

NEUROCOMPUTAÇÃO E SIG NO SUPORTE AO PLANEJAMENTO URBANO

Marcos Aurélio Santos da Silva¹

Palavras-chave: Redes Neurais Artificiais, Análise Espacial de Área, Visualização de dados, Análise de Agrupamentos Espaciais, SIG.

1. Introdução

A capacidade para geração, armazenamento e recuperação de dados, com referência no espaço e no tempo, cresceu muito nos últimos anos. No entanto, a nossa capacidade de analisar este conjunto de dados, em várias escalas e com existência em unidades espaciais distintas, é bem menor que a nossa capacidade de produzi-lo. Os Sistemas de Informação Geográfica ampliaram suas capacidades funcionais para a análise desta massa de dados geograficamente referenciados, aqui chamados dados geoespaciais, mas ainda estão no início da efetiva integração dos métodos de exploração e análise dos mesmos. 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 (Câmara & Monteiro, 2001; Franzini et al., 2001).

O objetivo deste trabalho foi procurar oferecer, em um ambiente de processamento de informação geográfica integrado, uma metodologia que utilize a técnica de Redes Neurais Artificiais (RNA), mais precisamente dos Mapas Auto-Organizáveis de Kohonen – SOM (Kohonen, 2001), como uma ferramenta para análise exploratória de dados geoespaciais multivariados de área. O propósito é avaliar este tipo de rede como técnica possível para detecção de padrões, visualização e análise de componentes. São objetivos específicos deste trabalho: especificar quais métodos de análise exploratória do SOM podem ser aplicados aos dados geoespaciais e como esta aplicação deve ser conduzida; verificar a sensibilidade dos métodos às variações nos parâmetros livres da rede neural, principalmente quanto às dimensões da rede; pesquisar métodos automáticos de análise de agrupamentos em dados geoespaciais a partir do SOM; identificar a distribuição espacial e regimes específicos de dependência espacial global e local do fenômeno a partir do SOM. A metodologia foi aplicada no problema de mapeamento da

¹ Laboratório de Geotecnologias Aplicadas - Embrapa Tabuleiros Costeiros, Av. Beira Mar, 3250 – 49025-040 Aracaju – SE, Tel.: (79) 2261300 Fax: (79) 226 1369 e-mail: aurelio@cpatc.embrapa.br

exclusão/inclusão social intra-urbana em São José dos Campos – SP e os resultados comparados com os obtidos por Genovez (2002).

2. Metodologia

A SOM é uma RNA competitiva com duas camadas: a camada de entrada I e a de saída U. A entrada da rede corresponde a um vetor no espaço d -dimensional em \mathcal{R}^d , representado por $x_k = [\xi_{k1}, \dots, \xi_{kd}]^T$, $k = 1, \dots, n$, sendo n o número de vetores de entrada. Cada neurônio j da camada de saída possui um vetor de código w_j , também no espaço \mathcal{R}^d , associado ao vetor de entrada x_k , $w_j = [w_{j1}, \dots, w_{jd}]^T$ (Figura 1). A principal função da SOM é fazer uma aproximação não-linear dos dados de entrada de forma não-supervisionada e com base nas propriedades locais do conjunto de dados de forma que padrões próximos no espaço de entrada estejam associados a neurônios próximos na grade neural.

A metodologia consiste de 4 fases: na primeira os dados são selecionados e processados de maneira que os valores das variáveis possuam significado; posteriormente define-se um conjunto de redes SOM a serem avaliadas segundo métricas próprias (erro de quantização e topológico) ou visualmente; em terceiro, para cada tipo de análise usa-se uma técnica distinta associada a SOM, ver Figura 2; os resultados da fase anterior fornecerá diferentes visões para o mesmo conjunto de dados auxiliando tanto trabalhos de pesquisa como a elaboração de políticas públicas.

Para detecção de dados atípicos e análise geral da estrutura dos dados foi usada a matriz de distância unificada (Ultsch, 1993). A U-matriz utiliza os pesos dos neurônios da rede SOM, treinada, para gerar uma imagem bidimensional onde as zonas de similaridade são representadas por regiões com abaixo valores (azul) e separadas por regiões com baixos valores (vermelho), dissimilares.

