

Ensemble Neural Networks Using Interval Neutrosophic Sets and Bagging

Pawalai Kraipeerapun, Chun Che Fung and Kok Wai Wong
School of Information Technology,
Murdoch University, Australia
Email: {p.kraipeerapun, l.fung, k.wong}@murdoch.edu.au

Abstract

This paper presents an approach to the problem of binary classification using ensemble neural networks based on interval neutrosophic sets and bagging technique. Each component in the ensemble consists of a pair of neural networks trained to predict the degree of truth and false membership values. Uncertainties in the prediction are also estimated and represented using the indeterminacy membership values. These three membership values collectively form an interval neutrosophic set. In order to combine and classify outputs from components in the ensemble, the outputs of an ensemble are dynamically weighted and summed. The proposed approach has been tested with three benchmarking UCI data sets, which are ionosphere, pima, and liver. The proposed ensemble method improves the classification performance as compared to the simple majority vote and averaging methods which were applied only to the truth membership value. Furthermore, the results obtained from the proposed ensemble method also outperform the results obtained from a single pair of networks and the results obtained from a single truth network.

1. Introduction

In order to solve the problem of classification and prediction, an ensemble of accurate and diverse neural networks was found capable of providing better results than a single neural network [4]. In the normal process of utilizing neural network ensemble, each network component has to be trained and then the outputs obtained from all networks in the ensemble are combined. However, there are situations that outputs from the networks are differed from one another. Dietterich [2] suggested that if two classifiers produce different errors on new input data then both classifiers are considered to be “diverse”. Diversity in an ensemble of neural networks can be handled by manipulating input data or output data. An example of algorithm that manipulates the diversity using output data is error-correcting out-

put coding. Bagging and boosting are examples of such algorithms that manipulate diversity using input data. In the error-correcting output coding algorithm, a unique code-word which is a binary string of length n , is created for each class and used as distributed output representation [3]. Bagging provides diversity by randomly resampling the original training data into several training sets [1] whereas boosting provides diversity by manipulating each training set according to the performance of the previous classifier [8]. Furthermore, manipulating the input features can also provide diversity in the ensemble [2]. In addition, diversity can be provided by applying artificial training samples. Melville and Mooney [6] built a training set for each new classifier by adding artificially constructed samples to the original training data. In order to construct sample labels, they assigned the class label that disagrees with the current ensemble to the constructed sample label.

In this paper, ensemble diversity is created using bagging algorithm. Bagging is created based on bootstrap resampling. Each training set in an ensemble is generated from the original training data using random resampling with replacement. Each generated training set contains the same number of training patterns as the original training set. The outputs of the ensemble can be aggregated using averaging or majority vote. Combination of the outputs in the ensemble can improve the accuracy results. However, uncertainty still exists. In this paper, we apply interval neutrosophic sets [9] in order to represent uncertainty in the prediction.

This research follows the definition of interval neutrosophic sets defined by Wang et al. [9]. The membership of an element to the interval neutrosophic set is expressed by three values: truth membership, indeterminacy membership, and false membership. The three memberships are independent although in some special cases, they can be dependent. In this study, the indeterminacy membership depends on both truth and false memberships. The three memberships can be any real sub-unitary subsets and can represent imprecise, incomplete, inconsistent, and uncertain information. In this paper, the memberships are used to represent uncertainty information. For example, let A

