Knowledge Recovery for Continental-Scale Mineral Exploration by Neural Networks

Laurent Bougrain, Maria Gonzalez, Vincent Bouchot, Daniel Cassard, Andor Lips, Frédéric Alexandre, Gilbert Stein

Knowledge Recovery for Continental-Scale Mineral Exploration by Neural Networks

Laurent Bougrain,[1] Maria Gonzalez,[2] Vincent Bouchot,[2] Daniel Cassard,[2] Andor L. W. Lips,[2] Frédéric Alexandre,[1] and Gilbert Stein[2]

This study is concerned with understanding of the formation of ore deposits (precious and base metals) and contributes to the exploration and discovery of new occurrences using artificial neural networks. From the different digital data sets available in BRGM's GIS Andes (a comprehensive metallogenic continental-scale Geographic Information System) 25 attributes are identified as known factors or potential factors controlling the formation of gold deposits in the Andes Cordillera. Various multilayer perceptrons were applied to discriminate possible ore deposits from barren sites. Subsequently, because artificial neural networks can be used to construct a revised model for knowledge extraction, the optimal brain damage algorithm by LeCun was applied to order the 25 attributes by their relevance to the classification. The approach demonstrates how neural networks can be used efficiently in a practical problem of mineral exploration, where general domain knowledge alone is insufficient to satisfactorily model the potential controls on deposit formation using the available information in continent-scale information systems.

KEY WORDS: Artificial neural networks, variable selection, pruning algorithm, geographic information system, metallogenic research.

INTRODUCTION

In the last five years, continent-scale information systems have been constructed by BRGM to serve metallogenic and environmental research (GIS Andes, (Cassard, 2000); GIS Africa, (Milesi and others, 2001); GIS Central Europe, (Cassard and others, 2001); GIS Urals, Leistel and others, in press). These information systems are composed of spatially referenced geographical, geological, and mineral-deposit thematic layers, and their respective attribute data (see Fig. 1). They are used primarily to establish insights in the region's mineral potential, its past and future mining activities, and its related environmental vulnerability. The information systems can be exploited further to derive new rules between different attributes in relation to mineral deposit formation and the spatial distributions of the deposits. Different methods are applied to define a region's mineral potential, to assign weights to the different attributes, and to derive new rules. The methods range from data-driven approaches to knowledge-driven approaches and also involve hybrid approaches.

[1] LORIA/INRIA Lorraine, CORTEX, Campus scientifique, BP 239, 54506 Vandœure-lès-Nancy, France, e-mail: bougrain@loria.fr.

[2] BRGM, Mineral Resources Division, BP 6009, 45060 Orléans cedex02, France; e-mail: v.bouchot@brgm.fr.

MINERAL-POTENTIAL MAPPING

Mineral-potential mapping subdivides areas according to metal or mineral favorability (probability of deposit) and may be carried out in commercial exploration programs or in governmental mineral-resource assessments. The ability of a Geographic Information System to combine spatial data from different sources contributes significantly to identifying spatial associations of the data, and to using models for analysis and prediction of

