

Segmentação do SOM por Métodos de Agrupamentos Hierárquicos com Conectividade Restrita

José Alfredo Ferreira Costa

Departamento de Engenharia Elétrica – Centro de Tecnologia
Universidade Federal do Rio Grande do Norte - 59.072-970 - Natal - RN
E-mail: alfredo@dee.ufrn.br

Abstract - *The self-organizing map (SOM) forms a nonlinear projection from a high-dimensional data onto a regular (usually) two-dimensional grid and have been widely studied as a software tool for visualization of high-dimensional data. However, the automation of knowledge discovery in SOM is not straightforward. A complementary stage is need in trained SOMs when dealing with complex data clustering problems. This paper investigates the usage of contiguity constrained hierarchical clustering in SOM outputs. The objective is to find regions of neurons that are related classes of input data. The distributed prototype representation enables the discovery of complex geometries and even nonlinear separable data clusters in an unsupervised way. It is also shown that the algorithm can also estimate the probable number of clusters.*

1. Introdução

Presenciamos um aumento explosivo no uso da informação em meios digitais e em bancos de dados nas mais variadas aplicações, seja na indústria, no comércio, no governo, na área científica, etc. Devido à crescente disponibilidade de grandes massas de dados [1,2] armazenados em computadores, a necessidade de métodos que possam analisá-los de forma automática, ou não supervisionada, torna-se cada vez maior.

Dentre as principais técnicas de análise de dados, a análise de agrupamentos (ou *classificação automática - CA*) tem como objetivo descobrir a estrutura (ou classes, caso existam) de um determinado conjunto de dados não rotulados, $X = \{x_1, x_2, \dots, x_n\}$, onde cada objeto (ou amostra) x_i , $i = 1, \dots, n$, é descrito por p variáveis (atributos ou características). Deve-se diferenciar classificação automática da classificação de padrões obtida por um processo de treinamento supervisionado. Nesta última, o classificador é visto como um discriminador, ou máquina de decisão, a qual foi treinada com dados rotulados, i.e., com origem de classe conhecida, com o objetivo de maximizar a generalização para futuras amostras de dados não usadas durante o treinamento. Por outro lado, CA objetiva descobrir as relações entre os dados, a estrutura destes, e caso seja possível, um esquema de análise de futuras amostras [3].

Duas classes de métodos prevalecem: os métodos hierárquicos, que partem de uma matriz de similaridades e geram uma árvore das fusões entre os

agrupamentos; e os métodos particionais, que geram apenas uma partição dos dados, sendo mais utilizados em problemas de engenharia, como em quantização vetorial.

Apesar do processo ser autônomo, i.e., não supervisionado, na maioria das abordagens o usuário tem grande influencia sobre o resultado final do processo através da escolha de parâmetros iniciais. Além disso, a maioria dos métodos, por exemplo, o Expectation-Maximization (EM) [4], impõem geometrias pré-estabelecidas aos agrupamentos, além de possuírem custo computacional elevado.

Redes neurais competitivas tem sido aplicadas a problemas de agrupamento de dados [3-10]. Dentre as redes competitivas, o SOM – *self-organizing map* – é o modelo em maior uso na atualidade [5]. O SOM define um mapeamento de um espaço p -dimensional contínuo para um conjunto finito de vetores referência, dispostos na forma de um arranjo espacial regular, normalmente bidimensional. O objetivo principal do treinamento é reduzir a dimensionalidade dos sinais ao mesmo tempo em que busca-se preservar ao máximo a topologia do espaço de entrada.

Uma característica importante do SOM, derivada da quantização vetorial gerada pelo algoritmo, é que densidade dos neurônios em um mapa treinado é uma aproximação da densidade dos dados. Assim, é possível obter informações dos agrupamentos analisando as relações geométricas dos neurônios após o treinamento.

O uso do SOM como ferramenta de mineração de dados tem sido abordada tanto em pesquisa como em produtos tais como o WEBSOM [11] e Clementine [12]. Porém, ambos os aplicativos efetuam agrupamentos nos dados a partir de visualização, isto é, há a necessidade de intervenção do usuário que guia manualmente a escolha dos parâmetros e a segmentação da rede. Flexer [13], após várias simulações usando conjuntos de dados com estrutura conhecida, descreve que a ferramenta Clementine sempre escolhe número inadequado de agrupamentos. Em ambos os casos, Clementine e WEBSOM, o SOM foi utilizado como um instrumento de visualização para indicar tendências de agrupamentos.

