



APLICAÇÕES DO ÍNDICE SILHOUETTE DE VALIDAÇÃO INTERNA EM ANÁLISES DE AGRUPAMENTO DE SÉRIES TEMPORAIS: UMA REVISÃO DE LITERATURA

Inácio Puntel¹, Débora Spenassato², Ana Maria V. de Azambuja³, Raquel da F. Nicolette²

¹Mestrando em Ambientometria – IMEF/FURG, Rio Grande, Brasil
(rbkpuntel@gmail.com)

²Instituto de Matemática, Estatística e Física - IMEF/FURG, Rio Grande, Brasil

³Escola de Engenharia - EE/FURG, Rio Grande, Brasil

Resumo: Aplicações de algoritmos de agrupamento a séries temporais têm crescido nos últimos anos, e diversos estudos buscaram avaliar as metodologias empregadas. Porém, a fase de validação das análises tem recebido pouca atenção. Este trabalho tem por objetivo apresentar os resultados de uma revisão bibliográfica sobre as aplicações do Índice Silhouette no contexto da análise de agrupamento de séries temporais. Os resultados apontam que o índice não tem sido utilizado conforme sua proposta original.

Palavras-chave: Análise de cluster; Séries temporais; Índice Silhouette; Índices de validação.

INTRODUÇÃO

Análise de agrupamento é um conjunto de técnicas de mineração de dados que têm ganhado notoriedade nas últimas décadas (AGHABOZORGI ET AL., 2015; ARBELAITZ ET AL., 2013). A aplicação dessas técnicas à análise de séries temporais tem se tornado cada vez mais acessível com os avanços recentes em poder de processamento e armazenamento de dados (AGHABOZORGI ET AL., 2015; JAVED ET AL., 2020).

Nas ciências ambientais, o uso dessas técnicas tem se mostrado útil nos mais diversos contextos, com aplicações variando desde a descoberta de índices climáticos à identificação de anomalias em registros de atividade sísmica (STEINBACH ET AL., 2003; SHUMWAY, 2003). Os objetivos dessas aplicações geralmente caem em uma de duas categorias: a) identificação de padrões que possam ser usados para extração de informações sobre as variáveis de interesse e b) detecção de anomalias ou características inesperadas (AGHABOZORGI ET AL., 2015). A implementação dessas aplicações, por sua vez, varia em termos dos métodos utilizados, porém, podem ser resumidas em cinco componentes: método de representação, medida de distância, algoritmo de agrupamento, definição dos protótipos e avaliação (AGHABOZORGI ET AL., 2015). É sobre este último componente que este trabalho se debruça.

Esta avaliação é geralmente feita através de dois tipos de índices: índices de validação externa e de validação interna. A escolha entre os dois tipos de índices é feita principalmente baseada na

disponibilidade de rótulos de classe que permitam comparar os resultados com uma verdade exterior (AGHABOZORGI ET AL., 2015; ARBELAITZ ET AL., 2013; JAVED ET AL., 2020).

Os índices de validação externa avaliam os resultados comparando as partições produzidas pelo algoritmo com as partições “reais” do conjunto de dados, geralmente definidas por especialistas humanos (AGHABOZORGI ET AL., 2015; ARBELAITZ ET AL., 2013; JAVED ET AL., 2020). Existem diversos índices de validação que se encaixam nessa categoria (*e.g.* Adjusted Rand Index, Adjusted Mutual Information, Jaccard Score, F-measure etc.) – um sumário pode ser encontrado em Aghabozorgi *et al.*, (2015) – mas, por estarem fora do escopo deste trabalho, não será feita uma descrição mais detalhada.

Já os índices de validação interna são usados quando esses parâmetros externos estão indisponíveis, fazendo uso então de parâmetros internos aos resultados, como a homogeneidade intragrupos e a heterogeneidade entre grupos. Dado que geralmente esses parâmetros são desconhecidos, tais índices acabam sendo os mais frequentemente usados. Assim como os índices de validação externa, há vários índices que se encaixam nessa categoria. Em um experimento recente, Arbelaitz *et al.* (2013) avaliaram o desempenho de 30 destes índices tanto com conjuntos de dados sintéticos quanto naturais. Porém, para fins desta revisão, apenas a aplicação do Índice Silhouette às análises de agrupamento de séries temporais será investigada.



