

6

Clustering: Definindo Grupos Estratégicos

6.1

Introdução

Não é difícil deduzir que as 60 empresas distribuidoras de energia elétrica, operantes no SEB, apresentam características heterogêneas. Entretanto, é indispensável que o pesquisador garanta, antes de aplicar os métodos da empresa de referência, DEA e SFA, que as distribuidoras analisadas sejam comparáveis [83].

Para atender a este requisito, os dados das 60 concessionárias foram processados por uma Rede Neural auto-organizável. É interessante registrar que a razão da escolha da Rede Neural de Kohonen deve-se, principalmente, aos seguintes fatores:

- a) Obtenção de um mapa topologicamente ordenado das distribuidoras, o que facilita a visualização e a interpretação dos agrupamentos;
- b) Usualmente, é mais fácil interpretar a operação das redes não-supervisionadas através da análise dos pesos.

À luz deste cenário, é importante mencionar que as variáveis (atributos) escolhidas para o desenvolvimento deste estudo estão em concordância com o trabalho de Zanini [97]. São elas:

- i. Custo operacional (OPEX);
- ii. Quantidade de energia distribuída (MWh);
- iii. Total de unidades consumidoras atendidas;
- iv. Extensão da rede de distribuição (Km).

6.2

Análise de dados

A análise de dados é requerida em praticamente todas as aplicações computacionais.

Martin-del-Brio & Serrano-Cinca [59] mencionam que o pré-processamento dos dados de entrada é um importante aspecto que tem demonstrado apresentar bastante influência tanto no desenvolvimento quanto na configuração do mapa de Kohonen. Dessa forma, a escolha de alguma estratégia para tratamento dos dados está diretamente relacionada com o objetivo que se queira alcançar e, também, com a forma como os dados vão estar dispostos.

Quando as variáveis de entrada têm dimensões muito diferentes, apresentar os dados em uma escala linear é uma maneira de inicializar cada variável por igual, ou seja, todas têm o mesmo grau de importância. Desta forma, cada componente do vetor de entradas é colocado em uma escala de mesma dimensão, por exemplo [-1,1].

Ainda de acordo com Martin-del-Brio & Serrano-Cinca [59], é interessante normalizar o vetor de dados de maneira que ele apresente comprimento unitário. Este procedimento, bastante comum, é muito utilizado em mapeamento auto-organizável quando a medida usada é a correlação. No entanto, quando é utilizada a medida de distância euclidiana, este procedimento não é estritamente recomendado, exceto se a norma dos vetores for muito alta.

6.3

Resultados obtidos via *cluster analysis*

Os quatro (4) atributos (padronizados) foram processados em uma Rede Neural de Kohonen cujo mapeamento usado consiste de dezesseis (16) unidades arranjadas em uma grade quadrada 4 x 4 e a medida de similaridade utilizada foi a distância euclidiana. Como o número de *clusters* não é previamente conhecido, tipicamente este número é sobre-estimado, usando-se mais unidades do que o necessário [67].

Inicialmente, atribuiu-se aos pesos sinápticos pequenos valores aleatórios no intervalo [-1,1]. Posteriormente, a rede foi treinada por 2.500 iterações sendo que, em cada iteração, os 60 padrões são apresentados à rede em seqüência

aleatoriamente definida para a iteração em vigor. A função de decaimento exponencial foi usada para atualizar o raio da distância da unidade vencedora às suas unidades vizinhas [82]:

$$\text{dist} = \text{round} \left(0,35 + \text{dist}0 \cdot \exp \left(- \left(\frac{\text{dec} \cdot \text{iter}}{\text{Niter}} \right) \right) \right) \quad (6.3.1)$$

onde:

- round = função que executa o arredondamento para o número inteiro mais próximo;
- dist0 = 3, distância inicial;
- exp = função exponencial;
- dec = constante de decaimento;
- iter = iteração atual;
- Niter = número total de iterações, 2.500.

A taxa de aprendizado η é atualizada em cada iteração da seguinte forma:

$$\eta = \eta_0 \cdot \exp \left(- \left(\frac{\text{dec} \cdot \text{iter}}{\text{Niter}} \right) \right) \quad (6.3.2)$$

onde:

η = valor da taxa de aprendizado em cada iteração;

η_0 = valor inicial da taxa de aprendizado: 0,0058.

