

Utilização de mapas de Kohonen para agrupamento e visualização de dados de consumo de energia elétrica do campus Medianeira da UTFPR.

Using Kohonen maps for clustering and visualization of the UTFPR Medianeira electricity consumption data.

RESUMO

Como consequência da democratização tecnológica, atualmente armazenamos uma quantidade muito grande de dados. Analisar tamanha quantidade de dados de maneira manual e em tempo para que as informações ainda sejam válidas é uma tarefa impossível, portanto é necessário utilizarmos a inteligência computacional para nos auxiliar nessa tarefa. Os mapas de Kohonen constituem uma poderosa ferramenta de análise de dados, principalmente em aplicações de clusterização. Utilizando a *toolbox* do *software Matlab* foi possível a partir dos dados de consumo de energia elétrica do campus Medianeira da UTFPR encontrar quatro padrões de consumo em função de determinadas características de temperatura ambiente, período do ano e horário do dia. A caracterização dos padrões de consumo permite que a universidade contrate fornecimento de energia junto a concessionária de maneira mais personalizada, evitando multas e desperdícios monetários.

PALAVRAS-CHAVE: Redes Neurais. Mapas Auto Organizáveis. Matlab.

ABSTRACT

As a result of technological democratization, we currently store a very large amount of data. Analyzing such amount of data manually and in time for the information can still be useful is an impossible task, so it is necessary to use computational intelligence to assist us in this task. Kohonen maps are a powerful data analysis tool, especially in clustering applications. Using Matlab's toolbox "SOM", the electricity consumption data of the UTFPR Medianeira campus was analyzed was possible to find four consumption patterns according to certain characteristics of ambient temperature, period of the year and time of day. The characterization of consumption patterns allows the university to contract energy supply from the electric power utility company in a more personalized way, avoiding fines and monetary waste.

KEYWORDS: Neural Networks. Self-Organizing Maps. Matlab.

Henrique Richetti Bonatto
henriquebonatto@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Medianeira, Paraná, Brasil

Leandro Antonio Pasa
leandropasa@gmail.com
Universidade Tecnológica Federal do Paraná, Medianeira, Paraná, Brasil

Recebido: 19 ago. 2020.

Aprovado: 01 out. 2020.

Direito autoral: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



INTRODUÇÃO

Com a democratização da tecnologia, armazenamos atualmente uma quantidade massiva de dados. É praticamente impossível analisar tamanha base de dados de forma manual com tempo útil para que a informação analisada ainda seja válida para aplicações, portanto, é necessário utilizarmos de artifícios matemáticos como a inteligência computacional para nos auxiliar nessa tarefa.

Com o objetivo de analisar a existência de padrões de consumo de energia elétrica no campus de Medianeira da UTFPR utilizou-se um algoritmo que aplica inteligência artificial para a análise dos dados de temperatura e consumo obtidos durante o ano de 2019. Mediante os conceitos teóricos do método elaborado por Teuvo Kohonen, o conjunto de dados foi aplicado na *toolbox* "SOM" disponível para o software *Matlab*. O algoritmo tem como intuito implementar de maneira simplificada os mapas auto organizáveis de Kohonen, que nada mais são que redes neurais não supervisionadas do tipo SOM (*Self-Organizing Map*), um método baseado na técnica de aprendizado competitivo e constitui uma ferramenta poderosa para análises de clusters e correlações existentes em conjuntos de dados.

Os conceitos que embasam os mapas de Kohonen, os procedimentos utilizados para parametrização da rede neural e os resultados obtidos serão abordados com maior detalhamento nas seções que seguem.

REFERÊNCIAL TEÓRICO

Teuvo Kohonen (1982, 1997), propôs um método de redes neurais que mapeia um espaço vetorial multidimensional em um conjunto de neurônios organizados em um arranjo unidimensional ou bidimensional, capaz de atuar como uma poderosa ferramenta de clusterização de conjuntos de dados.