Os Planos de Componentes (PC) consistem numa técnica de análise da intensidade de cada componente em cada neurônio da rede (Vesanto, 1997; Kohonen, 2001). Esta análise, realizada através da coloração dos neurônios segundo a intensidade de seu componente, permite que possa ser avaliada a correlação entre componentes e o comportamento do fenômeno ao longo da rede neural. Também utilizamos os PC para mapear no mapa geográfico a distribuição espacial do fenômeno automaticamente.

Outro fator considerado na análise exploratória é o grau de dependência espacial e detecção de fragmentos urbanos, zonas de alto grau de similaridade, através da partição automática dos setores analisados segundo sua distribuição do mapa neural através dos algoritmos Costa-Netto (Costa & Andrade Netto, 2003 e CDbw (Halkidi & Vazirgiannis, 2002; Wu & Chow, 2004).

A metodologia proposta foi aplicada no problema de mapeamento da inclusão/exclusão social intra-urbana, em São José dos Campos – SP, através da análise dos dados associados aos setores censitários. A seleção dos dados baseou-se no método revisto quantitativo (Genovez, 2002). Os índices indicam maior (+1) ou menor (-1) inclusão social no espaço urbano. Para este estudo selecionou-se os índices de Distribuição de Renda dos Chefes de Família (ARENDR), Desenvolvimento Educacional (DESEDUCR), Estímulo Educacional (ESTDUCR), Longevidade (LONGR), Qualidade Ambiental (QAMBR), Conforto Domiciliar (QDOMR), Mulheres não Alfabetizadas (MANALFR) e Concentração de Mulheres Chefes de Família (CMCHFR).

Para processamento dos dados foi desenvolvido, no escopo deste trabalho, o sistema CASAA – *Connectionist Approach for Spatial Analysis of Areal Data*, que lê e processa dados geoespaciais armazenados em bancos de dados geográficos em formato **TerraLib** (DPI/INPE, terralib.dpi.inpe.br). A visualização dos mapas geográficos foi realizada através do sistema **TerraView** (DPI/INPE).

3. Análise dos Resultados e Conclusões

A Figura 3 mostra a U-matriz para a rede SOM 20x15 após o processo de aprendizagem. A partir da U-matriz observa-se uma grande área homogênea (azul) ao centro e que destacam-se duas áreas nos cantos superiores separadas das demais por duas regiões de dissimilaridade (em vermelho). Fazendo o mapeamento dos setores censitários associados aos neurônios correspondentes às regiões dos cantos superiores identificamos que as mesmas correspondem a regiões que apesar de estarem em zonas de inclusão ou exclusão possuem indicadores altos de exclusão e inclusão, respectivamente (Figura 4).

A Figura 5 mostra como cada componente analisado está distribuído na rede 20x15. Deste PC observa-se que: as variáveis ARENDR e DESEDUCR possuem um elevado grau de similaridade visual, um forte indício que ambas podem estar estatisticamente correlacionadas. Também observa-se que existem mais setores com maior nível de inclusão na variável DESEDUCR do que na variável ARENDR. Todavia, existem mais setores no setor de exclusão na variável ARENDR do que na variável DESEDUCR; as variáveis LONGR e QAMBR contribuem muito pouco para a diferenciação entre os setores censitários, uma vez que possuem grandes áreas homogêneas no Mapa, com destaque para a variável LONGR; Não existe correlação visual entre as variáveis MANALFR e MCHFR. Rotulando-se os neurônios segundo a direção da distribuição da exclusão/inclusão social no mapa neural e fazendo o mapeamento para o mapa geográfico tem-se que a distribuição do fenômeno é centro-periferia (Figura 6).

Através do índice CDbw de validação de partição de dados por meio do algoritmo Costa-Netto chegou-se ao particionamento da Figura 7. Fazendo-se o mapeamento dos setores associados aos neurônios no mapa geográfico observa-se que aproximadamente 60% dos setores vizinhos no espaço de atributos também são ‘vizinhos’ no espaço físico, o que evidencia uma forte dependência espacial. Também foi possível detectar fragmentos urbanos em regiões de inclusão ou exclusão social urbana, como pode ser observado na área em destaque da Figura 8. Nesta área de exclusão social identificou-se vários subgrupos distintos entre si e com alto grau de similaridade local.

Foi observado que os resultados obtidos nestes experimentos são bastante próximos daqueles alcançados por Genovez (2002) através de métodos estatísticos e de análise espacial estatística, sendo que o método neural proposto oferece mais perspectivas de análise exploratória de maneira não-supervisionada e semi-automática.

