AVALIAÇÃO DO INCREMENTO DE CLASSES NO ALGORITMO K-MÉDIAS PARA IDENTIFICAÇÃO DE ÁREAS DE SILVICULTURA

Luis Henrique De Araujo Gouvêa [1]

Aline Biasoli Trentin [2]

Resumo:

Um dos elementos mais importantes da paisagem é a vegetação, pois ela proporciona a conservação dos solos, manutenção dos recursos hídricos, e fornece abrigo para a fauna. Desta forma, conhecer a vegetação e sua distribuição espacial é de suma importância a fim de adotar práticas de manejo e conservação. A obtenção de dados, a partir de sensores multiespectrais, a bordo de satélites, revolucionou a forma de monitorar e mapear a vegetação. Uma forma de analisar e manipular estes dados é através do processamento digital de imagens. A classificação não supervisionada de imagens é um dos métodos de avaliação da distribuição das classes de alvos dominantes em uma determinada área, sendo o método de agrupamento K-médias, um dos mais utilizados. Assim, o objetivo deste trabalho foi avaliar o incremento no número de classes no método de classificação de agrupamento não supervisionado K-médias, para a identificação de áreas de silvicultura na Microrregião Campanha Central - RS. Foram utilizadas três cenas obtidas pelo sensor OLI do satélite de observação da terra Landsat 8, de agosto de 2017. O processamento inicial dos dados (mosaico das cenas e recorte da área de estudo) foi executado no software QGIS 2.18. Para a interpretação dos alvos foi adotada uma composição de imagem falsa cor utilizando as bandas 6R, 5G e 3B. Para testar o número de classes mais adequado para a identificação de áreas de silvicultura utilizou-se a ferramenta de classificação automática de imagens "K-Means Clustering for Grids" disponível no módulo "Interative Minimum Distance (Forgy, 1965)" do software SAGA 2.1.2., onde se gerou três mapas temáticos com 5, 10 e 15 classes. Estes produtos foram submetidos a uma avaliação visual, e os resultados demonstraram que o algoritmo k-médias foi capaz de reconhecer padrões na paisagem, porém não diferenciou efetivamente a silvicultura de outros tipos de vegetação em todos os incrementos. Sendo assim, considera-se que este método de classificação pode ser utilizado de forma exploratória, a fim de reconhecer as características da paisagem e não para separar completamente um alvo do outro. Ao fim deste trabalho recomenda-se testar outros tipos de classificadores e comparar os resultados para determinar qual, o método mais eficiente para separar a classe silvicultura dos outros tipos de vegetação.

Palavras-chave: geoprocessamento, sensoriamento remoto, uso e cobertura da terra

Modalidade de Participação: Iniciação Científica

AVALIAÇÃO DO INCREMENTO DE CLASSES NO ALGORITMO K-MÉDIAS PARA IDENTIFICAÇÃO DE ÁREAS DE SILVICULTURA

[1] Aluno de graduação. luishenriquegouvea@yahoo.com.br. Autor principal

[2] Docente. alinetrentin@unipampa.edu.br. Orientador

Avaliação do incremento de classes no algoritmo K-Médias para identificação de áreas de silvicultura

1. INTRODUÇÃO

A vegetação é um dos elementos mais importantes de uma paisagem. Ela proporciona a conservação dos solos, manutenção dos recursos hídricos, e fornece abrigo para a fauna. Sendo assim, é de suma importância conhecer sua distribuição espacial, a fim de obter informações para adotar práticas de manejo e conservação previamente planejados.

Tradicionalmente a aquisição de informações sobre a vegetação era realizada através de levantamentos in loco, o que demanda um alto investimento em recursos humanos e financeiros. Sendo assim, uma nova forma de monitorar a vegetação é extraindo informações dos dados obtidos por sensores a bordo de satélites de observação terrestre. A exemplo disso, a mais de 40 anos a missão Landsat, atualmente coordenado pelo *United States Geological Survey* (USGS), através de seus satélites, fornecem continuamente dados da superfície terrestre, os quais podem ser utilizados na agricultura, geologia, silvicultura, planejamento urbano, educação, cartografia, mapeamento e estudos sobre as mudanças globais (EUA, 2017).

Atualmente a análise desses dados é feita através de técnicas de processamento digital de imagens (PDI), entre as quais, a utilização de algoritmos computacionais para realçar e classificar automaticamente alvos de interesse. Um dos métodos de avaliação rápida da distribuição das classes de alvos dominantes em uma área é a classificação não supervisionada (MENESES e ALMEIDA, 2012).

