Fog Forecasting Using Self Growing Neural Network "CombNET-II"
— A Solution for Imbalanced Training Sets Problem —

Anto Satriyo Nugroho, Susumu Kuroyanagi, and Akira Iwata
Dept.of Electrical and Computer Eng.,Nagoya Institute of Technology
Gokiso-cho, Showa-ku, Nagoya, 466-8555 Japan.
E-mail:anto@mars.elcom.nitech.ac.jp

Abstract

This paper proposes a method to solve problem that comes with imbalanced training sets which is often seen in the practical applications. We modified Self Growing Neural Network CombNET-II to deal with the imbalanced condition. This model is then applied to practical application which was launched in '99 Fog Forecasting Contest sponsored by Neurocomputing Technical Group of IEICE, Japan. In this contest, fog event should be predicted every 30 minutes based on the observation of meteorological condition. As the result of the contest, CombNET-II achieved the highest accuracy among the participants and was chosen as the winner of the contest. The advantage of this model is that the independency of the branch networks contribute to an effective way of training and the time can be reduced.

1 Introduction

Studies on artificial neural networks are conducted for a long time and various kinds of architectures are being proposed. However working on the application of neural network is still challenging, since many of conditions in the real life do not always satisfy the requirements of the model. As example is that the real-life condition does not always provide balanced information for every case(class). This condition will make several difficulties, when a backpropagation trained neural networks are chosen to solve the problem. Anand [1] reported that backpropagation algorithm converges very slowly for two-class problems in which most of the exemplars belong to one dominant class. Previous studies have been dedicated to deal with this problem, and several algorithms were proposed[1][2].

This imbalanced condition was also found in '99 Fog Forecasting Contest which was sponsored by Neurocomputing Technical Group (NC-TG) of IEICE, Japan. In this contest, meteorological observation was held in location where the observed supercooling-fog event is around 0.3% per year, which is the highest among airports in Japan.

We proposed a solution by Self Growing Neural Network "CombNET-II" which was modified to deal with imbalanced training sets problem. Original version of CombNET-II was dedicated to solve problem which involved large scale of categories to be classified such as Kanji character recognition[3]. Previous studies[3][4] showed that this model achieved a high performance for very complicated classification task, while the network can be decomposed into several modules which are independent of each other. This decomposition contributes to the reduction of time that is needed to train the entire training sets, as long as several machines are used to train the modules independently. Recently, this model has been implemented for industrial products[5] and yet the model is still challenging to be applied for general purposes.

2 Architecture of CombNET-II

CombNET-II is a large scale neural network which consists of 2 parts : Stem Network and Branch Networks as shown in Fig.1. Stem Network employs Self Growing Learning Procedure[3][4] which performs vector quantization of the vector space of the training data. Fig.2 shows the outline of this algorithm. It can be implemented as either an unsupervised type[3][4] or supervised type[5] depending on the characteristics of the problem to be solved.

The vector space is divided into several sub-spaces and then each sub-space becomes the input vector space for one branch network. Branch networks are 3 layer neural networks which are trained by backpropagation to do a finer classification for each sub-space. The number of branch network modules is as many as the number of generated stem neurons.

