

Segmentação da diversidade espaço-temporal da agricultura brasileira por meio de uma abordagem baseada em aprendizado de máquina

SANTOS, Flávio Emanuel de Oliveira¹; SILVA, Marcos Aurélio Santos da²; MATOS, Leonardo Nogueira³; MOURA, Fábio Rodrigues de⁴; DOMPIERI, Márcia Helena Galina⁴

¹ Graduando em Ciência da Computação, bolsista PBIC/FAPITEC/FUNTEC/CAPES, Embrapa Tabuleiros Costeiros, Aracaju, SE.

² Cientista da Computação, doutor em Computação, pesquisador da Embrapa Tabuleiros Costeiros, Aracaju, SE.

³ Cientista da Computação, doutor em Engenharia Elétrica, professor da Universidade Federal de Sergipe, São Cristóvão, SE.

³ Economista, doutor Economia Agrícola, professor da Universidade Federal de Sergipe, São Cristóvão, SE.

⁴ Estatística e Geógrafa, doutora em Organização do Espaço, pesquisador da Embrapa Territorial, Campinas, SP.

Resumo – O presente trabalho teve como objetivo segmentar os municípios Brasileiros de acordo com seus padrões espaço-temporais de diversidade agrícola. O índice de diversidade foi definido para oito categorias e calculado através do índice de entropia de Shannon a partir de estimativas anuais (1999-2018) da produção agrícola do IBGE. O método de segmentação proposto é baseado em Mapas Auto-Organizáveis de Kohonen e em algoritmos de análise de agrupamentos como o k-médias e o kml3d. Os resultados mostraram que o método proposto mostrou-se efetivo para a segmentação dos dados espaciais em painel da diversidade agropecuária brasileira, auxiliando na identificação de tendências e na proposição de uma regionalização de acordo com o perfil de diversidade municipal

Termos para indexação: mapa auto-organizável, dados espaciais em painel, visualização de dados, análise espacial, análise de trajetória.

Introdução

O Brasil apresenta forte heterogeneidade espacial quanto à distribuição das diversas atividades agropecuárias em seu território (Sales; Rodrigues, 2019). Esta característica exige que a elaboração de políticas públicas que promovam o desenvolvimento sustentável do espaço rural leve em consideração as particularidades de cada lugar ou região para que seja efetiva. Uma das formas de avaliar essa heterogeneidade é por meio de indicadores de diversidade agropecuária como os utilizados por Dessi et al. (2019). Estes índices de diversidade, quando medidos ao longo do tempo, fornecem informações relevantes sobre as tendências de diversificação ou de concentração agropecuária, importantes para a elaboração de políticas públicas.

Neste trabalho foram definidos oito indicadores de diversidade baseados na entropia de Shannon (Shannon, 1948) para 5570 municípios brasileiros que abrangem os temas efetivo de animais (incluindo o número de vacas ordenhadas), área plantada com culturas temporárias, valor da produção de culturas permanentes e temporárias, produção animal, extrativismo vegetal, silvicultura e aquicultura a partir das estimativas anuais realizadas pelo IBGE entre 1999 e 2018 (IBGE, 2021). Para a análise de tendência da diversidade agropecuária brasileira este trabalho propõe um método de segmentação baseado na rede neural artificial não-supervisionada Mapa Auto-Organizável de Kohonen (Kohonen, 2013) de forma a dividir os municípios em grupos com tendências homogêneas em termos de diversidade.

Material e métodos

A diversidade para cada categoria l , ano t e município p é definido pela Equação 1, onde X_{ltpi} representa o valor da estimativa do IBGE para a categoria l , ano t , município p e variável i , sendo m o número total de variáveis para a categoria l .

$$Diversidade_{ltp} = - \sum_{i=1}^m \left[\left(\frac{X_{ltpi}}{\sum_{j=1}^m X_{ltpj}} \right) \log_m \left(\frac{X_{ltpj}}{\sum_{j=1}^m X_{ltpj}} \right) \right] \quad (1)$$

Na tabela 1 tem-se as estatísticas básicas para os oito índices de diversidade onde se observa forte variação dos indicadores relacionados a aquicultura, extrativismo vegetal e silvicultura. A categoria efetivo de animais apresentou os maiores valores para média e valor máximo.

