

Classificação dos terminais portuários brasileiros utilizando métodos de análise multivariada de dados.

Moacyr Machado Cardoso-Junior <moacyr@ita.br>
Armando Gonçalves Madeira Junior <madeira@ita.br>
Tiago José Menezes Gonçalves <tiagojmg@ita.br>
Rodrigo Arnaldo Scarpel <rodrigo@ita.br>

Resumo: Os serviços portuários são elementos de vital importância para o crescimento econômico de um país. No caso do Brasil este fato ganha maior relevância pelo fato do país ser um exportador de commodities, sendo a maioria composta por carga seca. Neste sentido, este artigo propõe um modelo baseado em técnicas de geração de agrupamentos para classificar os portos de forma a ressaltar as similaridades entre estes através da análise de seus indicadores operacionais. Desta forma, torna-se possível à ANTAQ (Agência Nacional de Transportes Aquaviários), como órgão fiscalizador da atividade, comparar as diferentes práticas adotadas entre portos de mesma natureza, contribuindo para o aprimoramento dos serviços portuários.

Palavras-chave: Agrupamentos; Métricas de Validação; Logística e Transportes; Self Organizing Maps.

Classification of port terminals in Brazil using multivariate data analysis methods.

Abstract: Port services are vital elements for economic growth of a country. In Brazil this fact takes on major importance because the country is an exporter of commodities, mostly consisting of dry cargo. Therefore, this article proposes a model based in methods for cluster generation to classify the ports in order to emphasize the similarities between them by analyzing their operational indicators. Thus it, became possible to ANTAQ (National Agency for Waterborne Transport), as a supervisory body activity, comparing the different practices adopted between ports of the same nature, contributing to their improvement of best port services.

Keywords: Clusters; Metrics Validation; Logistic and Transport; Self Organizing Maps.

1. Introdução

No Brasil, aproximadamente 98% do comércio exterior necessita de serviços portuários, totalizando uma movimentação de cerca de 100 bilhões de dólares por ano. Esses dados demonstram a importância da infraestrutura portuária para as empresas que necessitam de utilizar modais de transportes vantajosos, como o marítimo.

Devido ao Brasil se destacar no cenário internacional como exportador de commodities, o transporte de carga seca no modal marítimo representa um modal de vital importância para o desenvolvimento do país.

A eficiência das instalações de manuseio de carga e os serviços portuários têm influência decisiva sobre a competitividade do transporte de mercadorias. Quando os armadores dos navios enfrentam dificuldades devido ao baixo nível de eficiência do porto ou problemas de congestionamento, começam a procurar soluções alternativas. Uma dessas soluções é repassar os custos adicionais de espera do navio aos clientes por meio de sobretaxa à tarifa básica do frete. Outra é procurar portos alternativos com maior eficiência no manuseio de cargas (ARRUDA; NOBRE JÚNIOR; MAGALHÃES, 2008).

A ANTAQ (Agência Nacional de Transportes Aquaviários) atua como órgão de regulação e supervisão dos portos organizados e terminais privativos, além de implementar as políticas formuladas pelo CONIT – Conselho Nacional de Integração de Políticas de Transporte (ANTAQ, 2011b). Para dar suporte quantitativo de forma a cumprir sua missão institucional, esta agência desenvolveu um Sistema Permanente de Acompanhamento de Preços e Desempenho Operacional dos Serviços Portuários (SDP), o que possibilita relatórios periódicos minuciosos sobre este setor.

Na análise constante do relatório periódico da ANTAQ (ANTAQ, 2011a) é apresentada uma divisão dos portos que movimentam cargas secas conforme a sua natureza: ferro, trigo, milho, fertilizantes, farelo de soja, cromita, coque, bauxita, caulim, alumina, açúcar. Porém, tal classificação inibe a adoção de práticas consideradas adequadas através da comparação entre portos que possuam similaridades operacionais. Isto ocorre porque, utilizando a metodologia de avaliação da ANTAQ, não é possível analisar as melhores práticas de um porto de grande porte (que possui alto índice de atratividade) comparando-as com um porto de pequeno porte que processem a mesma natureza de carga seca.

A contribuição deste trabalho é a agregação dos algoritmos geradores de agrupamentos hierárquicos e não hierárquicos e as métricas de validação de agrupamentos com uma aplicação prática na classificação dos portos brasileiros que processam cargas secas, considerando para tal, os dados operacionais constantes do relatório periódico da ANTAQ: frequência de navios, quantidade movimentada, atendimento ao tráfego, consignação média, produtividade média e tempo médio de espera.

Este trabalho encontra-se estruturado da seguinte forma: a seção 2 apresenta a definição do número de classes, as métricas de validação interna e de estabilidade de agrupamentos hierárquicos e não hierárquicos; a seção 3 apresenta as análises e resultados obtidos com a utilização de algoritmos de agrupamentos não hierárquicos, o K-medóides e os mapas auto-organizados de atributos, também conhecidos por mapas de Kohonen; e por fim, a seção 4 apresenta as considerações finais.

