Revista de Informática Teórica e Aplicada - RITA - ISSN 2175-2745
Vol. 28, Num. 02 (2021) 50-65

RESEARCH ARTICLE

Optimization inspired on herd immunity applied to non-hierarchical grouping of objects

Otimização inspirada na imunidade de rebanho aplicada ao agrupamento não hirárquico de objetos

Alfredo Silveira Araújo Neto[1]*

Abstract: Characterized as one of the most important operations related to data analysis, one non-hierarchical grouping consists of, even without having any information about the elements to be classified, establish upon a finite collection of objects, the partitioning of the items that constitute it into subsets or groups without intersecting, so that the elements that are part of a certain group are more similar to each other than the items that belong to distinct group. In this context, this study proposes the application of a meta-heuristic inspired by herd immunity to the determination of the non-hierarchical grouping of objects, and compares the results obtained by this method with the answers provided by four other grouping strategies, described in the literature. In particular, the resulting arrangements of the classification of 33 benchmark collections, performed by the suggested algorithm, by the metaheuristic inspired by the particle swarm, by the genetic algorithm, by the *K-means* algorithm and by the meta-heuristic inspired by the thermal annealing process, were compared under the perspective of 10 different evaluation measures, indicating that the partitions established by the meta-heuristic inspired by the herd immunity may, in certain respects, be more favorable than the classifications obtained by the other clustering methods.
Keywords: Cluster analysis — metaheuristic — bio-inspired computing

Resumo: Caracterizando-se com uma das mais importantes operações relacionadas à análise de dados, um agrupamento não hierárquico consiste em, mesmo sem dispor de quaisquer informações acerca dos elementos a classificar, estabelecer sobre uma coleção finita de objetos, o particionamento dos itens que a constituem em subconjuntos ou grupos sem interseção, de forma que os elementos que fazem parte de um determinado grupo são mais semelhantes entre si, do que os itens que pertencem a grupo distinto. Neste contexto, este estudo propõe a aplicação de uma meta-heurística inspirada na imunidade de rebanho à determinação do agrupamento não-hierárquico de objetos, e compara os resultados obtidos por este método com as respostas fornecidas por outras quatro estratégias de agrupamento, descritas na literatura. Em particular, os arranjos resultantes da classificação de 33 coleções de referência, realizada pelo algoritmo sugerido, pela meta-heurística inspirada no enxame de partículas, pelo algoritmo genético, pelo algoritmo *K-means* e pela meta-heurística inspirada no processo térmico de recozimento, foram comparados sob a perspectiva de 10 medidas de avaliação distintas, indicando que as partições estabelecidas pela meta-heurística inspirada na imunidade de rebanho podem, sob determinados aspectos, ser mais favoráveis do que as classificações obtidas pelos outros métodos de agrupamento.
Palavras-Chave: Análise por agrupamentos — meta-heurística — computação bioinspirada

[1] *Techway Informática Ltda., Fortaleza - Ceará, Brasil*
***Autor correspondente**:alfredosilveira@yahoo.com.br
DOI: http://dx.doi.org/10.22456/2175-2745.107478 • **Received:** 12/09/2020 • **Accepted:** 26/07/2021

1. Introdução

Os rápidos avanços das tecnologias relacionadas à coleta e ao armazenamento de dados em formato eletrônico têm permitido às organizações o acúmulo de um volume de informações demasiadamente elevado. Entretanto, por consequência da alta dimensionalidade, complexidade e, ocasionalmente, distribuição geográfica heterogênea desses dados digitais, a obtenção de inferências úteis com base nesses grandes bancos de informações tem-se apresentado como uma atividade bastante difícil de ser empreendida, que frequentemente exige a aplicação de métodos não usuais. Desta forma e consoante

[1], o emprego de técnicas sofisticadas de análise de dados, capazes de atuar sobre grandes respositórios a fim de identificar padrões consistentes que, de outra forma, permaneceriam desconhecidos, tem sido frequentemente experimentado.

Uma das mais importantes operações relacionadas à análise de dados, consiste em classificar ou agrupar os objetos em um conjunto de categorias ou grupos, de forma que os elementos relacionados a um mesmo grupo apresentem características equivalentes, de acordo com um determinado critério. Com efeito, verifica-se que a classificação desempenha um importante e indispensável papel no desenvolvimento humano, haja vista que para melhor compreender um novo objeto ou fenômeno, as pessoas frequentemente procuram identificar as características descritivas dos mesmos, no intuito de comparará-las com àquelas que pertencem a objetos ou fenômenos conhecidos [2].

Fundamentalmente, os sistemas de classificação podem ser decompostos em sistemas supervisionados e sistemas não supervisionados. A classificação supervisionada emprega uma coleção de objetos previamente qualificados, e o problema consiste em agrupar novos objetos, ainda não categorizados, com base nas informações obtidas por meio dos elementos já classificados. Na classificação não supervisionada, também denominada de análise de agrupamentos ou análise exploratória de dados, não existem rótulos de dados dispóniveis, e o objetivo é decompor uma coleção finita de objetos não categorizados, em um conjunto finito de grupos "naturais". O emprego de técnicas de análise de agrupamentos provém da necessidade de se investigar dados de natureza desconhecida sem que exista qualquer conhecimento prévio acerca dos mesmos, promovendo a divisão dos objetos em uma coleção de subgrupos relativamente homogêna, com base em uma medida, frequentemente referenciada como dissimilaridade, distância ou similaridade. Desde que a medida de dissimilaridade constitui um aspecto essencial para os procedimentos de classificação, a sua seleção deve ser empreendida de forma criteriosa e diligente, inclusive considerando a natureza e a escala dos atributos empregados na representação dos elementos. Com efeito, uma das mais difundidas métricas empregadas para determinar a dissimilaridade entre objetos retratados por intermédio de atributos contínuos, consiste na distância Euclidiana, a qual é denotada pela expressão

$$d(o_i,o_j) = \left[\sum_{k=1}^{d}((o_{i,k}-o_{j,k})^2\right]^{(1/2)} = \| o_i - o_j \|^2 \quad (1)$$

onde $o_i = (a_{i1}, a_{i2}, \ldots, a_{id})$ e $o_j = (a_{j1}, a_{j2}, \ldots, a_{jd}) \in \mathbb{R}^d$, e, a_{id} e a_{jd} são denominados atributos, dimensões ou características [2, 3, 4, 5].

Diferentes críterios de agrupamento, algoritmos de agrupamento distintos, ou ainda parâmetros distintos empregados para o mesmo algoritmo, podem resultar em partições completamente desiguais, a exemplo dos seres humanos, que quando classificados conforme sua etnia, região, situação econômica, grau de instrução, peso ou altura, podem originar grupos integralmente distintos [2, 4].

De acordo com [6, 4], os problemas de classificação não supervisionada podem ser categorizados em não exclusivos e exclusivos. Na classificação não exclusiva, cada objeto pode estar associado simultaneamente a mais de grupo, assim, mediante uma relação na qual um mesmo elemento compartilha alguma fração de diferentes grupos, a função de pertinência que denota o vínculo entre o objeto e o grupo, admite valores no intervalo compreendido entre 0 e 1. Sob outra perspectiva, na classificação exclusiva, cada elemento da coleção de objetos, que nesta circunstância vincula-se a exatamente um subconjunto ou agrupamento, é associado por intermédio de uma relação de pertinência binária, na qual o valor 1 indica que o objeto está contido no grupo, enquanto que o valor 0 denota que o objeto não está incluído. Adicionalmente e de acordo com o tipo de estrutura imposta sobre os dados, os problemas de classificação exclusivos são subdivididos em hierárquicos e não-hierárquicos. As classificações ou agrupamentos hierárquicos, correspondem a um procedimento que tem como objetivo estabelecer uma sequência de partições aninhadas com base nas dissimilaridades entre os objetos, enquanto que as categorizações ou agrupamentos não-hierárquicos, consistem na divisão da coleção de objetos em subconjuntos sem interseção, de modo que cada elemento esteja precisamente em um subconjunto.

O problema do agrupamento não-hierárquico pode ser formalmente definido como se segue. Dado um conjunto de objetos $O = \{o_i\}, i = 1, \ldots, n$, no qual cada elemento é descrito por intermédio de um vetor d-dimensional $o_i = (a_{i1}, a_{i2}, \ldots, a_{id}) \in \mathbb{R}^d$, tem-se que um método de agrupamento não-hierárquico procura estabelecer uma K-partição de O, $G = \{g_j\}$, $j = 1, \ldots, K$ $(K < n)$, de modo que: cada grupo, g_j, contenha pelo menos um objeto, $g_j \neq \emptyset$, $j = 1, \ldots, K$; cada objeto, o_i, seja pertinente a um determinado grupo, $\bigcup_{j=1}^{K} g_j = O$; cada objeto, o_i, seja pertinente a exatamente um grupo, $g_{j1} \bigcap g_{j2} = \emptyset$, $j1 \neq j2$. O valor K pode ou não ser conhecido e um critério de avaliação, classificado como local ou global, deve ser adotado. Um critério global representa cada grupo por meio de um protótipo e associa os objetos aos grupos segundo o protótipo mais similar. Sob outra perspectiva, um critério local constitui os grupos utilizando a própria estrutura dos dados, a exemplo de grupos que podem ser estabelecidos a partir da identificação de regiões de alta densidade no espaço de objetos, ou por meio da associação de um objeto e seus K vizinhos mais próximos, a um mesmo grupo em particular [2, 7].

A solução teórica para o problema do agrupamento não-hierárquico é direta e de fácil entendimento: define-se um critério, avaliam-se todas as partições constuídas de K grupos e por fim seleciona-se a partição que otimiza o critério estipulado. A primeira dificuldade desta abordagem reside em estabelecer um critério que represente o conceito de grupos em uma formulação matemática adequada, haja vista que os critérios de avaliação são bastante dependentes dos parâmetros

do problema, devem ser simples sob a perspectiva do esforço computacional, porém complexos o suficiente para refletir as diversas estruturas dos dados. A segunda dificuldade com este método consiste no número de partições, que é excessivamente elevado mesmo para um moderado número de objetos, tornando imprática a avaliação de um critério, mesmo que simples, em relação a todas as partições. De fato, o número de diferentes partições de n objetos em K grupos é determinado pela expressão

$$S(n,K) = \frac{1}{K!}\sum_{j=1}^{K}(-1)^{K-j}\binom{K}{j}(j)^n \tag{2}$$

e, embora existam somente 34.105 partições distintas de 10 objetos em quatro grupos, este número eleva-se para aproximadamente 11.259.666.000 se 19 objetos precisam ser particionados em quatro grupos [4].

