

ARTIGO ORIGINAL

Otimização inspirada nas ondas aquáticas aplicada ao agrupamento não-hierárquico de objetos

Optimization inspired by water waves applied to non-hierarchical grouping of objects

Alfredo Silveira Araújo Neto ¹

¹Techway Informática Ltda.

*alfredosilveira@yahoo.com.br

Recebido: 20/07/2021. Revisado: 19/04/2022. Aceito: 30/04/2022.

Resumo

Caracterizado como uma das condutas relacionadas à análise exploratória de dados, um método de agrupamento não hierárquico consiste em um procedimento apto a classificar uma coleção de objetos em um subconjunto finito de grupos ou classes, de forma que os objetos que pertencem a um grupo são mais semelhantes entre si do que os objetos compreendidos por um grupo distinto. Nesta circunstância, este estudo propõe a aplicação de uma meta-heurística inspirada no comportamento das ondas aquáticas, à determinação do agrupamento não hierárquico de objetos, e compara os resultados obtidos por este método com as respostas alcançadas por outras seis estratégias de agrupamento. De modo específico, os erros quadráticos auferidos pelo algoritmo sugerido, ao classificar 29 coleções de referência foram, por intermédio do teste dos postos sinalizados de Wilcoxon, confrontados com os resultados obtidos pelas meta-heurísticas enxame de partículas, algoritmo genético, colônia de abelhas artificiais e recozimento simulado, e com as respostas determinadas pelo algoritmo *K-means* e por uma variação da meta-heurística inspirada nas ondas aquáticas que incluía um operador de decaimento, indicando que, na maioria das circunstâncias, o algoritmo proposto foi capaz de obter classificações mais congruentes do que as estabelecidas pelos outros métodos de particionamento.

Palavras-Chave: Análise por agrupamentos; computação bioinspirada; meta-heurística

Abstract

Characterized as one of the strategies related to exploratory data analysis, a non-hierarchical grouping method consists of a procedure able to classify a collection of objects into a finite subset of groups or classes, such that objects belonging to a group are more similar to each other than objects comprised by a distinct group. In this circumstance, this study proposes the application of a meta-heuristic inspired by the behavior of water waves, to the determination of the non-hierarchical grouping of objects, and compares the results obtained by this method with the responses achieved by six other grouping strategies. Specifically, the squared errors obtained by the suggested algorithm, when classifying 29 benchmark collections were, through the Wilcoxon signed-rank test, compared with the results obtained by the meta-heuristics particle swarm, genetic algorithm, artificial bee colony and simulated annealing, and with the responses determined by the *K-means* algorithm and by a variation of the meta-heuristic inspired by the water waves that included a decay operator, indicating that, mostly, the proposed algorithm was able to obtain more congruent classifications than those established by other partitioning methods.

Keywords: Bio-inspired computing; cluster analysis; metaheuristic

1 Introdução

Numerosas disciplinas admitem a prática de classificar objetos de acordo com a similaridade percebida entre os mesmos como uma de suas premissas essenciais, sob a perspectiva de que organizar dados em grupos de elementos semelhantes, representa uma das condutas mais apropriadas em proporcionar a compreensão e a aquisição de conhecimento. A análise de agrupamentos consiste no estudo de métodos aptos a classificar ou agrupar objetos, os quais podem ser caracterizados por intermédio de uma coleção de medidas ou atributos, ou ainda por meio do relacionamento que encerram entre si. Por não recorrer a qualquer referência prévia acerca dos itens a classificar, a análise de agrupamentos, ou classificação não supervisionada, diferencia-se da análise discriminante, ou classificação supervisionada, por ter como propósito fornecer uma organização congruente dos objetos, e não por pretender estabelecer critérios que possibilitem categorizar elementos ainda não identificados, segundo informações obtidas a partir de objetos antecipadamente classificados (Jain and Dubes, 1998).

Formalmente, na classificação supervisionada, a categorização de um conjunto de objetos $O = \{o_i\}, i = 1, \dots, n$, com $o_i \in \mathbb{R}^d$, em um conjunto discreto de classes $C = \{c_j\}, j = 1, \dots, m$, é delineada por meio de uma função matemática $c = c(x, v)$, onde v consiste em um vetor de parâmetros ajustáveis cujos valores são determinados por intermédio de um algoritmo de aprendizagem indutiva, que tem como objetivo promover a minimização do risco funcional da relação de correspondência que determina a categoria à qual um objeto ainda não classificado deve ser relacionado, a partir da análise de um conjunto finito de exemplos $(x_k, c_k), k = 1, \dots, l$, onde l corresponde a cardinalidade do conjunto disponível de objetos categorizados. Já na classificação não supervisionada, também denominada análise de agrupamentos ou ainda análise exploratória de dados, não existem itens categorizados à disposição, e o objetivo consiste em subdividir um conjunto finito de objetos em um conjunto finito de grupos "naturais", ao invés de proporcionar a caracterização de elementos ainda não observados por intermédio de uma relação preditiva (Xu and Wunsch, 2009).

A análise de agrupamentos, que emerge da necessidade de se realizar a exploração de dados de natureza desconhecida, para os quais ainda não se dispõe de quaisquer informações, pode, adicionalmente, ser empregada a fim de promover a classificação não supervisionada de coleções de objetos mais extensas, sobretudo nas circunstâncias em que a atribuição dos objetos às suas respectivas classes torna-se impraticável por consequência da elevada quantidade de objetos a categorizar. Por intermédio da análise de agrupamentos, uma coleção de objetos pode ser desassociada em um número finito de grupos relativamente homogêneos, com base em uma medida de semelhança ou similaridade, de forma que a semelhança entre os objetos que pertencem a um grupo é maior do que a semelhança entre objetos que pertencem a grupos distintos. Critérios de agrupamento ou algoritmos de agrupamento diferentes, ou ainda o mesmo algoritmo de agrupamento com uma seleção distinta de parâmetros, podem originar partições completamente desiguais, não obstante sejam aplicados

na classificação da mesma coleção de objetos. Desta forma, embora existam arranjos de critérios ou de métodos que podem ser apropriados a uma grande variedade de circunstâncias, não é possível estabelecer, a princípio, um que seja adequado a qualquer condição (Xu and Wunsch, 2009).

De acordo com Xu and Wunsch (2009), Bishop (1995), Jain and Dubes (1998), Jain et al. (2000, 1999), um procedimento de análise de agrupamentos compreende fundamentalmente as etapas descritas a seguir:

- Seleção ou extração de características: a seleção das características consiste em designar dentre as propriedades disponíveis, àquelas que sejam mais relevantes em retratar de forma distintiva os objetos da coleção a classificar, ao passo que a extração das características utiliza operações de transformação com o propósito de estabelecer novas propriedades úteis, a partir das originalmente existentes. Embora a operação de extração seja eventualmente apta a caracterizar os objetos por intermédio de atributos mais apropriados em representar as estruturas subjacentes aos dados, essa conduta pode, sob outra perspectiva, determinar características que não sejam fisicamente interpretáveis, por não preservar o significado natural dos atributos escolhidos. Tanto a seleção quanto a extração de atributos que sejam favoráveis em distinguir objetos pertencentes a grupos distintos e que sejam simples de obter e interpretar, são de elevada relevância para a efetividade dos métodos de análise de agrupamentos, haja vista que escolhas apropriadas podem reduzir o custo e facilitar o processo de classificação, além de contribuir para uma melhor compreensão dos resultados alcançados;
- Seleção ou elaboração do algoritmo de agrupamento: a maioria dos algoritmos de agrupamento está implícita ou explicitamente relacionada com a medida de similaridade utilizada para determinar a semelhança entre os objetos a classificar, desta forma, uma vez que o índice de semelhança tenha sido estabelecido, a análise de agrupamentos pode ser estruturada como um problema de otimização, cujos resultados serão dependentes do próprio índice e da função de custo ou função critério a ser minimizada ou maximizada pelo método de classificação. Ainda que uma grande variedade de algoritmos de agrupamento tenha sido desenvolvida, com o propósito de endereçar problemas de diversas naturezas, verifica-se que não há uma estratégia universal, capaz de identificar soluções apropriadas para quaisquer circunstâncias. Consequentemente, uma avaliação criteriosa e diligente das características do problema a solucionar torna-se essencial para que a seleção ou a elaboração do método de agrupamento mais adequado possa ser empreendida;
- Avaliação do agrupamento: considerando que, por intermédio da aplicação de um determinado algoritmo de agrupamento, uma dada coleção de objetos pode invariavelmente ser decomposta em subcoleções, que métodos de agrupamentos distintos ou ainda o mesmo método com parâmetros diferentes, quando empregados a um mesmo conjunto de dados, podem originar classificações completamente desiguais, e que os resultados obtidos não necessariamente correspondem aos grupos naturais originalmente existentes, a adoção de

critérios de avaliação apropriados constitui uma conduta sobremaneira relevante, capaz de oferecer maior confiança quanto à integridade dos resultados finais. Os critérios escolhidos devem ser objetivos, não devem ser influenciados pelo método de análise de agrupamentos adotado, e devem ser aptos em fornecer indicações quanto ao endereçamento adequado de questões importantes, tais como o número de grupos presente na coleção, se as classificações obtidas são relevantes sob o aspecto prático, ou ainda se a escolha de um dado algoritmo de agrupamento em detrimento de outros disponíveis foi de fato oportuna. Em geral, os critérios de avaliação são categorizados em externos, internos ou relativos. Os índices de avaliação externos são baseados em estruturas pré-estabelecidas, que são utilizadas como um padrão a fim de avaliar as partições obtidas, enquanto que os índices de avaliação internos examinam as subcoleções resultantes segundo os dados em seu formato original, sem recorrer a quaisquer informações prévias acerca dos itens a classificar. Os índices relativos, por sua vez, efetuam uma comparação direta entre classificações distintas, com o propósito de fornecer a indicação de qual seria a mais conveniente em distinguir com acerto as características inerentes aos objetos da coleção segmentada;

- Interpretação dos resultados: consiste no procedimento de análise e de investigação dos resultados alcançados, por parte dos interessados. Compreende a utilização das subcoleções juntamente com outras evidências experimentais no intuito de procurar fornecer informações relevantes acerca dos itens classificados.

