Uncertainty in Mineral Prospectivity Prediction

Pawalai Kraipeerapun[1], Chun Che Fung[2], Warick Brown[3], Kok Wai Wong[1], and Tamás Gedeon[4]

[1] School of Information Technology, Murdoch University, Australia
{p.kraipeerapun, k.wong}@murdoch.edu.au
[2] Centre for Enterprise Collaboration in Innovative Systems, Australia
l.fung@murdoch.edu.au
[3] Centre for Exploration Targeting, The University of Western Australia, Australia
wbrown@cyllene.uwa.edu.au
[4] Department of Computer Science, The Australian National University, Australia
tom@cs.anu.edu.au

Abstract. This paper presents an approach to the prediction of mineral prospectivity that provides an assessment of uncertainty. Two feed-forward backpropagation neural networks are used for the prediction. One network is used to predict degrees of favourability for deposit and another one is used to predict degrees of likelihood for barren, which is opposite to deposit. These two types of values are represented in the form of truth-membership and false-membership, respectively. Uncertainties of type error in the prediction of these two memberships are estimated using multidimensional interpolation. These two memberships and their uncertainties are combined to predict mineral deposit locations. The degree of uncertainty of type vagueness for each cell location is estimated and represented in the form of indeterminacy-membership value. The three memberships are then constituted into an interval neutrosophic set. Our approach improves classification performance compared to an existing technique applied only to the truth-membership value.

1 Introduction

The prediction of new mineral deposit location is a crucial task in mining industry. In recent years, Geographic Information System (GIS) and neural networks have been applied in many applications for mineral prospectivity prediction [1,2,3]. Several sources of data such as geology, geochemistry, and geophysics are involved in the prediction. Data collected from these sources always contains uncertainty. Hence, the predicted mineral deposit locations also contain some degrees of uncertainty. There are several types of uncertainty such as error, inaccuracy, imprecision, vagueness, and ambiguity [4,5]. This paper deals with two types of uncertainty, which are uncertainty of type error and uncertainty of type vagueness. Error can happen from several aspects such as measurement, data entry, processing, lacking of knowledge about data, or lacking of ability in measurement [5]. This study deals with error occurred in the process of prediction. Vagueness refers to boundaries that cannot be defined precisely [5]. In

I. King et al. (Eds.): ICONIP 2006, Part II, LNCS 4233, pp. 841–849, 2006.

this study, the locations are known, but uncertain existence of favourability for deposit. Some locations have one hundred percent of favourability for deposits. Some locations have zero percent of favourability for mineral deposits. These cells are determined as non-deposit or barren cells. Most locations have degrees of favourability between these two extremes. For each location, we cannot predict the exact boundary between favourability for deposit and likelihood for barren. Vagueness or indeterminable information always occurs in the boundary zone.

This paper presents a method using GIS data and neural networks for predicting the degree of favourability for mineral deposit, degree of likelihood for barren, and degree of indeterminable information in the mineral prospectivity prediction. Instead of considering only uncertainty in the boundary between both degrees of favourability for deposit and barren, we also consider uncertainty of type error in the prediction of both degrees. A multidimensional interpolation method is used to estimate these errors. In order to represent the three degrees for each location, an interval neutrosophic set [6] is used to express them. The basic theory of an interval neutrosophic set is described in the next section.

The rest of this paper is organized as follows. Section 2 presents the basic theory of interval neutrosophic sets. Section 3 explains proposed methods for mineral prospectivity prediction and quantification of uncertainties using GIS data, neural networks, interval neutrosophic sets, and a multidimensional interpolation. Section 4 describes the GIS data set and the results of our experiments. Conclusions and future work are presented in Section 5.

2 Interval Neutrosophic Set

The membership of an element to the interval neutrosophic set is expressed by three values: t, i, and f. These values represent truth-membership, indeterminacy-membership, and false-membership, respectively. The three memberships are independent. In some special cases, they can be dependent. These memberships can be any real sub-unitary subsets and can represent imprecise, incomplete, inconsistent, and uncertain information [7]. In this paper, the three memberships are considered to be dependent. They are used to represent uncertainty information. This research follows the definition of interval neutrosophic sets that is defined in [7]. This definition is described below.

Let X be a space of points (objects). An interval neutrosophic set in X is defined as:

$$\begin{array}{c} A = \{x(T_A(x), I_A(x), F_A(x))|x \in X \wedge \\ T_A : X \longrightarrow [0,1] \wedge \\ I_A : X \longrightarrow [0,1] \wedge \\ F_A : X \longrightarrow [0,1]\} \end{array} \tag{1}$$

where

T_A is the truth-membership function,
I_A is the indeterminacy-membership function, and
F_A is the false-membership function.

