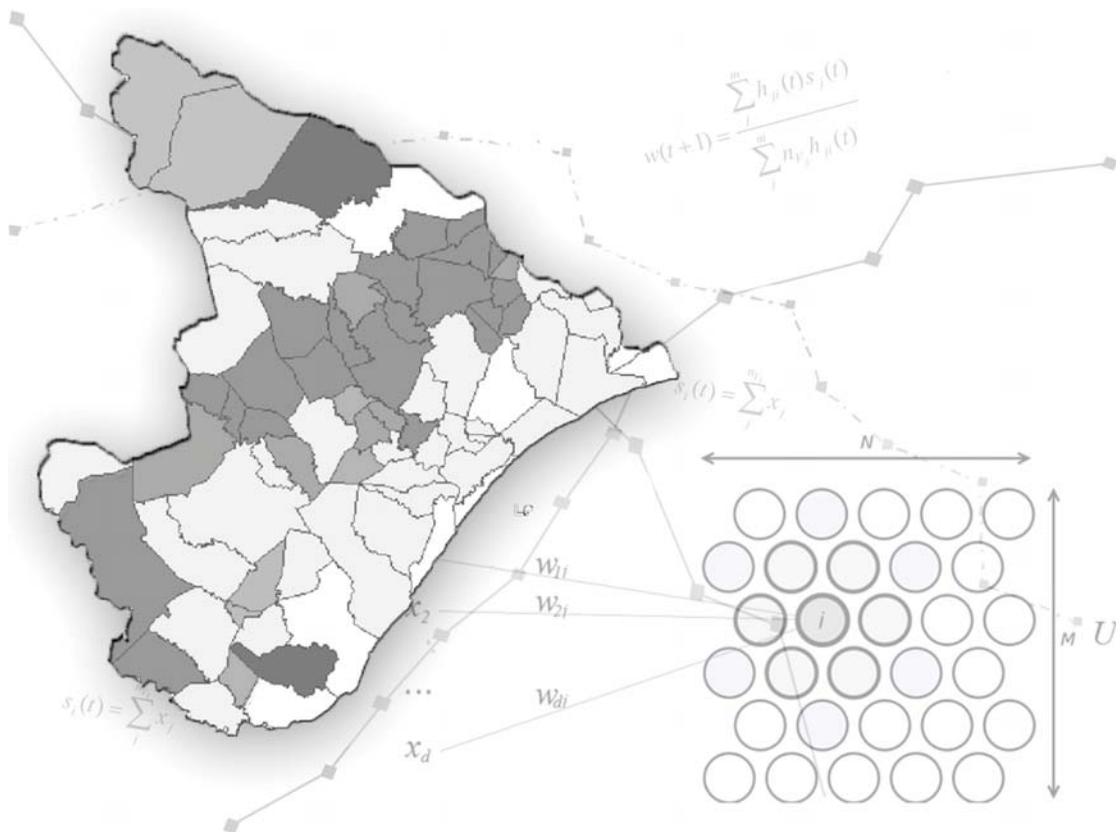


TerraSOM: Sistema para Análise de Dados Geoespaciais Agregados por Área Baseado na Rede Neural do Tipo Mapa Auto-Organizável de Kohonen





Empresa Brasileira de Pesquisa Agropecuária
Centro de Pesquisa Agropecuária dos Tabuleiros Costeiros
Ministério da Agricultura, Pecuária e Abastecimento

ISSN 1678-1961

Junho, 2015

Boletim de Pesquisa e Desenvolvimento 65

TerraSOM: Sistema para Análise de Dados Geoespaciais Agregados por Área Baseado na Rede Neural do Tipo Mapa Auto- Organizável de Kohonen

Marcos Aurélio Santos da Silva
Renato José Santos Maciel
Leonardo Nogueira Matos
Márcia Helena Galina

Aracaju, SE
2015

Embrapa Tabuleiros Costeiros

Av. Beira Mar, 3250

49025-040 Aracaju, SE

Fone: (79) 4009-1344

Fax: (79) 4009-1399

www.cpatc.embrapa.br

www.embrapa.com.br/fale-conosco

Comitê Local de Publicações da Embrapa Tabuleiros Costeiros

Presidente: *Marcelo Ferreira Fernandes*

Secretária-executiva: *Raquel Fernandes de Araújo Rodrigues*

Membros: *Ana Veruska Cruz da Silva Muniz, Élio César Guzzo, Hymerson Costa Azevedo, Josué Francisco da Silva Junior, João Gomes da Costa, Julio Roberto de Araujo Amorim, Viviane Talamini e Walane Maria Pereira de Mello Ivo*

Supervisão editorial: *Raquel Fernandes de Araújo Rodrigues*

Normalização bibliográfica: *Josete Melo Cunha*

Imagens da capa: *Marcos Aurélio Santos da Silva*

Editoração eletrônica: *Nathalie de Góis Paula*

1ª Edição (2015)

On-line (2015)

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Tabuleiros Costeiros

Silva, Marcos Aurélio Santos da

TerraSOM: sistema para análise de dados geoespaciais agregados por área baseado na rede neural do tipo mapa auto organizável de Kohonen / Marcos Aurélio Santos da Silva et al. – Aracaju : Embrapa Tabuleiros Costeiros, 2011.

36 p. il; color. (Boletim de Pesquisa / Embrapa Tabuleiros Costeiros, ISSN 1678-1961; 65).

Disponível em: <http://www.bdpa.cnptia.embrapa.br>

1. Geossistema. 2. Informação geográfica. 3. Análise de dados. I. Maciel, Renato José Santos. II. Matos, Leonardo Nogueira. III. Márcia Helena Galina. IV. Título. V. Série.

CDD 621.367 8

©Embrapa 2015

Sumário

Resumo	5
Abstract	6
Introdução	7
Mapas Auto-organizáveis	8
Aprendizagem padrão ou sequencial.....	9
Aprendizagem em lote.....	10
Avaliação da qualidade da aprendizagem.....	11
Segmentação da rede SOM baseada em partição de Grafos	12
Índices de validação dos agrupamentos	13
Índice Davies-Bouldin.....	14
Índice CDbw.....	15
Usando os vetores de código como vetores de referência no CDbw...	17
Validando o particionamento do SOM.....	17
Estudo de Caso	18

Conclusão.....	24
-----------------------	-----------

Referências.....	24
-------------------------	-----------

Anexo - Tutorial de uso do plugin TerraSOM

Primeiros passos.....	27
-----------------------	----

Conectando-se a um Banco de Dados.....	28
--	----

Criando um Banco de Dados.....	28
--------------------------------	----

Importando Dados.....	29
-----------------------	----

Inicializando o plugin TerraSOM.....	30
--------------------------------------	----

Configurando e Treinando a Rede SOM.....	31
--	----

Segmentando a Rede SOM.....	33
-----------------------------	----

Resultados.....	34
-----------------	----

TerraSOM: Sistema para Análise de Dados Geoespaciais Agregados por Área Baseado na Rede Neural do Tipo Mapa Auto-Organizável de Kohonen

*Marcos Aurélio Santos da Silva*¹

*Renato José Santos Maciel*²

*Leonardo Nogueira Matos*³

*Márcia Helena Galina*⁴

Resumo

Este trabalho apresenta o sistema de análise exploratória de dados geoespaciais agregados por área baseado na rede neural não-supervisionada do tipo Mapa Auto-organizável de Kohonen, sistema TerraSOM. O sistema explora as diversas possibilidades de partição semi-automática dos dados a partir de um conjunto de algoritmos e índices de validação dos grupos. O sistema funciona acoplado ao Sistema de Informação Geográfica TerraView, versão 3.5.0 para o sistema operacional Linux. Para demonstração do sistema, foi analisado o problema de agregação municipal no estado de Sergipe, Brasil, a partir de doze variáveis referentes ao fator uso da terra com dados relativos ao censo agropecuário de 1996.

Palavras-chave: sistema de informação geográfica, análise espacial de área, análise multivariada de agrupamentos, geocomputação.

¹ Bacharel em Ciência da Computação, M.Sc. em Computação Aplicada, pesquisador da Embrapa Tabuleiros Costeiros, Aracaju, SE.

² Bacharel em Ciência da Computação, analista da Embrapa Milho e Sorgo, Sete Lagoas, MG.

³ Bacharel em Computação, PhD. em Engenharia Elétrica, professor do Departamento de Ciência da Computação da Universidade Federal de Sergipe (UFS), São Cristóvão, SE.

⁴ Bacharel em Geografia e Tecnologia da Informática, doutora em Organização do Espaço, pesquisadora da Embrapa Tabuleiros Costeiros, Aracaju, SE.

