



Aplicabilidade de Técnicas de Clusterização e Máquinas de Vetores de Suporte para Previsão de Radiação Solar em Regiões Tropicais

Hendriigo Batista da Silva

Universidade Federal de Minas Gerais
Av. Antonio Carlos, 6627 - Pampulha - Belo Horizonte, MG, Brazil
hendrigobatista@ufmg.br

Leonardo P. Santiago

Copenhagen Business School
Solbjerg Pl. 3, 2000 Frederiksberg, Dinamarca
ls.om@cbs.dk

RESUMO

A geração de energia solar terá um impacto significativo na matriz de eletricidade nos próximos anos. Assim, a previsão da radiação solar a curto prazo desempenhará um papel estratégico ao dar suporte a um grid elétrico cada dia mais conectado e intermitente. Este artigo estende para regiões tropicais a aplicação da técnica de clusterização, classificação e regressão, recentemente proposta na literatura e que apresentou uma boa habilidade de previsão na Europa. Regiões tropicais geralmente apresentam menor acurácia nas previsões nos modelos numéricos correntes devido à instrumentação mais escassa. Foram avaliados a previsibilidade das fases desta abordagem em dois cenários utilizando dados auferidos no dia anterior ao do dia da previsão. Foi analisada como a substituição das variáveis do dia anterior pela previsão equivalente para o dia posterior impacta na habilidade de previsão do modelo. Verificou-se como sua acurácia aumenta, à medida em que se melhora a previsão destas novas entradas.

PALAVRAS CHAVE. Previsão de Radiação Solar. Máquinas de Vetores de Suporte. Análise de Cluster.

ÁREA PRINCIPAL: EN - PO na Área de Energia.

ABSTRACT

Solar generation will have a significant impact in the electricity supply mix in the upcoming years. Thereby the short-term solar radiation forecast will play a strategic rule in supporting an even more connected and intermittent electrical grid. This paper extends to tropical regions the applicability of the clustering, classification and regression technique, recently proposed in the literature and that presented good results in Europe. Tropical regions usually present lower accuracy in the current numerical models predictions, due to a more sparse instrumentation. We evaluate the predictability of the phases of this approach in two scenarios, by utilizing data from the previous day of the prediction one. Furthermore, we analyzed how the replacement of these variables for the prediction of those equivalent in the prediction day impacts in the forecast skill of the model. We verified how the accuracy increases as the predictions of these new entries improves.

KEYWORDS. Solar Radiation Prediction. Support Vector Machine. Cluster Analysis.

MAIN AREA: EN - OR in the Energy Sector.



1. Introdução

O crescente aumento da participação da geração solar na matriz elétrica demandará níveis de acurácia cada vez maiores na previsão da radiação solar a curto prazo. Quanto maior for a acurácia de um modelo de previsão para geração solar, mais barata poderá ser a energia elétrica despachada no grid.

O custo de ajuste de outras instalações de geração de energia para balancear o grid em sistemas conectados com fontes solares aumenta com a diminuição do horizonte de ajuste, de acordo com Bacher et al. [2009]. Conseqüentemente, uma previsão precisa da radiação para usinas solares e microgrids leva a uma melhor alocação de diversos recursos energéticos e diminui o custo total de operação de todo o sistema conectado. De acordo com Reikard [2009], previsões de radiação de curto prazo são necessárias para planejamento operacional, programação de backup, compras de energia de curto prazo, planejamento de uso de reservas e gerenciamento de pico de carga.

De acordo com Aguiar et al. [2015], a radiação é uma das principais variáveis que impactam na produção de energia solar em uma região. Sua previsão é de vital importância e é largamente discutida na literatura. Por exemplo, Inman et al. [2013], Besharat et al. [2013], Heinemann et al. [2006] e Diagne et al. [2012] fazem uma revisão dos principais modelos de previsão de radiação presentes na literatura. Meteorologistas lidam com a previsão de radiação através de diferentes abordagens, dependendo do horizonte de previsão (resolução temporal) e cobertura geográfica (resolução espacial). Em resumo, Diagne et al. [2012] define qual delas é melhor de acordo com escalas temporais e espaciais. Há duas principais abordagens a se considerar dentre todos eles: modelos físicos baseados em equações físicas e condições iniciais da atmosfera e modelos estatísticos baseados em análise e reconhecimento de padrões em séries temporais de dados históricos de radiação.