A classificação não supervisionada é o processo pelo qual um analista submete um conjunto de dados espectrais à ação de um determinado algoritmo a fim de agrupar os dados em classes espectrais sem o conhecimento prévio da existência ou nome dessas classes (RICHARDS e JIA, 2006; NOVO, 2011). Mesmo existindo vários métodos de agrupamento de dados (XU et al., 2005), a maioria dos sistemas computacionais dedicados ao processamento digital de imagem utilizam os métodos de agrupamentos particionais, como o K-médias e o ISODATA (BALL et al., 1967).

De acordo com Jain (2010), mesmo proposto a mais de 50 anos, o método de agrupamento K-médias continua sendo muito utilizado, por ser simples eficiente e facilmente aplicável. Este método minimiza a soma dos quadrados sobre todos os k grupos, utilizando a métrica euclidiana, computando a distância entre pontos e o centro do grupo de dados. O método de agrupamento K-médias requer que o usuário especifique três parâmetros: número de classes K, iniciação e distância do agrupamento (Tibshirani et al., 2001).

A semelhança espectral entre os alvos e a dificuldade de obtenção de pixels puros para cada classe, podem ser fatores que dificultam a separabilidade de tipologias de vegetação (Moreira, 2007). Vasconcellos e Benedetti (2011) analisando a dinâmica temporal da cobertura florestal na microrregião Campanha Central a

partir da técnica de classificação supervisionada MaxVer (Máxima Verossimilhança), não obtiveram resultados positivos para diferenciar as áreas de silvicultura das áreas de vegetação nativa. Desta forma, justifica-se este trabalho, que tem por objetivo avaliar o incremento no número de classes no método de classificação de agrupamento não supervisionado K-médias, para a identificação de áreas de silvicultura na Microrregião Campanha Central - RS.

2. METODOLOGIA

A Microrregião Geográfica Campanha Central localizada na Mesorregião Geográfica Sudoeste Rio-Grandense, foi a área escolhida para a realização deste trabalho. Com aproximadamente 1.729.582 ha pertencentes ao Bioma Pampa, é composta pelos municípios de Santa Margarida do Sul, São Gabriel, Rosário do Sul e Santana do Livramento. A economia da microrregião é caracterizada pela pecuária extensiva, rizicultura, silvicultura e fruticultura. De acordo com o zoneamento ambiental do Rio Grande do Sul a região possui influência de sete unidades de paisagem caracterizadas pela fitofisionomia campestre do pampa gaúcho com formações Estepe-Gramíneo Lenhosa, Estepe-Parque e a presença da Floresta Estacional Decidual, de relevo plano com várzeas amplas e em algumas áreas, matas de galeria associadas aos cursos d´água.

Para cobrir toda área de estudo foi necessário a utilização de três cenas (Tabela 1) Landsat 8, com resolução espacial de 30m, fornecidas pelo Instituto Nacional de Pesquisas Espaciais (INPE) através da Divisão de Geração de Imagens (DGI). Para o processamento inicial dos dados utilizou-se o *software* QGIS 2.18.0, onde se realizou o mosaico das imagens, e o recorte utilizando uma camada vetorial do limite da microrregião estudada. A fim de facilitar o trabalho de interpretação dos alvos foi adotada uma composição de imagem falsa cor utilizando as bandas 6R, 5G e 3B.

Tabela 1: Cenas do satélite Landsat 8 que foram utilizadas neste trabalho.

Data	Órbita/Ponto	Nuvens
14/08/2017	224/81	0,01%
07/08/2017	223/82	0,81%
07/08/2017	223/81	0,62%

Para testar o número de classes mais adequado para a identificação de áreas de silvicultura utilizou-se a ferramenta de classificação automática de imagens "*K-Means Clustering for Grids*" disponível no módulo "*Interative Minimum Distance* (Forgy, 1965)" do software SAGA 2.1.2. Foram elaborados três mapas temáticos utilizando 5, 10 e 15 classes, sendo que para cada produto o algoritmo realizou 150 interações, pois de acordo com JAIN (2010), quanto maior o número de interações menor é a variação no produto final. Após o processamento os produtos gerados foram submetidos a uma avaliação visual.

