

Análise de agrupamentos usando Mapas de Kohonen segmentados por morfologia matemática e o índice de validação CDbw

Márcio L. Gonçalves¹, Márcio L. de Andrade Netto², José Alfredo F. Costa³, Jurandir Zullo Jr.⁴

¹PUC Minas Campus de Poços de Caldas - E-mail: marcio@pucpcaldas.br

²DCA/FEEC/UNICAMP – E-mail: marcio@dca.fee.unicamp.br

³DEE – Centro de Tecnologia - UFRN – E-mail: alfredo@ufrnet.br

⁴CEPAGRI/UNICAMP – E-mail: jurandir@cpa.unicamp.br

Abstract

This work presents a new strategy of automatic clusters analysis using Kohonen's Maps segmented by mathematical morphology and the clustering validity index CDbw (Composing Density Between and Within Clusters)[12]. The strategy is presented as an improvement of the method proposed in [2]. The incorporation of a clustering validity index in the data partitioning process makes possible to analyze the quality of clusters that are generated by different segmentations of a trained SOM. Experimental results show that the new strategy allows to determine in more efficient way the correct partitioning of a data set.

1. Introdução

O Mapa Auto-organizável de Kohonen (SOM – *Self-Organizing Map*) tem sido utilizado constantemente em diversas aplicações de engenharia cobrindo áreas como reconhecimento de padrões, análise de imagens e texto, quantização vetorial, regressão, problema do caixeiro viajante e mineração de dados [1-4,6,8,10,13].

Embora o SOM seja empregado largamente como uma ferramenta de visualização de dados que apresentam dimensões elevadas, a sua aplicação em tarefas de classificação automática de dados ou de análise de agrupamentos tem sido significativa. Diferentemente de métodos supervisionados de reconhecimento de padrões, nos quais um conjunto rotulado de amostras é usado para treinar o classificador, o SOM realiza um mapeamento, via aprendizado não-supervisionado, de um espaço p -dimensional contínuo para um conjunto discreto de vetores de referência, ou neurônios, dispostos na forma

de um arranjo espacial regular, normalmente bidimensional. O SOM funciona como uma rede elástica ocupando o espaço p -dimensional de forma a representar da melhor maneira as regiões com maior densidade de pontos.

Uma vez que em um SOM tradicional a única informação de saída quando se apresenta um padrão, x , são os índices (i,j) do neurônio vencedor, c , e o erro de quantização, que pode ser dado pela distância $d(x,c)$, a rede necessita de ferramentas adicionais para que uma interpretação adequada sobre o seu mapa de saída seja realizada. Diferentemente de outras aplicações, nas quais o SOM é usado como uma ferramenta de auxílio visual para indicar tendências de agrupamentos, em tarefas de análise de agrupamentos (ou classificação automática), cujo objetivo básico é descobrir a estrutura inerente dos dados, a detecção de agrupamentos através dos neurônios do SOM deve ser feita sem a intervenção de um usuário.

Na literatura podemos encontrar alguns trabalhos propondo mecanismos e procedimentos para analisar e interpretar de maneira automática os neurônios de um SOM treinado. Costa e Netto [2] apresentaram um método de segmentação do SOM aplicando 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. Vesanto e Alhoniemi [5] utilizaram métodos hierárquicos aglomerativos e o algoritmo K-médias para detectar agrupamentos de neurônios em um SOM treinado. No entanto, além do processo de descoberta dos agrupamentos exigir muita interação com o usuário, não caracterizando, portanto, um método totalmente automático, os autores utilizaram como critério para