Costa e Netto [6,7,10] apresentaram um método de segmentação do SOM usando técnicas de morfologia matemática sobre uma imagem de distâncias entre neurônios adjacentes. O resultado do algoritmo são regiões conectadas de neurônios rotuladas, e que definem no espaço de entrada geometrias complexas e

não paramétricas. O método foi aplicado a diversos conjuntos de dados com formatos e dimensões variadas [7, 10]. O método também foi aplicado com sucesso em segmentação de imagens multi-espectrais [19].

Em outro artigo, Costa e Netto [9] propuseram a segmentação do SOM utilizando a idéia de particionamento de grafos, através da eliminação de conexões (arestas) inconsistentes entre neurônios vizinhos. Tanto o número de agrupamentos quanto os membros das classes são determinados automaticamente pelo algoritmo. Aplicações com sucesso do algoritmo foram também obtidas por Silva *et al* [14] na área de detecção de fragmentos urbanos utilizando dados geoespaciais.

Vesanto e Alhoniemi [8] descreveram o uso de técnicas de agrupamentos hierárquicos para aglomerar neurônios, porém há muita iteração do usuário com o resultado final obtido.

Este artigo descreve a aplicação de métodos de agrupamento hierárquico em neurônios de um mapa treinado. O objetivo é agrupar neurônios semelhantes, descobrindo não só o número mas a geometria das regiões de neurônios relacionadas aos agrupamentos de dados. Esta técnica, baseada na idéia da representação distribuída dos protótipos da classe [7], permitem a descoberta de clusters complexos e não linearmente separáveis, de forma não supervisionada. A idéia é baseada em [18], no qual o autor utilizou algoritmo hierárquico com conectividade restrita, de forma a agrupar neurônios similares, no espaço de pesos, mas também vizinhos no espaço de saída. O artigo descreve o uso dos métodos baseados nos critérios de centróide e da mínima variância, porém, preveram que podem ocorrer problemas de reversão, dificultando a interpretação dos resultados. No nosso caso, utiliza-se o método das ligações simples (LS) com restrição de contigüidade. Apresenta-se também um método para estimar o número de agrupamentos existente no conjunto de dados, baseado na estabilidade das fusões versus a distância entre os agrupamentos.

O restante do artigo é organizado da seguinte forma: a seção 2 aborda brevemente análise de agrupamentos, enquanto que a seção 3 descreve o método de proposto. Resultados são apresentados na 4, e a seção 5 apresenta as conclusões e considerações finais.

2. Análise de agrupamentos

A análise de agrupamentos tem como objetivo descobrir a estrutura (ou classes) de um determinado conjunto de dados não rotulados, onde cada objeto é descrito por p variáveis. Perguntas freqüentes incluem: (1) Existem subgrupos menos heterogêneos nos dados?; (2) Quantos subgrupos de fato existem (se é que existem)?; (3) Que objetos fazem parte de cada subgrupo?; (4) Uma vez encontrado um modelo para a estrutura dos dados, poderíamos gerar regras de decisão que possibilitassem a classificação de novas amostras?; Como tratar tipos de variáveis diferentes?

Tais perguntas, aparentemente simples, tornam-se bastante complexas à medida que os valores de n e p aumentam. Considera-se neste trabalho partições *crisp*, i.e., não há interseções entre as partições obtidas.

Um dos problemas é a própria definição do que é um agrupamento. A regra básica poderia ser descrita como 'agrupe dados em subgrupos de forma que os elementos constituintes de um subgrupo sejam mais similares entre si do que qualquer outro elemento alocado para outro subgrupo' [15, 17]. Esta regra aplica-se bem quando há agrupamentos disjuntos e relativamente homogêneos, ou elipsoidais. Uma outra definição [7], utilizada neste trabalho, é que um agrupamento é uma região do espaço, contínua, que contenha uma densidade de pontos relativamente elevada, separada de outras regiões densas por regiões com baixa densidade de pontos.

Duas classes de métodos prevalecem: os métodos hierárquicos e os particionais. Os primeiros partem de uma matriz de similaridades e geram uma árvore das fusões entre os agrupamentos, utilizando uma ultramétrica. Dividem-se ainda em aglomerativos (mais comuns) e divisivos. Os métodos aglomerativos consideram no início que os n objetos são n subgrupos e por meio de uniões sucessivas, uma por vez, chega-se a um único agrupamento contendo todos os objetos no final do processo. Os métodos divisivos consideram inicialmente um agrupamento de n objetos, e por divisões sucessivas chega-se a n subgrupos, cada um contendo um único objeto. Já os métodos particionais geram apenas uma partição dos dados, sendo mais utilizados em problemas de engenharia, como em quantização vetorial [1,5,7].