2. Definição do número de classes para portos de cargas secas

O número de técnicas formais para a determinação do número ideal de agrupamentos passa de trinta. Estes autores avaliaram a habilidade de tais métodos indicarem o número ideal de agrupamentos com dados simulados e suas conclusões são de que não se pode generalizar, uma vez que dependem da estrutura desconhecida dos dados, assim como do algoritmo utilizado (MILLIGAN; COOPER, 1985; EVERITT; LANDAU; MORVEN, 2001).

Para determinação do número ideal de classes foi utilizada neste trabalho a comparação de vários métodos de geração agrupamento, hierárquicos e não hierárquicos juntamente com métricas de validação do número de classes encontradas.

Os agrupamentos hierárquicos são gerados com base em uma matriz de dissimilaridade, onde se utilizou a distância Euclidiana. Além disso, foi utilizado o método de

Ward, também conhecido como algoritmo de variância mínima, que busca aglomerar as observações de forma a minimizar a variância total (ver THEODORIDIS; KOUTROUMBAS, 2009). Estes métodos buscam maximizar a distância entre agrupamentos diferentes e minimizar as distâncias intra-agrupamento.

No caso dos métodos não hierárquicos, foi definido inicialmente o número de agrupamentos para particionar os dados, sendo utilizados posteriormente os métodos K-Médias, K-Médias Fuzzy, K-Medóide (PAM – *Partitioning Around Medoids*). O K-Médias Fuzzy é uma variante no qual a idéia é que cada objeto pertence a todos os agrupamentos com um dado grau de pertinência. (WEBB, 2002). O K-Medóides difere do K-Médias pelo fato de definir o centróide de cada agrupamento coincidente com uma das instâncias (THEODORIDIS; KOUTROUMBAS, 2009) e por fim foram utilizados os mapas auto-organizados de atributos (SOM).

A equação 1 apresenta a formulação por programação matemática do algoritmo K-Médias (ver WEBB, 2002).

$$FO = \text{Min} \sum_{i=1}^n \sum_{c=1}^k z_{ic} \left[\sum_{j=1}^p (x_{ij} - m_{cj})^2 \right]^{\frac{1}{2}} \quad (1)$$

$$S.A. \sum_{c=1}^k z_{ic} = 1 \quad i = 1, \dots, n$$

em que:

$$z_{ic} = \begin{cases} 1, & \text{se o ponto } i \text{ pertencer ao agrupamento } c \\ 0, & \text{caso contrario} \end{cases}, \quad c = 1, \dots, k \quad e \quad i = 1, \dots, n$$

$$m_{cj} = \frac{\sum_{i=1}^n z_{ic} x_{ij}}{\sum_{i=1}^n z_{ic}} \quad c = 1, \dots, k \quad e \quad j = 1, \dots, p$$

sendo m_{cj} o centróide do agrupamento c na dimensão j .

Os mapas auto-organizados de atributos, conhecidos como *Self Organizing Maps* (SOM), representam um tipo especial de quantificação vetorial, na qual existe uma ordem ou topologia imposta aos vetores codificados. O objetivo do SOM é representar dados de alta dimensionalidade em uma estrutura de baixa dimensão, usualmente uma ou duas dimensões que capturam a estrutura original dos dados. Agrupamentos distintos de dados são mapeados a diferentes agrupamentos de vetores codificados em uma matriz, sendo o SOM citado como um método para a exploração de dados (WEBB, 2002).

Os autores Wehrens e Buydens (2007) comparam o SOM proposto por Kohonen em 2001 ao Escalonamento Multidimensional (MDS). Este método, ao invés de tentar reproduzir distâncias como no MDS, busca reproduzir uma dada topologia, ou seja, tenta manter os mesmos vizinhos. Desta forma, se dois objetos estão próximos em alta dimensionalidade, então a posição dos mesmos em baixa dimensão também será similar, sendo que o SOM mapeia os objetos em uma grade regular de unidades.

2.2 Métricas para definição do número de classes

Conforme proposto por Brock *et al.* (2010), as métricas utilizadas para validação do número de agrupamentos ou classes nas quais os dados foram particionados são divididas em dois grandes grupos: Internas e de Estabilidade.

Em relação à validação interna, foram selecionadas medidas relacionadas à conectividade, a compactação e a separação dos agrupamentos gerados. A conectividade refere-se à extensão na qual uma instância está alocada a um mesmo agrupamento dos seus vizinhos mais próximos. Já a compactação avalia a homogeneidade do agrupamento, usualmente calculando a variância intra-agrupamento. Por fim, a separação quantifica o grau de separação dos agrupamentos, usualmente pela medida da distância dos centróides.

A conectividade é definida conforme a equação 2.

$$con(C) = \sum_{i=1}^N \sum_{j=1}^M x_i nn_{i(j)} \quad (2)$$

Onde:

N representa o número total de observações;

M representa o número de dimensões;

$nn_{i(j)}$ consiste no j -ésimo vizinho mais próximo da instância i na dimensão j ; e

$x_i nn_{i(j)} = 0$ se i e j estiverem no mesmo agrupamento ou $1/j$ caso contrário.