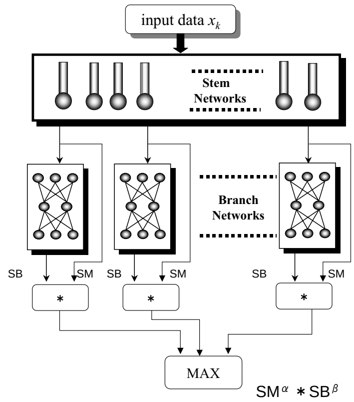

Figure 1: Structure of CombNET-II

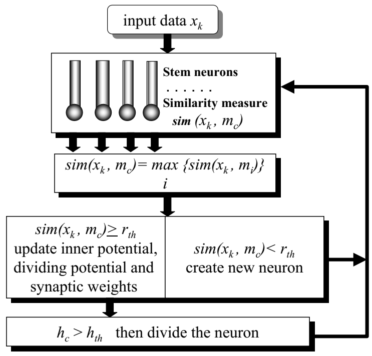

Figure 2: Self Growing Algorithm

3 Fog Forecasting

The database which was used in this experiment was provided by Forecast Department of Japan Meteorological Agency[7]. This database consists of 12 years worth of meteorological information from 1984 until 1995, while the observation was conducted every 30 minutes. This observation was held by Shin Chitose Meteorological Observatory Station, and being used to support the aircraft transportation. The observation was located at : Long.141.70°E, 42.77°N Lat., 25 m above sea level. Note that the observed supercooling-fog events in this location is around 0.3% per year, which is the highest among airports in Japan. "Supercooling-fog" is defined as a condition where the fog appears while the temperature is under 0°C and the range of visibility is less than 1000 m. The original data were expressed as telegraphic messages, which were converted into numeric expressions by Numerical Forecasting Section of Japan Meteorological Agency. The observation included 26 items as shown in Tab.1.

No.	item	No.	item	No.	item
1	Year	10	Max. Inst. Wind Speed [m/s]	19	Cloud Shape (2nd layer)
2	Month	11	Change of Wind (1) [°]	20	Cloud Height (2nd layer)
3	Date	12	Change of Wind (2) [°]	21	Cloudiness (3rd layer)
4	Time	13	Range of Visibility [m]	22	Cloud Shape (3rd layer)
5	Atmospheric Pressure [hPA]	14	Weather	23	Cloud Height (3rd layer)
6	Temperature [°C]	15	Cloudiness (1st layer)	24	Cloudiness (4th layer)
7	Dew Point temperature [°C]	16	Cloud Shape (1st layer)	25	Cloud Shape (4th layer)
8	Wind Direction [°]	17	Cloud Height (1st layer)	26	Cloud Height (4th layer)
9	Wind Speed [m/s]	18	Cloudiness (2nd layer)		

Table 1: Meteorological informations which are provided in the database

Year	1984	1985	1986	1987	1988	1989
number of fog event	467	426	314	275	282	251
number of no fog event	16961	17033	17130	17172	17260	17218
ratio	1:36.3	1:40	1:54.5	1:62.4	1:61.2	1:68.6
Year	1990	1991	1992	1993	1994	1995
number of fog event	220	220	389	211	298	288
number of no fog event	17199	17272	17163	17301	17211	17222
ratio	1:78.1	1:78.5	1:44.1	1:82	1:57.8	1:59.8

Table 2: Ratio of number of patterns of each class

The database was then classified into 2 classes : "fog event" and "no fog event" depending on the value of "range of visibility" and "weather". In the database fog event is defined as condition when "range of visibility" is less than 1000 m and "weather" shows the appearance of fog. The classification results for all years are shown in Tab.2.

Tab.2 shows that the ratio of two classes is extremely imbalanced, which leads to a difficult condition when backpropagation-trained neural networks are used. Anand [1] reported that the major disadvantage of backpropagation is the slow rate of convergence of net output error, because the negative gradient vector computed by backpropagation for an imbalanced training set does not initially decrease the error for the subordinate class (class which has fewer samples). To deal with this imbalanced condition we proposed a solution by the use of modified CombNET-II[6].

4 Strategy to deal with Imbalanced Training Sets Problem

To deal with imbalanced training sets problem, we apply the Self Growing Algorithm to perform vector quantization of the dominant class (class which has larger number of samples). The purpose of this algorithm is to partition the vector space of dominant class based on the statistical distribution condition.

Let the training set $T = T_0 \cup T_1$, where each class C_k contains samples as follows :

$$T_k = \{x_j^{(k)} : j = 1, \cdots, n_k\} \qquad for\ k = 0, 1 \tag{1}$$

"0 " stands for "no fog event" class while "1" stands for "fog event". n_k represents the number of class k. Tab.2 shows that $n_0 \gg n_1$.