2.1 Métodos Hierárquicos Aglomerativos

As técnicas aglomerativas operam sobre uma matriz de similaridades (ou dissimilaridades), D_{ij} ($i, j = 1, 2, \dots, n$), produzindo uma seqüência de partições dos dados, P^n, P^{n-1}, \dots, P^1 . A primeira, P^n , consiste de n agrupamentos contendo um elemento apenas, e a última, P^1 , consiste de um agrupamento contendo todos os n objetos [15]. O que basicamente diferencia os métodos é a escolha da ultramétrica a ser adotada. O procedimento geral pode ser descrito em poucos passos:

1. Início: Cada agrupamento C_1, C_2, \dots, C_n contém um único objeto.
2. Determine o par de agrupamentos distintos (C_i, C_j), $i \neq j$, com maior (menor) grau de similaridade (dissimilaridade).
3. Forma-se um novo agrupamento pela união dos agrupamentos C_i e C_j , i.e., $C_k = C_i \cup C_j$. Calculam-se as novas medidas de similaridade (dissimilaridade) entre o novo agrupamento C_k e todos os outros restantes (D_{ki}).

Os passos 2 e 3 são executados ($n-1$) vezes, até que todos os objetos estejam em um único agrupamento.

Pode-se desejar parar a fusão dos agrupamentos em um determinado nível, w , do dendrograma, sem que todos os objetos estejam totalmente agrupados, reduzindo o número total de passos. A cada fusão (passo 3), diminui-se o número total de agrupamentos em 1.

A fórmula (eq. 1) de Lance & Williams [16] permite a determinação das novas distâncias entre o agrupamento formado (C_k) e todos os outros l agrupamentos, D_{kl} , onde $C_k = C_i \cup C_j$ e C_l é um outro agrupamento qualquer. A eq. 1 tem a vantagem de precisar, em cada estágio da análise, apenas das informações da matriz de dissimilaridades do estágio anterior, e funciona para vários métodos aglomerativos.

$$D_{kl} = \alpha_i \cdot D_{ki} + \alpha_j \cdot D_{jk} + \beta \cdot D_{ij} + \gamma \cdot |D_{ik} - D_{jk}| \quad (1)$$

Os parâmetros α_i , α_j , β e γ definem cada um dos métodos (ver tabela 1 para alguns métodos).

Tabela 1: Especificações de sete métodos de agrupamentos hierárquicos

Método	α_i	β	γ
Ligação simples (LS)	0.5	0	-0.5
Ligação completa (LC)	0.5	0	0.5
Média de ligações (ML)	$\frac{ i }{ i + j }$	0	0
McQuitty	0.5	0	0
Mediana das ligações (MdL)	0.5	-0.25	0
Centróide	$\frac{ i }{ i + j }$	$-\frac{ i \cdot j }{(i + j)^2}$	0
Ward	$\frac{ i + k }{ i + j + k }$	$-\frac{ k }{ i + j + k }$	0

onde $|i|$ é a cardinalidade do agrupamento i . No método de vizinhos mais próximos ou ligações simples (LS), a ultra-métrica é a distância mínima de cada par de objetos formados por um objeto pertencente a C_i e outro pertencente a C_k :

$$D_{i \cup k, j} = \min\{D_{i, j}, D_{k, j}\}$$

para todos os objetos $j \neq i, k$. Uma deficiência deste método é 'efeito de cadeia', o qual possibilita a descoberta de agrupamentos alongados no espaço, mas tem a desvantagem de conectar agrupamentos que idealmente deveriam estar separados.

No caso do método das ligações completas (LC) a dissimilaridade entre dois agrupamentos, C_i e C_k , é definida como a maior das dissimilaridades dentro de cada par de objetos formados por um objeto pertencente a C_i e outro pertencente a C_k , isto é, para todos os objetos $j \neq i, k$:

$$D_{i \cup k, j} = \max\{D_{i, j}, D_{k, j}\}$$

Neste caso o par de elementos mais afastados entre os agrupamentos $C_i \cup k$ e C_j serão usados no cálculo da dissimilaridade resultante. A tendência deste algoritmo é formar vários agrupamentos de tamanhos pequenos e compactos, i.e., com grande homogeneidade. Um outro inconveniente é a sensibilidade à presença de valores discrepantes. Detalhes de outros métodos podem ser vistos em [1,7,8,15-17].

As desvantagens principais de métodos hierárquicos são: (a) fusão de agrupamentos em um determinado estágio não pode ser corrigido em estágios subsequentes; (b) em geral, eles requerem espaço de memória de ordem $O(N^2)$ e tempo de processamento $O(N^3)$, onde N é o número de registros (cardinalidade) do conjunto de dados; (c) os resultados podem ser difíceis de interpretar, especialmente para conjuntos de dados grandes; (d) para se determinar onde cortar o dendrograma (qual o número de agrupamentos ideal), ou onde parar o método, há necessidade de um segundo critério [17].