Dentre deste último grupo, tem-se modelos autorregressivos (AR), autorregressivos com média móvel (ARMA), redes neurais artificiais (ANN), máquinas de vetores de suporte (SVM) e algoritmos genéticos (GA). Goh e Tan [1977] introduziram o uso de séries temporais estocásticas para prever dados solares e o uso de técnicas de inteligência artificial tem aumentado ao longo do tempo para previsões meteorológicas. Por exemplo, Kuligowski e Barros [1998] utilizam modelos de inteligência artificial para prever a chuva em algumas regiões, obtendo níveis razoáveis de previsibilidade. Marzban e Stumpf [1996] por sua vez utiliza um modelo de ANN para prever a formação de tornados em Oklahoma considerando 23 diferentes características de entrada observáveis. Maqsood et al. [2004] aplica modelos de ANN para obter previsões do tempo baseadas em dados de temperatura, velocidade do vento e umidade, obtendo previsões com relativa acurácia para algumas formulações. Mangalova e Agafonov [2014] apresenta uma abordagem não-paramétrica de k-vizinhos mais próximos (k-NN) para previsão de geração de energia eólica. Silva [2014] também apresenta um modelo para previsão de geração eólica. Neste caso, é abordado um modelo que considera árvores de decisão e modelos de regressão linear. Nagy et al. [2016] apresenta como a previsão probabilística pode obter bons resultados trabalhando com preditores para energia renovável. Há também um conjunto de técnicas de previsão probabilística apresentadas em Hong et al. [2014], Hong et al. [2016] e Hong e Fan [2016], que lidam com previsões para geração eólica, solar e de carga. Com relação à previsão solar, Reikard [2009] apresenta conclusões comparando previsões considerando regressão em logs, ARIMA, modelos de componentes não-observados, ANN e modelos híbridos. Huang e Perry [2016] utilizam gradiente boosting para treinar o modelo para converter saídas NWP para previsões pontuais de energia solar. Hammer et al. [2001] utiliza um combinado de análise estatística e imagens de satélite em sua abordagem. Sfetsos e Coonick [2000] apresentam bons resultados na aplicação de técnicas de inteligência artificial e Ahmad et al. [2015] propõe um modelo de ANN para prever radiação solar em cidades da Nova Zelândia. Outros artigos recentes também abordam o aprendizado de máquina para previsão da radiação solar, como Li et al. [2016] e Lima et al. [2016].

Recentemente, modelos estão sendo propostos baseados em uma combinação híbrida de diferentes ferramentas e habilidades como Numerical Weather Prediction (NWP), análise estatística,



aprendizado de máquina e vetores de imagens de satélite. Gala et al. [2016] apresenta algumas vantagens de um modelo misto de NWP global com aprendizado de máquina para previsão da radiação local. Aguiar et al. [2015] propõe um modelo de ANN para prever a radiação solar utilizando medidas no solo e imagens de satélite, melhorando a acurácia. Akarslan e Hocaoglu [2016] apresenta uma abordagem híbrida adaptativa que utiliza diferentes técnicas de previsão dependendo da estação do ano e de índices de claridade. Aryaputera et al. [2015] lida com a previsão de curto-prazo da radiação solar utilizando krigagem espaço-temporal, obtendo bons resultados for alguns horizontes de tempo. Além destas abordagens, Jiménez-Pérez e Mora-López [2016] propõem um modelo de duas fases combinando algoritmos de clustering, árvores de decisão, ANN e SVM para obter previsões para valores horários do próximo dia para radiação solar com razoável acurácia.

Entretanto, poucos destes modelos foram desenvolvidos para regiões tropicais. Além disso, uma das barreiras mais importantes para países em desenvolvimento como o Brasil para aumentar a penetração de fontes solares e eólica é a falta de informações confiáveis e expertise nacional, como pontuado na pesquisa conduzida por Martins e Pereira [2011]. Uma das recomendações científicas e tecnológicas apresentadas nesta pesquisa foi a necessidade de adaptação para regiões tropicais dos modelos numéricos de alta performance das regiões de alta latitude. Esta limitação é devido principalmente à escassez de dados e ao tratamento geral de processos diabáticos. Krishnamurti et al. [1991] argumenta que grandes lacunas nos dados sobre partes terrestres e oceânicas de regiões tropicais impactam na inicialização dos fluxos de superfície de umidade e calor sensível, precipitação e nuvens nos modelos numéricos. Ao contrário de regiões de alta latitude, iniciar um modelo de previsão com grande erro na distribuição de umidade tem um impacto significativo em baixas latitudes. Apesar de diferentes modelos de NWP ao redor do mundo estarem assimilando dados de satélite para diminuir essa escassez de dados sobre regiões tropicais, os dados atualmente disponíveis estão longe de já serem explorados no seu potencial máximo. Oliveira et al. [2002] também argumenta que a atividade das nuvens em regiões tropicais como o Brasil torna o problema de representação dos efeitos das nuvens em modelos numéricos particularmente severo, aliado ao fato da rede de medição solar ser esparsa e na maioria das vezes localizada apenas em áreas urbanas.

Em linha com esta constatação e de forma a avaliar como uma previsão baseada nos dados históricos pode ser abordada em regiões tropicais, este artigo investiga o desempenho da previsão de radiação solar. São aplicadas para a cidade de Belo Horizonte as técnicas de clusterização, classificação e regressão propostas em Jiménez-Pérez e Mora-López [2016] e discutidos os resultados da sua aplicação para regiões de baixas a médias latitudes. Também é avaliado como previsões podem ser melhoradas se um conjunto de previsões com boa acurácia para dados meteorológicos do próximo dia pudessem ser utilizados como dados de entrada do modelo, ao invés dos dados do dia anterior.

