

Segmentação do SOM baseada em particionamento de grafos

José Alfredo Ferreira Costa¹ e Márcio Luiz de Andrade Netto²

¹ Departamento de Engenharia Elétrica – Centro de Tecnologia
Universidade Federal do Rio Grande do Norte - 59.072-970 - Natal - RN

²Dept. de Engenharia de Computação e Automação - Faculdade de Engenharia Elétrica e de Computação
Universidade Estadual de Campinas - 13083-970 - Campinas - SP

E-mails: alfredo@ufrnet.br, marcio@dca.fee.unicamp.br

Abstract

Clustering methods have been studied and applied in a diversity of problems involving multidimensional data. The objective is to classify N unlabeled objects in a P -dimensional space into groups based on their similarities. Difficulties include determining the real number of categories and a metric which optimally adapt to data. Conventional methods, such as k -means, may impose a structure on data rather than finding it. This paper focuses the usage of self-organizing feature map (SOM) as a clustering tool. Although SOM had been applied as visualization tool of high-dimensional data some additional procedures are required to enable a meaningful cluster's interpretation. It is shown that the map can be partitioned by analyzing inconsistent neighboring relations between neurons. The results are sets of connected neurons that represent data clusters. The number of clusters and its membership neurons are determined by the algorithm.

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 [25] armazenados em computadores, a necessidade de métodos que possam analisá-los de forma automática, ou não supervisionada, torna-se cada vez maior.

Mineração de Dados (*data mining*) é um campo relativamente novo de pesquisa que tem sido descrita como conjuntos de técnicas para a extração de conhecimento e procura de padrões úteis, e previamente desconhecidos, em grandes volumes de dados multidimensionais [7,15]. Aplicações podem ter diferentes objetivos, como por exemplo, a determinação de classes de objetos similares. Pode-se fazer uma síntese do banco de dados observando os objetos representantes de cada subgrupo, para confirmar, ou não, hipóteses a respeito dos dados em questão. Pode-se também formular hipóteses sobre a estrutura dos

dados e determinar esquemas de classificação de novas amostras.

Dentre as principais técnicas de análise de dados, a análise de agrupamentos (ou classificação automática) 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). 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 influência sobre o resultado final do processo através da escolha de parâmetros iniciais. Além disso, a maioria dos métodos, por exemplo estatísticos como o Expectation-Maximization (EM) [1] impõem geometrias pré-estabelecidas aos agrupamentos, além de possuírem custo computacional elevado.

Redes neurais competitivas tem sido aplicadas a problemas de agrupamento de dados [2-19,21]. Dentre as redes competitivas, o SOM – *self-organizing map* – é o modelo em maior uso na atualidade.

O SOM define um mapeamento de um espaço p -dimensional contínuo para um conjunto finito de vetores referência, cada um representado por um neurônio da rede, dispostos na forma de um arranjo espacial regular, normalmente bidimensional. Todos os neurônios recebem os mesmos p sinais de entrada e o objetivo principal do treinamento é reduzir a dimensionalidade dos sinais, buscando a melhor preservação possível da topologia do espaço de entrada.

O uso do SOM como ferramenta de mineração de dados tem sido abordada tanto em pesquisa como em produtos tais como o WEBSOM [17] e Clementine [18]. 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

[19], 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 apenas como um instrumento de visualização para indicar tendências de agrupamentos. Costa e Netto [3-7] 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 [3-8,10]. Recentemente Vesanto e Alhoniemi [22] descreveram o uso de técnicas de agrupamentos hierárquicos para aglomerar neurônios, porém há muita interação do usuário com o resultado final obtido.

Este artigo descreve um método no qual o mapa é particionado eliminando conexões entre neurônios vizinhos, em mapas treinados, de forma a detectar o número e componentes de agrupamentos em um mapa de Kohonen. O espaço de saída é considerado um grafo e são definidos critérios de arestas inconsistentes. Como resultado do processamento temos conjuntos de neurônios rotulados que representam os agrupamentos dos dados. Tanto o número de agrupamentos quanto os membros das classes são determinados automaticamente pelo algoritmo.

