre os modelos.

Para os modelos de Box & Jenkins, não há separação entre treinamento e validação e os parâmetros são estimados a partir dos dois conjuntos como um todo. Já para todos os outros métodos de regressão, a repartição dos conjuntos de treinamento e validação é feita de forma aleatória no contexto da validação cruzada. A Figura 2 exemplifica esse procedimento.

Exploraremos também o desempenho de todos os métodos aqui citados para diferentes percentuais de treinamento (incluímos também a parte de validação quando assim nos referirmos adiante) e teste, mostrando como cada um deles é capaz de generalizar na presença de mais ou menos pontos.

B. Modelos de Aprendizado Multitarefa

Modelos de aprendizado multitarefa lineares nos parâmetros costumam compartilhar de uma estrutura comum de



Separação Randômica

Fig. 2. Exemplo de separação dos dados em conjuntos de treinamento, validação e teste

otimização. Dada uma função de perda, adicionamos um elemento de regularização sobre os parâmetros e buscamos sua minimização. Assim, seja $\Theta \in \mathbb{R}^{M \times T}$ a matriz contendo em cada coluna o vetor de parâmetros da tarefa t , $\Theta_t \in \mathbb{R}^M$, de tal forma que o preditor linear obtido é dado por $\hat{\mathbf{Y}}_t = \mathbf{X}_t \Theta_t$, então escrevemos o problema final de aprendizado multitarefa da seguinte forma:

$$\Theta = \arg \min_{\Theta} \mathcal{L}(\Theta) + \mathcal{R}(\Theta) \quad (2)$$

Sendo $\mathcal{L}(\Theta)$ a função de perda original, dada, para problemas de regressão, por:

$$\mathcal{L}(\Theta) = \sum_{t=1}^T \|\mathbf{Y}_t - \mathbf{X}_t \Theta_t\|_2^2 \quad (3)$$

enquanto $\mathcal{R}(\Theta)$ é o termo de regularização que busca compartilhar informações entre as tarefas durante o treinamento de forma a obter T modelos com melhor capacidade de generalização.

Embora muitas propostas para o termo de regularização existam na literatura, escolhemos estudar uma das mais consolidadas.

Autores como Evgeniou & Pontil (2004) utilizam uma estrutura em grafo para representar o relacionamento entre as tarefas. Define-se, então, o conjunto \mathcal{E} de arestas $\mathbf{e}^{(j)} \in \mathbb{R}^T$, em que duas tarefas (nós do grafo) x e y estão ponderadamente relacionadas se $\mathbf{e}_x^{(j)} = \alpha$ e $\mathbf{e}_y^{(j)} = -\beta$, com $\alpha, \beta > 0$. O grafo completo é estruturado na matriz $\mathbf{G} = [\mathbf{e}^{(1)}, \mathbf{e}^{(2)}, \dots, \mathbf{e}^{(|\mathcal{E}|)}]$. O termo de regularização implementado a seguir penaliza, portanto, a diferença ponderada entre todos os pares conectados.

$$\mathcal{R}(\Theta) = \|\Theta \mathbf{G}\|_F^2 = \sum_{j=1}^{|\mathcal{E}|} \|\Theta \mathbf{e}^{(j)}\|_2^2 \quad (4)$$

O problema de otimização reduz-se, então, a:

$$\Theta = \arg \min_{\Theta} \mathcal{L}(\Theta) + \lambda_1 \|\Theta \mathbf{G}\|_F^2 + \lambda_2 \|\Theta\|_1 \quad (5)$$



onde o último termo corresponde à norma ℓ_1 de Θ e é utilizado para induzir esparsidade ao modelo. Já os hiperparâmetros λ_1 e λ_2 controlam a penalização dos regularizadores que impomos ao aprendizado, e são selecionados via validação cruzada.

Aqui, adotamos a Correlação de Pearson entre as séries temporais para estabelecer uma medida de similaridade entre tarefas, construindo as arestas do grafo da seguinte forma: Sejam x e y duas tarefas, então definimos uma aresta $e^{(j)} \in \mathcal{E}$ tal que $e_x^{(j)} = -e_y^{(j)} = \text{corr}(Y_i^x, Y_i^y)$ e $e_k^{(j)} = 0, k \neq x, y$. Sendo as séries temporais tomadas somente nos seus conjuntos de treinamento (para o caso das séries diferenciadas, trocamos Y_i^t por S_i^t). Nos experimentos que se seguem, utilizaremos a sigla MTL para indicar a aplicação dessa técnica.