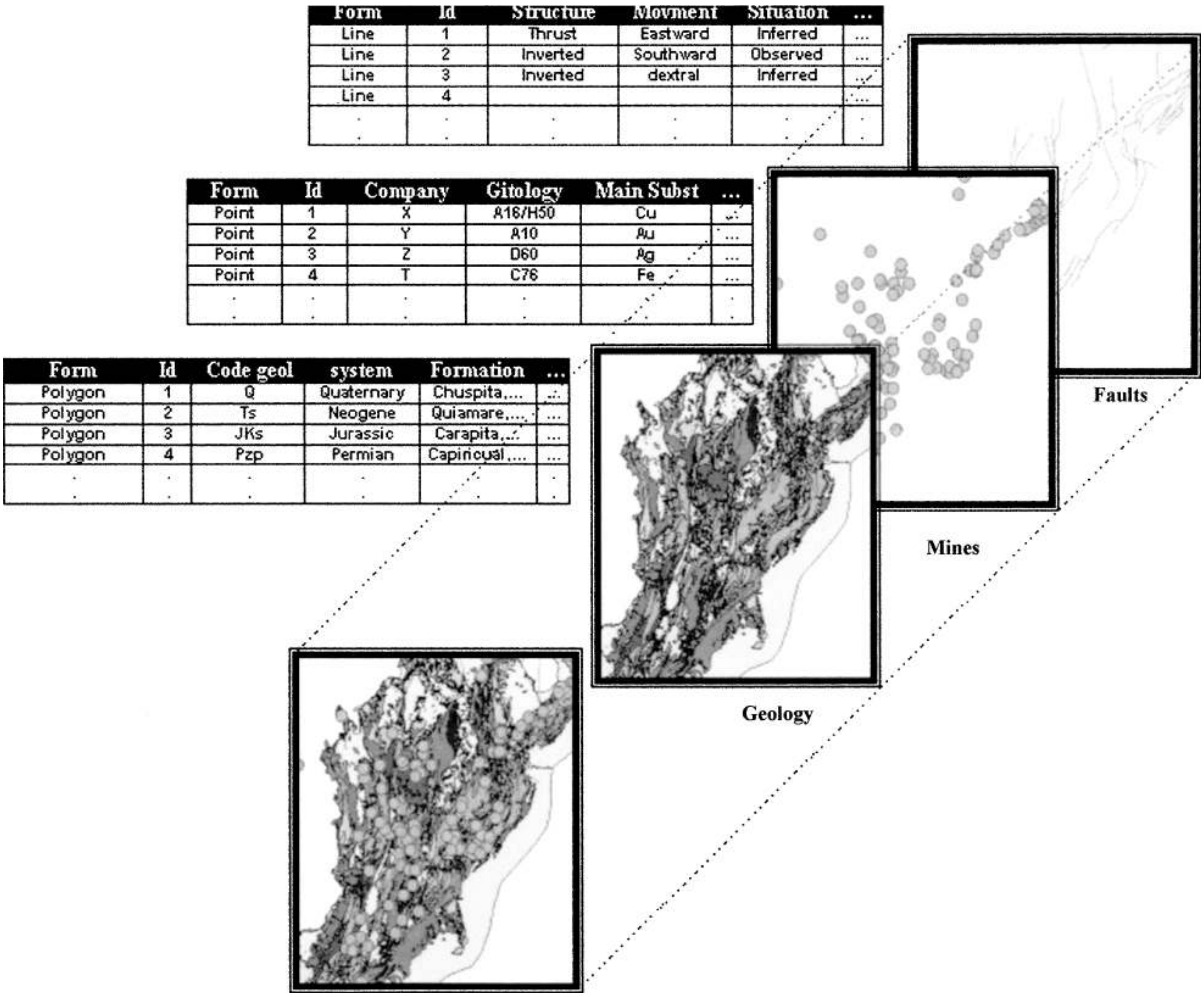

Figure 1. Composite "Metallogenic" layer with its constituent layers: geology (polygons), mines (points), and faults (lines), and related attributes tables.

the spatial phenomena (Bonham-Carter, Agterberg and Wright, 1989; Bonham-Carter, 1994; Burrough and McDonnell, 1998). The dedication of a GIS to mineral-potential mapping is classically a three-step approach (Bonham-Carter, 1994; Knox-Robinson and Wyborn, 1997): (i) data identification and organization, (ii) data quantification and processing, and (iii) data integration and modeling. The evolutions of Data Mining and Artificial Neural Networking techniques allow an alternative approach. The latter represents an adaptive computing system that can learn from the data in order to conduct specific tasks. Unlike the conventional three-step approach, a neural network can analyze all data simultaneously.

In the past decade, mineral-potential mapping using a GIS has been applied by various research groups and involved different ore deposit models. Models for predicting mineral potential, based on statistical relationships or on heuristic relationships, are examples of empirical models in contrast to conceptual models (Knox-Robinson and Wyborn, 1997). The assignments of weights to the different attributes can either be carried out using statistical criteria, or the weights can be estimated on the basis of expert opinion. The approaches can be subdivided into data-driven and knowledge-driven models (Bonham-Carter, 1994; Harris and others, 2001). In data-driven modeling, the various input maps are combined using methods such as logistic regression, weights of evidence, or neural networks. Data-driven approaches require that "a prior" knowledge exists in the form of known mineral deposits or occurrences and barren areas in the study area. Knowledge-driven approaches, which rely on the geologist to weight the importance of different data layers, may include the use of fuzzy logic, Bayesian probability and Dempster-Shafer belief theory (Bonham-Carter, 1994). Conceptual mineral deposit models, containing all typical characteristics of a certain type of deposit, may contribute in data selection and data modeling, in helping to decide which features to enhance and extract as evidence, and in helping to decide how to weigh the relative importance of evidence.

The principal advantages of weights of evidence modeling is that the method is well defined, reproducible, objective (avoids the subjective choice of weighting factors), and provides a quantitative measure of confidence (Raines, 1999). Fuzzy sets (Zadeh, 1965) have been introduced to handle with inexact concepts in a definable way (e.g. Burrough and McDonnell, 1998). Fuzzy set theory has been applied to mineral exploration (e.g., An, Moon, and Rencz, 1991) demonstrating that the method can represent adequately and manipulate imprecise and incomplete information. Data mining is defined as a process of extracting implicit unknown and potentially useful data from a GIS (Salleb and Vrain, 2000). Similar to other statistical approaches, a neural network requires the construction of a training set that is representative of both deposits and nondeposits in order to allow discrimination between the two modalities.