Os mapas de Kohonen tem sua estrutura topológica dividida em dois conjuntos: o espaço de entrada e de saída, que são conectados por meio de ligações sinápticas. Os neurônios no espaço de saída estão dispostos em um *grid* (matriz-U) conectados de forma com que as relações com os neurônios vizinhos descrevam a topologia do mapa, podendo ser esta retangular ou hexagonal, a matriz-U é a representação dos valores médios dos pesos dos neurônios. O processo de formação dos mapas se dá por meio de quatro etapas: inicialização, processo competitivo, cooperativo e adaptação sináptica. Cada neurônio possui um vetor de pesos sinápticos, e durante a etapa de competição, os vetores de entrada são apresentados a rede de forma aleatória, assim os neurônios competem pelo direito de representar esse vetor por meio da distância Euclidiana, o vencedor, ou seja, o neurônio cujo vetor de pesos sinápticos mais se assemelha ao vetor de entrada é denominado BMU (*Best Matching Unit*). Na etapa de cooperação, as vizinhanças são formadas nas redondezas dos BMUs, onde os demais neurônios que se assemelham o BMU se agrupam, quanto mais próximos, mais semelhantes. Assim na etapa de adaptação, o neurônio vencedor e seus vizinhos tem os pesos sinápticos ajustados de forma a potencializar a semelhança com os vetores de entrada, quando mais próximo do BMU, mais os neurônios vizinhos serão ajustados por meio da equação (1), Marcio (2009, p.58).

$$w_j(n+1) = w_j(n) + \eta(n)h_{j,i(x)}(n)(x(n) - w_j(n)) \quad (1)$$

Em que $\eta(n)$ é a taxa de aprendizagem e $h_{j,i(x)}(n)$ é a função que representa a topologia escolhida para a vizinhança do BMU ($i(x)$), ambas as funções variam de acordo com as épocas de treinamento. Seja $h_{j,i(x)}$ uma função gaussiana do tipo:

$$h_{j,i(x)} = \exp\left(-\frac{d_{j,i}^2}{2\sigma^2}\right) \quad (2)$$

A função $h_{j,i(x)}$ decresce com o aumento da distância lateral $d_{j,i}$, tendendo a zero quando $d_{j,i} \rightarrow \infty$. σ é a função do raio de vizinhança, que decai com o passar do tempo, usualmente por meio de uma função exponencial.

MATERIAIS E MÉTODOS

Utilizando o *software Matlab*, pode-se realizar o processamento de conjuntos de dados com o método de Kohonen por meio da *toolbox SOM*. O algoritmo que se encontra na *toolbox* tem como intuito implementar os mapas de maneira simplificada. Os dados referentes as temperaturas registradas foram obtidos mediante consulta ao banco de dados do Simepar – Sistema Meteorológico do Paraná e os dados de consumo de energia foram obtidos diretamente da leitura da conta da concessionária de energia, Chevalier e Elis Weiss (2019, p.40). O conjunto de dados do estudo em questão é composto por quatro variáveis: dia do ano, temperatura média, minuto absoluto e consumo de energia, e as informações foram adquiridas em intervalos de 15 minutos entre 01/01/2019 a 18/09/2019.

Segundo Vesanto e Alhoniemi (2000), o tamanho do mapa pode ser estimado a partir da equação (3), em que T é a quantidade de neurônios e n o número de amostras.

$$T = 5 \cdot \sqrt{n} \quad (3)$$

Caso a quantidade de neurônios seja pequena (menor que 1000), Kaski (1997) propõe que o número de neurônios seja igual ao dos dados de entrada, no conjunto de dados em questão $T = 1580$, portanto maior que mil. Para Faria et al. (2010, p.16) a razão entre os lados do mapa deve ser igual à raiz quadrada da razão entre os dois maiores autovalores da amostra, contudo, autovalores só são encontrados em matrizes quadradas e o conjunto de dados utilizados no desenvolvimento deste trabalho não tem essa característica, assim para determinar a dimensão do mapa utilizou-se que a razão do lado maior pelo lado menor tem que resultar em 1, ou seja, um mapa quadrado, com tamanho de 40×40 . Considerou-se o raio inicial como sendo $\frac{1}{8}$ do lado maior e o raio final como padrão 1, Faria et al. (2010, p.16).