A análise de agrupamentos é uma atividade humana importante, que se inclui no cotidiano dos indivíduos e que se manifesta por intermédio de um processo ininterrupto de aprendizagem. Com efeito, ainda no início da infância, as crianças aprendem a realizar a distinção entre animais, objetos ou mesmo entre homens e mulheres, a partir da elaboração de esquemas semiconscientes de classificação, continuamente aprimorados. Em ciências, a análise de agrupamentos exerce, de forma semelhante, atribuições relevantes. Em astronomia, estrelas são classificadas em categorias, conforme a luminosidade e segundo a temperatura de sua superfície. Em ciências sociais, pessoas são agrupadas consoante o seu comportamento e as suas preferências particulares. Em mercadologia, a análise de agrupamentos procura identificar clientes com necessidades semelhantes a fim de, eventualmente, permitir a determinação de segmentos de mercado. De modo equivalente, a aplicação de métodos de análise de grupos contribui, em geografia, para a classificação de regiões com características semelhantes, na medicina, para a identificação da incidência de determinados tipos de câncer, em química, para a classificação de elementos com mesmo número atômico, assim como em história, para o agrupamento de achados arqueológicos. As técnicas de análise de agrupamentos podem ser utilizadas ainda não somente para identificar estruturas já existentes nos dados, mas também para estabelecer estruturas particulares em coleções que precisam ser subdivididas de forma razoável e coerente, a exemplo das regiões de um país, que podem exigir, segundo determinados critérios, uma classificação

que auxilie o estabelecimento de iniciativas para expansão dos serviços de telefonia (Kaufman and Rousseeuw, 2005).

Além da introdução, este trabalho compreende cinco seções adicionais, as quais estão organizadas conforme indicado a seguir. A Seção 2 refere os métodos de análise de agrupamentos, e especifica como essas técnicas podem ser utilizadas na determinação da subdivisão de uma coleção de objetos em classes ou categorias. A Seção 3 descreve o método meta-heurístico inspirado no comportamento das ondas oceânicas em águas de baixa profundidade e expõe como os fenômenos ondulatórios associados podem ser estruturados a fim estabelecer uma estratégia capaz de ser aplicada à resolução de problemas de otimização. A Seção 4 refere as adequações efetuadas sobre o algoritmo original com o propósito de aplicá-lo à determinação do agrupamento não hierárquico. A Seção 5 apresenta os experimentos de avaliação que foram conduzidos com a intenção de comparar a meta-heurística inspirada nas ondas aquáticas com outros métodos de particionamento descritos na literatura. Por fim, a Seção 6 relata, consoante os resultados dos experimentos, as conclusões obtidas.

2 Métodos de análise de agrupamentos

Um algoritmo ou método de agrupamento, que tem como propósito subdividir uma coleção de objetos em grupos, subcoleções ou categorias, de modo que os objetos que pertencem a um grupo são mais semelhantes entre si do que os objetos que pertencem a grupos diferentes, pode originar classificações não hierárquicas ou hierárquicas. Formalmente, dado um conjunto de objetos $O = \{o_i\}, i = 1, \dots, n$, onde $o_i = (a_{i1}, a_{i2}, \dots, a_{id}) \in \mathbb{R}^d$, com cada medida a_{id} denominada atributo, dimensão ou característica, um método de agrupamento não hierárquico procura estabelecer uma K -partição de O , $C = \{c_j\}, j = 1, \dots, K$, com $K \leq n$, tal que $c_j \neq \emptyset, j = 1, \dots, K, \bigcup_{j=1}^K c_j = O$ e $c_l \cap c_m = \emptyset, l, m = 1, \dots, K$, com $l \neq m$. Por outro lado, um algoritmo de agrupamento hierárquico procura determinar uma estrutura aninhada de O , $H = \{h_j\}, j = 1, \dots, Q$, com $Q \leq n$, tal que se $c_l \in h_p, c_m \in h_q$ e $p > q$, então $c_l \subset c_m$ ou $c_l \cap c_m = \emptyset \forall l, m \neq p, q = 1, \dots, Q$. A partição obtida por um método de agrupamento não hierárquico é denominada rígida quando cada objeto está associado exclusivamente a um grupo, embora admita-se além disso que os itens da coleção possam estar relacionados simultaneamente a todos os K grupos, com um grau de pertinência $u_{j,i} \in [0, 1], j = 1, \dots, K, i = 1, \dots, n$. Nesta circunstância, tem-se uma partição designada como difusa, na qual o valor $u_{j,i}$, que deve satisfazer às restrições $\sum_{j=1}^K u_{j,i} = 1 \forall i$ e $\sum_{i=1}^n u_{j,i} < n \forall j$, denota o coeficiente de relacionamento entre o i -ésimo objeto e o j -ésimo grupo (Hansen and Jau-mard, 1997, Xu and Wunsch, 2009).

As técnicas de agrupamento hierárquico são mais frequentemente utilizadas em ciências biológicas, sociais e de comportamento, com o objetivo de auxiliar, por exemplo, o estabelecimento de taxonomias, enquanto que os métodos de agrupamento não hierárquico ou de particionamento, são comumente empregados em aplicações do âmbito da engenharia, ou ainda em circunstâncias que de-

mandam uma manipulação eficiente de coleções extensas, constituídas por um número demasiadamente elevado de elementos. O número de grupos K , entre os quais os n objetos da coleção O devem ser subdivididos por um algoritmo de particionamento, pode ou não ser conhecido previamente, e um critério de avaliação deve ser admitido a fim de possibilitar determinar a qualidade do agrupamento estabelecido. Ainda que a solução teórica do problema do agrupamento não hierárquico, que simplesmente consiste em selecionar um critério, avaliar todas as possíveis partições contendo K grupos, e por fim selecionar aquela que otimiza o critério escolhido, seja direta e de fácil entendimento, observa-se que a admissão desse tipo de abordagem conduz a duas dificuldades essenciais. A primeira, reside em estabelecer um critério que retrate o conceito de grupos por intermédio de uma formulação matemática apropriada, enquanto que a segunda dificuldade consiste no número de partições, que é excessivamente elevado mesmo para um número moderado de objetos, tornando imprática a avaliação de um critério, a despeito de simples, em relação à todas as partições. Com efeito, embora existam somente 34.105 partições distintas de 10 objetos em quatro grupos, esse número eleva-se para aproximadamente 11.259.666.000, se 19 objetos necessitam ser particionados em quatro grupos, tornando a enumeração exaustiva computacionalmente inviável, mesmo para um número diminuto de grupos (Jain and Dubes, 1998).

Com o propósito de evitar o encargo de ter que considerar e examinar, todos os arranjos em que n objetos podem ser subdivididos em K grupos, a função critério pode ser avaliada apenas para um subconjunto das partições aceitáveis. Assim, uma das abordagens mais frequentes consiste em determinar uma partição inicial e, em seguida e iterativamente, mover os objetos de um grupo para o outro, em um esforço a fim de melhorar o critério de avaliação. Como cada partição sucessiva é resultado de uma perturbação no arranjo anterior, somente um reduzido número de configurações é analisado, desta forma, ainda que algoritmos fundamentados neste método sejam computacionalmente eficientes, frequentemente convergem para mínimos locais da função critério. Uma outra conduta, que da mesma maneira tem como intenção evitar que um número demasiado de partições tenha que ser investigado, corresponde à utilização de técnicas de programação dinâmica, capazes de identificar e rejeitar a maioria das partições inadequadas, e ainda aprimorar de forma satisfatória o valor da função critério. Por intermédio dessa estratégia, se, por exemplo, 19 objetos precisam ser subdivididos em quatro grupos, menos de 2% do número total de partições necessita ser avaliado, entretanto ainda que esse procedimento favoreça uma redução significativa do custo de processamento, verifica-se que essa diminuição não é suficiente para viabilizar a sua utilização em circunstâncias práticas (Jain and Dubes, 1998).

Uma das condutas mais frequentemente empregadas pelos métodos de agrupamento não hierárquico consiste em, para um número estabelecido de grupos K , minimizar o valor da função critério determinado pelo erro quadrático. Dessa maneira, supondo-se uma coleção de objetos $O = \{o_i\}$, $i = 1, \dots, n$, onde $o_i = (a_{i1}, a_{i2}, \dots, a_{id}) \in \mathbb{R}^d$, que de alguma forma foi particionada em um conjunto $C = \{c_j\}$, $j = 1, \dots, K$, de modo que c_j contém n_k objetos e que

cada objeto esteja situado precisamente em um grupo, isto é Eq. (1),

$$\sum_{k=1}^K n_k = n, \quad (1)$$

considerando-se adicionalmente que o vetor média, ou centro de c_j , seja definido como o centróide do grupo e estabelecido mediante a Eq. (2)

$$m^{(k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} o_i^{(k)}, \quad (2)$$

onde $o_i^{(k)}$ corresponde ao i -ésimo objeto pertencente ao grupo c_j , e que o erro quadrático de c_j , determinado pela soma do quadrado da distância Euclidiana entre cada objeto do grupo e o seu centro $m^{(k)}$, seja computado por meio da Eq. (3)

$$e_k^2 = \sum_{i=1}^{n_k} (o_i^{(k)} - m^{(k)})^T (o_i^{(k)} - m^{(k)}), \quad (3)$$

tem-se que o erro quadrático da partição, constituída por K grupos, é determinado por intermédio da Eq. (4)

$$E_k^2 = \sum_{k=1}^K e_k^2. \quad (4)$$

Minimizar o valor estabelecido pela Eq. (4) constitui um problema otimização combinatória NP -difícil mesmo para $K = 2$, sendo assim, a utilização de algoritmos de aproximação, capazes de identificar soluções admissíveis com o uso limitado dos recursos computacionais, tem sido frequentemente empreendida. Dentre as técnicas aptas a proporcionar de maneira satisfatória a resolução de problemas de otimização, situam-se as meta-heurísticas, que por utilizarem uma associação entre escolhas aleatórias e o conhecimento histórico acerca dos resultados até então alcançados durante o seu processo iterativo, inibem a ocorrência de uma convergência prematura e a consequente obtenção de respostas inapropriadas. Por serem estratégias simples, robustas e flexíveis, que de forma eficiente investigam o conjunto dos resultados viáveis, as meta-heurísticas podem ser aplicadas a uma grande variedade de problemas, a fim de determinar soluções de boa qualidade em um tempo de processamento aceitável (Jain and Dubes, 1998, Aloise et al., 2009, Gonzalez, 2007).