3 Mineral Prospectivity Prediction and Quantification of Uncertainty

In this study, gridded map layers in a GIS database are used to predict mineral prospectivity. Fig.1 shows our proposed model that consists of GIS input layers, two neural networks, and a process of indeterminacy calculation. The output of this model is an interval neutrosophic set in which each cell in the output consists of three values: deposit output, indeterminacy output, and non-deposit output which are truth-membership, indeterminacy-membership, and false-membership values, respectively.

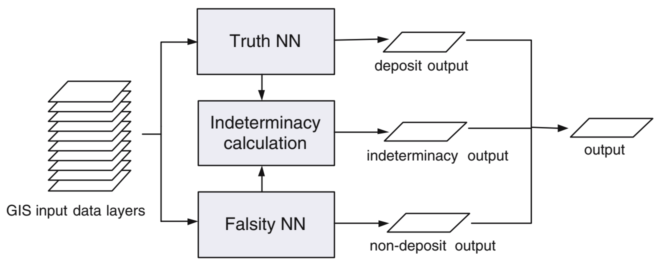


Fig. 1. Uncertainty model based on the integration of interval neutrosophic sets (INS) and neural networks (NN)

In the proposed model, the truth NN is a feed-forward backpropagation neural network. This network is trained to predict the degree of favourability for deposit, which is the truth-membership value. The falsity NN is also a feed-forward backpropagation neural network in which its architecture and all properties are the same as the architecture and all properties used for the truth NN. The only difference is that the falsity NN is trained to predict degree of likelihood for barren using the complement of target outputs used for training data in the truth NN. For example, if the target output used to train the truth neural network is 0.9, its complement is 0.1. Fig.2 shows our training model. It consists of two neural networks: truth NN and falsity NN. Errors produced from both neural networks will be used to estimate uncertainties in the prediction for the new input data.

Fig.3 shows uncertainty estimation in the prediction of truth-membership for the new data set or unknown data. The errors produced from the truth NN are plotted in the multidimensional feature space of the training input patterns. Thus, uncertainties of the new input patterns can be estimated using multidimensional interpolation. Estimated uncertainty in the barren prediction are also calculated in the same way as the estimated uncertainty for deposit. The errors produced from the falsity NN are plotted in the multidimensional feature space of the training input patterns. A multidimensional interpolation is then used to estimate uncertainty for the prediction of false-membership or degree of likelihood for barren. These two estimated uncertainties will be used in the dynamically

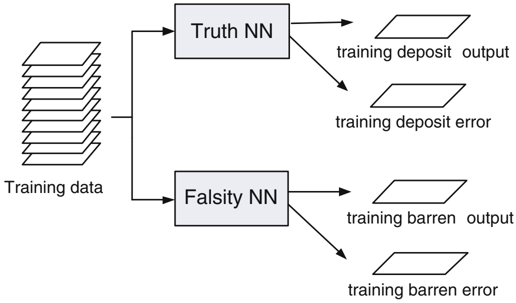


Fig. 2. Two neural networks used for training

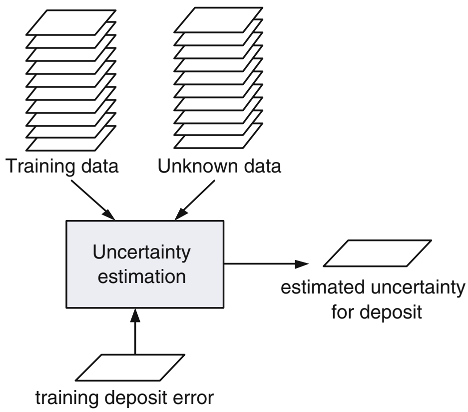


Fig. 3. Uncertainty estimation for mineral deposit prediction

weighted combination between truth-membership and false-membership for the binary classification later.