O Índice Silhouette foi proposto por Peter J. Rousseeuw como uma ferramenta gráfica para auxiliar a interpretação e validação de análises de agrupamento (ROUSSEEUW, 1987). Os escores do índice são calculados a partir de uma partição obtida através de um algoritmo de agrupamento utilizado e das distâncias entre os objetos. Para cada objeto no banco de dados é calculado um valor S_i entre -1 e 1, e um gráfico é produzido pela combinação desses valores (ROUSSEEUW, 1987). O valor S_i é obtido através da equação 1.

$$S_i = \frac{b_i - a_i}{\max \{a_i, b_i\}} \quad (1),$$

onde a_i representa a distância média entre o objeto i e todos os outros objetos pertencentes ao seu grupo e b_i representa a distância média entre o objeto i e todos os outros objetos de seu grupo vizinho. O grupo vizinho é encontrado calculando a distância média de i em relação aos objetos de cada grupo diferente do seu e escolhendo a menor distância média. Isto é, b_i representa a segunda melhor opção de alocação do objeto i dada uma partição com K grupos. Ou seja, o uso do Índice Silhouette requer a existência de pelo menos dois grupos (ROUSSEEUW, 1987).

Os valores S_i podem então ser plotados em ordem decrescente para cada grupo, formando uma silhueta, conforme a figura 1. De modo geral, silhuetas mais largas apontam grupos bem definidos.

Valores S_i próximos a 1 sugerem que o objeto i está alocado a um grupo semelhante a si. Valores próximos a 0 sugerem que o objeto i está na borda de seu grupo, e que sua distância média em relação ao grupo vizinho não é grande o suficiente para dar certeza de seu grupo “real”. Valores próximos a -1 sugerem que o objeto i é em média mais parecido com seu grupo vizinho que com membros do grupo ao qual foi alocado, indicando que uma realocação pode ser necessária.

A partir disso pode ser computada também a *largura média* da silhueta de cada grupo, que nada mais é que a média dos valores S_i para cada grupo. Esta medida permite comparar o quão bem definido cada grupo está em uma determinada solução. De forma semelhante, pode ser computada a *largura média total* da silhueta através da média dos valores S_i para o conjunto de dados inteiro. É devido a esta característica do Índice que ele pode ser usado para estimar o número K que melhor representa os grupos naturais existentes no conjunto de dados. Isto é feito computando a *largura média total* para diferentes valores de K e identificando qual produz a silhueta média mais larga.

Porém, Rousseeuw (1987) não recomenda que a *largura média total* seja utilizada unicamente para a tomada de decisão, recomendando que os gráficos sejam consultados para entender esses resultados. Paralelamente, o autor exemplifica cenários em que independentemente do valor de K , a maior *largura média total* obtida continua sendo baixa. Nesses casos, a recomendação é que a escolha da análise de agrupamento como método de análise de dados seja revista (ROUSSEEUW, 1987).

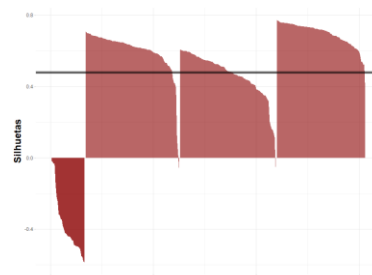


Figura 1. Gráfico de silhuetas com quatro grupos.

Como ainda não há um algoritmo de agrupamento que produza consistentemente resultados ótimos em diferentes contextos (PAL, BISWAS, 1997; ARBELAITZ ET AL., 2013; JAVED ET AL., 2020), a avaliação dos grupos produzidos acaba tendo um papel fundamental na interpretação dos resultados obtidos. Porém, a inexistência de uma teoria que oriente o processo de avaliação torna esta uma tarefa difícil (ARBELAITZ ET AL., 2013; BRUN ET AL., 2007). O fato de que, assim como os algoritmos de agrupamento, não há um índice de validação que apresente performance consistentemente melhor que a dos outros (ARBELAITZ ET AL., 2013) também contribui para as dificuldades deste processo.