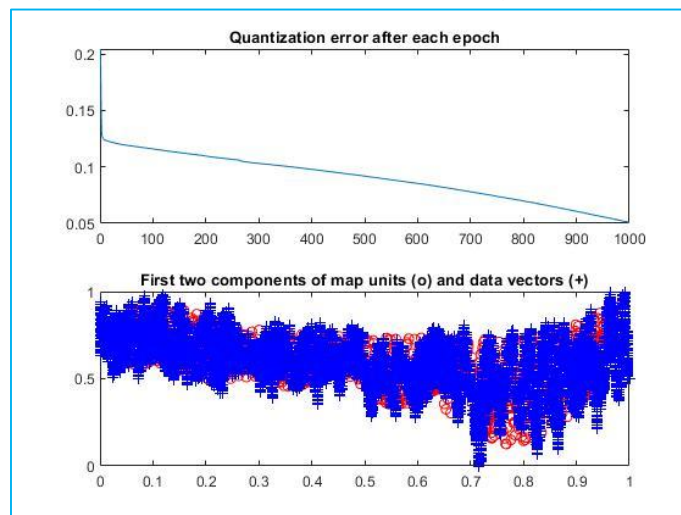
RESULTADOS E DISCUSSÕES

A partir dos parâmetros mencionados no capítulo anterior, a rede foi configurada e treinada. O processo de treinamento foi repetido 10 vezes, e em 80% dos treinamentos o resultado foi o mesmo. Por meio de figura 1, podemos verificar como ocorreu o processo iterativo do treinamento da rede e como os

neurônios do mapa se distribuíram de acordo com os vetores de entrada durante 1000 épocas de iteração. As circunferências vermelhas representam os neurônios do mapa treinado e as cruzes azuis representam os vetores de entrada. Observa-se que o erro decaiu de forma aguda logo nas primeiras iterações e foi estabilizando até chegar a: 0,0507. Analisando o erro de quantização (Qerr), que corresponde a uma métrica de avaliação do ajuste do mapa ao conjunto de dados de entrada, Miguel (2018, p.63), quanto menor o valor do erro, mais os vetores de pesos se assemelham com os vetores de entrada. O erro topográfico (Terr) é utilizado para verificar a qualidade em que o algoritmo SOM preserva a topologia do conjunto de dados, quanto menor, mais semelhante é o mapa gerado Miguel (2018, p.63).

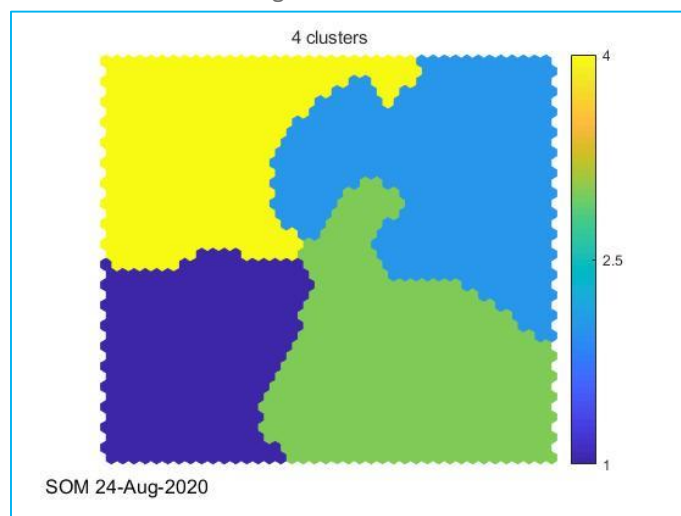
O erro topográfico do algoritmo aplicado ao conjunto de dados foi de 0,0356, indicando grande semelhança entre os vetores. Analisando a curva do erro na Figura 1, ele poderia ser diminuído aumentando as épocas de treinamento:

Figura 1 – Processo iterativo e erro de quantização.



Fonte: Elaborado pelo autor.