fusão ou separação de grupos de dados, o índice Davies-Bouldin [7], que baseia-se nos centróides dos agrupamentos, não permitindo uma avaliação adequada de estruturas não hiperesféricas. Costa e Netto [3] também propuseram um método de segmentação do SOM baseado em particionamento de grafos. A estratégia consiste em eliminar ramos inconsistentes que correspondem a ligações entre neurônios vizinhos que não satisfazem pré-requisitos relacionados com informações de distância entre neurônios, erro de quantização e atividade. Silva et al. [8] aplicaram com sucesso o método de segmentação baseado em particionamento de grafos juntamente com índices de validação de agrupamentos na análise de dados geoespaciais. No trabalho de Kiang [9], o autor utilizou um método hierárquico aglomerativo empregando critérios de variância mínima para junção ou divisão de grupos de neurônios. O algoritmo necessita recalcular os centróides sempre que dois grupos são unidos e apresenta boa aplicabilidade apenas para agrupamentos com formas hiperesféricas ou hiperelipsoidais. Recentemente, Wu e Chow [14] propuseram também um método hierárquico aglomerativo para analisar a saída do SOM utilizando um índice de validação de agrupamentos. A aplicação do método em algumas bases de dados artificiais e reais mostra que o seu desempenho é superior em relação a outras técnicas de agrupamentos.

Através desses trabalhos publicados observa-se, portanto, a existência de diferentes propostas para analisar agrupamentos de dados a partir da segmentação automática ou semi-automática dos neurônios de um SOM treinado, cada uma delas com metodologias de agrupamentos diferenciadas (hierárquicas ou particionais) e com vantagens e desvantagens sob diferentes aspectos. Especificamente, pode-se observar também que a incorporação de índices de validação de agrupamentos pode ajudar no processo, certamente quando índices adequados são utilizados.

Nesse sentido, o presente trabalho apresenta um método alternativo para a análise de agrupamentos empregando a técnica de segmentação do SOM baseada em morfologia matemática, proposta por Costa e Netto [2], em conjunto com o índice de validação de agrupamentos CDbw [12], que trata de maneira eficiente agrupamentos com formatos arbitrários. O método proposto apresenta um aprimoramento da técnica de análise de agrupamentos apresentada em [2], uma vez que a incorporação de um índice de validação permite avaliar a qualidade de diferentes configurações de agrupamentos dos dados que são geradas por diferentes partições do conjunto de neurônios do SOM,

possibilitando assim determinar de forma mais eficiente aquela que corresponde a partição mais adequada.

Este artigo apresenta-se organizado da seguinte forma: a seção 2 descreve brevemente o algoritmo SOM e a seção 3 apresenta o índice de validação de agrupamentos CDbw. A seção 4 apresenta a estratégia de análise de agrupamentos proposta. A seção 5 mostra alguns resultados experimentais, e a seção 6 apresenta as conclusões obtidas.

2. Mapa Auto-organizável de Kohonen

O SOM é uma rede neural artificial que define um mapeamento de um espaço de entrada p -dimensional contínuo para um conjunto finito de vetores referências (neurônios), que são dispostos em um arranjo topológico, geralmente bidimensional [1].

A rede consiste essencialmente de duas camadas de neurônios: uma camada de entrada I e uma camada de saída U (camada de Kohonen). A entrada da rede corresponde a um vetor p -dimensional, \mathbf{x} , geralmente no espaço \mathfrak{R}^p . Todas as p componentes do vetor de entrada alimentam cada um dos neurônios do mapa. Cada neurônio i pode ser representado então por um vetor de pesos $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{ip}]^T$, também no espaço p -dimensional.

Para cada padrão de entrada um neurônio é escolhido o vencedor, c , usando o critério de maior similaridade:

$$\|\mathbf{x} - \mathbf{w}_c\| = \min_i \{\|\mathbf{x} - \mathbf{w}_i\|\}$$

onde $\|\cdot\|$ representa a distância Euclidiana. Os pesos do neurônio vencedor, juntamente com os pesos dos seus neurônios vizinhos, são ajustados de acordo com a seguinte equação:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + h_c(t)[\mathbf{x}(t) - \mathbf{w}_i(t)]$$

onde t indica a iteração do processo de treinamento, $\mathbf{x}(t)$ é o padrão de entrada e $h_c(t)$ é o núcleo de vizinhança em torno do neurônio vencedor.