Tabela 1. Estatísticas básicas para os oito índices de diversidade considerando todas as observações para todo o período de 1999 a 2018.

Sigla	Categoria	DP	Média	CV(%)	Mediana	Max.	<i>m</i>
DIV.EFETIVO	Efetivo de animais, incluindo nº de vacas ordenhadas	0,18	0,48	37,50	0,52	0,87	11
DIV.PLANT.T	Área plantada com culturas temporárias	0,074	0,32	23,13	0,33	0,51	33
DIV.VL.T	Valor da produção de culturas temporárias	0,13	0,28	46,43	0,3	0,68	33
DIV.VL.P	Valor da produção de culturas permanentes	0,16	0,2	80,00	0,19	0,72	38
DIV.VL.PRODANI	Valor da produção da produção animal	0,18	0,26	69,23	0,26	0,88	6
DIV.AQU.VL	Valor da produção da aquicultura	0,085	0,026	326,92	0	0,66	24
DIV.EXTV.VL	Valor da produção do extrativismo vegetal	0,1	0,088	113,64	0,041	0,49	44
DIV.SILV.VL	Valor da produção da silvicultura	0,14	0,084	166,67	0	0,75	15

O método proposto, baseado em Ling e Delmelle (2016), consiste em sete etapas, sendo as duas primeiras a geração dos dados espaciais em painel a partir dos dados brutos do IBGE descritos acima. Na terceira etapa uma Rede Neural Artificial (RNA) não supervisionada Mapa Auto-Organizável de Kohonen (SOM, do inglês *Self-Organizing Maps*) é treinada a partir do conjunto de dados (5570 municípios, 20 anos e, conseqüentemente 111400 observações), ordenando-os na grade neural (Figura 1). Na quarta etapa a grade neural é segmentada (método k-médias sobre os pesos da RNA SOM) em zonas homogêneas sendo que cada município-ano estará associado a somente um neurônio da grade. Na quinta etapa essas zonas homogêneas são caracterizadas a partir da interpretação dos Planos de Componentes gerados a partir dos pesos da RNA SOM, assim é possível verificar se um determinado município-ano está associado a áreas com maior ou menor diversidade agropecuária. Na sexta etapa é definido o trajeto do município na grade neural conectando cada observação-ano segundo a ordem cronológica. Dessa forma cada município estará associado a uma trajetória na grade neural. Na sétima e última etapa essas trajetórias são segmentadas a partir do algoritmo kml3d (Genoulini et al., 2015) e os resultados são projetados no mapa geográfico para análise da distribuição espacial dos grupos. Os dois processos de segmentação dos dados usam métricas de qualidade de partição de dados Sillhouette (Rouseeew, 1987), Davies-Bouldin (Davies; Bouldin, 1979) e Caliński-Harabasz (Caliński; Harabasz, 1974) para auxiliar a definição do número *k* de grupos.