A Figura 6.1 mostra os padrões similares descobertos pela Rede Neural de Kohonen:

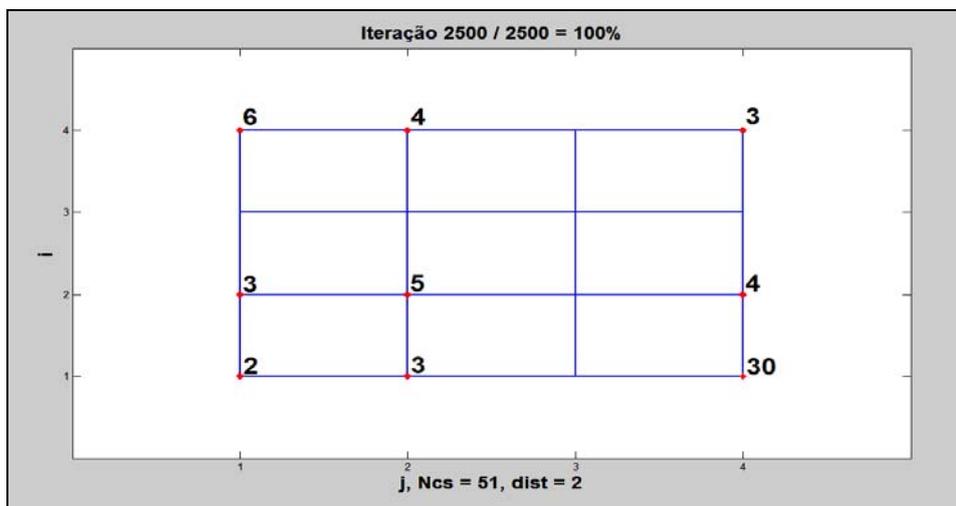


Figura 6.1: Resultado de um processo de *clustering* em uma simulação.

Os resultados de cada *cluster*, mostrados na Figura 6.1, são decodificados na Tabela 6.2:

Tabela 6.2 – *Clusters* obtidos pela Rede de Kohonen em uma simulação.

<i>Cluster</i>	Concessionárias
C1	[1 9 17 41 53]
C2	[2 15 22 28]
C3	[3 4 5 6 7 11 16 20 24 25 26 31 33 34 36 37 42 43 46 47 48 49 50 52 55 56 57 58 59 60]
C4	[8 30 45 54]
C5	[10 14 51]
C6	[12 19 27 32 38 44]
C7	[13 29 35]
C8	[18 40]
C9	[21 23 39]

Neste ponto, é pertinente apresentar os resultados obtidos através da Rede Neural de Kohonen via simulação de Monte Carlo (MCRNK). Para tal, realizou-se 5.000 simulações resultando em 48 *clusters* distintos. A Figura 6.2 mostra o histograma dos *clusters* encontrados:

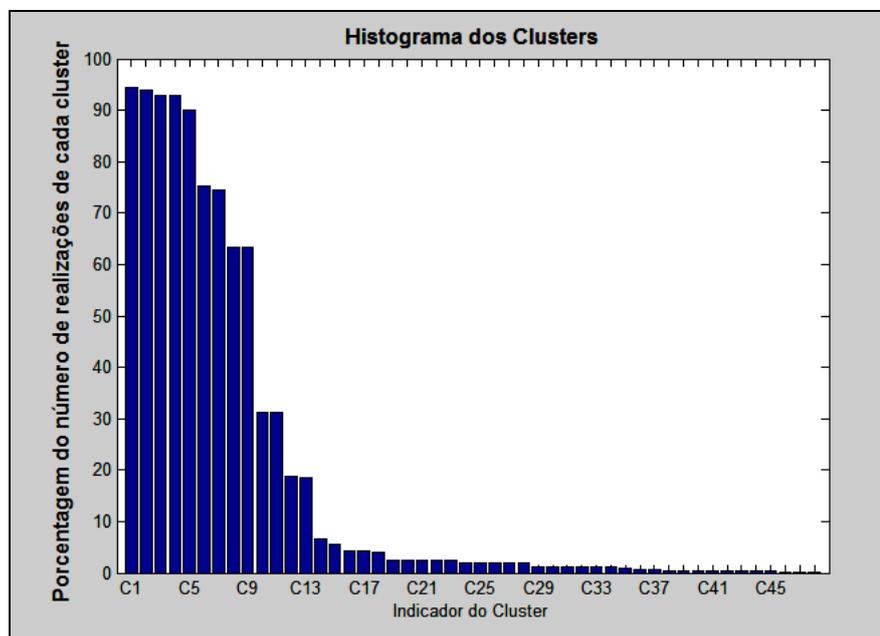


Figura 6.2: Histograma dos *clusters*.

Com base em uma análise pictórica do histograma (Figura 6.2) buscou-se elaborar a Tabela 6.3 cujo objetivo é apresentar a frequência e também a composição de cada *cluster*.

Tabela 6.3 – Sequência ordenada dos *clusters* mais freqüentes.