4. Referências Bibliográficas

Câmara, G.; Monteiro A. Geocomputation Techniques for Spatial Analysis: Is it the case for health data sets? **Caderno de Saúde Pública**, v. 17, n. 5, p. 1059-1081, 2001.

Costa, J.A.F.; Andrade Netto, M. L. Segmentação do SOM baseada em particionamento de grafos. **Congresso Brasileiro de Redes Neurais**. 2003.

Franzini, L.; Bolchi, P.; Diappi, L. Self Organizing Maps: A clustering neural method for urban analysis. **Proc. of the V Recontres de Théo Quant**. 2001. p. 1-15.

Genovez, P. C. **Território e Desigualdades: Análise Espacial Intra-Urbana no estudo da dinâmica de exclusão/inclusão social no espaço urbano em São José dos Campos-SP**. Dissertação – INPE, Dezembro 2002.

Halkidi, M.; Vazirgiannis, M. Clustering validity assessment using multi-representatives. **Proc. of SETN Conference**, Thessaloniki, Grécia. 2002.

Kohonen, T. **Self-Organizing Maps**. Springer. 3^a ed. 2001.

Ultsch, A. **Knowledge extraction from self-organizing neural networks**. In: Opitz, O. ed. Information and Classification. Springer, 1993.

Vesanto, J. **Data mining techniques based on the Self-Organizing Map**. Dissertação – Helsinki University of Technology, Maio, 1997.

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**, v. 37, p. 175-188, 2004.

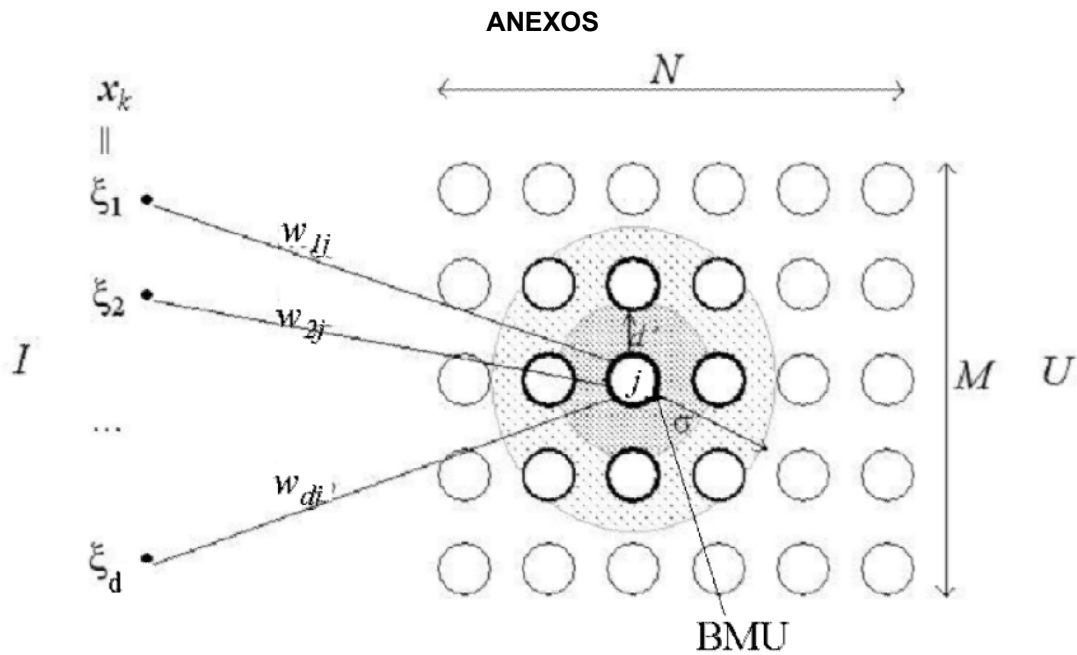


Figura 1. Exemplo de um Mapa Auto-Organizável bidimensional $N \times M$, com entrada x_k .