A conectividade tem valores que variam entre $0 \leq con(C) \leq \infty$, sendo que quanto menor seu valor, melhor será a estrutura proposta pelo algoritmo (EVERITT; LANDAU; MORVEN, 2001).

Devido ao fato de que a compactação e a separação apresentarem tendências opostas (sendo que a primeira aumenta com o número de agrupamentos e a segunda diminui), uma opção é unir as duas métricas de forma a proporcionar uma análise mais equilibrada. Duas medidas que representam uma combinação não linear de compactação e separação são respectivamente o índice de Dunn e a largura de Silhouette (EVERITT; LANDAU; MORVEN, 2001).

O índice Dunn consiste na razão entre a menor distância entre instâncias que não estejam no mesmo agrupamento e a maior distância intra-agrupamento. O valor do índice Dunn varia de 0 a 1 e quanto mais próximo de 1 melhores são os resultados dos os agrupamentos (BROCK *et al.* 2010).

Já na largura de Silhouette (KAUFFMAN; ROUSSEEUW, 1990), para cada instância i é definida um índice $S(i) \in [-1, +1]$ que mede a diferença entre $b(i)$ e $a(i)$, onde $a(i)$ é a dissimilaridade média da instância i em relação ao seu agrupamento e $b(i)$ é a dissimilaridade média da instância i em relação a todas as instâncias no agrupamento mais próximo. Quando $S(i)$ está próximo de +1 a instância i está mais próxima do seu agrupamento do que do agrupamento vizinho, e, portanto representa uma boa alocação. Um valor de $S(i)$ próximo de -1 indica que a instância está mal alocada. Os autores da proposta indicam ainda que valores acima de 0,5 representam um bom resultado e que valores abaixo de 0,2 podem indicar ausência de estruturação clara dos dados.

Cabe ressaltar que, conforme Everitt, Landau e Morven (2001) recomendam, não é prudente confiar em apenas uma das métricas de validação interna para selecionar o número ideal de agrupamentos, tendo sido seguido este princípio neste trabalho.

Em relação à validação do número de agrupamentos utilizando o conceito da estabilidade, este se baseia na retirada de uma coluna por vez da matriz inicial e na comparação dos agrupamentos formados antes e depois desta retirada. As métricas utilizadas neste trabalho foram: Proporção Média de Não Sobreposição (*Average Proportion of Non-overlap* - APN), Distância Média (*Average Distance* - AD), Distância Média entre Centros dos Agrupamentos (*Average Distance between Means* - ADM) e Figura de Mérito (*Figure of Merit* - FOM).

A APN mede a proporção média de instâncias não alocadas ao mesmo agrupamento quando a matriz completa é utilizada e quando uma coluna é retirada. Seja $C^{i,0}$ o agrupamento que contém a instância i usando a matriz completa e $C^{i,l}$ quando uma coluna l foi retirada, considerando o número total de agrupamentos igual a K , sendo definida conforme equação 3.

$$APN(K) = \frac{1}{MN} \sum_{i=1}^N \sum_{l=1}^M \left(1 - \frac{n(C^{i,l} \cap C^{i,0})}{nC^{i,l}} \right) \quad (3)$$

Com $APN(K) \in [0,1]$, sendo que valores próximos a zero correspondem a um agrupamento consistente (BROCK *et al.*, 2010).

A AD mede a distância média entre as instâncias de um mesmo agrupamento baseada na matriz completa e com a retirada de uma coluna, sendo é definida conforme equação 4.

$$AD(K) = \frac{1}{MN} \sum_{i=1}^N \sum_{l=1}^M \frac{1}{n(C^{i,0})n(C^{i,l})} \left[\sum_{i \in C^{i,0} \wedge j \in C^{i,l}} dist(i, j) \right] \quad (4)$$

Com $AD(K) \in [0, \infty]$, sendo que valores próximos a zero representam melhores resultados.

A ADM mede a distância média entre centros dos agrupamentos, e é definida conforme a equação 5.

$$ADM(K) = \frac{1}{MN} \sum_{i=1}^N \sum_{l=1}^M dist(\bar{X}_{C^{i,l}}, \bar{X}_{C^{i,0}}) \quad (5)$$

Com $ADM(K) \in [0, \infty]$, sendo que da mesma forma que AD, os valores mais próximos de 0 representam melhores resultados.

Por fim, a FOM mede a variância média intra-agrupamento das instâncias na coluna suprimida, onde o agrupamento está baseado nas colunas restantes. Para tanto é estimado o erro médio usando previsões baseadas nas médias dos agrupamentos. Para uma coluna l deixada de fora, o cálculo é realizado conforme a equação 6.

$$FOM(K) = \left(\frac{1}{N} \sum_{k=1}^K \sum_{i \in CK(l)} dist(x_{i,l}, \bar{x}_{CK(l)}) \right)^{\frac{1}{2}} \quad (6)$$

Onde $x_{i,l}$ é o valor da i -ésima instância da l -ésima coluna no agrupamento $CK(l)$ e $\bar{x}_{CK(l)}$ é a média do agrupamento. Finalmente FOM é multiplicado por um fator $(N/(N-K))^{1/2}$, para minimizar sua tendência de decrescer à medida que K cresce. O escore final é calculado pela média de todas as colunas removidas, e os valores obtidos estão entre zero e ∞ , com os menores valores representando a melhor configuração (BROCK *et al.*, 2010).