The proposed algorithm is stated as follows :

1. Stem Network is trained to divide input vector space of T^0 into R sub-spaces : $\{T^{(0,r)} : r = 1, \cdots, R\}$, where
$$T^0 = T^{(0,1)} \cup T^{(0,2)} \cup \cdots \cup T^{(0,R)} \tag{2}$$
and $\{x_i^{(0,r)} : i = 1, \cdots, n^{(0,r)}\}$ are the patterns belong to $T^{(0,r)}$. Each sub-space is represented by a stem neuron μ_r as a mean vector of the patterns belongs to the sub-space of interest.
$$\mu_r = \frac{1}{n^{(0,r)}} \sum_{i=1}^{n^{(0,r)}} x_i^{(0,r)} \tag{3}$$
2. Each generated subspace is then unified with sub ordinate class.
$$T'_r = T^{(0,r)} \cup T^1 \qquad where\ r = 1, 2, \cdots, R \tag{4}$$
3. T'_r is used to train the branch networks. The number of branch networks equal with the generated stem neurons.

By applying the algorithm, the original problem is turned to more balanced problems to be solved by branch networks. Another advantage of this algorithm is that branch networks can be trained independently. This will reduce the time needed to train all modules as long as several machines are possible to use.

To do prediction, test pattern is presented to all stem neurons, and the distance between test vector x and stem neurons μ_r are calculated. Few of the closest neurons are selected, and test vector is presented to the branch networks of the selected neurons. Final score Z is obtained by the next equation :

$$Z = SM^{1-\alpha} \times SB^{\alpha} \tag{5}$$

$$0.0 \leq \alpha \leq 1.0$$

SM stands for Score of Matching which is defined as similarity between test vector and stem neuron.

$$sim(x, \mu_r) = \frac{x \cdot \mu_r}{|x||\mu_r|} \tag{6}$$

SB stands for Score of Branch which is defined as the maximum score of the branch layers. The prediction result is the classification result of branch network corresponding to the stem neuron μ_r which gives the highest score of Z.

5 Experiments and Discussion

This section presents the results of the experiments to evaluate the performance of CombNET-II to solve the imbalanced training sets problem. In this experiment, some data were selected from the database as training set, while the remains were used for performance evaluation. Therefore, we made training set which consisted of meteorological data from 1988, 1990, 1991, 1992, 1993. Since the database contained lots of dummies which were expressed as -1 (for temperature, dummy was expressed as 999 to avoid ambiguity with -1°C) , normalization to the interval value of [-1,+1] was appropriate choice while the dummies were turned into 0. Dummy in this database means that there was no observation. After this classification, we found that number of fog event patterns and no fog event patterns of training set were 1322 and 86195 respectively. In this experiment we dropped "year","month","date","time", "range of visibility", "weather"(as requirement in the fog forecasting contest, prediction should be done based on the meteorological information where "range of visibility" and "weather" are excluded), and used the other meteorological informations of the last 1 hour 30 minutes to create one input vector. It is because our experiments showed that inviting the previous information tend to give a better performance.

Data of "no fog event" were then presented to stem network. Stem network applied Self Growing Algorithm[3][4], which performed vector quantization to the vector space of dominant class. The number of sub-spaces was controlled by adjusting 2 parameters of this algorithm, i.e inner potential threshold h_{th} and similarity r_{th}. In this experiment, the dominant class was partitioned into 18 spaces which were represented by 18 stem neurons. The number of sub-spaces in this experiment was not optimized, since the optimization would need a long time. Each partition was then unified with the "fog event class", and was presented to branch network. The number of neurons for input layer, hidden layer and output layer of branch networks were 80, 100 and 2 respectively.