TerraSOM: a System for Aggregated Geospatial Data Analysis Using Kohonen Self-Organizing Map neural Network

Abstract

This work presents a system for exploratory spatial analysis of geospatial data aggregated by area based on an unsupervised neural network Self-Organizing Map, TerraSOM system. This system explores data partition by a semiautomatic approach from a set of algorithms and clustering validity indexes. It works as a plugin of the Geographical Information System TerraView, version 3.5.0 for linux. In order to show how it works, it was applied on the problem of municipal aggregation or proximity analysis in Sergipe, Brazil. It was used twelve land use variables taken from the Brazilian agricultural census in 1996.

Index terms: *geographic information system, aerial spatial analysis, clustering, geocomputation.*

Introdução

A capacidade para geração, armazenamento e recuperação de dados, com referência no espaço e no tempo, cresceram muito nos últimos anos. Contribuíram para isto: a ampliação da oferta de dados de satélites em várias resoluções espaciais, espectrais e temporais; a oferta de Mapas Urbanos Básicos digitais (MUB) para diversas cidades; e a possibilidade de coleta direta de dados posicionais com o uso de sistemas GPS (Global Positioning Systems). Observa-se também certa facilidade de acesso a um conjunto bem mais amplo de dados demográficos e ambientais, como é o caso dos levantamentos censitários, realizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

As tecnologias da informação que lidam com essas bases de dados, em particular a tecnologia dos SGBDs (Sistemas Gerenciadores de Bancos de Dados) e a de Sistemas de Informação Geográfica (SIG) permitiram acomodar parte desta capacidade geradora de dados posicionais, com a possibilidade de armazenamento duradouro e com sua recuperação simples, mais eficiente e facilitada. No entanto, a nossa capacidade de analisar este conjunto de dados em várias escalas e em unidades espaciais distintas, é bem menor que a nossa capacidade de produzi-lo.

Várias décadas depois de seu surgimento, os SIGs ampliaram suas capacidades funcionais para a análise desta massa de dados geograficamente referenciados, aqui chamados dados geoespaciais. Porém, ainda estão no início da efetiva integração dos métodos de exploração e análise destes dados. Estes métodos têm surgido dentro de áreas já tradicionais, como os métodos estatísticos, assim como em áreas emergentes, como aquelas ligadas à Inteligência Artificial e Sistemas Complexos, entre outras (HEWITSON; CRANE, 1994; FISCHER; GETIS, 1996; OPENSHAW; OPENSHAW, 1997; COUCLELIS, 1998; OPENSHAW; ABRAHART, 2000; CÂMARA; MONTEIRO, 2001; KOGA, 2003).

Diante da necessidade de ampliação da capacidade funcional para análise de dados geoespaciais, a criação de ferramentas estatísticas ou computacionais para atuarem em conjunto com os SIGs é de suma importância. Estas ferramentas podem, dentre outros aspectos, aumentar o escopo de atuação do SIG e melhorar sua eficiência como instrumento de análise exploratória sobre dados multidimensionais. Neste contexto, foi desenvolvido o software TerraSOM que tem como principal objetivo avaliar agregados espaciais a partir de um algoritmo baseado nas Redes Neurais Artificiais.

Este trabalho apresenta técnicas de análise espacial de área baseadas em redes neurais e o software TerraSOM, que é um plugin desenvolvido para ser utilizado em conjunto com o TerraView. O TerraSOM implementa uma Rede Neural Artificial (RNA) não supervisionada chamada Mapa Auto-organizável de Kohonen (SOM - Self-Organized Maps). Os resultados obtidos pelo TerraSOM são exibidos de forma gráfica pelo software TerraView, o que permite uma clara visualização dos resultados. O uso do plugin será demonstrado a partir da análise exploratória de agregados municipais no estado de Sergipe, tendo como base doze variáveis relativas à utilização de terras em 31/12/2005, obtidas por meio do censo agrícola, ano 2006 (IBGE).

Mapas Auto-organizáveis

O Mapa Auto-Organizável de Kohonen é uma RNA (Rede Neural Artificial) com duas camadas (KOHONEN, 2001): a camada de entrada I e a de saída U . A entrada da rede corresponde a um vetor no espaço d -dimensional em \mathfrak{R}^d , representado por $\mathbf{x}_k = [x_1, x_2, \dots, x_d]^T$, $k = 1, 2, \dots, n$, sendo n o número de vetores de entrada. Cada neurônio j da camada de saída possui um vetor de código \mathbf{w}_j , também no espaço \mathfrak{R}^d , associado ao vetor de entrada \mathbf{x}_k , $\mathbf{w}_j = [w_{j1}, w_{j2}, \dots, w_{jd}]^T$.

Os neurônios da camada de saída estão interconectados por uma relação de vizinhança que descreve a estrutura do mapa. Por exemplo, na Figura 1, tem-se um mapa com a camada de saída U , bidimensional, retangular, de dimensões $N \times M$. Nessa figura, somente estão representados os vetores de código \mathbf{w} , conectados ao neurônio j .

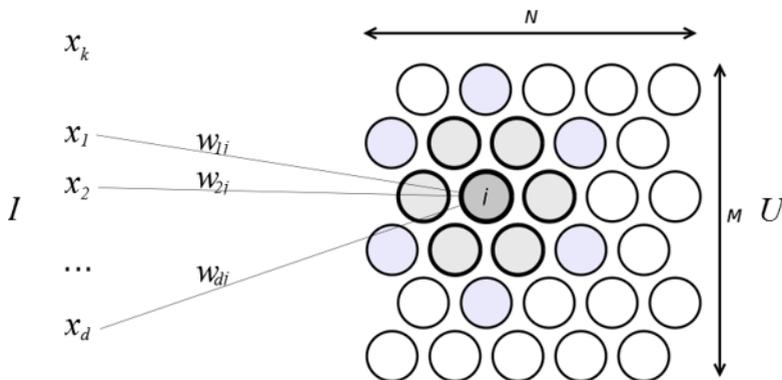


Figura 1. Exemplo de um Mapa Auto-organizável bidimensional $N \times M$, com entrada \mathbf{x}_k .

Existem diferentes topologias para estruturação de um Mapa Auto-organizável, sendo que a estrutura mais comum é a de duas dimensões. Na Figura 2, vê-se uma rede SOM unidimensional (a), uma rede bidimensional com organização hexagonal dos neurônios, com 6-vizinhos (b) e uma rede bidimensional com disposição retangular dos neurônios, com 4-vizinhos (c).

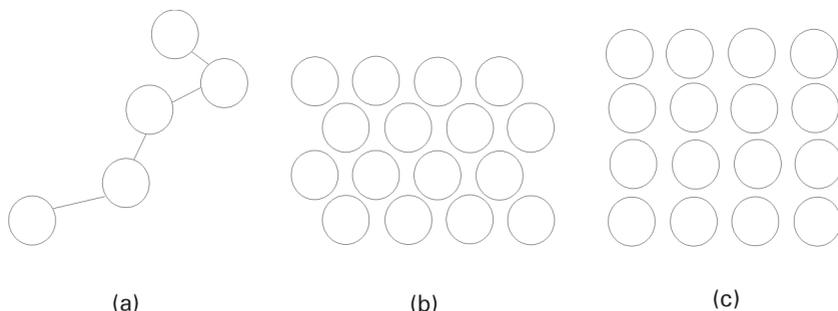


Figura 2. Diferentes topologias para estruturação dos mapas auto-organizáveis: a) unidimensional; b) bidimensional hexagonal; c) bidimensional retangular.

Desde o seu surgimento, em 1982, o SOM vem sendo aplicado numa ampla variedade de problemas de engenharia, medicina, etc. Destacam-se as potencialidades de visualização de dados multivariados, análise de agrupamentos, mineração de dados, descoberta de conhecimento e compressão de dados (KOHONEN, 2001).

Aprendizagem padrão ou sequencial

O algoritmo básico de treinamento do SOM consiste de três fases. Na primeira fase, competitiva, os neurônios da camada de saída competem entre si, denominado algum critério, geralmente a distância euclidiana, para encontrar um único vencedor, também chamado de BMU (*Best Match Unit*). Na segunda fase, cooperativa, é definida a vizinhança deste neurônio. Na última fase, adaptativa, os vetores de código do neurônio vencedor e de sua vizinhança são ajustados de forma a se aproximarem do BMU.

A relação de vizinhança entre os neurônios é estabelecida segundo uma determinada função. O principal objetivo da função de vizinhança é controlar o nível de atuação dos neurônios em torno do neurônio vencedor do processo competitivo. Seguindo o modelo neurobiológico tem-se que o nível de atuação dos neurônios vizinhos decai à medida que o mesmo se distancia do BMU.