A adoção de meta-heurísticas à resolução do agrupamento não hierárquico tem sido assiduamente verificada na literatura, e inclui métodos inspirados em fenômenos físicos, tendo como exemplo o resfriamento simulado (Selim and Alsultan, 1991), em estratégias estocásticas, tal como a busca tabu (Al-Sultan, 1995), além de algoritmos originados a partir de analogias relacionadas a compor-

tamentos evolucionários e biológicos, a exemplo do algoritmo genético (Maulik and Bandyopadhyay, 2000), do algoritmo inspirado na capacidade de ecolocalização dos morcegos (Aboubi et al., 2016), dos algoritmos de otimização fundamentados em colônias de formigas (Inkaya et al., 2015), em colônias de abelhas artificiais (Karaboga and Ozturk, 2011, Zou et al., 2010), em enxames de partículas (Cura, 2012, Cui et al., 2005), em enxames bioluminescentes (Isimeto et al., 2017), entre outros.

3 Otimização inspirada nas ondas aquáticas

O método de otimização inspirado nas ondas aquáticas orienta-se pelo comportamento das ondas oceânicas em águas de baixa profundidade e provê uma estratégia apta a ser empregada à resolução de problemas de otimização combinatória (Zheng, 2015, Zheng et al., 2019). Constitui um algoritmo que tem sido frequentemente abordado na literatura e que tem sido admitido ao endereçamento de problemas de diversas naturezas, a exemplo da estruturação de redes neurais artificiais (Zhou et al., 2019), do aprimoramento de métodos de classificação (Ibrahim et al., 2020), da alocação do espectro de frequências em sistemas de radiodifusão (Singh et al., 2019), do planejamento da produção de itens manufaturados (Zhao et al., 2020), do planejamento do percurso de veículos sub-aquáticos autônomos (Yan et al., 2021), entre outros.

Pressupondo a existência de um problema de otimização P , cujo propósito consiste em maximizar uma função critério f , tem-se, por analogia, que o conjunto de soluções viáveis, $S' = \{s'_i\}$, $i = 1, \dots, m$, com $s_i = (w_{i1}, w_{i2}, \dots, w_{id}) \in \mathbb{R}^d$, corresponde ao leito oceânico e que a adequação de uma solução particular $s \in S'$ de P é inversamente proporcional à profundidade, de forma que quanto menor for a distância entre s e a superfície da água, mais apropriado será o valor $f(s)$. Assim como observado em alguns algoritmos fundamentados em comportamentos evolucionários, o método inspirado na propagação das ondas oceânicas de superfície mantém uma população de soluções admissíveis ou "ondas", estabelecidas aleatoriamente e caracterizadas pela altura $h \in \mathbb{Z}^+$ e pelo comprimento $\lambda \in \mathbb{R}^+$. Na inicialização, cada uma das ondas que constituem a população de soluções, $S = \{s_i\}$, $i = 1, \dots, n$, $n \ll m$, manipulada pelo algoritmo, admite como altura, h , e como comprimento, λ , respectivamente, h_{max} e 0,5, enquanto que durante o procedimento iterativo de aperfeiçoamento da função critério, três tipos de operações são empreendidas: propagação, refração e quebra (Zheng, 2015, Zheng et al., 2019).

A cada iteração do algoritmo, o operador de propagação origina uma nova onda s'_i , modificando o valor de cada uma das dimensões da solução s_i por intermédio da Eq. (5)

$$w'_{ij} = w_{ij} + U(-1, 1)\lambda L(j), \quad j = 1, \dots, d \quad (5)$$

onde $U(-1, 1)$ corresponde a um número aleatório situado no intervalo $[-1, 1]$, e $L(j)$ retrata o comprimento da j -ésima dimensão de s_i . Como, à medida em que a onda se desloca das regiões de maior profundidade para as águas de menor profundidade, a sua altura aumenta e o seu comprimento diminui, a qualidade da solução modificada, obtida

por meio do mecanismo de propagação é, com relação à função critério, comparada com a solução que a originou. Desta forma, se $f(s'_i) > f(s_i)$, s'_i substitui s_i na população de soluções e tem a sua altura, h , definida como h_{max} . Caso contrário, a solução s_i é preservada, mas tem a sua altura, h , decrementada em uma unidade, a fim reproduzir a dissipação de energia da onda, por consequência da resistência inercial e do atrito com o leito oceânico (Zheng, 2015, Zheng et al., 2019).

No decorrer da execução do algoritmo, qualquer onda s_i cuja altura, h , tenha sido decrementada até zero é submetida ao procedimento de refração, o qual consiste em, a partir de s_i , estabelecer uma nova solução s'_i , modificando cada uma das coordenadas da solução original por meio da Eq. (6)

$$w'_{ij} = N\left(\frac{w_j^* + w_{ij}}{2}, \frac{|w_j^* - w_{ij}|}{2}\right), \quad j = 1, \dots, d \quad (6)$$

onde w_j^* corresponde a j -ésima dimensão da melhor solução até então obtida, e $N(\mu, \sigma)$ consiste em um número Gaussiano aleatório, de média μ e desvio padrão σ . Adicionalmente e imediatamente após a refração, a solução s'_i , recém estabelecida, tem a sua altura, h' , definida como h_{max} e o seu comprimento, λ' , determinado pela Eq. (7)

$$\lambda' = \lambda \frac{f(s_i)}{f(s'_i)} \quad (7)$$

onde $f(s_i)$ e $f(s'_i)$ retratam, respectivamente, a adequação das soluções s_i e s'_i , com relação à função critério f , e λ denota o comprimento da onda s_i (Zheng, 2015, Zheng et al., 2019).

Nas praias oceânicas, à medida em que uma onda da superfície da água se desloca para posições em que a profundidade é inferior a um determinado limite mínimo, a velocidade da sua exterioridade superior passar a exceder a velocidade da própria onda. Consequentemente, a crista, ou o topo da onda, torna-se continuamente mais íngreme e mais íngreme, até que a onda quebra e continua a se deslocar em direção à frente inclinada da praia. Durante as iterações do algoritmo, admite-se por analogia que a operação de quebra, a qual é empreendida somente quando a solução, s'_i , determinada por meio do operador de propagação é mais apropriada do que a melhor solução, s^* , até então obtida, consiste na seleção aleatória de k coordenadas de s'_i e na posterior atualização dos valores das dimensões escolhidas por intermédio da Eq. (8)

$$w''_{ij} = w'_{ij} + N(0, 1)\beta L(j) \quad j \in [1, d] \quad (8)$$

onde β corresponde ao coeficiente de quebra e k denota um número aleatório disposto no intervalo compreendido entre 1 e o valor k_{max} , preestabelecido. As k soluções, s_i'' , recém determinadas pela operação de quebra, são subsequentemente comparadas com a melhor solução até aquele momento obtida. Desta forma, se nenhuma das ondas s_i''

for mais apropriada do que s^* , esta é preservada. Caso contrário, s^* é substituída pela melhor dentre as k soluções s_i'' que são mais adequadas do que s^* (Zheng, 2015, Zheng et al., 2019).

A cada iteração do método e logo após a aplicação do operador de propagação, e eventualmente dos operadores de refração e quebra, as ondas, s_i , que constituem o conjunto de soluções, S , manipulado pelo algoritmo, têm o seu comprimento, λ , modificado por meio da Eq. (9)

$$\lambda = \lambda \alpha^{-(f(s_i) - f_{\min} + \theta) / (f_{\max} - f_{\min} + \theta)} \quad (9)$$

onde $f_{\max}$ e $f_{\min}$ correspondem, respectivamente, ao maior e ao menor valor de adequação das soluções em relação à função critério f , α expressa o coeficiente de redução do comprimento de onda, e $\theta \in \mathbb{R}^+$ retrata um valor constante diminuto que tem como propósito evitar divisões por zero (Zheng, 2015, Zheng et al., 2019).

Algoritmo 1: Otimização inspirada nas ondas aquáticas

```

1 Estabeleça  $f(s)$ ,  $n$ ,  $h_{\max}$ ,  $\lambda$ ,  $\alpha$ ,  $\beta$ , e  $k_{\max}$ ;
2 Inicialize aleatoriamente um conjunto de soluções
   $S = \{s_i\}$ ,  $i = 1, \dots, n$ ;
3 enquanto o critério de interrupção não for satisfeito
  faça
4   para cada  $s_i \in S$  faça
5     Com base na Eq. (5), aplique o operador de
      propagação sobre  $s_i$  e obtenha  $s_i'$ ;
6     se  $f(s_i') > f(s_i)$  então
7       se  $f(s_i') > f(s_i^*)$  então
8         Com base na Eq. (8), aplique o
          operador de quebra sobre  $s_i'$  e
          obtenha  $s_i''$ ;
9         se  $f(s_i'') > f(s_i^*)$  então
10          Substitua  $s_i^*$  por  $s_i''$ ;
11       Substitua  $s_i$  por  $s_i'$ ;
12     Defina a altura,  $h$ , de  $s_i$  igual a  $h_{\max}$ ;
13   senão
14     Decrementa a altura,  $h$ , de  $s_i$  em uma
      unidade;
15     se  $h = 0$  então
16       Com base nas Eqs. (6) e (7), aplique o
        operador de refração sobre  $s_i$  e
        obtenha  $s_i'$ ;
17       Substitua  $s_i$  por  $s_i'$ ;
18     Defina a altura,  $h$ , de  $s_i$  igual a  $h_{\max}$ ;
19   Com base na Eq. (9), atualize o comprimento,
       $\lambda$ , de  $s_i$ ;

```

De acordo com Zheng (2015), Zheng et al. (2019), a estratégia de otimização inspirada nas ondas aquáticas pode ser estabelecida conforme o algoritmo 1. Durante a execução do método, o operador de propagação efetua por

meio das ondas de maior aptidão, uma investigação de menor abrangência entre os elementos que constituem o conjunto de soluções viáveis, S' , ao mesmo tempo em que, mediante as ondas de menor aptidão, empreende uma avaliação de maior abrangência entre os elementos de S' . O operador de refração aperfeiçoa a diversidade do conjunto, S , utilizado pelo algoritmo, evitando que o procedimento de investigação do espaço de soluções permaneça estagnado, reduzindo por consequência a perspectiva de uma convergência prematura. Por fim, o operador de quebra proporciona uma avaliação complementar de áreas potencialmente promissoras, haja vista que efetua uma busca aleatória na circunvinhaça da melhor solução, s^* , até então alcançada.