Com o propósito de evitar esta súbita elevação combinatória, a função critério pode ser avaliada somente para um subconjunto admissível de partições. Neste sentido, a abordagem mais frequente consiste em promover a otimização da função critério por meio do emprego de um método iterativo, que inicia com uma partição preliminar e desloca os objetos de um grupo para o outro com o objetivo de melhorar o parâmetro de avaliação. A mais intuitiva e mais reiterada função critério empregada pelas técnicas de agrupamento não-hierárquico consiste no erro quadrático, que para uma K-partição de $O = \{o_i\}, i = 1,\ldots,n$ em $G = \{g_j\}$, $j = 1,\ldots,K$ $(K \leq n)$, pode ser estabelecido por meio da igualdade

$$e^2(O,G) = \sum_{j=1}^{K}\sum_{i=1}^{n_j} \parallel o_i^j - m_j \parallel^2 \tag{3}$$

onde o_i^j denota o i-ésimo objeto associado ao j-ésimo grupo, n_j expressa o número de objetos do j-ésimo grupo e

$$m_j = \frac{1}{n_j}\sum_{i=1}^{n_j} o_i \tag{4}$$

retrata a média ou protótipo do j-ésimo grupo [4, 3].

Muitos dos problemas de otimização, sejam de natureza prática ou teórica, consistem em investigar e identificar o melhor arranjo de um dado conjunto de variáveis a fim de alcançar um determinado resultado. As variáveis manipuladas durante o procedimento de identificação, podem ser de natureza discreta ou contínua, e, dentre as estratégias que se propõem a examinar os arranjos retratados por variávies contínuas, situam-se os problemas de otimização combinatória. Formalmente, dados um conjunto de variáveis $V = \{v_i\}, i = 1,\ldots,n$, sujeitas a determinadas restrições e pertencentes aos domínios $D_1 \ldots D_n$, uma função critério $f : D_1 \times \cdots \times D_n \rightarrow \mathbb{R}^+$ a ser minimizada e um conjunto S de todos os resultados viáveis, solucionar um problema de otimização combinatória $P = (S, f)$ consiste em identificar uma solução $s^* \in S$ que conduz ao valor mínimo do critério de avaliação, ou seja, $f(s^*) \leq f(s)$, $\forall s \in S$. Por conta de sua importância prática, os problemas de otimização combinatória têm sido frequentemente endereçados por intermédio de técnicas exatas e aproximativas. As estratégias exatas, a despeito de serem aptas a invariavelmente identificar soluções que resultam no valor mínimo da função critério, podem eventualmente requerer um tempo de processamento indamissível. Em contrapartida, os métodos aproximativos, não obstante sejam capazes de determinar soluções a um custo de processamento significativamente menor, se comparados às estratégias exatas, são inaptos em assegurar que os resultados alcançados conduzam ao valor mínimo do critério de avaliação estebelecido [8].

Minimizar a função critério determinada pela equação 3 constitui um problema *NP*-difícil mesmo para $K = 2$, desta forma e considerando que a maioria das operações de particionamento pode ser eventualmente caracterizada como problemas de otimização combinatória, o esforço computacional torna-se crítico quando coleções de objetos mais extensas precisam ser examinadas [4, 3, 9].

Além da introdução, este trabalho compreende quatro seções adicionais, as quais estão organizadas conforme indicado a seguir. A seção 2 descreve o método meta-heurístico inspirado na interrupção da cadeia de transmissão de uma doença contagiosa e refere como o processo de contaminação, recuperação e, eventualmente, óbito, dos indivíduos de uma população suscetível a uma infecção, pode ser estruturado a fim estabelecer uma estratégia capaz de ser aplicada à resolução de problemas de otimização combinatória. A seção 3 refere as adequações efetuadas sobre o algoritmo original com o propósito de aplicá-lo à determinação do agrupamento não hierárquico. A seção 4 apresenta os experimentos de avaliação que foram conduzidos com a intenção de comparar a meta-heurística inspirada na imunidade de rebanho com outros métodos de agrupamento descritos na literatura. Por fim, a seção 5 relata, conforme os resultados dos experimentos, as conclusões obtidas.

2. Otimização inspirada na imunidade de rebanho

As meta-heurísticas representam algoritmos de aproximação, capazes de obter resultados satisfatórios quando aplicados à resolução de problemas de otimização. Constituem estratégias simples, robustas e flexíveis, muitas vezes inspiradas em fenômenos naturais, que eficientemente investigam o conjunto de resultados viáveis e que podem ser empregadas a uma grande variedade de questões, a fim de determinar soluções de boa qualidade em um tempo de processamento aceitável[10, 8].

O algortimo de otimização inspirado na imunidade de rebanho consiste em uma meta-heurística que reproduz o processo de aquisição de imunidade de uma população, quando submetida ao contato com um vírus para o qual os indivíduos que a constituem ainda não possuem qualquer proteção natural

[11].

De acordo com [11, 12], os vírus são agentes infecciosos que se propagam aceleradamente entre os seres de uma população e que, em geral, são contidos por intermédio de uma vacina, capaz de fornecer imunidade ativa contra um microrganismo causador de doenças. Porém, tendo em conta que em determinadas situações a manifestação de um vírus novo, para o qual ainda não há nenhuma vacina disponível, impossibilita o estabelecimento de uma imunidade generalizada entre os indivíduos, as organizações de saúde admitem, com o propósito de promover a interrupção da contaminação, a aplicação de duas condutas. A primeira, institui o saneamento, a desinfecção, o isolamento social dos indivíduos infectados e a reclusão dos indivíduos sadios com os quais os indivíduos infectados tiveram algum tipo de contato, pelo período máximo de incumbação da doença. Já a segunda, faz uso do princípio de imunidade de rebanho a fim de controlar a disseminação do agente infeccioso.

A imunidade de rebanho é fundamentada na percepção de que os vírus são parasitas que invariavelmente necessitam de um organismo hospedeiro para se multiplicar. Desta forma, se há na população uma quantidade considerável de indivíduos imunes à infecção, o vírus deixa de ser transmitido aos indivíduos que ainda não se contaminaram, o que, por consequência, resulta na eliminação da circulação do agente infeccioso entre aqueles que constituem a população. Um indivíduo ainda não contaminado pelo vírus é denominado de indivíduo suscetível, enquanto que um indivíduo já infectado é nomeado como indivíduo transmissor, que poderá se recuperar e portanto se tornar imune ou, infelizmente, morrer. Considerando que fatores desfavoráveis conduzem a uma resposta imunológica menos eficaz, a presença de outras doenças, a exemplo de diabetes e cancer, ou ainda de problemas cardiovasculares, que em geral acometem os indivíduos de idade mais avançada, faz com que os idosos eventualmente infectados tenham uma menor perspectiva de restabelecimento [11, 13].

Usualmente, a proporção de indivíduos imunes que deve existir em uma população, a fim de imprimir a erradicação do contágio descontrolado de uma doença, é determinada por intermédio da expressão $(R_o - 1)/R_o$, onde R_o representa o número médio de infecções secundárias que uma pessoa contaminada pode ocasionar, em uma população constituída por indivíduos suscetíveis. Desta maneira, se, por exemplo, o valor de R_o estiver entre 2 e 3, a imunidade de rebanho somente será alcançada quando a razão dos indivíduos resistentes ao vírus estiver no intervalo compreendido entre 50% e 67%. Assim, ainda que indivíduos transmissores sejam capazes de propagar o vírus, a existência de uma parcela menor de indivíduos ainda suscetíveis proporcionará a interrupção da circulação da infecção, e é fundamentada neste princípio, que a modelagem matemática do algoritmo inspirado na imunidade de rebanho é constituída [11, 13, 14].

Os indivíduos que compreendem a população manipulada pelo algoritmo são classificados como suscetíveis, infectados ou imunes, e, a fim de proporcionar aplicação do método sob a perspectiva de uma estratégia de otimização, institui-se na população uma divisão determinada conforme a seguinte proporção. O número de indivíduos suscetíveis é inicialmente evelado, enquanto que a quantidade de indíviduos infectados, que primeiramente é reduzida, expande-se acelaradamente, se as recomendações de distanciamento social não forem observadas, até que todos os infectados estejam imunes, por terem se recuperado, ou mortos. A princípio, não há indivíduos imunes e sua quantidade, que é dependente dos infectados que se recuperam, eleva-se de forma que nas últimas iterações do algoritmo passem a constituir a maior proporção, o que por consequência propicia o desaparecimento da epidemia. Observa-se adicionalmente que o distanciamento social, admitido pelas organizações de saúde como uma estratégia eficaz em reprimir a propagação do vírus, é estruturada como uma operação que afere a diferença entre um dado indivíduo e outro qualquer, selecionado da população, que nesta circunstância poderá ser suscetível, infectado ou imune [11].