3. Resultados

Todas as métricas descritas anteriormente foram aplicadas ao conjunto de dados de indicadores operacionais na busca pela melhor estruturação dos dados em termos de classes. Para tanto foram testadas de 2 a 6 classes, com os métodos de formação de agrupamentos hierárquicos Ward e não-hierárquicos K-Médias, Fuzzy K-Médias, K-Medóides e Kohonen (*Self Organizing Maps*). Foi utilizado o pacote *clValid* do *software* R versão 2.9, implementado por Brock *et al.* (2010).

Os resultados obtidos para a validação interna estão na Tabela 1, sendo que a abreviação “NA” significa que o algoritmo não convergiu no limite arbitrado de 500 iterações.

Tabela 1 – Resultado da validação interna do número ideal de agrupamentos.

Método de agrupamento	Métrica	2 Classes	3 Classes	4 Classes	5 Classes	6 Classes
Hierárquico – Ward	Conectividade	9.3587	12.2155	17.4575	27.3413	27.5913
	Dunn	0.0489	0.0583	0.0583	0.0280	0.0680
	Silhouette	0.3230	0.3774	0.4225	0.3698	0.3760
K-Médias	Conectividade	14.8028	15.1635	21.9044	32.4437	32.6937
	Dunn	0.0376	0.0543	0.0521	0.0263	0.0770
	Silhouette	0.3343	0.3855	0.4334	0.3980	0.4042
K-Medóides (PAM)	Conectividade	13.2313	12.9635	21.2306	32.0226	32.0226
	Dunn	0.0376	0.0554	0.0528	0.0201	0.0607
	Silhouette	0.3393	0.3776	0.4203	0.3794	0.3971
K-Médias Fuzzy	Conectividade	13.8790	46.0829	61.0619	66.3742	NA
	Dunn	0.0376	0.0252	0.0137	0.0171	NA
	Silhouette	0.3312	0.3398	0.2861	0.2788	NA
Kohonen (SOM)	Conectividade	6.5369	16.4901	21.9044	32.4437	40.6468
	Dunn	0.7697	0.0543	0.0521	0.0263	0.0263
	Silhouette	0.8079	0.3849	0.4334	0.3980	0.3728

Fonte: Elaborado pelos autores.

Pode-se observar que os melhores resultados apontam para apenas duas classes, sendo que o algoritmo de agrupamento *Self Organizing Maps* foi o que produziu os melhores resultados nas 3 métricas de validação interna.

Na Tabela 2 são apresentados os resultados para as métricas de estabilidade de agrupamentos.

Tabela 2 – Resultado das métricas de estabilidade de agrupamentos.

Método de agrupamento	Métrica	2 Classes	3 Classes	4 Classes	5 Classes	6 Classes
Hierárquico – Ward	APN	0.0354	0.1246	0.1284	0.2020	0.2799
	AD	2.1719	1.8942	1.6433	1.5375	1.4614
	ADM	0.3600	0.4670	0.3490	0.4637	0.5615
	FOM	0.9954	0.9026	0.8240	0.8170	0.7901
K-Médias	APN	0.0407	0.0706	0.1135	0.1797	0.2718
	AD	2.1717	1.8316	1.6393	1.5185	1.4492

	ADM	0.3737	0.3114	0.3577	0.4365	0.5502
	FOM	0.9958	0.8387	0.8315	0.8224	0.7944
<i>K-Medóides (PAM)</i>	APN	0.0387	0.1038	0.1604	0.1554	0.2726
	AD	2.1795	1.8650	1.7010	1.5461	1.4439
	ADM	0.3962	0.3356	0.4787	0.4809	0.5156
	FOM	0.9994	0.8380	0.8351	0.8314	0.8006
<i>K-Médias Fuzzy</i>	APN	0.0965	0.1784	0.2510	0.2575	NA
	AD	2.1552	1.9803	1.9077	1.8139	NA
	ADM	0.2635	0.4303	0.6369	0.6237	NA
	FOM	0.9930	0.9606	0.9622	0.9366	NA
<i>Kohonen (SOM)</i>	APN	0.0815	0.0357	0.1076	0.1768	0.2086
	AD	2.2765	1.8159	1.6071	1.5456	1.4775
	ADM	0.2387	0.2835	0.2850	0.5093	0.5677
	FOM	0.8612	0.7884	0.7314	0.7877	0.7968

Fonte: Elaborado pelos autores.

Para as métricas de estabilidade interna dos clusters o número de agrupamentos não convergiu, sendo que tivemos valores ótimos para 2 classes nas métricas APN, com o algoritmo de geração de agrupamentos hierárquicos Ward, e na métrica ADM com SOM. A métrica FOM indicou 4 classes para SOM e a métrica AD apontou para 6 agrupamentos utilizando K-Medóides (PAM).

