GeoDMA – A software for geographical pattern recognition

Thales Sehn Korting, Leila Fonseca, Gilberto Camara

Brazil's National Institute for Space Research
Image Processing Division
Av. dos Astronautas 1758
São José dos Campos, SP, Brazil

Abstract

This work presents a novel software for pattern recognition applied to geographical and remote sensing data. GeoDMA stands for Geographical Data Mining Analyst, and is a free software designed to include data mining techniques to aid domain-experts in the detection of land cover and land use patterns in geographical databases. For this purpose, GeoDMA includes spatial and spectral features extraction, visual data analysis and supervised and unsupervised classification algorithms. GeoDMA has already been applied to detect intra-urban classes of land cover and to detect patterns of deforestation in the Brazilian Amazon.

Keywords: remote sensing, data mining, decision trees, self-organizing maps, neural networks

Software availability

Software name: GeoDMA
Contact: `tkorting@dpi.inpe.br`
Year first available: 2009
Hardware: IBM compatible PC
Operation System: Linux/Windows
Program language: C++
Availability: free from `http://geodma.sf.net/`

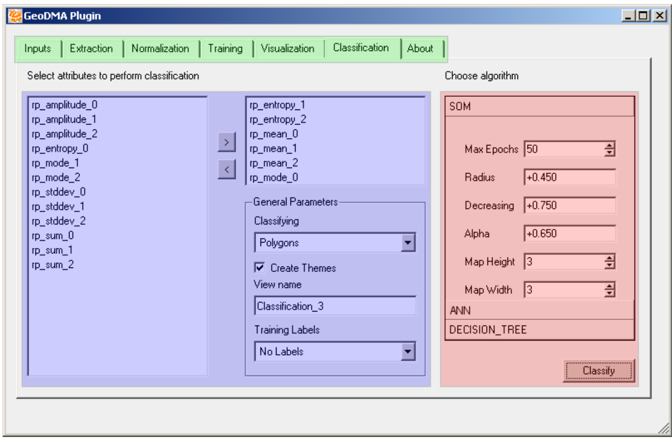

Figure 1: GeoDMA's GUI, Classification tab. Green area contains data mining steps, Blue area shows the available object features, and the Red area describes the functions to be applied into the data.

1. Introduction

In this paper we describe a novel software for remote sensing image analysis based on data mining techniques, called Geographical Data Mining Analyst – GeoDMA. It has been developed by the Image Processing Division at INPE (Brazil's National Institute for Space Research).

Tools for easy data manipulation and pattern recognition are useful in many remote sensing applications, like urban planning or deforestation detection. The increasingly amount of remote sensing data brings the urgent need of computational tools to deal with it. GeoDMA is a data mining software designed to be as generic as possible. Its intent is to detect patterns in geographical information databases. The system provides a flexible and friendly Graphical User Interface – GUI, as shown in Figure 1.

2. Mining geographical data

GeoDMA was built to work with different types of remote sensing data, like satellite images or cadastral data. When the user provides an image, it is segmented into a set of regions, and they are classified based on spectral and spatial features. Haralick and Shapiro [1] have established that a good segmentation should separate the image into simple regions with homogeneous behavior. It must represent commonly found targets in images, like roofs [2], streets [3] and trees in the urban context. In agricultural [4] and forest applications [5], the regions shall be related to crops and deforested areas, respectively.

Two techniques for image segmentation are available in GeoDMA. The first one is the region growing algorithm, based on [6]. The second algorithm is based on [7], in which the average image objects size must be free adaptable to the scale of interest. If the user provides cadastral data, it is possible to classify using different features, including topological and contextual relations. The system

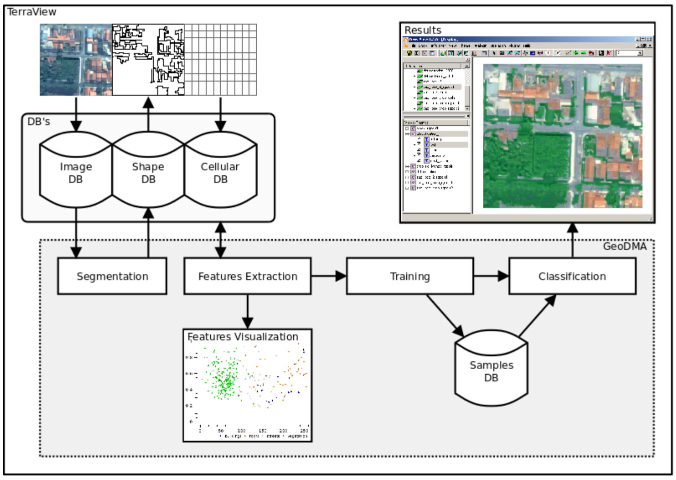

Figure 2: GeoDMA's framework. From the database with images or objects, GeoDMA is able to perform feature extraction, training for supervised classification and visualization of the results.

works with landscape metrics as well, gathering cellular spaces with raster or vector information. Figure 2 shows GeoDMA's framework.