C. Dados e Métricas de Performance

O conjunto de dados a ser utilizado é composto por $T = 10$ séries de velocidade do vento de sítios eólicos localizados em Miami, Flórida, Estados Unidos. Elas foram extraídas da ferramenta *Wind Integration National Dataset (WIND) Toolkit* (Draxl et al., 2015), que consiste em uma base de dados meteorológicos gerados de 5 em 5 minutos de 2007 até 2013 em milhares de sítios eólicos no território estadunidense. As velocidades são capturadas a uma altitude de 100 metros e são expressas em metros por segundo (m/s). Nós, entretanto, focaremos este experimento apenas nos três primeiros meses do ano de 2012, que resultam em, aproximadamente, 24 mil pontos para cada série temporal.

Esses pontos são fatiados, então, em 30 conjuntos C_k com 400 pontos para cada uma das 10 séries, intervalados entre si por 400 pontos, de forma que cada um deles tenha a menor correlação possível com os outros, o que será necessário quando formos realizar os testes de significância para comparar os modelos.

Para cada C_k aplicamos os modelos de aprendizado independente e multitarefa, descritos previamente, nas cinco seguintes porcentagens de treinamento 10%, 27.5%, 45%, 67.5% e 80%, avaliando a performance dos modelos nos respectivos conjuntos de teste.

A fim de podermos avaliar as suas performances e compará-las entre si, é necessário estabelecer métricas para o desempenho dos modelos nas 10 tarefas dentro das 5 porcentagens de treinamento para cada um dos 30 recortes.

Adotamos, então, duas métricas: uma sensível à escala, RMSE e outra relativa, RELRMSE, descritas abaixo.

Seja \hat{Y}_i^t a predição de horizonte 1 (5 minutos) do ponto i do conjunto de testes da série t , então definimos a raiz quadrada do erro quadrático médio (RMSE na sigla inglesa) como:

$$RMSE^t = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{Y}_i^t - Y_i^t)^2} \quad (6)$$

onde N é o número de pontos do conjunto de teste na respectiva porcentagem.

Um procedimento comum em predição de séries temporais é comparar o erro da predição obtida com o modelo proposto com aquele obtido por um método de referência. Aqui, adotaremos o método da persistência (ou método ingênuo), em que $\hat{Y}_i^t = Y_{i-1}^t$. Assim, definimos o RELRMSE como:

$$RELRMSE^t = \frac{RMSE^t}{RMSE_{pers}^t} \quad (7)$$

Fazemos, portanto, a média dessas métricas em relação às 10 tarefas. Vale ressaltar que métricas sensíveis à escala não são adequadas para serem tomadas em média no contexto de múltiplas séries devido ao fato de que um erro relativamente pequeno para uma determinada série pode acabar ponderando com mais força a média final caso seja muito maior em escala que o das outras séries.

Dada a proximidade entre as tarefas aqui consideradas, essa disparidade não se faz tão evidente, e utilizamos os $RMSE^t$ com intuito de auxiliar o cálculo do $RELRMSE^t$. Já no caso das métricas relativas, diversos autores (Bergmeir & Benítez, 2012) referendam o seu uso no contexto deste experimento.

IV. RESULTADOS E DISCUSSÃO

Essas métricas médias são tomadas no contexto de cada percentual de treino/teste para cada um dos recortes C_k . Assim, dispomos de métricas de performance dos modelos em 30 amostras diferentes, o que nos permite realizar o Teste de Friedman, seguido pelo Teste de Fisher (post-hoc) (Pereira et al., 2015), para determinar a diferença de desempenho entre os modelos. Utilizamos o RELRMSE pelas razões já citadas e reportamos os resultados na Tabela 1.

Tabela 1. Resultado dos testes de significância utilizando o RELRMSE.