O algoritmo 1 possui como parâmetros o número, n , de elementos da população de soluções, a altura máxima, $h_{\max}$, das ondas que constituem a população, o coeficiente de redução de comprimento, α , além do coeficiente de quebra, β , e do número máximo de direções de quebra $k_{\max}$. Valores mais elevados do parâmetro $h_{\max}$ conduzem a um comportamento no qual as ondas permanecem por mais tempo entre os elementos que constituem a população de soluções, em contrapartida, valores menos elevados de $h_{\max}$, por exemplo 1 ou 2, proporcionam um incremento da diversidade das soluções, por aumentarem a frequência com que as ondas são substituídas por novas soluções, originadas por intermédio do operador de refração. Valores mais elevados do parâmetro α permitem uma exploração mais abrangente do conjunto de soluções viáveis, S' , enquanto que valores menos elevados de α conduzem a uma avaliação mais particularizada de S' . Adicionalmente, valores reduzidos do parâmetro β juntamente com valores elevados do parâmetro $k_{\max}$ proporcionam um incremento da velocidade de convergência do algoritmo, mediante o aprimoramento da melhor solução obtida, s^* , a partir da avaliação das soluções localizadas nas proximidades de s^* (Zheng, 2015).

Conforme avaliações empíricas, uma população de soluções que compreenda entre 5 e 10 elementos é suficiente para obtenção de resultados satisfatórios. De modo semelhante, recomenda-se que o valor de $h_{\max}$ seja 5 ou 6, que os valores dos parâmetro α e β , estejam, respectivamente, nos intervalos compreendidos entre 1,001 e 1,01, e entre 0,001 e 0,01, e que o valor do parâmetro $k_{\max}$ seja o mínimo entre 12 e $d/2$, onde d corresponde à dimensão do problema (Zheng, 2015).

4 Otimização inspirada nas ondas aquáticas aplicada ao agrupamento não hierárquico de objetos

Fundamentado no algoritmo referido na Seção 3, o método de agrupamento não hierárquico de objetos proposto neste estudo e descrito por meio do algoritmo 2, incorporou algumas particularidades, as quais foram decorrentes da natureza do problema endereçado.

As parâmetros de execução do algoritmo foram acrescentados o conjunto de objetos $O = \{o_i\}$, $i = 1, \dots, n$, com $o_i = (a_{i1}, a_{i2}, \dots, a_{id}) \in \mathbb{R}^d$, assim como o número de grupos K nos quais O deveria ser subdividido, de forma a obter a K -partição $C = \{c_j\}$, $j = 1, \dots, K$ ($K \leq n$). As ondas

$s = (w_1, w_2, \dots, w_{d'}) \in \mathbb{R}^{d'}$, $d' = dK$, da população foram retratadas pelos protótipos ou médias dos K grupos. Consequentemente, os valores posicionados no intervalo $[1, d]$ referiam-se às coordenadas da média do primeiro grupo, os valores dispostos no intervalo $[d + 1, 2d]$ às coordenadas do protótipo do segundo grupo, e assim sucessivamente até o intervalo $[(k - 1)d + 1, dK]$, que denotava as coordenadas da média do K -ésimo grupo.

Algoritmo 2: Otimização inspirada nas ondas aquáticas aplicada ao agrupamento não hierárquico de objetos

```

1  Estabeleça  $f(s)$ ,  $n$ ,  $h_{max}$ ,  $\lambda$ ,  $\alpha$ ,  $\beta$ ,  $k_{max}$ ,  $O$  e  $K$ ;
2  Inicialize aleatoriamente um conjunto de soluções
    $S = \{s_i\}$ ,  $i = 1, \dots, n$ ;
3  enquanto o critério de interrupção não for satisfeito
   faça
4     para cada  $s_i \in S$  faça
5         Com base na Eq. (5), aplique o operador de
           propagação sobre  $s_i$  e obtenha  $s'_i$ ;
6         se  $1/f(s'_i) > 1/f(s_i)$  então
7             se  $1/f(s'_i) > 1/f(s_i^*)$  então
8                 Com base na Eq. (8), aplique o
                   operador de quebra sobre  $s'_i$  e
                   obtenha  $s''_i$ ;
9                 se  $1/f(s''_i) > 1/f(s_i^*)$  então
10                    Substitua  $s_i^*$  por  $s''_i$ ;
11             Substitua  $s_i$  por  $s'_i$ ;
12             Defina a altura,  $h$ , de  $s_i$  igual a  $h_{max}$ ;
13         senão
14             Decrementa a altura,  $h$ , de  $s_i$  em uma
               unidade;
15         se  $h = 0$  então
16             Com base nas Eqs. (6) e (7), aplique o
               operador de refração sobre  $s_i$  e
               obtenha  $s'_i$ ;
17             Substitua  $s_i$  por  $s'_i$ ;
18             Defina a altura,  $h$ , de  $s_i$  igual a  $h_{max}$ ;
19     Com base na Eq. (9), atualize o comprimento,
        $\lambda$ , de  $s_i$ ;
    
```

A operação de inicialização dos n elementos da população, consistia, para cada onda, na seleção aleatória de K objetos de O , a fim de retratar os protótipos ou médias dos grupos, e na posterior distribuição dos objetos não selecionados como protótipos entre os grupos conforme a menor dissimilaridade entre o objeto e o protótipo, determinada por meio da distância Euclidiana, de forma que o i -ésimo objeto, o_i , era relacionado ao j -ésimo grupo, de protótipo $m^{(j)}$, quando $[(o_i - m^{(j)})^T(o_i - m^{(j)})]^{(1/2)} \leq [(o_i - m^{(l)})^T(o_i - m^{(l)})]^{(1/2)}$, $\forall i = 1, 2, \dots, n$, $\forall j = 1, 2, \dots, K$ e $\forall l = 1, 2, \dots, K$, com $o_i \neq m^{(j)}$, $o_i \neq m^{(l)}$ e $m^{(j)} \neq m^{(l)}$. As coordenadas da média ou protótipo de

cada grupo, recalculadas após a distribuição dos objetos segundo a menor dissimilaridade entre o objeto e o protótipo, eram determinadas por intermédio da Eq. (2). Ademais, como o agrupamento não hierárquico de objetos foi abordado como um problema de otimização, que tinha como objetivo identificar os protótipos que mais adequadamente representavam os grupos da partição, a função critério, a minimizar, foi estabelecida por intermédio da Eq. (4), enquanto que as estruturas de seleção incluídas nas linhas 6, 7 e 9 do algoritmo original foram modificadas para que as condições avaliadas passassem a considerar o inverso da função critério.

O procedimento de aperfeiçoamento do valor da função critério, calculado para cada uma das soluções da população, foi efetuado por meio da modificação das coordenadas dos protótipos dos grupos, mediante a aplicação dos operadores de propagação, refração e quebra, com a posterior distribuição dos objetos entre os grupos, considerando a menor dissimilaridade entre o objeto e o protótipo alterado. A fim de evitar a permanência de soluções que eventualmente contivessem grupos vazios, entre os itens da população, uma operação de ajuste que, nestas circunstâncias, adicionava um valor constante excepcionalmente elevado ao resultado da função critério, foi acrescentado.

Os demais parâmetros de execução do algoritmo, foram estabelecidos conforme admitido nos trabalhos de Zheng (2015), Zheng et al. (2019), assim, a altura máxima, h_{max} , o comprimento de onda inicial, λ , o coeficiente de redução do comprimento, α , o coeficiente de quebra, β , e o número máximo de direções de quebra k_{max} , eram nessa ordem, iguais a 6, 0,5, 1,0026, 0,0026 e ao valor mínimo entre 12 e $d'/2$, onde $d' = dK$ correspondia à dimensão do problema. Adicionalmente, o número de ondas da população de soluções, n , e o número máximo de iterações até que a execução do algoritmo fosse descontinuada, foram determinados conforme Zheng (2015), Zheng et al. (2019), Zhu et al. (2012), Xie et al. (2019), Maulik and Bandyopadhyay (2000). Desta forma, os valores desses parâmetros nas simulações conduzidas neste estudo eram, respectivamente, iguais a 10 e 200.

5 Experimentos computacionais

Com o objetivo de avaliar o comportamento do procedimento de otimização inspirado nas ondas aquáticas, quando aplicado ao agrupamento não hierárquico de objetos, um conjunto de experimentos foi empreendido. Vinte e nove coleções, constituídas por objetos de variadas dimensões e com números de grupos conhecidos, foram submetidas ao particionamento determinado pela meta-heurística inspirada nas ondas aquáticas, e, os resultados obtidos, foram subsequentemente comparados com os resultados estabelecidos pelas meta-heurísticas enxame de partículas, recozimento simulado, algoritmo genético, colônia de abelhas artificiais e pelo método K -means, descritos, respectivamente, em Cui et al. (2005), Selim and Alsultan (1991), Maulik and Bandyopadhyay (2000), Zou et al. (2010) e MacQueen (1967). Adicionalmente, e com a intenção de proporcionar uma maior abrangência à análise das respostas fornecidas pelo algoritmo sugerido, a avaliação conduzida neste trabalho incluiu as soluções obtidas por meio do algoritmo descrito no estudo de Kaur

and Kumar (2021), que de maneira semelhante empregou a meta-heurística inspirada nas ondas aquáticas ao agrupamento de coleções de objetos, mas que acrescentou modificações aos operadores de propagação e de refração, com o propósito de, supostamente, inibir a ocorrência de uma convergência prematura e assim possibilitar a obtenção de soluções mais apropriadas do que aquelas que seriam auferidas pelo algoritmo em sua forma canônica. Em particular, no trabalho desenvolvido por Kaur and Kumar (2021), a propagação das ondas da população de soluções foi alterada para considerar a melhor solução até então calculada, enquanto que a refração foi revisada para adicionar um operador de decaimento, que incluída em sua formulação a diferença entre o valor de aptidão da solução original e o valor de adequação da solução recém determinada.