3. Descrição do Método Proposto

A premissa básica do uso de um métodos hierárquicos no agrupamento de neurônios do SOM é que deve-se respeitar as relações topológicas do espaço de saída. O algoritmo, baseado em uma ultra-métrica escolhida, como a dos vizinhos mais próximos, deve impor restrições às fusões de neurônios à conectividade escolhida para o mapa (relações de adjacência dos neurônios).

Considera-se que o treinamento do SOM tenha ocorrido com sucesso. O conjunto de dados X é representado por um conjunto de neurônios m_i , $i = 1, \dots, N$, onde cada m_i é um vetor no espaço p -dimensional de atributos e N é o número total de neurônios da rede. Considere um mapa bi-dimensional com tamanho 10×10 (fig. 1). Cada neurônio possui um endereço bidimensional, ex. (2,2) ou unidimensional, ex. 12, de acordo com a posição no espaço de saída, e tem como vizinhos $\{(1,2), (2,1), (2,3), (3,2)\}$ ou $\{(11), (2), (22), (13)\}$.

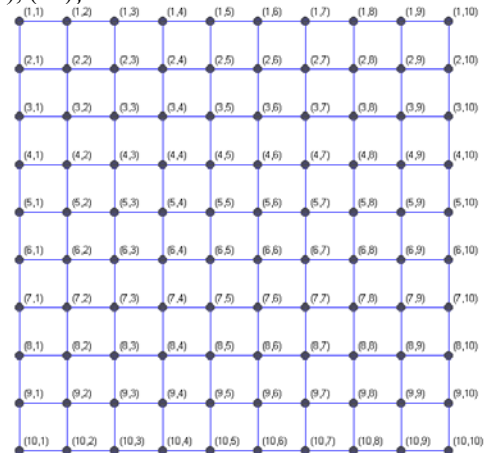


Figura 1: índices dos neurônios no espaço de saída.

A figura 2 ilustra a matriz de dissimilaridade para um mapa, contendo 100 neurônios, topologia 2-D. Como descrito na seção anterior, a ultra-métrica do método das ligações simples é a menor dissimilaridade entre os elementos constituintes dos dois agrupamentos, C_i e C_k , a todos os outros $C_j, j \neq i, k$.

Porém, em nosso caso, as distâncias permitidas para fusões são restritas à vizinhança imediata do agrupamento. A restrição de contigüidade também foi levada em consideração em métodos baseados em morfologia matemática [7]. A figura 3 ilustra as vizinhanças permitidas para o caso de adjacência 4-vizinhos, onde limitamos aos 25 primeiros neurônios da figura 2 para facilitar a visualização. Cada elemento desta matriz representa existência (em preto) ou não de uma adjacência imediata (distância = 1). O método proposto efetua fusões neurônios e agrupamentos de neurônios usando a equação 3, restrita a matriz de adjacências ilustrada na figura 3.

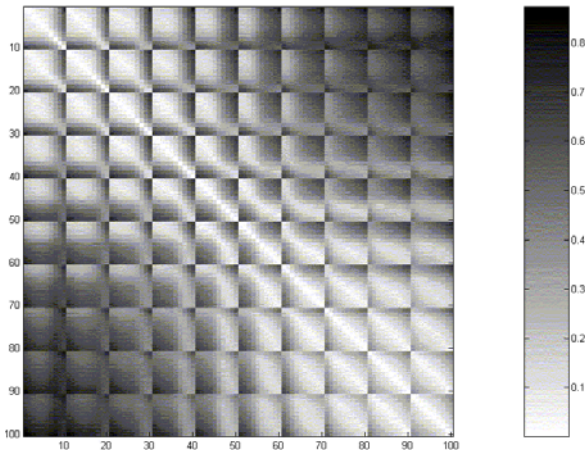


Fig. 2: matriz de dissimilaridade para um mapa contendo 100 neurônios, topologia 2-D, 10x10.

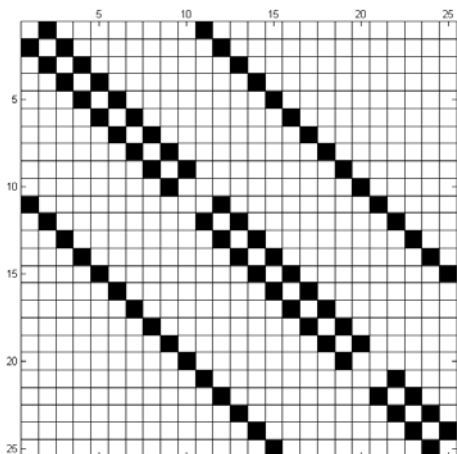


Fig. 3: Matriz de adjacências para vizinhança 4-conectado

A figura 4 ilustra o dendrograma para um mapa utilizado em um problema com três agrupamentos [7]. No eixo x estão os neurônios enquanto em y está a distância de fusão. Pode-se obter um gráfico da seqüência de fusões versus a distância destas fusões. Isto é mostrado na figura 5. Nota-se saltos no final do

gráfico, na parte direita do gráfico, entre 90 e 100, indicando uma necessidade de elevar bastante a distância (eixo y) para ocorrer uma nova fusão. Estas dissimilaridades são utilizadas para determinar o número de agrupamentos em métodos hierárquicos.