Para lidar com esses problemas, passou-se a adotar a prática de utilizar diversos índices em conjunto, buscando pontos de concordância como indício de otimização (ARBELAITZ ET AL., 2013). Apesar de engenhosa, esta abordagem ainda sofre com o fato de nem todos os índices serem comparáveis entre si, ou mesmo comparáveis consigo mesmo em diferentes contextos (ARBELAITZ ET AL., 2013). Ainda que existam experimentos que avaliem o desempenho dos índices de validação interna em análises de agrupamento convencionais (ARBELAITZ ET AL., 2013; BRUN ET AL., 2007; DURAN-ROSAL ET AL., 2018), estes trabalhos focam na acurácia do uso destes índices para a estimação do parâmetro K em algoritmos de agrupamento nos quais o número de grupos deve ser provido *a priori*, oferecendo poucos *insights* sobre a interpretação dos escores obtidos através desses índices. Dentro do contexto da análise de agrupamento de séries temporais, não foram



encontrados estudos desse cunho à época da escrita deste artigo.

Tendo em vista as questões abordadas acima, o objetivo desta revisão de literatura é identificar as aplicações que têm sido dadas ao Índice Silhouette de validação interna no contexto da análise de agrupamento de séries temporais, buscando também identificar estudos que possam atestar a acurácia do índice para esta aplicação. Justifica-se o estudo pela necessidade de se analisar a adequação dos métodos de validação utilizados, cuja preocupação também é expressa na literatura sobre análise de agrupamento de séries temporais, de forma a avançar o conhecimento metodológico sobre o tema.

MATERIAL E MÉTODOS

Para a realização desta revisão, buscou-se artigos científicos nas bases de dados ScienceDirect, SagePub, Wiley Online Library e SciELO. As palavras-chave usadas foram “Silhouette Index” AND “time series clustering”, em qualquer lugar do texto. Na base de dados SciELO, buscou-se também as palavras “Índice Silhouette” AND “agrupamento de séries temporais”.

Ao todo foram encontrados 31 artigos na base de dados Science Direct; 5 na Wiley Online Library; e 0 em ambos SagePub e SciELO, publicados até 2021. Foi realizada a leitura dos resumos dos artigos selecionados e, apesar de dois dos artigos encontrados se tratarem de revisões de literatura, optou-se por mantê-los para a leitura completa. Passou-se, então, para a leitura completa dos artigos, a partir da qual foram excluídos os estudos que apenas mencionavam as palavras-chave utilizadas, sem efetivamente tratar do tema.

A seguir da leitura completa dos 36 artigos, foram excluídos 11 artigos por não se encaixarem no escopo da revisão, isto é, não aplicarem o Índice Silhouette ao contexto da análise de agrupamento de séries temporais, resultando num total de 25 artigos selecionados para a análise.

Foram então computadas as seguintes características: revista onde o artigo foi publicado; área do conhecimento, segundo as palavras-chave escolhidas; aplicação do índice como estimador do parâmetro K ; aplicação do índice para validação das soluções produzidas; utilização do gráfico de silhuetas; e uso de outros índices em conjunto.

RESULTADOS E DISCUSSÃO

Ao longo desta seção será apresentada uma sumarização dos resultados obtidos, seguida de uma discussão dos resultados dos trabalhos que melhor ajudaram a responder os objetivos desta revisão. A tabela 1 apresenta a distribuição dos artigos de acordo com a área do conhecimento a qual

pertencem. A classificação foi feita através dos objetivos dos estudos.