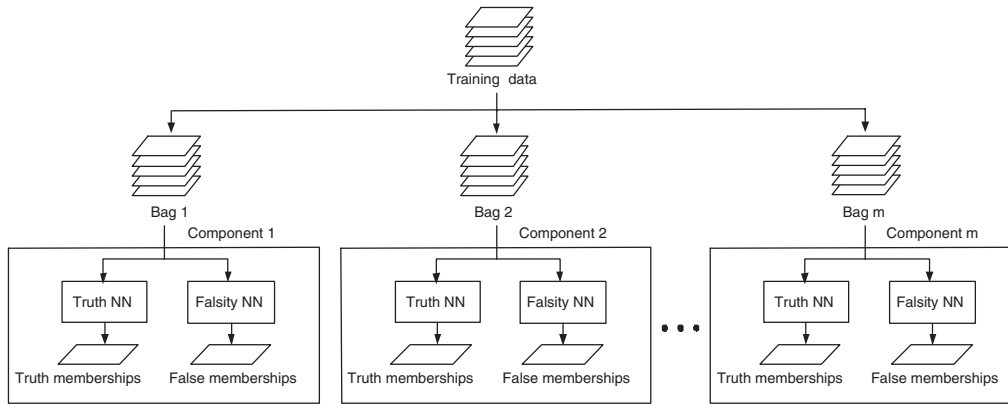


Figure 1. The proposed training model based on the integration of interval neutrosophic sets with bagging neural networks.

be an interval neutrosophic set, then $x(75, \{25, 35, 40\}, 45)$ belongs to A means that x is in A to degree of 75%, x is uncertain to degrees of 25% or 35% or 40%, and x is not in A to degree of 45%. The definition of an interval neutrosophic set is described below.

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

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

where

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

In this paper, we create a pair of neural networks for each component in the ensemble. In each pair, two networks are opposite to each other. Both neural networks are trained with the same bag of data but disagree in the output targets. The first network predicts degrees of truth membership whereas the second network predicts degrees of false membership. The predicted outputs from both networks are supposed to be complement to each other. However, both predicted outputs may not completely complement to each other. Uncertainty may occur in the prediction. In this study, we represent uncertainty in the form of indeterminacy membership value. The three memberships form an interval neutrosophic set and are used for decision making in the binary classification.

The rest of this paper is organized as follows. Section 2 explains the proposed method for the binary classification

with the assessment of uncertainty using interval neutrosophic sets and bagging. Section 3 describes the data set and the results of our experiments. Conclusions and future work are presented in Section 4.

2. Binary classification using interval neutrosophic sets, ensemble neural network, and bagging

In our previous papers [5], we integrated neural networks with interval neutrosophic sets in order to classify mineral prospectivity into deposit or barren cell. A pair of neural networks was created to predict degree of truth and false membership values. The predicted truth and false membership values were then compared to give us the classification results. Uncertainties in the classification were calculated as the difference between the truth and false membership values and were represented using indeterminacy membership values. We found that interval neutrosophic sets can represent uncertainty information and support the classification quite well. In this paper, we extend the work from our previous paper [5] by applying ensemble neural networks, interval neutrosophic sets, and a bagging technique to the problem of binary classification. Figure 1 shows the proposed training model that applies interval neutrosophic sets and a bagging technique to the ensemble neural network. Each component in the ensemble consists of a pair of neural networks, which are the truth neural network (Truth NN) and the falsity neural network (Falsity NN). The truth network is trained to predict degrees of truth membership. The falsity network is trained to predict degrees of false membership. Both networks are based on the same architecture. They apply the same bag for training. The difference between both networks is that the falsity network is trained

with the complement of the target output values presented to the truth network. In the training phase, each bag of data presented to each component in the ensemble is created using bootstrap resampling. In this study, each bootstrap sample or bag of data is created by random selection of input patterns from the training data set with replacement. Each bag contains the same number of training patterns as the original data set. Hence, m bags of data are applied to m pairs of truth and falsity neural networks.