De acordo com [11], as etapas do método estruturado consoante o princípio da imunidade de rebanho, expressas no algoritmo 1, podem ser descritas conforme a seguir:

1. Estabelecer a função critério e os demais parâmetros de execução: nesta etapa o problema de otimização é formulado no contexto de uma relação matemática, a qual consiste em minimizar $f(x)$ $x \in [lb, ub]$, onde $f(x)$ denota a função critério ou razão de imunidade, calculada para cada indivíduo $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$, e $x_i \in [lb_i, ub_i]$ retrata o gene, atributo ou variável de decisão que caracteriza cada indivíduo. Os parâmetros de execução são constituídos pelo número preliminar de indivíduos infectados, npi, número máximo de iterações do algoritmo, nmi, número de indivíduos da população, nip, e número de genes de cada indivíduo, d, além da taxa de propagação, tp, a qual determina a celeridade com o que vírus contamina os indivíduos da população, e da idade máxima dos indivíduos infectados, im, a qual demarca o limite da existência que um indivíduo contaminado e não recuperado pode possuir, até que seja considerado morto;

2. Inicializar a população de indivíduos: compreende a inicialização aleatória da população, armazenada em uma matriz de indivíduos I bidimensional $d \times nip$

$$I = \begin{bmatrix} x_1^1 & x_2^1 & \ldots & x_d^1 \\ x_1^2 & x_2^2 & \ldots & x_d^2 \\ \vdots & \vdots & \ldots & \vdots \\ x_1^{nip} & x_2^{nip} & \ldots & x_d^{nip} \end{bmatrix}$$

na qual cada linha j de I retrata um indivíduo x^j cujos atributos são determinados por intermédio da expressão $x_i^j = lb_i + (ub_i - lb_i)U(0,1)$, $\forall i = 1, 2, \ldots d$, onde lb_i e ub_i denotam respectivamente os limites inferior e superior dos valores que o atributo x_i pode admitir e $U(0,1)$

corresponde a um número aleatório situado no intervalo $[0,1]$. Adicionalmente, o valor da função critério é determinado para cada indivíduo, assim como um vetor de status S de dimensão nip

$$S = \begin{bmatrix} s^1 \\ s^2 \\ \vdots \\ s^{nip} \end{bmatrix}$$

com a ressalva de que para S, npi de seus elementos serão aletaoriamente inicializados com um, a fim de designar os indivíduos inicialmente infectados pelo vírus, enquanto que os demais terão valor zero, estabelecendo os indivíduos suscetíveis à contaminação;

3. Promover a evolução da imunidade de rebanho: consiste na operação que imprime o melhoramento da capacidade defensiva dos indivíduos x^j, por intermédio de operações que, eventualmente, empreendem alterações sobre os genes x_i^j de x^j, conforme o distanciamento social, que por sua vez é sujeito a três critérios dependentes da taxa de propagação da infecção tp,

$$x_i^j(t+1) = \begin{cases} x_i^j(t) & r \geq tp \\ C(x_i^j(t)) & r < tp/3 \\ N(x_i^j(t)) & r < 2tp/3 \\ R(x_i^j(t)) & r < tp \end{cases} \quad (5)$$

onde r denota um número pertencente ao intervalo $[0,1]$, selecionado aleatoriamente. Se a escolha de r resultar em número situado no intervalo $[0, tp/3)$ o novo valor do gene x_i^j será modificado por um distanciamento social determinado pela diferença entre o valor do gene e o gene obtido de um indivíduo infectado, ou seja,

$$x_i^j(t+1) = C(x_i^j(t)), \forall i = 1,2,\ldots d \quad (6)$$

onde

$$C(x_i^j(t)) = x_i^j(t) + r(x_i^j(t) - x_i^c(t)), \forall i = 1,2,\ldots d \quad (7)$$

e $x_i^c(t)$ expressa um gene selecionado aleatoriamente dentre os indivíduos infectados x^c, com base no vetor de status S, de forma que $c = \{i | s^i = 1\}$. Se r estiver no intervalo $[tp/3, 2tp/3)$ o novo valor do gene x_i^j será alterado por um distanciamento social definido pela diferença entre o valor do gene e o gene obtido de um indivíduo suscetível, isto é,

$$x_i^j(t+1) = N(x_i^j(t)), \forall i = 1,2,\ldots d \quad (8)$$

onde

$$N(x_i^j(t)) = x_i^j(t) + r(x_i^j(t) - x_i^m(t)), \forall i = 1,2,\ldots d \quad (9)$$

e $x_i^m(t)$ retrata um gene selecionado aleatoriamente dentre os indivíduos suscetíveis x^m, considerando o vetor de status S, de forma que $m = \{i | s^i = 0\}$. Por fim, se r estiver no intervalo $[2tp/3, tp)$ o novo valor do gene x_i^j será modificado por um distanciamento social estabelecido pela diferença entre o valor do gene e o gene obtido de um indivíduo imune, ou seja,

$$x_i^j(t+1) = R(x_i^j(t)), \forall i = 1,2,\ldots d \quad (10)$$

onde

$$R(x_i^j(t)) = x_i^j(t) + r(x_i^j(t) - x_i^v(t)), \forall i = 1,2,\ldots d \quad (11)$$

e $x_i^v(t)$ consiste em um gene selecionado do indivíduo imune x^v mais apto, em termos da função da critério ou razão de imunidade, com base no vetor de status S, de modo que $f(x^v) \leq f(x^j), \forall j \in \{k | s^k = 2\}$.

4. Atualizar a imunidade de rebanho: a razão de imunidade $f(x^j(t+1))$ de cada indivíduo modificado $x^j(t+1)$ é calculada e comparada com a razão de imunidade $f(x^j(t))$ do indivíduo que o originou. Se $f(x^j(t+1)) < f(x^j(t))$, então o indivíduo modificado substiutui o indivíduo que o originou. Caso contrário, não há substituição e a idade a^j do indivíduo que originou o indivíduo modificado é incrementada em um. Adicionalmente, o status s^j do novo indivíduo x^j, seja ele resultado de uma substituição ou não, é modificado por intermédio da operação $s^j = 1$, se $f(x^j(t+1)) < (x^j(t+1))/\Delta f(x)$, $s^j = 0$ e $w(x^j(t+1)) = 1$, ou pela atribuição $s^j = 2$, se $f(x^j(t+1)) > f(x^j(t+1))/\Delta f(x)$ e $s^j = 1$, onde $\Delta f(x)$ denota o valor médio da função critério ou razão de imunidade da população, $\sum_{i=1}^{nip} f(x_i)/nip$, e w consiste em uma função binária que resulta em um, quando o novo indivíduo possuir qualquer um dos seus genes originado de um dos indivíduos infectados. A razão de imunidade de cada indivíduo será, por meio das iterações do algoritmo, modificada conforme as operações dependentes do distanciamento social, com a tendência de que os indivíduos com melhor ação defensiva contra o vírus sejam preservados, conduzindo a população a aproximar-se da proporção que determina a imunidade de rebanho.

5. Promover o óbtido dos indivíduos irrecuperáveis: se a razão de imunidade modificada, $f(x^j(t+1))$, de um indivíduo x^j infectado, $s^j = 1$, não for aproprimorada

após um determinado número de iterações subsequentes, conforme especificado em *im*, isto é, $a^j \geq im$, então o indivíduo será considerado morto. Para substituí-lo, o algoritmo conceberá, aleatoriamente, um novo indivíduo, ou seja, $x_i^j(t+1) = lb_i + (ub_i - lb_i)U(0,1)$, $\forall i = 1,2,\ldots d$. Adicionalmente, e a fim de proporcionar uma maior diversificação da população, a idade a^j e o status s^j do novo indivíduo serão assinalados com zero;

6. Interromper a execução: o algoritmo irá repetir as operações descritas nas etapas 3, 4 e 5 até que o critério de interrupção, em geral determinado pelo número máximo de iterações, seja alcançado. Ao final da execução, os indivíduos suscetíveis e imunes irão prevalecer na população, ao passo que os indivíduos infectados irão se extinguir.

Com o propósito de estabelecer o melhor arranjo dos parâmetros de execução do método descrito em [11], os seus autores conduziram um conjunto de ensaios no qual 23 funções matemáticas, frequentemente referidas na literatura, foram avaliadas. Os resultados auferidos sugeriram que o número preliminar de indivíduos infectados, *npi*, o número máximo de iterações, *nmi*, o número de indivíduos da população, *nip*, a taxa de propagação da infecção, tp, e a idade máxima dos indivíduos infectados, *im*, deveriam ser, respectivamente, 1, 100.000, 30, 0,05 e 100. Adicionalmente, uma comparação suplementar que teve como intenção determinar a melhor estratégia de distanciamento social a ser adotada durante as operações que promoviam a evolução da imunidade de rebanho, revelou que as alterações que consideravam indivíduos imunes, deveriam selecionar os novos genes de forma aleatória ao invés de, como havia sido inicialmente proposto, sempre fundamentar-se no indivíduo imune melhor avaliado em termos da função critério. Por conseguinte, o valor do fator $x_i^v(t)$, na expressão 11, passou a ser retratado por um gene selecionado aleatoriamente dentre os indivíduos imunes x^v, considerando o vetor de status S, de modo que $v = \{i | s^i = 2\}$.

3. Otimização inspirada na imunidade de rebanho aplicada ao agrupamento não hirárquico de objetos

Fundamentado no algoritmo referido na seção 2, 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:

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

Algoritmo 1: Otimização inspirada na imunidade de rebanho

```
1  Estabeleça f(x), npi, nmi, nip, n, tp e im;
2  Inicialize x_i^j = lb_i + (ub_i − lb_i)U(0,1),
     ∀i = 1,2,...,d e ∀j = 1,2,...,nip;
3  Calcule f(x^j) e atribua s^j = 0 e a^j = 0,
     ∀j = 1,2,...,nip;
4  enquanto t ≤ nmi faça
5      para j = 1 até nip faça
6          w(x^j(t+1)) = 0;
7          para i = 1 até d faça
8              se r < tp/3 então
9                  x_i^j(t+1) = C(x_i^j(t));
10                 w(x^j(t+1)) = 1;
11             senão se r < 2tp/3 então
12                 x_i^j(t+1) = N(x_i^j(t));
13             senão se r < tp então
14                 x_i^j(t+1) = R(x_i^j(t));
15             senão
16                 x_i^j(t+1) = x_i^j(t);
17         se f(x^j(t+1)) ≤ f(x^j(t)) então
18             x^j(t) = x^j(t+1);
19         senão
20             a^j = a^j + 1;
21         se f(x^j(t+1)) < f(x^j(t+1))/Δf(x) e
             s^j = 0 e w(x^j(t+1)) = 1 então
22             s^j = 1;
23             a^j = 1;
24         se f(x^j(t+1)) > f(x^j(t+1))/Δf(x) e
             s^j = 1 então
25             s^j = 2;
26             a^j = 0;
27         se a^j ≥ im e s^j = 1 então
28             x_i^j(t) = lb_i + (ub_i − lb_i)U(0,1),
                 ∀i = 1,2,...,d;
29             s^j = 0;
30             a^j = 0;
31     t = t + 1;
```

- Os indivíduos $x = (x_1, x_2, \ldots, x_{d'}) \in \mathbb{R}^{d'}$, $d' = dK$, da população foram retratados pelos protótipos ou médias dos K grupos. Consequentemente, os genes posicionados no intervalo $[1,d]$ referiam-se às coordenadas da média do primeiro grupo, os genes 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;