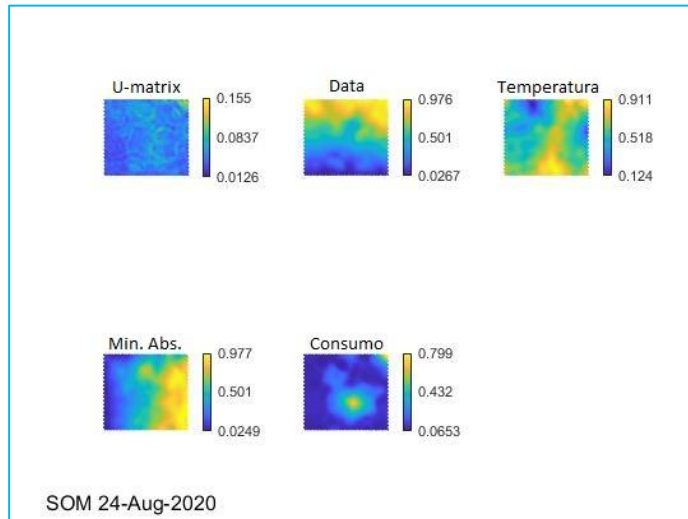
Figura 2 – Clusters.



Fonte: Elaborado pelo autor.

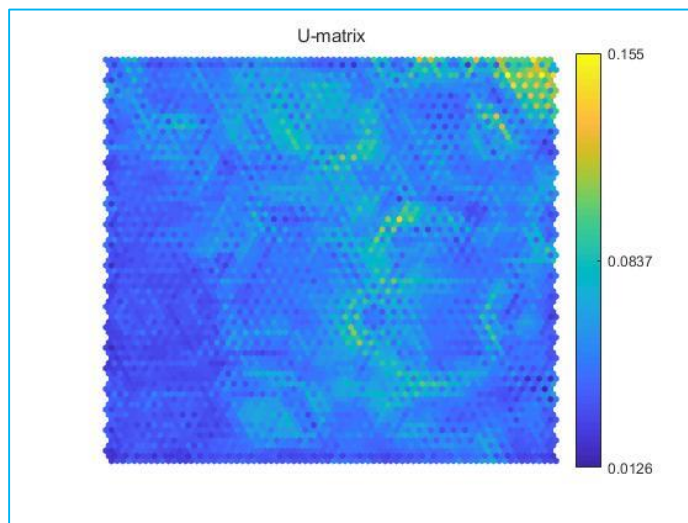
Na figura 2, podemos verificar que o algoritmo identificou a existência de 4 *clusters* para o conjunto de dados analisado, isso nos indica que existem 4 padrões de consumo de energia elétrica no campus. As características de cada um dos perfis de consumo podem ser analisadas a partir da figura 3.

Figura 3 –Mapas individuais das variáveis.



Fonte: Elaborado pelo autor.

Figura 4 –Matriz U.



Fonte: Elaborado pelo autor.

Quando comparamos os mapas das variáveis individuais na figura 3, com a Matriz U da figura 4 podemos notar por exemplo que os dois *clusters* superiores correspondem a padrões de consumos relacionados a dias da segunda metade do período analisado, possuindo um alto valor para dias do ano. Na tabela 1 podemos verificar quais as características dos 4 *clusters*, sendo o primeiro o localizado no canto superior esquerdo, o segundo localizado no canto superior direito, o terceiro no canto inferior direito e o quarto no canto inferior esquerdo.

Tabela 1 – Características dos *Clusters*.

Cluster	Dia do Ano	Temperatura Média	Minuto Absoluto	Consumo
1	MÉDIO	BAIXO	BAIXO	BAIXO
2	ALTO	MÉDIO	ALTO	MÉDIO
3	BAIXO	ALTO	ALTO	ALTO
4	BAIXO	MÉDIO	BAIXO	BAIXO

Fonte: Autoria própria (2020).

Pela tabela 1 é possível verificar que o cluster 3 é referente aos dias de verão na cidade de Medianeira, com ponto máximo de consumo nos horários logo após ao almoço. O *cluster 1* provavelmente representa o período no inverno em que ocorrem as férias escolares, o *cluster 2* os demais dias do ano após férias e o *cluster 4* os meses de outono na cidade.