Modelos	Vitórias (V), Derrotas (D) e Rank Médio (R) por Porcentagem de Treinamento				
	10%	27.5%	45%	67.5%	80%
ARMA	V=1 D=6 R=7.4	V=0 D=4 R=7.2	V=1 D=4 R=6.7	V=2 D=2 R=6.4	V=3 D=1 R=6.1
OLS	V=0 D=6 R=8.4	V=0 D=4 R=7.2	V=2 D=2 R=6.2	V=3 D=1 R=5.9	V=3 D=1 R=5.6
Ridge	V=0 D=7 R=9.3	V=0 D=6 R=8.2	V=0 D=7 R=8.2	V=0 D=8 R=8.3	V=0 D=8 R=8.0
LASSO	V=0 D=6 R=9.0	V=0 D=6 R=8.2	V=0 D=5 R=7.7	V=0 D=7 R=8.1	V=0 D=7 R=7.8
MTL	V=0 D=6 R=8.2	V=2 D=2 R=6.2	V=4 D=1 R=4.9	V=3 D=1 R=5.0	V=3 D=0 R=4.5
ARIMA	V=3 D=1 R=4.7	V=1 D=6 R=5.6	V=3 D=1 R=6.2	V=1 D=6 R=6.3	V=2 D=1 R=6.2
OLS Dif.	V=5 D=1 R=4.7	V=5 D=1 R=4.9	V=3 D=1 R=5.5	V=3 D=1 R=5.0	V=3 D=0 R=4.8
Ridge Dif.	V=5 D=1 R=4.2	V=5 D=1 R=4.7	V=4 D=1 R=4.7	V=3 D=1 R=4.9	V=3 D=1 R=5.2
LASSO Dif.	V=5 D=0 R=3.4	V=6 D=1 R=4.1	V=6 D=0 R=4.4	V=4 D=1 R=4.6	V=3 D=1 R=5.1
MTL Dif.	V=9 D=0 R=2.4	V=10 D=0 R=2.3	V=9 D=0 R=2.8	V=10 D=0 R=2.7	V=8 D=0 R=3.2
Persistência	V=5 D=1 R=4.4	V=0 D=5 R=5.5	V=0 D=8 R=8.6	V=0 D=8 R=8.9	V=1 D=6 R=9.4

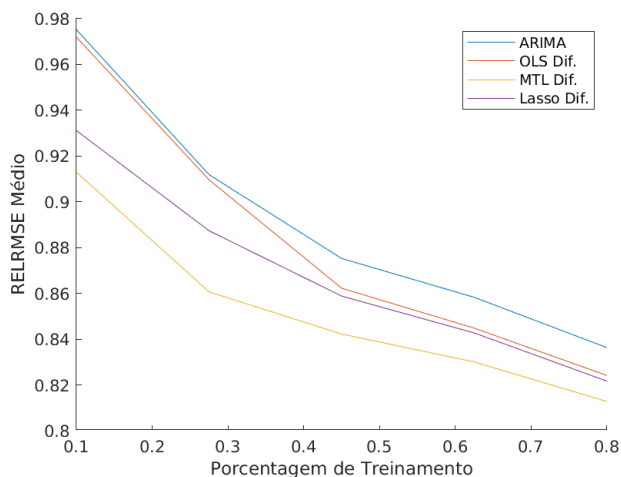


Fig. 3. Gráfico do RELRMSE médio vs percent. de treinamento.

Nessa tabela, descrevemos para cada porcentagem de treinamento o número de vitórias e derrotas estatisticamente significativas em relação aos outros modelos, bem como a posição (*rank* em inglês) média de cada método, em que um menor RELRMSE significa uma maior performance do modelo e, portanto, uma posição mais alta (numericamente menor) na classificação.

Em linhas gerais, podemos observar que os métodos baseados nas séries diferenciadas são capazes de generalizar melhor na tarefa de predição. Isso se deve ao fato de que eles são capazes de absorver o comportamento da variação das séries sem ficar presos a diferentes platôs de média que elas possam apresentar. Vemos, também, que o modelo proposto de aprendizado multitarefa, quando aplicado nas séries diferenciadas, é capaz de se sobressair em todas as porcentagens de treinamento experimentadas, apresentando o menor *rank* médio e não perdendo de nenhum outro modelo.

Apresentamos, então, na Figura 3, o gráfico com o RELRMSE médio (tomado nas 30 amostras) *versus* a porcentagem de treinamento para os seguintes modelos: MTL, OLS e LASSO sobre séries diferenciadas e ARIMA.