Como pode ser observado, a área do conhecimento melhor representada foi a de Energia, reunindo a maioria dos artigos encontrados. Os artigos que compuseram esta categoria ocupam-se da construção de modelos para previsão da produção de energia elétrica (AZIMI ET AL., 2016), predição dos níveis de radiação solar (WU, CHAN, 2013) e classificação dos padrões de consumo de energia elétrica (MANIAR, ABHYANKAR, 2019; BISCARRI ET AL., 2017; BOURDEAU ET AL., 2020; FERREIRA ET AL., 2013; FERREIRA ET AL., 2015; LIU ET AL., 2020; RAJABI ET AL., 2019; RUIZ ET AL., 2020; TURECZEK ET AL., 2019; WEN ET AL., 2019; YANG ET AL., 2017).

A Tabela 2, por sua vez, apresenta a sumarização das revistas nas quais os artigos selecionados foram encontrados, e mostra a presença de revistas focadas na publicação de trabalhos da área de energia, sistemas e ciências ambientais. A Tabela 2 também apresenta os anos de publicação dos artigos analisados. É possível verificar que nenhum dos artigos precede o ano de 2013, apesar de não ter sido estabelecido nenhum limite temporal na busca dos artigos. Isso, provavelmente, é um reflexo do crescimento recentemente observado no uso de análises de agrupamento como um método de análise exploratória de séries temporais nos últimos anos.

Tabela 1. Área do conhecimento dos artigos selecionados

Área	Total
Energia	13
Computação	3
Oceanografia	2
Recursos Hídricos	2
Geologia	1
Finanças	1
Paleontologia	1
Estatística	1
Comp. Consumidor	1

No que diz respeito as aplicações dadas ao índice, 19 dos 25 artigos analisados utilizaram o índice para avaliação e validação dos resultados, 10 artigos o utilizaram tanto para validar os resultados quanto



como estimador do parâmetro K (número de grupos), e 6 artigos utilizaram o índice apenas como estimador de K . No total, apenas 4 artigos utilizaram as silhuetas como uma ferramenta gráfica (AVNI ET AL., 2015; FERREIRA ET AL., 2013; FERREIRA ET AL., 2015; SERRA, ZÁRATE, 2015) para a visualização da qualidade dos grupos. A maior parte dos artigos fez uso apenas da largura média total como medida de qualidade para um determinado valor de K . Dentre os 16 artigos que utilizaram o índice como estimador do parâmetro K , 15 o fizeram apenas através da largura média total da silhueta para os diferentes valores de K , desconsiderando a largura média das silhuetas dos grupos formados no processo de escolha. Serra e Zárate (2015), apesar de utilizarem o gráfico de silhuetas para ilustrar o resultado de uma solução, não o consideraram na estimação do parâmetro K . Por sua vez, Avni et al. (2015) não somente utilizaram o gráfico de silhuetas na estimação de K , como apontaram que o número de grupos sugerido pela maximização da silhueta média falhou em produzir grupos suficientemente distintos, optando-se por procurar por K na vizinhança do valor sugerido.

Suominen (2018), em seu trabalho sobre análise de agrupamento de séries temporais obtidas através de imagens de satélite de águas costeiras, utilizou o Índice Silhouette para estimar o número ótimo de grupos K para os dados. O autor encontrou que independentemente do valor K , a largura média total das silhuetas não alcançou um valor alto, tampouco variou substancialmente (SUOMINEN, 2018). Esse tipo de comportamento da largura média total das silhuetas é tido como uma evidência de ausência de uma estrutura de grupos no conjunto de dados analisado, sendo sugerida a inspeção visual das silhuetas individuais para buscar as causas desse resultado e possivelmente justificar a busca de uma outra abordagem e metodologia de análise (ROUSSEUW, 1987). De fato, apesar de o autor não apresentar uma inspeção das silhuetas, uma mudança na metodologia adotada revelou que a estrutura de grupos contida nos dados não favorecia uma divisão em grupos no sentido clássico, mas ainda assim grupos podiam ser encontrados (SUOMINEN, 2018).

Biscarri et al. (2017), por outro lado, ao comparar o desempenho de diversos algoritmos na classificação de consumidores de energia elétrica, apresentou resultados nos quais o número ótimo de grupos K sugerido pela maximização da largura média total das silhuetas concorda com o número sugerido por especialistas humanos, produzindo grupos com padrões de consumo distintos. Após análise dos resultados, os autores chegaram à conclusão de que os grupos produzidos separavam os consumidores entre noturnos, diurnos e comerciais (BISCARRI et

al., 2017), porém, não foram realizadas análises mais extensas que investigassem as soluções produzidas por outros valores de K , como feito em Avni et al. (2015).