Ao fim do processo de treinamento da rede e confecção dos gráficos, o algoritmo retorna também valores para o índice de Davies-Bouldin, nele observa-se que a posição no vetor com o menor número tem a maior probabilidade de corresponder ao número real de *clusters* do conjunto de entrada, nesse caso o menor valor encontra-se na posição 3 do vetor, correspondendo a 4 *clusters*.

CONCLUSÕES

A evolução acelerada da tecnologia computacional gera atualmente uma quantidade massiva de dados, portanto, é necessário que a análise desse volume de informação aconteça também de forma acelerada, realizando filtros para que o tempo de processamento não seja desperdiçado com variáveis que influenciam pouco ou nada nos resultados. Os mapas de Kohonen constituem uma poderosa ferramenta para análises de *clusters* e redução da dimensão dos conjuntos de dados.

A análise dos dados de consumo de energia elétrica no campus de Medianeira da UTFPR em função da temperatura no período de 01/01/2019 a 18/09/2019 foi apresentado nesse trabalho. As métricas de erro de quantização, erro topográfico são baixos o suficiente para afirmar-se que a rede neural conseguiu preservar a topologia dos dados de entrada e reproduziu similarmente os vetores de pesos sinápticos aos vetores de entrada. Foi encontrado quatro padrões de consumo de energia durante o período analisado e as características de cada padrão em função de suas variáveis.

É necessário para aprimoramento do trabalho analisar a possível existência de correlação entre o consumo energético e as estações do ano, considerando também o fluxo de pessoas utilizando as instalações da instituição, além de expandir o período de análise. Desse modo, a caracterização dos perfis de consumo será cada vez mais detalhada, possibilitando criar tendências para que os contratos de demanda de energia do campus com a concessionária possam ser dimensionados de maneira mais personalizada, evitando desperdícios por consumo abaixo o contratado e gastos com multas por consumo além do contratado.

REFERÊNCIAS

KOHONEN, Teuvo. **Self-Organizing Maps**. 3. ed. New York: Springer, 2000. 500 p.

CARVALHO, Clarissa Chevalier; WEISS, Elis Regina. **PREVISÃO DE CARGA UTILIZANDO REDES NEURAIAS ARTIFICIAIS PARA DEFINIÇÃO DO PONTO ÓTIMO DE CONTRATAÇÃO DE DEMANDA DE ENERGIA**. 2019. 65 f. TCC (Graduação) - Curso de Engenharia Elétrica, Engenharia Elétrica, Universidade Tecnológica Federal do Paraná, Medianeira, 2019. Cap. 3.

GONÇALVES, Márcio Leandro. **Métodos de Classificação Não-supervisionada de Imagens de Sensoriamento Remoto usando Mapas Auto-organizáveis de Kohonen**. 2009. 455 f. Tese (Doutorado) - Curso de Engenharia Elétrica, Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2009. Cap. 3.

VESANTO, Juha; ALHONIEMI, Esa. "Clustering of the Self-Organizing Map" *IEEE Trans. Neural Networks*, vol. 11, pp. XXX–XXX, May 2000.

KASKI, S., LAGUS, K.: Comparing self-organizing maps. In: Vorbrüggen, J.C., von Seelen, W., Sendhoff, B. (eds.) *Artificial Neural Networks - ICANN 96*. LNCS, vol. 1112, pp. 809–814. Springer, Heidelberg (1996).

FARIA, Elisângela Lopes de; ALBUQUERQUE, Marcelo Portes; AFONSO, Jorge Luiz Gonzalez; ALBUQUERQUE, Márcio Portes; CAVALCANTE, Jose Thadeu Pinto. **Introdução ao Toolbox de redes neurais de kohonen**. Rio de Janeiro: Cbpf, 2010. 25 p.

DAVIES, D. L., D. W. BOULDIN. "A Cluster Separation Measure" *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. PAMI-1, No. 2, 1979, pp. 224–227.