Multiclass Classification using Neural Networks and Interval Neutrosophic Sets

PAWALAI KRAIPEERAPUN CHUN CHE FUNG KOK WAI WONG
School of Information Technology
Murdoch University
Perth, Western Australia
AUSTRALIA

Abstract:- This paper presents a new approach to the problem of multiclass classification. The proposed approach has the capability to provide an assessment of the uncertainty value associated with the results of the prediction. Two feed-forward backpropagation neural networks, each with multiple outputs, are used. One network is used to predict degrees of truth membership and another network is used to predict degrees of false membership. Indeterminacy membership or uncertainty in the prediction of these two memberships is also estimated. Together these three membership values form an interval neutrosophic set. Hence, a pair of single multiclass neural networks with multiple outputs produces multiple interval neutrosophic sets. We experiment our technique to the classical benchmark problems including balance, ecoli, glass, lenses, wine, yeast, and zoo from the UCI machine learning repository. Our approach improves classification performance compared to an existing technique which applied only to the truth membership created from a single neural network with multiple outputs.

Key–Words:- multiclass classification, uncertainty, interval neutrosophic sets, multiclass neural network, feedforward backpropagation neural network

1 Introduction

Multiclass neural network classification involves building neural networks that map the input feature vector to the network output containing more than two classes [1]. In general, there are two existing neural network architectures used to classify multiple classes. The first approach is to build multiple binary neural networks in which each network can be modeled independently. One advantage of using this technique is that different features can be applied to train different neural networks [2]. However, each neural network is trained only based on local knowledge which may produce overlaps or gaps in the classification boundary zone [1]. The second approach is the implementation of a single neural network with multiple outputs. The complexity of this approach is usually high [3]. However, the classification boundaries are sharp [1]. In order to avoid uncertainty in the classification boundary zone, the second approach is applied in this paper.

The multiple outputs of a single neural network can be modeled using a distributed output code in which each class is assigned a unique codeword, which is a binary string of length n. The columns of the codewords should neither identical nor complementary in order to avoid error correlation [4]. There are various techniques to define a codeword [3, 4, 5]. One of the models using a simple codeword is One-Against-All neural networks (OAA). The length of the codeword used in this model is equal to the number of classes. For a k-class neural network, the codeword for the i-th class can be defined with the length k. The bit in the codeword at the i-th position is equal to 1, and the rest is equal to 0. In the testing phase, a sample is assigned to the i-th class if the network output at the i-th position has the highest confidence value. In order to keep our approach simple, we apply this model to our proposed model.

Hansen and Salamon [6] suggested that ensemble of accurate and diverse neural networks gives better results and less errors than a single neural network. Diversity can be conducted by manipulating input data or output data. Designing the codewords for multiclass neural network is an example of manipulating the diversity using output data [7]. Examples of the ensemble diversity using input data are bagging [8] and boosting [9] neural networks. Bagging provides diversity by randomly resampling the original training data into several training sets whereas boosting provides diversity by manipulating each training set according to the performance of the previous classifier. Diversity can also be provided by using artificial training samples. Melville and Mooney [10] built each training set for a new committee by adding artificially constructed samples to the original training data. They assigned the class label that disagrees with the current ensemble to the constructed sample label.

In this paper, we implement diversity neural network ensembles using two networks which are opposite to each other. Two multiclass neural networks are trained with the same input feature vectors but disagree in the target codewords. The first network predicts the degrees of truth membership and the second network predicts the degrees of false membership. The boundary between these two predicted outputs may not be sharp. Uncertainty can occur in the boundary zone. This paper also estimates this uncertainty and represents it in the form of indeterminacy membership. These three memberships form an interval neutrosophic set [11]. The final decision of the classification is decided from these three memberships.

The rest of this paper is organized as follows. Section 2 presents the basic theory of interval neutrosophic sets. Section 3 explains the proposed model for the muticlass neural network classification with the assessment of uncertainty using interval neutrosophic sets. Section 4 describes the data set and the results of our experiments. Conclusions and future work are presented in Section 5.

