

QUANTIFICATION OF VAGUENESS IN MULTICLASS CLASSIFICATION BASED ON MULTIPLE BINARY NEURAL NETWORKS

PAWALAI KRAIPEERAPUN, CHUN CHE FUNG, KOK WAI WONG

School of Information Technology, Murdoch University
South Street, Murdoch, 6150, Western Australia
E-MAIL: {p.kraipeerapun | l.fung | k.wong}@murdoch.edu.au

Abstract:

This paper presents an innovative approach to solve the problem of multiclass classification. One-against-one neural networks are applied to interval neutrosophic sets (INS). INS associates a set of truth, false and indeterminacy membership values with an output. Multiple pairs of the truth binary neural network and the false binary neural network are trained to predict multiple pairs of the truth and false membership values. The difference between each pair of truth and false membership values is considered as vagueness in the classification and formed as the indeterminacy membership value. The three memberships obtained from each pair of networks constitute an interval neutrosophic set. Multiple interval neutrosophic sets are then created and used to support decision making in multiclass classification. We have applied our technique to three classical benchmark problems including balance, wine, and yeast from the UCI machine learning repository. Our approach has improved classification performance compared to an existing one-against-one technique which applies only to the truth membership values.

Keywords:

Multiclass classification; Feed-forward backpropagation neural network; Vagueness

1. Introduction

Vagueness is normally expected in real world problems. For instance, one cannot exactly define how many grains of sand constitute a heap. This is an example of the Sorites paradox which can be explained using the following questions. Is one grain of sand a heap? The answer is 'no'. If one grain is added, is it turned into a heap? The answer is 'no'. A grain is added one at a time until we have n grains of sand. If n grains of sand are not a heap, and one grain is added. Is it turned into a heap? The answer is still 'no'. However, if n is a very large number, e.g., many millions, the correct result should turn into a heap. The initial condition is true and the following sequence is correct, but the conclusion is false. This situation is called

the Sorites paradox. If a concept is Sorites susceptible, then it should be modeled as a vague concept [1]. In [2], Duckham argued that vagueness deals with the concept of boundaries which cannot be defined precisely. From the previous example, the exact number of n cannot be defined precisely. This means we cannot define the exact boundary for a heap. However, boundary can be considered as a transition zone instead of a single value. Dilo et al. [3] categorized vague objects into three types: vague point, vague line, and vague region. Vague point is a finite set of disjoint sites with known location, but the existence of the sites may be uncertain. In our study, we apply the concept of vague point defined in [3] to our approach.

In general, multiclass neural network classification can be implemented using a single neural network with multiple outputs or multiple binary neural networks. In the first case, the output value is compared to different threshold values attributing to different classes or bins. In the latter case, each network determines a class. Both approaches concentrate only on the "truth" output of the network. In general, the output from the model is compared to a certain threshold value in order to determine whether the input vector is associated with the class. In practice, the threshold values may not be defined precisely. It is Sorites susceptible. Vagueness therefore exists. In multiclass classification, the input features are known but the degree of the existence of the output is uncertain. Therefore, the output of the neural network can be considered as vague point.

In this study, instead of considering only the truth output obtained from a single neural network, we have considered both truth and false output values predicted from a pair of truth and falsity neural networks. These values are then used to deal with the issue of vagueness. Moreover, applying both truth and falsity networks can also increase diversity in neural network ensembles thereby increasing the performance. Diversity can be described as disagreement of classifiers [4]. Hansen and Salamon [5] also suggested that ensemble of accurate and diverse neural

networks provide better results than a single neural network. There are several techniques to manage diversity [6][7][8]. In our study, we deal with diversity by the manipulation of a pair of output targets that are complementary to each other.

In our previous paper [9], we have dealt with the issues of vagueness in multiclass neural network classification using a pair of neural networks with multiple outputs. The first network predicts the truth membership value whereas the second network predicts the false membership value which is supposed to be complement to the truth membership value. The difference between both membership values is considered as the vagueness value. We found that using two opposite neural networks can improve the classification performance compared to the existing technique that deals only with the truth membership values.

