

Aplicação da técnica de mineração de dados por meio do algoritmo J48 para definição de limiares de imagens de sensoriamento remoto

Rodrigo Rodrigues Antunes e Israel Rodrigues Gonçalves

Resumo: O objetivo principal da pesquisa foi demonstrar a eficiência da aplicação da técnica de mineração de dados no processo de classificação orientada a objetos de imagem de sensoriamento remoto. Para a mineração de dados, utilizou-se o sistema minerador *open source* WEKA, que constitui em sua estrutura o algoritmo J48. Foi utilizada um recorte de imagem urbana do sensor GeoEye-1, que possui resolução espacial de 50 cm. O resultado da pesquisa foi uma árvore de decisão com valores de limiares de cada classe definida.

Palavras-chave: Mineração de dados, sensoriamento remoto, árvore de decisão.

Abstract: The main objective of this research was to demonstrate the efficiency of applying data mining technique in the process of object-oriented classification of remote sensing image. For the data mining, the WEKA mining system was used, which the algorithm J48 is in its structure. A piece of urban image from 50cm spatial resolution sensor called GeoEye-1 was used. The result of this research was a decision tree with threshold values from each defined class.

Keywords: Data mining, remote sensing, decision tree.

Introdução

Uma importante tecnologia que vem sendo destaque no mundo é a mineração de dados. Os volumes de dados digitais nas empresas, sejam elas públicas ou privadas, estão cada vez maiores e mais complexos. Muitas informações acabam ficando escondidas, por falta de ferramentas adequadas para explorar os dados. O mesmo acontece com os dados geográficos. De acordo com Guidini e Ribeiro (2006), devido à facilidade crescente de aquisição de imagens e coleta de dados para aplicações, o volume de dados armazenados cresce muito além da capacidade humana de processá-los para extrair informações relevantes e conhecimentos úteis à tomada de decisão. Em aplicações geográficas, além dos dados alfanuméricos a serem processados, surge a necessidade de se lidar com geometrias e suas relações, determinando áreas de contato, de sobreposição etc., e com os aspectos temporais, que podem ser contínuos ou discretos. Estas dificuldades levam, naturalmente, à aplicação de técnicas de mineração de dados como solução para extrair conhecimento destes sistemas.

Para Miller e Han (2009), métodos analíticos espaciais tradicionais foram desenvolvidos em uma época em que a coleta de dados tinha alto custo e o poder computacional era limitado. O crescente volume e a diversidade de dados geográficos digitais limitam as técnicas de análises espaciais tradicionais, provocando escassez de informações a partir de pequenos e

homogêneos conjuntos de dados. O processo de “garimpagem” de dados é chamado de *Knowledge Discovery Databases* (KDD) e pode ser visto como processo de descoberta de novas correlações, padrões e tendências significativos por meio de análise minuciosa de grandes conjuntos de dados escondidos. A mineração de dados é uma técnica utilizada para a realização de KDD. São chamadas de mineração de dados todas as técnicas que permitem extrair conhecimento de uma massa de dados que, de outra maneira, permaneceria escondida nas grandes bases (QUONIAN, 2010).

Materiais

Para esta pesquisa utilizou-se dados exportados oriundos das amostragens de classes realizadas no software de classificação orientada a objetos IntelIMAGE. Essas amostragens foram realizadas após a segmentação feita em um recorte de uma imagem do sensor GeoEye-1 da região central da cidade de Goianésia, Goiás. Esse sensor possui 4 bandas espectrais e sua resolução espacial é de 50 cm.

Junto com as amostras foram exportados atributos dos objetos como: brilho, compacidade, razão, média e outros.



**Fig. 1 - Recorte da região central de Goianésia, Goiás.
Sensor GeoEye-1 (imagem de 2014).**

As amostras foram das seguintes classes de interesses: cobertura com telhado metálico, amianto, cerâmica (clara e escura) piscina e solo (exposto e outro).

Para a mineração de dados foi utilizado o software minerador WEKA, versão 3.6. Esse sistema minerador é open source, desenvolvido em java, pela Universidade de Waikato, de Nova Zelândia. A primeira versão foi lançada em 1999.

O algoritmo J48 que disponibiliza uma árvore de decisão por meio de um conjunto de dados, foi implementado no sistema WEKA. Segundo Mathur, Kumar e Santos (2012), esse algoritmo foi um aprimoramento do algoritmo ID3, desenvolvido por Ross Quinlan em 1983.

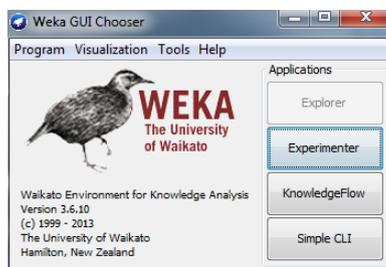


Fig. 2 – Tela de entrada do sistema minerador de dados WEKA.