NEURAL NETWORKS APPLIED TO GIS ANDES

GIS Andes, A Continental-Scale GIS

GIS Andes is a homogeneous information system of the entire Andes Cordillera, covering an area of 3.83 million km^2 and extending for some 8,500 km from the Guajira Peninsula (northern Colombia) to Cape Horn (Tierra del Fuego). Conceived as a tool for both the mining sector, where it is an aid to minerals exploration and development, and the academic sector where it is an aid to developing new metallogenic models, GIS Andes is based on original syntheses and compilations (Cassard, 1999).

The following different layers of the system are available and can be combined in any way that the user sees fit:

- ➢ Geographic: a geographic base (DCW®);
- ➢ DEM: digital elevation models with a structural analysis of the topography;
- ➢ Imagery: SPOT 4 VEGETATION® images;
- ➢ Geological synthesis: geological map of the Andes at 1:2,000,000 scale;
- ➢ Geologic map coverage: more than 1,100 georeferenced maps;
- ➢ Seismic: more than 50,000 seismic records, with modeling of the subduction zone (for example, Eg. 2);
- ➢ Volcanic: data on Holocene volcanism;
- ➢ Gravimetric: the Bouguer anomaly calculation; isostatic correction and corresponding residual anomalies; vertical gradient calculation and structural analysis; gravity modeling of the Nazca plate;
- ➢ Heat flow: a base with oceanic and continental data;
- ➢ Geochemistry: a database containing 3,935 whole-rock analyses;
- ➢ Ore deposits: linked to a Database under Access® and using a new metallogenic lexicon;
- ➢ Mineralogy, fluid inclusions, isotopes: data on the main ore deposits of the Cordillera.

The data used by the artificial neural networks in this study were extracted from these data layers.

Attributes and classes

The present study included an analysis of the GIS Andes data that enabled a specific model to be compiled for recent Andean ore deposits (15 to 2 Ma) linked to the recent evolution of the subduction zone (Fig. 2) that underlies the entire cordillera and may be a continental-scale ore localizing phenomenon. This specific model concerns the development of highly prospected ore deposits (Au, Ag, Cu . . .) belonging to epithermal-porphyry systems related to active margins. The model revealed a set of attributes considered to have, more or less, a relationship to the spatial and temporal distribution of the recent epithermal-porphyry Andean deposits.

The main attributes retained were:

- Type and age of the country rock hosting the deposits,
- Proximity of the deposit to a regional fault distinguished by its strike,

- Density and focal depth of earthquakes immediately below the deposit,
- Proximity of active volcanoes (indicative of active hydrothermal systems),
- Geometry of the subduction zone: depth of subducted plate below the ore deposit, dip of the subduction zone immediately below the deposit, distance to the subduction trench,
- Position (longitude, latitude, altitude) of the deposit.

To train the network optimally, two variables were selected for the study; that is, "deposit" and "barren".

Under the "deposit" variable, a total of 398 recent (2 to 15 Ma) Andean Au, Ag, Cu, Mo, Sb, Hg, Pb, and Zn -bearing deposits (epithermal-porphyry type), documented as points, were extracted from GIS Andes.

The "barren" database contains 243 entries. Although the notion of "barren site or nondeposit" has no geological meaning or significance as such, it is nevertheless of extreme importance to an optimal learning exercise of the network. For a correct "learning," the neural network needs to be supplied with a database of examples that are not deposits and that, by their opposition or contrast to deposits, enable the network to adapt its weight to the task of discriminating a deposit from a nondeposit. If a deposit covers about 1 km^2 on average, then any randomly selected 1 km^2 cell in the study area has a high probability of belonging to the barren population because known deposits represent 398/2.307 million km^2 or about 0.017% of possible cells. The remaining 99.983% of the cells actually represent barren areas, areas possibly containing "not well defined" (age and/or type) deposits and occurrences not taken into account in this study, and a small proportion of undiscovered deposits. Thus, the method adopted to create the barren database consisted in selecting arbitrary points from the areas well outside the mining districts and where, as far as known, no deposits have been discovered. Because these points were georeferenced, their values relative to the retained attributes could be obtained by crossing the different layers of GIS Andes.