Seja $h_{j,i}$ a vizinhança topológica centrada no neurônio i e com um conjunto de neurônios cooperativos $J, j \in J$. Seja $d_{i,j}$ a distância lateral entre o neurônio vencedor i e o neurônio j . Para que $h_{j,i}$ atenda aos requisitos neurobiológicos, a vizinhança tem que ser simétrica em relação ao ponto de valor máximo ($d_{i,j} = 0$), e $h_{j,i}$ deve decair monotonicamente com o aumento da distância lateral ($d_{i,j}$), decaindo para próximo de 0 quando $d_{i,j} \rightarrow \infty$.

A função gaussiana $h_{j,i} = \exp(-d_{i,j}^2/2\delta^2)$ satisfaz estas exigências e é invariante à translação. O termo δ representa o raio da vizinhança topológica e o grau que os neurônios vizinhos do BMU participam do processo de aprendizagem adaptativa. A rede SOM converge mais rapidamente com este tipo de função de vizinhança (LO et al., 1991; ERWIN et al., 1992; LO et al., 1993). A função $\delta(t)$ é definida como sendo monotonicamente decrescente em relação ao tempo (discreto), $\delta(t) = \delta(0)\exp(-t/\tau_i)$, sendo τ_i uma constante.

Seja Ξ o conjunto dos vetores de entrada composto por $\mathbf{x}_k, k = 1, 2, \dots, n$, tem-se o algoritmo de aprendizagem padrão ou sequencial, como segue:

a) Os vetores de código, $\mathbf{w}_j = [w_{j1}, w_{j2}, \dots, w_{jp}]^T$, são iniciados linearmente (KOHONEN, 2001, p. 142).

b) Para cada tempo discreto t

- Para todo $\mathbf{x}_k \in \Xi, k = 1, \dots, n$, para o tempo discreto t , encontre o neurônio vencedor c segundo a distância euclidiana:

$$c = \arg \min_j \{ \|\mathbf{x}_k - \mathbf{w}_j\| \}, j = 1, 2, \dots, m \quad (1)$$

onde m corresponde ao número de neurônios na rede. A ordem de apresentação dos vetores de entrada deve ser aleatória.

- Os vetores de código \mathbf{w}_j do neurônio vencedor e dos seus vizinhos são, então, atualizados segundo a equação:

$$w_{ji}(t+1) = w_{ji}(t) + \alpha(t)h(t)[x_{ik}(t) - w_{ji}(t)] \quad (2)$$

Onde $\alpha(t)$ é uma função que determina a taxa de aprendizagem na iteração t e $h(t)$ é a função que determina a vizinhança entre o neurônio vencedor c e seus vizinhos.

Aprendizagem em lote

O algoritmo de aprendizagem em lote difere da aprendizagem sequencial quanto à forma de atualização dos vetores de código, quanto à ausência da taxa de aprendizagem $\alpha(t)$ e quanto a não obrigatoriedade de apresentação aleatória dos vetores de entrada. Nesse algoritmo, os vetores de código são atualizados ao final de cada passo ou tempo discreto. Em cada passo o conjunto de dados é particionado de acordo com as regiões de Voronoi dos vetores de código do Mapa neural, definido segundo o critério de proximidade do vetor de código com o conjunto de dados. Os vetores de código w podem ser atualizados a partir das equações que seguem (VESANTO, 2000):

$$s_i(t) = \sum_j^{n_{V_i}} x_j \quad (3)$$

onde s_i representa o somatório dos vetores de entrada relativos à região de Voronoi V_i e n_{V_i} corresponde ao número de amostras do conjunto de Voronoi do neurônio i .

$$w(t+1) = \frac{\sum_j^m h_j(t) s_j(t)}{\sum_i^m n_{V_j} h_j(t)} \quad (4)$$

Avaliação da qualidade da aprendizagem

Existe um conjunto razoável de mecanismos de avaliação da qualidade do Mapa gerado após o processo de aprendizagem. Escolheu-se duas das métricas mais utilizadas, o erro da quantização vetorial e o erro topológico (KOHONEN, 2001).

O erro de quantização (E_q) corresponde à média do erro correspondente à diferença entre o vetor de características x_k e o vetor de código w_{BMU} , vetor de código vencedor no processo competitivo para o vetor x_k :

$$E_q = \frac{\sum_{k=1}^n \|x_k - w_{BMU}\|}{n} \quad (5)$$

O erro topológico (E_t) procura avaliar o quanto a estrutura da grade aproxima os vetores de entrada próximos entre si. Considerando que, para cada vetor x_k tem-se o BMU como o primeiro neurônio na ordem de competição na grade, o BMU2 corresponderá ao segundo neurônio nesta escala. Assim, o erro topológico corresponderá ao percentual de vetores cujo BMU e BMU2 não são vizinhos na grade:

$$E_t = \frac{1}{n} \sum_{k=1}^n u(x_k) \quad (6)$$

onde $u(x_k)$ corresponde a 1, se o BMU e BMU2 não são vizinhos, e 0 caso contrário.

Segmentação da rede SOM baseada em partição de Grafos

Costa e Andrade Netto (2003) propõem um método para segmentação do mapa baseado no particionamento de grafos. O algoritmo baseia-se em informações geométricas de distância entre os neurônios, no erro de quantização e no nível de atividade do neurônio. A estratégia é considerar que todos os neurônios fazem parte de um grafo não orientado, totalmente conectado e, a partir de regras heurísticas, eliminar conexões inconsistentes entre neurônios vizinhos, restando grupos conectados, representando agrupamentos distintos.

Para um Mapa Auto-organizável “treinado”, tem-se o algoritmo Costa-Netto como segue:

- a) Obter as distâncias entre os pesos dos neurônios adjacentes i e j , $d(w_i, w_j)$; e a atividade de cada neurônio i , $H(i)$.

b) Para cada par de neurônios adjacentes i e j , a aresta será considerada inconsistente se:

- A distância entre os pesos exceder em dois a distância média dos outros neurônios adjacentes a i ou a j .

- Os dois neurônios adjacentes i e j possuírem atividade (H) abaixo de 50% da mínima permitida (H_{min}), ou um dos neurônios for inativo ($H(i) = 0$); $H_{min} = wH_{med}$, sendo que $0,1 \leq w \leq 0,6$ e $H_{med} = n/m$.

- A distância entre os centróides dos conjuntos de dados associados aos neurônios i e j exceder em duas vezes a distância entre os pesos $d(\mathbf{w}_i, \mathbf{w}_j)$.

c) 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 .

d) Atribuir um código distinto para cada conjunto de neurônios conectados. Remover componentes conectados pequenos (com menos de 3 neurônios, por exemplo).

O algoritmo é independente da U-matriz e da dimensionalidade da grade do Mapa, o que o torna mais genérico que a proposta de segmentação baseada na U-matriz (COSTA, 1999). O algoritmo faz uso de alguns limiares empíricos definidos por meio de experimentações, porém, consegue particionar os dados usando as informações inerentes ao Mapa treinado, como a distância entre os neurônios, o erro de quantização e o nível de atividade.

Índices de validação dos agrupamentos

Para validar o particionamento dos dados gerados pelo algoritmo de segmentação, baseado no particionamento de grafos, usou-se dois índices, já aplicados aos Mapas Auto-Organizáveis. O índice Davies-Bouldin (DAVIES; BOULDIN, 1979), usado para auxiliar o processo de definição do número de agrupamentos corretos (VESANTO; ALHONIEMI, 2000; PARK et al., 2003) e o índice CDbw (HALKIDI; VAZIRGIANNIS, 2002) usado numa aplicação semelhante à anterior (WU; CHOW, 2004).

Índice Davies-Bouldin

O cálculo do índice Davies-Bouldin é independente do número de agrupamentos e do método de partição dos dados. O índice Davies-Bouldin é dado por:

$$\sum_{k=1}^c \max_{c \neq l} \left\{ \begin{array}{l} S_c(Q_k) + S_c(Q_l) \\ d_{ce}(Q_k, Q_l) \end{array} \right\} \quad (7)$$

onde c é o número de agrupamentos, $S_c(Q_k)$ representa a distância intra-agrupamento (dispersão interna do agrupamento Q_k), baseado na distância para o centróide, $d_{ce}(Q_k, Q_l)$ representa a distância entre os agrupamentos Q_k e Q_l , também baseado na distância entre os centróides. $S_c(Q_k)$ é dado por:

$$S_c(Q_k) = \left\{ \frac{1}{N_k} \sum_{j=1}^{N_k} |\mathbf{x}_j - \mathbf{c}_k|^q \right\}^{1/q} \quad (8)$$

onde $x_j \in Q_k$, N_k é o número de mostras no agrupamento Q_k e $\mathbf{c}_k = (1/N_k) \sum_{\mathbf{x}_i \in Q_k} \mathbf{x}_i$ d_{ce} é dado por:

$$d_{ce}(Q_k, Q_l) = \left\{ \sum_{k=1}^d |c_{ki} - c_{kj}|^p \right\}^{1/p} \quad (9)$$

onde:

d corresponde à dimensão do vetor \mathbf{x}_k .