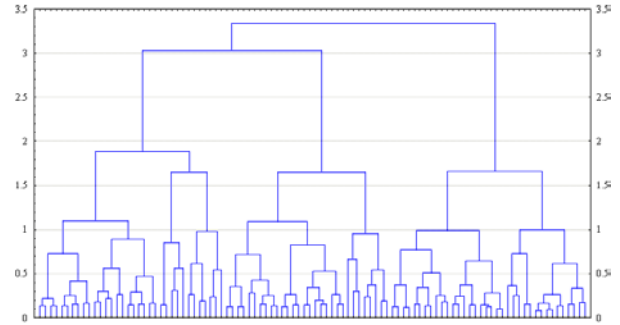


Figura 4: dendrograma para um mapa em um problema com três agrupamentos [7]

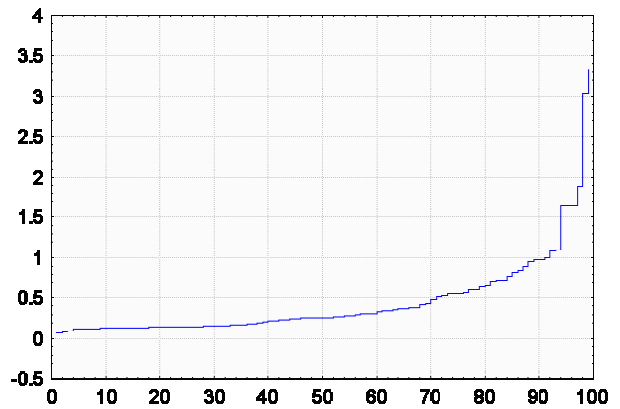


Figura 5: gráfico da seqüência de fusões (eixo x) versus a distância destas fusões (eixo y).

A seguir, apresentam-se os passos do algoritmo proposto. Dado o mapa treinado, e uma matriz de rótulos associados aos neurônios (inicialmente nula), executar os procedimentos abaixo:

Algoritmo proposto (SL-CR):

1. Enquanto existir neurônio não rotulado, executar os passos 2 a 6:
2. Encontrar a menor distância $D(i,j)$ entre dois neurônios i e j adjacentes;
3. Se os dois neurônios não possuem rótulos atribua um novo rótulo aos dois neurônios;
4. Se um dos neurônios já possuem rótulo L , o neurônio não rotulado recebe o rótulo L .
5. Se ambos os neurônios já possuem rótulos L_1 e L_2 , crie um novo rótulo L_3 e atribua aos neurônios sob rótulo L_1 e L_2 .
6. Para cada fusão, armazene N_{cl} (o número de agrupamentos que contenham pelo menos 3 neurônios) e a distância de fusão $D(i,j)$.
7. Encontre a região de maior estabilidade (número de agrupamentos versus a distância de fusão entre os neurônios).

8. A partição obtida é dada pela último passo da região de maior estabilidade.

Caso ainda existam neurônios não rotulados, e deseje rotular, usar o algoritmo de expansão de rótulo, descrito a seguir.

Algoritmo de Expansão de Rótulos:

1. Encontrar a menor distância $D(i,j)$ entre dois neurônios i e j adjacentes;
2. Se os dois neurônios não possuem rótulos atribua um novo rótulo temporário aos dois neurônios;
3. Se um dos neurônios já possuem rótulo L , o neurônio não rotulado recebe o rótulo L .
4. Se um dos agrupamentos de neurônios possui rótulo inicialmente dado L_1 , e o outro um rótulo temporário L_{2t} , o conjunto de neurônios com rótulo L_{2t} receberá os rótulos L_1 .

Basicamente o algoritmo, como o nome informa, expande o rótulo aos neurônios vizinhos à medida que aumentamos a distância, ou o raio de influência dos neurônios já rotulados. O algoritmo é similar ao SL-CR porém não é permitido a fusão entre dois agrupamentos com rótulos já estabelecidos. A fusão é permitida apenas quando um agrupamento possui rótulos inicialmente definidos e o outro possui um rótulo temporário.