If the degree of favourability for deposit or truth-membership value is high then the degree of likelihood for barren or false-membership value should be low, and the other way around. For example, if the degree of favourability for deposit is 1 and the degree of likelihood for barren is 0, then the boundary between these two values is sharp and the uncertainty of type vagueness is 0. However, the values predicted from both neural networks are not necessary to have a sharp boundary. For instance, if both degrees predicted from the truth NN and the falsity NN for the same cell is equal, then this cell contains the highest uncertainty value, which is 1. Therefore, uncertainty in the boundary zone can be calculated as the difference between these two values. If the difference between these two values is high then the uncertainty is low. If the difference is low then the uncertainty is high. In this paper, uncertainty in the boundary between these two values is represented by the indeterminacy-membership value. Let C be an output GIS layer. $C = \{c_1, c_2, ..., c_n\}$ where c_i is a cell at location i. Let $T(c_i)$ be a truth-membership value at cell c_i. Let $I(c_i)$ be an indeterminacy-membership value at cell c_i. Let $F(c_i)$ be a false-membership value at cell c_i. For each cell, the indeterminacy membership value ($I(c_i)$) can be defined as follows:

$$I(c_i) = 1 - |T(c_i) - F(c_i)| \tag{2}$$

After three membership values are created for each cell, the next step is to classify the cell into either deposit or barren. Both truth-membership and false-membership are used in the classification. The estimated uncertainty of type error in the prediction of truth- and false-memberships are also integrated into the truth- and the false-membership values to support certainty of the classification. The less uncertainty in the prediction, the more certainty in the classification. Let $e_t(c_i)$ be an estimated uncertainty of type error in the prediction of the truth-membership at cell c_i. Let $e_f(c_i)$ be an estimated uncertainty of type error in the prediction of the false-membership at cell c_i. We determine the weights dynamically based on these estimated uncertainties. The weights for the truth- and false-membership values are calculated as the complement of the errors estimated for the truth- and false-membership, respectively. These weights are considered as the degrees of certainty in the prediction. In this paper, the certainty in the prediction of the false-membership value is considered to be equal to the certainty in the prediction of non-false-membership value, which is the complement of the false-membership value. Let $w_t(c_i)$ and $w_f(c_i)$ be the weights of the truth- and false-membership values, respectively. The output $O(c_i)$ of the dynamic combination among the truth-memberships, the false-memberships, and their uncertainties of type error can be calculated using equations below.

$$O(c_i) = (w_t(c_i) \times T(c_i)) + (w_f(c_i) \times (1 - F(c_i))) \tag{3}$$

$$w_t(c_i) = \frac{1 - e_t(c_i)}{(1 - e_t(c_i)) + (1 - e_f(c_i))} \tag{4}$$

$$w_f(c_i) = \frac{1 - e_f(c_i)}{(1 - e_t(c_i)) + (1 - e_f(c_i))} \tag{5}$$

In order to classify the cell into either deposit or barren, we compare the output to the threshold value. A range of threshold values are determined and compared to the output for each cell. The best threshold value that can produce the best accuracy in the classification will be selected for the mineral prospectivity prediction.

4 Experiments

4.1 GIS Data Set

The data set used in this study contains ten GIS layers in raster format. Each layer represents different variables which are collected and preprocessed from various sources such as geology, geochemistry, and geophysics in the Kalgoorlie region of Western Australia. An approximately 100×100 km area is divided into a grid of square cells of 100 m side. Each layer contains 1,254,000 cells. Each grid cell represents a single attribute value which is scaled to the range [0, 1]. For example, a cell in a layer representing the distance to the nearest fault contains a value of distance scaled to the range [0, 1]. Each single grid cell is classified into deposit or barren cell. The cells containing greater than 1,000 kg total contained gold are labeled as deposits. All other cells are classified as non-deposits or barren cells. In this study, the co-registered cells in the GIS input layers are used to constitute the input feature vector for our neural network model. We use only 268 cells in this experiment in which 187 cells are used for training and 81 cells are used for testing. For training data, we have 85 deposit cells and 102 barren cells. For testing data, we have 35 deposit cells and 46 barren cells.

4.2 Experimental Methodology and Results

Two feed-forward backpropagation neural networks are created in this experiment. The first neural network is used as the truth NN to predict degree of favourability for deposit and another network is used as the falsity NN to predict degree of likelihood for barren. Both networks contain ten input-nodes, one output node, and one hidden layer constituting of 20 neurons. The same parameter values are applied to the two networks and both networks are initialized with the same random weights. The only difference is that the target values for the falsity NN are equal to the complement of the target values used to train the truth NN.