O restante do artigo é organizado da seguinte forma: a seção 2 descreve-se brevemente o algoritmo SOM e a seção 3 descreve o método de segmentação proposto. Resultados são apresentados na 4 e a seção 5 são apresentadas as conclusões.

2. O Mapa de Kohonen

O SOM define um mapeamento de um espaço p -dimensional contínuo para um conjunto finito de vetores referência, ou neurônios, 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 [2].

Cada neurônio i é representado por um vetor de pesos $\mathbf{m}_i = [m_{i1}, m_{i2}, \dots, m_{ip}]^T$ onde p é a dimensão dos vetores de entrada. Para cada padrão de entrada um neurônio é escolhido o vencedor, c , usando o critério de maior similaridade,

$$\|\mathbf{x} - \mathbf{m}_c\| = \min_i \{\|\mathbf{x} - \mathbf{m}_i\|\} \quad (1)$$

onde $\|\cdot\|$ representa a distância Euclidiana. Os pesos do neurônio vencedor, bem como os pesos dos neurônios

compreendidos em sua vizinhança N_c , são atualizados de acordo com a equação

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h_{ci}(t)[\mathbf{x}(t) - \mathbf{m}_i(t)] \quad (2)$$

onde t indica a iteração, $\mathbf{x}(t)$ é o padrão de entrada fornecido de forma aleatória na iteração e $h_{ci}(t)$ inclui dois fatores: o núcleo de vizinhança em torno do neurônio vencedor e a constante de aprendizado, α . Ambos decaem com o tempo, de acordo com funções previamente determinadas.

Uma variação do SOM convencional é o algoritmo em lote, que o torna insensível a seqüência de apresentação dos dados em cada época. As contribuições de cada padrão são acumuladas e ao final de cada época é feita a atualização dos pesos [2].

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 [23]. Assim, é possível obter informações dos agrupamentos analisando as relações geométricas dos neurônios após o treinamento. Para efeito de visualização, Ultsch [24] descreveu a *U-matrix* que a partir é uma imagem das distâncias entre neurônios (pesos) adjacentes e o resultado é projetado no espaço de saída, correspondendo a uma imagem que pode ser plotada como superfície em 3D cuja topografia revela a configuração dos neurônios obtida pelo treinamento. Pode-se abstrair vales e montanhas, os primeiros correspondendo a regiões de neurônios similares, enquanto que montanhas refletem a dissimilaridade entre neurônios vizinhos e podem ser associadas a regiões de fronteiras de agrupamentos.

Pelo fato de geralmente a *U-matrix* ser uma imagem relativamente complexa, principalmente em problemas de análise de dados reais, seu uso em geral tem sido restrito a visualização, ou seja, uma ferramenta de auxílio na separação manual dos agrupamentos em um mapa treinado. Costa e Netto [3-5] propuseram métodos eficientes baseados em morfologia matemática para segmentar a *U-matrix* e obter informações sobre os agrupamentos de neurônios, que refletem os agrupamentos dos dados. Mesmo tendo sido aplicado com sucesso em várias bases de dados, sintéticas ou reais, nosso objetivo é ter métodos alternativos às abordagens de morfologia matemática (*watershed*) especialmente quando a dimensão do espaço de saída do mapa seja superior a 3. A seguir discute-se uma nova maneira de segmentar mapas treinados.

3. Segmentação do SOM baseada em particionamento de grafos

Esta seção ilustra como o SOM pode ser particionado, usando informações geométricas de distância entre neurônios, erro de quantização e atividade (número de vezes que o neurônio vence ou número de padrões

associados ao neurônio). A estratégia é unir informações e através de regras heurísticas eliminar ramos inconsistentes [20], i.e., ligações entre neurônios vizinhos que não satisfaçam aos pré-requisitos de um agrupamento de neurônios.