Este artigo é organizado da seguinte forma. A seção 2 explica a metodologia utilizada, como detalhes das técnicas de análise de cluster e aprendizado de máquina para classificação e regressão. Na seção 3, os dados utilizados são detalhados e como estas observações foram abordadas nessas técnicas de inteligência artificial. Na seção 4, os resultados e estatísticas de previsão são apresentados e discutidos em três diferentes cenários de previsibilidade de variáveis de entrada e na seção 5, é apresentada a conclusão e possibilidades de trabalhos futuros.

2. Metodologia

Nesta seção é definida a metodologia que foi utilizada para previsão da radiação solar. Foi aplicada análise de cluster e máquinas de vetores de suporte para a tarefa de reconhecimento de padrões e para prever a radiação solar para o próximo dia, através do índice de claridade.

2.1. Análise de Cluster

De acordo com Jain et al. [1999], a análise de cluster é um processo de agrupamento de dados baseados em uma medida de similaridade. Com cálculos de erros quadráticos, foi escolhida a abordagem de k-médias, como na abordagem recente apresentada em Jiménez-Pérez e Mora-López [2016]. Ela consiste em selecionar k centroides aleatoriamente e alocar cada observação a



seu centroide mais próximo. Logo após, a próxima iteração calcula novas posições de centroides baseadas na distância média de todas as observações associadas a cada centroide. A nova posição do centroide passa a modificar as observações associadas a cada cluster. Após várias iterações, o centroide e os clusters se estabilizam, finalizando o algoritmo com k diferentes grupos definidos.

A figura 1 exemplifica uma versão bidimensional, x_1 e x_2 , com quatro diferentes clusters. Este artigo lida com uma versão com oito dimensões, também com quatro diferentes clusters.

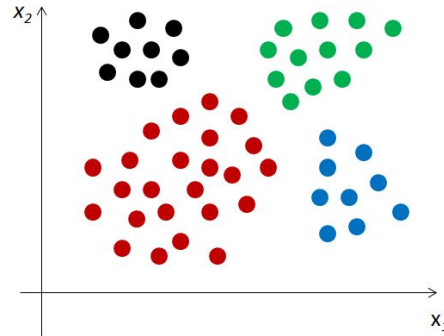


Figura 1: Análise de cluster em um espaço bidimensional com quatro clusters.

2.2. Máquinas de Vetores de Suporte (SVM)

Máquinas de Vetores de Suporte ou Support Vector Machines (SVM) é uma técnica de inteligência artificial ou aprendizado de máquina utilizada para mineração de dados ou reconhecimento de padrões em um banco de dados. A teoria por trás da SVM foi primeiramente estudada por Vapnik e Chervonenkis [1964]. De acordo com esta teoria, um grupo de dados representativos pode ser utilizado para treinar uma máquina e parâmetros ótimos podem assim serem obtidos. Esta teoria caracteriza propriedades de aprendizagem de máquinas que possibilitam sua boa generalização para parâmetros não utilizados no treinamento, de acordo com Smola e Schölkopf [2004]. Pode ser utilizada para propósitos de classificação ou regressão.

2.2.1. SVM - Classificação

Modelos de SVM para classificação são utilizados que se diferenciar e rotular diferentes classes de observações definindo-se parâmetros de um hiperplano que separa regiões distintas em um espaço n -dimensional com base nas classes definidas. O uso de curvas de separação não-lineares é possível porque, de acordo com Cortes e Vapnik [1995], vetores de entrada são mapeados não-linearmente para um espaço de entrada com dimensões bem maiores.

Nesta abordagem, foi utilizada uma versão denominada SVM de margem suave. Ela permite flexibilidade em lidar com outliers de todas as classes de separação. Os parâmetros ótimos (w e b) do hiperplano de separação são obtidos como solução do modelo de otimização. O parâmetro C regula como outliers devem ser lidados ou desconsiderados e ξ_i indica quanto a observação i viola seu lado da margem. Para mais detalhes da abordagem de SVM de margem suave, veja Cortes e Vapnik [1995]. Cada vetor de entrada x_i é rotulado com um valor de saída y_i , considerando m observações de treinamento i :

$$\min_{w,b,\epsilon} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \quad (1)$$

s.a

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad \forall i \quad (2)$$

$$\xi_i \geq 0 \quad \forall i \quad (3)$$

O hiperplano ótimo obtido separa observações de classes diferentes e pode ser usado para generalizações com dados não utilizados para treinamento. A figura 2a exemplifica uma versão simplificada com duas dimensões de entrada, x_1 e x_2 .