Para $p = 2$ tem-se d_{ce} como a distância euclidiana entre os centróides dos dois agrupamentos. S_c representa a raiz q -ésima do q -ésimo momento dos pontos no agrupamento k em torno da média. Se $q = 1$, S_c torna-se a média euclidiana dos vetores no agrupamento k em relação ao centróide deste grupo. Se $q = 2$, S_c torna-se o desvio padrão da distância das amostras com relação ao centróide do grupo.

Vesanto e Alhoniemi (2000) usaram o índice Davies-Bouldin com $p = 2$ e $q = 2$ para avaliação da partição dos dados feita através do Mapa Auto-organizável. Neste caso, o SOM foi usado como um redutor do volume de dados a ser particionado. Após esta redução, dois métodos de partição tradicionais, o k-médias e o método hierárquico aglomerativo, foram aplicados, separadamente, para encontrar os agrupamentos. O índice Davies-Bouldin foi usado como critério de junção ou separação de agrupamentos nos algoritmos de partição.

Índice CDbw

O índice CDbw - *Compose Density between and within clusters* (HALKIDI; VAZIRGIANNIS, 2002; WU; CHOW, 2004) também baseia-se na medição das distâncias intra e inter-agrupamentos, porém, enfatiza as características geométricas de cada agrupamento, identificando eficientemente agrupamentos com formatos arbitrários. A característica geométrica do agrupamento é representada através do uso de vetores representativos de cada agrupamento. Ao invés de usar o centróide como referência, usa-se um conjunto de vetores. Isto permite que o índice avalie corretamente estruturas não hiperesféricas, o que não ocorre com o índice Davies-Bouldin.

Para um conjunto de dados particionados em c agrupamentos, define-se um conjunto de pontos representativos $V_i = \{v_{i1}, v_{i2}, \dots, v_{in}\}$ para o agrupamento i , onde: r_i representa o número de pontos de representação para o agrupamento i . Para cada componente ρ do agrupamento i tem-se que o desvio padrão $stdev(i)$ é dado por:

$$stdev^\rho(i) = \sqrt{\sum_{k=1}^{n_i} (\mathbf{x}_k^\rho - \bar{\mathbf{x}}_i^\rho)^2 / (n_i - 1)} \quad (10)$$

onde n_i representa o número de amostras no agrupamento i , $\mathbf{x}_k \in Q_p$ e $\bar{\mathbf{x}}_i$ ou média da amostra do i -ésimo agrupamento. A média do desvio padrão é dada por:

$$stdev = \frac{1}{c} \sqrt{\sum_{i=1}^c \|stdev(i)\|^2} \quad (11)$$

A densidade intra-agrupamento é definida como:

$$Intra_dens(c) = \frac{1}{c} \sum_{i=1}^c \frac{1}{r_i} \sum_{j=1, j \neq i}^{r_i} density(\mathbf{v}_{ij}), c > 1 \quad (12)$$

O termo $density(\mathbf{v}_{ij})$ é definido como:

$$density(v_{ij}) = \sum_{l=1}^{n_i} f(\mathbf{x}_l, \mathbf{v}_{ij})$$

onde $x_i \in Q_p$, \mathbf{v}_{ij} é a j -ésima representação do i -ésimo agrupamento, e $f(x_p, \mathbf{v}_{ij})$ é dado por 1, se $\|\mathbf{x}_i - \mathbf{v}_{ij}\| \leq stdev$, ou 0 caso contrário.

A densidade inter-agrupamento é dada por:

$$Inter_dens(c) = \sum_{i=1}^c \sum_{j=1, j \neq i}^c \frac{\|close_rep(i) - close_rep(j)\|}{\|stdev(i)\| + \|stdev(j)\|} density'(\mathbf{v}_{ij}), c > 1, c \neq n \quad (13)$$

onde: $close_rep(i)$ e $close_rep(j)$ representam o par de pontos de representação mais próximos entre o agrupamento i e o j , \mathbf{v}_{ij} é o ponto médio entre este par de pontos. A função $density(\mathbf{v}_{ij})$ é dada por:

$$density'(\mathbf{v}_{ij}) = \sum_{k=1}^{n_i+n_j} f'(\mathbf{x}_k, \mathbf{v}_{ij})$$

onde: $\mathbf{x}_k \in Q_i$ ou $r_k \in Q_j$, e $f(x_k, \mathbf{v}_{ij})$ é dado por 1 se $\|\mathbf{x}_k - \mathbf{v}_{ij}\| \leq (\|stdev(i)\| + \|stdev(j)\|)$, ou 0 caso contrário.

A separação dos agrupamentos é dada por:

$$Sep(c) = \sum_{i=1}^c \sum_{j=1, j \neq i}^c \frac{\|close_rep(i) - close_rep(j)\|}{1 + Inter_dens(c)}, c > 1 \quad (14)$$

O índice CDbw é definido por:

$$CDbw(c) = Intra_dens(c) * Sep(c) \quad (15)$$

Usando os vetores de código como vetores de referência no CDbw

Uma boa partição dos dados é indicada para valores altos do índice. A complexidade $O(n)$ é favorável para dados geoespaciais (HALKIDI; VAZIRGIANNIS, 2002). Uma questão importante a ser considerada neste algoritmo é a definição dos vetores de referência para cada agrupamento. Segundo Halkidi e Vazirgiannis (2002) este processo é iterativo. Primeiro escolhe-se o ponto mais distante da média do agrupamento, posteriormente o ponto mais distante do ponto anterior é escolhido e assim sucessivamente.

Os vetores de referência para o cálculo do CDbw podem ser encontrados de forma iterativa a partir do conjunto de dados particionado (HALKIDI; VAZIRGIANNIS, 2002). Todavia, estes autores não definem o critério de parada para o algoritmo de criação dos vetores de referência. Ou seja, o número de vetores de referência, para cada agrupamento, tem de ser definido empiricamente para servir como critério de parada, caso contrário todos os vetores poderiam ser escolhidos como vetores de referência.

Validando o particionamento do SOM

Para o caso de partição dos dados através do SOM têm-se os vetores de código como uma aproximação da distribuição dos dados de entrada sendo, portanto, vetores representativos dos dados amostrais. Logo, pode-se usar os vetores de código do SOM particionado como vetores de referência dos seus respectivos agrupamentos. Isto simplifica o processo de cálculo do CDbw para o caso de partição dos dados com o SOM.

A adequação desta abordagem dependerá da relação entre o número n de vetores de entrada e o número m de neurônios. Para m/n muito pequeno, pode-se ter uma deficiência em número de neurônios para representação de cada agrupamento. Para m/n muito grande, tem-se o inverso.

O algoritmo de detecção automática de agrupamentos baseado na partição do SOM (COSTA; ANDRADE NETTO, 2003) separa os padrões, mas não garante que todos os vetores de entrada serão rotulados. Por exemplo, dados atípicos podem não ser rotulados devido a alguma restrição do Item "b" do algoritmo de segmentação da rede SOM Costa-Netto. Esse problema pode ser solucionado usando-se o critério do vizinho mais próximo para a rotulação de todos os neurônios especializados do Mapa, evitando que o cálculo dos índices de validação sejam comprometidos.

O processo de avaliação dos agrupamentos pode ser realizado da seguinte forma:

a) Define-se um conjunto de redes que serão testadas e, para cada rede, deve-se:

- Efetuar o treinamento da rede.
- Aplicar o algoritmo Costa-Netto.
- Rotular todos os neurônios através do método do vizinho mais próximo.
- Calcular os índices Davies-Bouldin e CDbw.

b) Escolhem-se as redes candidatas com os melhores valores dos índices.

Estudo de Caso

Para uma simples demonstração da aplicação do método será analisada, de forma exploratória, a agregação municipal no estado de Sergipe a partir de variáveis relativas à utilização de terras de acordo com o censo agrícola 2006 realizado pelo IBGE. Foram consideradas as seguintes variáveis, em hectares: lavouras permanentes (LAVPERM); lavouras temporárias (LAVTEMP); lavouras, área plantada com forrageiras para corte (LAVFORR); pastagens naturais (PASTNAT); pastagens plantadas e degradadas (PASTPLANTD); pastagens plantadas e em boas condições (PASTPLANTB); matas ou florestas naturais destinadas à preservação permanente ou reserva legal (MATASLEGAL); matas ou florestas naturais excluindo as áreas de preservação permanente e as em sistemas agroflorestais (MATAS); sistemas agroflorestais (SAFS); tanques, lagos, açudes ou área de águas públicas para exploração de aquicultura (TANQUES); construções benfeitorias ou caminhos (CONSTRU); terras inaproveitáveis para agricultura ou pecuária (TERRASINAP).