Um grafo $G(V, A)$ é um conjunto finito não vazio V e um conjunto E de pares não ordenados de elementos distintos de V . Os elementos de V são os vértices e os de E são as arestas de G , respectivamente. Duas arestas que possuem um extremo comum são chamadas de adjacentes. Grafo é uma representação gráfica das relações existentes entre elementos de dados. Para nossos propósitos, trabalharemos com grafos não direcionados (i.e., há simetria nas relações de distância dos neurônios) e a matriz de adjacências (A) é dada pela topologia do espaço de saída da rede. Os vértices são os pesos dos neurônios no espaço p -dimensional de entrada. Um grafo com n_v vértices possui matriz (Booleana) de adjacências quadrada com tamanho $n_v \times n_v$, onde cada elemento (i, j) de A indica a existência de uma relação entre os elementos i e j .

Pode-se definir um grafo de limiar (*threshold graph*), $G(V)$, como um grafo não direcionado, sem pesos, e onde as arestas entre os vértices i e j são preservadas caso a distância entre eles seja inferior a um limiar, v .

O algoritmo proposto é realizado em duas etapas distintas. A primeira é a detecção de centros de elevada ativação, que inclusive podem e têm sido utilizados como marcadores para a *watershed* (algoritmo SL-SOM [3-5]). Em seguida os componentes conectados são rotulados e há um processo opcional de expansão dos rótulos no grafo.

Algoritmo - Eliminação de arestas inconsistentes

1. Dado um mapa treinado, obtenha as distâncias entre os pesos dos neurônios adjacentes i e j , $d(w_i, w_j)$, a atividade ou número de padrões associados a cada neurônio i , $H(i)$.

2. Para cada par de neurônios adjacentes i e j , a aresta (i, j) é considerada inconsistente caso ocorra as condições:

2.1. Se a distância entre os pesos excede em 2 a distância média dos outros neurônios adjacentes a i ou a j ;

2.2. Se os dois neurônios adjacentes i e j possuem atividade (H) abaixo de 50% da mínima permitido (H_{min} , ver texto), ou um dos neurônios for inativo ($H(i) = 0$);

2.3. Se a distância entre os centróides dos conjuntos de dados associados aos neurônios i e j exceder em 2 vezes a distância entre os pesos $d(w_i, w_j)$.

3. Remoção dos ramos (arestas) inconsistentes. Para cada aresta (i, j) considerada inconsistente resultará

em uma conexão nula no endereço (i, j) da matriz de adjacência A . Ramos consistentes recebem entrada 1 no endereço (i, j) de A .

4. Atribuir um código distinto para cada conjunto de neurônios conectados. Remover componentes conectados pequenos (com menos de 3 neurônios).

O resultado prático do algoritmo é uma poda no espaço de saída do mapa, no que tange questões de vizinhança ou de adjacências. Caso existam agrupamentos esperam-se como saída grafos independentes derivados do mapa, que representam subconjuntos de neurônios sob um mesmo rótulo. Pode-se, caso necessário, expandir os rótulos dos subconjuntos sobre os neurônios e ramos não rotulados.

O motivo da existência da condição 2.1 pode ser vista olhando a figura 1. Caso a distância entre i e j seja superior a duas vezes a distância média entre os neurônios adjacentes a i e j , desconsiderando o ramo (i, j) têm-se um ramo inconsistente. O valor 2 têm sido usado após várias simulações e tem mostrado funcionar em vários casos, porém estudos estão sendo realizados para tornar este critério mais independente de um peso fixo para o limiar.

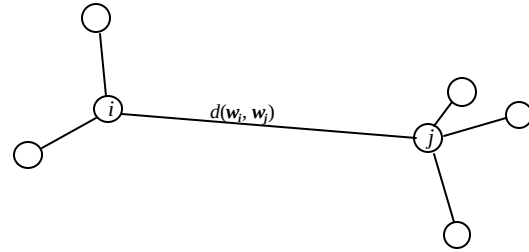


Figura 1: Ilustração de ramo inconsistente. A distância entre os pesos dos neurônios i e j excede em mais de 2 a média das distâncias dos neurônios adjacentes a i (excluindo j) ou média das distâncias dos neurônios adjacentes a j (excluindo i).