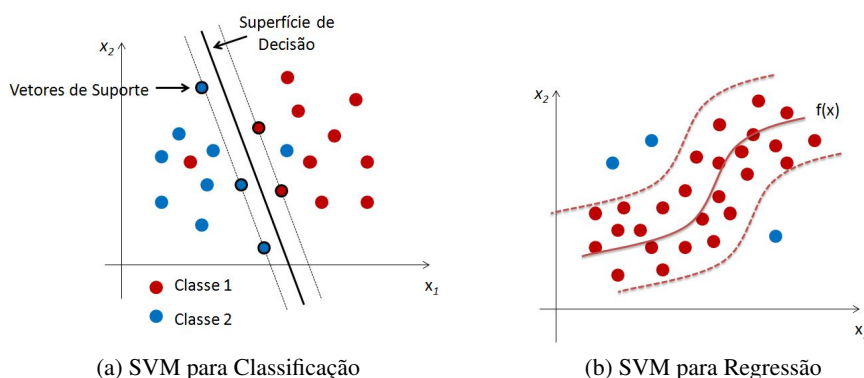


Figura 2: SVM para Classificação e Regressão

2.2.2. SVM - Regressão

SVM para regressão foi primeiramente introduzida por Vapnik [1995]. É similar à versão de classificação, mas de acordo com Smola e Schölkopf [2004], o objetivo da técnica é obter uma função $f(x)$ que possui no máximo um desvio ϵ do alvo y_i para todas as observações de treinamento, e ao mesmo tempo ser uma função o mais plana possível. De forma similar ao problema da classificação, pode-se ter uma versão de margem suave do modelo. Nesta versão, é permitida a uma observação de treinamento outlier violar a distância ϵ da função $f(x)$.

Neste modelo, w e b são os parâmetros ótimos da função $f(x)$. O parâmetro C também regula como os outliers devem ser levados em consideração e ϵ_i indica quanto a observação i viola seu lado da margem, sendo ξ_i e ξ_i^* variáveis de folga introduzidas na versão de margem suave. Cada vetor de entrada x_i é rotulado com um valor de saída y_i , considerando m observações de treinamento i . De acordo com Smola e Schölkopf [2004], o modelo de otimização é definido como:

$$\min_{w,b,\xi_i} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i + \xi_i^* \quad (4)$$

s.t.

$$y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i \quad \forall i \quad (5)$$

$$\langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \quad \forall i \quad (6)$$

$$\xi_i, \xi_i^* \geq 0 \quad \forall i \quad (7)$$

Após obter a função ótima $f(x)$, este modelo pode ser generalizado para observações não utilizadas no treinamento. A figura 2b exemplifica uma versão simplificada com duas dimensões de entrada, x_1 e x_2 .

3. Dados

Esta análise foi realizada em um banco de dados de diferentes variáveis meteorológicas horárias, obtidas da estação automática do INMET em Belo Horizonte (Pampulha - latitude 19°53'S, longitude 43°58'W, 869m de elevação). A figura 3 mostra os dados de radiação solar utilizados para treinar e testar esta abordagem de 9 da manhã até 4 da tarde, de 1º de janeiro de 2012 até 31 de dezembro de 2016, totalizando 1.827 dias ou 14.616 observações horárias sobre radiação solar.

Estes dados foram utilizados em conjunto com variáveis de temperatura, umidade e pressão atmosférica para se identificar padrões e poder inferir um valor para a radiação solar a cada hora no próximo dia. Uma abordagem recente proposta nessa linha e que apresentou uma alta habilidade de previsão foi apresentada em Jiménez-Pérez e Mora-López [2016] e aplicada em Málaga (Espanha). Neste artigo exploramos como esta abordagem espanhola se comporta em regiões tropicais.

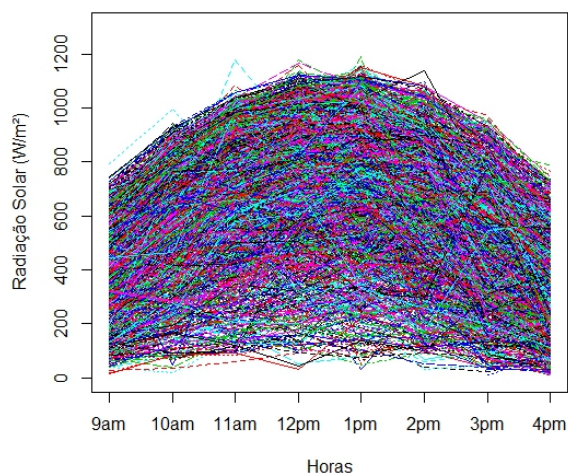


Figura 3: Radiação Solar em Belo Horizonte (Pampulha) de 1º de janeiro de 2012 até 31 de dezembro de 2016 na estação meteorológica automática do INMET.

Ela se baseia em técnicas de clusterização, classificação e regressão. A partir dos dados obtidos, é calculada uma série temporal horária de índices de claridade. Este índice é a fração entre a radiação medida na estação e a radiação extraterrestre recebida acima da atmosfera em determinado horário e local. A radiação extraterrestre não sofre interferência das nuvens e foi calculada com base no modelo descrito por Iqbal [2012], considerando a latitude de Belo Horizonte. A figura 4 mostra as observações sazonais da radiação solar de 2012 a 2016 (1827 dias) às 12h para a cidade. Por este gráfico, pode-se ver como as nuvens e outras condições atmosféricas podem impactar significativamente na radiação medida no solo, e consequentemente na energia solar gerada.

