Model Size Selection for Prototype-Based Classifiers using Large Geometric Margin Minimum Classification Error Training

David Ha† Hideyuki Watanabe†† Shigeru Katagiri† Miho Ohsaki†
†Doshisha University
††National Institute of Information and Communications Technology

1. Introduction

To avoid the over-training problem in pattern classifier developments, we propose a new training scheme, which is based on the Large Geometric Margin Minimum Classification Error (LGM-MCE) training [1], for selecting an optimal class model size for classifiers.

2. LGM-MCE training

The LGM-MCE training is characterized by the following geometric-margin-based misclassification measure:

$$D(\vec{x},\Lambda) \approx \frac{d_y(\vec{x},\Lambda)}{\|\nabla_{\vec{x}} d_y(\vec{x},\Lambda)\|}$$

where $\vec{x}$ is a training sample, $d_y(\vec{x},\Lambda)$ is a functional margin misclassification measure, and Λ is a class model. Maximizing the geometric margin is shown to be effective in avoiding the over-training [2]. Utilizing this effect, the LGM-MCE training conducts the concurrent optimization, i.e., the minimization of classification error counts and the maximization of geometric margin values, aiming at the accurate classification of future samples.

3. Proposed method

In the LGM-MCE training, the smoothness of the classification error count loss plays a key role in facilitating the concurrent optimization, and a procedure for automatically optimizing the smoothness was successfully developed [3]. Using this recent advancement, we formalize a method for automatically optimizing the class model size as shown in Figure 1. For presentation clarity, we adopt in the formalization a multi-prototype classifier that handles fixed-dimension vector patterns. The method aims at finding a minimal size ensuring sufficient accuracy on the training data while avoiding overfitting, by introducing some geometric misclassification tolerance justified by the smoothness. Hence the size is incremented from small values then stopped once the geometric margin is sufficient.

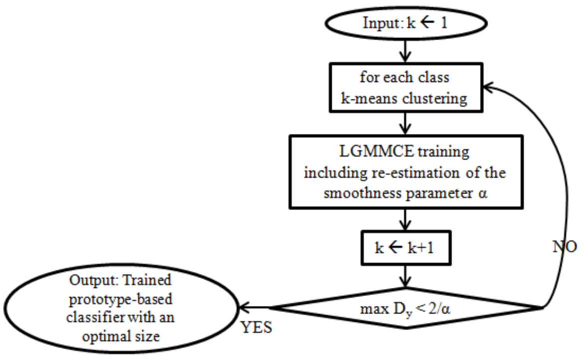

Figure 1. Proposed model size selection algorithm.

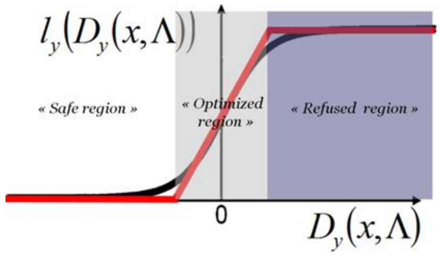

Figure 2. Justification of the geometric misclassification tolerance.

4. Results

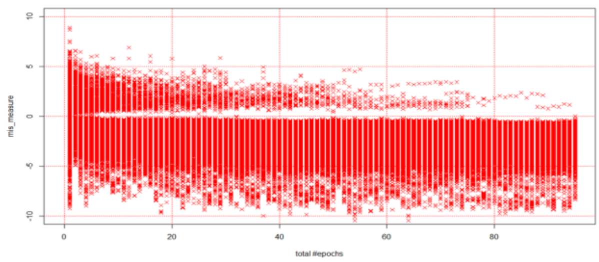

Figure 3. Evolution of the training samples misclassification measures against model size selection epoch described in Figure 1.

To confirm the feasibility of selecting the model size based on a geometric misclassification tolerance, experiments were conducted on the letter recognition dataset by using 4000 samples for the training set (Figure 3). It can be seen that misclassification measures are decreasing overall, which confirms that a higher model size can ensure fewer misclassifications on the training set. More importantly, misclassification measures are efficiently shifted to the negative region, which clearly demonstrates the effectiveness of the LGM-MCE training. Moreover, once the model size is sufficiently high, the misclassification measures can be shifted below the tolerance threshold, hence the selected model size in this case is around 80 prototypes per class.

5. Conclusions and future works

So far, the method performs reasonably well on the tested datasets. However the tolerance criterion based on the smoothness value seems slightly restrictive, hence a future work would consist in a more accurate estimation of the smoothness parameter.

Acknowledgements: This work was supported in part by JSPS Grants-in-Aid for Scientific Research No. 26280063 and MEXT-Supported Program "Driver-in-the-Loop".

Bibliography

[1] H. Watanabe, et al. (2013). J. Sig. Proc. Sys., vol. 74, pp. 297-310.
[2] Bishop, C. (2006). Pattern Recognition and Machine Learning, pp 325-345. Boston: Springer.
[3] H. Watanabe, et al. (2013). J. Sig. Proc. Sys., vol. 74, pp. 311-322.