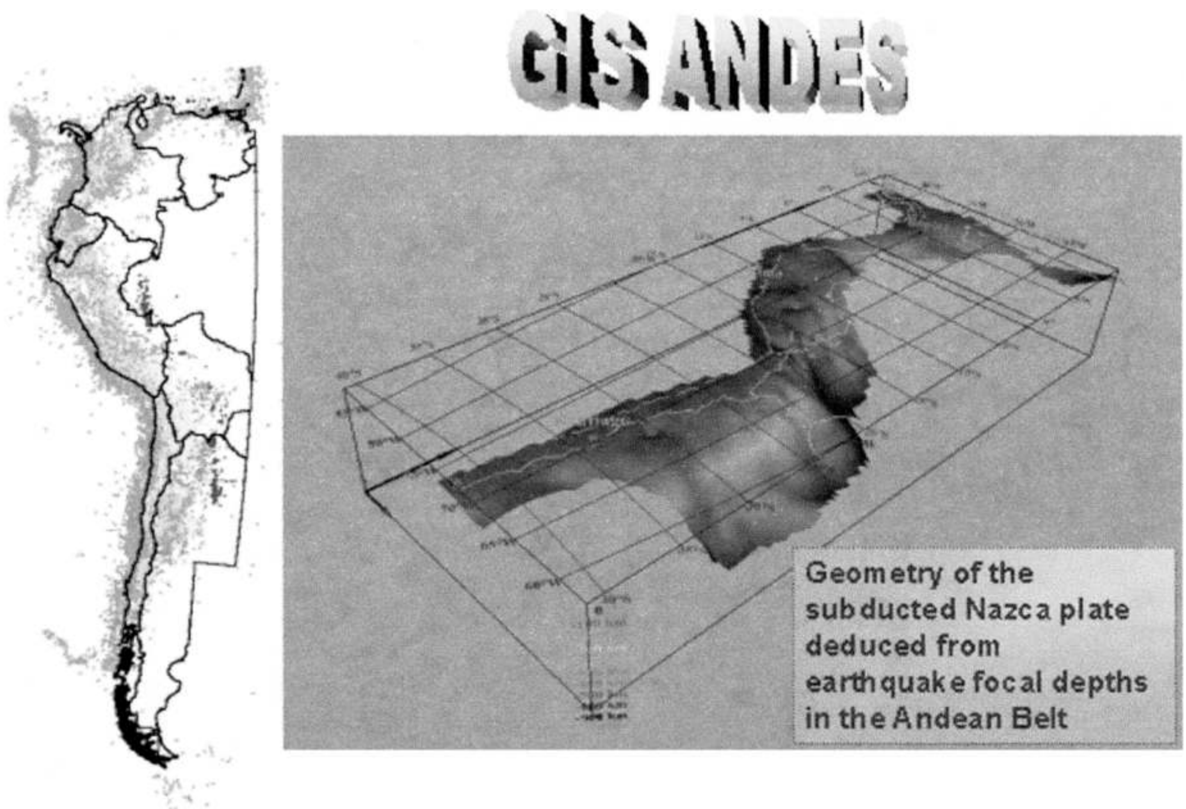

Figure 2. Seismicity layer of GIS Andes and modeling of subduction zone.

In the end, the data set comprises 641 examples (398 epithermal-porphyry deposits and 243 barren sites) available for the neural network's learning phase.

Classification

Artificial neural networks constitute good tools for this type of computational modeling. Artificial neural networks, such as multilayer perceptrons, can learn how to classify patterns described by a set of attributes into different classes using a set of examples as supervised learning. In this situation, the correct class is given and a function calculates the difference or error between the desired class and the estimate of the network. Then a learning algorithm, for example the back-propagation algorithm, uses this difference to modify the weights of the network with the aim of reducing the error. This process continues until all of the training data have been learned by the network.

A part of the database should be used to learn the model (learning set) and the other part should be used to test its performance in a generalization test set (i.e. how the network classifies patterns it has never seen during the learning stage). In this way, it is possible to handle the risk of overtraining of the network. This phenomenon occurs when the number of weights is too large compared to the number of learning examples that allows the network to learn idiosyncrasies of the examples without learning the general model that really links the inputs to the outputs. Results obtained by the network can fluctuate with different selected learning examples and the size of the leaning set. To improve the confidence in the results, it is mandatory to test the performance of the network using different learning sets. The cross-validation is a special situation of this procedure. The database is split up into *n* subsets. Each subset is used as a test set, the other ones are gathered into the learning set. Finally, the average of the results is used as a measure of the performance of the model.

For a n-classification task by artificial neural networks, the target class should be coded in a n-binary output vector with a value of 1 for the class index and 0 for the others. Using the Gibbs' activation function [Eq. (1)], also termed the softmax function, the outputs of the network predict the *a posteriori* probabilities that a pattern (sample) belongs to each class.

$$p(k_j|x) = \frac{e^{a_j}}{\sum_j e^{a_j}} \tag{1}$$

where k_j is the class identified by the output neuron j, $j \in [1, n]$ and a_j is the potential (i.e. the input) of this neuron.

Various multilayer perceptrons were applied with 25 input variables, one hidden layer and 2 output units using a softmax function to predict the *a posteriori* probability that a new site belongs to the "deposit" or to the "barren" class.

The final objective to be attained was twofold:

(1) Improve and quantify the established "deposit" model, after ranking the input attributes on the neural network;
(2) Prepare intermediate-term predictive mineral maps once the network had completed its learning.