In the test phase, the test data is applied to each component in the ensemble. From our testing model, each pair of the truth and falsity networks predict n pairs of the truth and false membership values where n is the total number of patterns. For each pair of the truth and false membership value, the truth membership value is supposed to be complement to the false membership value. For example, if the truth membership value is 1, the false membership value is supposed to be 0. If the difference between these two values is 1 then the uncertainty value will be 0. However, the predicted truth membership value is not necessary to be one hundred percent complement to the predicted false membership value. Uncertainty may occur in the prediction. For instance, if the truth membership value is 0.5 and the false membership value is also a value of 0.5 then the uncertainty value will be 1. Consequently, we compute the uncertainty value as the difference between the truth membership and false membership values. If the difference between these two values is high then the uncertainty is low. On the other hand, if the difference is low then the uncertainty value is high. Figure 2 shows the relationships among the truth membership, false membership, and uncertainty values. In this paper, we represent uncertainty value in the form of indeterminacy membership value. Hence, the output obtained from each component is represented as an interval neutrosophic set. The three memberships created from each component can be defined as the following.

Let X_j be the j -th output at the j -th component, where $j = 1, 2, 3, \dots, m$. Let A_j be an interval neutrosophic set in X_j . A_j can be defined as

$$A_j = \{x(T_{A_j}(x), I_{A_j}(x), F_{A_j}(x)) \mid x \in X_j \wedge \begin{aligned} T_{A_j} : X_j &\longrightarrow [0, 1] \wedge \\ I_{A_j} : X_j &\longrightarrow [0, 1] \wedge \\ F_{A_j} : X_j &\longrightarrow [0, 1] \}, \end{aligned} \quad (2)$$

$$I_{A_j}(x) = 1 - |T_{A_j}(x) - F_{A_j}(x)|, \quad (3)$$

where

- T_{A_j} is the truth membership function,
- I_{A_j} is the indeterminacy membership function,
- F_{A_j} is the false membership function.

In order to combine the outputs obtained from all components for each input pattern, the truth membership values

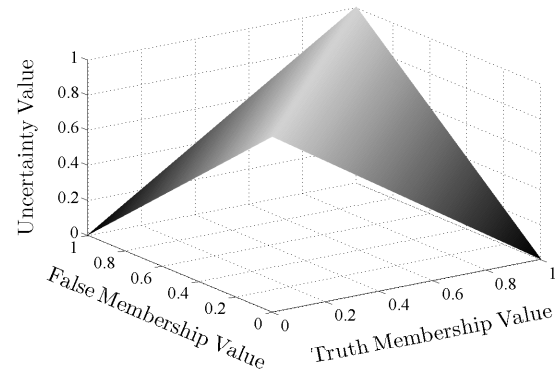


Figure 2. Relationships among the truth membership value, false membership value, and uncertainty value.

are dynamically weighted average. Also, the false membership values obtained from all components are dynamically weighted average. After that, the average truth membership and the average false membership values are compared in order to classify the input pattern into a binary class. In this study, the weight is dynamically created based on the indeterminacy membership value. The more weight means the more certainty in the prediction. We calculate the certainty as the complement of the indeterminacy membership value. Let $P(x_i)$ be an average truth membership value based on weights. Let $Q(x_i)$ be an average false membership value based on weights. Let $W_j(x_i)$ be the weight based on the indeterminacy membership value at the j -th component. P , Q and W can be defined as the following.

$$P(x_i) = \sum_{j=1}^m (W_j(x_i) \times T_{A_j}(x_i)) \quad (4)$$

$$Q(x_i) = \sum_{j=1}^m (W_j(x_i) \times F_{A_j}(x_i)) \quad (5)$$

$$W_j(x_i) = \frac{1 - I_{A_j}(x_i)}{\sum_{j=1}^m (1 - I_{A_j}(x_i))} \quad (6)$$

After the average truth membership and the average false membership values are computed for each input pattern, these two values are compared. If the average truth membership value is greater than the average false membership value ($P(x_i) > Q(x_i)$) then the input pattern is classified as a value 1. Otherwise, the input pattern is classified as a value 0.

3. Experiments

3.1. Data set

Three data sets from UCI Repository of machine learning [7] are used for binary classification in this paper. Table 1 shows the characteristics of these three data sets. The size of training and testing data used in our experiments are also shown in this table.

Table 1. Data sets used in this study.