- A operação de inicialização dos nip elementos da população, consistia, para cada indivíduo, 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 dissimilaridade entre o objeto e o protótipo, determinada por meio da expressão 1, de forma que i-ésimo objeto, o_i, era relacionado ao j-ésimo grupo, de protótipo m_j, quando $d(o_i,m_j) \leq d(o_i,m_l)$, $\forall i = 1,2,\ldots,n$, $\forall j = 1,2,\ldots,K$ e $\forall l = 1,2,\ldots,K$, 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 expressão 4. 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 ou razão imunidade a minimizar, computada para cada indivíduo, foi, conforme o trabalho de [15], definida como sendo igual a uma variação do erro quadrático referido na equação 3 e estabelecida por intermédio da expressão

$$e^{2'}(O,G) = \frac{1}{K}\sum_{j=1}^{K}\frac{1}{n_j}\sum_{i=1}^{n_j} \parallel o_i^j - m_j \parallel^2 \qquad (12)$$

- A evolução da imunidade de rebanho foi efetuada por meio da modificação das coordenadas dos protótipos dos grupos, às quais correspondiam aos genes do indivíduo, mediante a aplicação das operações estabelecidas pela expressão 5, 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 indivíduos da população, uma operação de ajuste que, nestas circunstâncias, adicionava um valor constante excepcionalmente elevado ao resultado da função critério, foi acrescentado. Por meio dessa conduta, caracterizada pelas linhas 17 e 18 do algoritmo 2, a razão de imunidade $f(x^j(t+1))$ do indivíduo modificado $x_i^j(t+1)$ passava a ser demasiadamente inadequada. Desta forma, as soluções retratadas por itens nessa condição, convertiam-se em indivíduos irrecuperáveis que, por irem a óbito, eram substituídos por novas soluções, originadas aleatoriamente.

Os parâmetros de execução do algoritmo, à exceção do número máximo de iterações, nmi, foram estabelecidos conforme sugerido por [11], assim, o número preliminar de indivíduos infectados, npi, o número de indivíduos da população, nip, a taxa de propagação, tp, e a idade máxima dos indivíduos infectados, im, eram, nesta ordem, iguais a 1, 30, 0,05 e 100. De fato, ainda que o valor de nmi indicado por [11] tenha sido 100.000, uma avaliação complementar dos estudos de [16] e [17], que, de forma semelhante, também utilizaram meta-heurísticas fundamentadas no aproprimoramento de populações de soluções, à determinação do particionamento de coleções de objetos, permitiu inferir que um número máximo de iterações igual a 200 seria apropriado. Deste modo, nos ensaios conduzidos neste trabalho, este foi o valor admitido para nmi.

4. Experimentos computacionais

Com o propósito de avaliar o comportamento do procedimento de otimização inspirado na imunidade de rebanho, quando aplicado ao agrupamento não hierárquico de objetos, um conjunto de experimentos foi empreendido. Em particular, 33 coleções, constituídas por objetos de variadas dimensões e com números de grupos conhecidos, foram submetidas ao particionamento determinado pela meta-heurística inspirada na imunidade de rebanho, e, os resultados obtidos, foram subsequentemente comparados com os resultados estabelecidos pelas meta-heurísticas enxame de partículas, recozimento simulado, algoritmo genético e pelo método *K-means*, descritos, respectivamente, nos trabalhos de [18], [19], [20] e [21].

Todas as estratégias de particionamento, que tiveram como função critério a minimizar o erro quadrático modificado, retratado pela expressão 12, foram codificadas por meio da linguagem de programação Microsoft Visual Basic .NET, e, adicionalmente, 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 imunidade de rebanho, referidos na seção 3, os seguintes critérios foram assinalados para os demais algoritmos. Para o enxame de partículas e consoante ao estabelecido em [18, 16, 17, 11], o número máximo de iterações, o número de partículas, o coeficiente de inércia e os coeficientes de acelaração eram, nessa ordem, iguais a 200, 30, 0,72 e 1,49. Para o recozimento simulado e de acordo com o sugerido em [19], 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 algortimo genético e segundo a recomendações de [20, 16, 17, 11], o número máximo de gerações, o número de indivíduos da população, a probabilidade de recombinação e a probabilidade de mutação, eram, respectivamente, iguais a 200, 30, 0,8 e 0,001. Por fim, para o método *K-means* e de acordo com o observado em [18], o número máximo de iterações até que a execução fosse descontinuada foi estabelecido em 10.

Tendo em conta que, conforme [3, 2], a aplicação dos métodos de particionamento sobre um conjunto de itens sempre resultará na divisão dos objetos entre os grupos, mesmo

Algoritmo 2: Otimização inspirada na imunidade de rebanho aplicada ao agrupamento não hierárquico de objetos

Estabeleça $f(x)$, npi, nmi, nip, n, tp, im, O e K;
Inicialize x^j com K objetos selecionados aleatoriamente de O, $\forall j = 1,2,\ldots,nip$;
Calcule $f(x^j)$ e atribua $s^j = 0$ e $a^j = 0$, $\forall j = 1,2,\ldots,nip$;
enquanto $t \leq nmi$ **faça**
  para j = *1 até* nip **faça**
    $w(x^j(t+1)) = 0$;
    para i = *1 até* d' **faça**
      se $r < tp/3$ **então**
        $x_i^j(t+1) = C(x_i^j(t))$;
        $w(x^j(t+1)) = 1$;
      senão se $r < 2tp/3$ **então**
        $x_i^j(t+1) = N(x_i^j(t))$;
      senão se $r < tp$ **então**
        $x_i^j(t+1) = R(x_i^j(t))$;
      senão
        $x_i^j(t+1) = x_i^j(t)$;
    se *há grupos vazios na partição retratada pelo indivíduo* $x^j(t+1)$ **então**
      $f(x^j(t+1)) = f(x^j(t+1)) + \infty$;
    se $f(x^j(t+1)) \leq f(x^j(t))$ **então**
      $x^j(t) = x^j(t+1)$;
    senão
      $a^j = a^j + 1$;
    se $f(x^j(t+1)) < f(x^j(t+1))/\Delta f(x)$ *e* $s^j = 0$ *e* $w(x^j(t+1)) = 1$ **então**
      $s^j = 1$;
      $a^j = 1$;
    se $f(x^j(t+1)) > f(x^j(t+1))/\Delta f(x)$ *e* $s^j = 1$ **então**
      $s^j = 2$;
      $a^j = 0$;
    se $a^j \geq im$ *e* $s^j = 1$ **então**
      Inicialize $x^j(t)$ com K objetos selecionados aleatoriamente de O;
      $s^j = 0$;
      $a^j = 0$;
  $t = t + 1$;

que estes sejam inexistentes ou ainda de pouca relevância para o contexto do problema que se deseja determinar, este trabalho admitiu, além da função critério, o emprego de outras medidas estatísticas aptas a expressar a qualidade das partições obtidas. Em particular, índices de validação externos, passíveis de serem aplicados sobre coleções de objetos antecipadamente classificadas, além de índices de validação internos e relativos, capazes de serem empregados sobre coleções classificadas ou não, foram adotados.

O primeiro conjunto de objetos submetido aos experimentos, originado de [22, 23] e referido na tabela 1, era constituído por coleções de itens previamente classificados, desta forma, além da função critério, os índices de validação de agrupamentos entropia, pureza, silhueta, Rand, Jaccard, Fowlkes-Mallows, Γ statistic, Davies-Bouldin e Dunn, descritos em [1, 24, 25, 26, 27, 4, 28, 29], foram considerados na análise das partições. O segundo conjunto, retratado na tabela 2 e também originado de [22, 23], compreendia coleções de objetos que não estavam antecipadamente classificados, sendo assim, além da função critério, apenas os valores dos índices silhueta, Davies-Bouldin e Dunn foram calculados.

Table 1. Características das coleções classificadas.

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
Balance	625	4	3
CMC	1.473	9	3
Dermatology	366	34	6
WDBC	569	30	2

Para cada coleção de objetos, os métodos de particionamento foram executados dez vezes, e, os valores médios das medidas de avaliação, determinados ao final de cada processamento, foram comparados. Com o propósito de auxiliar a aferição dos resultados, as medidas que admitiam valores extrínsecos ao intervalo $[0,1]$ foram normalizadas e uma tabela de escores, que determinava 1 para o melhor resultado e também 1 para qualquer resultado distinto do melhor, desde que estivesse até 5% aquém do valor mais apropriado, foi elaborada para cada conjunto de experimentos. Os escores alcançados pelos métodos de particionamento, ao se analisar os valores das medidas de avaliação, foram somados, sendo o parecer mais favorável atribuído à estratégia que obtivesse a maior pontuação.

As tabelas 3 e 4 demonstram os escores obtidos por cada um dos métodos de agrupamento, quando empregados na organização dos objetos previamente classificados, enquanto que as tabelas 5 e 6 retratam os escores alcançados quando da aplicação dos métodos aos objetos que não estavam antecipadamente classificados. A análise dos valores das tabelas 3 e 4, permite observar que, para as coleções classificadas, as meta-heurísticas enxame de partículas, algoritmo genético,

Table 2. Características das coleções não classificadas.

Nome	Objetos	Dimensões	Grupos
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
Dim128	1.024	128	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
S4	5.000	2	15
Thyroid	215	5	2

recozimento simulado e imunidade de rebanho, obtiveram, respectivamente, 39, 84, 1 e 58 escores, enquanto que o método *K-means* alcançou 77 escores. De forma semelhante, a avaliação dos valores das tabelas 5 e 6 possibilita verificar que, para as coleções não classificadas, as meta-heurísticas enxame de partículas, algoritmo genético, recozimento simulado e imunidade de rebanho, alcançaram, respectivamente, 28, 63, 0 e 60 escores, ao passo que o algoritmo *K-means* obteve 37 escores. Os valores consolidados retratam, por fim, que as meta-heurísticas enxame de partículas, algoritmo genético, recozimento simulado e imunidade de rebanho, auferiram, nesta ordem, 67, 147, 1 e 118 escores, enquanto que o método *K-means* atingiu 114 escores, sugerindo, portanto, que o algoritmo genético seria a estratégia mais apropriada a realizar o particionamento das coleções de objetos admitidas neste trabalho. Os resultados unificados expressam ainda que a meta-heurística inspirada na imunidade de rebanho constituiu o segundo melhor método, acompanhado, nesta ordem, do *K-means*, do exame de partículas e do recozimento simulado.