3. RESULTADOS e DISCUSSÃO

Ao observarmos a Figura 1 podemos constatar que o algoritmo quando configurado para agrupar cinco classes e 150 interações conseguiu diferenciar dois tipos de vegetação, uma constituída de árvores e arbustos (cluster 3) e outra por vegetação campestre (cluster 5). A classe 3a é constituída principalmente pelo componente arbóreo-arbustivo possuindo áreas de floresta estacional, mata ciliar e áreas de silvicultura, porém o algoritmo também classificou nesta classe áreas alagadas em pousio da cultura do arroz. A classe 5a possui áreas de campo nativo, pastagem e até mesmo silvicultura. Podemos perceber que a silvicultura está presente nas duas classes de vegetação criadas pelo algoritmo. A confusão ocorrida no 5-médias pode ser devido à idade dos povoamentos florestais. Povoamentos jovens se confundem com áreas campestres ou agricultáveis, pois o seu dossel ainda está incompleto (Moreira, 2007), enquanto povoamentos mais maduros são confundidos com a vegetação mais densa (mata ciliar e floresta estacional), devido à característica densa do dossel da floresta. Gonçalves (2008) testando o algoritmo K-médias configurado para 4 classes e 100 interações, também constatou confusão entre as classes geradas em sua área de estudo.

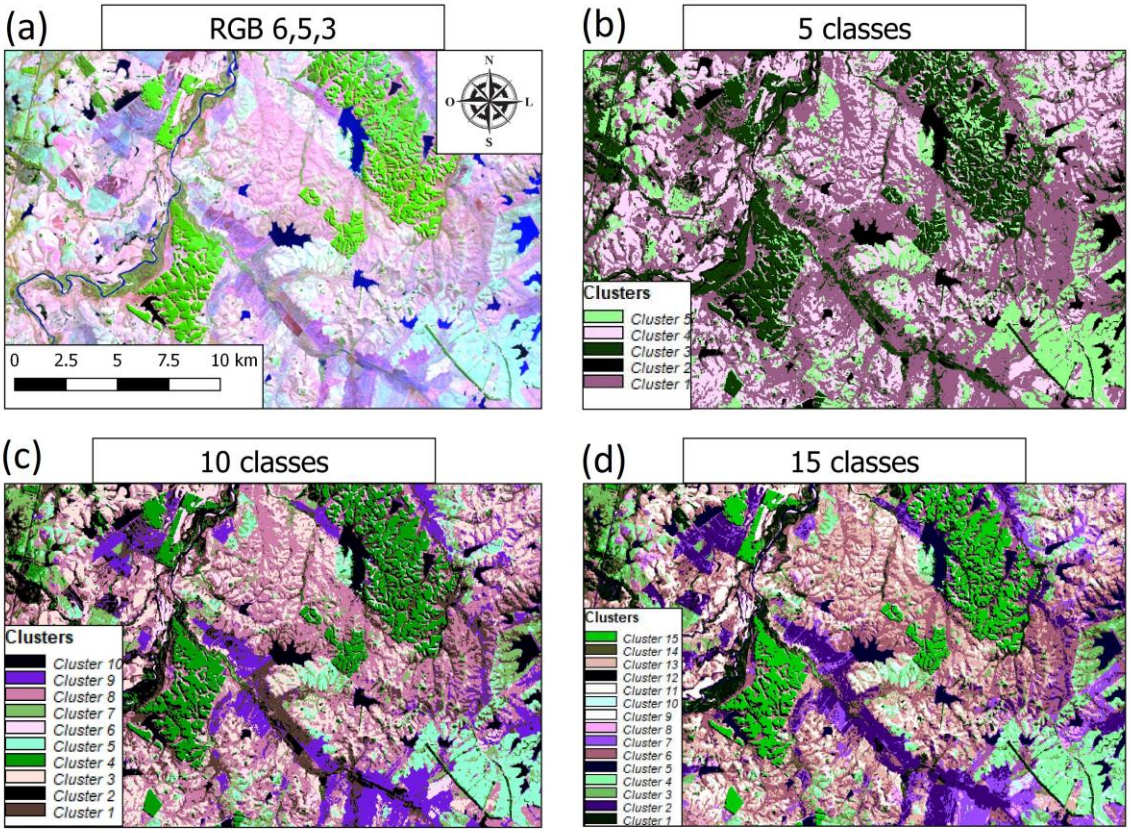


Figura 1: Amostras dos produtos gerados neste trabalho: (a) composição falsa-cor RGB 6,5,3; (b) classificação K-médias com 5 classes; (c) classificação k-médias com 10 classes; (d) classificação k-médias com 15 classes.