As variáveis lavouras para cultivo de flores (LAVFLOR), florestas plantadas com essências florestais (MATASESS) e terras degradadas (TERRASD) foram excluídas da análise por apresentarem percentuais muito pequenos para a ocupação do solo nos municípios de Sergipe. Os valores em hectares foram transformados de forma a expressarem a fração de ocupação municipal correspondente àquela variável. Posteriormente extraiu-se a raiz cúbica de cada valor para contornar o problema de processamento computacional de valores muito pequenos. Foi atribuído o valor zero para observações ausentes ou quando não há mais de três observações para aquela variável no município. Foram definidas doze configurações de rede neural uni e bidimensional conforme ilustrado na Tabela 1.

Optou-se por realizar os experimentos a partir da aprendizagem em lote, grade hexagonal, função de vizinhança gaussiana e número de iterações de aprendizagem fixo em 1.000. Após realizar os procedimentos de treinamento da rede neural, partição da rede neural a partir do algoritmo Costa-Netto, rotulagem de todos os neurônios através do método do vizinho mais próximo e cálculo dos índices Davies-Bouldin e CDBw chegou-se aos resultados da Tabela 1.

Observando-se as curvas do erro de quantização E_q e do número de agrupamentos c , segundo a sequencial de experimentos e* em ordem do número $m=N*M$ de neurônios, tem-se que, quanto menor a rede, maior será o erro de quantização e mais reduzido será o número de agrupamentos (Figura 3).

Tabela 1. Para cada experimento E foi definido o tamanho da rede MAO ($N \times M$), onde $m = N \times M$. Também foram avaliados o número de agrupamentos c , os erros de quantização e topológicos, E_q e E_t , além dos índices de validação Davies-Bouldin e CDbw.

E	m	N	M	c	E_q	E_t	Davies-Bouldin	Davies-Bouldin (2)	CDbw
e1	5	1	5	4	0,354	0,173	5,372	2,003	11,585
e2	15	1	15	12	0,285	0,093	4,547	1,717	9,942
e3	20	1	20	16	0,281	0,133	4,164	1,574	9,487
e4	24	6	4	17	0,284	0	4,208	1,55	8,285
e5	60	10	6	39	0,246	0,04	2,465	0,893	11,787
e6	70	10	7	46	0,234	0,027	2,343	0,854	16,486
e7	80	10	8	49	0,233	0,053	2,333	0,836	7,925
e8	120	12	10	58	0,209	0,04	1,643	0,588	0
e9	150	15	10	61	0,19	0,026	1,367	0,504	0
e10	180	15	12	66	0,174	0,04	1,315	0,458	9,909
e11	200	20	10	67	0,159	0,026	1,185	0,408	11,82
e12	375	25	15	72	0,103	0,013	0,579	0,219	445,618

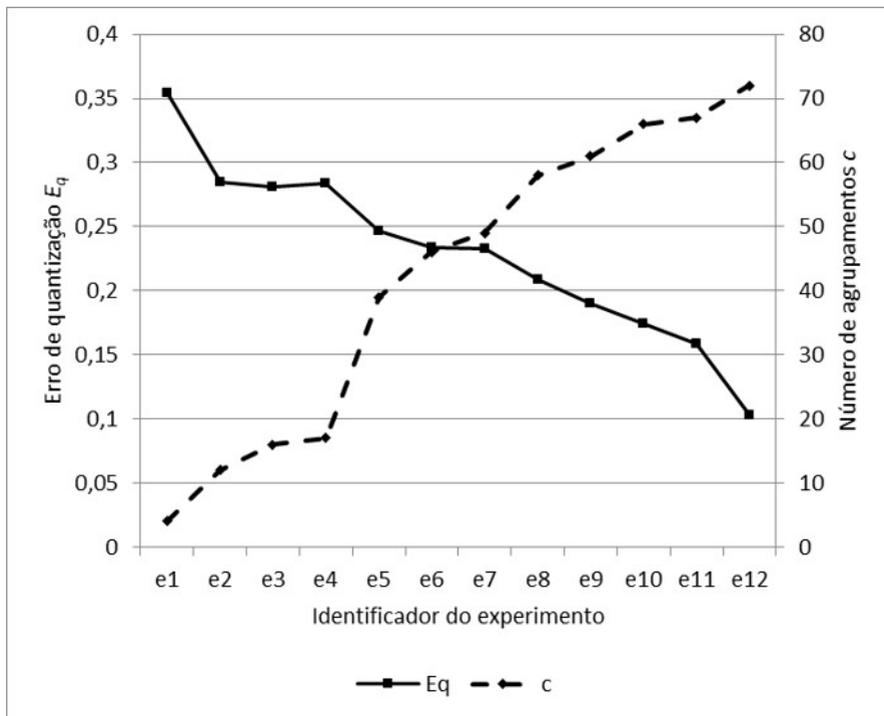


Figura 3. Curvas do erro de quantização E_q e número de agrupamentos c encontrados para cada experimento realizado. Os experimentos estão ordenados em ordem crescente do número $m = N.M$ de neurônios.

Os resultados da Tabela 1 devem ser interpretados com parcimônia. O número de agrupamentos c deve ser o primeiro parâmetro a ser observado. Cada aplicação exigirá ou limitará o número ideal de partição dos dados. Para o estudo de caso em questão, não faz sentido considerar valores de c muito altos, por exemplo, acima de vinte. Os erros de quantização, topológico e os índices de validação devem ser tomados unicamente como indicadores da qualidade da partição. O melhor valor para índice CD_{bw} ocorre para o experimento e12, no entanto o número de agrupamentos gerados é igual a 72, quase o número total de observações.

Para o estudo de caso considerou-se como ideal partições entre 10 e 20 agrupamentos. Observando a Tabela 1, constata-se que os experimentos e2, e3 e e4 geraram, respectivamente, doze, dezesseis e dezessete agrupamentos com

valores próximos para os erros de quantização, topológico e para os índices de validação. Optou-se pela análise do experimento e2 que apresenta o maior valor para o índice CDbw uma vez que estudos anteriores (SILVA et al., 2010; SILVA, 2004; SILVA; MONTEIRO; MEDEIROS, 2004a, 2004b) mostram que o índice CDbw caracteriza melhor a agregação no espaço de atributos de dados censitários. Os doze grupos foram espacializados e podem ser analisados por meio da Figura 4. Observa-se que os municípios foram equitativamente distribuídos nos grupos, sendo que o menor grupo contém quatro municípios e o maior onze.

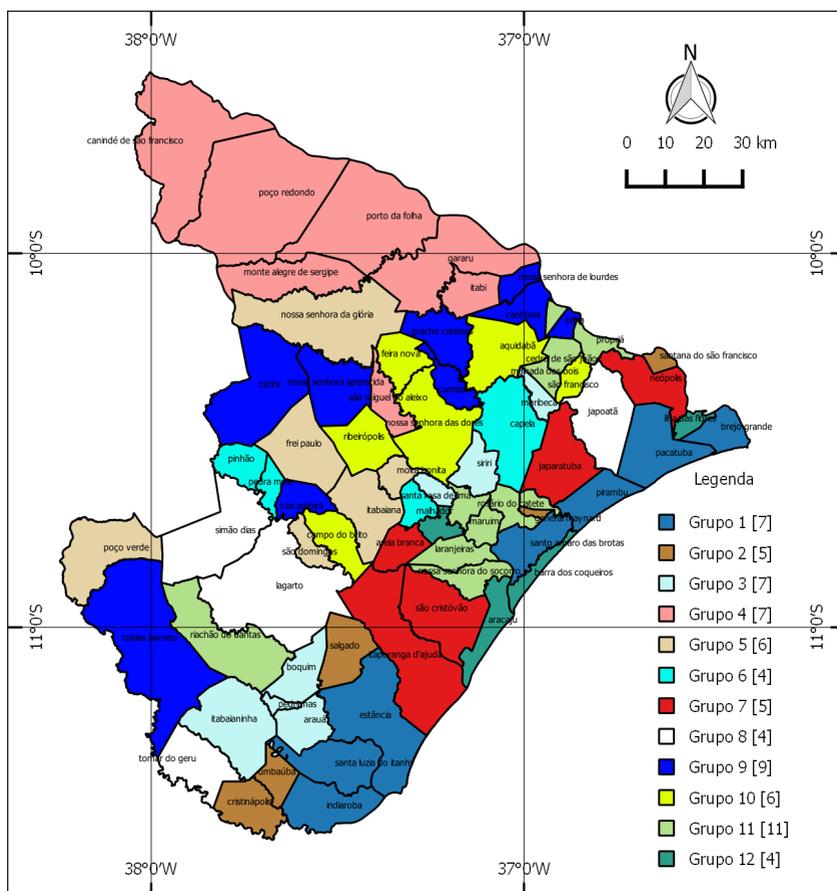


Figura 4. Mapa de agregados municipais em Sergipe gerados a partir da segmentação automática da rede MAO unidimensional (1x15).