Name	ionosphere	pima	liver
No. of Class	2	2	2
No. of Feature	34	8	6
Feature Type	numeric	numeric	numeric
Size of Samples	351	768	345
Size of Training Data	200	576	276
Size of Test Data	151	192	69

3.2. Experimental methodology and results

In this experiment, three data sets named ionosphere, pima, and liver from UCI Machine Learning Repository are applied to our model. Each data set is split into a training set and a testing set. The sizes of both sets are shown in table 1. For each training data set, thirty bags are created using bootstrap resampling with replacement and applied to thirty components in the ensemble. For each component, a pair of feed-forward backpropagation neural networks is trained in order to predict degree of truth membership and degree of false membership values. In this paper, we want to focus on our technique that aims to increase diversity by creating a pair of opposite networks in each component in the ensemble. Therefore, all networks in each ensemble apply the same parameter values and are initialized with the same random weights. The only difference for each pair of networks is that the target outputs of the falsity network are equal to the complement of the target outputs used to train the truth network. In the ionosphere data set, all networks in the ensemble have the same architecture which composes of thirty-four input units, a single output unit, and one hidden layer constituting of sixty-eight neurons. In the pima data set, all networks compose of eight input units, a single output unit, and one hidden layer constituting of sixteen neurons. In the liver data set, all networks compose of six input units, a single output unit, and one hidden layer constituting of twelve neurons.

In the test phase, after both truth and false membership values are predicted, the indeterminacy memberships are then computed using an equation 3. In order to combine the output from the networks within the ensemble, we apply our

Table 2. Average classification results for the test data set obtained by applying the proposed methods and the existing methods

	Technique	Ionosphere %correct	Pima %correct	Liver %correct
Bagging	dynamically weighted average	98.21	77.97	74.13
	Simple averaging	98.01	77.16	69.93
	Simple majority vote	98.15	76.38	71.30
Single	$T_j > F_j$	96.42	74.74	66.52
	$T_j > 0.5$	93.54	70.49	62.68

technique described in the previous section. In this paper, we do not consider the optimization of the prediction but concentrate only on the improvement of the prediction. In the experiment, we try twenty ensembles for each UCI data set. Each ensemble includes thirty different bags of training set. For each data set, the classification accuracy results obtained from all twenty ensembles are averaged and shown in Table 2. Furthermore, we also compare the average results obtained from our bagging technique (row 2) among the average results obtained from the existing bagging techniques (row 3-4), the existing technique using a single pair of networks (row 5), and the existing technique using only a single truth neural network (row 6).

In the third row of Table 2, the results obtained from the simple averaging technique are shown. In this technique, only the truth neural network constitutes each component in an ensemble. The truth membership values obtained from all components are averaged and then compared to the threshold value of 0.5. If the average result is greater than the threshold value then the input pattern is classified as a value 1. Otherwise it is classified as a value 0. In this technique, twenty ensembles are created for each data set, and the average results are shown.

In the simple majority vote technique, only the truth neural networks constitute an ensemble. The truth membership value obtained from each network is compared to the threshold value of 0.5. If the truth membership value is greater than the threshold value then the input pattern is classified as a value 1. Otherwise it is classified as a value 0. After that, all results are voted for each input pattern. If at least half of the results yield a value 1 then the input pattern is classified as a value 1. Otherwise, it is classified as a value 0. In this technique, twenty ensembles are created for each data set. The average results are shown in the fourth row.

Table 3. Total number of correct and incorrect outputs predicted from the proposed technique for the test set of pima data.

Uncertainty		Number of cell		%correct
value	level	correct	incorrect	
0.6940-0.9998	High	21	17	55.26
0.3881-0.6939	Med	52	19	73.24
0.0821-0.3880	Low	79	4	95.18

In the fifth row, the technique presented in our previous paper [5] is applied. A single pair of neural networks is trained. After that, the predicted truth and false membership values are compared in order to classify the binary class. If the truth membership value is greater than the false membership value then the input pattern is classified as a value 1. Otherwise, it is classified as a value 0. In this technique, we try twenty pairs of neural networks with twenty different randomized training sets for each data set. All twenty results are then averaged. The average result belonging to each data set is shown in Table 2.