4. Resultados

Para ilustrar o método, utilizou-se de um conjunto de dados com 500 pontos com duas classes oriundas de densidades Gaussianas, cada uma com 200 pontos, e 100 pontos oriundos de ruído aleatório (densidade uniforme), apresentado em [9]. A figura 6 mostra o conjunto de dados juntamente com uma grade de neurônios (com topologia 2-D e tamanho 10×10) após treinamento.

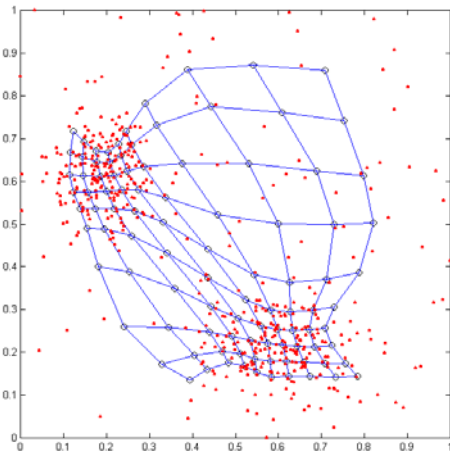


Figura 5 - Dados (pontos) e a grade de neurônios (com tamanho 10×10) após treinamento

À medida que vão ocorrendo as fusões, códigos vão sendo atribuídos aos agrupamentos. A figura 6 ilustra

um passo intermediário do algoritmo, onde a distância é 0.0404 e existem duas grandes regiões de neurônios sendo formadas.

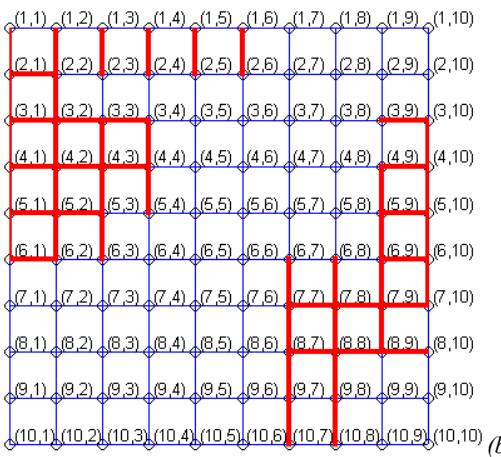
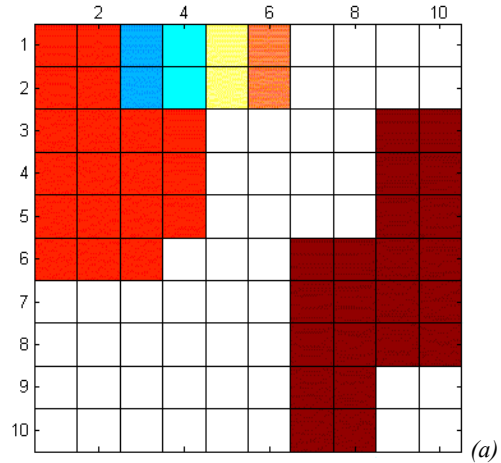


Figura 6: ilustração de um passo do algoritmo, na distância 0.0404.

A figura 7 ilustra o número de agrupamentos efetivos à medida que aumenta-se a distância nas fusões entre os neurônios. Considera-se um agrupamento efetivo, ou válido, aquele contendo pelo menos 3 neurônios agrupados. Algumas conexões, como o neurônio (1,5) com o (2,5), na figura 6, não estão sendo consideradas por estarem abaixo do limiar aceito como agrupamento, porém também recebem um código, pois no passo seguinte ele pode receber mais um neurônio (que ainda não esteja rotulado), passando, assim, a ser um agrupamento válido.

Nota-se, na figura 7, grande estabilidade para um número de agrupamentos igual a 2, que é o correto, aproximadamente entre as distâncias 0.05 e 0.107. A partição para o final da região de estabilidade é mostrada na figura 8. Note que existem alguns neurônios que não foram rotulados, e 4 que foram rotulados porém possuem menos de 3 neurônios, e não constituem agrupamentos efetivos. A figura 9 ilustra a informação das conexões entre neurônios utilizando o espaço original de entrada dos dados. Pode-se, caso queira, seguir com o algoritmo de expansão de rótulos, pela distância mínima restrita à vizinhança, e impedindo

que neurônios já rotulados em agrupamentos efetivos possam se unir.

