

DESCOBERTA DE CONHECIMENTO SOBRE DADOS GEOESPACIAIS CENSITÁRIOS COM O AUXÍLIO DE MAPAS AUTO-ORGANIZÁVEIS

LEITE, A. Y. C. M.¹; MATOS, L. N.²; SILVA, M. A. S. DA³

Resumo: O objetivo deste trabalho é descobrir conhecimento sobre dados dos municípios do estado de Sergipe extraídos do censo agropecuário realizado em 1996 feito pelo IBGE (Instituto Brasileiro de Geografia e Estatística) utilizando Mapas Auto-Organizáveis, um tipo de rede neural artificial (RNA), para auxiliar na identificação de agrupamentos espaciais e correlação entre as diversas variáveis, visto que os Mapas Auto-Organizáveis tem sido amplamente utilizados para análise de conjuntos complexos de dados, como dados censitários multivariados. O resultado obtido através dos Planos de Componentes mostrou que cada grupo de municípios de Sergipe utiliza suas terras predominantemente de uma maneira, ou seja, ou para pastagens ou lavoura etc.

Palavras-Chave: Mapas Auto-Organizáveis, Dados Geoespaciais, Descoberta de Conhecimento

DISCOVERY OF KNOWLEDGE ABOUT CENSUAL GEOESPATIAL DATA USING SELF-ORGANIZING MAPS

Abstract: The purpose of this paper it's to discover knowledge about the data related to the cities of the state of Sergipe and extracted from the farming census made in 1996 by the IBGE (Brazilian Institute of Geography and Statistics) using Self-Organizing Maps, a kind of artificial neural network (ANR), in order to help the identification of spatial groupings and correlation between the different variables, since the Self-Organizing Maps have been widely used for analysis of complex sets of data, such as multi-varied censual data. The result accomplished using Plans of Components showed that each group of cities of Sergipe utilizes its lands mainly in only one way, that is., either for pastures or farming etc.

Keywords: Self-Organizing Maps; Geospatial Data; Knowledge Discovery.

Introdução: Atualmente o volume de dados produzidos é muito grande, o que tornou complexa a análise dos mesmos. Para viabilizar a análise exploratória destes dados são utilizados métodos

¹ Alan Yves C. Martins Leite, Graduando em Ciência Da Computação, Universidade Federal de Sergipe - UFS, alanyves@yahoo.com.br.

² Drº Leonardo Nogueira Matos, Cientista da Computação, Universidade Federal de Sergipe - UFS.

estatísticos (análise de fator, k-médias, análise de componentes principais) e métodos de áreas emergentes como Inteligência Artificial, Redes Neurais Artificiais etc.

O foco deste trabalho é apresentar como uma RNA, mais especificamente os Mapas Auto-Organizáveis de Kohonen (Kohonen's *Self-Organizing Map* ou simplesmente SOM), auxiliam na análise exploratória de dados complexos, como dados geoespaciais censitários. A partir da visualização de Planos de Componentes é possível de maneira eficiente verificar correlações entre diversas variáveis. O objetivo é identificar como quatro variáveis censitárias se correlacionam, se existe ou não municípios com comportamentos atípicos, e se há grupos de municípios com características semelhantes (homogêneos).

Material e Métodos: O SOM é uma RNA competitiva, com uma camada de entrada e outra de saída, onde cada neurônio da camada de saída é associado a um vetor da camada de entrada e uma de suas principais aplicações é a descoberta de agrupamentos de dados.