Training phase was performed in UNIX environment, and several Pentium II 400 MHz Solaris 2.6 machines were used. To train all the modules, we needed only 3 days (5 machines were used in the experiment), while multilayer perceptron which was trained by backpropagation needed around 3 weeks to learn the entire data of training set. The major advantage of this model is that the decomposition of the task to several branches permits us to use several machines simultaneously which lead to a time efficiency. We also attempted to solve the same problem using nearest neighbor classifier, where all data of the training set were used as templates. The result was also compared with backpropagation trained 3 layered MLP which took a same structure as branch network and was trained at 3000 times iteration.

Fig.3 shows the correct prediction rate achieved by CombNET-II at several value of α. We analyzed the correct prediction rate of each class, and found that the correct prediction rate for "fog event" class increased as the value of α increased, while for "no fog event" showed the opposite. Parameter α was chosen by taking the value which gave the smallest number of wrong predicted patterns for both categories, as the requirement of the contest. $\alpha = 0$ was chosen since it gave the smallest number of wrong predictions. It implied that the score Z depended only to the similarity value of input test vector and stem neurons. In other words, the winner was the stem neuron μ_r which had the largest similarity with the input pattern x. Therefore, the forecasting result was the classification result of branch network corresponding to the winner μ_r.

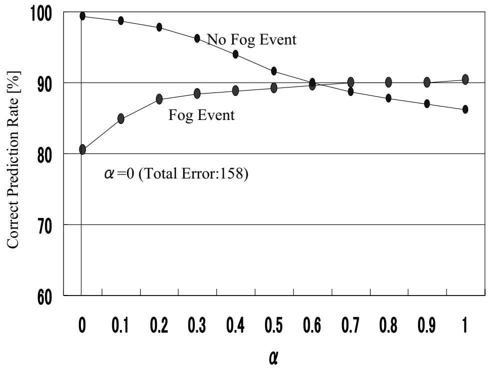

Figure 3: Result for each class(CombNET-II)

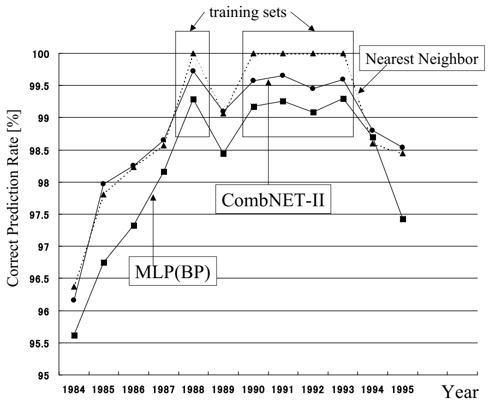

Figure 4: Performance comparison

The results are shown at Fig.4 and Tab.3. The figure inside the brackets is the number of wrong predictions. Note that the prediction rate is calculated as ratio between the number of right predictions for all classes and the entire data.

Year	1984	1985	1986	1987
CombNET-II	96.16(669)	97.97(355)	98.26(304)	98.65(236)
Nearest Neighbor Classf.	96.37(633)	97.80(385)	98.23(308)	98.57(249)
MLP(BP)	95.62(763)	96.75(567)	97.32(467)	98.16(321)
Year	1989	1994	1995	
CombNET-II	99.10(158)	98.80 (210)	98.54(256)	
Nearest Neighbor Classf.	99.06(164)	98.60(246)	98.39(282)	
MLP(BP)	98.44(272)	98.70(219)	97.43(450)	

Table 3: Correct Prediction Rate [%] of CombNET-II, NN-Classf. and MLP for test sets

The result shows that CombNET-II gave better correct prediction rate for test sets than MLP and nearest neighbor classifier. We analyzed the correct prediction rate for "fog event class" of CombNET-II and nearest neighbor classifier, and found that CombNET-II achieved higher rate than nearest neighbor classifier. For test data of 1989, CombNET-II achieved 80.48% while nearest neighbor classifier was 60.56%. This condition also occured for all years of test set. This result shows that score calculation which involved SM gave success predictions for patterns of fog event class which were very closed to "no fog event" class. Nearest neighbor classifier could not avoid misclassification for such patterns, since the distance calculations were linearly performed. This result also shows that the forecasting fog event is really a hard problem, since nearest neighbor classifier failed to do fine predicition, even the entire data of training set were used as templates. Therefore, this fact shows that CombNET-II promises a better approach for classification task than nearest neighbor.