O sistema microsoft excel foi utilizado para organizar os dados exportados pelo InterIMAGE e prepará-lo para leitura do sistema WEKA.

Metodologia

Para alcançar os resultados, foram cumpridas as etapas metodológicas apresentadas na Figura 3:

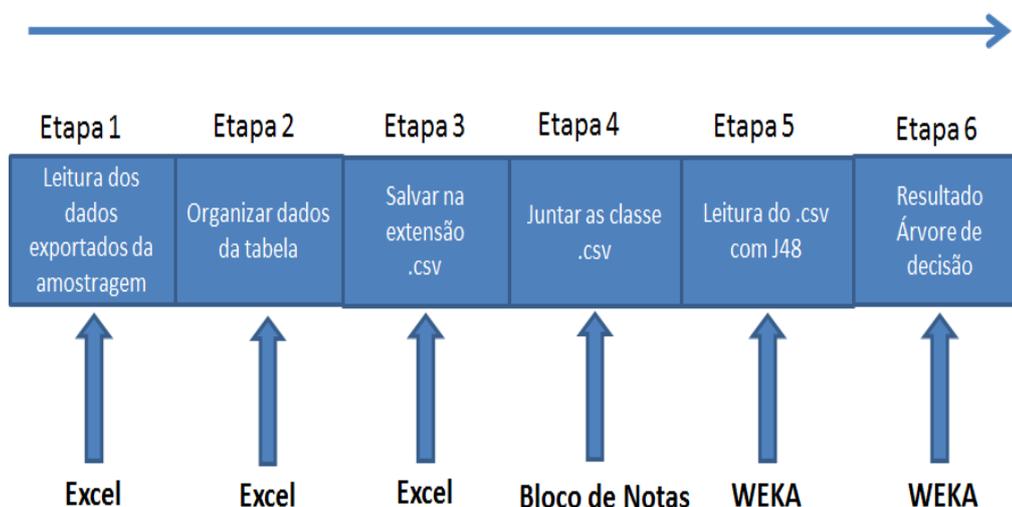


Fig. 3 - Esquema das etapas do trabalho.

Etapa 1: cada amostragem de classe exportada pelo software InterIMAGE foi salva na extensão .dbf. Essa extensão possibilitou a leitura pelo excel.

Etapa 2: foi preciso organizar a tabela, pois o arquivo exportado possui algumas colunas com informações que são desprezíveis e precisam ser excluídas.

Etapa 3: foi salvo o arquivo com a extensão .csv, pois o sistema minerador WEKA somente lê arquivos com essa extensão.

Etapa 4: depois de cada arquivo salvo na extensão .csv, foi juntado todos em um arquivo somente, conforme pode ser visto na figura 4.

The screenshot shows a Notepad window with a CSV file. The columns are: angle, brightness, circle, class, compacity, div4_1, div4_3, entropy1, entropy2, entropy3, entropy4, maxpixel1, maxpixel2. The data rows consist of numerical values followed by the class name, such as 'amiante' or 'ceramica_clara'.

Fig. 4 – juntada dos dados por meio do bloco de nota para leitura do minerador WEKA.

Etapa 5: após a juntada dos arquivos .csv de cada classe, a leitura foi feita pelo sistema WEKA, por meio do algoritmo J48, conforme figura 5.

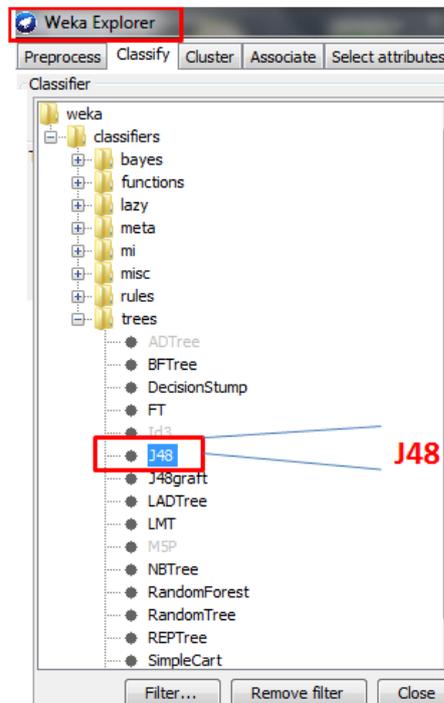


Fig. 5 – leitura do arquivo .csv por meio do algoritmo J48

Etapa 6: o minerador WEKA, por meio do algoritmo J48, disponibilizou a árvore de decisão com valores de limiares intercalados.

Resultado

O resultado apresentou uma árvore de decisão com valores de limiares intercalados. Esses valores são os limites de cada classe de interesse.

A Figura 6 é uma porção da árvore de decisão criada pelo minerador WEKA.