A condição 2.2 têm sido estabelecida da seguinte forma: dado um mapa e o seu conjunto de dados, obtém-se H (matriz de ativação) e o H_{med} , este último como o tamanho do conjunto de dados $|X|$ dividido pelo número total de neurônios $|n|$. H_{min} é definido como uma fração de H_{med} , i.e., $H_{min} = s * H_{med}$, onde s é um valor entre 0.1 e 0.6. Por exemplo, caso $|X| = 500$ e $|n| = 100$ temos $H_{med} = 5$. Caso s seja 0.5, teremos ativação mínima permitida (H_{min}) como 50% de H_{med} , que no nosso caso é 2.5. Para o ramo permanecer ativo ambos os neurônios i e j devem possuir atividade superior a H_{min} .

A condição 2.3 estabelece que, sendo D_i e D_j os conjuntos de dados associados aos neurônios i e j , e c_i e c_j , os respectivos centróides de D_i e D_j , a distância entre os pesos $d(w_i, w_j)$ não deve ser superior a duas vezes a distância entre os centróides, $d(c_i, c_j)$, i.e.:

Se $d(w_i, w_j) > 2 * d(c_i, c_j) \rightarrow$ ramo (i, j) é inconsistente

Esta condição, apesar de empírica, está relacionada a questões de quantização da informação. É inconsistente a ligação entre dois neurônios representando dois conjuntos de dados vizinhos (ou que façam parte do mesmo agrupamento) tenham distância superior a duas vezes a distância dos centros de massa dos conjuntos de dados.

A checagem para eliminação de arestas inconsistentes é feita para todos os pares de neurônios adjacentes, dada a topologia e padrão de conexão do espaço de saída desejados. Componentes conectados efetivos são aqueles que possuem pelo menos 3 ou mais neurônios. Os que possuem menos são eliminados, i.e., recebem código nulo, e não são considerados como agrupamentos de neurônios.

4. Resultados

Para ilustrar o método descrito na seção anterior geramos 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). A figura 2 mostra o conjunto de dados juntamente com uma grade de neurônios (com topologia bidimensional e tamanho 10×10) após treinamento. Para melhor visualização da quantização vetorial do mapa, a figura 3 ilustra o diagrama de Voronoi, a grade de neurônios e as ligações dos dados aos neurônios vencedores. A atividade (H) de cada neurônio é mostrada na figura 4 como uma barra vertical, e a figura 5 apresenta as distâncias entre neurônios adjacentes, também como barras na vertical, sobrepostas à configuração da grade de neurônios. Olhando as figuras 3 a 5, percebe-se que as regiões com baixa atividade possuem também, em geral, distâncias relativas maiores entre neurônios adjacentes. Porém pode existir neurônios dentro de regiões densas que possuam atividade nula, como é o caso do neurônio (2,3) (ver figura 3), que possui célula de Voronoi sem padrão associado. Esse tipo de problema pode ocorrer principalmente quando o conjunto de dados tem tamanho relativamente pequeno, e a dinâmica de criação do mapa pode alocar um, ou mais neurônios, em uma posição que satisfaça esta condição ($H(i) = 0$).

A figura 6 ilustra os componentes conectados (linhas mais espessas) após aplicação das condições 2.1 a 2.3 sobre o mapa apresentado na figura 2. Note que há dois componentes com 2 neurônios apenas, e no agrupamento na parte superior esquerda há um neurônio sem qualquer ligação aos vizinhos, devido a sua atividade, $H(i)$ ser nula. O parâmetro s usado foi 0.3 (ver seção 3, condição 2.2).

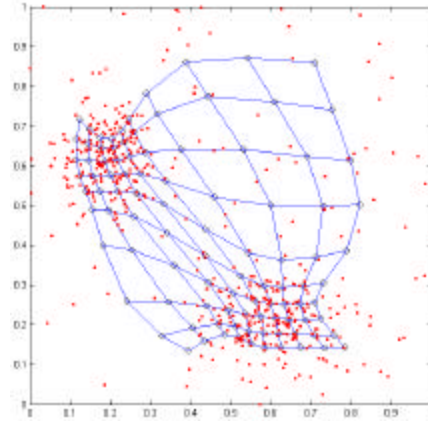


Figura 2 - Dados (pontos) e a grade de neurônios (com tamanho 10×10) após treinamento