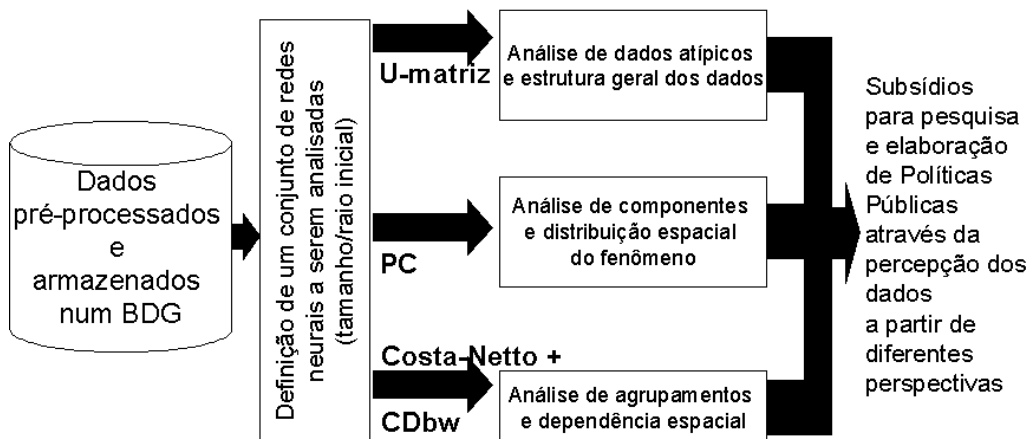


Figura 2. Metodologia de análise exploratória de dados geoespaciais com SOM.

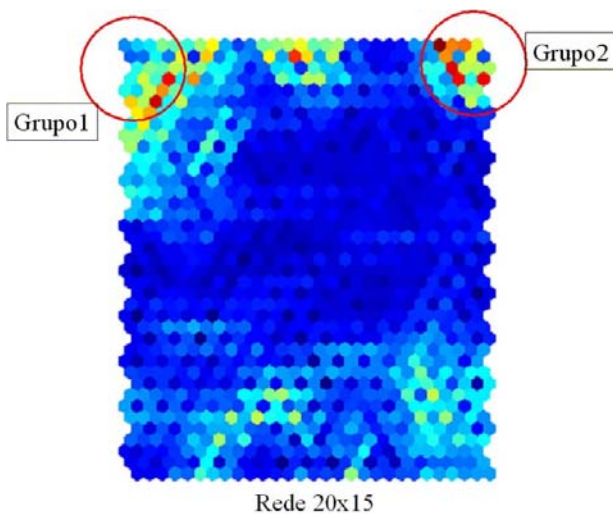


Figura 3. U-matriz gerada para a rede 20×15 .

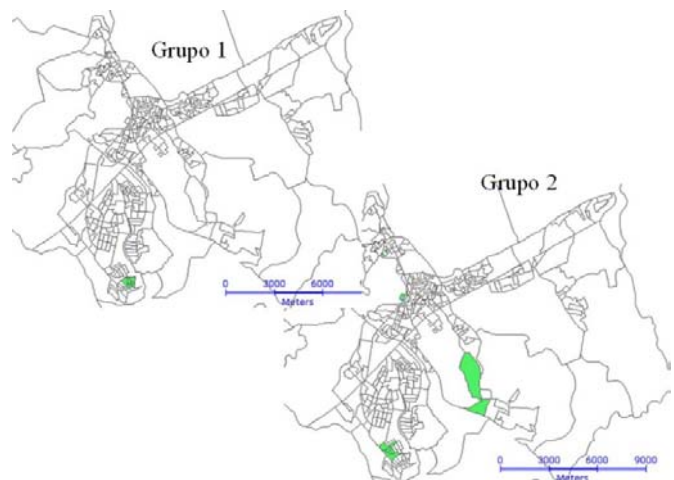


Figura 4. Mapas dos setores censitários identificados como setores atípicos.

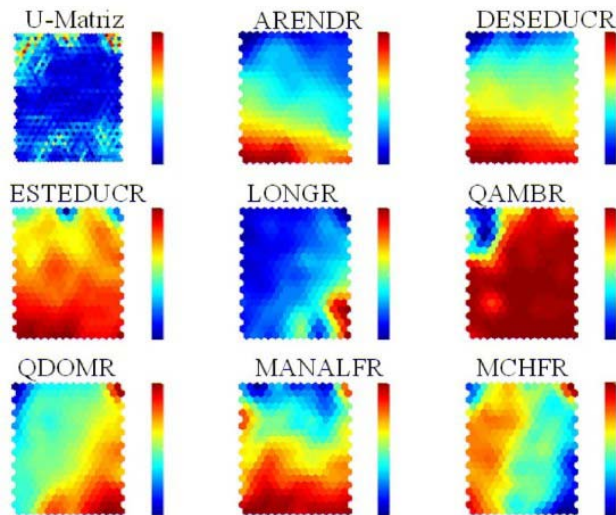


Figura 5. Planos de Componentes para a rede 20x15.