Automatic attribute selection

This real-world problem is so complex that, at the moment, the general domain knowledge does not allow one to understand fully or predict ore deposit formation. Good performances obtained by artificial neural networks can lead us to extract additional knowledge which had not yet been incorporated in existing ore deposit models. With the aim of reducing the size of the model and extracting knowledge about which attributes to use, we used a pruning method (Reed, 1993; Cibas and others, 1996). The operating procedure of the pruning method is to remove connections or neurons, or both, which have less influence on the system answer. The goals to reach are many. From a performance point of view, reducing the number of variables (weights) reduces the chances of overtraining as generalization performance is improved. The size of the model is adapted to the complexity of the task to be modeled. From an implementation point of view, the smaller the model is, the faster it will operate. Finally, from a knowledge extraction point of view, removing connections allows one to extract rules and removing input units allowing selection of the most useful attributes. According to the goal and the method, the removing process can be applied to a connection, a hidden unit or an input unit. The idea common to many pruning methods is to build an approximation of the error surface in the neighborhood of a local minimum by a Taylor expansion [Eq. (2)] to study the variation resulting from a removed weight.

$$E = \sum_i \frac{\partial E}{\partial w_i} \ w_i + \frac{1}{2} \sum_i \sum_j \frac{\partial^2 E}{\partial w_i \partial w_j} \ w_i \ w_j + O(\| \ W\|^3) \tag{2}$$

where E is the error of the neural network, W is the vector containing the weights and w_i is one specific weight.

Some assumption are made to simplify the computation. First, the weight variation is fixed to obtain the suppression of the weight influence, $w_i = -w_i$. We assume that after learning we are in a local minimum, so $\forall i \, \partial E / \partial w_i = 0$. We assume that around this minimum, the error surface is approximately quadratic, so $O(\| \; W\|^3)$ are insignificant terms. In these conditions, Eq. (3) contains terms with second order derivatives only, which are, by definition, the coefficients of the Hessian matrix **H**:

$$E = \frac{1}{2} \sum_i \sum_j \frac{\partial^2 E}{\partial w_i \partial w_j} w_i w_j \tag{3}$$

The computational complexity is quadratic according to the number of weights. Therefore, a network with one thousand weights needs to compute one million terms at each iteration.

Le Cun, Denker, and Solla (1990) have suggested assuming that the Hessian matrix is diagonal to reduce the number of derivatives to compute. That amounts to making the assumption that the error variation that occurs when several connections are removed is equal to the sum of the error variations resulting from each being removed. Then:

$$E = \frac{1}{2} \sum_i \frac{\partial^2 E}{\partial w_i \partial w_i} (w_i)^2 = \sum_i \quad E(w_i) \tag{4}$$

To measure the influence of a weight, they define the saliency, s_i of a connection w_i as:

$$s_i = \quad E(w_i) = \frac{1}{2} \frac{\partial^2 E}{\partial w_i^2} (w_i)^2 \tag{5}$$

This method is a good compromise between simplicity and efficiency.

Different strategies exist to decide how many connections to remove. The best thing to do is to remove only one connection at a time and then to retrain the network before pruning again. But to go faster, we can decide to remove a small fixed number of connection or, better, a small percentage of the current number of connections. In any instant, the computations are available for a small variation around the actual location.

RESULTS

Discrimination performances

Several multilayer perceptrons were used with 25 inputs, 2 outputs and between 5 and 15 units in the hidden layer (Fig. 3).

The best performances in generalization for these different architectures are similar to each other and can be resumed by the following table (Table 1).

The results we obtain are metal dependant. The system is able to determine correctly when a site is a gold deposit in 92%, a silver deposit in 93%, and a copper deposit in 88 of the situations. When the number of examples in the database is too small, such as for antimony, molybdenum or mercury (resp. 31, 25, and 5 examples), the performances are less effective because from a statistical point of view the network discovers it can obtain good performance just by classifying all unknowns as a "barren site" (e.g. it obtains only 31/641 = 5% error if it classes all antimony sites as barren sites). Hence, the global measure of misclassified patterns should be considered to be complete only with the information given by the confusion matrix, especially when the number of patterns available for each class is unequal.

Variable selection

The optimal brain damage algorithm was applied to select the most relevant attributes by reducing by 5% the number of connections after each training (Fig. 4 illustrates many advantages of this approach).

The approach produced the following ordered list of relevant attributes:

1. Shortest distance to fault orientations between N 135°E and N 157.5°E;
2. Earthquake density;
3. Longitude;
4. Presence of Cenozoic host rock;
5. Shortest distance to an active volcano;
6. Altitude;
7. Presence of Mesozoic host rock;
8. Shortest distance to fault orientations between N 0°E and N 22.5°E;
9. Shortest distance to fault orientations between N 112.5°E and N 135°E.