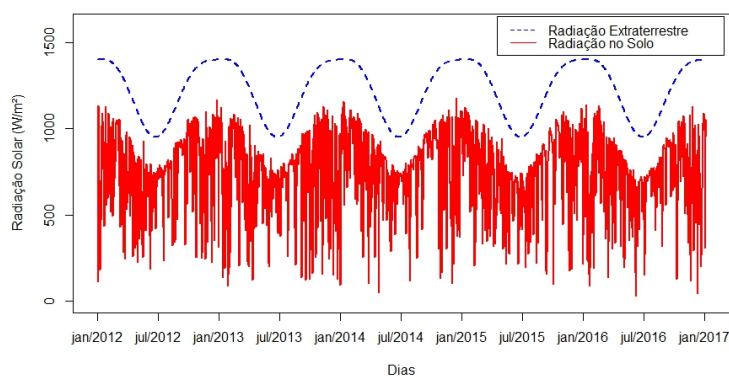


Figura 4: Radiação horária medida e radiação horária extraterrestre às 12h de 2012 a 2016 (1827 dias)

Este índice de claridade capta como as nuvens podem tornar a geração solar intermitente. A figura 5 mostra o índice de claridade para o mesmo dia às 12h. Pode-se observar que não há mais as tendências sazonais observadas na figura 4 quando se considera apenas o índice de claridade. Isto ocorre devido ao fato de esta sazonalidade desaparecer na razão entre duas variáveis igualmente impactadas por esta variação.

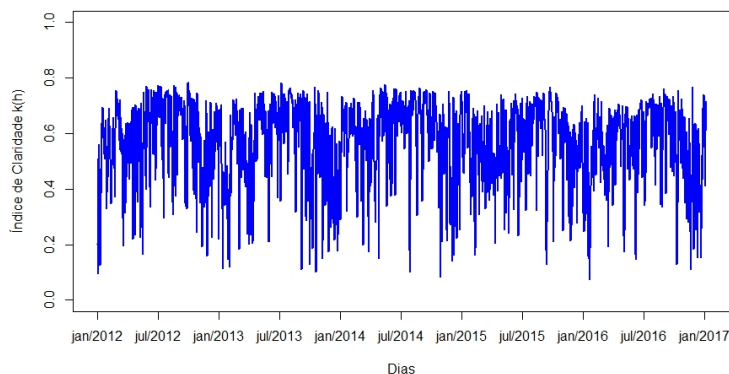


Figura 5: Índice horário de claridade às 12h de 2012 até 2016 (1827 dias)

De forma a deixar todas as observações com magnitude comparável e captar apenas o padrão de variação do formato diário do índice de claridade, foi eliminada a tendência ao se subtrair o índice horário de claridade pelo respectivo índice diário daquele dia. Esta medida diária é obtida ao se dividir a radiação total diária recebida pela radiação extraterrestre ao longo deste dia.

Logo após foi realizada uma análise de cluster no conjunto de dados de oito dimensões (oito horas) para se obter quatro diferentes centroides de índices de claridade ao longo do dia, definindo quatro diferentes padrões: céu limpo, céu nublado todo o dia, céu nublado principalmente pela manhã, céu nublado principalmente pela tarde.

Foram utilizadas variáveis como temperatura, umidade relativa e pressão atmosférica como dados de entrada para se prever a radiação solar do próximo dia. Esta tarefa foi realizada pela aplicação das técnicas de inteligência artificial que obtiveram os melhores resultados nos experimentos de Jiménez-Pérez e Mora-López [2016]: SVM para classificação para prever o cluster do próximo dia e SVM para regressão para prever o índice de claridade diário para o próximo dia. Finalmente, este índice é somado a cada centroide e multiplicado pela radiação solar extraterrestre para se obter um valor estimado de radiação para cada hora do dia.

Buscando avaliar o desempenho das previsões, optou-se por particionar os dados em dados para treinamento (ano de 2012 até 2015) e dados para teste (ano de 2016).

4. Resultados e Discussão

O código utilizado para os cálculos foi programado na linguagem R. Na análise de cluster, foi utilizada a função 'kmeans' do pacote 'stats' e para SVM, foi utilizada a função 'svm' do pacote 'e1071'. O primeiro resultado obtido foram os 4 centroides da análise de cluster do conjunto de dados de índices de claridade, apresentados na figura 6. Pode-se observar que o cluster 1 agrupa principalmente dias nublados pela tarde, o cluster 2 dias com períodos nublados principalmente pela manhã. Os clusters 3 dias com céu claro e o cluster 4 os dias irregulares com períodos nublados ao longo de todo o dia. Pode-se observar que os dados de Belo Horizonte apresentam padrões similares aos dados auferidos em Málaga (Espanha) e apresentados em Jiménez-Pérez e Mora-López [2016]. Por outro lado, há algumas variações importantes, principalmente a presença do cluster 4, cujo formato não se observa no caso espanhol.

Para testar a habilidade de previsão desta abordagem, foram simulados e analisados 3 cenários de previsão. No primeiro cenário, os clusters e o índice de claridade diário k_d são previstos e no segundo cenário apenas k_d é previsto. Já nos terceiro cenário foi realizada uma análise da utilização de dados estocásticos como variáveis de entrada do modelo de previsão de k_d . Foi realizada uma análise de como o aumento da incerteza da previsão dos dados de variáveis de entrada impactaria na habilidade de previsão do modelo.