Uma propriedade importante do SOM, derivada da quantização vetorial gerada pelo algoritmo, é aproximar a densidade dos dados de entrada pela densidade de neurônios em um mapa treinado. Isso possibilita obter informações dos agrupamentos analisando as relações geométricas dos neurônios. Ultsch [15] descreveu a matriz de distâncias unificadas, a U -matrix, que permite obter uma imagem das distâncias entre neurônios adjacentes. Essa imagem pode ser visualizada como uma superfície 3D, na qual "vales" correspondem a neurônios que são similares. Valores altos na U -matrix, que correspondem as

“montanhas” na imagem da superfície 3D, refletem a dissimilaridade entre neurônios vizinhos e podem ser associados a regiões de fronteiras de agrupamentos. Em Costa e Netto [2], os autores apresentaram uma técnica eficiente baseada em morfologia matemática para segmentar a *U-matrix* e obter informações sobre os agrupamentos de neurônios. Conforme será descrito na seção 4, a mesma técnica é utilizada no presente trabalho para realizar diferentes segmentações na *U-matrix*.

3. Índice de Validação de Agrupamentos

Na literatura podemos encontrar diferentes métodos para investigar a validação de agrupamentos de dados [7,12,16]. A implementação da maioria dos algoritmos de validação exige alto custo computacional, principalmente quando o número de agrupamentos e a quantidade dos dados de entrada são muito grandes. Alguns índices são dependentes dos dados e/ou do número de agrupamentos. Existem índices que utilizam os centróides de subconjuntos de dados em seus algoritmos, enquanto que outros usam todos os pontos de cada subconjunto. Quando os agrupamentos apresentam geometrias ou formatos arbitrários, o uso de um conjunto de vetores para representar cada agrupamento gera melhores resultados que a utilização de apenas um único vetor (centróide) como representante de cada grupo de dados, além de apresentar um custo computacional menor em comparação com os algoritmos que consideram todos os pontos de cada agrupamento [16].

Halkidi e Vazirgiannis [12] propuseram um índice de validação, o CDbw (*Composing Density Between and With Clusters*), que apresenta algumas características que o torna mais eficiente em relação a outros índices. O método baseia-se em dois conceitos importantes: a densidade intra-agrupamento e a separação entre os agrupamentos. A característica geométrica do agrupamento é representada pelo uso de vetores representativos, isso permite que o índice avalie de maneira correta estruturas não hiperesféricas, o que não ocorre em outros índices [7]. Além disso, a complexidade computacional do método é $O(n)$ [12], que é aceitável para grandes conjuntos de dados.

O cálculo do índice CDbw é descrito a seguir. Para um conjunto de dados particionados em c agrupamentos, define-se um conjunto de pontos representativos $V_i = \{v_{i1}, v_{i2}, \dots, v_{ir}\}$ para o agrupamento i , onde r representa o número de pontos de representação do agrupamento i . A densidade intra-agrupamento é definida por:

$$Intra(c) = \frac{1}{c} \sum_{i=1}^c \frac{1}{r_i} \sum_{j=1}^{r_i} density(v_{ij}), c > 1$$

com $density(v_{ij}) = \sum_{l=1}^{n_i} f(x_l, v_{ij})$, onde x_l pertence ao i -ésimo agrupamento, v_{ij} é a j -ésima representação do agrupamento i , e $f(x_l, v_{ij})$ é dado por 1, se $\|x_l - v_{ij}\| \leq stdev$, ou 0 caso contrário. $stdev$ corresponde a média dos desvios padrões de cada um dos c agrupamentos.

A densidade inter-agrupamento é dada por:

$$Inter(c) = \sum_{i=1}^c \sum_{j=1}^c \frac{\|close_rep(i) - close_rep(j)\|}{\|stdev(i)\| + \|stdev(j)\|} density(v_{ij}),$$

onde $c > 1$, $close_rep(i)$ e $close_rep(j)$ representam o par de pontos de representação mais próximos entre o agrupamento i e j , v_{ij} é o ponto médio entre este par de pontos, $stdev(i)$ e $stdev(j)$ representam os desvios padrões dos agrupamentos i e j , e $density(v_{ij}) = \sum_{k=1}^{n_i+n_j} f(x_k, v_{ij})$, onde x_k pertence ao agrupamento i ou j , e $f(x_k, v_{ij})$ é dado por 1 se $\|x - v\| \leq (\|stdev(i)\| + \|stdev(j)\|)$, ou 0 caso contrário.