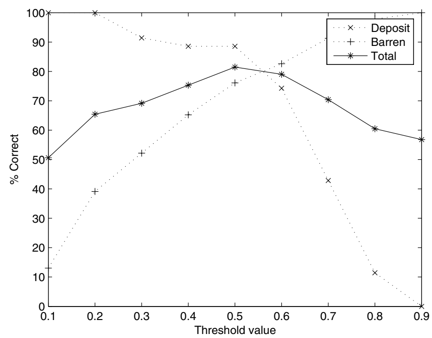
In order to estimate uncertainty of type error for the test data set, errors produced from the truth NN are plotted in the input feature space. In this study, only 60 patterns from the input training data are plotted in the input feature space because of memory limitations of the computer used in the experiment. A multidimensional interpolation [8] is then used to estimate uncertainty of type error for the test data set in the prediction of degree of favourability for deposit. We use multidimensional nearest neighbour interpolation function in Matlab to interpolate these errors. The estimation of uncertainty of type error for the prediction of degree of likelihood for barren is also calculated using the same technique as the error estimation for the deposit prediction.

In order to calculate uncertainty of type vagueness, which is the indeterminacy-membership value, equation 2 is used to compute this kind of uncertainty for each pattern in the test data set. After we created the three memberships: truth-membership, indeterminacy-membership, and false-membership for each pattern, these three memberships are then constituted into an interval neutrosophic set.

After the three memberships are determined for each pattern in the test data sets, the next step is to classify each pattern into deposit or barren cells. The truth-memberships, the false-memberships, and their uncertainties of type error are combined into a single output used for the binary classification. The dynamic combination can be computed using equation 3. After that, the classification is done by comparing each dynamic combination output to a threshold value. In this paper, threshold values are ranged from 0.1 to 0.9 in steps of 0.1. These threshold values are tested with each output to seek for the best threshold value

Table 1. Classification accuracy for the test data set obtained by applying a range of threshold values to the output of dynamic weighted combination among truth-membership, false-membership, and uncertainties of type error. (Right: graphical representation of data in this table)

Threshold value	Deposit %correct	Barren %correct	Total %correct
0.1	100.00	13.04	50.62
0.2	100.00	39.13	65.43
0.3	91.43	52.17	69.14
0.4	88.57	65.22	75.31
0.5	88.57	76.09	81.48
0.6	74.29	82.61	79.01
0.7	42.86	91.30	70.37
0.8	11.43	97.83	60.49
0.9	0.00	100.00	56.79



that produces the best accuracy in the classification. For each cell c_i in the classification, if the truth-membership value is greater than the threshold value then the cell is classified as a deposit. Otherwise, it is classified as barren. Table 1 shows classification accuracy for the test data set obtained by applying a range of threshold values to the output of dynamic combination. We found that the maximum of the total correct cell in the classification is 81.48 percent. Hence, the optimal threshold value used in this classification is determined to be 0.5.

In this paper, we do not consider the optimization of the prediction, but our purpose is to test a new approach that provides an estimate of uncertainty in the prediction. We compare our classification results with those obtained using the traditional method for binary classification. In the traditional approach, only truth-membership values are used in the comparison. If the cell has the truth-membership value greater than the threshold value then the cell is classified as deposit. Otherwise, the cell is classified as barren. Table 2 shows classification accuracy for the test data set obtained by applying a range of threshold values to the only truth-membership values. The maximum of the total correct cell in the traditional classifications is 80.25 percent. Therefore, the optimal threshold value used in this traditional classification is determined to be 0.5.

The results from our proposed classification using the dynamic combination represent 1.23 percent improvement over those obtained using the traditional classification applied only the truth-membership values. Table 3 shows samples of individual predicted cell types and their uncertainties of type error and vagueness resulted from our proposed model for the test data set. The individual predicted cell types for the traditional approach are also shown in this table in the last column. These samples are shown that our proposed model has an advantage of quantification of uncertainty in the prediction. For example, the actual cell type for the cell in the first row of this table is a deposit cell, but it is predicted to be a barren cell. The traditional approach cannot explain about uncertainty in this prediction, but our approach can explain that the cell is predicted to be a

Table 2. Classification accuracy for the test data set obtained by applying a range of threshold values to the truth-membership values. (Right: graphical representation of data in this table)

Threshold value	Deposit %correct	Barren %correct	Total %correct
0.1	100.00	30.43	60.49
0.2	94.29	47.83	67.90
0.3	91.43	56.52	71.61
0.4	91.43	63.04	75.31
0.5	88.57	73.91	80.25
0.6	77.14	80.43	79.01
0.7	54.29	89.13	74.07
0.8	22.86	95.65	64.19
0.9	2.86	100.00	58.02