In this paper, we extend our previous multiclass classification task by applying a pair of opposite neural networks to the second technique of multiclass neural network classification which is multiple binary neural networks. In general, there are two basic approaches to deal with multiple binary neural networks. These approaches are one-against-all and one-against-one neural networks. In one-against-all approach, k binary neural networks are created to classify a k -class problem, where $k > 2$. Each neural network is trained using the same training data but different target outputs. The i -th label of the i -th neural network is set to '1' and the rest is set to '0'.

In one-against-one approach, $k(k-1)/2$ binary neural networks are created to classify a k -class problem. Each neural network is trained using training data that contains only the i -th label and the j -th label, where $1 \leq i, j \leq k$. The outputs from all networks are then voted in order to classify the input features into multiple classes. One-against-all approach can cause unbalance data among individual neural networks whereas one-against-one approach can cause tie more often but its major advantage over the one-against-all approach is that it provides redundancy that can make the system more generalized [10]. Hence, one-against-one approach is applied to this paper.

In order to represent vague objects, we apply interval neutrosophic sets to represent them. In this research, we follow the definition of interval neutrosophic sets defined by Wang et al. [11]. 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. 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 of type vagueness. For example, let A 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 vague 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:

$$A = \{x(T_A(x), I_A(x), F_A(x)) \mid x \in X \wedge T_A : X \rightarrow [0,1] \wedge I_A : X \rightarrow [0,1] \wedge F_A : X \rightarrow [0,1]\} \quad (1)$$

where

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

2. Multiclass classification using one-against-one neural networks and interval neutrosophic sets

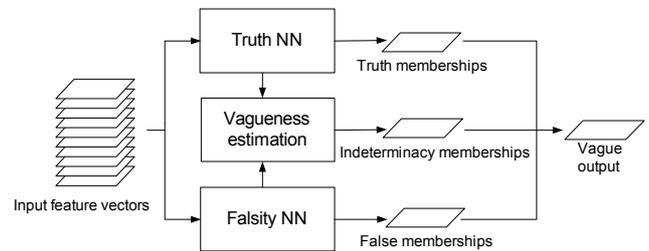


Figure 1. Neural network model based on interval neutrosophic sets

In our proposed one-against-one neural networks, $k(k-1)/2$ components are created. Each component consists of a pair of binary neural networks. Fig.1 represents the proposed component that consists of a set of input feature vectors, a pair of opposite neural networks (Truth NN and Falsity NN), vagueness estimation, three memberships, and a vague output. In each component, both binary neural networks are trained with the same training data from two classes. The truth network is trained to predict degrees of truth membership whereas the falsity network is trained to predict degrees of false membership. Both networks apply the same architecture. However, the falsity network is trained with the complement of the target output values

presented to the truth network. If the output target of the truth network belongs to class i then the target value of the truth network is set to '1' and the target value of the falsity network is set to '0'. In contrast, if the output target of the truth network belong to class j then the target value of the truth network is set to '0' and the target value of the falsity network is set to '1'.

For each pair of membership values in the testing phase, the boundary between both predicted outputs will be sharp if the value of truth membership is 1 and the value of false membership is 0, or vice versa. However, both membership values may not completely complement to each other. Vagueness can occur. This paper deals with this vagueness by considering the difference between these two membership values. If the difference between these two values is high then the vagueness value is low. In contrast, if the difference is low then the vagueness value is high. In this paper, we represent a vagueness value in the form of an indeterminacy membership value. After the three memberships are created, a vague output is then built for each component. In this study, a vague output is represented in the form of an interval neutrosophic set. Therefore, each cell in each vague output consists of three membership values which can be defined as the following.