A separação entre os agrupamentos é definida por:

$$Sep(c) = \sum_{i=1}^c \sum_{j=1}^c \frac{\|close_rep(i) - close_rep(j)\|}{1 + Inter(c)}, c > 1$$

O índice CDbw é dado então por:

$$CDbw(c) = Intra(c) * Sep(c).$$

Experimentos em [12] demonstram que uma boa partição dos dados é indicada por valores altos do índice.

No método de análise de agrupamentos proposto neste trabalho (seção 4), o índice CDbw foi utilizado para avaliar a qualidade de diferentes configurações de agrupamentos dos dados. Diferentemente da proposta apresentada em [12], que encontra os vetores de referência dos agrupamentos de forma iterativa a partir do conjunto de dados, em nossa abordagem os próprios vetores de pesos dos neurônios do SOM são utilizados como os vetores de referência dos seus respectivos agrupamentos, simplificando assim o cálculo do índice CDbw.

4. Metodologia proposta

A estratégia de análise de agrupamentos proposta neste trabalho procura encontrar a melhor partição para um conjunto de dados amostrais a partir da análise de diferentes segmentações realizadas sobre a *U-matrix* (descrita brevemente na seção 2).

Costa e Netto [2] propuseram um método eficiente baseado em morfologia matemática para realizar a segmentação da *U-matrix*. O método emprega o algoritmo de segmentação de imagens, *watershed*, utilizando uma imagem de marcadores para regularizar o processo de segmentação. Essa mesma estratégia também é aplicada em nossa proposta, porém ao invés de definirmos apenas uma imagem de marcadores para segmentar a *U-matrix* (como apresentado em [2]), várias imagens de marcadores são determinadas e, portanto, diferentes segmentações da *U-matrix* são obtidas. Cada uma dessas segmentações é associada aos neurônios do SOM, obtendo-se assim diferentes partições do mapa de neurônios que definem diferentes configurações de agrupamentos dos dados. Para cada uma das diferentes configurações de agrupamentos é aplicado o índice de validação CDbw, que serve como critério de decisão para a escolha da melhor partição dos dados. Aquela partição que apresenta o maior valor para o índice CDbw é considerada a melhor.

Considerando a *U-matrix* de um SOM treinado dada pela imagem U com 256 níveis de cinza, os seguintes passos são efetuados para se obter o conjunto de imagens de marcadores para U :

1. Filtragem: obter uma imagem U_1 suavizando a imagem U através da remoção de pequenas depressões com área inferior a 3 pixels;
2. Para $k = 1, \dots, f_{max}$, onde f_{max} é o nível de cinza máximo na imagem U_1 , criar as imagens binárias U_2^k correspondendo a conversões de U_1 usando k como valor de limiar;
3. Obter o número de regiões conectadas de U_2^k , para cada valor de k , N_{rc}^k ;
4. Construir o gráfico $k \times N_{rc}^k$, e através dele obter o conjunto S_k composto por todos os valores de k que correspondem a inícios de seqüências contíguas e constantes de números de regiões conectadas com tamanhos superiores a 3;
5. Obter o conjunto de todas as imagens de marcadores, $S_m = \{U_2^{k_1}, U_2^{k_2}, \dots, U_2^{k_n}\}$, onde k_1, k_2, \dots, k_n são os elementos de S_k obtidos no passo 4.

Embora o procedimento usado para encontrar os marcadores da *U-matrix* seja semelhante ao apresentado em [2], em nossa abordagem os passos 4 e 5 determinam um conjunto de imagens de marcadores da *U-matrix* considerando todas as seqüências contíguas e constantes de regiões conectadas. Em todos os experimentos foram considerados apenas as seqüências com tamanhos maiores que 3.

Sendo assim, a estratégia geral de particionamento de um conjunto de dados, a partir de um SOM treinado com sucesso, pode ser resumida como segue:

1. Obtenha a *U-matrix* a partir do SOM treinado;
2. Encontre o conjunto, $S_m = \{U_2^{k_1}, U_2^{k_2}, \dots, U_2^{k_n}\}$, de imagens de marcadores para a *U-matrix*;
3. Para cada imagem de marcadores, $U_2^{k_i}$, $i=1,2,\dots,n$, faça:
 - 3.1. Aplicar o algoritmo watershed sobre a *U-matrix*;
 - 3.2. Rotular as regiões conectadas da imagem segmentada no passo 3.1;
 - 3.3. Realizar a cópia dos rótulos obtidos no passo 3.2 para os neurônios do SOM associados a cada pixel da *U-matrix*;
 - 3.4. Rotular o conjunto de dados de entrada utilizando o SOM rotulado no passo 3.3;
 - 3.5. Aplicar o índice de validação de agrupamentos CDbw sobre o conjunto de dados rotulados obtido no passo 3.4, usando os vetores de pesos do SOM como vetores referência dos agrupamentos.
4. Escolher a partição de dados que apresenta o maior valor para o índice CDbw.