<i>Cluster</i>	Porcentagem	Concessionárias
C1	94,48 %	[13 15 22 28 35]
C2	94,04 %	[30 54]
C3	93,00 %	[12 19 27 32 38 44]
C4	92,96 %	[9 17 18 40]
C5	89,96 %	[2 29]
C6	75,38 %	[10 14 53]
C7	74,36 %	[1 41 51]
C8	63,38 %	[8 16 21 23 39 45]
C9	63,38 %	[3 4 5 6 7 11 20 24 25 26 31 33 34 36 37 42 43 46 47 48 49 50 52 55 56 57 58 59 60]
C10	31,10 %	[6 8 16 21 23 39 45]
C11	31,10 %	[3 4 5 7 11 20 24 25 26 31 33 34 36 37 42 43 46 47 48 49 50 52 55 56 57 58 59 60]
C12	18,66 %	[10 53]
C13	18,60 %	[1 14 41 51]
C14	6,56 %	[12 19 32 38 44]
C15	5,52 %	[3 4 5 6 7 11 16 20 24 25 26 31 33 34 36 37 42 43 46 47 48 49 50 52 55 56 57 58 59 60]
C16	4,40 %	[23 39 45]
C17	4,40 %	[8 21 30 40 54]
C18	4,08 %	[2 27 29]
C19	2,48 %	[2 35]
C20	2,48 %	[10 15 28]
C21	2,48 %	[1 41 51 53]
C22	2,48 %	[9 14 17 18]
C23	2,48 %	[13 22 27 29]
C24	1,92 %	[2]
C25	1,92 %	[18 28]
C26	1,92 %	[1 17 41 53]
C27	1,92 %	[9 10 14 51]
C28	1,92 %	[13 15 22 29 35]
C29	1,12 %	[10 14 51]
C30	1,12 %	[13 29 35]
C31	1,12 %	[2 15 22 28]
C32	1,10 %	[18 40]
C33	1,10 %	[1 9 17 41 53]
C34	1,08 %	[9 18 40]
C35	1,02 %	[1 17 41 51]
C36	0,72 %	[21 23]
C37	0,72 %	[8 30 39 45 54]
C38	0,44 %	[10]
C39	0,44 %	[27 29]
C40	0,44 %	[9 14 18]
C41	0,44 %	[30 40 54]
C42	0,44 %	[1 17 41 51 53]
C43	0,44 %	[2 12 19 32 38 44]
C44	0,40 %	[21 23 29]
C45	0,40 %	[8 30 45 54]
C46	0,06 %	[1 14 17 41 51]
C47	0,02 %	[18 40 53]
C48	0,02 %	[1 9 17 41]

O entendimento da Tabela 6.3 é direto. Como exemplo, seja a concessionária número 6; em 63,38% das realizações (simulações), ela pertence ao *cluster* 9 (C9). No entanto, ela também pode, com 31,10 %, pertencer ao *cluster* 10 (C10).

Para uma melhor inspeção visual, o histograma é truncado nos 13 *clusters* mais freqüentes. A Figura 6.3 mostra este processo:

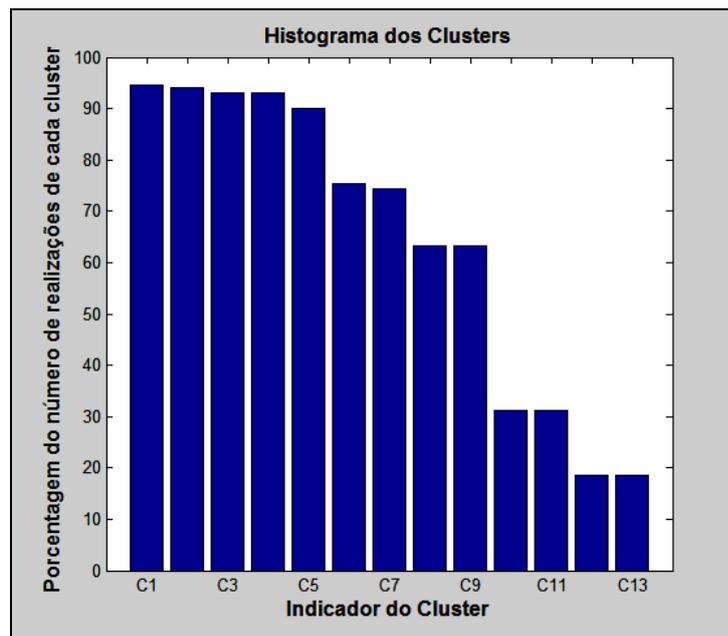


Figura 6.3: Histograma dos 13 *clusters* mais freqüentes.

Pela Tabela 6.3, percebe-se que, com exceção do *cluster* 9 (que contém um número expressivo de empresas), há vários *clusters* constituídos de poucas empresas. Baseado na análise dos pesos sinápticos, foi possível obter três (3) agrupamentos. Com este procedimento, pode-se verificar que cada *cluster* passa a ser composto de um número significativo de empresas com características similares, conforme informado na Tabela 6.4:

Tabela 6.4 – Agrupamentos finais.

<i>Cluster</i>	<i>Concessionárias</i>
C1	[1 8 10 14 16 21 23 30 39 41 45 51 53 54]
C2	[2 9 12 13 15 17 18 19 22 27 28 29 32 35 38 40 44]
C3	[3 4 5 6 7 11 20 24 25 26 31 33 34 36 37 42 43 46 47 48 49 50 52 55 56 57 58 59 60]

Vale informar que estes agrupamentos representam uma importante etapa na obtenção dos próximos resultados.