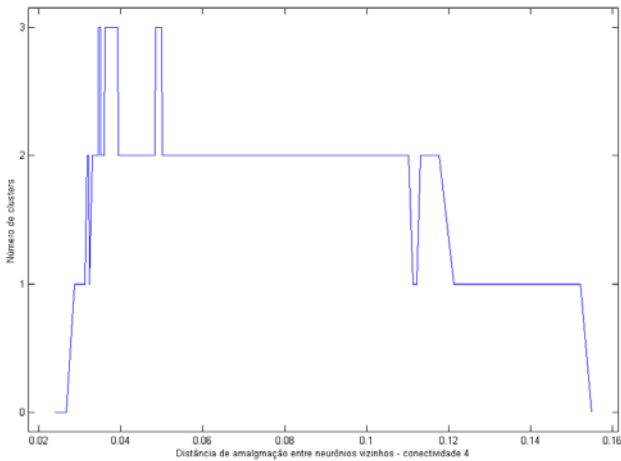


Figura 7: Número de agrupamentos efetivos à medida que aumenta-se a distância nas fusões entre os neurônios.

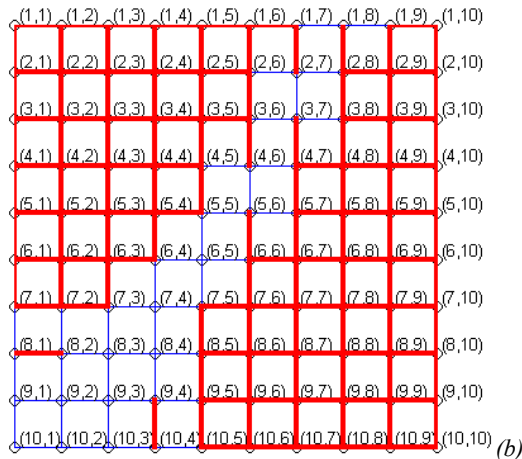
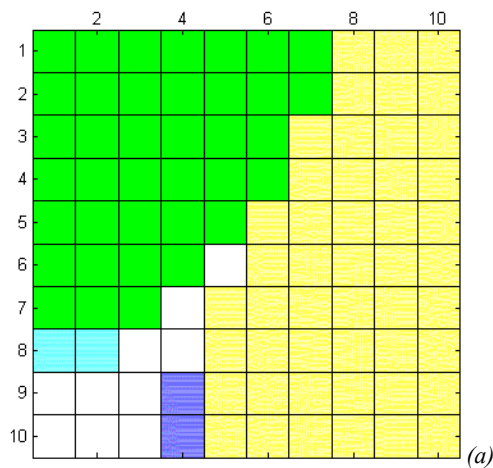


Figura 8: Partição obtida para distância de fusão entre neurônios = 0.107, final da maior região de estabilidade do gráfico apresentado na fig. 7.

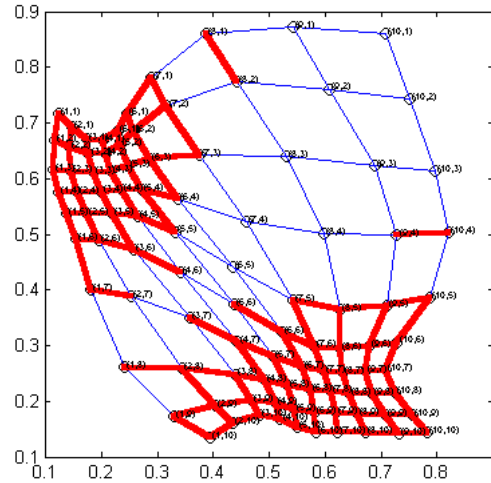


Figura 9: Informação das conexões entre neurônios utilizando o espaço original de entrada dos dados.

5. Conclusões

Apesar do extensivo uso do SOM para visualização e agrupamento de dados [1,5-14], poucos são os métodos de auxílio para análise de agrupamentos efetiva de um mapa. Abordagens para segmentação do SOM incluem o uso de técnicas baseados em morfologia matemática [3,6,7], particionamento de grafos [9], e baseados em agrupamentos de dados [8,18].

O objeto deste artigo foi a investigação e implementação de métodos de agrupamento hierárquico nos neurônios em um mapa treinado, porém com restrições de contigüidade. O objetivo é agrupar neurônios similares, no espaço de pesos, mas também vizinhos, passo a passo, no espaço de saída. Tendo não apenas um neurônio, mas um conjunto rotulado de neurônios, formando uma representação distribuída da classe, possibilita encontrar geometrias complexas [3,7], até mesmo não linearmente separáveis. Foi utilizado o método das ligações simples (LS), com conectividade restrita por contigüidade no espaço de saída, tendo apresentado resultados satisfatórios, e em conformidade com os outros métodos utilizados [3,6,7,9].