Devido a esta indefinição do número de classes, e também pelo fato de não existirem agrupamentos previamente definidos, optou-se por estudar as potenciais classes por dois métodos a fim de compreender melhor a estrutura dos dados. A primeira abordagem utilizada foi a do SOM e a segunda foi utilizando K-Medóides (PAM). O algoritmo do SOM foi implementado conforme Wehrens e Buydens (2007).

A figura 1 mostra o mapa SOM para dois clusters, onde em uma classe foram definidos os portos CK, CL e CS (ver tabelas de 3 a 6), e na outra classe todos os demais portos. Ressalta-se que esta configuração foi a que produziu os melhores valores na métrica de validação interna dos agrupamentos.

Na figura 2 apresenta-se à esquerda um mapa de qualidade do ajuste, isto é, a distância média do objeto para o vetor de códigos daquela unidade. Quanto menor a distância, melhor aquele objeto é representado pela unidade. Assim a classe dos portos CK, CL e CS é a que apresenta o pior resultado em termos de qualidade do ajuste. No mapa de atributos fica evidenciado que a classe dos portos descritos anteriormente apresenta relevância para todos os atributos na composição do agrupamento.

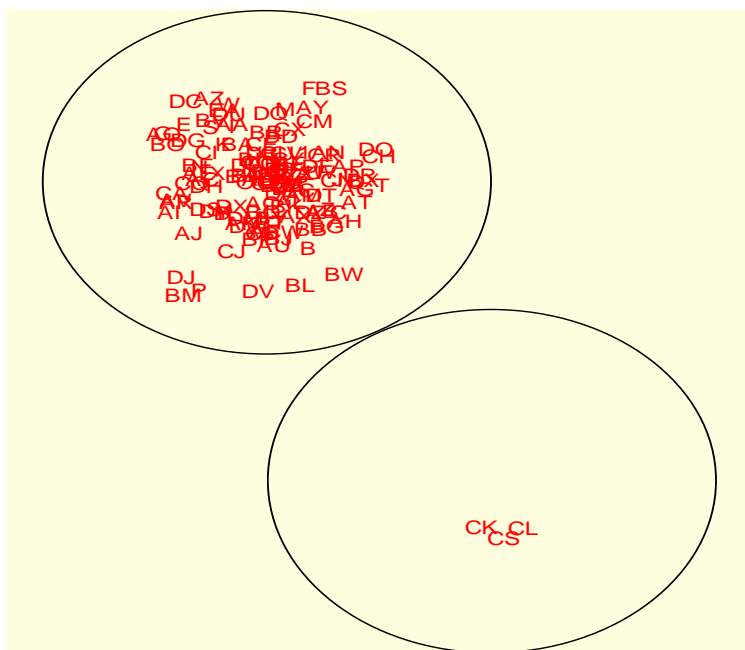


Figura 1 – Mapa auto-organizado de Atributos (SOM) para os Portos de Carga Seca para duas classes. Fonte: Elaborado pelos autores.

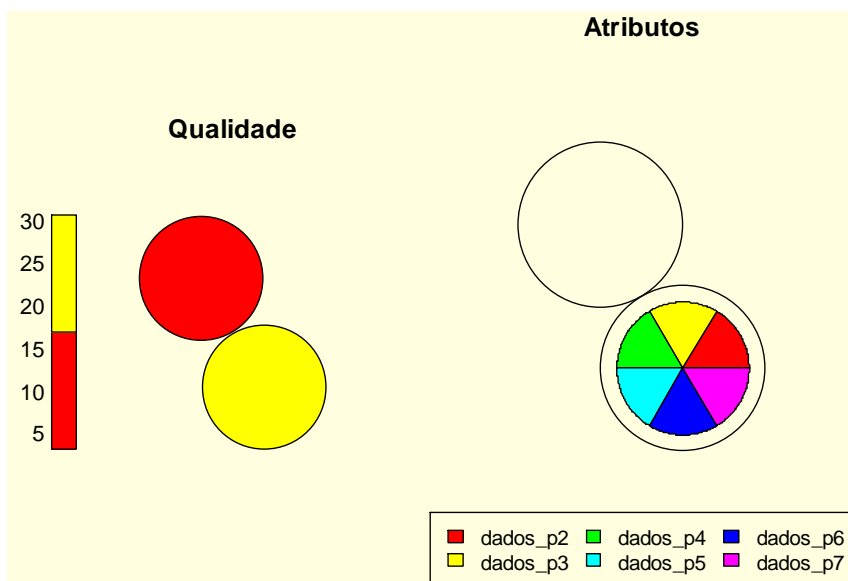


Figura 2 – Mapa de Qualidade do ajuste e dos Atributos, pelo método SOM para duas classes. Fonte: Elaborado pelos autores.