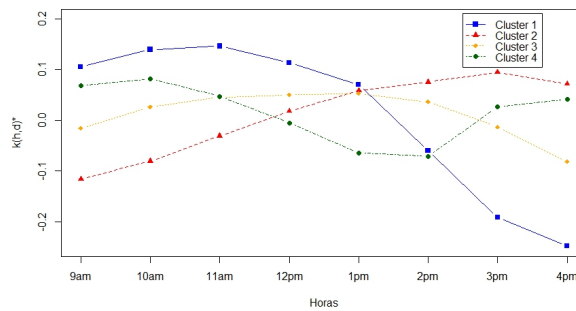


Figura 6: Valores dos centroides de cada cluster.

4.1. Cenário 1: prever cluster e k_d

Neste cenário, foram utilizadas variáveis meteorológicas como temperatura, umidade e pressão atmosférica do dia atual para previsão do cluster (SVM-Classificação) e do índice diário de claridade (SVM-Regressão) de amanhã, que foram as técnicas de aprendizado de máquina que alcançaram os melhores resultados nos experimentos apresentados por Jiménez-Pérez e Mora-López [2016]. Assim como no estudo, foi aplicado o SVM com kernel radial e $\gamma = 0.001$ para a regressão.

As estatísticas para as previsões de radiação estão na tabela 1. Foi obtido o erro médio absoluto (MAE), o erro médio absoluto percentual (MAE%) e a raiz do erro quadrático médio (RMSE) e seu valor percentual (RMSE%), de acordo com as respectivas equações (8), (9), (10) e (11). Em um conjunto de m observações horárias de previsão, X_t representa os valores reais e \hat{X}_t os valores previstos no modelo.

$$MAE = \frac{\sum_{t=1}^m |X_t - \hat{X}_t|}{m} \quad (8)$$

$$MAE(\%) = \frac{\sum_{t=1}^m |X_t - \hat{X}_t|}{\sum_{t=1}^m X_t} 100(\%) \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^m (X_t - \hat{X}_t)^2}{m}} \quad (10)$$

$$RMSE(\%) = \sqrt{\frac{\sum_{t=1}^m (X_t - \hat{X}_t)^2}{m} \frac{m \cdot 100(\%)}{\sum_{t=1}^m X_t}} \quad (11)$$

Tabela 1: Estatísticas de erro - Cenário 1

Subconjuntos	MAE	MAE(%)	RMSE	RMSE(%)
Erros de Treinamento	138	23.2	191	32.2
Erros de Teste	139	24.1	192	33.2



A medida de habilidade de previsão do modelo ou *forecast skill* s desta abordagem foi calculada de acordo com a equação (12).

$$s = 1 - \frac{RMSE_{model}}{RMSE_{persistence}} \quad (12)$$

Esta medida s é calculada através da comparação com um modelo base para avaliar quão boa é a abordagem proposta. Este modelo base é o modelo de persistência, representado por $RMSE_{persistence}$, e foi calculado considerando as previsões de radiação para o próximo dia como sendo exatamente iguais aos valores auferidos hoje. Após comparar ambos os modelos, obteve-se $s = 0.157$ neste cenário.

4.2. Cenário 2: prever k_d

Neste segundo cenário, foi tomado como premissa que a tarefa de prever o cluster para o próximo dia poderia ser realizada de forma bem precisa por terceiros. Com apenas quatro diferentes situações associadas ao padrão de nuvens, institutos meteorológicos locais, usando imagens de satélite, modelos de NWP e técnicas no estado-da-arte poderiam realizar esta tarefa de forma bem precisa. Portanto, foi assumida uma perfeita previsibilidade para os clusters e esta situação hipotética de previsão de clusters foi explorada. Assim, o modelo passa a contar apenas com os erros de previsão do índice diário de claridade da regressão por SVM.

As estatísticas para a previsão de radiação são apresentadas na tabela 2. Para este cenário, tem-se um aumento esperado na habilidade de previsão para $s = 0.232$, mostrando que a incerteza dos cluster impacta em 0.075 na magnitude de s .

Tabela 2: Estatísticas de erro - Cenário 2 (100% de previsibilidade de clusters)

Subconjunto	MAE	MAE(%)	RMSE	RMSE(%)
Erros de Treinamento	127	21.4	176	29.7
Erros de Teste	127	22.4	174	30.2

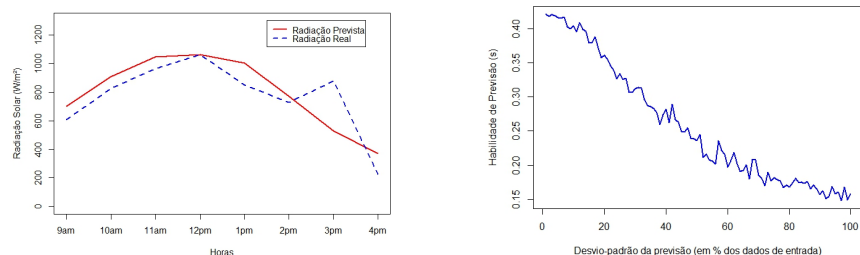
4.3. Cenário 3: prever k_d com variáveis de entrada do próximo dia

Neste cenário, foi avaliada em uma situação hipotética quão bom poderia ser as previsões de radiação se além dos centroides, as variáveis de entrada utilizadas no modelo de Jiménez-Pérez e Mora-López [2016] como temperatura, umidade e pressão atmosférica fossem os dados do dia da previsão e não do dia anterior. Em outras palavras, buscou-se avaliar neste cenário a habilidade máxima de previsão que este conjunto de dados de entrada poderia alcançar com uma previsão perfeita dos dados de amanhã e uso destes dados como entradas da previsão. As estatísticas são apresentadas na tabela 3. Quando a previsibilidade do conjunto de entrada aumenta, a habilidade de previsão atinge um máximo de $s = 0.415$ para os dados históricos de Belo Horizonte.