Uma avaliação modificada, tendo em consideração somente a função critério, possibilita observar que as meta-heurísticas enxame de partículas, algoritmo genético, recozimento simulado e imunidade de rebanho, além do algoritmo *K-means*, antigiram, nesta ordem, 23, 22, 0, 30 e 11 escores, revelando, por meio deste aspecto, uma melhor adequação da estratégia inspirada na imunidade de rebanho ao problema. Uma análise semelhante, admitindo de forma particular os escores dos índices de validação de agrupamentos, indica que o algortimo *K-means* foi o mais apropriado de acordo com as medidas entropia, Rand, e Γ statistic, que o algoritmo genético obteve melhores resultados para os índices pureza, silhueta, Fowlkes-Mallows, Davies-Bouldin e Dunn, que de acordo com os escores da medida Jaccard, o algoritmo genético e a meta-heurística inspirada na imunidade de rebanho tiveram resultados idênticos e mais acertados, e denota que, exceto pela função critério e pelo índice de validação Jaccard, o algoritmo inspirado na imunidade de rebanho não obteve um parecer mais favorável para nenhuma das outras medidas de análise. Com efeito, a imunidade de rebanho foi a segunda melhor estratégia ao se considerar os critérios silhueta, Fowlkes-Mallows, Davies-Bouldin e Dunn, mas apenas o terceiro melhor método de particionamento, sob a perspectiva dos índices entropia, pureza, Rand e Γ statistic.

Uma análise adicional, que teve como propósito empregar um teste estatístico não paramétrico, sobre os resultados que foram obtidos nos experimentos, conduziu aos procedimentos descritos no trabalho de [30]. Em particular, o teste do sinal, indicado para avaliar diferenças consistentes entre pares de observações, foi utilizado para confrontar o algoritmo inspirado na imunidade de rebanho, que nesta circunstância foi considerado o método de referência, com as demais estratégias de particionamento. Com efeito e conforme [30], uma forma elementar de comparar o desempenho de algoritmos consiste em estabelecer um indicador, e, a partir deste, verificar qual procedimento obtém os resultados mais apropriados. Desta forma, o valor da função critério computado por cada algoritmo pode, por exemplo, ser aferido, sendo considerado o "vencedor", o método que tiver alcançado a resposta mais adequada. Por intermédio dessa conduta, se há n comparações a realizar entre dois algoritmos quaisquer, o primeiro algoritmo será considerado significativamente melhor do que o segundo, com $p < 0,05$, se obtiver pelo menos $n/2 + 1,96\sqrt{n}/2$ vitórias sobre o segundo.

Considerando que, no primeiro conjunto de ensaios e para cada método de agrupamento, 10 medidas distintas foram computadas, que 13 coleções de objetos previamente classificadas foram submetidas ao particionamento, e que para cada coleção, os algortimos foram executados 10 vezes, tem-se, por medida, 130 comparações possíveis entre o algoritmo inspirado na imunidade de rebanho e cada um dos demais métodos. Assim e tendo em conta a expressão $n/2+1,96\sqrt{n}/2$, verifica-se que o método inspirado na imunidade de rebanho será qualificado como o mais apropriado em relação ao algoritmo com o qual estiver sendo confrontado, se alcançar pelo menos 77 vitórias. De forma semelhante, tem-se, para o segundo conjunto de experimentos, que o algortimo inspirado na imunidade de rebanho será interpretado com o mais adequado, ao obter, por medida, pelo menos 114 vitórias, haja vista que nesta avaliação 4 indicadores foram utilizados, 20 coleções não classificadas foram segmentadas, e que, para cada coleção, os métodos foram processados 10 vezes.

Table 3. Avaliação dos métodos de agrupamento considerando as coleções classificadas e as medidas função critério, entropia, pureza, silhueta e Rand. EP: Exame de partículas; AG: Algoritmo genético; KM: *K-means*; RS: Recozimento simulado; IR: Imunidade de rebanho; EEP: Escores do exame de partículas; EAG: Escores do algoritmo genético; EKM: Escores do *K-means*; ERS: Escores do recozimento simulado; EIR: Escores da imunidade de rebanho.

Medida	Coleção	EP	AG	KM	RS	IR	EEP	EAG	EKM	ERS	EIR
Função Critério	Aggregation	**0,9940**	**0,9747**	**0,9724**	0,0000	**1,0000**	1	1	1	0	1
	Compound	**0,9912**	**0,9701**	0,9491	0,0000	**1,0000**	1	1	0	0	1
	Pathbased	**0,9562**	0,8481	0,8396	0,0000	**1,0000**	1	0	0	0	1
	Spiral	0,9115	0,8987	0,8957	0,0000	**1,0000**	0	0	0	0	1
	D31	**0,9723**	**0,9964**	**0,9934**	0,0000	**1,0000**	1	1	1	0	1
	R15	**0,9732**	**0,9904**	**0,9676**	0,0000	**1,0000**	1	1	1	0	1
	Jain	0,8957	0,8275	0,8221	0,0000	**1,0000**	0	0	0	0	1
	Flame	0,8235	0,7660	0,7598	0,0000	**1,0000**	0	0	0	0	1
	Iris	**0,9866**	**0,9516**	0,9436	0,0000	**1,0000**	1	1	0	0	1
	Balance	**1,0000**	0,6532	0,6489	0,0000	0,9320	1	0	0	0	0
	CMC	**0,9759**	**0,9825**	**0,9817**	0,0000	**1,0000**	1	1	1	0	1
	Dermatology	0,8974	**0,9548**	0,9441	0,0000	**1,0000**	0	1	0	0	1
	WDBC	0,9207	0,6738	0,6921	0,0000	**1,0000**	0	0	0	0	1
Entropia	Aggregation	0,9410	**1,0000**	**0,9882**	0,0000	0,9343	0	1	1	0	0
	Compound	**0,9677**	**0,9934**	**0,9713**	0,0000	**1,0000**	1	1	1	0	1
	Pathbased	0,7419	**0,9979**	**1,0000**	0,0000	0,6991	0	1	1	0	0
	Spiral	0,0497	0,0000	0,0151	0,1210	**1,0000**	0	0	0	0	1
	D31	0,8956	**0,9738**	**0,9616**	0,0000	**1,0000**	0	1	1	0	1
	R15	0,8600	0,8971	0,8848	0,0000	**1,0000**	0	0	0	0	1
	Jain	0,5932	**1,0000**	**0,9961**	0,0000	0,2495	0	1	1	0	0
	Flame	0,9341	**1,0000**	**0,9860**	0,0000	0,5308	0	1	1	0	0
	Iris	0,8144	**1,0000**	0,9440	0,0000	0,7769	0	1	0	0	0
	Balance	0,6039	**1,0000**	**0,9678**	0,0000	0,6273	0	1	1	0	0
	CMC	0,8492	**1,0000**	**0,9934**	0,0000	0,4656	0	1	1	0	0
	Dermatology	0,8978	0,6017	**1,0000**	0,0000	0,7097	0	0	1	0	0
	WDBC	0,1922	0,9457	**1,0000**	0,0000	0,0834	0	0	1	0	0
Pureza	Aggregation	**0,9853**	**1,0000**	**0,9905**	0,0000	**0,9936**	1	1	1	0	1
	Compound	0,9400	**0,9658**	0,9393	0,0000	**1,0000**	0	1	0	0	1
	Pathbased	0,6545	**0,9964**	**1,0000**	0,0000	0,6296	0	1	1	0	0
	Spiral	0,3962	0,0000	0,1474	0,5135	**1,0000**	0	0	0	0	1
	D31	0,7873	**0,9533**	0,9230	0,0000	**1,0000**	0	1	0	0	1
	R15	0,7740	0,8379	0,8115	0,0000	**1,0000**	0	0	0	0	1
	Jain	0,0000	**0,9510**	**1,0000**	0,0023	0,0000	0	1	1	0	0
	Flame	0,8448	**1,0000**	**0,9624**	0,0000	0,5428	0	1	1	0	0
	Iris	0,6171	**1,0000**	0,8980	0,0000	0,5197	0	1	0	0	0
	Balance	0,6020	**0,9592**	**1,0000**	0,0000	0,6830	0	1	1	0	0
	CMC	0,9403	0,8806	**1,0000**	0,0000	0,3134	0	0	1	0	0
	Dermatology	0,8871	0,5264	**1,0000**	0,0000	0,6672	0	0	1	0	0
	WDBC	0,0017	**0,9590**	**1,0000**	0,0000	0,0017	0	1	1	0	0
Silhueta	Aggregation	0,9337	**1,0000**	**0,9910**	0,0000	**0,9744**	0	1	1	0	1
	Compound	**0,9643**	0,9417	0,9370	0,0000	**1,0000**	1	0	0	0	1
	Pathbased	0,7067	**1,0000**	**0,9989**	0,0000	0,7262	0	1	1	0	0
	Spiral	**0,9958**	**0,9992**	**1,0000**	0,0000	0,5670	1	1	1	0	0
	D31	0,7802	**0,9603**	0,9382	0,0000	**1,0000**	0	1	0	0	1
	R15	0,7406	0,8775	0,8126	0,0000	**1,0000**	0	0	0	0	1
	Jain	0,9385	**1,0000**	**0,9993**	0,0000	0,7117	0	1	1	0	0
	Flame	0,9025	**1,0000**	**0,9997**	0,0000	0,6605	0	1	1	0	0
	Iris	**0,9577**	**1,0000**	**0,9816**	0,0000	0,9433	1	1	1	0	0
	Balance	0,5776	**1,0000**	**0,9919**	0,0000	0,5614	0	1	1	0	0
	CMC	0,9395	**1,0000**	**0,9891**	0,0000	0,7372	0	1	1	0	0
	Dermatology	0,9332	**0,9798**	**0,9687**	0,0000	**1,0000**	0	1	1	0	1
	WDBC	0,1066	**1,0000**	**0,9917**	0,0000	0,0575	0	1	1	0	0
Rand	Aggregation	**0,9629**	**0,9953**	**0,9817**	0,0000	**1,0000**	1	1	1	0	1
	Compound	0,9333	0,9351	**1,0000**	0,0000	0,9110	0	0	1	0	0
	Pathbased	0,5947	**0,9972**	**1,0000**	0,0000	0,5650	0	1	1	0	0
	Spiral	0,9430	**0,9709**	**0,9709**	**1,0000**	0,0000	0	1	1	1	0
	D31	0,6636	0,9238	0,8885	0,0000	**1,0000**	0	0	0	0	1
	R15	0,6469	0,7294	0,7371	0,0000	**1,0000**	0	0	0	0	1
	Jain	0,1363	**0,9820**	**1,0000**	0,0000	0,0100	0	1	1	0	0
	Flame	0,8362	**1,0000**	**0,9558**	0,0000	0,5633	0	1	1	0	0
	Iris	0,6172	**1,0000**	0,8945	0,0000	0,5522	0	1	0	0	0
	Balance	0,0322	**1,0000**	**0,9936**	0,0000	0,4702	0	1	1	0	0
	CMC	0,9351	**0,9558**	**1,0000**	0,7481	0,0000	0	1	1	0	0
	Dermatology	0,0000	0,8882	**1,0000**	0,9430	0,3925	0	0	1	0	0
	WDBC	0,0000	0,9465	**1,0000**	0,0558	0,0046	0	0	1	0	0
Total							**14**	**44**	**42**	**1**	**29**