Na figura 3 é apresentada a classificação para 4 classes, conforme indicado pela métrica de estabilidade FOM. No caso de 4 classes, novamente percebe-se que os portos CS, CK e CL aparecem juntos em uma mesma classe.

Na figura 4 é apresentada a qualidade do ajuste e o mapa de atributos. Novamente percebe-se que a classe dos portos CS, CK e CL é que apresenta o pior ajuste, enquanto as outras três classes apresentam desempenhos similares. Na questão dos atributos, verificou-se que na segunda classe predomina o atributo Atendimento, na terceira predominaram o

Atendimento (em menor grau) e o Tempo de Espera Médio, e finalmente na última não houve predominância de algum atributo em especial.

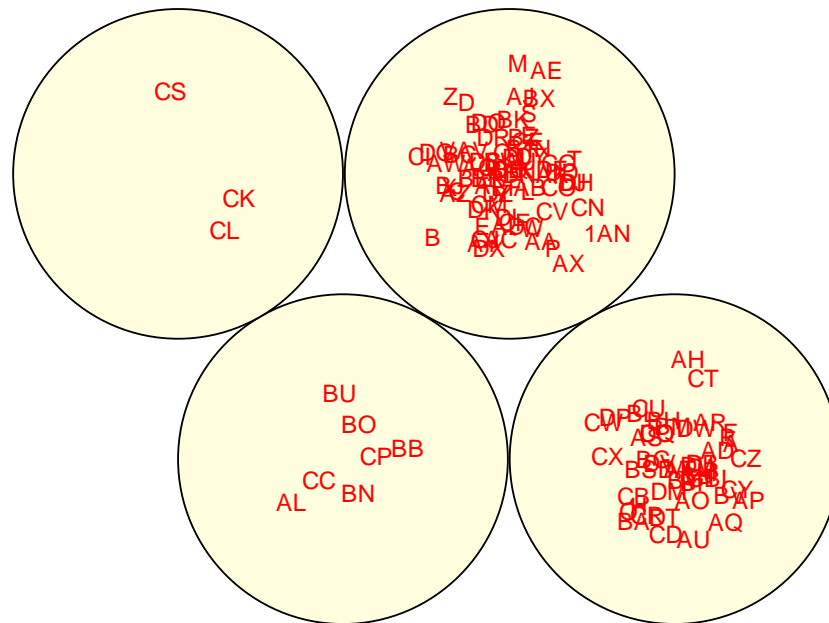


Figura 3 – Mapa auto-organizado de Atributos (SOM) para os Portos de Carga Seca, para 4 classes. Elaborado pelos autores.

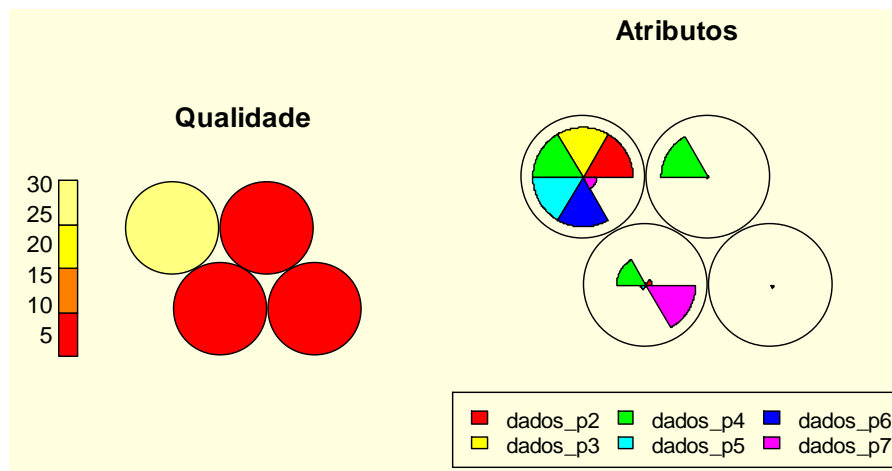


Figura 4 – Mapa de Qualidade do ajuste e dos Atributos, pelo método SOM para quatro classes. Fonte: Elaborado pelos autores.

As tabelas de 3 a 6 apresentam o extrato da classificação dos portos segundo o algoritmo K-Medóides para 6 classes de portos. Convém destacar que a classe 6 foi representada por apenas um porto (Porto Velho - Cais Publico, com descritor CS). A classe 5 apresentou como componentes somente os portos TUP CVRD Tubarão (CK) e TUP Ponta da Madeira (CL). A análise mostra ainda que a classe 5 está isolada das demais, ou seja, o diâmetro da classe é menor do que a separação das demais classes.

Na figura 5 pode-se visualizar o *clusplot*, ou seja, as duas primeiras componentes principais, que explicam 68% da variação. Pode-se visualizar as 6 classes e o isolamento da classe 5. Na mesma figura pode-se ainda verificar o ajuste pelo diagrama Silhouette, onde

quanto mais próximo de 1 melhor é o ajuste. Verifica-se que de forma geral os objetos estão bem alocados às classes.

Tabela 3 – Resultado do Agrupamento pelo algoritmo K-Medóides para classe 1 para os 13 primeiros portos.