2 Interval Neutrosophic Sets

In our previous papers [12, 13], we applied an interval neutrosophic set to the problem of binary classification. We found that an interval neutrosophic set can represent uncertainty information and support the binary classification quite well. In order to expand our approach to represent uncertainty in the multiclass classification, an interval neutrosophic set is also used in this paper.

An interval neutrosophic set is an instance of a neutrosophic set, which is generalized from the concept of classical set, fuzzy set, interval valued fuzzy set, intuitionistic fuzzy set, interval valued intuitionistic fuzzy set, paraconsistent set, dialetheist set, paradoxist set, and tautological set [14]. 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. 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 [14]. In this paper, the memberships are used to represent uncertainty information. This research follows the definition of interval neutrosophic sets that is defined in [11]. This definition is described below.

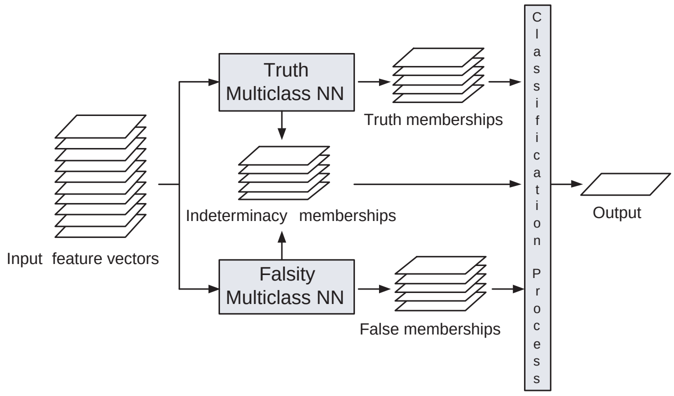

Figure 1: Multiclass neural network model based on interval neutrosophic sets (INS).

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

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

where

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

3 The Proposed Multiclass Classification

In this paper, ensemble of two multiclass neural networks with multiple outputs are created to classify multiple classes. We apply one-against-all neural network model to the two neural networks. Fig.1 shows our proposed model that consists of a set of input feature vectors, two multiclass neural networks, three memberships, a classification process, and a final output.

In this experiment, the Truth Multiclass NN is a feed-forward backpropagation neural network with multiple outputs. This network is trained to predict degrees of truth membership. For a k-class truth neural network, the length of a codeword is equal to k. The codeword for the i-th class has a bit at the i-th position equal to 1, and the rest is equal to 0. The Falsity Multiclass NN is also a feed-forward backpropagation neural network with multiple outputs. It has the same architectures and properties as the ones used for the truth neural network. The only difference is that the falsity network is trained to predict degree of false membership using the complement of target codewords used for training data in the truth network. For example, if the codeword used to train the truth network for the i-th class at the i-th bit is equal to 1 and the rest is equal to 0, then the codeword used to train the falsity network for the i-th class at the i-th bit is equal to 0 and the rest is equal to 1.

For a given unknown input pattern, let T_j be the truth membership of the j-th output for the truth network. Let F_j be the false membership of the j-th output for the falsity network. These two predicted outputs are supposed to be opposite. If the truth membership value is high then the false membership value should be low, and vise versa. Otherwise, uncertainty occurs in the prediction of these two outputs. Hence, one way to represent the degrees of uncertainty in the prediction or indeterminacy membership values can be calculated as the difference between the truth and false membership values. If the difference between these two values is high then the uncertainty is low. In contrast, if the difference between both values is low then the uncertainty is high. Let I_j be the indeterminacy membership of the j-th output. The indeterminacy membership value can be calculated as $I_j = 1 - |T_j - F_j|$.

The three memberships form an interval neutrosophic set. Let A_j be an interval neutrosophic set of the j-th output. A_j can be defined as $A_j = \{x(T_{A_j}(x), I_{A_j}(x), F_{A_j}(x))\}$ where T_{A_j} is the truth membership function of the j-th output, I_{A_j} is the indeterminacy membership function of the j-th output, and F_{A_j} is the false membership function of the j-th output.