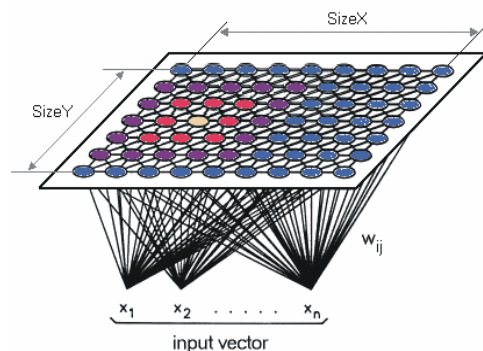


Figura 1 - Topologia da Rede SOM

Os Planos de Componentes possibilitam visualizar as correlações entre os atributos de entrada do mapa. Para cada atributo é gerado um Plano de Componente, que é uma representação gráfica (codificação de cores) dos valores de cada variável no vetor de pesos. Neurônios com cores similares representam características (índices) similares. Foi utilizada uma escala de cores do azul para o vermelho, onde quanto mais escuro o tom de azul significa índices mais baixos e quanto mais escuro o tom de vermelho significa índices mais altos.

Os dados referentes aos municípios de Sergipe, obtidos do censo agropecuário 1995/1996 realizado pelo IBGE, foram selecionados e padronizados gerando as entradas utilizadas no SOM. Ao total foram quatro variáveis que corresponde ao fator Utilização da Terra.

Foi utilizado um pacote de funções que implementam os algoritmos da rede SOM para o Matlab chamado SOM Toolbox para visualização dos Planos de Componentes.

³ Marcos Aurélio Santos da Silva, Cientista da Computação – Embrapa Tabuleiros Costeiros.

Depois de iniciar e apresentar os dados às redes SOM, procedemos à fase de descoberta de conhecimento através da análise dos Planos de Componentes, e rotulando os neurônios com os nomes dos municípios, o que possibilitou identificar correlações entre os atributos de entrada.

Para geração de um mapa geográfico de Sergipe, colorido de acordo com os resultados obtidos dos Planos de Componentes, foi utilizado o SIG (Sistema de Informações Geográficas) Terraview do INPE (Instituto Nacional de Pesquisas Espaciais).

Resultados e Discussão: Através dos planos de componentes foi possível extrair conhecimento importante acerca dos dados censitários, revelados pela existência de correlação entre as variáveis envolvidas, o que não seria possível fazer pela inspeção dos dados em estado bruto. O uso de planos de componentes facilita a realização de cruzamento de informação, ainda que de modo subjetivo, por um analista humano. Os mapas gerados não quantificam o grau de similaridade entre atributos, mas apresentam visualmente as relações existentes entre eles.

Logo abaixo se encontra os Planos de Componentes relacionados com atributos relativos à utilização da terra:

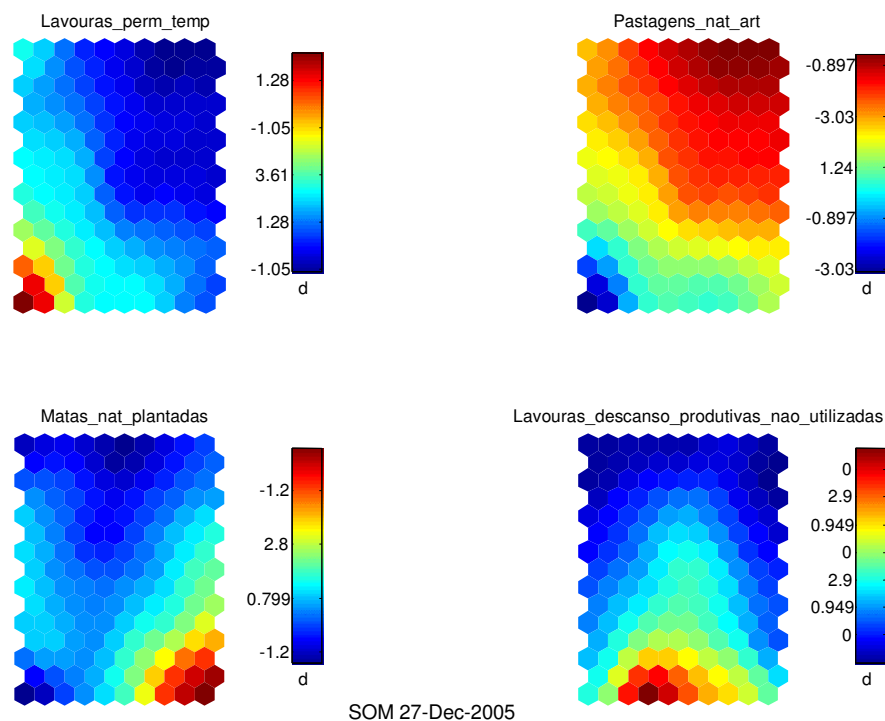


Figura 2 - Plano de Componentes

Pode-se observar que as variáveis se complementam, onde há altos índices de pastagens, as outras variáveis (lavouras_perm_temp, matas_nat_plantadas e lavouras_descanso_produtivas_não_utilizada) apresentam índices baixos, e que grande parte dos municípios utiliza as terras para pastagens.

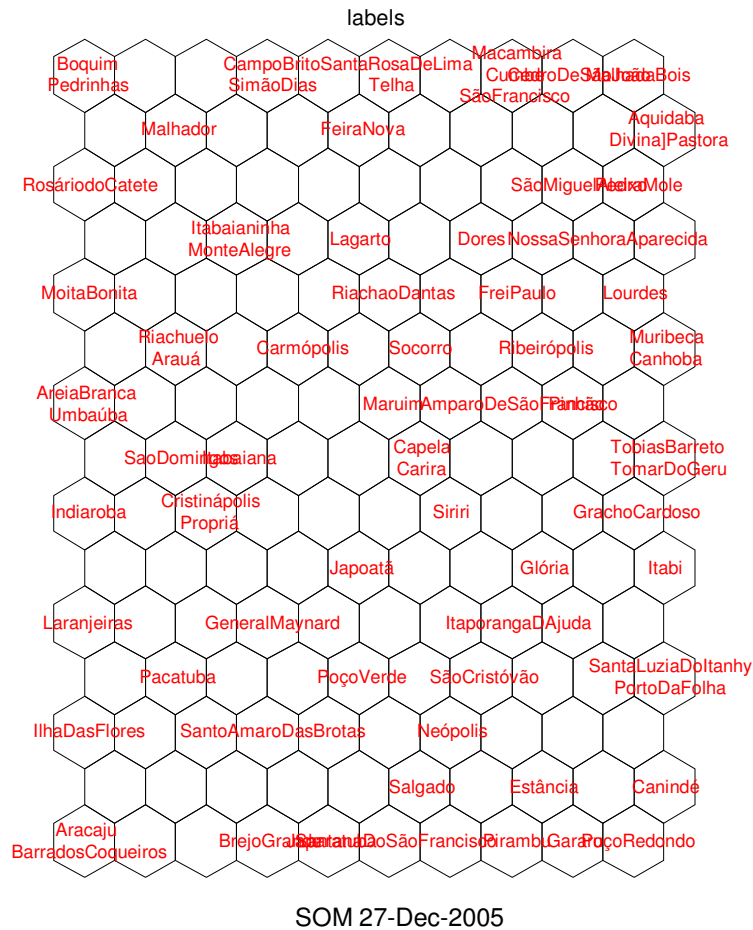


Figura 3 - Grade com os rótulos (nomes dos municípios)