Nome	Classe	Produto	Porto
Maceió Cais Público	1	açúcar	A
Santos - Cargill	1	açúcar	F
Santos - Teaçú 1	1	açúcar	G
Santos - Teaçú 2	1	açúcar	H
Santos - Teaçú 3	1	açúcar	I
Santos - Cais Público	1	bauxita	O
Vila do Conde - Cais Público	1	caulim	R
TUP Manaus	1	coque	AD
Vitoria - Capuaba	1	coque	AH
Paranaguá - Múltiplo Uso	1	farelo soja	AM
Rio Grande - Bianchini	1	farelo soja	AO
Rio Grande - Termasa	1	farelo soja	AP
Rio Grande - Terminal Bunge	1	farelo soja	AQ

Fonte: Elaborado pelos autores.

Tabela 4 – Resultado do Agrupamento pelo algoritmo K-Medóides para classe 2 para os 10 primeiros portos.

Nome	Classe	Produto	Porto
Maceió Empat	2	açúcar	B
Paranaguá - Múltiplo Uso	2	açúcar	C
Porto Velho - Cais Público	2	açúcar	D
Recife - Cais Público	2	açúcar	E
Aratu - Cais Público	2	alumina	J
Vila do Conde - Cais Público	2	alumina	K
Recife - Cais Público	2	barrilha	L
São Sebastião - Cais Público	2	barrilha	M
Aratu - Cais Público	2	bauxita	N
Aratu - Cais Público	2	coque	S

Fonte: Elaborado pelos autores.

Tabela 5 – Resultado do Agrupamento pelo algoritmo K-medóides para Classe 3 para os 10 primeiros portos.

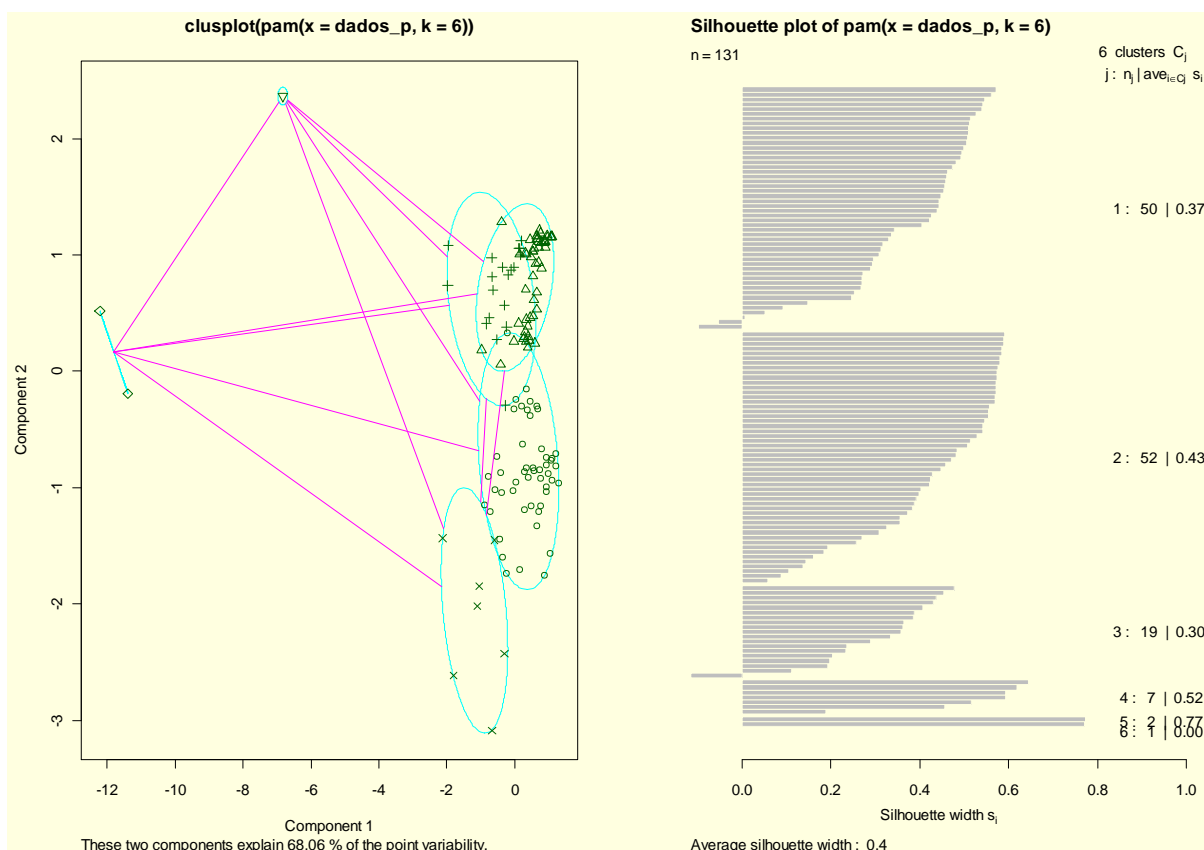
Nome	Classe	Produto	Porto
TUP Trombetas	3	bauxita	P
Vila do Conde - Cais Público	3	bauxita	Q
Fortaleza - Cais Público	3	coque	V
Imbituba - Cais Público	3	coque	W
Recife - Cais Público	3	coque	Y
Santos - Cosipa	3	coque	AB
TUP CVRD Praia Mole	3	coque	AE
TUP USIMINAS	3	coque	AF
Vitoria - Praia Mole	3	coque	AI
Rio Grande - Termasa	3	milho	BZ

Fonte: Elaborado pelos autores.