In the last row, a single neural network is trained in order to provide the truth membership value. The output of the network is then compared to the threshold value of 0.5. The input pattern is assigned a value 1, if the output is greater than the threshold value. Otherwise it is assigned a value 0. Similar to the previous techniques, twenty neural networks are trained with twenty different randomized training sets. All twenty predicted results are averaged. The average results are shown in Table 2.

From the experiments, we found that the technique of the comparison between the truth and false membership values gives us better performance compared to the technique using the threshold value for the classification. We also found that the bagging technique improves the classification performance as compared to the technique that applies only a single pair of opposite networks or only a single network. Furthermore, our experiments show that the results obtained from the proposed ensemble technique (row 2) outperform the results obtained from the other existing techniques used in this paper.

In addition, our approach has an ability to represent uncertainty in the classification. For each input pattern, uncertainty in the classification can be calculated as the difference between $P(x)$ and $Q(x)$, which are the average truth membership value and the average false membership value based on weights, respectively. This value can be used to support the confidence in the classification. For example, table 3 shows the ranges of uncertainty in the classification of pima data set. Uncertainty values are categorized into three levels: High, Med, and Low. This table represents the total

number of correct and incorrect outputs predicted from the proposed ensemble technique. The table shows that most of the outputs that have low level of uncertainty are correctly classified. Hence, this uncertainty level can be used as an indicator in order to support the decision making.

4. Conclusion and future work

This paper has applied a pair of opposite neural networks to input patterns derived from bagging technique for the prediction of the truth membership and false membership values. A pair of networks constitutes a component in the ensemble. The difference between each pair of the truth and false membership values gives us an uncertainty value or an indeterminacy membership value. The three memberships form an interval neutrosophic set and are used for dynamically weighted averaging. The advantage of our approach over a simple averaging and majority vote approaches is that the indeterminacy membership values provide an estimate of the uncertainty of the classification. In addition, our experimental results indicate that our proposed ensemble technique improves the classification performance compared to the existing techniques. In the future, we will apply our technique to the problem of multiclass classification.

References

- [1] L. Breiman. Bagging Predictors. *Machine Learning*, 24(2):123–140, 1996.
- [2] T. G. Dietterich. Ensemble Methods in Machine Learning. In J. Kittler and F. Roli, editors, *Multiple Classifier Systems*, volume 1857 of *Lecture Notes in Computer Science*, pages 1–15. Springer, 2000.
- [3] T. G. Dietterich and G. Bakiri. Solving Multiclass Learning Problems via Error-Correcting Output Codes. *Journal of Artificial Intelligence Research*, 2:263–286, 1995.
- [4] L. K. Hansen and P. Salamon. Pattern Analysis and Machine Intelligence. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 12, pages 993–1001, October 1990.
- [5] P. Kraipeerapun, C. C. Fung, W. Brown, and K. W. Wong. Mineral Prospectivity Prediction using Interval Neutrosophic Sets. In V. Devedzic, editor, *Artificial Intelligence and Applications*, pages 235–239. IASTED/ACTA Press, 2006.
- [6] P. Melville and R. J. Mooney. Constructing Diverse Classifier Ensembles using Artificial Training Examples. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 505–512, 2003.
- [7] D. Newman, S. Hettich, C. Blake, and C. Merz. UCI Repository of machine learning databases, 1998.
- [8] H. Schwenk and Y. Bengio. Boosting Neural Networks. *Neural Computation*, 12(8):1869–1887, 2000.
- [9] H. Wang, D. Madiraju, Y.-Q. Zhang, and R. Sunderraman. Interval neutrosophic sets. *International Journal of Applied Mathematics and Statistics*, 3:1–18, March 2005.