Table 4. Avaliação dos métodos de agrupamento considerando as coleções classificadas e as medidas Jaccard, Fowlkes-Mallows, Γ statistic, Davies-Bouldin e Dunn . EP: Exame de partículas; AG: Algoritmo genético; KM: *K-means*; RS: Recozimento simulado; IR: Imunidade de rebanho; EEP: Escores do exame de partículas; EAG: Escores do algoritmo genético; EKM: Escores do *K-means*; ERS: Escores do recozimento simulado; EIR: Escores da imunidade de rebanho.

Medida	Coleção	EP	AG	KM	RS	IR	EEP	EAG	EKM	ERS	EIR
Jaccard	Aggregation	0,9387	0,9250	0,9093	0,0000	**1,0000**	0	0	0	0	1
	Compound	**0,9948**	0,8804	**1,0000**	0,0000	0,9146	1	0	1	0	0
	Pathbased	0,8330	**0,9982**	**1,0000**	0,0000	0,8023	0	1	1	0	0
	Spiral	0,0627	0,0000	0,0088	0,0389	**1,0000**	0	0	0	0	1
	D31	0,6382	0,8890	0,8521	0,0000	**1,0000**	0	0	0	0	1
	R15	0,5950	0,7217	0,6583	0,0000	**1,0000**	0	0	0	0	1
	Jain	0,2125	**0,9776**	**1,0000**	0,0000	0,4243	0	1	1	0	0
	Flame	0,9025	**1,0000**	0,9464	0,0000	0,8699	0	1	0	0	0
	Iris	0,6890	**1,0000**	0,9148	0,0000	0,6492	0	1	0	0	0
	Balance	**1,0000**	0,4996	0,4895	0,0000	0,9014	1	0	0	0	0
	CMC	0,4218	0,4972	0,4637	0,0000	**1,0000**	0	0	0	0	1
	Dermatology	**1,0000**	0,3717	0,6404	0,0000	0,7758	1	0	0	0	0
	WDBC	0,2208	**0,9527**	**1,0000**	0,0000	0,3038	0	1	1	0	0
Fowlkes-Mallows	Aggregation	0,9495	**0,9571**	0,9452	0,0000	**1,0000**	0	1	0	0	1
	Compound	**0,9935**	0,9198	**1,0000**	0,0000	0,9393	1	0	1	0	0
	Pathbased	0,8863	**0,9987**	**1,0000**	0,0000	0,8587	0	1	1	0	0
	Spiral	0,0562	0,0000	0,0072	0,0378	**1,0000**	0	0	0	0	1
	D31	0,7686	0,9339	0,9103	0,0000	**1,0000**	0	0	0	0	1
	R15	0,7457	0,8181	0,7880	0,0000	**1,0000**	0	0	0	0	1
	Jain	0,2158	**0,9807**	**1,0000**	0,0000	0,4315	0	1	1	0	0
	Flame	0,9128	**1,0000**	**0,9542**	0,0000	0,9088	0	1	1	0	0
	Iris	0,7689	**1,0000**	0,9340	0,0000	0,7484	0	1	0	0	0
	Balance	**1,0000**	0,4988	0,4901	0,0000	0,8399	1	0	0	0	0
	CMC	0,4120	0,4820	0,4500	0,0000	**1,0000**	0	0	0	0	1
	Dermatology	**1,0000**	0,3601	0,6083	0,0000	0,7494	1	0	0	0	0
	WDBC	0,2836	**0,9640**	**1,0000**	0,0000	0,3979	0	1	1	0	0
Γ Statistic	Aggregation	**0,9529**	**0,9697**	**0,9572**	0,0000	**1,0000**	1	1	1	0	1
	Compound	**0,9736**	0,9264	**1,0000**	0,0000	0,9285	1	0	1	0	0
	Pathbased	0,7772	**0,9982**	**1,0000**	0,0000	0,7502	0	1	1	0	0
	Spiral	0,1202	0,0000	0,0343	0,2704	**1,0000**	0	0	0	0	1
	D31	0,7667	0,9336	0,9097	0,0000	**1,0000**	0	0	0	0	1
	R15	0,7424	0,8151	0,7862	0,0000	**1,0000**	0	0	0	0	1
	Jain	0,2889	**0,9869**	**1,0000**	0,2221	0,0000	0	1	1	0	0
	Flame	0,8265	**1,0000**	**0,9559**	0,0000	0,5150	0	1	1	0	0
	Iris	0,7189	**1,0000**	0,9198	0,0000	0,6892	0	1	0	0	0
	Balance	0,6189	**1,0000**	**0,9911**	0,0000	0,8009	0	1	1	0	0
	CMC	0,8442	**0,9638**	**1,0000**	0,0000	0,0906	0	1	1	0	0
	Dermatology	**0,9918**	0,5237	**1,0000**	0,0000	0,7897	1	0	1	0	0
	WDBC	0,0259	**0,9530**	**1,0000**	0,1299	0,0000	0	1	1	0	0
Davies-Bouldin	Aggregation	**0,9991**	**1,0000**	**1,0000**	0,0000	**0,9997**	1	1	1	0	1
	Compound	**0,9930**	**0,9943**	**0,9931**	0,0000	**1,0000**	1	1	1	0	1
	Pathbased	**0,9638**	**1,0000**	**0,9975**	0,0000	**0,9913**	1	1	1	0	1
	Spiral	**0,9921**	**1,0000**	**0,9984**	0,0000	**0,9906**	1	1	1	0	1
	D31	**0,9963**	**0,9995**	**0,9992**	0,0000	**1,0000**	1	1	1	0	1
	R15	**0,9873**	**0,9958**	**0,9918**	0,0000	**1,0000**	1	1	1	0	1
	Jain	**0,9848**	**0,9850**	**0,9845**	0,0000	**1,0000**	1	1	1	0	1
	Flame	**0,9892**	**0,9869**	**0,9865**	0,0000	**1,0000**	1	1	1	0	1
	Iris	**0,9836**	**1,0000**	**0,9932**	0,0000	**0,9961**	1	1	1	0	1
	Balance	**0,9865**	**1,0000**	**0,9996**	0,0000	**0,9972**	1	1	1	0	1
	CMC	**0,9987**	**1,0000**	**0,9998**	0,0000	**0,9979**	1	1	1	0	1
	Dermatology	**0,9841**	**0,9954**	**0,9882**	0,0000	**1,0000**	1	1	1	0	1
	WDBC	0,3471	**1,0000**	**0,9786**	0,0000	0,8618	0	1	1	0	0
Dunn	Aggregation	0,7556	**1,0000**	0,9407	0,0000	0,7852	0	1	0	0	0
	Compound	0,7353	**1,0000**	**0,9647**	0,0000	0,7471	0	1	1	0	0
	Pathbased	**0,9883**	**0,9854**	**1,0000**	0,0000	0,6696	1	1	1	0	0
	Spiral	0,6897	0,4483	0,7155	0,0000	**1,0000**	0	0	0	0	1
	D31	0,4207	0,7724	0,7310	0,0000	**1,0000**	0	0	0	0	1
	R15	0,1354	0,5184	0,1662	0,0000	**1,0000**	0	0	0	0	1
	Jain	**1,0000**	0,6425	0,8792	0,0000	0,4444	1	0	0	0	0
	Flame	**1,0000**	0,6961	0,8137	0,0000	0,6765	1	0	0	0	0
	Iris	0,3925	**1,0000**	0,8305	0,0000	0,4793	0	1	0	0	0
	Balance	0,3464	**1,0000**	**1,0000**	0,0000	0,2793	0	1	1	0	0
	CMC	**0,9594**	**0,9926**	**1,0000**	0,0000	0,7768	1	1	1	0	0
	Dermatology	0,5017	**1,0000**	0,7686	0,0000	0,6075	0	1	0	0	0
	WDBC	0,1768	**1,0000**	0,8110	0,0000	0,1341	0	1	0	0	0
Total							**25**	**40**	**35**	**0**	**29**

Table 5. Avaliação dos métodos de agrupamento considerando as coleções não classificadas e as medidas Função Critério e Silhueta. EP: Exame de partículas; AG: Algoritmo genético; KM: *K-means*; RS: Recozimento simulado; IR: Imunidade de rebanho; EEP: Escores do exame de partículas; EAG: Escores do algoritmo genético; EKM: Escores do *K-means*; ERS: Escores do recozimento simulado; EIR: Escores da imunidade de rebanho.