Todas as estratégias de agrupamento, que tiveram como função critério a minimizar o erro quadrático, retratado pela Eq. (4), foram codificadas por meio da linguagem de programação Microsoft Visual Basic .NET. Em adição, os experimentos de determinação das partições foram conduzidos em um microcomputador que dispunha do sistema operacional Microsoft Windows 10, memória RAM de 8GB e processador Intel i5 de 1,80GHz. Além dos parâmetros de execução pertinentes à meta-heurística inspirada na propagação das ondas aquáticas, referidos na Seção 4, os seguintes critérios foram assinalados para os demais algoritmos. Para o enxame de partículas e consoante ao estabelecido em Cui et al. (2005), Maulik and Bandyopadhyay (2000), Zhu et al. (2012), Xie et al. (2019), o número de partículas, o número máximo de iterações, o coeficiente de inércia e os coeficientes de aceleração eram, nessa ordem, iguais a 10, 200, 0,72 e 1,49. Para o recozimento simulado e de acordo com o sugerido em Selim and Alsultan (1991), a quantidade de perturbações na solução inicial, utilizada para determinar a temperatura preliminar do sistema, foi estabelecida em 100, enquanto que o número de iterações na temperatura, a probabilidade de seleção de uma solução circunvizinha e o fator de resfriamento eram, nessa ordem, iguais a 50, 0,95 e 0,8. Para o algoritmo genético e segundo a recomendações de Maulik and Bandyopadhyay (2000), Zhu et al. (2012), Xie et al. (2019), o número de indivíduos da população, o número máximo de gerações, a probabilidade de recombinação e a probabilidade de mutação, eram, respectivamente, iguais a 10, 200, 0,8 e 0,001. Para a colônia de abelhas artificiais e de acordo com as sugestões de Maulik and Bandyopadhyay (2000), Zhu et al. (2012), Xie et al. (2019), o número de abelhas do exame, o número máximo de iterações, o limite máximo de iterações sem melhoria, até que uma solução representada por uma abelha operária fosse substituída por uma solução calculada aleatoriamente, eram, nessa ordem, iguais a 10, 200 e 10. Para o método de otimização inspirado nas ondas aquáticas que incluía o operador de decaimento, e consoante as orientações de Kaur and Kumar (2021), Maulik and Bandyopadhyay (2000), Zhu et al. (2012), Xie et al. (2019), o número de ondas da população, o número máximo de iterações, a altura máxima das ondas, o comprimento de onda inicial, o coeficiente de redução do comprimento, o coeficiente de quebra e o número máximo de direções de quebra, eram, respectivamente, iguais a 10, 200, 12, 0,5, 1,2, 0,01 e 12. Por fim, para o método *K-means* e de acordo com o observado em Cui et al. (2005), o número máximo

de iterações, até que a execução do algoritmo fosse descontinuada, foi estabelecido em 10.

A Tabela 1, refere as características das coleções de objetos originadas de Fränti and Sieranoja (2018), Dua and Graff (2017), que foram utilizadas nos ensaios de avaliação dos algoritmos de particionamento. Inicialmente e para cada um dos conjuntos de itens, os métodos de agrupamento foram executados dez vezes, e, subsequentemente, os valores médios dos erros quadráticos resultantes das subdivisões dos objetos entre os grupos foram comparados, segundo orientações descritas em Field (2009), Derac et al. (2011). De modo específico, os resultados obtidos por cada estratégia de classificação, foram examinados por intermédio do teste dos postos sinalizados de Wilcoxon, indicado para circunstâncias em que, para um dado ensaio experimental, há dois conjuntos de valores a comparar, originados do mesmo grupo de participantes, submetido a duas condições distintas. Em particular, os participantes dos ensaios foram retratados pelas coleções de objetos a classificar, os resultados corresponderam ao erro quadrático associado ao particionamento alcançado, enquanto que as condições foram representadas pelos métodos de particionamento, neste contexto examinados aos pares.

Tabela 1: Características das coleções

Nome	Objetos	Dimensões	Grupos
Aggregation	788	2	7
Compound	399	2	6
Pathbased	300	2	3
Spiral	312	2	3
D31	3.100	2	31
R15	600	2	15
Jain	373	2	2
Flame	240	2	2
Iris	150	4	3
Wine	178	13	3
Yeast	1.484	8	10
Breast	699	9	2
Glass	214	9	7
Dim032	1.024	32	16
Dim064	1.024	64	16
A1	3.000	2	20
A2	5.250	2	35
A3	7.500	2	50
Dim2	1.351	2	9
Dim3	2.026	3	9
Dim4	2.701	4	9
Dim5	3.376	5	9
Dim6	4.051	6	9
S1	5.000	2	15
S2	5.000	2	15
S3	5.000	2	15
Balance	625	4	3
CMC	1.473	9	3
Dermatology	366	34	6

Na análise conduzida, a estratégia sugerida, isto é, a meta-heurística inspirada nas ondas aquáticas em seu formato canônico, foi estabelecida como sendo o algoritmo de referência, que teve os seus resultados confrontados com as repostas fornecidas pelas demais estratégias de classi-

ficação. A hipótese experimental foi a de que a utilização do algoritmo proposto iria resultar em uma classificação cujo erro quadrático seria, sob o aspecto estatístico, significativamente menor do que erro obtido ao se utilizar qualquer um dos demais algoritmos. De modo contrário, a hipótese nula foi a de que a utilização do algoritmo sugerido não iria resultar em um erro quadrático inferior, ou ainda que não iria ocasionar a obtenção de um resultado estatisticamente significativo, ainda que inferior ou superior ao alcançado pela estratégia com a qual estivesse sendo comparado. Desta forma, uma análise estatística não paramétrica, fundamentada no teste dos postos sinalizados de Wilcoxon, foi empreendida a fim atestar, para cada coleção e para cada par de algoritmos, se a hipótese experimental seria ou não aceita.

Tendo em conta as circunstâncias em que os ensaios foram efetuados, havia por coleção e para cada par de algoritmos em análise, 20 observações, constituídas pelos erros quadráticos associados às 10 execuções independentes do primeiro e do segundo métodos, as quais foram confrontadas conforme descrito a seguir. Inicialmente e a fim de auxiliar a avaliação e a demonstração dos resultados, os valores do conjunto correspondente às respostas auferidas pelo par de algoritmos foram normalizados, de forma que o resultado mais apropriado fosse representado por 1,0000, enquanto que resultado menos favorável fosse retratado por 0,0000. Posteriormente, e seguindo as orientações descritas em Field (2009), para o teste dos postos sinalizados de Wilcoxon, a estatística teste, T , além da média, μ , e do desvio-padrão, σ , associados a T , foram determinados. Subsequentemente, com base em T , μ e σ , e ainda consoante as orientações de Field (2009), o escore- z , assim como o tamanho do efeito r foram calculados, sendo a hipótese experimental aceita quando o algoritmo sugerido obtinha a maior média normalizada do erro padrão, juntamente com um escore- z cujo valor absoluto fosse superior a 1,96 ($p < 0,05$). O tamanho do efeito, r , ainda que não determinante para aceitar ou não a hipótese experimental, permitiu complementar avaliação dos resultados ao indicar, segundo Cohen (1988), se a importância do resultado havia sido pequena, média ou grande, conforme o valor auferido para o mesmo tivesse sido, respectivamente, maior do que 0,10, 0,30 ou 0,50.

O exame dos valores indicados na primeira linha da Tabela 2, que contém os resultados da comparação entre o algoritmo sugerido e a meta-heurística inspirada no enxame de partículas, permite de modo específico observar uma circunstância em que a hipótese experimental não foi aceita. De fato, ainda que para a coleção Aggregation o valor médio do erro quadrático do algoritmo de referência, 0,7391, tenha sido superior ao valor médio do erro quadrático do método com o qual foi confrontado, 0,5294, o escore- z , -1,6818, de valor absoluto inferior a 1,96 indicou a ausência de significância estatística na diferença observada, determinando assim a rejeição da hipótese experimental de que o erro obtido pela meta-heurística inspirada nas ondas aquáticas seria, de forma expressiva, mais apropriado do que erro alcançado pela estratégia fundamentada no enxame de partículas. De modo contrário, a segunda linha da Tabela 2, retrata uma condição de aceite da hipótese experimental, na qual os valores do escore- z e dos erros quadráticos médios do método proposto e do

algoritmo com o qual foi comparado, foram, para a coleção Compound, respectivamente iguais a -2,2934, 0,8800 e 0,5236. A importância do efeito, 0,5128, adicionalmente demonstrada, possibilita assinalar ainda que a utilização do algoritmo inspirado nas ondas aquáticas, para subdividir a coleção Compound proporcionou, além de um erro quadrático mais apropriado, uma expressiva alteração nos resultados obtidos, quando comparados às respostas fornecidas pela meta-heurística fundamentada no enxame de partículas.