Para ilustrar a aplicabilidade da metodologia proposta resultados experimentais com dados de sensoriamento remoto são apresentados na próxima seção.

5. Resultados Experimentais

Uma imagem composta pelas bandas espectrais 3, 4 e 5 do satélite Landsat-5 TM [17], mostrando a cidade de Manaus, e o encontro dos Rios Negro e Solimões foi utilizada nos experimentos. Uma composição colorida das 3 bandas da imagem, que possui resolução espacial de 512x496 pixels, é apresentada na figura 1.



Figura 1 – Composição colorida da imagem de satélite.

A cena apresenta 4 grandes classes de cobertura terrestre: vegetação (classe 1), área urbana (classe 2),

água 1 (classe 3) e água 2 (classe 4). Uma amostragem de 2500 pixels selecionados de maneira aleatória e uniforme foi feita sobre a imagem. O gráfico da figura 2 apresenta as amostras de pixels plotadas no espaço 3D onde é possível visualizar as 4 classes de padrões. É possível notar também que as classes não apresentam fronteiras bem definidas, tornando mais difícil o processo de particionamento das mesmas.

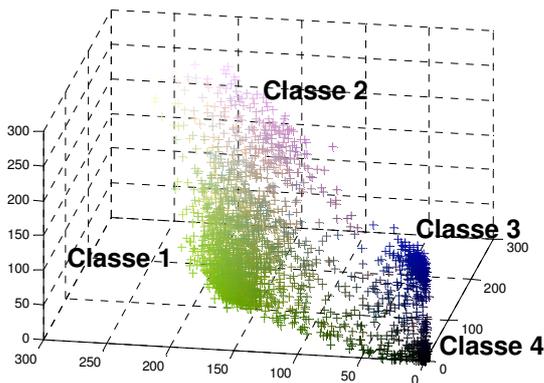


Figura 2 – Pontos amostrais da imagem.

Um SOM retangular com tamanho 15x10 foi treinado. Foram realizadas 500 épocas de treinamento e os raios de vizinhança inicial e final usados foram 12 e 1, respectivamente. A figura 3 mostra a *U-matrix* na forma de superfície obtida a partir do conjunto de neurônios do SOM treinado. O gráfico do número de regiões conectadas versus o limiar *k* da *U-matrix* é apresentado na figura 4.

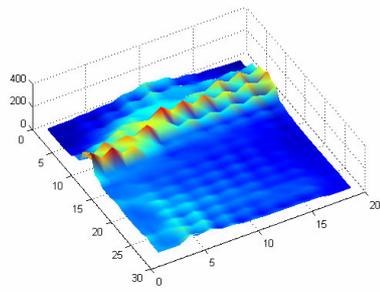


Figura 3 – *U-matrix* (3D).

Conforme pode ser observado no gráfico da figura 4, foram obtidos oito valores de *k* (7, 24, 37, 94, 107, 114, 121, 163) que correspondem a inícios de seqüências contíguas e constantes de números de regiões conectadas com tamanhos superiores a 3. Para cada um dos valores de *k* foi determinada então a imagem de marcadores da *U-matrix* e aplicada a estratégia de particionamento dos dados. No gráfico da figura 5 é possível visualizar o valor do índice CDbw que foi calculado para cada uma das partições de dados geradas a partir dos oito valores de *k*. A partição

correspondente ao valor de *k* igual a 94 foi aquela que apresentou o maior valor para o índice CDbw e foi considerada, portanto, a melhor partição. Observando o gráfico da figura 4 nota-se que o número de agrupamentos é 4 para o valor de *k* igual a 94, o que coincide com a quantidade de classes presentes na imagem.