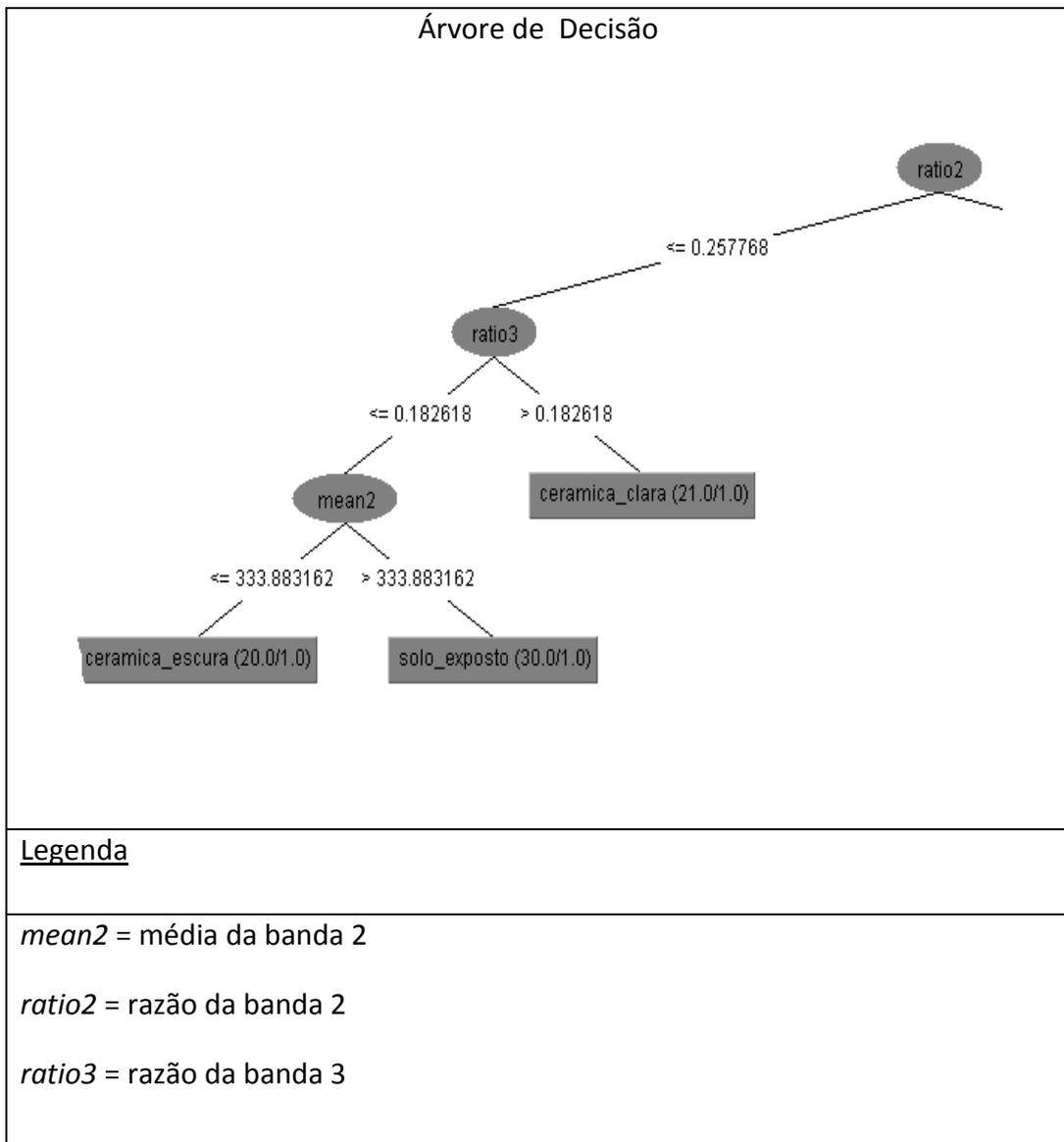


Fig. 6 - Árvore de decisão. Fonte: Sistema WEKA.

Considerações finais

A técnica de mineração de dados por meio do algoritmo J48 implementado no sistema minerador WEKA automatizou a definição de valores de limiares por meio de uma árvore de decisão. Esses valores podem ser expressados de forma intercalada e aplicados em uma classificação orientada a objetos de imagem de sensoriamento remoto.

Referências

MILLER, Harvey J. Miller; HAN Jiawei. **Geographic data mining and knowledge discovery**. 2. ed. Minneapolis, Minnesota, USA: Chapman & Hall/CRC, 2009.

MATHUR, Nishant; KUMAR, Sumit; KUMAR Santosh e JINDAL Rajni. **The base strategy for ID3 algorithm of data mining using havrda and charvat entropy based on decision tree**. International Journal of Information and Electronics Engineering, v. 2, n. 2, março/2012.

QUONIAN, Luc et al. **Inteligência obtida pela aplicação de *data mining* em base de teses francesas sobre o Brasil**. Ci. Inf., Brasília, v. 30, n. 2, p. 20-28, maio/ago. 2001.