Duran-Rosal et al. (2018), em um experimento no qual, dentre outros objetivos, buscava identificar entre um grupo de índices de validação interna qual obtinha o melhor desempenho no contexto da segmentação de séries temporais, encontraram que o Índice Silhouette apresentou o desempenho mais consistente quando aplicado a uma série sintética, resultado semelhante ao encontrado por Arbelaiz et al. (2013) no contexto de análises de agrupamento convencionais. A partir desse resultado, os autores escolheram o Silhouette como índice de validação interna para suas análises com séries temporais reais sem repetir a comparação feita com a série sintética. Tendo em vista que no artigo de Arbelaiz et al. (2013) foi encontrado que o desempenho do índice decaía quando aplicado a séries reais, as afirmações de Duran-Rosal et al. (2018) sobre o Índice Silhouette devem ser tomadas com precaução.

Ferreira et al. (2013) utilizaram o Índice Silhouette para avaliar as soluções produzidas pelo algoritmo de classificação proposto no artigo em comparação com o algoritmo Fuzzy C-Means. A largura média total das silhuetas foi utilizada para a comparação em conjunto com o gráfico das silhuetas para inspeção visual. Porém, apesar do gráfico de silhuetas revelar apenas um grupo bem definido e uma largura média total relativamente baixa (0,28) – ambos sendo fatores que sugerem a necessidade de uma investigação mais aprofundada (ROUSSEUW, 1987) –, os resultados obtidos pelo índice são tidos como atestado da qualidade do algoritmo sem a realização de testes adicionais.

Avni et al. (2015) utilizaram o Índice Silhouette para estimar o número ótimo de grupos K em análises de agrupamento de séries temporais no contexto da modelagem da demanda de água em Israel. Os autores encontraram resultados particularmente interessantes, segundo os quais o número ótimo de grupos K sugerido pela maximização da largura média total das silhuetas produziu grupos compostos por séries heterogêneas, enquanto uma solução com largura média total significativamente mais baixa (0,411 para 0,277) produziu grupos com padrões de consumo melhor definidos (AVNI ET AL., 2015).

Em relação ao uso de outros índices de validação interna em conjunto com o Silhouette, os mais frequentes foram o Índice Dunn, aparecendo em 48% dos artigos, o Índice Daves-Bouldin, aparecendo em 43% dos artigos e o Índice Calinski & Harabasz, que apareceu em 33% dos artigos.



Tabela 2. Revistas onde os artigos selecionados foram publicados.

Revista	N. de Artigos	Artigos
Expert Systems with Applications	4	Biscarri et al., 2017; Iorio et al., 2017; Ruiz et al., 2020; Serra, Zárate, 2015
Energy & Buildings	4	Bourdeau et al., 2020; Liu et al., 2021; Tureczek et al., 2019; Yang et al., 2017
Energy Conversion and Management	2	Azimi et al., 2016; Wu, Chan, 2013
Electrical Power and Energy Systems	2	Ferreira et al., 2013; Ferreira et al., 2015
Neurocomputing	2	Langone et al., 2014; Pérez-Ortiz et al., 2017
Journal of Applied Geophysics	1	Delforge et al., 2020
Information Sciences	1	Durán-Rosal, 2018
Renewable and Sustainable Energy Reviews	1	Rajabi et al., 2020
Ecological Indicators	1	Suominen, 2018
Fuzzy Sets and Systems	1	Vilar et al., 2018
Journal of Cleaner Production	1	Wen et al., 2019
Water Research	1	Xing, Sela, 2019
Environmetrics	1	Alvarez-Esteban, 2016
Water Resources Research	1	Avni et al., 2015
IET Smart Grid	1	Maniar, Abhyankar, 2019
Expert Systems	1	Wilson et al., 2017

