A Comparison of Crossover Operators in Neural Network Feature Selection with Multiobjective Evolutionary Algorithms

Christos Emmanouilidis and Andrew Hunter

Centre for Adaptive Systems,
School of Computing, Engineering and Technology
University of Sunderland,
St. Peter's Campus, Sunderland, SR6 ODD, UK
{christos.emmanouilidis, andrew.hunter}@sunderland.ac.uk

Abstract

Genetic Algorithms are often employed for neural network feature selection. The efficiency of the search for a good subset of features, depends on the capability of the recombination operator to construct building blocks which perform well, based on existing genetic material. In this paper, a commonality-based crossover operator is employed, in a multiobjective evolutionary setting. The operator has two main characteristics: first, it exploits the concept that common schemata are more likely to form useful building blocks; second, the offspring produced are similar to their parents in terms of the subset size they encode. The performance of the novel operator is compared against that of uniform, 1 and 2-point crossover, in feature selection with probabilistic neural networks.

1 INTRODUCTION

Evolutionary algorithms (EAs) are increasingly employed in neural network modelling for tasks, ranging from evolution of connection weights, network architectures and learning rules to evolution of inputs, control parameters and ensembles of networks [1]. In particular, EAs have been used to aid the selection of feature subsets in various classification tasks (e.g. [2] [3]). Recently, the use of Multi Objective Evolutionary Algorithms (MOEA) has been suggested for feature selection [4]. In addition, a novel, commonality-based crossover operator has been introduced, called Subset Size Oriented Common Features (SSOCF) operator [5]. When put in a multiobjective evolutionary setting, the SSOCF operator can facilitate the search for good subsets of features. This is achieved, first by preserving building blocks with promising performance, and second by promoting useful population diversity across the range of Pareto optimal solutions. In multiobjective optimisation, a key concept is that of Pareto optimality. Solutions are compared against each other in terms of Pareto dominance, i.e. a solution is dominant over another only if it has better performance in at least one criterion and non-inferior performance in all criteria. A solution is said to be Pareto optimal if it cannot be dominated by any other solution in the search space. In complex search spaces, wherein exhaustive search is infeasible, it is very difficult to guarantee Pareto optimality. Therefore, instead of the true set of optimal solutions (Pareto Set), one usually aims to derive a set of non-dominated solutions with objective values as close as possible to the objective values (Pareto Front) of the Pareto Set. Feature selection is well-suited to multiobjective optimisation. In the simplest case, it involves two objectives: feature subset size minimisation and performance maximisation. In this paper, a variation of the niched Pareto GA (NPGA) [6] is employed. This is known to be a fast MOEA [7], since tournament domination is determined by a random subsample of the population. However, any MOEA could be employed in this setting. Details of the MOEA employed in this work can be found in [5]. This paper examines the performance of the SSOCF operator against *n*-point crossover operators in multiobjective evolutionary feature selection.

2 SUBSET SIZE-ORIENTED COMMON FEATURES CROSSOVER

Common uniform or *n*-point crossover operators can be disruptive, since they may result in breaking up useful building blocks. When the aim is to identify good subsets of features for different subset sizes, common crossover operators can have an additional negative side effect. A standard crossover operating on two individuals, coding subsets of size n and m, tends to yield offspring with complexity approximately $(n+m)/2$. Therefore, the EA tends to explore mostly medium-sized subsets, while the edges of the non-dominated front are less well explored. In [5], the SSOCF crossover operator is introduced, a commonality-based operator, which helps preserving building blocks of promising performance. It also yields offspring populations with relatively even distribution, across the range of the Pareto front, while it does not require mating restrictions. It exploits the concept that preserving the maximal common schema of two parents results in a more creative recombination strategy, compared to standard crossover. This concept has been recently termed the Commonalty-Based Crossover Framework [8].

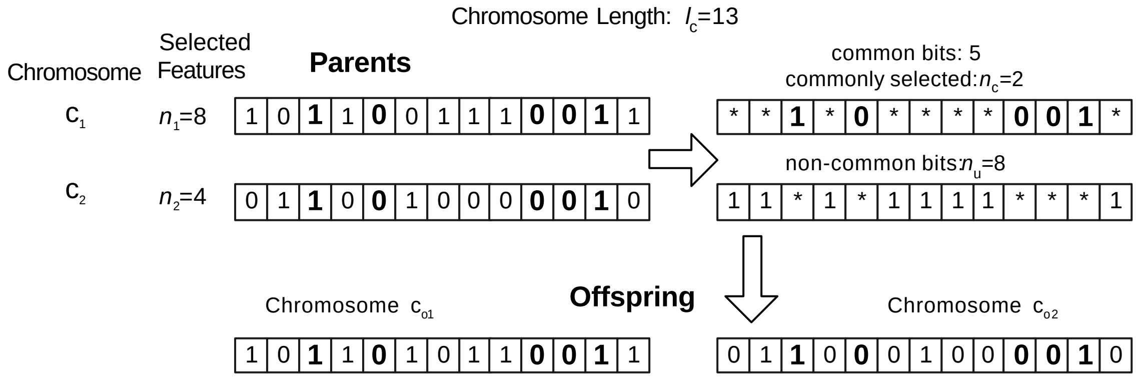

Figure 1: Example of the functionality of the Subset Size Oriented Common Features Operator