Instead of using only the truth membership, we apply the three memberships of an element in an interval neutrosophic set to classify multiple classes. Hence, the predicted binary string is created from the truth, indeterminacy, and false membership values. In order to create the predicted binary string with the length k where k is the number of class, each bit representing each output in the multiple outputs is considered. For the j-th output, if $T_j > F_j$ then the bit in the binary string at the j-th bit is equal to 1. Otherwise, the bit is equal to 0. In the classification, the predicted binary string will be matched to the truth codeword. In order to match these two strings, the predicted bit string must have only one bit equal to 1 and the rest is equal to 0. However, if all bits are equal to 0 then the bit that has the highest indeterminacy membership value will be changed from 0 to 1. If there are more than one bit equal to 1 then the bit that has a value 1 with the minimum indeterminacy membership value will be assigned a value 1, and the rest will be 0. The unknown input pattern is assigned to the i-th class if its predicted binary string matches the codeword that has the i-th bit equal to 1. The degree of uncertainty in the classification can be expressed using the average indeterminacy membership values of the predicted multiple outputs.

4 Experiments

4.1 Data set

Seven data sets from UCI Repository of machine learning data sets [15] are employed in this paper. Table 1 summaries the characteristics of these seven data sets.

Table 1: Data sets used in this study.

Name	No. of Classes	No. of Features	Feature Type	Size of Samples
balance	3	4	numeric	625
ecoli	8	7	numeric	336
glass	7	9	numeric	214
lenses	3	4	nominal	24
wine	3	13	numeric	178
yeast	10	8	numeric	1484
zoo	7	16	numeric, nominal	101

4.2 Experimental methodology and results

In this paper, seven data sets from UCI Repository are used to test our proposed model. These data sets are balance-scale, ecoli, glass, lenses, wine, yeast, and zoo. Each data set is split into a training set containing 80% of the data and a testing set containing 20% of the data. For each UCI data set used in this experiment, twenty pairs of feed-forward backpropagation neural networks with multiple outputs are trained with twenty different randomized training sets.

For each pair of neural networks, the first network is used as the Truth Multiclass NN to predict degree of truth membership and another network is used as the Falsity Multiclass NN to predict degree of false membership. The number of input-node and output-node for each network are equal to the number of features and the number of classes, respectively. Both networks include one hidden layer constituting of $2n$ neurons where n is the number of features. The same parameter values are applied to the two networks and

Table 2: Average classification accuracy for the test data set obtained by applying the proposed model using the three memberships memberships (TIF) and the existing model using only the truth (T) memberships.

Name	%correct (TIF)	%correct (T)
balance	94.32	93.12
ecoli	69.78	61.69
glass	62.67	48.60
lenses	79	84
wine	96.53	94.72
yeast	56.80	56.64
zoo	94.32	93.12

both networks are initialized with the same random weights. The only difference is that the target codewords for the falsity network are equal to the complement of the target codewords used to train the truth network. The indeterminacy membership value is calculated using the different between the truth and false membership values for each pair of network.

After the three membership values are determined for training and test sets, the truth, indeterminacy, and false membership values are used to create the predicted binary string using our proposed technique explained in the previous section. After that, the predicted binary strings are matched to the truth codewords in order to classify multiple classes. For each UCI data set, twenty classification results are averaged. The average percentage of the correct classification results for the test data are shown in Table 2. In this table, the results from our proposed model are compared to the results from the existing one-against-all (OAA) neural network model that applies to only the truth memberships for the multiclass classification. The table shows that six results produced from our technique outperform the results produced from the existing technique.