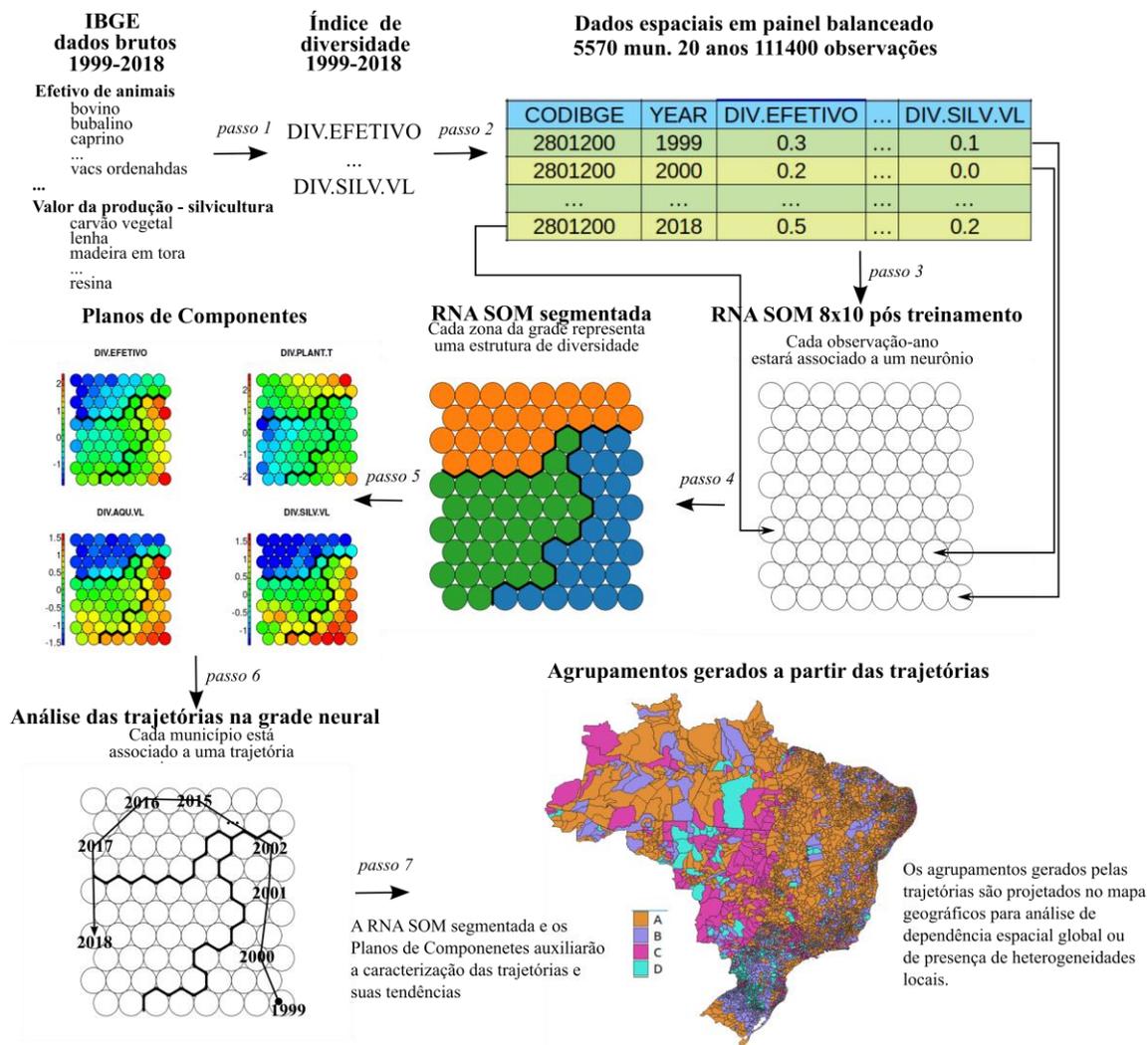


Figura 1. Método proposto para segmentação de dados espaciais em painel sobre a diversidade da agricultura brasileira.

Fonte: elaborado pelos autores.

Resultados e Discussão

A segmentação foi realizada a partir de uma RNA SOM 25x30 gerando oito grupos de trajetórias na grade neural projetados no mapa geográfico conforme a Figura 2. A partir da interpretação das trajetórias médias de cada um desses grupos foi possível identificar que quatro deles apresentam tendências de modificação do perfil da diversidade dos municípios entre 1999 e 2018 (grupos B, D, E e G), enquanto os demais mostram tendência à manutenção do mesmo estado de diversidade ao longo do tempo (grupos A, C, F e H). A partir do padrão global da distribuição dos agrupamentos é possível dividir o Brasil em três regiões: Região I, caracterizada pela forte diversidade em termos de efetivo de animais e baixa diversidade para valores da produção de culturas permanentes e temporárias, e aquicultura; Região II caracterizada pela tendência à diminuição da diversidade do efetivo de animais; e Região III que apresenta uma distribuição mais equânime do número de municípios por grupo e apresenta forte diversidade em termos de efetivo de animais, valor da produção de culturas permanentes e área plantada com culturas temporárias.