Tabela 2: Onda aquática versus enxame de partículas

Coleção	T	z	r	μ_{oa}	μ_{ep}
Aggregation	11	-1,6818	-0,3761	0,7391	0,5294
Compound	5	-2,2934	-0,5128	0,8800	0,5236
Pathbased	0	-2,8031	-0,6268	0,0572	0,8753
Spiral	7	-2,0896	-0,4673	0,0780	0,2185
D31	0	-2,8031	-0,6268	0,9069	0,2291
R15	0	-2,8031	-0,6268	0,9391	0,3135
Jain	10	-1,7838	-0,3989	0,2051	0,4025
Flame	16	-1,1722	-0,2621	0,3884	0,5070
Wine	6	-2,1915	-0,4900	0,5630	0,3222
Yeast	0	-2,8031	-0,6268	0,5812	0,1347
Breast	3	-2,4973	-0,5584	0,0335	0,5461
Iris	0	-2,8031	-0,6268	0,1182	0,4835
Glass	10	-1,7838	-0,3989	0,4778	0,7763
Dim032	0	-2,8031	-0,6268	0,8951	0,4944
Dim064	3	-2,4973	-0,5584	0,5928	0,3080
A1	0	-2,8031	-0,6268	0,8541	0,2759
A2	0	-2,8031	-0,6268	0,8646	0,1373
A3	0	-2,8031	-0,6268	0,9167	0,1406
Dim2	4	-2,3953	-0,5356	0,7096	0,3506
Dim3	0	-2,8031	-0,6268	0,7686	0,3051
Dim4	8	-1,9876	-0,4444	0,6549	0,3559
Dim5	0	-2,8031	-0,6268	0,8014	0,3834
Dim6	0	-2,8031	-0,6268	0,9272	0,2940
S1	0	-2,8031	-0,6268	0,9658	0,1955
S2	0	-2,8031	-0,6268	0,9435	0,2046
S3	0	-2,8031	-0,6268	0,8987	0,1595
Balance	8	-1,9876	-0,4444	0,0334	0,1880
CMC	0	-2,8031	-0,6268	1,0000	0,4897
Dermatology	0	-2,8031	-0,6268	0,8452	0,0925

T : estatística teste; z : escore- z ; r : tamanho do efeito; μ_{oa} : média do valor normalizado do erro quadrático - onda aquática; μ_{ep} : média do valor normalizado do erro quadrático - enxame de partículas. Os valores em negrito indicam circunstâncias em que a hipótese experimental foi aceita.

Uma análise complementar do conteúdo da Tabela 2, por intermédio de uma perspectiva que consolida as circunstâncias em que a hipótese experimental foi aceita ou não, possibilita observar que a forma canônica do algoritmo inspirado nas ondas aquáticas, foi mais apropriada do que a meta-heurística fundamentada no enxame de partículas em 20 (68,97%) oportunidades, que em 5 (17,24%) ocasiões a meta-heurística inspirada no enxame de partículas foi a mais congruente, e que em 4 (13,79%) comparações não houve significância estatística entre as diferenças verificadas. Os valores da Tabela 2 assinalam ainda que o tamanho do efeito foi grande em 18 (90,00%) das 20 ocasiões em que o algoritmo sugerido foi o mais apropriado, e que foi médio em apenas duas oportunidades. Os

resultados da [Tabela 2](#) assinalam portanto, que houve uma predominância do algoritmo proposto em ser, conforme o erro quadrático, mais apropriado do que a meta-heurística fundamentada no enxame de partículas, ao realizar a subdivisão das coleções de objetos admitidas, em um conjunto finito de grupos.

Tabela 3: Onda aquática versus algoritmo genético

Coleção	T	z	r	μ_{oa}	μ_{ag}
Aggregation	3	-2,4973	-0,5584	0,6704	0,3599
Compound	0	-2,8031	-0,6268	0,9649	0,4649
Pathbased	0	-2,6656	-0,5961	0,4754	0,2344
Spiral	0	-2,8031	-0,6268	0,3331	0,0205
D31	11	-1,6818	-0,3761	0,7210	0,6013
R15	0	-2,8031	-0,6268	0,9154	0,6212
Jain	0	-2,6656	-0,5961	0,2288	0,0060
Flame	0	-2,8031	-0,6268	0,7365	0,1058
Wine	0	-2,8031	-0,6268	0,3722	0,1317
Yeast	0	-2,8031	-0,6268	0,6011	0,2502
Breast	0	-1,3417	-0,3000	0,2000	0,0000
Iris	5	-1,5213	-0,3402	0,2584	0,0944
Glass	7	-2,0896	-0,4673	0,7009	0,3697
Dim032	11	-1,6818	-0,3761	0,7220	0,4392
Dim064	6	-2,1915	-0,4900	0,3027	0,5646
A1	11	-1,6818	-0,3761	0,6982	0,4412
A2	22	-0,5606	-0,1254	0,4209	0,3231
A3	23	-0,4587	-0,1026	0,5900	0,5097
Dim2	7	-1,5403	-0,3444	0,6912	0,6172
Dim3	25	-0,2548	-0,0570	0,7324	0,6791
Dim4	23	-0,4587	-0,1026	0,6841	0,6371
Dim5	4	-1,9604	-0,4384	0,7252	0,5366
Dim6	0	-2,6656	-0,5961	0,8687	0,3912
S1	0	-2,8031	-0,6268	0,9342	0,4138
S2	10	-1,4809	-0,3311	0,9050	0,6797
S3	11	-1,6818	-0,3761	0,7708	0,5468
Balance	0	-2,8031	-0,6268	0,8910	0,2045
CMC	0	-2,8031	-0,6268	1,0000	0,5250
Dermatology	1	-2,7011	-0,6040	0,7019	0,1119

T: estatística teste; z: escore-z; r: tamanho do efeito; μ_{oa} : média do valor normalizado do erro quadrático - onda aquática; μ_{ag} : média do valor normalizado do erro quadrático - algoritmo genético. Os valores em negrito indicam circunstâncias em que a hipótese experimental foi aceita.

A avaliação do conteúdo da [Tabela 3](#), que demonstra os resultados da comparação entre a forma canônica da meta-heurística inspirada nas ondas aquáticas e o algoritmo genético, permite constatar que em 16 (55,17%) oportunidades o algoritmo sugerido foi o mais adequado, com um efeito caracterizado como grande em 14 (87,50%) das 16 ocasiões em que se sobressaiu. As repostas da [Tabela 3](#) indicam adicionalmente que somente em 1 (3,45%) comparação, o algoritmo genético assinalou o melhor resultado, e que em 12 (41,38%) ocasiões não houve significância estatística na diferença entre os erros quadráticos. O teor da [Tabela 3](#) assinala por conseguinte, o predomínio da meta-heurística inspirada nas ondas aquáticas em obter partições mais congruentes do que algoritmo genético, ainda que em uma proporção menor à verificada na comparação com o enxame de partículas, sobretudo por consequência de uma quantidade maior diferenças sem relevância estatística.

Tabela 4: Onda aquática versus K-means

Coleção	T	z	r	μ_{oa}	μ_{km}
Aggregation	5	-2,2934	-0,5128	0,7105	0,4336
Compound	0	-2,8031	-0,6268	0,9711	0,3561
Pathbased	2	-2,5992	-0,5812	0,6528	0,3701
Spiral	5	-2,2934	-0,5128	0,6858	0,4836
D31	6	-2,1915	-0,4900	0,7210	0,4599
R15	0	-2,8031	-0,6268	0,8883	0,2638
Jain	16	-1,1722	-0,2621	0,3453	0,2194
Flame	0	-2,8031	-0,6268	0,7512	0,1708
Wine	0	-2,8031	-0,6268	0,2920	0,0259
Yeast	0	-2,8031	-0,6268	0,6835	0,2952
Breast	1	-2,7011	-0,6040	0,9386	0,6950
Iris	5	-2,2934	-0,5128	0,8122	0,6206
Glass	0	-2,8031	-0,6268	0,7748	0,3594
Dim032	0	-2,8031	-0,6268	0,8346	0,4543
Dim064	0	-2,8031	-0,6268	0,5261	0,3386
A1	5	-2,2934	-0,5128	0,6300	0,3306
A2	6	-2,1915	-0,4900	0,6022	0,3431
A3	19	-0,8664	-0,1937	0,5872	0,4384
Dim2	0	-2,8031	-0,6268	0,8679	0,4887
Dim3	0	-2,8031	-0,6268	0,8237	0,4437
Dim4	4	-2,3953	-0,5356	0,8468	0,5775
Dim5	0	-2,8031	-0,6268	0,8883	0,6267
Dim6	0	-2,8031	-0,6268	0,9423	0,3856
S1	0	-2,8031	-0,6268	0,9509	0,3822
S2	0	-2,8031	-0,6268	0,9146	0,3918
S3	1	-2,7011	-0,6040	0,7828	0,4057
Balance	0	-2,8031	-0,6268	0,9401	0,2583
CMC	0	-2,8031	-0,6268	1,0000	0,8043
Dermatology	0	-2,8031	-0,6268	0,7189	0,1028

T: estatística teste; z: escore-z; r: tamanho do efeito; μ_{oa} : média do valor normalizado do erro quadrático - onda aquática; μ_{km} : média do valor normalizado do erro quadrático - K-means. Os valores em negrito indicam circunstâncias em que a hipótese experimental foi aceita.

A observação dos valores da [Tabela 4](#), que contém os resultados da comparação entre o método sugerido e o algoritmo K-means, indica que a forma canônica da meta-heurística inspirada nas ondas aquáticas foi, na maioria das ocasiões, a estratégia mais apropriada, por ter obtido de forma significativa e em 27 (93,10%) das 29 comparações, a partição mais congruente. Em adição, as respostas da [Tabela 4](#) demonstram que somente em 2 (6,90%) oportunidades não houve diferença estatisticamente significativa entre os métodos de agrupamento, que em nenhuma circunstância o algoritmo K-means alcançou uma resposta mais adequada, e que, em 25 (92,59%) das 27 ocasiões nas quais o algoritmo proposto foi superior, um efeito grande foi verificado.