Tabela 3: Estatísticas de erro - Cenário 3 (100% de previsibilidade de clusters e da entrada)

Subconjuntos	MAE	MAE(%)	RMSE	RMSE(%)
Erros de Treinamento	99.5	16.8	135	22.7
Erros de Teste	97.6	16.9	133	23

A figura 7a mostra um exemplo no conjunto de testes da radiação prevista e da radiação realmente obtida no mesmo dia neste cenário, em um cluster com aumento da nebulosidade na parte da tarde. A figura 7b mostra como a habilidade de previsão s diminui à medida que esta previsão das variáveis de entrada de temperatura, umidade e pressão atmosférica se torna mais imprecisa. O comparativo começa com uma situação de 0% de desvios em relação aos valores auferidos ou 100% de previsibilidade e termina com um desvio-padrão da ordem de magnitude do valor absoluto da variável.



(a) Exemplo de radiação prevista e real no conjunto de teste para um dia

(b) Diminuição da habilidade de previsão (s) com o aumento da variabilidade da previsão das variáveis de entrada.

Figura 7: Previsão de Radiação e Impacto das Incertezas na Habilidade de Previsão (s)

5. Conclusão

Uma recente abordagem de previsão de radiação solar foi aplicada neste artigo para a cidade brasileira de Belo Horizonte, localizada em uma região de baixa a média latitude. Este modelo foi primeiro introduzido em Jiménez-Pérez e Mora-López [2016] e consiste em prever índices de claridade para o próximo dia como resultado de um processo de previsão de aprendizado de máquina de duas fases. Primeiro, quatro diferentes clusters de padrões de radiação são definidos no conjunto de dados. Estas observações são índices horários de claridade cuja tendência é removida pelos respectivos índices diários de claridade. Variáveis meteorológicas são utilizadas para treinar um SVM para prever o cluster do próximo dia ou qual tipo de tempo fará no próximo dia: limpo, nublado, nublado pela manhã ou nublado a tarde. A outra fase envolve prever o índice de claridade diário do próximo dia e adicioná-lo ao respectivo centroide previsto, obtendo uma estimativa do índice de claridade horário e, portanto, da radiação solar para o próximo dia.

O modelo no primeiro cenário apresentou uma habilidade de previsão de $s = 0.157$. Além disso, foram explorados outros dois cenários hipotéticos e avaliado como a habilidade de previsão cresce, mostrando o impacto de cada fase e do conjunto de dados de entrada na incerteza do modelo. O objetivo dessa simulação é traçar um limite superior e verificar a acurácia máxima que cada fase poderia atingir com o aumento da previsibilidade dos dados de entrada. No segundo cenário, foi assumido que clusters poderiam ter uma alta taxa de previsibilidade com apenas 4 situações ou classes a serem previstas. Observou-se que esta situação aumenta a habilidade de previsão para $s = 0.232$. No terceiro cenário, utilizou-se os valores das variáveis meteorológicas do dia posterior como dados de entrada, considerando que estes parâmetros pudessem ser perfeitamente previstos em uma situação hipotética. Para esta situação, obteve-se o valor de $s = 0.415$. Também verificou-se como é o comportamento decrescente na habilidade de previsão à medida que esses dados de variáveis de entrada vão se tornando cada vez menos precisos.

Para pesquisas futuras, pode-se avaliar diferentes configurações de conjuntos de entrada e seu impacto na habilidade de previsão (s), incluindo análises de como dados de outras estações meteorológicas poderiam ser utilizados para melhorar a habilidade de previsão do modelo para Belo Horizonte. Além disso, modelos que retratem e discutem os impactos no sistema elétrico da diminuição da incerteza na previsão de geração também devem ser considerados.

Agradecimentos

Esta pesquisa foi parcialmente financiada pela Fundação de Pesquisa de Minas Gerais (FAPEMIG), Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) do Ministério da Educação, e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) do Ministério da Ciência, Tecnologia, Inovações e Comunicações, Brasil. Também agradecemos ao Instituto Nacional de Meteorologia (INMET) do Ministério da Agricultura, Pecuária



e Abastecimento, por prover os dados utilizados nesta pesquisa da sua rede de estações meteorológicas automáticas.