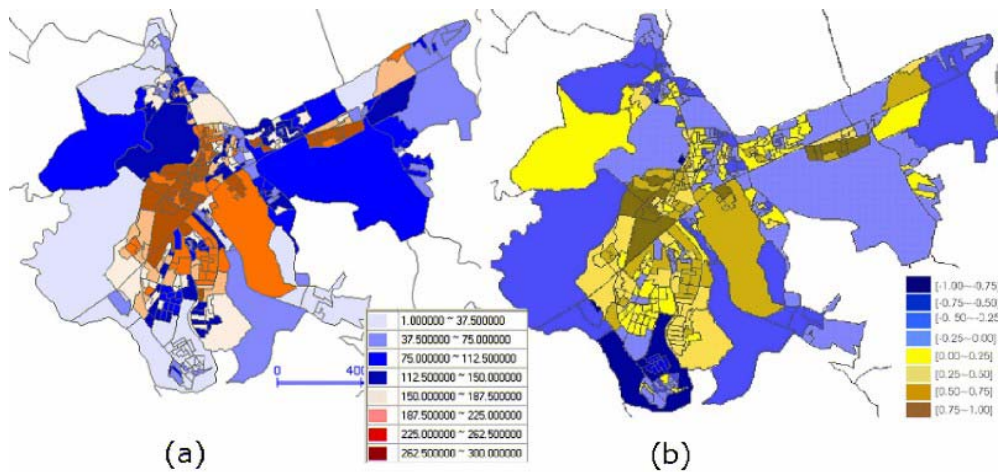


Figura 6. “a” Mapa gerado a partir da rotulação, no sentido vertical, da grade de neurônios, baseada na distribuição dos Planos de Componentes. “b” Mapa baseado no lex reviso, FONTE: (Genovez, 2002).

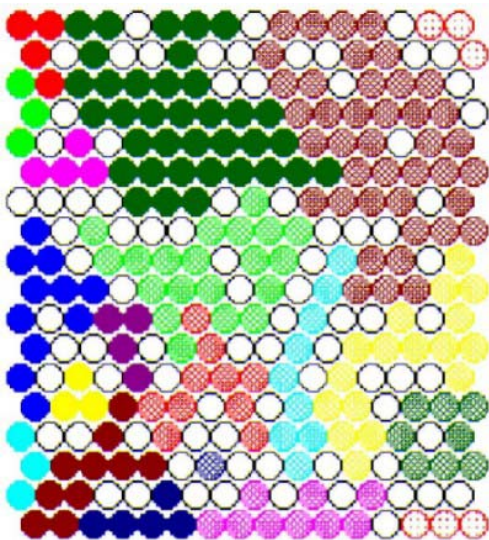


Figura 7. Mapa 18x16 particionado segundo o índice Cdbw.

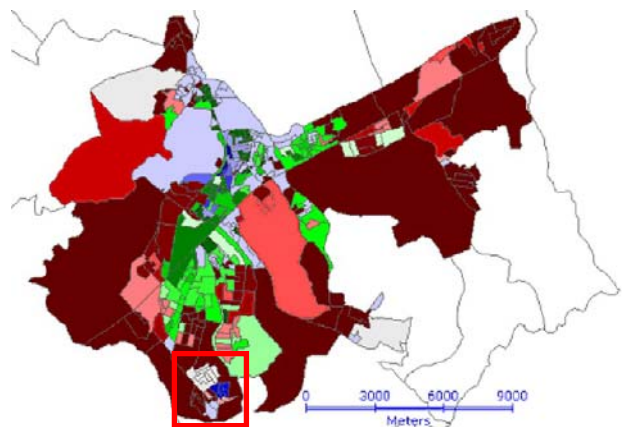


Figura 8. Mapa dos setores censitários gerados a partir do SOM particionado segundo o algoritmo Costa-Netto e validação do índice CDbw. Em destaque, o setor sul da área urbana onde pode-se verificar que o algoritmo identificou dentro de uma área de exclusão sub-agrupamentos que podem ser caracterizados como fragmentos urbanos