ELSEVIER

Computers & Geosciences 34 (2008) 127–143

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

Support vector machine for 3D modelling from sparse geological information of various origins

Alex Smirnoff*, Eric Boisvert, Serge J. Paradis

Natural Resources Canada, Earth Sciences Sector, 490, de la Couronne Québec, Que., Canada G1K 9A91

Received 26 July 2006; received in revised form 1 December 2006; accepted 28 December 2006

Abstract

Three-dimensional (3D) geological models are a powerful way of visualization, analysis and interpretation of geological information. However, manual modelling with available GIS tools is a challenging and time-consuming task. Here we propose the use of the support vector machine (SVM) in order to automate the creation of such models. We experiment with various input data and hyperparameters in order to demonstrate that the SVM can be efficiently applied in 3D geological reconstructions overcoming some limitations of previously used methods.

Keywords: SVM; Support vector machine; Gocad; 3D modelling; Geological reconstruction

1. Introduction

Often geologists are faced with a variety of information requiring generalization and analysis. Three-dimensional (3D) modelling software packages such as Gocad® of Earth Decision Sciences have proven an excellent means for data presentation and interpretation. The modelling procedure normally requires reconstruction of individual geological layers using surfaces interpolated from control points with subsequent fusion of these units into a single model. The popular interpolation techniques include inverse distance weighting (IDW), discrete smooth interpolation (DSI) and various flavors of kriging preceded by semi-variogram analysis.

The above multi-step procedure can easily become a tedious and time-consuming task when a complex geomodel is considered. In addition, the traditional interpolation methods assume reasonable areal coverage of the input data. Therefore, there is a strong need for an algorithm that would automate the process of model construction by eliminating the need for building individual geological interfaces. This algorithm should also be able to handle cases when only a few pieces of information on regional geology are available. These might be occasional borehole data, a surface geology map, several cross-sections through multiple geological strata re-created from well drilling or seismic profiling or any combination of the above. Finding such an algorithm and testing its performance on available datasets was the objective of this study.

*Corresponding author. Tel.: +1 418 654 3716; fax: +1 418 654 2615.

E-mail address: alsmirno@nrcan.gc.ca (A. Smirnoff).

doi:10.1016/j.cageo.2006.12.008

A number of relevant algorithms have been proposed over the recent years. Some of them are capable of re-connecting cross-sections in 3D space and thus re-creating the original body shape. These are primarily known from shape-based modelling and various morphing techniques. In medicine, such an approach has been applied to reconstructions of various human organs from multiple ultrasound sections (e.g., Treece et al., 1998, 1999, 2000). Treece et al. (1998) demonstrated that the success of these reconstructions varies depending on the number of input sections, their position, shape of the original body and the applied interpolation technique. This method, however, performs adequately when complete sections of a single body are considered as the only input data and normally requires a large number of those in order to reproduce a reasonable 3D shape.

Another set of more versatile algorithms has been applied specifically in geology. Courrioux et al. (2001) reviewed these methods and pointed out their limitations. As an alternative, they proposed 3D space partitioning between geological units based on Voronoi cells. The most interesting aspect of this approach is its comprehensive and automatic character. All volumes are built automatically and simultaneously, which allows the process of construction of individual geological interfaces to be avoided. However, this approach requires preliminary discretization of data along the interfaces. In addition, the final results were shown to be sensitive to the way the data have been discretized. The method also requires subsequent smoothing (Courrioux et al., 2001).

Another automation technique was described by Chiles et al. (2004). It is based on the potential-field theory. A set of iso-potential surfaces representing contacts between geological units is drawn on a scalar potential field previously interpolated by universal cokriging from input structural data. This method is best suited for layered geology and just like the previous one, heavily relies on contact information and orientation data (strike and dip). Both algorithms found their implementation in 3D GeoModeller developed by Bureau de Recherches Géologiques et Minières of France (BRGM; Lajaunie et al., 1997), which resulted in certain inherited limitations of this software. Hence, to our knowledge, no universal algorithm capable of automatically building a multi-unit 3D model from sparse data of diverse origin has been applied in geology so far.

Here we propose the use of the support vector machine (SVM), a tool routinely applied in the field of image analysis and pattern recognition, for solving this task. The SVM is becoming increasingly popular and has successfully been used to solve classification and regression problems in biology (e.g., Noble et al., 2005), hydrology (e.g., Yu et al., 2004), medicine (e.g., El-Naqa et al., 2002) and environmental science (e.g., Gilardi et al., 1999). Sharifzadeh et al. (2005) tested the SVM along with other classification methods for creating 2D thematic maps from labelled geospatial information. Their results clearly showed that the SVM outperforms such methods as nearest-neighbor and discriminant analysis.

In this study, after a brief introduction of the SVM algorithm, we demonstrate that it allows us to simultaneously (all layers at once) combine sparse data from different sources without complicated pre- or post-processing and supplementary geological information, thus overcoming certain limitations of the previously described methods. To do so, we experiment with two datasets including various numbers of geological units in study areas with different topographies and geometries. A comparison with an interpreted model built through sequential layer interpolation shows that multiple units with similar shape, surface area and volume can be reconstructed in a single SVM prediction step.

2. Methodology

2.1. The SVM algorithm

The SVM algorithm is based on the Statistical Learning Theory described by Vapnik (1995). It uses a set of examples with known class information to build a linear hyperplane separating samples of different classes. This initial dataset is known as a training set and every sample within it is characterized by features upon which the classification is based. In the machine learning theory, this is known as supervised learning as opposed to unsupervised learning when no a priori class information is available. In more complicated, non-linear cases, the task of discovering the best separator is turned into a linear task by transferring input data into a higher-dimensional space known as the feature space. Various kernel functions are normally employed for this transfer.