Furthermore, we also compare our results to the results produced from [16]. In [16], Draghici created the constraint based decomposition (CBD) technique, a constructive neural network technique guaranteed the convergence and can deal with both binary and multiclass problems. In his experiment, some data sets from UCI machine learning repository were applied and each data set was randomly split into 80% training set and 20% test set. He compared the result obtained from CBD and results obtained from other existing machine learning techniques by reporting the average results for test data over five trails. In order to compare the result obtained from our proposed technique to the results obtained from other neural and non-neural machine learning techniques, we compare our results to some of the existing results obtained from his experiment in [16]. Table 3 shows classification accuracy comparison between our proposed technique (column 2) and several existing techniques obtained from [16] (column3-13). These existing techniques include C4.5, C4.5 using classification rules (C4.5r), incremental decision tree induction (ITI), linear machine decision tree (LMDT), learning vector quantization (LVQ), induction of oblique trees (OCI), Nevada backpropagation (NEVP), k-nearest neighbors with k=5 (K5), Q*, and constraint based decomposition (CBD).

In addition, our approach has an ability to represent uncertainty in the classification. Uncertainty in the classification for each input pattern can be calculated as the average of indeterminacy membership values produced from the multiples outputs of the truth and falsity neural networks.

Table 4 shows samples of individual predicted output and their uncertainties resulted from our proposed model for the test set of ecoli data set. The individual predicted outputs for the traditional approach applying only the truth membership values are also shown in this table under the heading (one-against-all). Uncertainty of individual predicted output can be used to enhance and support the confidence in the classification. For example, the actual value for the output in the fifth row of this table is *im*, but our proposed model classifies this output as *imU*, which is wrong. However, uncertainty for this output is 0.2452 which is very high comparing to the maximum uncertainty which is 0.2503. Hence, the decision makers can classify the unknown patterns by using uncertainty value to support their confidence in the classification.

Considering the last row of data from this table, both predicted outputs classified from our approach and the traditional approach are incorrectly classified. The traditional approach cannot provide uncertainty information for this classification, but our approach can explain that the output is misclassified with the uncertainty value 0.2455. Hence, the decision-maker can use this information to support the confidence in decision making. The table has shown a comparison of the uncertainty values as Low, Med and High. It can be seen that in all cases that when the uncertainty values are Low, a correct result is predicted for both approaches. When the values are Medium, 3 out of 4 of the prediction from the traditional approach are wrong where as the proposed approach yield correct results. Finally, when the uncertainty value is High, the prediction has been wrong and attention from the operators should be drawn in such cases.

Table 3: Classification accuracy comparison between our proposed technique (column TIF) and several existing techniques from Draghici [16].

Name	TIF	C4.5	C4.5r	ITI	LMDT	CN2	LVQ	OC1	NEVP	K5	Q*	CBD
balance	94.32	64.61	75.01	76.76	93.27	80.89	89.54	92.5	91.04	83.96	69.21	90.08
glass	62.67	70.23	67.96	67.49	60.59	70.23	60.69	57.72	44.08	69.09	74.78	68.37
wine	96.53	91.09	91.9	91.09	95.4	91.09	68.9	87.31	95.41	69.49	74.35	94.44
zoo	94.32	90.27	90	90.93	96.61	91.91	91.42	66.68	92.86	67.64	74.94	94.29

Table 4: Sample outputs from the traditional classifications based on truth membership values (One-against-all) compared to the proposed model for the test set of ecoli data set (proposed OAA).

Actual value	Predicted value (One-against-all)		Predicted value (proposed OAA)		Uncertainty value	
cp	cp	✓	cp	✓	0.0013	Low
cp	cp	✓	cp	✓	0.0004	Low
im	cp	×	im	✓	0.1368	Med
im	im	✓	im	✓	0.0007	Low
im	im	✓	imU	×	0.2452∗	High
im	cp	×	cp	×	0.2503∗	High
imU	im	×	imU	✓	0.1551	Med
om	omL	×	om	✓	0.1242	Med
pp	pp	✓	pp	✓	0.1286	Med
pp	cp	×	cp	×	0.2455∗	High
	5/10 correct		7/10 correct		3 errors are marked by high uncertainty values	