Um método, baseado na estabilidade das fusões versos a distância entre os agrupamentos, foi utilizado para estimar o número de classes existente no conjunto de dados. Vê-se na figura 9, que desconsiderando os agrupamentos não válidos [(8,1), (8,2)] e [(9,4), (10,4)], que possuem apenas 2 neurônios apenas, considerados não efetivos, o método proposto obteve o número correto de agrupamentos para o problema, 2, e os conjuntos, ou regiões rotuladas descrevem aproximadamente a geometria do espaço de entrada. O método também aplicado a outras bases de dados, como a Iris e Wine, apresentando resultados similares à segmentação do SOM por morfologia matemática [7].

Extensões do algoritmo incluem sua implementação hierárquica com particionamento recursivo de conjuntos de dados e a geração de uma árvore de mapas rotulados, de forma similar ao *Tree-Structured SL-SOM* [3,6,7] e a

adoção de critérios de validação, com o objetivo de tornar mais robustos a detecção do número estimado de classes [19].

Tree-Structured Self-Organizing Maps. *IGARSS'05 (2005 IEEE International Geoscience And Remote Sensing Symposium)*, July, 25-29 Seoul, Korea. Accepted.

Referências Bibliográficas

- [1] Murtagh, F. (2000). Clustering massive data sets, in J. Abello et al. (Eds.), *Handbook of Massive Data Sets*, Kluwer.
- [2] Cios, K. J., & Kurgan, L. (2002), Trends in Data Mining and Knowledge Discovery, In: Pal N.R., Jain, L.C. and Teoderesku, N. (Eds.), *Knowledge Discovery in Advanced Information Systems*, Springer.
- [3] Costa, J.A.F., & Netto, M. L. A., "Clustering of complex shaped data sets via Kohonen maps and mathematical morphology". In: *Proc. of the SPIE*, B. Dasarathy (Ed.), Vol. 4384, pp. 16-27, 2001.
- [4] Bishop, C.M., *Neural Networks for Pattern Recognition*, Oxford: Oxford University Press, 1995.
- [5] Kohonen, T., *Self-Organizing Maps*, 2nd Ed., Berlin: Springer, 1997.
- [6] Costa, J.A.F. and Netto, M.L.A., Estimating the Number of Clusters in Multivariate Data by Self-Organizing Maps. *Intl. Journal of Neural Systems*, vol. 9, pp. 195-202, 1999.
- [7] Costa, J. A. F., *Classificação Automática e Análise de Dados por Redes Neurais Auto-Organizáveis*. Tese de Doutorado, Unicamp, SP, 1999.
- [8] Vesanto, J. and Alhoniemi, E. (2000). Clustering of the Self-Organizing Map, *IEEE Trans. on Neural Networks*, 11, (3), pp. 586-602.
- [9] Costa, J.A.F., e Netto, M.L.A. (2003) Segmentação do SOM Baseada em Particionamento de Grafos. *Anais do VI Congresso Brasileiro de Redes Neurais*, São Paulo, Junho de 2003, pp. 451-456.
- [10] Costa, J.A.F. and Netto, M.L.A., A new tree-structured self-organizing map for data analysis. In: *Proc. of the Intl. Joint Conf. on Neural Networks*, Washington, DC, July 2001, pp. 1931-1936.
- [11] Kohonen, T. (1998). Self-Organization of Very Large Document Collections: State of the Art. In: *Proc. the 8th ICANN*, Skoevde, Sweden, pp.65-74.
- [12] *Clementine User Guide*, Integral Solutions Limited, 1998.
- [13] Flexer, A. (1997). Limitations of Self-organizing Maps for Vector Quantization and Multidimensional Scaling, in Mozer M.C., et al.(eds.), *Advances in Neural Information Processing Systems 9*, MIT Press/Bradford Books, pp.445-451.
- [14] Silva, M.A.S., Monteiro, A.M.V., and Medeiros, J.S. (2004). Semi-Automatic Geospatial Data Clustering by Self-Organizing Maps. In: *Anais do Simpósio Brasileiro de Redes Neurais*, São Luiz, MA, outubro de 2004.
- [15] Everitt, B.S., 1993. *Cluster Analysis*, 3rd Edition. Heinemann Educational, London.
- [16] Lance, G.N. & Williams, W. T. (1967). A general theory of classificatory sorting strategies: 1. Hierarchical systems. *Computer Journal*, vol. 9, pp. 373-380.
- [17] Milligan, G.W. and Cooper, M.C. (1985), An Examination of Procedures for Determining the Number of Clusters in a Data Set, *Psychometrika*, 50, 159-179.
- [18] Murtagh, F. (1995). Interpreting the Kohonen self-organizing feature map using contiguity-constrained clustering. *Pattern Recognition Letters*, v.16, pp. 399-408.
- [19] Gonçalves, M., Netto, M.L.A., Costa, J.A.F., e Zullo, J. (2005). Automatic Remotely Sensed Data Clustering by