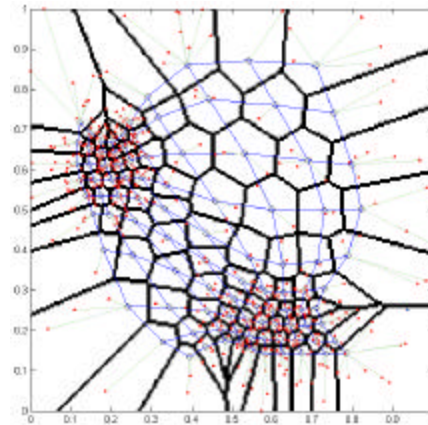


Figura 3 - Grade de neurônios, diagrama de Voronoi e dados ligados aos neurônios vencedores

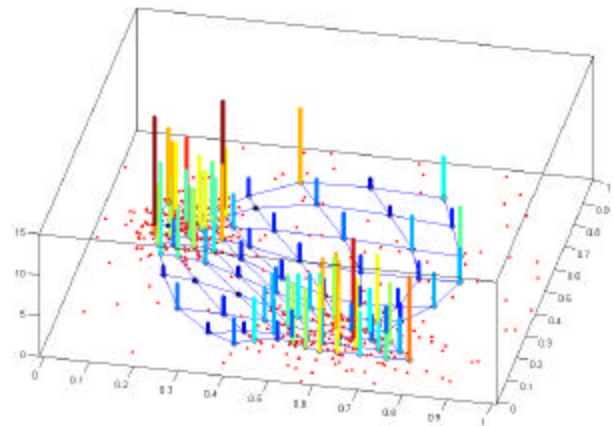


Figura 4 - Atividade (H) de cada neurônio (barras verticais) sobrepostas à configuração obtida após treinamento.

A figura 7 ilustra o mapa após eliminação de componentes conectados menores que 3 neurônios, e a figura 8 ilustra como ficou na grade de neurônios. Uma outra visão, mostrada na figura 9, temos que os

agrupamentos de neurônios aproximam a densidade real das classes existentes (agrupamentos) de dados.

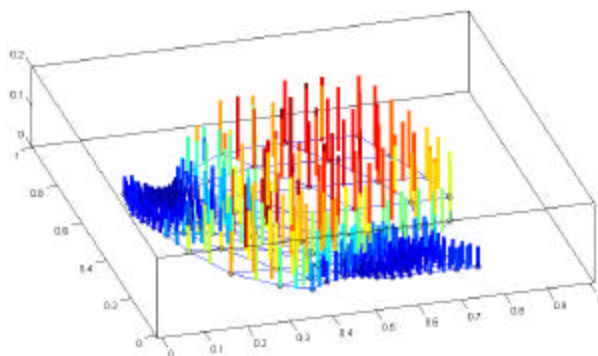


Figura 5 - distâncias entre neurônios adjacentes (barras na vertical) sobrepostas à configuração da grade de neurônios obtida após treinamento.

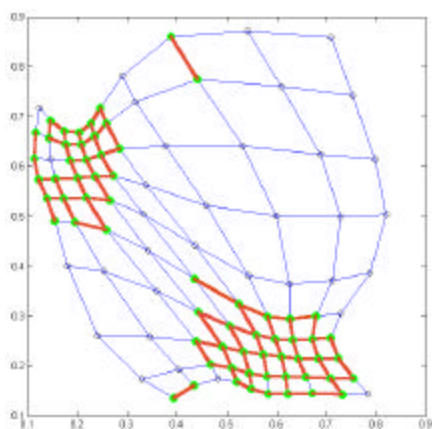


Figura 6 - Componentes conectados (linhas mais espessas) após aplicação das condições 2.1 a 2.3 sobre o mapa apresentado na figura 2.

5. Conclusões

Apesar do extensivo uso do SOM para visualização e agrupamento de dados, poucos são os métodos de auxílio para análise efetiva de um mapa. Abordagens para segmentação do SOM incluem o uso de técnicas baseadas em morfologia matemática [3-5].