This order presents the elements which occur to be most influential out of the list of 25 different items in potentially controlling gold mineralization:

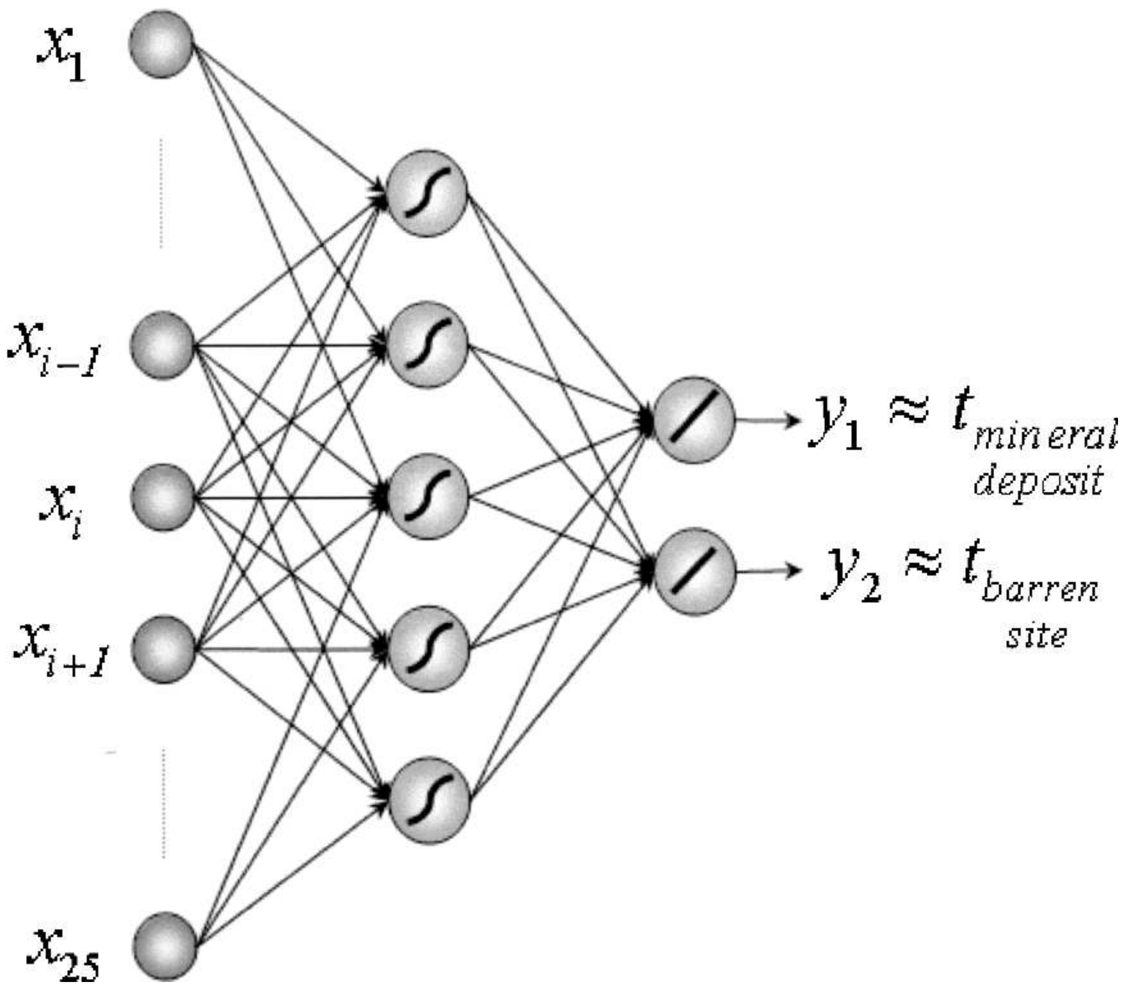

Figure 3. Neural network architecture used to discriminate patterns of two different classes (e.g. mineral deposits and barren sites or gold deposits and no gold deposits) and described by 25 attributes using supervised learning algorithm such as back-propagation. Hidden units have hyperbolic tangent function as activation function and output units have Gibbs' activation function to predict *a posteriori* probabilities that sample belongs to each class.

(i) variables 2 and 5 could underline the role of unsteady and long-lived crustal structures, at least since the Miocene; (ii) variables 3 and 6 probably are linked to the selection and location of barren sites that are located preferentially outside the mineralized zones at lower altitudes; (iii) variable 4 indicates that gold concentrations are controlled by Tertiary magmatism which represents the preferred but nonexclusive host rocks of gold mineralization (e.g. variable 7); and (iv) NW-striking faults (variables 1 and 9) play an important role as well as approximately N-S (variable 8) structures.

Table 1. Confusion matrix in generalization obtained by a multilayer perceptron using all attributes and 10 units on the hidden layer. In this example, test set contains 47 epithermal-porphyry deposits and 82 barren sites and it leads to 84% of correct classification.

