Operator Adaptation in Structure Optimization of Neural Networks

Christian Igel
Institut für Neuroinformatik
Ruhr-Universität Bochum
44780 Bochum, Germany
Christian.Igel@neuroinformatik.ruhr-uni-bochum.de

Martin Kreutz
ZN Vision Technologies AG
44781 Bochum, Germany
Martin.Kreutz@zn-ag.de

The ability of an evolutionary algorithm (EA) to adapt its search strategy during the optimization process is a central concept in evolutionary computation, because (i) the best setting of an EA is not known *a priori* for a given task, (ii) the optimal search strategy is normally not constant during the evolutionary process. Thinking in terms of search (or generating) distributions, population and search strategy as well as selection and strategy adaptation are of equal importance. We propose a derandomized algorithm that adapts operator probabilities on population level. A detailed description, theoretical and more empirical results, and in particular references to related work can be found in [1].

Let Φ be a fitness function to maximize, Ω the set of (asexual) variation operators employed, and $p_o^{(t)}$ the probability at generation t that $o \in \Omega$ is chosen. Further, let $O_o^{(t)}$ contain all offsprings produced in generation t by application of operator o, e.g., if individual $\boldsymbol{g}$ has been generated by consecutive application of the operators o_i and o_j then $\boldsymbol{g}$ is added to $O_{o_i}^{(t)}$ and $O_{o_j}^{(t)}$. The adaptation of operator probabilities is based on a function q that measures the value of a single modification by an operator. One possible choice for this measure is the *local delta* or *credit* given by $q(\boldsymbol{g}) := \max\{\Phi(\boldsymbol{g}) - \Phi(\boldsymbol{g}_{\text{best}}), 0\}$, where $\boldsymbol{g}_{\text{best}}$ is the best individual in the current population. Replacing $\Phi(\boldsymbol{g}_{\text{best}})$ with the fitness of the parent of $\boldsymbol{g}$ yields an alternative measure called *benefit*. The generation dependent quality of $o \in \Omega$ is defined as $q_o^{(t)} := 1/\left|O_o^{(t)}\right| \sum_{\boldsymbol{g} \in O_o^{(t)}} q(\boldsymbol{g})$. The operator probabilities are adjusted every τ generations. Let $q_{\text{all}} := \sum_{i=0}^{\tau-1} \sum_{o' \in \Omega} q_{o'}^{(t-i)}$. Then we set

$$\tilde{p}_o^{(t+1)} := \begin{cases} \zeta/q_{\text{all}} \cdot \sum_{i=0}^{\tau-1} q_o^{(t)} + (1-\zeta) \cdot \tilde{p}_o^{(t)} & \text{if } q_{\text{all}} > 0 \\ \zeta/|\Omega| \qquad + (1-\zeta) \cdot \tilde{p}_o^{(t)} & \text{else} \end{cases}$$

for all $o \in \Omega$. Then

$$p_o^{(t+1)} := p_{\min} + (1 - |\Omega| \cdot p_{\min}) \frac{\tilde{p}_o^{(t+1)}}{\sum_{o' \in \Omega} \tilde{p}_{o'}^{(t+1)}} \ .$$

Herein $\zeta \in]0,1]$ controls a momentum effect and $p_{\min} < 1/|\Omega|$ serves as a lower bound on the operator probabilities. We initialize $\tilde{p}_o^{(0)} = p_o^{(0)}$ (e.g., $p_o^{(0)} = 1/|\Omega|$) for all $o \in \Omega$. This adaptation algorithm itself has free parameters, but (i) their number is reduced compared to the number of parameters that are adapted, (ii) to our experience, the new parameters are very robust, and (iii) sensible guidelines for their choice exist.

The proposed adaptation scheme has proven to be beneficial in structure optimization of neural networks (NNs). In our experiments, the goal is to find an NN that solves a real-world classification task and has as few as possible degrees of freedom (DOF). It turns out that (i) structure optimization with operator adaptation performs statistically significantly better than optimization without operator adaptation, (ii) the *benefit* tends to give better results than the *local delta*, and (iii) the operator probabilities change in an intuitive way during evolution. Early in the search process, operators that add DOF to the NN perform better. But after NNs that solve the classification task evolved, operators that reduce the DOF are preferred: First, complete nodes are pruned, then the NNs are fine-tuned by removing single connections.

References

[1] C. Igel and M. Kreutz. Operator adaptation in evolutionary computation and its application to structure optimization of neural networks. Technical Report IRINI 2001-03, Institut für Neuroinformatik, Ruhr-Universität Bochum, 44780 Bochum, Germany, 2001.