

# Pattern Recognition with the Decomposed Multilayer Perceptron

S. Lucas, Z. Zhao, G. Cawley and P. Noakes

## Abstract

The Letter reports the benefits of decomposing the multilayer perceptron (MLP) for pattern recognition tasks. Suppose there are  $N$  classes, then instead of employing 1 MLP with  $N$  outputs,  $N$  MLPs are used each with a single output. In practice, this allows fewer hidden units to be used than would be employed in the single MLP. Furthermore, it is found that decomposing the problem in this way allows convergence in fewer iterations, and it becomes straight forward to distribute the training over as many workstations as there are pattern classes. The speedup is then linear in the number of pattern classes, assuming there are as many processors as classes. If there are more classes than processors, then the speedup is linear in the number of processors. It is shown that on a difficult handwritten OCR problem, the results obtained with the decomposed MLP are slightly superior than those for the conventional MLP, and obtained in a fraction of the time.

**Introduction:** This Letter reports the benefits of decomposing the MLP for pattern recognition tasks. Suppose we have  $N$  classes, then instead of employing 1 MLP with  $N$  outputs, we use  $N$  MLPs, each with a single output. This raises at least two interesting questions.

1. Suppose we have a large network of workstations that lie mostly idle overnight. Can we easily exploit this power to learn our problem faster?
2. Can we improve recognition rate on difficult problems by simplifying the learning task?

The answer to the first of these questions turns out to be yes, whereas initial results suggest a maybe in the answer to the second.

The decomposition idea occurs naturally in personal verification tasks (e.g. speaker verification [1]), but seems to be seldom reported for recognition problems, and is certainly not the norm. The results presented below suggest that it should be.

Figure 1: Standard MLP with one output for each of  $N$  classes

**Decomposition:** The decomposition is conceptually straightforward. Instead of having a single MLP with one output (Fig. 1), we just have  $N$  decomposed MLPs (DMLP), each with a single output (Fig. 2). But what about the hidden units? Naïvely, we might suppose that if we needed  $H$  units in the MLP, then perhaps we could get by with  $(H/N)$  units in the DMLP. Clearly, this simplistic reasoning fails for problems such as the encoder, where if we regard each output as a class, then we only require  $\log_2 N$  hidden units. However, for difficult pattern recognition tasks we find that such minimal hidden layers do not perform at all well.

**(i) Connections per presentation:** The time taken to process a pattern is (to a good approximation) proportional to the number of connections in the network. Given a fully-connected MLP with  $I$  inputs,  $H$  hidden and  $O$  output nodes the number of connections is

$$C = (I \times H) + (H \times O)$$

Two useful measures of performance are (millions of) connections per second (MCPS) and connection updates per second (MCUPS).

Figure 2: Decomposed MLP with single output and smaller hidden layer

## References

- [1] J. Oglesby and J. Mason. Optimisation of neural models for speaker identification. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pages 393–396, Toronto, Canada, 1990.