Embora nenhuma informação explicitamente espacial tenha sido incluída na análise, os agregados municipais (Figura 4) mostram clara dependência espacial, ou seja, municípios próximos ou vizinhos apresentam perfis de ocupação do solo similares, com algumas exceções. A Figura 5 mostra os perfis de uso da terra do Brasil, do Nordeste e de Sergipe. Observa-se que o menor estado da federação destaca-se no percentual do território ocupado por lavouras permanentes, áreas plantadas com forrageiras para corte, pastagens naturais e pastagens plantadas e em boas condições.

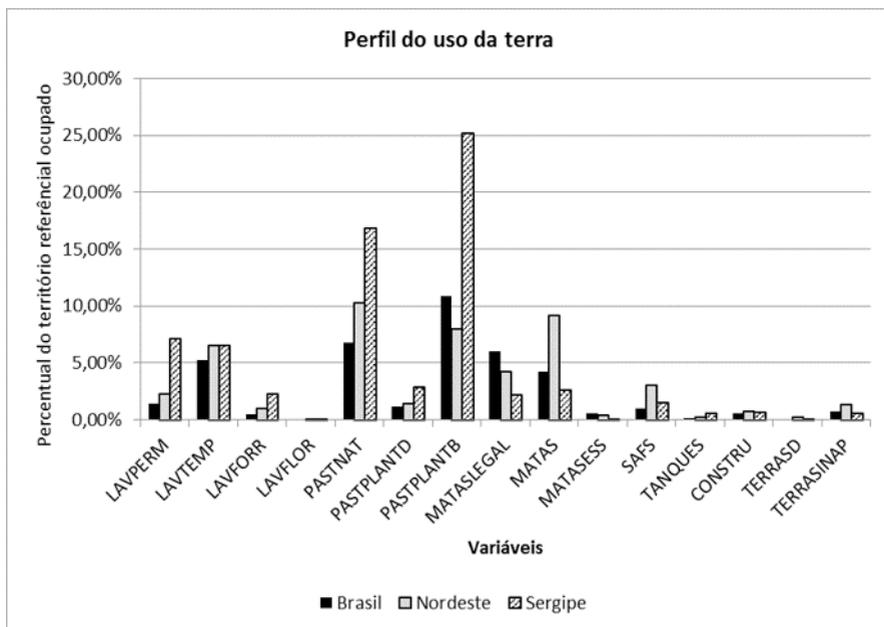


Figura 5. Perfis do uso da terra no Brasil, no Nordeste e em Sergipe.

Cada grupo gerado automaticamente pelo algoritmo pode ser caracterizado por meio da análise da média dos perfis do uso da terra dos municípios componentes. Para efeito ilustrativo tomemos como exemplo a análise dos perfis médios dos grupos um, quatro, onze e doze (Figura 6). Os grupos são formados pelos seguintes municípios: grupo 1, Brejo Grande, Estância, Indiaroba, Pacatuba, Pirambu, Santa Luzia do Itanhy e Santo Amaro das Brotas; grupo 4, Canindé de São Francisco, Gararu, Itabi, Monte Alegre de Sergipe, Poço Redondo, Porto da Folha e São Miguel do Aleixo; grupo 11, Amparo de São Francisco, Carmópolis, Cedro de São João, Divina Pastora, Laranjeiras,

Malhada dos Bois, Maruim, Nossa Senhora do Socorro, Propriá, Riachão do Dantas e Rosário do Catete; grupo 12, Aracaju, Barra dos Coqueiros, Ilha das Flores e Riachuelo.

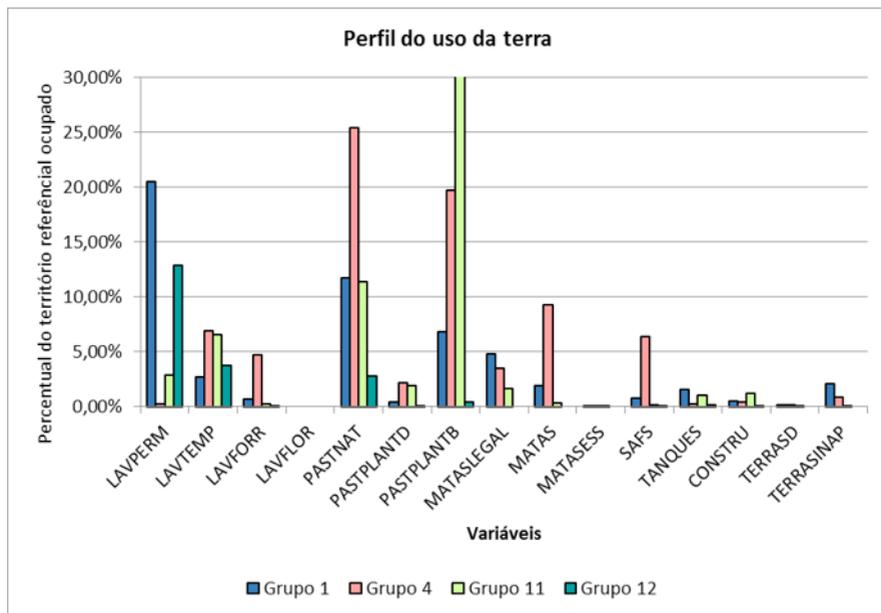


Figura 6. Perfis do uso da terra dos grupos um, quatro, onze e doze.

O grupo um é caracterizado por ter mais de 20% do seu território ocupado com lavouras permanentes e é constituído por municípios localizados na zona litorânea. O grupo quatro reúne os municípios com uma média de mais de 25% do território ocupado por pastagens naturais, este grupo apresenta forte dependência espacial, sendo que apenas o município de São Miguel do Aleixo não faz fronteira com os demais. O grupo quatro corresponde aos municípios localizados no Semiárido sergipano. Mais de 30%, em média, do território dos municípios do grupo onze é ocupado por pastagens plantadas em boas condições, e se organizam espacialmente em três grupos. O grupo doze é formado por municípios com pouca diversidade em termos de uso da terra, sendo que se destaca a lavoura permanente (12,86%) e a temporária (3,73%).

Conclusão

O método implementado por meio do plugin TerraSOM e da biblioteca de classes SOMCode permite que sejam avaliadas diversas possibilidades de partição dos dados tendo em vista a análise exploratória de dados geoespaciais agregados por área baseado na rede neural não-supervisionada do tipo Mapa Auto-Organizável de Kohonen. Como encaminhamentos futuros, sugere-se uma incrementação do método por meio do acoplamento do ajuste de pesos dos componentes no processo de segmentação, permitindo assim a informação sobre a importância de cada componente na constituição dos grupos.

Referências

CÂMARA, G.; MONTEIRO, A. Geocomputation Techniques for Spatial Analysis: Is it the Case for Health data Sets? **Caderno de Saúde Pública**, Rio de Janeiro, v. 17, n. 5, p. 1059–1081, set./out. 2001.

COSTA, J. A. F. **Classificação automática e análise de dados por redes neurais auto organizáveis**. 1999. Tese (Doutorado) – UNICAMP, Faculdade de Engenharia Elétrica e de Computação, São Paulo.

COSTA, J. A. F.; ANDRADE NETTO, M. L. Segmentação do SOM baseada em particionamento de grafos. In: CONGRESSO BRASILEIRO DE REDES NEURAIAS, 4., 2003, São Carlos. **Anais...** São Carolo, 2003.

COUCLELIS, H. Geocomputation in context. In: LONGLEY, P.; BROOKS, S.; MCDONNELL, R. et al. (Ed.). **Geocomputation, a primer**. [S.l.]: John Wiley and Sons, 1998.

DAVIES, D. L.; BOULDIN, D. W. A cluster separation measure. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. PAMI-1, p. 224–227, 1979.

ERWIN, E.; OBERMAYER, K. K.; SCHULTER, K. I. Self-organizing maps: stationary states, metastability and convergence rate. **Biological Cybernetics**, New York, v. 67, p. 35–45, 1992.

FISCHER, M.; GETIS, A. **Recent developments in spatial analysis**. New York: Springer, 1996.

HALKIDI, M.; VAZIRGIANNIS, M. Clustering validity assessment using multirepresentatives. In: SETN CONFERENCE, 2002, Thessaloniki. **Proceeds...** Greece, 2002.