Tab.4 summarized the comparison of CombNET-II, MLP and Nearest Neighbor Classifier. CombNET-II needed only 150,876 synapses which was much smaller than nearest neighbor classifier. The size was 18 times larger than MLP's. However, task decomposition of CombNET-II allowed the reduction of the time which was needed to train, since several machines were used to train branch modules simultaneously.

	CombNET-II	Nearest Neighbor Classifier	MLP (BP)
Correct Prediction Rate (for 1989 data)	99.10 %	99.06 %	98.44 %
Num. of wrong predictions (for 1989 data)	158	164	272
Number of synapses(training)	150876	7001360	8302
Ratio	18.2	843.3	1.0
Number of synapses(prediction)	9742	7001360	8302
Ratio	1.2	843.3	1.0
Network Structure	Stem Neuron : 18 Branch : 80-100-2	2 classes (87517 templates)	80-100-2

Table 4: Comparison of CombNET-II, Nearest Neighbor Classifier, and MLP(BP)

6 Conclusion

This paper proposes a solution for imbalanced training sets classification problem by applying Self Growing Neural Network CombNET-II. This model was implemented for fog forecasting problem, and the proposed method showed a good result when compared to nearest neighbor classifier and MLP. Another advantage is that the model permits task decomposition which leads the time efficiency. We participated in "1999 Fog Forecasting Contest" which was sponsored by Neurocomputing Technical Group of IEICE Japan as of part of 1999 IEICE General Conference, and CombNET-II achieved the best result among the participants[7].

Future work will be focused to improve the correct prediction rate of the model for subordinate class.

7 Acknowledgements

The authors wish to express their appreciation to Forecast Department of Japan Meteorological Agency for providing meteorological database.

References

[1] R. Anand, K.G. Mehrotra, C.K. Mohan, and S.Ranka, "An Improved Algorithm for Neural Network Classification of Imbalanced Training Sets", *IEEE Trans.Neural Networks*, vol.4, pp.962-969, 1993.

[2] B.L. Lu and M. Ito, "Decomposition and Parallel Learning on Imbalanced Classification Problems by Min-Max Modular Neural Network", *ICONIP98-Kitakyushu*, vol.1, pp.199-202, 1998.

[3] K. Hotta, A. Iwata, H. Matsuo, and N. Suzumura, "Large Scale Neural Network "CombNET-II" ", *Tech. and Res.Rep. of IEICE*, Vol.75-D-II, no.3, pp.545-553, 1992.

[4] A. Iwata, Y. Suwa, Y. Ino, and N. Suzumura, "Hand-written Alpha-Numeric Recognition by a Self-Growing Neural Network "CombNET-II"", *IJCNN92-Baltimore*, June 7-11, 1992

[5] H. Kawajiri, T. Yoshikawa, J. Tanaka, Anto S.N. and A. Iwata, "Hand-written Numeric Character Recognition for Facsimile Auto-dialing by Large Scale Neural Network "CombNET-II"", In *Proc. of the 4th.EANN*, Gibraltar, pp.40-46, June 10-12,1998.

[6] Anto S.N., S. Kuroyanagi, A. Iwata, "Fog forecasting by "CombNET-II"", In *Proc. of 1999 IEICE General Conference*, Yokohama, March 25-28,1999, pp.323-324

[7] Explanation about meteorological database and the result of Fog Forecasting Contest (in Japanese) can be seen at
http://www.bpe.es.osaka-u.ac.jp/~shouno/IEICE_NC/