Commonality-based operators have been previously employed for feature selection in [9] and [10]. In the former (CF/RSC algorithm), the non-common features are discarded and any features additional to the common are inserted as the result of mutation. In the latter (CHC algorithm), half of the differing bits are crossed at random [11], and therefore this operator also tends to average the number of selected bits. In both CF/RSC and CHC the aim is to identify a single solution. Here, instead of aiming at a single solution, we seek to obtain a range of solutions across the Pareto front. In the simplest case, these are non-dominated solutions in a two-dimensional complexity-performance space. The SSOCF operator utilises the subset size of each mating parent as the desirable target state for each offspring. The functionality of the SSOCF operator is illustrated in Figure 1. Both offpring preserve the common features of their parents. The non-shared features are inherited by the offspring corresponding to the *i*th parent with probability $(n_i - n_c / n_u)$, where n_i is the subset size of the *i*th parent, n_c is the number of commonly selected features across both mating partners and n_u is the number of non-shared selected features. Those non-shared features which are not inherited by the first offspring are inherited by the second. The SSOCF operator has no effect when one parent is a subset of its mating partner. In such cases, any potential modification is the result of consequent mutation. The SSOCF operator lends itself to a simple mutation adaptation strategy, while there is no need to adapt the crossover rate. This strategy is described in detail in [5].

3 FITNESS ASSIGNMENT WITH PROBABILISTIC NEURAL NETS

In classification, performance can be assessed in terms of the misclassification rate. In this paper, feature selection is treated as a multiobjective optimisation problem, in the Pareto sense. The objectives are subset size minimisation and performance maximisation. We consider a dual modelling performance criterion consisting of the estimated misclassification rate and the cost function. The former is common regardless of the classifier and the training algorithm employed, whereas the latter depends on the choice of classifier and algorithm. The major computational cost, associated with the use of EAs for feature selection, is in the feature subset evaluation. Probabilistic neural networks (PNNs) have modest computational requirements for reasonably small data sets [12]. They are based on simple kernel density estimation, equivalent to Parzen windows. PNNs use Bayes rule to estimate the posterior class probabilities, that an input vector **x** corresponds to the class $?_i$. The primary performance measure in our experiments is the estimated misclassification rate, while the secondary is a sum squared error form [5]. Classifiers built without some of the useful features carry an *omission bias*. A second type of bias, which is more difficult to handle, is the *selection bias*. This occurs as a result of the data-dependent nature of the subset selection process. Selection bias becomes more of a problem when the ratio of the number of training patterns to the number of potential predictor variables is small. A simple way of reducing selection bias is by resampling. Here, ten different random splits of the available data set are employed, each into three subsets. The first is employed for training; the second for assessing the impact of different subsets of inputs during the MOEA feature selection. The third data set (the test set) is kept aside for independent evaluation of the final models. Fitness assignment during the MOEA search is performed by taking the average fitness over the different validation sets. A three-element fitness vector is passed to the MOEA. The first two values, the misclassification rate and the feature subset size, are the primary objectives to be minimised. The third value is the cost function and is treated as a secondary cost term, only employed to compare individuals achieving the same misclassification rate. An additional benefit of the resampling is that it reduces the effect of the noise in fitness evaluation.

4 EXPERIMENTAL INVESTIGATION

We compare the performance of the SSOCF operator against that of standard *n*-point crossover on a benchmarking data set of considerable dimensionality, the ionosphere dataset [13]. It consists of 351 patterns, with 34 attributes and one output with two classes. Ten random permutations of this data set are employed. Each one is split in 3 subsets. The training set consists of 176 patterns, the validation set 88 and the evaluation set 87. In addition, the best non-dominated solutions found by forward selection and backwards elimination are compared against those found by MOEA. In terms of computational costs the MOEA is considerably more expensive than sequential methods. The sequential procedures always continue from the subset having the best performance at each step. The following settings are employed:

? Four different runs have been performed with each one of the SSOCF, uniforn, 1 and 2-point crossover.

? N-point crossover settings: crossover rate: 0.85, mutation rate: 1/ l_c . (l_c is the chromosome length)

? triangular sharing function; sharing threshold is l_c / 4.

? tournament group size: 3; sampling group size: 20.

? PNN smoothing factor: 0.2.

? Initial population: uniform distribution across the feature subset size and the different features.

? Parent population size: 200.

? Sequential feature selection: the best subsets found by both forward selection and backwards elimination.