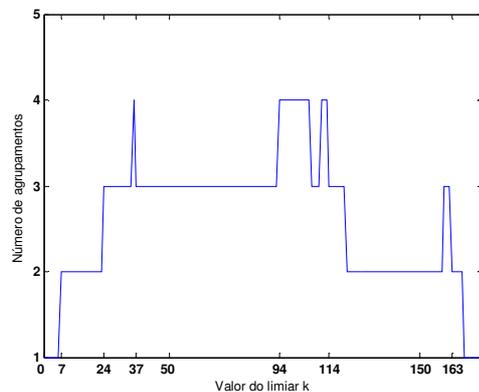


Figura 4 – Gráfico do número de regiões conectadas versus o limiar *k* da *U-matrix*.

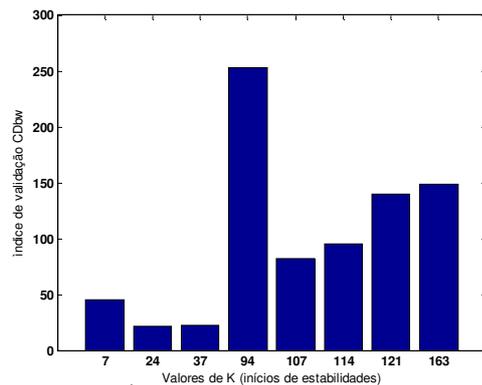


Figura 5 – Índice CDbw para as partições geradas a partir dos valores de *k* estáveis.

A figura 6 ilustra o particionamento dos pontos amostrais que foi gerado a partir do valor de *k* que apresentou o maior índice CDbw (cada cor representa um agrupamento diferente). Embora as classes de padrões apresentem estruturas não muito bem definidas, o algoritmo proposto neste trabalho, que não teve acesso a nenhuma informação sobre o número de agrupamentos e nenhuma intervenção do usuário (ou de um especialista em imagens), conseguiu de forma totalmente não-supervisionada identificá-las e particioná-las com êxito. O tempo gasto, desde o treinamento do SOM até a classificação final dos dados foi de aproximadamente 260 segundos, em uma máquina com processador AMD Athlon™ XP 2600+ 1,91GHZ e 512 de RAM.

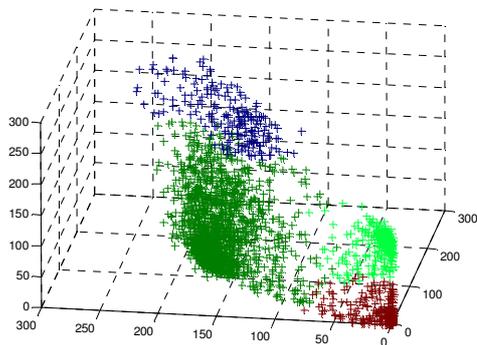


Figura 6 – Partição dos dados obtida a partir de $k = 94$.

6. Conclusões

Este trabalho apresentou uma nova estratégia de análise automática de agrupamentos usando mapas de Kohonen segmentados por morfologia matemática e o índice de validação de agrupamentos CDbw. A estratégia adotada apresenta-se como um aprimoramento do método proposto por Costa e Netto [2]. Ao invés de definirmos apenas uma imagem de marcadores para segmentar a *U-matrix* (como apresentado em [2]), várias imagens de marcadores são determinadas e, portanto, diferentes segmentações da *U-matrix* são obtidas. A estratégia apresentada considera então diferentes partições do SOM e avalia a qualidade dos agrupamentos que são gerados por eles aplicando um índice de validação adequado, podendo alcançar de maneira mais eficiente o particionamento ideal do conjunto de dados. No exemplo mostrado nesse artigo, vê-se na figura 4 que o número de agrupamentos é igual a 3 para o valor de k igual a 37, que corresponde ao início da seqüência que apresenta maior estabilidade, abordagem adotada em Costa e Netto [2]. Os dados pertencentes as classes 1 e 2 não foram particionados e fizeram parte do mesmo agrupamento, o que exige mais uma iteração do algoritmo para encontrar todas as 4 classes de padrões. Através da abordagem apresentada aqui foi possível encontrar de forma mais eficaz a partição correta do conjunto de dados (figura 6), avaliando diferentes configurações de agrupamentos que foram geradas por diferentes partições do SOM.