		Predicted Class	
		Deposit	Barren
Known Class	Deposit	66%	34%
	Barren	6%	94%

The absence in this list of the "nature of host rock" variable (e.g. volcanic lithologies), which appears at the 15th position (out of 25) illustrates a surprising result. However, despite the fact that some attributes, such as the longitude or the altitude, would not seem to be relevant, the application shows that the network can discriminate easily a barren site

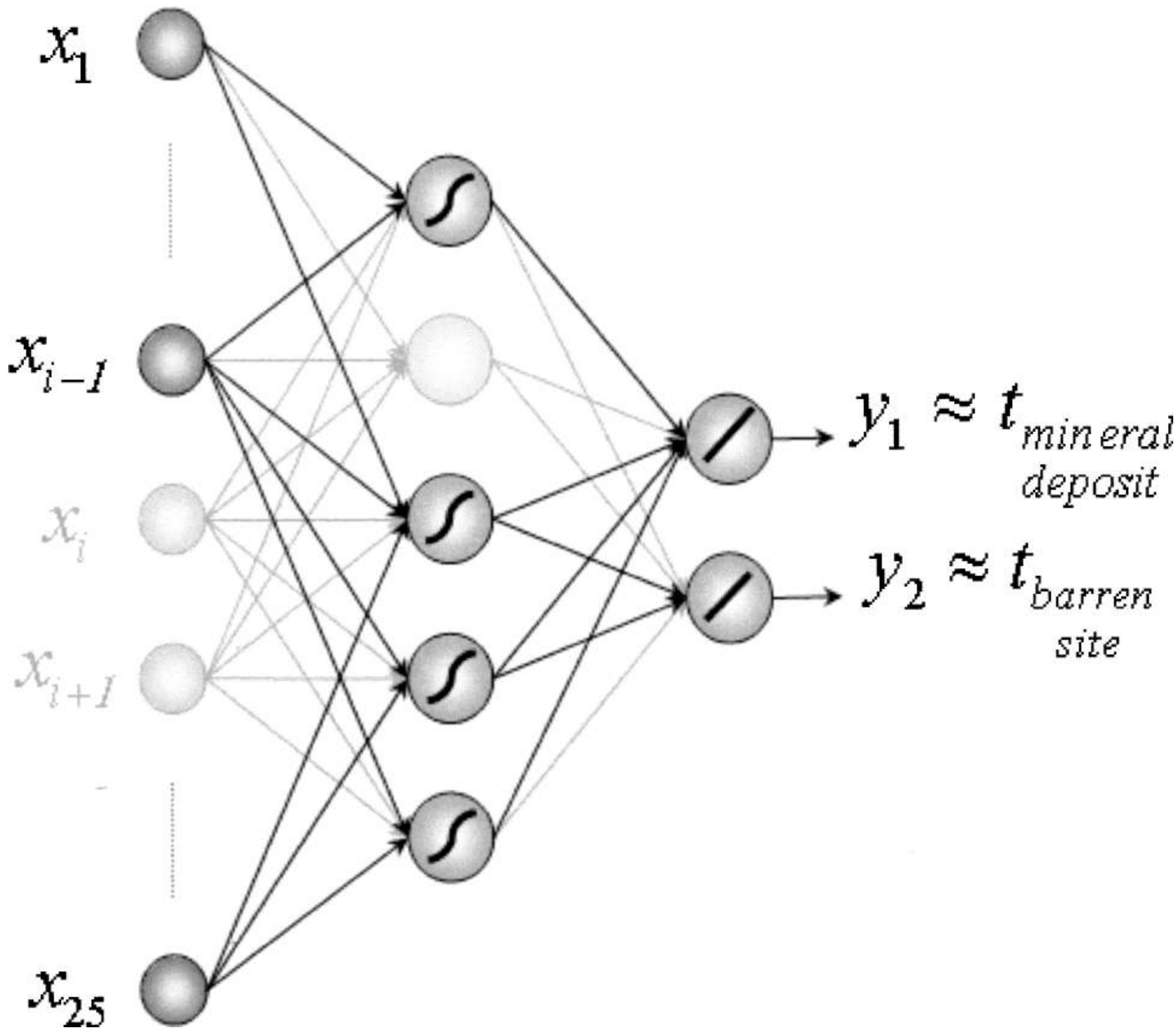

Figure 4. After pruning, many connections are removed. This reduces complexity model. Hence, it is possible to handle overtraining problem. Moreover, it is possible to extract knowledge; that is to select most useful attributes, to identify some meaningful combinations, and to extract rules (Remm and Alexandre, 2002).

from a mineralized one with a success rate of 84% (from Table 1). Careful selection of barren sites, as discussed next remains vital to the success of the network operation and is incorporated in current research.

CONCLUSIONS

Data mining performances are dependent heavily on the database and the efficiency of the applied data exploration routines. Selection of barren sites is a key factor in mineral-deposit exploration using machine-learning techniques. These sites should nearly resemble the deposit sites, in location and data description. For example, barren sites can be selected from deposits that do not contain the specific metal we are looking for. Nevertheless, our approach demonstrates how neural networks can be used efficiently in a practical problem of mineral exploration, where the general domain knowledge alone is insufficient to satisfactorily model the possible controls on deposit formation using the available information in a continental-scale information system.

ACKNOWLEDGMENT

The stylesheet was derived from those proposed by Damien Genthial for TALN'97 and Pierre Zweigenbaum for TALN'98. GIS Andes was developed within the context of two successive BRGM R&D projects – "Andes Metallogeny" and "Global Environmental and Metallogenic Syntheses" (GEMS), the latter being host to this study.