Let X_p be the p -th output at the p -th component, where $p = 1, 2, 3, \dots, k(k-1)/2$. Let A_p be an interval neutrosophic set in X_p . A_p can be defined as

$$A_p = \{x(T_{A_p}(x), I_{A_p}(x), F_{A_p}(x)) \mid x \in X_p \wedge$$

$$T_{A_p} : X_p \rightarrow [0,1] \wedge$$

$$I_{A_p} : X_p \rightarrow [0,1] \wedge$$

$$F_{A_p} : X_p \rightarrow [0,1]\},$$
(2)

$$I_{A_p}(x) = 1 - |T_{A_p}(x) - F_{A_p}(x)|,$$
(3)

where

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

After $k(k-1)/2$ vague outputs are created; a majority vote is applied in order to classify the input feature vector into multiple classes. In this paper, two voting techniques are proposed and described below.

1. Majority vote based on T>F

For each cell in each vague output, if the truth membership value is greater than the false membership value ($T(x) > F(x)$) then the cell is classified as class i . Otherwise it is classified as class j . After that, the majority vote is applied to all results for each input pattern. If there is only one class that has the highest number of votes then the final predicted output will be assigned to that class. However, if there is more than one class that has the same highest number of votes then the confidence value belonging to each output are considered in order to support the final decision making. We propose two techniques for choosing the class.

a. Randomness

In this technique, we select one of the classes that have the same highest number of votes by random.

b. Vagueness

In order to select the class, a vagueness value is used as the confidence level. The class that has the highest number of votes with the minimum average vagueness value will be chosen.

2. Majority vote based on averaging

For each cell in each vague output, the truth and the complement of the false membership values are averaged. The average output $O_p(x_h)$ at the cell x_h of the p -th vague output can be computed as follow.

$$O_p(x_h) = \frac{T_p(x_h) + (1 - F_p(x_h))}{2}$$
(4)

The cell is assigned to class i if the average output $O_p(x_h)$ is greater than the threshold value of 0.5. Otherwise, the cell is assigned to class j . After that, the majority vote is applied for each input pattern. Similar to the previous technique, if the tie occurs in the classification then the final decision can be made by the supporting of the randomness or considering a vagueness value belonging to each cell.

3. Experiments

3.1. Data sets

In this experiment, we apply three data sets from UCI Repository of machine learning [12] for multiclass classification. Table 1 shows the characteristics of these three data sets. The size of training and testing data used in this experiment are also shown in this table.

Table 1. Data set used in this study

Name	Balance	Wine	Yeast
No. of Class	3	3	10
No. of Feature	4	13	8
Feature Type	numeric	numeric	numeric
Size of Samples	625	178	1484
Size of Training Data	500	142	1186
Size of Testing Data	125	36	298

3.2. Experimental methodology and results

Three data sets named balance, wine, and yeast from UCI Repository are used in this experiment. Each data set is separated into a training set and a testing set. After that, the proposed one-against-one neural networks are applied to each training set. Therefore, each training set is reorganized into $k(k-1)/2$ sub training sets and each subset contains only two classes. Each sub training set is then applied to each pair of feed-forward backpropagation neural networks in order to predict degree of truth membership and degree of false membership values. In this paper, we focus on our approach that aims to increase diversity by creating a pair of opposite output targets. Hence, we apply the same parameter values and the same initial weight to all networks. The number of input-node for each binary neural network is equal to n , which is the number of input features. All networks contain one hidden layer constituting of $2n$ neurons. 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. After the truth and false membership values are predicted, an equation 3 is then used to compute vagueness or indeterminacy membership values.

After $k(k-1)/2$ vague outputs are created, a majority vote is applied. All majority vote techniques described in the previous section are then applied to the vague outputs. We compare the results obtained from our techniques to the results obtained from the existing one-against-one technique that applies only to the truth neural networks. In this existing technique, the threshold value of 0.5 is used to classify the output obtained from each network. After that, the majority vote is applied. If the tie occurs then the class

is selected by random.