Este trabalho apresentou uma nova maneira de particionar o mapa, de forma a obter informações como o número de agrupamentos e as classes de cada padrão. Há de se ressaltar a diferença do uso do mapa de Kohonen para agrupamentos neste tipo de abordagem da tradicional. É comum considerar um neurônio como representante exclusivo de um agrupamento e isso implica em considerarmos que a geometria da classe seja hiper-esférica, o que significa uma imposição de estrutura aos dados, ao invés da descoberta de tal estrutura.

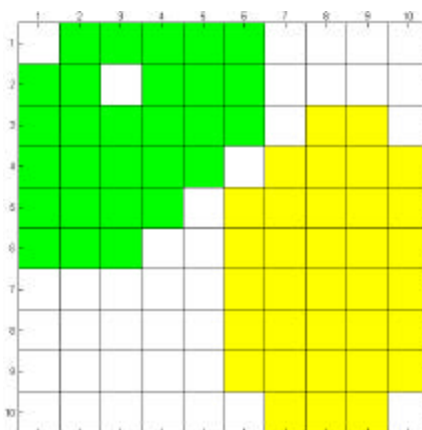


Figura 7 - Componentes conectados após eliminação de componentes menores que 3 neurônios.

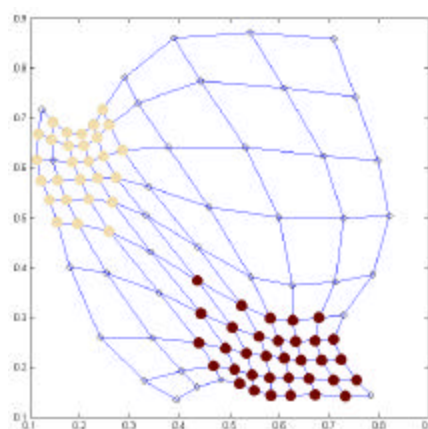


Figura 8 - Grade de neurônios com rótulos ilustrando os agrupamentos - componentes conectados após eliminação de componentes menores que 3 neurônios.

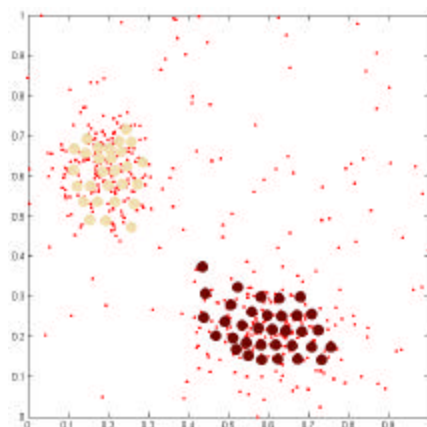


Figura 9 - Dados (pontos pequenos) e os agrupamentos de neurônios (rotulados).

O uso da representação distribuída da classe, i.e., usando vários neurônios, possibilita a detecção de

formas variadas, inclusive no mesmo conjunto de dados [7]. No exemplo mostrado neste artigo vê-se na figura 9 que os neurônios rotulados sob mesma cor aproximam a densidade dos agrupamentos reais dos dados, e importante: a influência do ruído, que não constituía um agrupamento efetivo, foi eliminada do mapa pelo algoritmo, e o número de agrupamentos correto foi determinado.

O algoritmo também foi aplicado a outros conjuntos de dados, como as bases de dados Iris e Wine, apresentando os resultados esperados e similares à segmentação do SOM por morfologia matemática [3-9], o que o torna mais uma ferramenta de análise de dados e descoberta automática de informações em bases de dados usando redes neurais.

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* [3,6-8], e o uso de novas regras com uso de vizinhança mútua e métodos hierárquicos com conectividade restrita.