A comparação entre os resultados do algoritmo inspirado nas ondas aquáticas em sua forma convencional, e as respostas do mesmo algoritmo, acrescido do operador de decaimento, encontra-se retratada por intermédio da [Tabela 5](#). As repostas indicadas na tabela referida, denotam que a hipótese experimental foi aceita em 13 (44,83%) oportunidades, que em 16 (55,17%) ocasiões não houve relevância estatística na diferença entre as subdivisões obtidas por cada método de particionamento, e que, em nenhuma circunstância, a meta-heurística inspirada nas ondas aquáticas, acrescida do operador de decaimento, obteve um erro quadrático mais apropriado. Os valores

Tabela 5: Onda aquática versus onda com decaimento

Coleção	T	z	r	μ_{oa}	μ_{od}
Aggregation	23	-0,4587	-0,1026	0,5017	0,4086
Compound	0	-2,8031	-0,6268	0,8961	0,6582
Pathbased	3	-2,3102	-0,5166	0,5034	0,2824
Spiral	1	-2,5471	-0,5696	0,3317	0,0436
D31	19	-0,8664	-0,1937	0,7210	0,6584
R15	0	-2,6656	-0,5961	0,7727	0,4335
Jain	0	-1,8258	-0,4083	0,2051	0,0000
Flame	0	-1,8258	-0,4083	0,2326	0,0459
Wine	25	-0,2548	-0,0570	0,2632	0,3501
Yeast	0	-2,8031	-0,6268	0,4963	0,1980
Breast	13	-1,4780	-0,3305	0,5649	0,6519
Iris	19	-0,4146	-0,0927	0,0556	0,1186
Glass	23	-0,4587	-0,1026	0,6856	0,5966
Dim032	0	-2,8031	-0,6268	0,7219	0,2510
Dim064	25	-0,2548	-0,0570	0,4493	0,5016
A1	13	-1,4780	-0,3305	0,6245	0,4683
A2	15	-1,2741	-0,2849	0,4429	0,3224
A3	9	-1,8857	-0,4217	0,3472	0,5826
Dim2	5	-0,6742	-0,1508	0,2746	0,2487
Dim3	6	-1,3522	-0,3024	0,4226	0,2520
Dim4	6	-1,6803	-0,3757	0,6104	0,6423
Dim5	6	-1,6803	-0,3757	0,6297	0,4474
Dim6	3	-2,1004	-0,4697	0,8240	0,4835
S1	2	-2,4286	-0,5431	0,8950	0,4696
S2	4	-1,9604	-0,4384	0,8745	0,4866
S3	8	-1,9876	-0,4444	0,7672	0,5174
Balance	0	-2,8031	-0,6268	0,9238	0,3011
CMC	0	-2,6656	-0,5961	1,0000	0,7445
Dermatology	4	-2,3953	-0,5356	0,7218	0,2442

T: estatística teste; z: escore-z; r: tamanho do efeito; μ_{oa} : média do valor normalizado do erro quadrático - onda aquática; μ_{od} : média do valor normalizado do erro quadrático - onda aquática com operador de decaimento. Os valores em negrito indicam circunstâncias em que a hipótese experimental foi aceita.

Tabela 5 indicam por consequência que, embora o algoritmo sugerido tenha alcançado resultados mais satisfatórios, inclusive com a ocorrência de um efeito grande em 10 (76,92%) das 13 ocasiões em que foi mais congruente, houve o predomínio da circunstância em que a diferença entre os erros obtidos por cada algoritmo não teve significância estatística.

A observação das respostas constantes da **Tabela 6**, que incluem os resultados da comparação entre a forma canônica do algoritmo inspirado nas ondas aquáticas, e a meta-heurística fundamentada nas colônias de abelhas artificiais, permitem inferir que o segundo algoritmo foi o mais apropriado. Com efeito, os resultados descritos na tabela indicada expressam que, em 14 (48,28%) oportunidades a estratégia inspirada nas colônias de abelhas artificiais obteve os erros quadráticos mais congruentes, que 8 (27,59%) circunstâncias o algoritmo inspirado nas ondas aquáticas foi o mais apropriado, e que em 7 (24,14%) ocasiões não houve diferença estatisticamente significativa entre os resultados. Os valores referidos na **Tabela 6** demonstram também que, em todas as circunstâncias nas quais o método fundamentado nas ondas aquáticas foi o mais apropriado, o efeito verificado foi grande, embora esta estratégia não tenha sido de forma predominante a mais bem avaliada em subdividir os itens das coleções de objetos entre os grupos.

Tabela 6: Onda aquática versus colônia de abelhas

Coleção	T	z	r	μ_{oa}	μ_{ca}
Aggregation	1	-2,7011	-0,6040	0,2709	0,7866
Compound	0	-2,8031	-0,6268	0,1514	0,8208
Pathbased	0	-2,8031	-0,6268	0,0234	0,9856
Spiral	0	-2,8031	-0,6268	0,0078	0,7274
D31	0	-2,8031	-0,6268	0,8913	0,2932
R15	0	-2,8031	-0,6268	0,8247	0,2531
Jain	0	-2,8031	-0,6268	0,0777	0,9846
Flame	0	-2,8031	-0,6268	0,0056	1,0000
Wine	0	-2,8031	-0,6268	0,2078	0,9524
Yeast	1	-2,7011	-0,6040	0,2471	0,7109
Breast	0	-2,8031	-0,6268	0,0014	0,6376
Iris	0	-2,8031	-0,6268	0,0468	0,9997
Glass	16	-1,1722	-0,2621	0,5187	0,6664
Dim032	19	-0,8664	-0,1937	0,4272	0,5077
Dim064	0	-2,8031	-0,6268	0,3983	0,8517
A1	0	-2,8031	-0,6268	0,8323	0,2783
A2	0	-2,8031	-0,6268	0,8040	0,0887
A3	0	-2,8031	-0,6268	0,8778	0,1499
Dim2	18	-0,5331	-0,1192	0,2746	0,3361
Dim3	17	-0,6516	-0,1457	0,4226	0,5282
Dim4	25	-0,2548	-0,0570	0,6104	0,6556
Dim5	20	-0,2962	-0,0662	0,4426	0,4575
Dim6	6	-0,9435	-0,2110	0,7000	0,9048
S1	0	-2,8031	-0,6268	0,9536	0,1634
S2	0	-2,8031	-0,6268	0,9339	0,2285
S3	0	-2,8031	-0,6268	0,8597	0,1482
Balance	0	-2,8031	-0,6268	0,0102	0,8608
CMC	0	-2,8031	-0,6268	0,0000	0,6466
Dermatology	0	-2,8031	-0,6268	0,3089	0,7183

T: estatística teste; z: escore-z; r: tamanho do efeito; μ_{oa} : média do valor normalizado do erro quadrático - onda aquática; μ_{ca} : média do valor normalizado do erro quadrático - colônia de abelhas artificiais. Os valores em negrito indicam circunstâncias em que a hipótese experimental foi aceita.

Por fim, a análise dos dados da **Tabela 7**, que expressa os resultados da comparação entre a forma convencional da meta-heurística inspirada nas ondas aquáticas e a meta-heurística inspirada no recozimento simulado, permite constatar que em 29 (100,00%) oportunidades o algoritmo sugerido foi o mais adequado, com um efeito caracterizado como grande em todas as ocasiões. As respostas da tabela referida assinalam desta maneira, o predomínio da meta-heurística inspirada nas ondas aquáticas em obter as partições mais congruentes, em uma proporção até então não observada, indicando por conseguinte que a técnica de agrupamento fundamentada no processo térmico de recozimento foi a menos apropriada dentre todas as que foram experimentadas.

6 Conclusões

Este trabalho propôs a aplicação do algoritmo inspirado nas ondas aquáticas à determinação do agrupamento não hierárquico de objetos. De modo específico, o particionamento de uma coleção de itens em um conjunto finito de grupos, que foi abordado como um problema de otimização, que tinha como propósito minimizar o erro quadrático da classificação obtida, foi realizado, pelo algoritmo sugerido e por outros seis métodos descritos na literatura,

Tabela 7: Onda aquática versus recozimento simulado

Coleção	T	z	r	μ_{oa}	μ_{rs}
Aggregation	0	-2,8031	-0,6268	0,9844	0,0111
Compound	0	-2,8031	-0,6268	0,9972	0,0509
Pathbased	0	-2,8031	-0,6268	0,9790	0,0347
Spiral	0	-2,8031	-0,6268	0,9908	0,0524
D31	0	-2,8031	-0,6268	0,9954	0,0030
R15	0	-2,8031	-0,6268	0,9946	0,0076
Jain	0	-2,8031	-0,6268	0,8941	0,0683
Flame	0	-2,8031	-0,6268	0,9786	0,0311
Wine	0	-2,8031	-0,6268	0,6147	0,0690
Yeast	0	-2,8031	-0,6268	0,8513	0,0068
Breast	0	-2,8031	-0,6268	0,9967	0,0175
Iris	0	-2,8031	-0,6268	0,9848	0,0855
Glass	0	-2,8031	-0,6268	0,8018	0,0805
Dim032	0	-2,8031	-0,6268	0,9769	0,0023
Dim064	0	-2,8031	-0,6268	0,9401	0,0044
A1	0	-2,8031	-0,6268	0,9927	0,0045
A2	0	-2,8031	-0,6268	0,9937	0,0015
A3	0	-2,8031	-0,6268	0,9966	0,0017
Dim2	0	-2,8031	-0,6268	0,9777	0,0108
Dim3	0	-2,8031	-0,6268	0,9746	0,0034
Dim4	0	-2,8031	-0,6268	0,9703	0,0022
Dim5	0	-2,8031	-0,6268	0,9736	0,0017
Dim6	0	-2,8031	-0,6268	0,9898	0,0014
S1	0	-2,8031	-0,6268	0,9972	0,0019
S2	0	-2,8031	-0,6268	0,9961	0,0011
S3	0	-2,8031	-0,6268	0,9937	0,0015
Balance	0	-2,8031	-0,6268	0,9957	0,0137
CMC	0	-2,8031	-0,6268	1,0000	0,0033
Dermatology	0	-2,8031	-0,6268	0,9649	0,0629

T: estatística teste; z: escore-z; r: tamanho do efeito; μ_{oa} : média do valor normalizado do erro quadrático - onda aquática; μ_{rs} : média do valor normalizado do erro quadrático - recozimento simulado. Os valores em negrito indicam circunstâncias em que a hipótese experimental foi aceita.

para 29 coleções de referência, com número de grupos conhecido.