Miller et al. related four main data mining steps, namely background, preprocessing, data mining and knowledge construction [8]. Such steps must be flexible and iterative, to improve the overall accuracy of every analysis. According to Takatsuka and Gahegan [9], the lack of a single package to support and integrate these steps over geographical information represents a noticeable bottleneck.

However, GeoDMA includes such steps. Preprocessing includes data selection, cleaning and reduction, searching for outliers and redundant data. With feature space visualization, it is possible to see data correlation and choose a set of features for best classification. Visualization supports data preprocessing, selection of data mining tasks and techniques [10]. Using the training tool, the domain-expert assigns representative data samples to each interest class.

Classification involves mapping data into specific categories whose cardinality is fewer than the number of data objects [8]. The Self-Organizing Maps algorithm [11], available in GeoDMA, has the ability to group or typify automatically data according to different properties [12]. Supervised algorithms uses a priori knowledge created in the training step. It has been proved that experts often feel more at ease providing samples than developing rigid model-based preference rules. GeoDMA incorporates two supervised algorithms, namely the structural classifier C4.5 [13], which uses a decision tree to split the data into different classes, and the Artificial Neural Networks, using the McCulloch-Pitts model [14].

3. Conclusions

Openshaw [15] stated that the development of geographical data mining tools are a major research theme. Han et al. stated that data mining and exploratory uses of information hold much promise [16]. In this scenario we presented GeoDMA, a free software developed to perform data mining analysis with geographical data.

Further research includes the implementation of multitemporal analysis, combining several periods in a time series. Within this approach we believe it is possible to infer automatically about land use and land cover changes occurred in the space.

References

[1] R. Haralick, L. Shapiro, Image segmentation techniques., Applications of Artificial Intelligence II., 1985, 548 (1985) 2–9.

[2] A. L. Chesnel, R. Binet, L. Wald, Object oriented assessment of damage due to natural disaster using very high resolution images, in: IEEE International Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007, 3736–3739, 2007.

[3] Y. He, H. Wang, B. Zhang, Color based road detection in urban traffic scenes, 2003 IEEE Intelligent Transportation Systems, 2003. Proceedings 1.

[4] A. Perez, J. Benlloch, F. Lopez, S. Christensen, Colour and shape analysis techniques for weed detection in cereal fields, Computers and electronics in agriculture 25 (2000) 197–212.

[5] M. Silva, M. Escada, G. Câmara, R. Souza, Remote-sensing image mining: detecting agents of land-use change in tropical forest areas, International Journal of Remote Sensing 29 (2008) 4803–4822.

[6] L. Bins, L. Fonseca, G. Erthal, F. Li, Satellite imagery segmentation: a region growing approach, Simpósio Brasileiro de Sensoriamento Remoto 8.

[7] M. Baatz, A. Schape, Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation, in: Wichmann-Verlag (Ed.), XII Angewandte Geographische Informationsverarbeitung, Heidelberg, 2000.

[8] H. J. Miller, J. Han, Geographic data mining and knowledge discovery, CRC Press, URL `http://books.google.com.br/books?id=fJTD26iYxpkC`, 2001.

[9] M. Takatsuka, GeoVISTA Studio: a codeless visual programming environment for geoscientific data analysis and visualization, Computers and Geosciences 28 (10) (2002) 1131–1144, ISSN 00983004, doi:\bibinfo{doi}{10.1016/S0098-3004(02)00031-6}.

[10] D. Keim, H. Kriegel, Using visualization to support data mining of large existing databases, Lecture Notes in Computer Science 871 (1994) 210.

[11] T. Kohonen, Self-Organizing Maps Springer, Springer, URL `http://books.google.com.br/books?id=e4igHzyfO78C`, 2001.

[12] A. Kalteh, P. Hjorth, R. Berndtsson, Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application, Environmental Modelling and Software 23 (2008) 835–845.

[13] J. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, CA, URL `http://books.google.com/books?id=HExncpjbYroC`, 1993.

[14] S. Haykin, Neural networks and learning machines, Prentice Hall, 2008.

[15] S. Openshaw, Geographical data mining: key design issues, Proceedings of GeoComputation 99.

[16] J. Han, M. Kamber, Data mining: concepts and techniques, Morgan Kaufmann, URL `http://books.google.com/books?id=AfL0t-YzOrEC`, 2006.