Pattern-based Simulation using Self Organized Maps

Snehamoy Chatterjee [a] and Roussos Dimitrakopoulos [b]

[a] *Department of Mining Engineering, NIT Rourkela, Orissa-769008, India*

[b] COSMO – Stochastic Mine Planning Laboratory; Department of Mining and Materials Engineering, McGill University, Montreal, Quebec, H3A 2A7, Canada.

Introduction

The simulation of complex geology and heterogeneous orebodies is always a difficult and challenging job. There are a number of multi-point simulation techniques proposed like *simpat* (Arpat and Cares 2007), *snesim* (Strebelle 2002), *filtersim* (Zhang et al. 2006), cumulant-based simulation (Mustapha and Dimitrakopoulos 2010), and wavelet-based simulation (Chatterjee et al. 2012). Out of these multi-point simulation techniques, *simpat*, *filtersim* and *wavelet*-based simulation are pattern simulation algorithms. In pattern-based multi-point simulations, multi-point conditioning data is compared with the patterns, a set of values spatially distributed over a given template of spatial locations, of the training image. The best matching pattern from the pattern database with the conditioning data event is searched. The computational time of such algorithms is very high. In this work, a self-organized maps (SOM) based algorithm is proposed for pattern-based simulation. The dimensional reduction of the pattern database is performed using the SOM. In this work, we have applied a topology preservation algorithm to select the optimum topology of SOM. The optimum cluster number was selected by calculating the Davies-Bouldin index (Davies and Bouldin 1979).

Method

Define $ti(u)$ as the value of the training image ***ti*** where $u \in G_{ti}$ and G_{ti} is the regular Cartesian grid discretizing the training image, $ti_T(u)$ indicates a specific multiple-point vector of $ti(u)$ within a template ***T*** centered at node u, that is

$$ti_T(u)=\left\{ti(u+h_1), ti(u+h_2), ..., ti(u+h_\alpha), ..., ti(u+h_{nT})\right\} \tag{1}$$

where, the h_α vectors are the vectors defining the geometry of the n_T nodes of template ***T*** and $\alpha = \{1, 2, ..., n_T\}$. The pattern database generated training image is defined as ***patdbT***.

The SOM is a neural network algorithm that can be used to reduce the amount of data and to project the data nonlinearly onto a lower-dimensional display. The basic mechanism in the SOM network is competitive learning. Each neuron or nodes *j=1, 2,..., L* has a n_T dimensional synaptic weight vector $W_j = [w_{j1}, ..., w_{jn_T}]^T$. Initially, the reference vector *Wj(0)* has randomly distributed components. Then the reference vector is compared with the input patterns $ti_T(u)$, and the distances between the input and reference patterns are derived. The best-matching unit (BMU) is the neuron whose reference pattern is nearest to the input. At the next step, the reference vectors are updated. The learning algorithm stops when the algorithm reaches its maximum iteration. The main characteristic of SOM mapping is the preservation of neighbourhood relations. The similar pattern vectors in the input space are mapped onto neighbouring locations in the output space. We have used quantization error (Q.E) to evaluate the quality of the map and select the optimal one to represent our patterns. The quantization error measures the average distance between each pattern and its best matching unit (BMU). Several indicators of topological errors have been used to control the conservation of topology (Goodhill and Sejnowski 1996). One of the indices for control the conservation of topology is the topographic error (T.E). This error measures the proportion of all pattern vectors for which first and second best-matching units (BMU) are not adjacent vectors. So the lower the topographic error, the better the Self-Organizing Map preserves the topology. In this work, we have used an error function (total error) which can accommodate both the error terms together. The prototypes of SOM are combined to form the actual clusters. We have applied a *k*-means clustering algorithm. To select the optimum cluster number, the Davies–Bouldin index (Davies and Bouldin 1979) is calculated.

After classifying the ***patdbT*** using the SOM-based classification algorithm, simulation is carried out. The similarity between the conditioning data and the class prototype are determined. A sequential simulation algorithm (Goovearts 1997) is used for pattern-based simulation. At each visited node, a conditioning data event is obtained by placing the same template used for pattern extraction from the training image, centering at the node to be simulated. The similarity between the conditioning data and class prototype are calculated by a distance function.