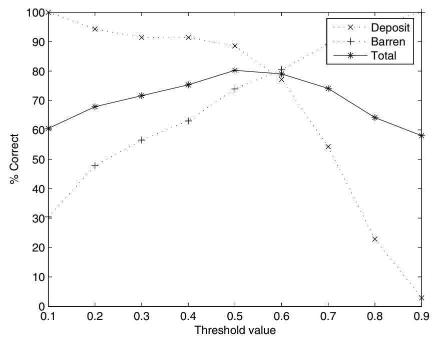


Table 3. Sample outputs from the proposed model for the test data set (columns 2-7) together with classifications based on dynamic combination (column 8) and traditional classifications based on truth-membership values (column 9)

Actual Cell Type	$T(c_i)$	$e_t(c_i)$	$F(c_i)$	$e_f(c_i)$	$I(c_i)$	Dynamic Combination $O(c_i)$	Predicted Cell Type $O(c_i) > 0.5$	Predicted Cell Type $T(c_i) > 0.5$
Deposit	0.40	0.04	0.70	0.16	0.70	0.35	Barren	Barren
Deposit	0.87	0.15	0.26	0.24	0.39	0.81	Deposit	Deposit
Deposit	0.69	0.14	0.65	0.14	0.95	0.53	Deposit	Deposit
Deposit	0.84	0.51	0.23	0.70	0.39	0.81	Deposit	Deposit
Barren	0.07	0.09	0.70	0.04	0.37	0.19	Barren	Barren
Barren	0.57	0.28	0.54	0.23	0.97	0.51	Deposit	Deposit
Barren	0.43	0.28	0.46	0.23	0.97	0.49	Barren	Barren
Barren	0.51	0.13	0.57	0.53	0.94	0.48	Barren	Deposit

barren cell with the uncertainty of 70 percent. Hence, the decision-maker can use this information to support the confidence in decision making.

Considering the last row of this table, the actual cell type for this cell is barren. Using the traditional approach, this cell is classified as a deposit which is a wrong prediction and there is no explanation of uncertainty in the prediction for this cell. Using our approach, this cell is classified as a barren, which is correct. We also know that the cell is barren with the uncertainty of 94 percent. We can see that uncertainty of type error in the prediction can enhance the classification. Therefore, the combination among the truth-membership, false-membership, and their uncertainties of type error gives the more accuracy in prediction.

5 Conclusions and Future Works

This paper represents a novel approach for mineral deposit prediction. The prediction involves ten GIS input data layers, two neural networks, an interval neutrosophic set, and a multidimensional interpolation. The co-register cells from GIS data are applied into two neural networks to produce the degrees of favourability for deposits (truth-membership) and the degrees of likelihood for barrens (false-membership). Two types of uncertainty in the prediction are estimated. These two kinds of uncertainty are error and vagueness. Estimated errors are computed using a multidimensional interpolation. Vagueness is calculated as the different between the truth- and false-membership values for each cell. This paper represents vagueness as the indeterminacy-membership. These three memberships are formed into an interval neutrosophic set. The goal of this paper is to quantify uncertainty in mineral prospectivity prediction. The more we know uncertainty information, the more certainty in decision making. In the future, we will apply this model to bagging and boosting neural networks.

References

1. Brown, W.M., Gedeon, T.D., Groves, D.I., Barnes, R.G.: Artificial neural networks: A new method for mineral prospectivity mapping. Australian Journal of Earth Sciences **47** (2000) 757–770
2. Skabar, A.: Mineral potential mapping using feed-forward neural networks. In: Proceedings of the International Joint Conference on Neural Networks. Volume 3. (2003) 1814–1819
3. Fung, C., Iyer, V., Brown, W., Wong, K.: Comparing the performance of different neural networks architectures for the prediction of mineral prospectivity. In: the Fourth International Conference on Machine Learning and Cybernetics (ICMLC 2005), Guangzhou, China (2005) 394–398
4. Duckham, M., Sharp, J.: Uncertainty and geographic information: Computational and critical convergence. In: Representing GIS. John Wiley, New York (2005) 113–124
5. Fisher, P.F.: Models of uncertainty in spatial data. In: Geographical Information Systems: Principles, Techniques, Management and Applications. 2 edn. Volume 1. John Wiley, Chichester (2005) 69–83
6. Wang, H., Madiraju, D., Zhang, Y.Q., Sunderraman, R.: Interval neutrosophic sets. International Journal of Applied Mathematics and Statistics **3** (2005) 1–18
7. Wang, H., Smarandache, F., Zhang, Y.Q., Sunderraman, R.: Interval Neutrosophic Sets and Logic: Theory and Applications in Computing. Neutrosophic Book Series, No.5. http://arxiv.org/abs/cs/0505014 (2005)
8. MATHWORKS: MATLAB Mathematics Version 7. The MathWorks Inc., Natick (2004)