Tabela 6 – Resultado do Agrupamento pelo algoritmo K-medóides, para Classes 4,5 e 6

Nome	Classe	Produto	Porto
Paranaguá - Corex	4	farelo soja	AL
Paranagua - Múltiplo Uso	4	fertilizantes	BB
Santos - TMG	4	fertilizantes	BN
Santos - Ultrafertil	4	fertilizantes	BO
Paranaguá - Corex	4	milho	BU
Santos - Corex	4	milho	CC
TUP CVRD Tubarão	5	ferro	CK
TUP Ponta da Madeira	5	ferro	CL
Paranaguá - Corex	4	ferro	CP
Porto Velho - Cais Público	6	ferro	CS

Fonte: Elaborado pelos autores.


 Figura 5 – Clusplot (duas componentes principais) das Classes (à esquerda) e o Diagrama Silhouette (à direita).
 Fonte: Elaborado pelos autores.

4. Considerações finais

Este trabalho teve como objetivo investigar a utilização de diversos algoritmos geradores de agrupamentos e suas métricas de validação para a classificação dos portos brasileiros que processam cargas secas. Para tanto, foram considerados os dados operacionais da ANTAQ referentes à frequência de navios, quantidade movimentada, atendimento ao tráfego, consignação média, produtividade média e tempo médio de espera.

A utilização de métricas de validação em conjunto com os diversos métodos de agrupamentos (hierárquicos e não hierárquicos) demonstrou ser eficiente para classificar os

portos que processam cargas secas, bem como definir o número ideal de agrupamentos levando em consideração os indicadores operacionais, diferentemente do critério atual da ANTAQ que contempla apenas a natureza da carga movimentada.

A partir do emprego da metodologia apresentada em conjunto com a análise do índice de atratividade de cada porto (realizada pela ANTAQ), é possível identificar em cada cluster o porto que se encontra na fronteira da eficiência, proporcionando uma priorização para identificar portos operacionalmente eficientes ou não. Com esta análise, torna-se possível aos portos que estão fora da fronteira de eficiência do agrupamento analisar as práticas dos portos considerados mais eficientes e adotar as práticas consideradas mais adequadas visando melhorar seu processo produtivo.

Por fim, ressalta-se que a aplicação desta metodologia como uma ferramenta para a agência reguladora (ANTAQ) pode gerar resultados interessantes na tarefa de fiscalização dos portos organizados e terminais de uso privativo, o que poderia vir a contribuir para a redução das tarifas de utilização do serviço e, em nível internacional, contribuir para o Brasil ter produtos mais competitivos em outros países do mundo.

5. Referências

ANTAQ - Agência Nacional de Transporte Aquaviário. Sistema de desempenho portuário – relatório técnico 2008. Disponível em <<http://www.antaq.gov.br/portal/desempenhportuario/documentos/relatoriodesempenhoportuario2008.pdf>>. Acesso em 20 fev. 2011a.

ANTAQ - Agência Nacional de Transporte Aquaviário. Desempenho portuário Disponível em <<http://www.antaq.gov.br/Portal/DesempenhoPortuario/Index.asp>>. Acesso em 20 Fev. 2011b.

ARRUDA, C. M.; NOBRE JÚNIOR, E. F.; MAGALHÃES, P. S. B. Método dos indicadores de desempenho proposto pela ANTAQ: uma aplicação ao terminal portuário do Pecém. In: Encontro Nacional de Engenharia de Produção, 28., 2008, Rio de Janeiro. *Anais...* Rio de Janeiro: Associação Brasileira de Engenharia de Produção, 2008.

BROCK, G.; PIHUR, V.; DATTA, S.; DATTA, S. *clValid: Validation of Clustering Results*. Disponível em <<http://www.louisville.edu/~g0broc01/research>>. Acesso em: 13 mar. 2010.

EVERITT, B. S.; LANDAU, S.; MORVEN, L. *Cluster Analysis*. Arnold: London, 2001.

KAUFMAN, L.; ROUSSEEUW, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons: New York, 1990.

MILLIGAN, G. W.; COOPER, M. An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, v. 50, p. 159-179, 1985.

THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition*. Elsevier: London, 2009.

WEBB, A. *Statistical Pattern Recognition*, 2. ed. John Wiley & Sons: West Sussex, 2002.

WEHRENS, R.; BUYDENS, L. M. C. Self- and super-organizing maps in R: The kohonen package. *Journal of Statistical Software*, v. 21, n. 5, 2007.