3. Application of the proposed method

The proposed algorithm is tested and validated by simulating known exhaustive categorical and continuous data sets. The results of our proposed approach are compared with *filtersim* results to make a valid comparison. Only one categorical unconditional simulation is presented herein; for more results see Chatterjee and Dimitrakopoulos (2011). A categorical unconditional simulation is performed for two-category data sets. The binary training image consists of sand and non-sand materials. The training image is presented in Figure 2 (a). The patterns from the training image are extracted using a 9×9 template. The generated patterns are then projected into a 2-dimensional plane using the self-organized map algorithm. The optimum map size for the binary training image is 14×14. The optimum cluster number 84. After classifying the patterns, an unconditional simulation is performed. The template size is 9×9 and the inner patch size is 5×5. The parameters used for the simulation in our approach and *filtersim* are the same except for the cluster number. However, to make a valid comparison, we have chosen 150 for the *filtersim* algorithm. The cluster number for our proposed approach is 83, which is optimally selected. Two different unconditionally generated realizations using our proposed method and *filtersim* are presented in Figure 2. It is observed from the figures that our proposed algorithm can reproduce the continuity of channels presented in the training image better than *filtersim*.

Conclusions

The projection of high dimensional data to a 2-dimensional space is performed in this paper using self organized maps to reduce the dimension. The SOM helps to reduce the dimensions of the data to visualise as well as to map the same types of patterns under the same reference vector. The main advantage of SOM over other multi-dimensional scaling (MDS) algorithms is that SOM preserves the topology of the original high-dimensional data. Topology preservation helps to map the patterns in such a way that the more similar patterns in high dimensional space will be placed in the neighbouring reference vectors in the 2-D SOM. Thus, topology preservation helps to guide the pattern classification of the projected data. One unconditional example is presented here to show the performance of our proposed approach. The comparative study with *filtersim* reveals that our proposed algorithm out performs the *filtersim* algorithm.

(a) Training Image (b) Realisation of our method (c) *Realisation of filtersim*

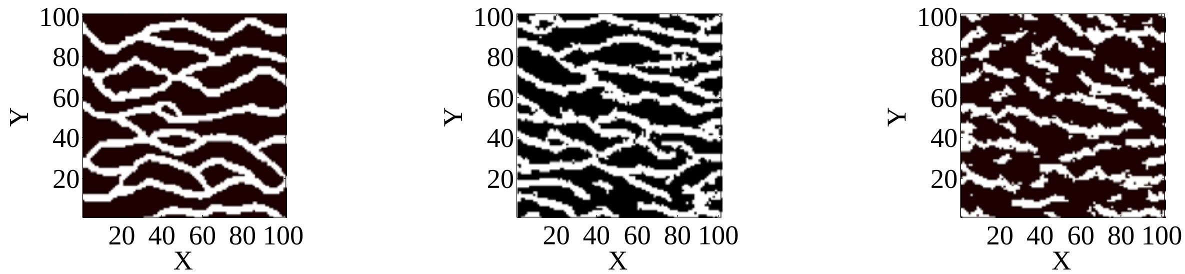

Figure 2 Training image and simulated realisations of our proposed method and *filtersim*

References

Arpat G, and Caers J (2007) Conditional simulation with patterns. Mathematical Geology 39 (2):177-203.

Chatterjee, S., Dimitrakopoulos, R. (2011) Multi-scale stochastic simulation with a wavelet-based approach. Computers & Geosciences doi:10.1016/j.cageo.2011.11.006

Chatterjee, S., Dimitrakopoulos, R. (2011) Pattern-based Simulation using Self Organized Maps, COSMO Research Report 5.

Davies DL, and Bouldin DW (1979) A cluster separation measure. IEEE Trans. Patt. Anal. Machine Intell. PAMI-1: 224–227.

Goodhill GJ, Sejnowski TJ. (1996) Quantifying neighbourhood preservation in topographic mappings. Proceedings of the 3rd Joint Symposium on Neural Computation. La Jolla. CA. 6: 61-82.

Goovaert P (1997) Geostatistics for Natural Resources Evaluation (Applied Geostatistics Series). Oxford University Press, Oxford.

Mustapha H, Dimitrakopoulos R (2010) High-order stochastic simulations for complex non-Gaussian and non-linear geological patterns, Mathematical Geosciences, 42 (5).

Strebelle S (2002) Conditional simulation of complex geological structures using multiplepoint statistics. Mathematical Geology 34 (1): 1-21.

Zhang T, Switzer P, and Journel A (2006) Filter-based classification of training image patterns for spatial simulation. Mathematical Geology 38(1): 63–80.