Os resultados alcançados por cada algoritmo, que foram analisados por intermédio do teste dos postos sinalizados de Wilcoxon, indicaram que o método proposto, isto é, fundamentado na estrutura convencional do algoritmo inspirado nas ondas aquáticas, foi, na maioria das circunstâncias, o mais apropriado. Em particular, o algoritmo sugerido se sobressaiu de forma estatisticamente significativa ao ser comparado às meta-heurísticas enxame de partículas, algoritmo genético e recozimento simulado, além de ter sido mais oportuno também do que o algoritmo *K-means*. A avaliação das respostas obtidas indicaram adicionalmente que, não houve significância estatística nas diferenças entre as classificações obtidas pela meta-heurística inspirada nas ondas aquáticas em sua forma canônica, e o mesmo método que incluía o operador de decaimento, e que o método proposto foi menos congruente do que a meta-heurística fundamentada nas colônias de abelhas artificiais, em subdividir as coleções de objetos manipuladas em um conjunto delimitado de grupos.

Como a análise conduzida sobre os resultados, foi fundamentada em técnicas estatísticas, admite-se que a comparação entre as respostas determinadas pelos métodos de particionamento foi efetuada de maneira mais rigorosa, do que se apenas as médias dos erros quadráticos estives-

sem sendo contrastadas diretamente, permitindo inclusive evidenciar circunstâncias em que não houve diferenças expressivas entre as classificações, que, de modo distinto, possivelmente não seriam assinaladas. Desta forma, considera-se o método proposto como uma estratégia tão congruente quanto outras já descritas na literatura, que se distingue por ser fundamentada apenas na estrutura convencional da metaheurística inspirada nas ondas aquáticas, e que, consoante os achados retratados neste estudo, não experimenta restrições significativas em razão dessa conduta.

Como limitações deste trabalho referem-se o número reduzido de coleções submetidas aos algoritmos de classificação, e a utilização de apenas uma medida, isto é, o erro quadrático, para contrapor as subdivisões estabelecidas por cada método. Como opções para estudos posteriores sugerem-se, o emprego de um volume maior de coleções, a aplicação de outros critérios aptos a precisar a congruência das partições determinadas pelos algoritmos, ou ainda a incorporação de variações nas estratégias admitidas pelos métodos de classificação especificados neste manuscrito.

Referências

- Aboubi, Y., Drias, H. and Kamel, N. (2016). Bat-clara: Bat-inspired algorithm for clustering large applications, *IFAC-PapersOnLine* 49(12): 243–248. <https://doi.org/10.1016/j.ifacol.2016.07.607>.
- Al-Sultan, K. S. (1995). A tabu search approach to the clustering problem, *Pattern Recognition* 28(9): 1443–1451. [https://doi.org/10.1016/0031-3203\(95\)00022-R](https://doi.org/10.1016/0031-3203(95)00022-R).
- Aloise, D., Deshpande, A., Hansen, P. and Popat, P. (2009). Np-hardness of euclidean sum-of-squares clustering, *Machine Learning* 75(2): 245–248. <https://doi.org/10.1007/s10994-009-5103-0>.
- Bishop, C. (1995). *Neural networks for pattern recognition*, Oxford University Press, New York, New York.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, Academic Press, New York.
- Cui, X., Potok, T. E. and Palathingal, P. (2005). Document clustering using particle swarm optimization, *Proceedings 2005 IEEE Swarm Intelligence Symposium, 2005. SIS 2005*, pp. 185–191. <https://doi.org/10.1109/SIS.2005.1501621>.
- Cura, T. (2012). A particle swarm optimization approach to clustering, *Expert Systems with Applications* 39(1): 1582–1588. <https://doi.org/10.1016/j.eswa.2011.07.123>.
- Derrac, J., García, S., Molina, D. and Herrera, F. (2011). A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, *Swarm and Evolutionary Computation* 1(1): 3–18. <https://doi.org/10.1016/j.swevo.2011.02.002>.
- Dua, D. and Graff, C. (2017). UCI machine learning repository. <http://archive.ics.uci.edu/ml>.

- Field, A. (2009). *Discovering Statistics Using SPSS*, SAGE Publications Ltd, California.
- Fränti, P. and Sieranoja, S. (2018). K-means properties on six clustering benchmark datasets, *Applied Intelligence* **48**(12): 4743–4759. <https://doi.org/10.1007/s10489-018-1238-7>.
- Gonzalez, T. F. (2007). *Handbook of Approximation Algorithms and Metaheuristics*, Chapman & Hall/CRC Computer & Information Science, Boca Raton.
- Hansen, P. and Jaumard, B. (1997). Cluster analysis and mathematical programming, *Mathematical Programming* **79**: 191–215. <https://doi.org/10.1007/BF02614317>.
- Ibrahim, A. M., Tawhid, M. and Ward, R. K. (2020). A binary water wave optimization for feature selection, *International Journal of Approximate Reasoning* **120**: 74–91. <https://doi.org/10.1016/j.ijar.2020.01.012>.
- Inkaya, T., Kayaligil, S. and Ozdemirel, N. E. (2015). Ant colony optimization based clustering methodology, *Applied Soft Computing* **28**: 301–311. <https://doi.org/10.1016/j.asoc.2014.11.060>.
- Isimeto, R., Yinka-Banjo, C., Uwadia, C. O. and Alienyi, D. C. (2017). An enhanced clustering analysis based on glowworm swarm optimization, *2017 IEEE 4th International Conference on Soft Computing Machine Intelligence (ISCMi)*, pp. 42–49. <https://doi.org/10.1109/ISCMi.2017.8279595>.
- Jain, A. K. and Dubes, R. C. (1998). *Algorithms for Clustering Data*, Prentice Hall, New Jersey.
- Jain, A. K., Duin, R. P. W. and Mao, J. (2000). Statistical pattern recognition: a review, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(1): 4–37. <https://doi.org/10.1109/34.824819>.
- Jain, A. K., Murty, M. N. and Flynn, P. J. (1999). Data clustering: A review, *ACM Computing Surveys* **31**(3): 264–323. <https://doi.org/10.1145/331499.331504>.
- Karaboga, D. and Ozturk, C. (2011). A novel clustering approach: Artificial bee colony (abc) algorithm, *Applied Soft Computing* **11**(1): 652–657. <https://doi.org/10.1016/j.asoc.2009.12.025>.
- Kaufman, L. and Rousseeuw, P. J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, Inc., Hoboken, New Jersey.
- Kaur, A. and Kumar, Y. (2021). A new metaheuristic algorithm based on water wave optimization for data clustering, *Evolutionary Intelligence*. <https://doi.org/10.1007/s12065-020-00562-x>.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, University of California Press, Berkeley, California, pp. 281–297.
- Maulik, U. and Bandyopadhyay, S. (2000). Genetic algorithm-based clustering technique, *Pattern Recognition* **33**(9): 1455–1465. [https://doi.org/10.1016/S0031-3203\(99\)00137-5](https://doi.org/10.1016/S0031-3203(99)00137-5).
- Selim, S. Z. and Alsultan, K. (1991). A simulated annealing algorithm for the clustering problem, *Pattern Recognition* **24**(10): 1003–1008. [https://doi.org/10.1016/0031-3203\(91\)90097-0](https://doi.org/10.1016/0031-3203(91)90097-0).
- Singh, G., Rattan, M., Gill, S. S. and Mittal, N. (2019). Hybridization of water wave optimization and sequential quadratic programming for cognitive radio system, *Soft Comput* **23**: 7991–8011. <https://doi.org/10.1007/s00500-018-3437-x>.
- Xie, H., Zhang, L., Lim, C. P., Yu, Y., Liu, C., Liu, H. and Walters, J. (2019). Improving k-means clustering with enhanced firefly algorithms, *Applied Soft Computing* **84**: 105763. <https://doi.org/10.1016/j.asoc.2019.105763>.
- Xu, R. and Wunsch, D. C. (2009). *Clustering*, IEEE Press, Piscataway, New Jersey.
- Yan, Z., Zhang, J. and Tang, J. (2021). Path planning for autonomous underwater vehicle based on an enhanced water wave optimization algorithm, *Mathematics and Computers in Simulation* **181**: 192–241. <https://doi.org/10.1016/j.matcom.2020.09.019>.
- Zhao, F., Zhang, L., Zhang, Y., Ma, W., Zhang, C. and Song, H. (2020). A hybrid discrete water wave optimization algorithm for the no-idle flowshop scheduling problem with total tardiness criterion, *Expert Systems with Applications* **146**: 113166. <https://doi.org/10.1016/j.eswa.2019.113166>.
- Zheng, Y.-J. (2015). Water wave optimization: A new nature-inspired metaheuristic, *Computers & Operations Research* **55**: 1–11. <https://doi.org/10.1016/j.cor.2014.10.008>.
- Zheng, Y.-J., Lu, X.-Q., Du, Y.-C., Xue, Y. and Sheng, W.-G. (2019). Water wave optimization for combinatorial optimization: Design strategies and applications, *Applied Soft Computing* **83**: 105611. <https://doi.org/10.1016/j.asoc.2019.105611>.
- Zhou, X.-H., Zhang, M.-X., Xu, Z.-G., Cai, C.-Y., Huang, Y.-J. and Zheng, Y.-J. (2019). Shallow and deep neural network training by water wave optimization, *Swarm and Evolutionary Computation* **50**: 100561. <https://doi.org/10.1016/j.swevo.2019.100561>.
- Zhu, W., Jiang, J., Song, C. and Bao, L. (2012). Clustering algorithm based on fuzzy c-means and artificial fish swarm, *Procedia Engineering* **29**: 3307–3311. <https://doi.org/10.1016/j.proeng.2012.01.485>.
- Zou, W., Zhu, Y., Chen, H. and Sui, X. (2010). A clustering approach using cooperative artificial bee colony algorithm, *Discrete Dynamics in Nature and Society*. <https://doi.org/10.1155/2010/459796>.