In this paper, we do not consider the optimization of the prediction but concentrate only on the improvement of the prediction. For each UCI data set, we try twenty runs with twenty different randomized training data sets. Each run provides the results obtained from the proposed majority vote based on averaging and $T > F$ as well as the existing majority vote based on only T . Twenty classification accuracy results obtained from each technique are averaged. The average results obtained from our approaches and the existing approach are compared and shown in Table 2. This table shows that the results obtained from the proposed techniques outperform the results obtained from the existing technique. We found that the technique of $T > F$ provides similar results comparing to the technique of averaging. This table also shows that two out of three results obtained from using the technique of vagueness as the confidence level provide better results than applying the technique of randomness.

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

Majority vote Technique		Balance %correct	Wine %correct	Yeast %correct
$T > F$	Random	95.52	95.69	58.19
	Vagueness	95.4	96.53	58.86
$\frac{T + (1 - F)}{2} > 0.5$	Random	95.52	95.69	58.19
	Vagueness	95.4	96.53	58.86
$T > 0.5$	Random	93.84	94.44	57.80

4. Conclusion and future work

In this paper, we integrate interval neutrosophic sets with one-against-one neural networks in order to classify the input features into multiple classes. In our approach, $k(k-1)/2$ pairs of the truth and falsity binary neural networks are constitute based on one-against-one technique. The outputs from each pair of networks are represented in the form of a vague output which contains the truth membership, indeterminacy membership, and false membership values. The indeterminacy membership value represents vagueness in the prediction which is the difference between the truth and false membership values. We found that vagueness values can support the decision making in the classification when the tie occurs. In addition,

our experimental results indicate that our proposed one-against-one technique improves the classification performance compared to the existing techniques. In the future, instead of considering only the uncertainty of type vagueness, we will focus on other types of uncertainty such as errors in the multiclass classification.

References

- [1] P. Fisher, "Sorites paradox and vague geographies," *Fuzzy Sets and Systems*, vol. 113, pp. 7–18, 2000.
- [2] M. Duckham, "Uncertainty and geographic information: Computational and critical convergence," in *Representation in a Digital Geography*. New York: John Wiley, 2002.
- [3] A. Dilo, P. Kraipeerapun, W. Bakker, and R. de By, "Storing and Handling Vague Spatial Objects," in *the 15th International Workshop on Database and Expert Systems Applications (DEXA 2004)*, 2004, pp.945–950.
- [4] 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.
- [5] 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.
- [6] T. G. Dietterich, "Ensemble Methods in Machine Learning," in *MultipleClassifier Systems*, ser. Lecture Notes in Computer Science, J. Kittler and F. Roli, Eds., vol. 1857. Springer, 2000, pp. 1–15.
- [7] M. M. Islam, X. Yao, and K. Murase, "A Constructive Algorithm for Training Cooperative Neural Network Ensembles," *IEEE Transactionson Neural Networks*, vol. 14, no. 4, pp. 820–834, July 2003.
- [8] G. Zenobi and P. Cunningham, "Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error." in *ECML*, ser. Lecture Notes in Computer Science, L. D. Raedt and P. A. Flach, Eds., vol. 2167. Springer, 2001, pp. 576–587.
- [9] P. Kraipeerapun, C. C. Fung, and K. W. Wong, "Uncertainty Assessment using Neural Networks and Interval Neutrosophic Sets for Multiclass Classification Problems," *WSEAS Transactions on Computers*, vol. 6, no. 3, pp. 463–470, March 2007.
- [10] 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.
- [11] H. Wang, D. Madiraju, Y.-Q. Zhang, and R. Sunderraman, "Interval neutrosophic sets," *International Journal of Applied Mathematics and Statistics*, vol. 3, pp. 1–18, March 2005.
- [12] D. Newman, S. Hettich, C. Blake, and C. Merz, "UCI Repository of machine learning databases," 1998. [Online]. Available: <http://www.ics.uci.edu/~mllearn/MLRepository.html>