HEWITSON, B.; CRANE, R. **Neural nets: applications in geography**. [S.l.]: Kluwer, 1994.

KOGA, D. **Medidas de cidades: entre territórios de vida e territórios vividos**. São Paulo: Cortez, 2003.

KOHONEN, T. **Self-organizing maps**. 3. ed. New York: Springer, 2001.

LO, Z.; FUJITA, M.; BAVARIAN, B. Analysis of neighborhood interaction in Kohonen neural networks. In: INTERNATIONAL PARALLEL PROCESSING SYMPOSIUM, 6., 1991. **Proceeds...** 1991.

LO, Z.; YU, Y.; BAVARIAN, B. Analysis of the convergence propoerties of topology preserving neural networks. **IEEE Transactions on Neural Networks**, New York, v. 4, p. 207–220, 1993.

OPENSHAW, S.; ABRAHART, S. **Geocomputation**. Londres: Taylor & Francis, 2000.

OPENSHAW, S.; OPENSHAW, C. **Artificial intelligence in geography**. San Francisco, CA: John Wiley, 1997.

PARK, Y. S.; CEREGHINO, R.; COMPIN, A. et al. Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. **Ecological Modelling**, Amsterdam, v. 160, p. 265–280, 2003.

SILVA, M. A. **Mapas Auto-organizáveis na análise exploratória de dados geoespaciais multivariados**. 2004. Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos.

SILVA, M. A. S.; MONTEIRO, A. M. V.; MEDEIROS, J. S. de. Visualization of geospatial data by component planes and u-matrix. In: Brazilian Symposium

on Geoinformatics, 6., 2004, Campos di Jordão. **Anais...** Campos do Jordão, 2004a.

SILVA, M. A. S.; MONTEIRO, A. M. V.; MEDEIROS, J. S. de. Semi-Automatic Geospatial Data Clustering by Self-Organizing Maps. In: BRAZILIAN SYMPOSIUM ON NEURAL NETWORKS, 6., 2004, São Luis. **Anais...** São Luis, 2004b.

VESANTO, J. **Using SOM in data mining**. 2000. Licentiate's thesis in the Helsinki University of Technology, Helsinki.

VESANTO, J.; ALHONIEMI, E. Clustering of the Self-Organizing Map. **IEEE Transactions on Neural Networks**, New York, v. 11, n. 3, p. 586–600, Mayo 2000.

WU, S.; CHOW, T. W. Clustering of the self-organizing map using a clustering validity index based on inter-cluster and intra-cluster density. **Pattern Recognition**, Ezmsford, NY, v. 37, p. 175–188, 2004.

Anexo - Tutorial de uso do plugin TerraSOM

Instalação do plugin

Copiar todos os arquivos contidos em <http://www.cpatc.embrapa.br/somcode/download/terrasom1.0.0.zip> para o diretório `terraView350/plug-ins`.

Primeiros passos

Inicialize o TerraView. Acompanhe na Figura 1 a tela inicial do TerraView: faça uma conexão com um Banco de Dados(BD). Clique no ícone (Database) ou Inicialize o TerraView. Acompanhe na Figura 1 a tela inicial do TerraView: faça uma conexão com um Banco de Dados(BD). Clique no ícone (Database) ou através do menu `Files::Database`. Após clicar em Database, a tela de conexão com o Banco de Dados aparecerá como mostra a Figura 2.

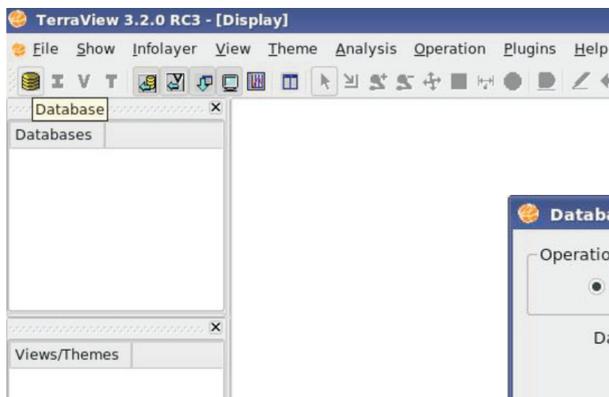


Figura 1. Inicializando o TerraView.

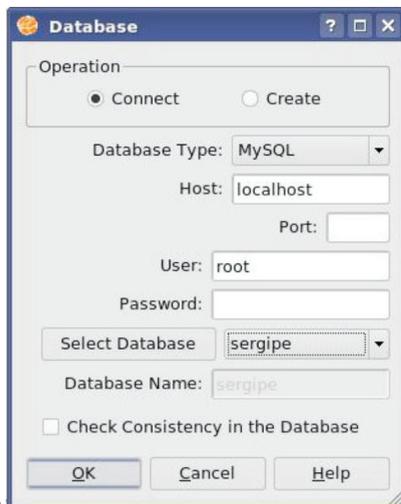


Figura 2. Conectando-se a um Banco de Dados.

Conectando-se a um banco de dados

Para se conectar ao Banco de Dados, devemos preencher algumas informações requeridas como mostra a Figura 2.

- Para se conectar a um banco já existente deve-se selecionar a opção *Connect* do menu *Operation*. Caso não haja nenhum Banco de Dados, na seção 4.3 será mostrado como criar um novo BD.

- Em *Database Type* escolhe-se o SGBD (Sistema Gerenciador de Banco de Dados) correspondente ao BD que será utilizado.

- Em *Host* deve-se indicar a localização do Banco de Dados. Neste campo deve ser colocado o endereço IP(Internet Protocol) correspondente à localização do BD. O campo referente a *Port* também é referente a localização do BD, aqui deve ser indicado a porta de acesso ao endereço indicado.

- Dados de acesso ao BD. Nos campos referentes à *User* e *Password*, deverá ser informado um nome de usuário do BD selecionado e a senha correspondente, para que assim o TerraView tenha acesso ao BD.

- Por fim, seleciona-se a base de dados que será utilizada. Clique em *Select Database*, e em seguida escolha o BD na lista ao lado.

Criando um banco de dados

Primeiro devemos clicar no ícone  (*Database*), ou seguir pelo menu *Files* → *Database*. Acompanhe a Figura 3.

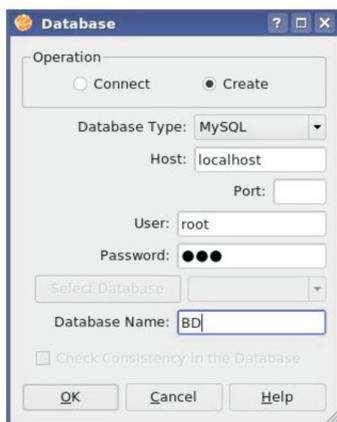


Figura 3. Criando um Banco de Dados.

- Selecione a opção *Create*, no menu *Operation*.
- Em seguida devem ser preenchidos os campos referentes à *Database Type*, *Host*, *Port*, *User* e *Password* tal como foi demonstrado na seção anterior. Lembrando que usuário e senha são os dados de acesso referente ao SGBD escolhido.
- Adiante devemos escolher o nome que será dado ao BD a ser criado. No campo referente à *Database Name*, informe o nome desejado para o BD a ser criado. Em seguida clique em *OK*. E assim, é criado um novo banco.

Importando dados

Nesta etapa, será mostrado como importar dados para o Banco de Dados.

- Primeiramente deve-se clicar em *Import*, o atalho para opção *Import* está representado pelo ícone **I**. Outra opção de acesso é através do menu *Files* → *Import*.
- Após abrir a janela de importação, selecione o arquivo que deseja importar, como mostra a Figura 4.

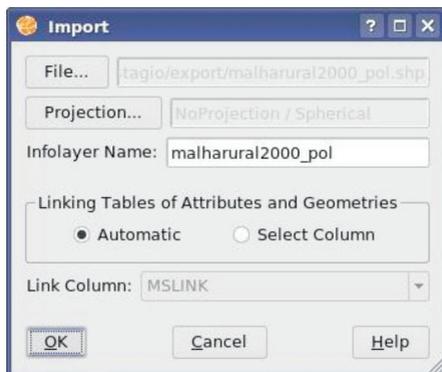


Figura 4. Importando Dados.

Os tipos de arquivos suportados são os seguintes: *Shape*, *MIF*, *SPRING-GEO* e *Atlas GIS BNA*. No exemplo apresentado aqui, o arquivo importado foi do tipo *Shape*.

- Após selecionar o arquivo desejado, clique em *OK*. Vai aparecer uma tela de confirmação, clique em *YES* para confirmar a importação do arquivo e *NO* para alterar ou voltar atrás e cancelar a importação. Após importar os dados com sucesso, os novos dados já poderão ser visualizados no TerraView.