Referências Bibliográficas

- [1] Bishop, C.M., *Neural Networks for Pattern Recognition*, Oxford: Oxford University Press, 1995.
- [2] Kohonen, T., *Self-Organizing Maps*, 2nd Edition, Berlin: Springer Verlag, 1997.
- [3] 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.
- [4] Costa, J. A. F., *Classificação Automática e Análise de Dados por Redes Neurais Auto-Organizáveis*. Tese de Doutorado, Unicamp, Campinas, SP, 1999.
- [5] Costa, J.A.F., & Netto, M. L. A., "Cluster Analysis Using Self-Organizing Maps and Image Processing Techniques". IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics, vol. 5, pp. 367-372, Tokyo, Japan.
- [6] Costa, J.A.F., & Netto, M. L. A., "Automatic Data Classification by a Hierarchy of Self-Organizing Maps". IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics, vol. 5, pp. 419-24, Tokyo, Japan.
- [7] Costa, J.A.F., & Netto, M. L. A., "Clustering of complex shaped data sets via Kohonen maps and mathematical morphology". In: *Proceedings of the SPIE, Data Mining and Knowledge Discovery*. B. Dasarathy (Ed.), Vol. 4384, pp. 16-27, 2001.
- [8] 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.
- [9] Costa, J.A.F., Barbalho, J., Dória Neto, A.D., e Netto, M.L.A., SOM Hierárquico aplicado a compressão de imagens. In: *Proc. Congresso Brasileiro de Redes Neurais*, Rio de Janeiro, abril de 2001.
- [10] Costa, J.A.F., e Netto, M.L.A., Segmentação automática de mapas de Kohonen. In: *Anais do CBA 2002 - Congresso Brasileiro de Automática*, Natal, RN, Setembro, 2002.
- [11] Barbalho, J., Dória Neto, A.D., Costa, J.A.F. and Netto, M.L.A., Hierarchical SOM applied to Image Compression. In: *Proc. of the Intl. Joint Conf. on Neural Networks*, Washington, DC, July 2001, pp. 442-447.
- [12] Barbalho, J., Costa, J.A.F., Dória Neto, A.D., e Netto, M.L.A., Estrutura hierárquica e dinâmica do algoritmo som aplicado à compressão de imagens. In: *Anais do CBA 2002 - Congresso Brasileiro de Automática*, Natal, RN, Setembro, 2002.
- [13] Vieira, F., Dória Neto, A.D., e Costa, J.A.F., An Efficient approach of the Salesman Travelling Problem Using Self-Organizing Maps, *Intl. Journal of Neural Systems* vol. 13, 2003, (in print).
- [14] Xavier, S., Dória Neto, A.D., Costa, J.A.F. and Netto, M.L.A., A neural hybrid system to large capacity memory association. In: *Proc. of the Intl. Joint Conf. on Neural Networks*, Washington, DC, July 2001, pp. 1174-1179.
- [15] 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.
- [16] Haykin, S., *Neural Networks A Comprehensive Foundation*, 2nd ed., Prentice-Hall, New Jersey, 1999.
- [17] Kohonen T., Self-Organization of Very Large Document Collections: State of the Art, in Niklasson L., et al.(eds.), *Proc. of the 8th International Conference on Artificial Neural Networks*, Skoevde, Sweden, September 2-4, Springer, pp.65-74, 1998.
- [18] *Clementine User Guide*, Integral Solutions Limited, 1998.
- [19] Flexer, A., On the use of self-organizing maps for clustering and visualization. in: *Proc. PKDD-99*, Prague, Springer-Verlag, 1999.
- [20] Jain, A.K. & Dubes, R.C., *Algorithms for Clustering Data*, Prentice Hall, Englewood Cliffs, N.J. 1988.
- [21] Jain, A.K., Murty, M.N. & Flynn, P.J., Data clustering: a review. *ACM Computing Surveys*, 31 (3), pp. 264 – 323, . 1999.
- [22] J. Vesanto and E. Alhoniemi, Clustering of the Self-Organizing Map, *IEEE Trans. on Neural Networks*, 11, (3), pp. 586-602, 2000.
- [23] Kaski, S., Nikkilä, J. and Kohonen, T, Methods for exploratory cluster analysis. In: *Proc. of SSGRR' 2000*, L'Aquila.
- [24] Ultsch, A., Self-Organizing Neural Networks for Visualization and Classification. In: O. Opitz et al. (Eds). *Information and Classification*. Springer, Berlin, 307-313, 1993.
- [25] F. Murtagh, Clustering massive data sets, in J. Abello, P.M. Pardalos and M.G.C. Reisende, Eds., *Handbook of Massive Data Sets*, Kluwer, 2000.