CONCLUSÃO

Os resultados obtidos por esta revisão serviram para dar uma ideia geral das aplicações do Índice Silhouette de validação interna. A maior parte dos artigos avaliados o utiliza para avaliação das soluções produzidas pelos algoritmos de agrupamento utilizados e apenas um quarto dos estudos que compõe esta revisão o utilizam somente para a estimação do número de grupos K existentes no conjunto de dados. Pouco menos da metade o utiliza para ambos os fins. Porém, a maioria dos estudos compartilha a ausência de uma discussão sobre o significado dos escores obtidos através do índice. De fato, a prática de comparar tais escores com os de outros índices, sem discutir nenhum, foi amplamente encontrada entre os artigos selecionados ao invés de uma interpretação dos resultados do Índice.

Paralelamente, poucos artigos tiveram a preocupação de justificar a escolha do Índice Silhouette (ou de qualquer dos índices utilizados) em função de características dos dados analisados, refletindo a falta de uma teoria que oriente o processo de avaliação dos resultados de análises de agrupamento (BRUN ET AL., 2007). Tampouco foram abordadas, na interpretação dos resultados, as limitações do Índice e as cautelas expressas no artigo original (ROUSSEUW, 1987).

A utilização das silhuetas como uma ferramenta gráfica encontrou pouquíssimo espaço entre os artigos selecionados, cedendo espaço para o uso da largura média total como uma medida de qualidade geral da solução produzida. Ou seja, o gráfico de silhuetas como recurso para auxiliar a interpretação dos resultados parece ter caído em



desuso no contexto da análise de agrupamento de séries temporais. De modo geral, o cenário atual apresenta o Índice apenas como uma medida genérica da qualidade dos agrupamentos produzidos.

Em conclusão, os resultados obtidos por esta revisão apontam que o Índice Silhouette de validação interna não apresenta desempenho consistente ao longo dos diferentes contextos encontrados, reforçando a necessidade de experimentos que investiguem sistematicamente o desempenho do Índice sob diferentes condições experimentais, a fim de identificar em que contextos os escores obtidos podem efetivamente ser considerados indicadores da qualidade das soluções produzidas pelos algoritmos de agrupamento.

AGRADECIMENTOS

Agradecimentos são devidos à Fundação de Amparo à Pesquisa do Estado do Rio Grande de Sul – FAPERGS, pelo financiamento do projeto de dissertação do qual faz parte esta revisão.