Medida	Coleção	EP	AG	KM	RS	IR	EEP	EAG	EKM	ERS	EIR
Função Critério	Wine	0,7023	0,5563	0,5036	0,0000	**1,0000**	0	0	0	0	1
	Yeast	**0,9817**	0,9151	0,8570	0,0000	**1,0000**	1	0	0	0	1
	Breast	**1,0000**	0,7094	0,6929	0,0000	0,7924	1	0	0	0	0
	Glass	**1,0000**	0,7908	0,5108	0,0000	**0,9760**	1	0	0	0	1
	Dim032	0,9246	**1,0000**	0,9086	0,0000	**0,9992**	0	1	0	0	1
	Dim064	0,9140	**1,0000**	0,9058	0,0000	**0,9816**	0	1	0	0	1
	Dim128	0,9189	**0,9935**	0,8978	0,0000	**1,0000**	0	1	0	0	1
	A1	**0,9774**	**0,9956**	**0,9905**	0,0000	**1,0000**	1	1	1	0	1
	A2	**0,9717**	**0,9969**	**0,9918**	0,0000	**1,0000**	1	1	1	0	1
	A3	**0,9713**	**0,9987**	**0,9967**	0,0000	**1,0000**	1	1	1	0	1
	Dim2	**0,9609**	**1,0000**	0,9226	0,0000	**0,9850**	1	1	0	0	1
	Dim3	**0,9642**	**1,0000**	0,9029	0,0000	**0,9910**	1	1	0	0	1
	Dim4	**0,9706**	**1,0000**	0,9184	0,0000	**0,9987**	1	1	0	0	1
	Dim5	**0,9597**	**1,0000**	0,9099	0,0000	**0,9956**	1	1	0	0	1
	Dim6	0,9408	**0,9882**	0,8962	0,0000	**1,0000**	0	1	0	0	1
	S1	**0,9579**	**0,9902**	**0,9753**	0,0000	**1,0000**	1	1	1	0	1
	S2	**0,9623**	**0,9864**	**0,9729**	0,0000	**1,0000**	1	1	1	0	1
	S3	**0,9705**	**0,9901**	**0,9855**	0,0000	**1,0000**	1	1	1	0	1
	S4	**0,9734**	**0,9954**	**0,9821**	0,0000	**1,0000**	1	1	1	0	1
	Thyroid	**1,0000**	0,3677	0,0000	0,7904	0,5840	1	0	0	0	0
Silhueta	Wine	0,7440	**1,0000**	**0,9781**	0,0000	0,7310	0	1	1	0	0
	Yeast	0,4373	**1,0000**	**0,9832**	0,0000	**0,9531**	0	1	1	0	1
	Breast	0,0000	**1,0000**	**0,9966**	0,2321	0,3737	0	1	1	0	0
	Glass	0,3000	**1,0000**	0,7610	0,0000	0,5288	0	1	0	0	0
	Dim032	0,7079	0,9317	0,7366	0,0000	**1,0000**	0	0	0	0	1
	Dim064	0,7433	**0,9666**	0,7891	0,0000	**1,0000**	0	1	0	0	1
	Dim128	0,7720	0,8507	0,7365	0,0000	**1,0000**	0	0	0	0	1
	A1	0,8199	**0,9690**	0,9281	0,0000	**1,0000**	0	1	0	0	1
	A2	0,8058	**0,9654**	0,9268	0,0000	**1,0000**	0	1	0	0	1
	A3	0,7644	**0,9873**	**0,9531**	0,0000	**1,0000**	0	1	1	0	1
	Dim2	0,8351	**1,0000**	0,8333	0,0000	0,9379	0	1	0	0	0
	Dim3	0,8778	**1,0000**	0,7404	0,0000	**0,9537**	0	1	0	0	1
	Dim4	0,8598	**1,0000**	0,7889	0,0000	**0,9992**	0	1	0	0	1
	Dim5	0,8408	**1,0000**	0,8064	0,0000	**0,9693**	0	1	0	0	1
	Dim6	0,7729	**0,9720**	0,7116	0,0000	**1,0000**	0	1	0	0	1
	S1	0,8129	**0,9580**	0,9004	0,0000	**1,0000**	0	1	0	0	1
	S2	0,8221	0,9429	0,8866	0,0000	**1,0000**	0	0	0	0	1
	S3	0,8256	**0,9777**	**0,9536**	0,0000	**1,0000**	0	1	1	0	1
	S4	0,7931	**0,9901**	**0,9648**	0,0000	**1,0000**	0	1	1	0	1
	Thyroid	0,1596	**1,0000**	**0,9988**	0,0000	0,5753	0	1	1	0	0
Total							**15**	**32**	**14**	**0**	**33**

Table 6. Avaliação dos métodos de agrupamento considerando as coleções não classificadas e as medidas Davies-Bouldin e Dunn. EP: Exame de partículas; AG: Algoritmo genético; KM: *K-means*; RS: Recozimento simulado; IR: Imunidade de rebanho; EEP: Escores do exame de partículas; EAG: Escores do algoritmo genético; EKM: Escores do *K-means*; ERS: Escores do recozimento simulado; EIR: Escores da imunidade de rebanho.

Medida	Coleção	EP	AG	KM	RS	IR	EEP	EAG	EKM	ERS	EIR
Davies-Bouldin	Wine	0,9450	**0,9587**	0,9439	0,0000	**1,0000**	0	1	0	0	1
	Yeast	**0,9717**	**1,0000**	**0,9978**	0,0000	**0,9965**	1	1	1	0	1
	Breast	0,6988	**1,0000**	**0,9936**	0,0000	**0,9626**	0	1	1	0	1
	Glass	0,9092	**1,0000**	**0,9699**	0,0000	**0,9608**	0	1	1	0	1
	Dim032	0,9334	**0,9873**	**0,9556**	0,0000	**1,0000**	0	1	1	0	1
	Dim064	0,8845	**0,9962**	**0,9703**	0,0000	**1,0000**	0	1	1	0	1
	Dim128	0,8183	**0,9692**	**0,9628**	0,0000	**1,0000**	0	1	1	0	1
	A1	**0,9936**	**0,9992**	**0,9979**	0,0000	**1,0000**	1	1	1	0	1
	A2	**0,9950**	**0,9994**	**0,9988**	0,0000	**1,0000**	1	1	1	0	1
	A3	**0,9952**	**0,9999**	**0,9995**	0,0000	**1,0000**	1	1	1	0	1
	Dim2	**0,9758**	**1,0000**	**0,9676**	0,0000	**0,9927**	1	1	1	0	1
	Dim3	**0,9837**	**1,0000**	**0,9674**	0,0000	**0,9955**	1	1	1	0	1
	Dim4	**0,9834**	**0,9984**	**0,9696**	0,0000	**1,0000**	1	1	1	0	1
	Dim5	**0,9828**	**1,0000**	**0,9675**	0,0000	**0,9957**	1	1	1	0	1
	Dim6	**0,9772**	**0,9962**	**0,9628**	0,0000	**1,0000**	1	1	1	0	1
	S1	**0,9914**	**0,9986**	**0,9964**	0,0000	**1,0000**	1	1	1	0	1
	S2	**0,9923**	**0,9979**	**0,9966**	0,0000	**1,0000**	1	1	1	0	1
	S3	**0,9941**	**0,9995**	**0,9989**	0,0000	**1,0000**	1	1	1	0	1
	S4	**0,9918**	**0,9997**	**0,9988**	0,0000	**1,0000**	1	1	1	0	1
	Thyroid	0,5687	**1,0000**	0,9435	0,0000	0,7562	0	1	0	0	0
Dunn	Wine	0,9308	0,7454	0,4684	0,0000	**1,0000**	0	0	0	0	1
	Yeast	0,7440	0,9280	**1,0000**	0,0000	0,9120	0	0	1	0	0
	Breast	0,3264	**1,0000**	**0,9725**	0,0000	0,3663	0	1	1	0	0
	Glass	0,3205	**1,0000**	0,7233	0,0000	0,3397	0	1	0	0	0
	Dim032	0,2021	0,8898	0,5328	0,0000	**1,0000**	0	0	0	0	1
	Dim064	0,0169	**1,0000**	0,0368	0,0000	0,0499	0	1	0	0	0
	Dim128	0,3128	0,8341	0,7346	0,0000	1,0000	0	0	0	0	1
	A1	0,4138	**1,0000**	0,9397	0,0000	0,8966	0	1	0	0	0
	A2	0,4286	**1,0000**	0,8265	0,0000	0,8163	0	1	0	0	0
	A3	0,3500	**0,9875**	**1,0000**	0,0000	0,8375	0	1	1	0	0
	Dim2	0,1156	**1,0000**	0,1161	0,0000	0,8346	0	1	0	0	0
	Dim3	0,1280	**1,0000**	0,0044	0,0000	0,4623	0	1	0	0	0
	Dim4	0,0109	0,9395	0,0067	0,0000	**1,0000**	0	0	0	0	1
	Dim5	0,0208	**1,0000**	0,1255	0,0000	0,7852	0	1	0	0	0
	Dim6	0,0121	**0,9616**	0,0093	0,0000	**1,0000**	0	1	0	0	1
	S1	0,2378	0,5676	0,4919	0,0000	**1,0000**	0	0	0	0	1
	S2	0,3455	0,5636	0,6364	0,0000	**1,0000**	0	0	0	0	1
	S3	0,4478	0,8209	**1,0000**	0,0000	0,8358	0	0	1	0	0
	S4	0,3692	**1,0000**	0,8154	0,0000	**0,9692**	0	1	0	0	1
	Thyroid	0,3811	0,7268	**1,0000**	0,0000	0,5683	0	0	1	0	0
Total							**13**	**31**	**23**	**0**	**27**

Table 7. Avaliação dos métodos de agrupamento considerando o teste do sinal, as coleções classificadas e as medidas função critério, entropia, pureza, silhueta, Rand, Jaccard, Fowlkes-Mallows, Γ statistic, Davies-Bouldin e Dunn. VEP: Vitórias da imunidade de rebanho diante do exame de partículas; DEP: Derrotas da imunidade de rebanho diante do exame de partículas; VAG: Vitórias da imunidade de rebanho diante do algoritmo genético; DAG: Derrotas da imunidade de rebanho diante do algoritmo genético; VKM: Vitórias da imunidade de rebanho diante do *K-means*; DKM: Derrotas da imunidade de rebanho diante do *K-means*; VRS: Vitórias da imunidade de rebanho diante do recozimento simulado; DRS: Derrotas da imunidade de rebanho diante do recozimento simulado.