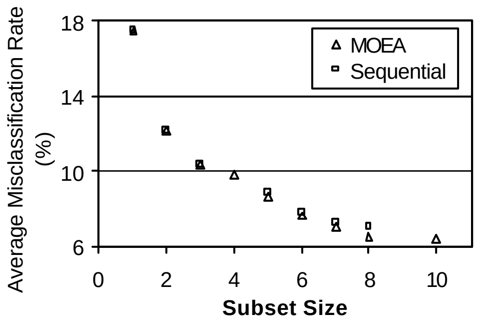

Figure 2: Non-dominated solutions

In both sequential and MOEA feature selection there are cases where an increase in the subset size does not improve performance. Experiments carried out have shown that the MOEA consistently finds a number of solutions missed by the sequential feature selection. In particular, the front identified by MOEA feature selection consists of 9 feature subsets, with 6 of them missed by sequential feature selection (*Figure 2*). The results are illustrated in *Figure 3*, where the average number of non-dominated solutions found, out of a non-dominated front of size 9 is shown for generation 25, 50, 100 and 150.

5 CONCLUSION

An experimental comparison of the commonality-based SSOCF operator against standard *n*-point crossover has been performed. All operators were employed in a MOEA feature selection setting. MOEA feature selection discovers a number of non-dominated solutions missed by both forward selection and backwards elimination. The results obtained provide strong evidence that the SSOCF operator can find a larger set of non-dominated solutions, compared to the *n*-point crossover. Moreover, these solutions are found at a much earlier stage of the MOEA feature selection process. Among *n*-point operators there is no clear winner with the exception of the uniform crossover, which appears to be more disruptive.

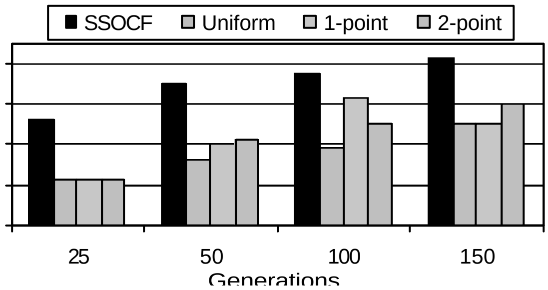

Figure 3: Average number of non-dominated solutions

References

[1]Y. Xin, (1999) "Evolving artificial neural networks," Proceedings of the IEEE, vol. 87, pp. 1423-47.

[2]J. Bala, J. Huang, H. Vafaie, K. DeJong, and H. Wechsler, (1995) "Hybrid learning using genetic algorithms and decision trees for pattern classification" Proceedings of International Joint Conference on Artificial Intelligence. vol.1. 20-25 Aug. 1995, Montreal, Canada. pp. 719-724

[3]A. Jain and D. Zongker, (1997) "Feature selection: evaluation, application, and small sample performance," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, pp. 153-158.

[4]C. Emmanouilidis, A. Hunter, C. MacIntyre, and C. Cox, (1999) "Selecting Features in Neurofuzzy Modelling Using Multiobjective Genetic Algorithms" Proceedings of ICANN'99, 9th International Conference on Artificial Neural Networks, 7-10 Sept. 1999. Edinburgh, UK, pp. 749-754.

[5]C. Emmanouilidis, A. Hunter, and J. MacIntyre, (2000) "A Multiobjective Evolutionary Setting for Feature Selection and a Commonality-Based Crossover Operator" CEC '2000. The 2000 Congress on Evolutionary Compuatation., San Diego, California, USA.

[6]J. Horn, N. Nafpliotis, and D. E. Goldberg, (1994) "A niched Pareto genetic algorithm for multiobjective optimization" Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence. June 1994, Orlando, FL, USA. Vol. 1, pp. 82-87

[7]C. A. C. Coello, (1999) "An updated survey of evolutionary multiobjective optimization techniques: state of the art and future trends" Proceedings of the 1999. Congress on Evolutionary Computation CEC99. 6-9 July 1999, Washington, DC, USA, Vol. 1. pp. 3-13.

[8]S. Chen and S. Smith, (1999) "Introducing a New Advantage of Crossover: Commonality-Based Selection" GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference.

[9]C. Chen, C. Guerra-Salcedo, and S. Smith, (1999) "Non-Standard Crossover for a Standard Representation -- Commonality-Based Feature Subset Selection" GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference.

[10] C. Guerra-Salcedo and D. Whitley, (1998) "Genetic Search for Feature Subset Selection: A Comparison Between CHC and GENESIS," SGA'98. Symposium on Genetic Algorithms, July 22 - 25, 1998, University of Wisconsin, Madison, USA.

[11] L. J. Eshelman and J. D. Schaffer, (1991) "Preventing Premature Convergence in Genetic Algorithms by Preventing Incest," ICGA 1991, Proceedings of the Fourth International Conference on Genetic Algorithms, University of California, San Diego, July 13-16, 1991, pp. 115-122.

[12] D. F. Specht, (1990) "Probabilistic neural networks," Neural Networks, vol. 3, pp. 109-18.

[13] C. Blake, Keogh, E. and Merz, C.J., "UCI repository of machine learning databases. http://www.ics.uci.edu/~mlearn/MLRepository.html. Irvine, CA: University of California, Dept. Information and Computer Science," 1998.