REFERÊNCIAS

- AGHABOZORGI, Saeed et al. Time-series clustering—a decade review. **Information Systems**, v. 53, p. 16-38, 2015.
- ALVAREZ-ESTEBAN, Pedro C. et al. Time series clustering using the total variation distance with applications in oceanography. **Environmetrics**, v. 27, n. 6, p. 355-369, 2016.
- ARBELAITZ, Olatz et al. An extensive comparative study of cluster validity indices. **Pattern Recognition**, v. 46, n. 1, p. 243-256, 2013.
- AVNI, Noa et al. Water consumption patterns as a basis for water demand modeling. **Water Resources Research**, v. 51, n. 10, p. 8165-8181, 2015.
- AZIMI, R. et al. A hybrid wind power forecasting model based on data mining and wavelets analysis. **Energy conversion and management**, v. 127, p. 208-225, 2016.
- BISCARRI, Felix et al. Electricity clustering framework for automatic classification of customer loads. **Expert Systems with applications**, v. 86, p. 54-63, 2017.
- BOURDEAU, Mathieu et al. Classification of daily electric load profiles of non-residential buildings. **Energy and Buildings**, v. 233, p. 110670, 2021.
- BRUN, Marcel et al. Model-based evaluation of clustering validation measures. **Pattern recognition**, v. 40, n. 3, p. 807-824, 2007.
- DELFORGE, Damien et al. Time-series clustering approaches for subsurface zonation and hydrofacies detection using a real time-lapse electrical resistivity dataset. **Journal of Applied Geophysics**, v. 184, p. 104203, 2021.
- DURÁN-ROSAL, Antonio Manuel et al. Simultaneous optimisation of clustering quality and approximation error for time series segmentation. **Information Sciences**, v. 442, p. 186-201, 2018.
- FERREIRA, Adonias M. S. et al. A new method for pattern recognition in load profiles to support decision-making in the management of the electric sector. **International Journal of Electrical Power & Energy Systems**, v. 53, p. 824-831, 2013.
- FERREIRA, Adonias M. S. et al. Pattern recognition as a tool to support decision making in the management of the electric sector. Part II: A new method based on clustering of multivariate time series. **International Journal of Electrical Power & Energy Systems**, v. 67, p. 613-626, 2015.
- IORIO, Carmela et al. A P-spline based clustering approach for portfolio selection. **Expert Systems with Applications**, v. 95, p. 88-103, 2018.
- JAVED, Ali et al. A benchmark study on time series clustering. **Machine Learning with Applications**, v. 1, p. 100001, 2020.
- LANGONE, Rocco et al. Incremental kernel spectral clustering for online learning of non-stationary data. **Neurocomputing**, v. 139, p. 246-260, 2014.
- LIU, Xue et al. A data mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data. **Energy and Buildings**, v. 231, p. 110601, 2021.
- MANIAR, Mudassir A.; ABHYANKAR, Abhijit R. Validity index based improvisation in reproducibility of load profiling outcome. **IET Smart Grid**, v. 2, n. 1, p. 131-139, 2019.
- PAL, Nikhil R.; BISWAS, J. Cluster validation using graph theoretic concepts. **Pattern Recognition**, v. 30, n. 6, p. 847-857, 1997.
- PÉREZ-ORTIZ, María et al. On the use of evolutionary time series analysis for segmenting paleoclimate data. **Neurocomputing**, v. 326, p. 3-14, 2019.
- RAJABI, Amin et al. A comparative study of clustering techniques for electrical load pattern segmentation. **Renewable and Sustainable Energy Reviews**, v. 120, p. 109628, 2020.
- ROUSSEEUW, Peter J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of computational and applied mathematics**, v. 20, p. 53-65, 1987.



- RUIZ, Luis G. Baca et al. A time-series clustering methodology for knowledge extraction in energy consumption data. **Expert Systems with Applications**, v. 160, p. 113731, 2020.
- SERRA, Ana P.; ZÁRATE, Luis E. Characterization of time series for analyzing of the evolution of time series clusters. **Expert systems with applications**, v. 42, n. 1, p. 596-611, 2015.
- SHUMWAY, Robert H. Time-frequency clustering and discriminant analysis. **Statistics & probability letters**, v. 63, n. 3, p. 307-314, 2003.
- STEINBACH, Michael et al. Discovery of climate indices using clustering. In: **Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining**, 2003. p. 446-455.
- SUOMINEN, Tapio. Applying MERIS time series and dynamic time warping for delineating areas with similar temporal behaviour in the northern Baltic Sea. **Ecological Indicators**, v. 95, p. 794-804, 2018.
- TURECZEK, Alexander Martin et al. Clustering district heat exchange stations using smart meter consumption data. **Energy and buildings**, v. 182, p. 144-158, 2019.
- VILAR, José A. et al. Quantile autocovariances: a powerful tool for hard and soft partitional clustering of time series. **Fuzzy Sets and Systems**, v. 340, p. 38-72, 2018.
- WEN, Lulu et al. A shape-based clustering method for pattern recognition of residential electricity consumption. **Journal of cleaner production**, v. 212, p. 475-488, 2019.
- WILSON, Jobin et al. Clustering short temporal behaviour sequences for customer segmentation using LDA. **Expert Systems**, v. 35, n. 3, p. e12250, 2018.
- WU, Ji; CHAN, Chee Keong. Prediction of hourly solar radiation with multi-model framework. **Energy conversion and management**, v. 76, p. 347-355, 2013.
- XING, Lu; SELA, Lina. Unsteady pressure patterns discovery from high-frequency sensing in water distribution systems. **Water research**, v. 158, p. 291-300, 2019.
- YANG, Junjing et al. k-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. **Energy and Buildings**, v. 146, p. 27-37, 2017.