A estratégia de particionamento de dados foi aplicada também a outras bases de dados artificiais e reais obtendo resultados satisfatórios [11]. O bom desempenho da estratégia deve-se também a adequação do índice de validação CDbw à representação distribuída dos agrupamentos através dos neurônios do SOM, permitindo encontrar uma boa representação para agrupamentos que apresentam geometrias variadas e complexas.

7. Referências Bibliográficas

- [1] Kohonen, T., *Self-Organizing Maps*, 2nd Edition, Berlin: Springer Verlag, 1997.
- [2] Costa, J.A.F., Netto, M. L. A., “Clustering of complex shaped data sets via Kohonen maps and mathematical morphology”, *Proceedings of the SPIE, Data Mining and Knowledge Discovery*, vol. 4384, pp. 16-27, 2001.
- [3] Costa, J.A.F., Netto, M.L.A., “Segmentação do SOM Baseada em Particionamento de Grafos”, *Congresso Brasileiro de Redes Neurais*, São Paulo, 2003.
- [4] Kohonen, T., Kaski, S., Lagus, K., Salojärvi, J., Honkela, J., Paatero, V., Saarela, A., “Self Organization of a Massive Document Collection”, *IEEE Transactions on Neural Networks*, vol. 11, (3), pp. 574-585, 2000.
- [5] Vesanto, J., Alhoniemi, E., “Clustering of the Self-organizing Map”, *IEEE Transactions on Neural Networks*, v. 11, (3), pp. 586-602, 2000.
- [6] Suganthan, P. N., “Pattern Classification using multiple hierarchical overlapped self-organising maps”, *Pattern Recognition*, vol. 34, pp. 2173-2179, 2001.
- [7] Davies, D. L., Bouldin, D. W., “A Cluster separation measure”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. PAMI-1, pp. 224-227, 1979.
- [8] Silva, M. A. S., Monteiro, A. M. V., Medeiros, J. S., “Semi-Automatic Geospatial Data Clustering by Self-Organizing Maps”, *Simpósio Brasileiro de Redes Neurais*, São Luiz, MA, 2004.
- [9] Kiang, M. Y., “Extending the Kohonen self-organizing map networks for clustering analysis”, *Computational Statistics & Data Analysis*, v. 38, pp. 161-180, 2001.
- [10] Gonçalves, M. L., Netto, M. L. A., Costa, J. A. F., Zullo Jr., J., “Automatic Remotely Sensed Data Clustering by Tree-structured Self-Organizing Map”, *Proc. of IEEE International Geosciences and Remote Sensing Symposium – IGARSS’05*, Korea, 2005.
- [11] Gonçalves, M. L., Netto, M. L. A., “Uma abordagem para análise de agrupamentos baseada em Mapas de Kohonen segmentados por Morfologia Matemática e índices de validação”. TR-002/05 DCA-FEEC-UNICAMP, 2005.
- [12] Halkidi, M., Vazirgiannis, M., “Clustering validity assessment using multi representatives”, *Proceedings of SETN Conference*, Thessaloniki, Grécia, 2002.
- [13] Vieira, F.C., Dória Neto, A.D., e Costa, J.A.F., “An Efficient Approach To The Travelling Salesman Problem Using Self-Organizing Maps”, *International Journal of Neural Systems*, vol. 13, pp. 59-66, 2003.
- [14] Wu, S., Chow, T. W. S., “Clustering of the self-organizing map using a clustering validity index based on inter-cluster and intra-cluster density”, *Pattern Recognition*, v. 37, pp. 175-188, 2004.
- [15] Ultsch, A., “Self-organizing Neural Networks for Visualization and Classification”, *Information and Classification*, Berlin, Springer-Verlag, pp. 307-313, 1993.
- [16] Bezdek, J. C., Pal, N. R., “Some new indexes of cluster validity”, *IEEE Trans. System, Man, and Cybern.*, v. 28 (3), pp. 301-315, 1998.
- [17] <http://www.dgi.inpe.br/html/imagens.htm>, Divisão de Geração de Imagens – INPE, recuperado em 15/02/05.