確率的フレームワークとサポートベクターマシンを用いた大規模パターン認識モデル CombNET-III
CombNET-III: a Support Vector Machine Based Large Scale Classifier with Probabilistic Framework

クグレ マウリシオ (PY)†, 黒柳 奨†, アント サトリヨ ヌグロホ‡, 岩田 彰†
Mauricio Kugler(PY), Susumu Kuroyanagi, Anto Satriyo Nugroho, Akira Iwata
† 名古屋工業大学大学工学研究科情報工学専攻
‡ 中京大学生命システム工学部
mauricio@kugler.com, bw@nitech.ac.jp, nugroho@life.chukyo-u.ac.jp, iwata@nitech.ac.jp

Abstract— CombNET-II was one of the first methods proposed for solving classification problems with large number of categories. This paper proposes an extension to this model, named CombNET-III, which replaces the MLP branch networks with multiclass SVMs and introduces a new probabilistic framework. The experiments show that CombNET-III outperforms both CombNET-II and a single multiclass SVM.

Keywords— **large scale classification problems, support vector machines, probabilistic framework, divide-and-conquer**

1 Introduction

Several research fields have to deal with very large classification problems, and many authors addressed those that present large number of samples. The majority of the large-scale classification methods, however, are not appropriate for problems containing large numbers of classes, for instance, classifying thousands of categories. This kind of problem usually also presents a large number of samples and/or features, as in the case of human-computer interface applications.

The CombNET-II model proposed by Hotta *et al.* [1] was one of the first divide-and-conquer based large scale classifiers specifically developed for dealing with classification problems composed by thousands of categories. It consists of a sequential clustering VQ based gating network (stem network) and several Multilayer Perceptron (MLP) based expert classifiers (branch networks). It has presented several good results in Chinese character (Kanji) recognition and some other specific applications. However, as the CombNET-II was originally developed for character recognition tasks, its application in different kinds of problems is not straightforward. Also, the Multilayer Perceptron (MLP) based branch networks presents a large processing time and problems of local minima.

The main objectives of this work are the improvement of the CombNET-II performance by the application of more modern pattern recognition algorithms and to develop a generic framework in order to enable its application in different scenarios. In order to accomplish this, a new model is introduced — the CombNET-III.

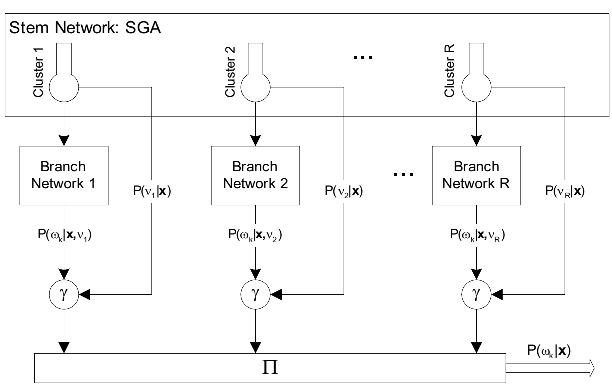

Figure 1: CombNET-III structure

2 Proposed Model: CombNET-III

The main modification to CombNET-II proposed in this paper is the substitution of the MLP branch networks by multiclass Support Vector Machines (SVM) [2] based branch networks. Moreover, as mentioned by many authors, a classifier should output posterior class probabilities to allow post processing [3, 4] and application in different scenarios. In order that, this paper introduces a new probabilistic framework for combining the stem and branch networks information.

The change from MLPs to SVMs is not straightforward. The SVMs unlimited output function and different output ranges among classifiers make the output combination inefficient [4]. Many approaches address the problem of converting the SVM output in a calibrated probability. In this paper, Platt's methodology [3] was used, which directly converts the function values to posterior probabilities by fitting the SVM output with a sigmoidal function.

As the binary classifiers corresponding to one class were trained with the same samples of that class, their output probabilities are not statistically independent. Thus, given a coding matrix $\mathbf{M}^{K\times H}$ in which K is the number of classes and H is the number of classifiers, $m_{k,h}=\{-1,0,+1\}$ and zero entries are interpreted as "don't care", the probability of class ω_k given an unknown sample $\mathbf{x}$ and a cluster ν_j is defined as the average probability outputted by the classifiers containing that class. The proposed decoding function hence becomes:

$$P\left(\omega_k\left|\mathbf{x},\nu_j\right.\right)=\frac{\sum\limits_{h:m_{k,h}\neq 0}P\left(y_{k,h}=m_{k,h}\left|\mathbf{x}\right.\right)}{\sum\limits_{h=1}^{H}\left|m_{k,h}\right|} \tag{1}$$

As a branch network cannot give any information about the categories that it was not trained to recognize, it is assumed that:

$$\omega_k \notin \nu_j \rightarrow P\left(\omega_k \left| \mathbf{x}, \nu_j \right.\right) = \frac{1}{2} \tag{2}$$