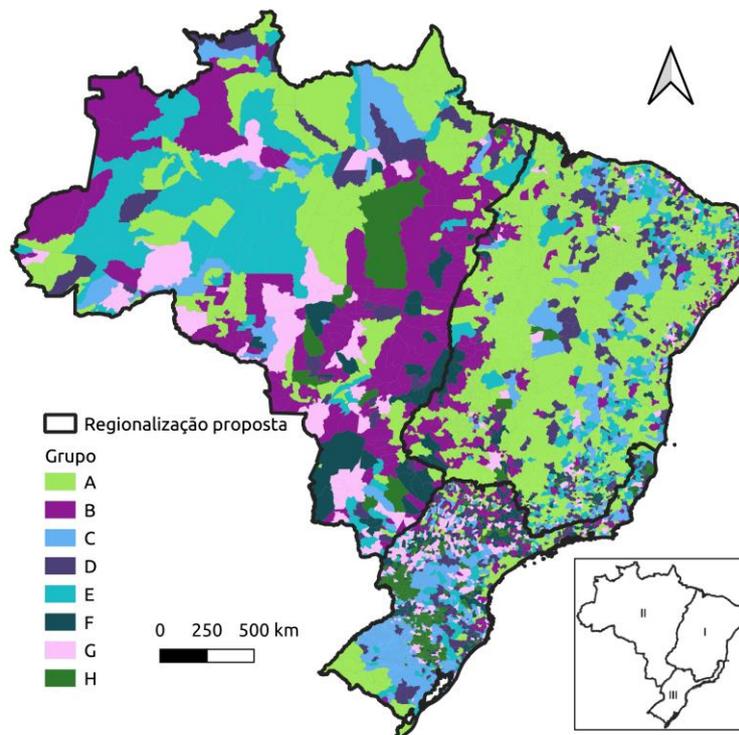


Figura 2. Segmentação dos municípios brasileiros a partir dos dados espaciais em painel da diversidade agropecuária segundo o índice de Shannon.

Fonte: elaborado pelos autores.

Conclusões

Conclui-se que o método proposto mostrou-se efetivo para a segmentação dos dados espaciais em painel da diversidade agropecuária brasileira, auxiliando na identificação de tendências e na proposição de uma regionalização de acordo com o perfil de diversidade municipal.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) e da FAPITEC por meio do edital nº 04/2019 PBIC/FAPITEC/SE/FUNTEC/CAPES.

Referências

- CALIŃSKI, T.; HARABASZ, J. A Dendrite Method for Cluster Analysis. **Communications in Statistics-theory and Methods**, v. 3, p. 1-27, 1974. doi:10.1080/03610927408827101.
- DAVIES, D. L.; BOULDIN, D. W. A Cluster Separation Measure". **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. PAMI-1, n. 2, pp. 224–227, 1979. doi:10.1109/TPAMI.1979.4766909.
- DESSIE, A.; ABATE, T.; MEKIE, T.; LIYEW, Y. Crop diversification analysis on red pepper dominated smallholder farming system: evidence from northwest Ethiopia. **Ecological Processes**, v. 8, n. 50, 2019.
- GENOLINI, C.; ALACOQUE, X.; SENTENAC, M.; ARNAUD, C. Kml and kml3d: R packages to cluster longitudinal data. **Journal of Statistical Software**, v. 65, n. 4, p. 1–34, 2015.
- IBGE. Sistema IBGE de recuperação automática (Tabelas 74, 94, 289, 291, 1612, 1613, 3939 e 3940). Disponível em: <https://sidra.ibge.gov.br>. Acesso em 15 jun. 2021.
- KOHONEN, T. Essentials of the self-organizing map. **Neural Networks**, v. 37, p. 52–65, 2013.
- LING, C.; DELMELLE, E. Classifying multidimensional trajectories of neighbourhood change: a self-organizing map and k-means approach. **Annals of GIS**, v. 22, n. 3, p. 173–186, 2016.

ROUSEEEEEUW, P. J. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Computational and Applied Mathematics*, v. 20, p. 53–65, 1987. doi:10.1016/0377-0427(87)90125-7.

SALES, C.; RODRIGUES, R. Espaço rural brasileiro: diversificação e peculiaridades. **Revista Espinhaço**, v. 8, n. 1, p. 54–65, 2019.

SHANNON, E. Mathematical theory of communication. **The Bell System Technical Journal**, v. 28, n. 4, p. 656–715, 1948.