REFERENCES

An, P., Moon, W.M., and Rencz, A., 1991, Application of fuzzy set theory to integrated mineral exploration: Can. Jour. Exploration Geophysics, v. 27, no. 1, p. 1–11.

Bonham-Carter, G.F., 1994, Geographic information systems for geoscientists-modelling with GIS: Pergamon, Oxford, 398 p.

Bonham-Carter, G.F., Agterberg, F.P., and Wright, D.F., 1989, Weights of evidence modelling: a new approach to mapping mineral potential, *in* Agterberg F.P., and Bonham-Carter, G.F., eds; Statistical Applications in Earth Sciences: Geol. Survey Canada Paper 89–9, p. 171–183.

Burrough, P.A., and McDonnell, R.A., 1998, Principles of geographic information systems (2nd edn.): Oxford Univ. Press, London, 333 p.

Cassard, D., 1999, GIS Andes: a metallogenic GIS of the Andes Cordillera: 4th Intern. Symp. Andean Geodynamics (Göttingen, Extended Abstracts): Institut de Recherche pour le Développement Publ., Paris, p. 147–150.

Cassard, D., 2000, GIS ANDES: A metallogenic GIS of the Andes Cordillera: IGC 31st Intern. Geological Congress, Rio de Janeiro, Brasil, Abstracts CD.

Cassard, D., Stein, G., Milesi, J.P., and Lips, A.L.W., 2001, GIS central Europe: the Metallogenic GIS of central and South-Eastern Europe: EUG XI, Strasbourg, France, Conf. Abst, v. 6, p. 556.

Cibas, T., Fogelman-Soulié, F., Gallinari, P., and Raudys, S., 1996, Variable selection with neural networks: Neurocomputing, v. 12, no. 2, p. 223–248.

Harris, J.R., Wilkinson, L., Heather, K., Fumerton, S., Bernier, M.A., Ayer, J., and Dahn, R., 2001, Application of GIS processing techniques for producing mineral prospectivity maps - a case study: mesothermal Au in the Wayze Greenstone Belt, Ontario, Canada: Natural Resources Research, v. 10, no. 2, p. 91–124.

Knox-Robinson, C.M., and Wyborn, L.A.I., 1997, Towards a holistic exploration strategy: using geographic information systems as a tool to enhance exploration: Australian Jour. Earth Sciences, v. 44, no. 4, p. 453–463.

Le Cun, Y., Denker, J.S., and Solla, S.A., 1990, Optimal brain damage, *in* Touretzky D.S., ed., Advances in Neural Information Processing Systems: Morgan Kaufmann, San Maeo, CA, p. 598–605.

Leistel, J.M., and the MinUrals Team, in press, MINURALS: Mineral Resources of the Urals - Origin, Development and Environmental Impact (abst.): EUG XI, Nice, France.

Milesi, J.P., Feybesse, J.L., Pinna, P., and Deschamps, Y., 2001, GIS AFRICA: a 1:2,000,000-scale tool for sustainable development (abst.): EUG XI, Strasbourg, France, Conf. Abst., v. 6, p. 556.

Raines, G.L., 1999, Evaluation of weights of evidence to predict epithermal-gold deposits in the Great Basin of the Western United States: Natural Resources Research, v. 8, no. 4, p. 257–276.

Reed, R., 1993, Pruning algorithms - a survey: IEEE transactions on Neural Networks, v. 4, no. 5, p. 740–747.

Remm, J.-F., and Alexandre, F., 2002, Knowledge extraction using artificial neural networks: application to radar target identification: Signal Processing. v. 82. no. 1. p. 117–120.

Salleb, A., and Vrain, C., 2000, An application of association rules discovery to geographic information systems: 4th European Conf. Principles of Data Mining and Knowledge Discovery (PKDD), Lyon, France, p. 613–618.

Zadeh, L.A., 1965, Fuzzy sets: Information and Control, v. 8, p. 338–353.

APPENDIX

PSEUDO-CODE for the Optimal Brain Damage Algorithm

```
Repeat
                                                        Training
  Repeat
    For each sample x do
    Present x as a new input of the network
    Compute the output values
    Compute the error
    For each weight w do
      Compute its update value w
      Compute its new value w+ w
    End for
  End for
Until the network is trained
                                                         Pruning
For each connection do
  Compute its saliency
End for
Arrange in growing order the connections according to
  their saliencies
Prune the first connection
For each hidden unit do
  If the unit is not linked anymore to at least
      one output unit
    Prune every input connection
    Remove the unit
  End if
  End for
For each input unit do
  If the unit is not linked anymore to at least
    one hidden unit
    Remove the unit
  End if
End for

Until the stopping criterion is reached (e.g. the error increases
compared with the previous one obtained by the last architecture
or the architecture is smaller enough to extract knowledge)
```