5 Conclusions and Future Works

In this paper, we apply an interval neutrosophic set to the multiclass classification. Two neural networks with multiple outputs are created for the prediction of the truth membership and false membership values. These two membership values are then used to calculate an indeterminacy membership value. The three membership values constitute an interval neutrosophic set. Multiple interval neutrosophic sets are created for multiple outputs and are used to classify the input patterns into multiple categories. Uncertainty in the classification is calculated from the average indeterminacy membership values produced from the multiple outputs for each pattern. The advantage of our proposed model over a traditional OAA approach is that the indeterminacy membership values provide an estimate of the uncertainty in the multiclass classification. Moreover, our experimental results indicate that our proposed model improves the classification performance compared to the existing OAA model applied only to the truth membership values. In the future, interpolation techniques will be applied to our approach in order to quantify uncertainty in the multiclass classification. Furthermore, we will apply our techniques to a real world problem of well log data analysis in the oil and gas industry.

References:

[1] G. Ou, Y. L. Murphey and L. A. Feldkamp, Multiclass Pattern Classification Using Neural Networks, in *Proceeding of the 17th International Conference on Pattern Recognition (ICPR)*, 2004, pp. 585–588.

[2] Y. L. Murphey and Y. Luo, Feature Extraction for a Multiple Pattern Classification Neural Network System, in *Proceeding of the 16th International Conference on Pattern Recognition (ICPR)*, 2002, pp. 220–223.

[3] R. Erenshteyn, P. Laskov, D. M. Saxe and R. A. Foulds, Distributed Output Encoding for Multi-Class Pattern Recognition, in *Proceeding of the 1oth International Conference on Image Analysis and Processing (ICIAP)*, 1999, pp. 229–234.

[4] T. G. Dietterich and G. Bakiri, Solving Multiclass Learning Problems via Error-Correcting Output Codes, *Journal of Artificial Intelligence Research*, Vol. 2, 1995, pp. 263–286.

[5] K. Crammer and Y. Singer, On the Learnability and Design of Output Codes for Multiclass Prob-

lems, *Machine Learning*, Vol. 47, No. 2-3, 2002, pp. 201–233.

[6] L. K. Hansen and P. Salamon, Pattern Analysis and Machine Intelligence, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, October 1990, pp. 993–1001.

[7] G. Brown, J. Wyatt, R. Harris and X. Yao, Diversity Creation Methods: A Survey and Categorisation, *Journal of Information Fusion*, Vol. 6, No. 1, 2005, pp. 5–20.

[8] L. Breiman, Bagging Predictors, *Machine Learning*, Vol. 24, No. 2, 1996, pp. 123–140.

[9] H. Schwenk and Y. Bengio, Boosting Neural Networks, *Neural Computation*, Vol. 12, No. 8, 2000, pp. 1869–1887.

[10] 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)*, 2003, pp. 505–512.

[11] H. Wang, D. Madiraju, Y.-Q. Zhang and R. Sunderraman, Interval neutrosophic sets, *International Journal of Applied Mathematics and Statistics*, Vol. 3, March 2005, pp. 1–18.

[12] P. Kraipeerapun, C. C. Fung and W. Brown, *Assessment of Uncertainty in Mineral Prospectivity Prediction Using Interval Neutrosophic Set*, Lecture Notes in Artificial Intelligence, Springer Verlag, LNAI 3802, 2005, pp. 1074–1079.

[13] P. Kraipeerapun, C. C. Fung, W. Brown and K. W. Wong, Mineral Prospectivity Prediction using Interval Neutrosophic Sets, in *Proceedings of IASTED International Conference on Artificial Intelligence and Applications*, 2006, pp. 235–239.

[14] H. Wang, F. Smarandache, Y.-Q. Zhang and R. Sunderraman, *Interval Neutrosophic Sets and Logic: Theory and Applications in Computing*, Neutrosophic Book Series, No.5. http://arxiv.org/abs/cs/0505014, May 2005.

[15] D. Newman, S. Hettich, C. Blake and C. Merz, UCI Repository of machine learning databases, 1998. [Online]. Available: http://www.ics.uci.edu/∼mlearn/MLRepository.html.

[16] S. Draghici, The constraint based decomposition (CBD) training architecture, *Neural Networks*, Vol. 14, No. 4–5, 2001, pp. 527–550.