Obs.: Antes de fazer a importação dos dados, uma base de dados já foi previamente criada.

Inicializando o plugin TerraSOM

Na barra de ferramentas superior do TerraView, inicialize o Plugin do TerraSOM.

- Clique em Plugins, e depois clique na opção TerraSOM, como mostra a Figura 5.

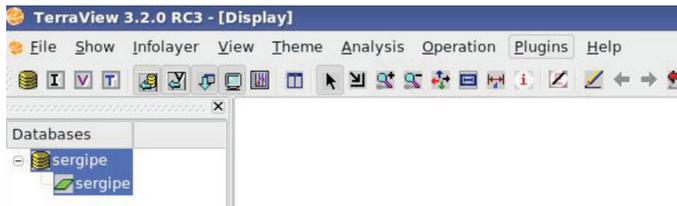


Figura 5. Plugins

- Em seguida, o plugin TerraSOM será exibido como mostra a Figura 6.



Figura 6. Plugin TerraSOM.

Para utilizar o plugin, o primeiro passo é clicar no ícone  , que representa a opção *Data Selection*, a Figura 6 indica a localização da opção. Em seguida deve-se selecionar as colunas de tabelas desejadas. Essas colunas serão o vetor de entrada da Rede Neural SOM.

- Selecione os campos da tabela desejada (Figura 7).

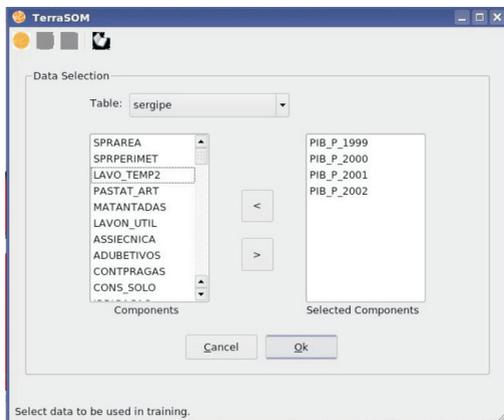


Figura 7. Seleção de dados.

Após selecioná-los, clique em *OK*.

Configurando e treinando a Rede SOM

Após selecionar as variáveis, e assim definir os dados de entrada, o próximo passo é configurar a Rede Neural SOM. Agora o botão de treinamento da rede deve estar habilitado. O ícone  representa a opção *Treinamento*.

- Clique no botão de treinamento. Aparecerá então, a tela de configuração da rede, como pode ser observado na Figura 8.

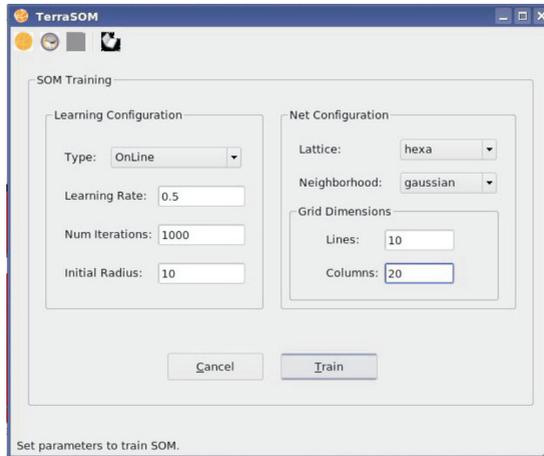


Figura 8. Configurar Rede Neural SOM.

- No menu *Learning Configuration*, será configurada a aprendizagem da Rede SOM. Em *Type*, selecione o tipo de aprendizagem. Pode ser *Batch (Em lote)* ou *Online (Sequencial)*.
- Informe a taxa de aprendizagem no campo referente a *Learning Rate*.
- Informe o número de iterações (épocas) da Rede SOM no campo referente a *Num Iterations*.
- Em *Initial Radius*, informe o raio de vizinhança inicial.
- No menu *Net Configuration*, será configurada a topologia da Rede SOM. No primeiro campo deve ser escolhido o tipo de Lattice, que pode ser hexagonal ou retangular.
- Em *Neighborhood* define-se a função de vizinhança.
- Por fim, define-se o tamanho da grade. Informe o número de linhas no campo referente a *Lines* e o número de colunas no campo referente a *Columns*.

É importante salientar que o tempo de processamento da Rede SOM é diretamente proporcional ao tamanho de sua grade e ao número de iterações.

Segmentando a Rede SOM

Após finalizar o processo de treinamento, o botão de agrupamento (clustering) ficará habilitado. O ícone  representa a opção *Clustering*. A Figura 9 mostra a tela de segmentação da rede.

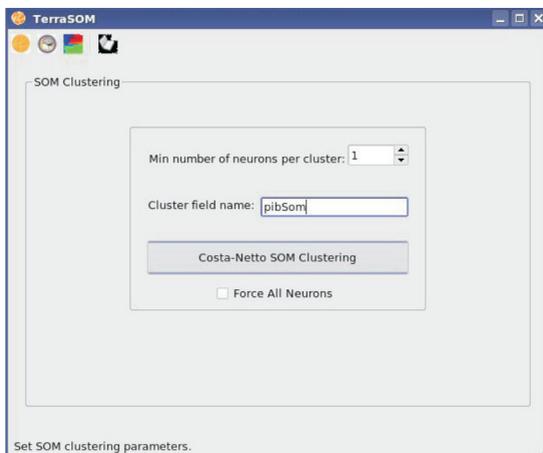


Figura 9. SOM Clustering.

Para configurar a segmentação da Rede SOM os seguintes passos deverão ser executados:

- Clique no botão de *Clustering* indicado pelo ícone . No primeiro campo referente ao menu *SOM Clustering*, informe o número mínimo de neurônios por agrupamento.
- No campo a seguir, referente a *Cluster Field Name*, informe um nome para esta nova coluna que será inserida no BD.
- Para executar a segmentação da Rede SOM, clique no botão *Costa-Netto SOM Clustering*.

Para uma melhor visualização dos agrupamentos formados, pode-se escolher uma cor para cada grupo.

- Clique com o botão direito do mouse sobre um *Theme*, da paleta Views/Themes e escolha a opção *Edit Legend*.
- Inserir um figurar associada ao tópico 1.
- No *Legend Editor*, no menu *Grouping Parameters*, na opção *Mode* escolha um dos modos de configuração da legenda. Em *Attribute* defina o coluna da tabela do BD que foi previamente gerada pelo TerraSOM (Figura10).

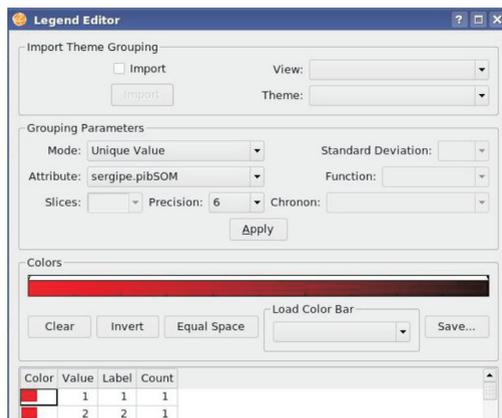


Figura 10. *Legend Editor*.

- Na parte inferior do *Legend Editor* poderá ser visto as legendas. Clique duas vezes sobre alguma cor da coluna *Color* para mudá-la.

Resultados

Para visualizar os resultados basta clicar no botão *Mostrar Resultados*, representado pelo ícone . Acompanhe, na figura 11, os resultados de experimentos realizados com o TerraSOM. Observe que na tela onde são exibidos os resultados, clicando no botão *Save Map*, há a opção de salvar o mapa auto-organizável (SOM) que acabou de ser treinado. Terminado o processamento do plugin TerraSOM, pode-se visualizar a localização espacial dos grupos através do TerraView.

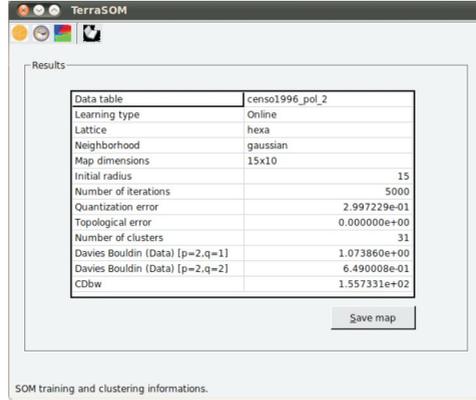


Figura 11. Resultados.

Embrapa

Tabuleiros Costeiros

Ministério da
**Agricultura, Pecuária
e Abastecimento**

GOVERNO FEDERAL
BRASIL
PÁTRIA EDUCADORA