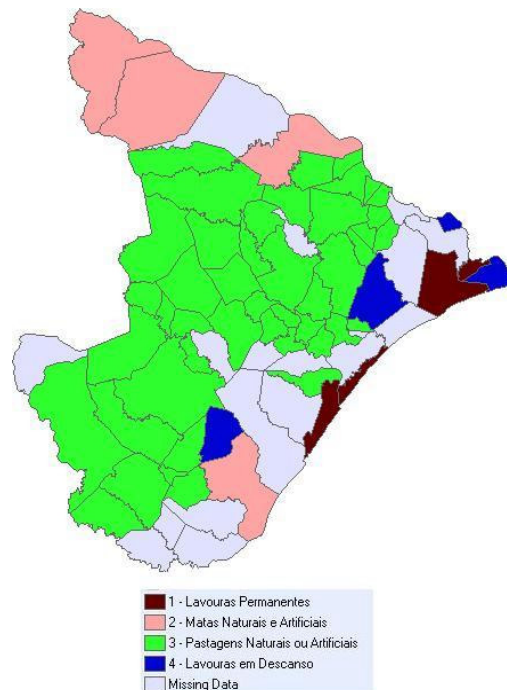


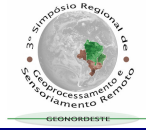
Figura 4 – Mapa de Sergipe Gerado no TerraView

Através dos rótulos é possível identificar a relação entre os atributos e os municípios. Sendo assim podemos observar que os municípios que utilizam suas terras para lavouras são Aracaju, Barra dos Coqueiros, Ilha das Flores, Pacatuba; os que mais apresentam áreas com Matas Naturais são Canindé, Poço Redondo, Gararu, Estância; a maioria dos municípios utiliza suas terras para pastagens, sejam naturais ou artificiais, destes os que apresentam os mais altos índices são Malhada dos Bois, Macambira, São Francisco, Aquidabã, Divina Pastora; e os municípios que utilizam suas terras para lavoura, sendo que estas estavam em descanso foram Brejo Grande, Salgado, Japarutuba e Santana do São Francisco.

Em breve deve ser realizado um novo censo agropecuário no Brasil, seria interessante realizar o mesmo processo e comparar os resultados obtidos com este para descobrir se houve mudanças significativas nos municípios de Sergipe.

Conclusões: Os aspectos positivos na utilização da rede SOM para análises de dados complexos são a facilidade de interpretar os Planos de Componentes e obter informações a partir deles, e que apesar do empirismo envolvido na geração de uma rede SOM, como dimensão da rede e número de épocas não há grandes variações de resultados para pequenas variações dos parâmetros.

Outro fator que deve ser ressaltado é a importância da interação da computação com outras áreas, geografia, economia, entre outras, otimizando suas tarefas, no caso deste trabalho a de analisar um



conjunto de dados complexos, é importante difundir como essas tecnologias podem ser empregadas para resolver certos problemas, como análise de um conjunto de dados complexo.

Algo que facilitaria o trabalho seria a integração do TerraView com o Matlab, se os dois ambientes fossem integrados o mapa geográfico estaria ligado à rede SOM, existindo uma funcionalidade nessa integração o mapa poderia ser colorido automaticamente de acordo com o resultado obtido da rede.

Referências Bibliográficas:

- Disponível em: <<http://www.dpi.inpe.br>>. Acesso em 11/08/2005.
- HEWITSON, B.; CRANE, R. **Neural Nets: applications in geography**. Kluwer, p. 53 - 70, 1994.
- KOHONEN, T. **Self-Organizing Maps**. Springer, 2001. Third Edition.
- KOUA, E. L. Using Self-Organization Maps for Information Visualization and Knowledge Discovery in Complex Geospatial Datasets. **Proceedings of the 21st International Cartographic Conference (ICC)**, Durban, África do Sul, 2003, p. 1694 - 1701.
- OPENSHAW, Stan; OPENSHAW, Christine. **Artificial intelligence in geography**. Jhon Wiley, p. 179 - 197, 1997.
- SILVA, Marcos Aurélio Santos ; **Mapas Auto-Organizáveis na Análise Exploratória de Dados Geospaciais Multivariados**. 2004. Teste (Mestrado) – INPE.
- VESANTO, J.; HIMBERG, J.; ALHONIEMI, E.; PARHANKANGAS, J. Self-Organization Map in Matlab: the SOM Toolbox. **Proceedings of the Matlab DSP Conference**, 1999, p. 35 - 40.