É possível notar que a abordagem MTL induz uma melhor generalização quando não dispomos de muitos dados para treinar o modelo, chegando a ganhos de performance em relação ao modelo clássico ARIMA da ordem de 6% nos menores percentuais.

V. CONCLUSÃO

Embora esse trabalho já tenha apresentado resultados sólidos e positivos em relação à adoção de métodos de aprendizado multitarefa na predição de velocidade do vento em usinas eólicas reais, muita investigação ainda se faz necessária. Vislumbramos produzir mais experimentos com

outras técnicas de MTL e outras medidas de similaridade entre as séries temporais. Vale ainda experimentar o uso desses métodos na predição a médio e longo horizonte.

E ressaltamos, também, o desejo desse projeto de pesquisa de aplicar a metodologia, aqui proposta, a usinas eólicas brasileiras, com a ambição de poder contribuir com o desenvolvimento e integração dessa importante fonte de energia renovável no território nacional.

REFERÊNCIAS BIBLIOGRÁFICAS

- Aguiar, N.S. "A multitask learning approach to automatic threshold selection in Pareto distributions", Tese de Mestrado, Faculdade de Engenharia Elétrica e de Computação, 2019.
- Argyriou, A.; Micchelli, C.A.; Pontil, M. "When Is There a Representer Theorem? Vector Versus Matrix Regularizers", *Journal of Machine Learning Research*, vol. 10, pp. 2507-2529, 2009.
- Baxter, J. "A model of inductive bias learning", *Journal of Artificial Intelligence Research*, vol. 12, pp. 149-198, 2000.
- Bergmeir C. and Benítez J. M. "On the use of cross-validation for time series predictor evaluation", *Information Sciences*, vol. 191, pp 192-213, 2012.
- Bishop, Christopher M. "Pattern Recognition and Machine Learning". New York :Springer, 2006.
- Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. "Time Series Analysis: Forecasting and Control", Wiley, 5th edition, 2015.
- Brockwell, P.J.; Davis, R.A. "Introduction to Time Series and Forecasting", Springer, 2nd edition, 2002.
- Brockwell, P.J.; Davis, R.A. "Time Series: Theory and Methods", Springer, 2nd edition, 1991.
- Caruana, R. "Multitask Learning", *Machine Learning*, vol. 28, no. 1, pp. 41-75, 1997.
- De Giorgi M. G., Ficarella A. and Tarantino M., "Error analysis of short term wind power prediction models", *Applied Energy*, vol. 88, pp. 1298-1311, 2011.
- Draxl, C., B.M. Hodge, A. Clifton, and J. McCaa. "The Wind Integration National Dataset (WIND) Toolkit." *Applied Energy*, 2015.
- Evgeniou, T. and Pontil, M. "Regularized multi-task learning", *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 109-117, 2004.
- Gonçalves, A.R.; Von Zuben, F.J.; Banerjee, A. "A Multitask Learning View on the Earth System Model Ensemble", *Computing in Science & Engineering*, vol. 17, pp. 35-42, 2015.
- Gonçalves, A.R.; Von Zuben, F.J.; Banerjee, A. "Multi-task Sparse Structure Learning with Gaussian Copula Models", *Journal of Machine Learning Research*, vol. 17, pp. 1-30, 2016.
- Hossain, J. "World Wind Resource Assessment Report", WWEA Technical Committee, World Wind Energy Association, 2014.
- Jung, J.; Broadwater, R. P. "Current status and future advances for wind speed and power forecasting", *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 762-777, 2014.
- Makridakis S., Spiliotis E. and Assimakopoulos V. "Statistical and Machine Learning forecasting methods: Concerns and ways forward". *PLoS ONE*, 2018.
- Pereira, D., Afonso, A. and Medeiros, F. "Overview of Friedman's Test and Post-hoc Analysis". *Communications in Statistics - Simulation and Computation*, vol. 44, pp. 2636-2653, 2015.
- Zhang, Y., Yang, Q. "A Survey on Multi-Task Learning", arXiv: 1707.08114v1, 2017.
- Zhou, J., Chen, J. and Ye, J. "MALSAR: Multi-task learning via structural regularization", *User's Manual*, 2012.

AGRADECIMENTOS

Agradecemos à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento desse projeto de iniciação científica sob número de processo: 2020/03913-0.