The cluster probability $P\left(\nu_j \left| \mathbf{x} \right.\right)$ represents the confidence of each branch network output, i.e, it weights between the branch network outputs and 1/2. Hence, the final posterior probability of the class ω_k given an unknown sample $\mathbf{x}$ is calculated as the product of the probability of class ω_k given by each branch network weighted by the respective cluster probability. Finally, the proposed framework final equation can be written as:

$$P\left(\omega_k \left| \mathbf{x} \right.\right) = c \prod_{j=1}^{R} \left[P\left(\nu_j \left| \mathbf{x} \right.\right)^{\gamma} P\left(\omega_k \left| \mathbf{x}, \nu_j \right.\right)^{1-\gamma} + \right.$$
$$\left. + \frac{1 - P\left(\nu_j \left| \mathbf{x} \right.\right)^{\gamma}}{2} \right] \tag{3}$$

where γ is a weighting factor and c adjusts the probabilities scale in order to ensure they are calibrated, summing to unity. The final structure of the CombNET-III is shown diagrammatically in Figure 1.

3 Experiments

The *Kanji400* database consists of a subset of the first 400 categories of the ETL9B database [1]. The performance of the proposed model CombNET-III was compared with the previous model CombNET-II, a single multiclass SVM and the k-NN method. All experiments were performed using in-house developed software packages.

Each categories contains 200 samples, from which 150 samples were used as the training set and 50 samples as the test set. The characters were resized by their largest dimension and the peripheral direction contributivity (PDC) feature extraction method was applied. For all classifiers except the k-NN, before the features normalization, each sample vector was independently normalized to a unitary maximal feature value and zero feature mean. Also, for all classifiers, the data was normalized to a zero mean and a unitary standard deviation.

For the CombNET-II experiments, the MLP neural networks were trained until the mean square error (MSE) was smaller than 10^{-4} or the iteration number exceeds 500, with learning rate equal to 0.9 and momentum 0.1, while the number of hidden neurons and the γ parameter were optimized for each experiment realization. In the case of the single SVM and the CombNET-III, the binary SVM classifiers had non-biased output, a One-versus-One output encoding and a Gaussian kernel function, whose parameter σ was optimized for each experiment realization. The soft-margin C parameter was fixed at 200.

Both divide-and-conquer models CombNET-II and CombNET-III used the same 12-cluster stem network, which was trained with similarity threshold and inner potential threshold respectively equal to -1 and 53, giving an accuracy of 78.70%. Figure 2 depicts the classification accuracy results for the proposed method and all compared methods. For the divide-and-conquer methods, it is also shown the branch networks average accuracy. The proposed model outperformed the other methods, reducing the single SVM error rate by around 16% and the previous model CombNET-II by around 26%.

[1] Available under request from http://www.is.aist.go.jp/etlcdb

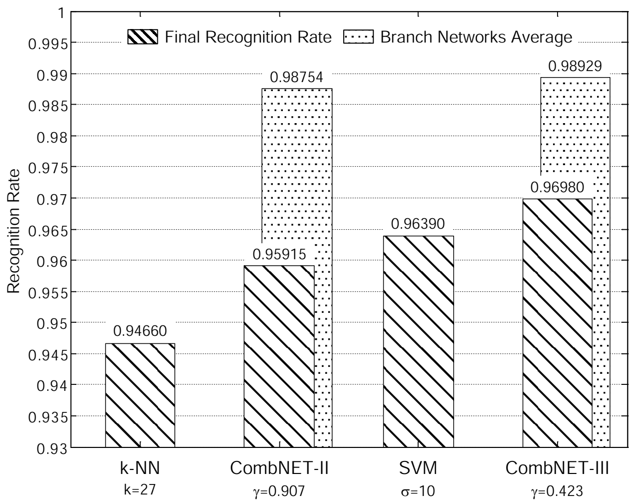

Figure 2: Recognition rate results comparison for the *Kanji400* database

4 Conclusions

This paper proposed an extension of the previous large scale classification model CombNET-II in order to increase classification accuracy and flexibility, the last one achieved by the introduction of a new probabilistic framework.

Substituting the MLP branch networks by multiclass SVMs with moderated outputs permitted the first objective to be accomplish. The local effect of the Gaussian kernel function reduces the interference between the clusters, as the SVM function value tends to be zero for outlier samples. Finally, the final classification accuracy of CombNET-III outperformed all the compared methods (k-NN, single SVM and CombNET-II), showing that the proposed framework and the use of SVM branch networks are effective.

References

[1] K. Hotta, A. Iwata, H. Matsuo, and N. Susumura, "Large scale neural network CombNET-II," IEICE Transactions on Information & Systems, vol.J75-D-II, no.3, pp.545–553, March 1992.

[2] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol.20, no.3, pp.273–297, 1995.

[3] J.C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in Advances in Large Margin Classifiers, ed. A.J. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, pp.61–74, MIT Press, Cambridge, MA, March 1999.

[4] A. Passerini, M. Pontil, and P. Frasconi, "New results on error correcting output codes of kernel machines," IEEE Transactions on Neural Networks, vol.15, no.1, pp.45–54, January 2004.