Quando configurado para 10 classes e 150 iterações o algoritmo dividiu a vegetação em quatro classes espectrais. A silvicultura foi identificada como pertencente à quarta classe (cluster 4), sendo que esta classe apresenta tanto silvicultura quanto áreas de floresta nativa. O algoritmo pode ter confundido alguns

pixels devido a semelhanças espectrais entre florestas plantadas e florestas nativas. Com o incremento para 15 classes o algoritmo foi capaz de identificar cinco padrões distintos na vegetação, onde a silvicultura está predominantemente inserida na décima quinta classe (cluster 15). Porém, pode-se constatar visualmente que, em menor quantidade, o algoritmo continua confundindo áreas de floresta estacional e mata ciliar, que apresentam o dossel mais denso, com áreas de silvicultura. Ou seja, os resultados obtidos confirmam a afirmação feita por Meneses e Almeida (2012), que classificadores não supervisionados são métodos de avaliação rápida da distribuição das classes de alvos dominantes em uma área, sendo capaz de identificar, mas não separar um alvo do outro.

4. CONSIDERAÇÕES FINAIS

Ao analisar o processamento conclui-se que a classificação automática através do método de agrupamento K-médias utilizando o algoritmo "*Interative Minimum Distance* (Forgy, 1965)" nas condições deste trabalho foi eficiente para a identificação de padrões. Porém o algoritmo não foi capaz de separar alvos que possuem comportamentos espectrais muito semelhantes. Desta forma, considera-se que este método de classificação, independente do número de classes (5, 10 e 15), não foi capaz de separar completamente a classe silvicultura dos outros tipos de vegetação. Sendo assim, o método de K-médias pode ser utilizado de uma forma exploratória, a fim de reconhecer padrões na paisagem. Recomenda-se testar outros tipos de classificação e comparar os resultados para determinar qual, o método mais eficiente para separar a classe silvicultura dos demais tipos de vegetação.

5. REFERÊNCIAS

BALL, G., HALL, D. A Clustering Technique for Summarizing Multivariate Data. Behavior Science, v. 12, pp. 153-155, 196

ESTADOS UNIDOS DA AMÉRICA. James R. Irons. Nasa. **Landsat Sience.** 2017. Disponível em: <https://landsat.gsfc.nasa.gov/>. Acesso em: 20 ago. 2017.

GONÇALVES, M. L. et al. CLASSIFICAÇÃO NÃO-SUPERVISIONADA DE IMAGENS DE SENSORES REMOTOS UTILIZANDO REDES NEURAIS AUTO-ORGANIZÁVEIS E MÉTODOS DE AGRUPAMENTOS HIERÁRQUICOS. **Revista Brasileira de Cartografia,** Brasil, v. 1, n. 60, p.1-13, abr. 2008.

JAIN, A. K.. Data clustering: 50 years beyond K-means. **Pattern Recognition Letters.** p. 651-666. 2010.

Meneses, P. R., & Almeida, T. De. **Introdução ao processamento de imagem de sensoriamento remoto. Brasília.** 2012. 266 p. Cnpq/Unb. https://doi.org/10.1590/S1413-294X2002000200020

MOREIRA, M. A. **Fundamentos do Sensoriamento Remoto e metodologias de aplicação**. 3 ed. 1° reimpressão. Viçosa: UFV, 2007.

NOVO, E. M. L. de M.; **Sensoriamento Remoto**: Princípios e aplicações. 4 ed. São Paulo: Blucher, 2010.

RICHARDS, J. A.; JIA, X. **Remote sensing digital image analysis.** 4.ed. Springer-Verlag, Berlin, 2006. 439 p.

VASCONCELLOS, B. N. de; BENEDETTI, A. C. P. DINÂMICA TEMPORAL DA COBERTURA FLORESTAL NA MICRORREGIÃO CAMPANHA CENTRAL DO RIO GRANDE DO SUL. **Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental,** Santa Maria, v. 4, n. 4, p.427-433, 2011. Disponível em: <https://periodicos.ufsm.br/geografia/article/view/24059>. Acesso em: 14 set. 2017.

XU, R., WUNSCH II, D. Survey of Clustering Algorithms. IEEE Transactions on Neural Networks. v. 16, n. 3, Maio, 2005.