Referências

- Aguiar, L. M., Pereira, B., David, M., Díaz, F., e Lauret, P. (2015). Use of satellite data to improve solar radiation forecasting with bayesian artificial neural networks. *Solar Energy*, 122:1309–1324.
- Ahmad, A., Anderson, T., e Lie, T. (2015). Hourly global solar irradiation forecasting for new zealand. *Solar Energy*, 122:1398–1408.
- Akarlan, E. e Hocaoglu, F. O. (2016). A novel adaptive approach for hourly solar radiation forecasting. *Renewable Energy*, 87:628–633.
- Aryaputera, A. W., Yang, D., Zhao, L., e Walsh, W. M. (2015). Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging. *Solar Energy*, 122:1266–1278.
- Bacher, P., Madsen, H., e Nielsen, H. A. (2009). Online short-term solar power forecasting. *Solar Energy*, 83(10):1772–1783.
- Besharat, F., Dehghan, A. A., e Faghih, A. R. (2013). Empirical models for estimating global solar radiation: A review and case study. *Renewable and Sustainable Energy Reviews*, 21:798–821.
- Cortes, C. e Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.
- Diagne, H. M., Lauret, P., e David, M. (2012). Solar irradiation forecasting: state-of-the-art and proposition for future developments for small-scale insular grids. In *WREF 2012-World Renewable Energy Forum*.
- Gala, Y., Fernández, Á., Díaz, J., e Dorronsoro, J. R. (2016). Hybrid machine learning forecasting of solar radiation values. *Neurocomputing*, 176:48–59.
- Goh, T. e Tan, K. (1977). Stochastic modeling and forecasting of solar radiation data. *Solar Energy*, 19(6):755–757.
- Hammer, A., Heinemann, D., Hoyer, C., e Lorenz, E. (2001). Satellite based short-term forecasting of solar irradiance: comparison of methods and error analysis. In *The 2001 EUMETSAT meteorological satellite data user's conference*, p. 677–684. Citeseer.
- Heinemann, D., Lorenz, E., e Girodo, M. (2006). Forecasting of solar radiation. *Solar energy resource management for electricity generation from local level to global scale*. Nova Science Publishers, New York.
- Hong, T. e Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3):914–938.
- Hong, T., Pinson, P., e Fan, S. (2014). Global energy forecasting competition 2012. *International Journal of Forecasting*, 30(2):357–363.
- Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., e Hyndman, R. J. (2016). Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting*, 32(3):896–913.
- Huang, J. e Perry, M. (2016). A semi-empirical approach using gradient boosting and k-nearest neighbors regression for gefcom2014 probabilistic solar power forecasting. *International Journal of Forecasting*, 32(3):1081–1086.



- Inman, R. H., Pedro, H. T., e Coimbra, C. F. (2013). Solar forecasting methods for renewable energy integration. *Progress in energy and combustion science*, 39(6):535–576.
- Iqbal, M. (2012). *An introduction to solar radiation*. Elsevier.
- Jain, A. K., Murty, M. N., e Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323.
- Jiménez-Pérez, P. F. e Mora-López, L. (2016). Modeling and forecasting hourly global solar radiation using clustering and classification techniques. *Solar Energy*, 135:682–691.
- Krishnamurti, T., Xue, J., Bedi, H., Ingles, K., e Oosterhof, D. (1991). Physical initialization for numerical weather prediction over the tropics. *Tellus B*, 43(4):53–81.
- Kuligowski, R. J. e Barros, A. P. (1998). Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks. *Weather and Forecasting*, 13(4):1194–1204.
- Li, J., Ward, J. K., Tong, J., Collins, L., e Platt, G. (2016). Machine learning for solar irradiance forecasting of photovoltaic system. *Renewable Energy*, 90:542–553.
- Lima, F. J., Martins, F. R., Pereira, E. B., Lorenz, E., e Heinemann, D. (2016). Forecast for surface solar irradiance at the brazilian northeastern region using nwp model and artificial neural networks. *Renewable Energy*, 87:807–818.
- Mangalova, E. e Agafonov, E. (2014). Wind power forecasting using the k-nearest neighbors algorithm. *International Journal of Forecasting*, 30(2):402–406.
- Maqsood, I., Khan, M. R., e Abraham, A. (2004). An ensemble of neural networks for weather forecasting. *Neural Computing & Applications*, 13(2):112–122.
- Martins, F. R. e Pereira, E. B. (2011). Enhancing information for solar and wind energy technology deployment in brazil. *Energy Policy*, 39(7):4378–4390.
- Marzban, C. e Stumpf, G. J. (1996). A neural network for tornado prediction based on doppler radar-derived attributes. *Journal of Applied Meteorology*, 35(5):617–626.
- Nagy, G. I., Barta, G., Kazi, S., Borbély, G., e Simon, G. (2016). Gefcom2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach. *International Journal of Forecasting*, 32(3):1087–1093.
- Oliveira, A. P., Escobedo, J. F., Machado, A. J., e Soares, J. (2002). Correlation models of diffuse solar-radiation applied to the city of sao paulo, brazil. *Applied Energy*, 71(1):59–73.
- Reikard, G. (2009). Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, 83(3):342–349.
- Sfetsos, A. e Coonick, A. (2000). Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. *Solar Energy*, 68(2):169–178.
- Silva, L. (2014). A feature engineering approach to wind power forecasting: Gefcom 2012. *International Journal of Forecasting*, 30(2):395–401.
- Smola, A. J. e Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222.
- Vapnik, V. e Chervonenkis, A. (1964). A note on one class of perceptrons. *Automation and remote control*, 25(1):103.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer-Verlag New York, Inc.