Medida	VEP	DEP	VAG	DAG	VKM	DKM	VRS	DRS
Função Critério	**114**	16	**128**	2	**130**	0	**130**	0
Entropia	59	71	44	86	45	85	**126**	4
Pureza	**84**	46	49	81	50	80	**117**	13
Silhueta	53	77	33	97	40	90	**128**	2
Rand	62	68	32	98	33	97	**78**	52
Jaccard	**87**	43	74	56	74	56	**130**	0
Fowlkes-Mallows	**89**	41	75	55	71	59	**130**	0
Γ Statistic	61	69	56	74	51	79	**102**	28
Davies-Bouldin	**103**	27	64	66	71	59	**130**	0
Dunn	65	65	38	92	47	83	**130**	0

Table 8. Avaliação dos métodos de agrupamento considerando o teste do sinal, as coleções não classificadas e as medidas Função Critério, Silhueta, Davies-Bouldin e Dunn. VEP: Vitórias da imunidade de rebanho diante do exame de partículas; DEP: Derrotas da imunidade de rebanho diante do exame de partículas; VAG: Vitórias da imunidade de rebanho diante do algoritmo genético; DAG: Derrotas da imunidade de rebanho diante do algoritmo genético; VKM: Vitórias da imunidade de rebanho diante do *K-means*; DKM: Derrotas da imunidade de rebanho diante do *K-means*; VRS: Vitórias da imunidade de rebanho diante do recozimento simulado; DRS: Derrotas da imunidade de rebanho diante do recozimento simulado.

Medida	VEP	DEP	VAG	DAG	VKM	DKM	VRS	DRS
Função Critério	**168**	32	**152**	48	**196**	4	**191**	9
Silhueta	**184**	16	108	92	**151**	49	**196**	4
Davies-Bouldin	**185**	15	**115**	85	**163**	37	**200**	0
Dunn	**169**	31	105	95	**123**	77	**200**	0

A tabela 7, que retrata por medida e para as coleções de objetos antecipadamente classificadas, o número de vitórias e derrotas obtidas pelo método inspirado na imunidade de rebanho, ao ser comparado individualmente com cada uma das demais estratégias de particionamento, permite observar que o algoritmo sugerido foi, de acordo com a função critério, substancialmente melhor do que os demais métodos, por ter obtido mais do que 77 vitórias em todas as comparações. De maneira análoga, verifica-se que a estratégia inspirada na imunidade de rebanho obteve respostas significativamente mais apropriadas do que o recozimento simulado em relação às medidas entropia, silhueta, Rand, Γ statistic e Dunn, mas que segundo os mesmos indicadores, não alcançou respostas substancialmente mais adequadas quando comparado ao exame de partírculas, ao algoritmo genético e ao *K-means*. A análise dos dados da tabela 7, possibilita verificar adicionalmente, que o algoritmo sugerido foi particularmente melhor do que o recozimento simulado e do que o enxame de partículas, segundo as medidas pureza, Jaccard, Fowlkes-Mallows e Davies-Bouldin, ainda que consoante os mesmos parâmetros, não tenha sido substancialmente melhor do que o algoritmo genético e do que o *K-means*.

A avaliação da tabela 8, que demostra por medida e para as coleções que não estavam antecipadamente classificadas, o número de vitórias e derrotas do algoritmo inspirado na imunidade de rebanho, possibilita distinguir que a estratégia sugerida foi, segundo a função critério e o índice de validação Davies-Bouldin, consideravelmente mais apropriada do que as demais, haja vista a obtenção de mais do que 114 vitórias em toda as confrontações. As informações da tabela 8, permitem observar também, que o método inspirado na imunidade de rebanho alcançou resultados consideravelmente melhores do que os obtidos pelo exame partículas, pelo *K-means* e pelo recozimento simulado, ao se considerar as medidas silhueta e Dunn, e que, conforme os mesmos parâmetros, a estratégia proposta não foi significativamente melhor do que o algoritmo genético.

A análise consolidada das tabelas que retrataram os escores obtidos por cada método, e das tabelas que apresentaram as vitórias e as derrotas dos algoritmos, permite observar uma predisposição do método inspirado na imunidade de rebanho em obter as melhores respostas sob a perspectiva da função critério, e em não alcançar os melhores resultados ao se incluir as demais medidas de avaliação. Desta forma e considerando que conforme [2, 4, 5], os índices de validação de agrupamentos, tais como os que foram utilizados neste estudo, têm como propósito fundamental indicar a consistência e a relevância das partições obtidas, sugere-se que a inferior adequação dos resultados do algoritmo inspirado na imunidade de rebanho, com relação a esses parâmetros, foi consequência da obtenção de arranjos que não eram tão congruentes aos dados, ainda que o valor da função critério tenha sido o mais apropriado.

5. Conclusões

Este trabalho retratou a aplicação de uma meta-heurística inspirada na aquisição da imunidade de rebanho por uma população de indivíduos, à determinação do agrupamento não hierárquico de objetos. De modo específico, o particionamento de uma coleção finita de itens em um conjunto de grupos, que foi abordado como um problema de otimização, que tinha como propósito minimizar o valor da função critério resultante da classificação obtida, foi realizado, pelo algoritmo sugerido e por outros quatro métodos descritos na literatura, para 33 coleções de referência, com número de grupos conhecido.

Os resultados alcançados por cada algoritmo, que foram analisados por intermédio de 10 medidas de avaliação distintas, indicaram que o método inspirado na imunidade de rebanho, foi o mais apropriado sob a perspectiva da função critério, haja vista que segundo esse aspecto, o algoritmo sugerido foi invariavelmente melhor do que todos os demais. Os resultados dos ensaios assinalaram além disso que, ao se considerar os critérios indicativos da adequação das classificações obtidas em relação à estrutura dos dados em seu formato original, o algoritmo inspirado na imunidade de rebanho não foi sistematicamente mais congruente do que as outras estratégias de agrupamento, ainda que em algumas circunstâncias, tenha obtido respostas equivalentes ou até mesmo mais acertadas.

Contribuição do autor

Alfredo Silveira Araújo Neto: Concepção e elaboração da pesquisa; Análise e interpretação dos resultados; Redação do manuscrito.

Referências

[1] TAN, P. N.; STEINBACH, M.; KUMAR, V. *Introduction to Data Mining*. Boston: Pearson Education, Inc., 2006.

[2] XU, R.; WUNSCH, D. C. *Clustering*. Piscataway, New Jersey: IEEE Press, 2009.

[3] JAIN, A. K.; MURTY, M. N.; FLYNN, P. Data clustering: A review. *ACM Computing Surveys*, v. 31, n. 3, p. 264–323, 1999.

[4] JAIN, A. K.; DUBES, R. C. *Algorithms for Clustering Data*. New Jersey: Prentice Hall, 1998.

[5] EVERITT, B. S. et al. *Cluster Analysis*. London: John Wiley & Sons, Ltd, 2011.

[6] SINGH, S.; SRIVASTAVA, S. Review of clustering techniques in control system. *Procedia Computer Science*, v. 173, p. 272–280, 2020. International Conference on Smart Sustainable Intelligent Computing and Applications under ICITETM2020.

[7] WIERZCHON, S. T.; KLOPOTEK, M. A. *Algorithms of Cluster Analysis*. Warsaw, Poland: Institute of Computer Science, Polish Academy of Sciences, 2015.

[8] BLUM, C.; ROLI, A. Metaheuristics in combinatorial optimization. *ACM Computing Surveys*, v. 35, n. 3, p. 268–308, 2003.

[9] JAIN, A. K. Data clustering: 50 years beyond k–means. *Pattern Recognition Letters*, v. 31, n. 8, p. 651–666, 2010.

[10] ALIA, O.; MANDAVA, R. The variants of the harmony search algorithm: an overview. *Artificial Intelligence Review*, v. 36, n. 1, p. 49–68, 2011.

[11] AL-BETAR, M. A. et al. Coronavirus herd immunity optimizer (chio). *Neural Computing and Applications*, v. 33, p. 5011–5042, 2020. Disponível em: ⟨https://doi.org/10.1007/s00521-020-05296-6⟩.

[12] TURNER, R. *Essentials of Microbiology*. United Kingdom: Ed-Tech Press, 2020.

[13] RASMUSSEN, A. L. Vaccination is the only acceptable path to herd immunity. *Med*, v. 1, n. 1, p. 21–23, 2020.

[14] FINE, P.; EAMES, K.; HEYMANN, D. L. Herd immunity: A rough guide. *Clinical Infectious Diseases*, v. 52, n. 7, p. 911–916, 2011.

[15] FORSATI, R. et al. Efficient stochastic algorithms for document clustering. *Information Sciences*, v. 220, p. 269–291, 2013.

[16] ZHU, W. et al. Clustering algorithm based on fuzzy c-means and artificial fish swarm. *Procedia Engineering*, v. 29, p. 3307–3311, 2012.

[17] XIE, H. et al. Improving k-means clustering with enhanced firefly algorithms. *Applied Soft Computing*, v. 84, p. 105763, 2019.

[18] CUI, X.; POTOK, T. E.; PALATHINGAL, P. Document clustering using particle swarm optimization. In: *Proceedings 2005 IEEE Swarm Intelligence Symposium, 2005. SIS 2005.* Pasadena, California: IEEE, 2005. p. 185–191.

[19] SELIM, S. Z.; ALSULTAN, K. A simulated annealing algorithm for the clustering problem. *Pattern Recognition*, v. 24, n. 10, p. 1003–1008, 1991.

[20] MAULIK, U.; BANDYOPADHYAY, S. Genetic algorithm-based clustering technique. *Pattern Recognition*, v. 33, p. 1455–1465, 2000.

[21] MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Berkeley, California: University of California Press, 1967. p. 281–297.

[22] FRÄNTI, P.; SIERANOJA, S. *K-means properties on six clustering benchmark datasets*. 2018. 4743–4759 p. Disponível em: ⟨http://cs.uef.fi/sipu/datasets/⟩.

[23] DUA, D.; GRAFF, C. *UCI Machine Learning Repository*. 2017. Disponível em: ⟨http://archive.ics.uci.edu/ml⟩.

[24] MANNING, C. D.; RAGHAVAN, P.; SCHüTZE, H. *Introduction to Information Retrieval*. USA: Cambridge University Press, 2008.

[25] ROUSSEEUW, P. 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.

[26] RAND, W. M. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, Taylor & Francis, v. 66, n. 336, p. 846–850, 1971.

[27] FOWLKES, E. B.; MALLOWS, C. L. A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association*, Taylor & Francis, v. 78, n. 383, p. 553–569, 1983.

[28] DAVIES, D. L.; BOULDIN, D. W. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1, n. 2, p. 224–227, 1979.

[29] DUNN, J. C. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, Taylor & Francis, v. 3, n. 3, p. 32–57, 1973.

[30] DERRAC, J. et al. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation*, v. 1, n. 1, p. 3–18, 2011.