Multi-Optimization Improves Genetic Programming Generalization Ability

Leonardo Vanneschi
Dipartimento di Informatica,
Sistemistica e Comunicazione
University of Milano-Bicocca,
Milan, Italy

vanneschi@disco.unimib.it

Denis Rochat
Computer Systems
Department
University of Lausanne,
Lausanne, Switzerland

Denis.Rochat@unil.ch

Marco Tomassini
Computer Systems
Department
University of Lausanne,
Lausanne, Switzerland

Marco.Tomassini@unil.ch

Categories and Subject Descriptors

I.2.m.c [**Artificial Intelligence**]: Miscellaneous: evolutionary computing and genetic algorithms

General Terms

Algorithms, Performance

Keywords

generalization, genetic programming, symbolic regression

Generalization is one of the most important performance evaluation criteria for artificial learning systems, in particular for supervised learning. While a large amount of literature and of well established results exist concerning the issue of generalization for many non-evolutionary Machine Learning strategies, like for instance Support Vector Machines, this issue in Genetic Programming (GP) has not received the attention it deserves and only recently, few papers dealing with the problem of generalization have appeared (see for instance [1, 2, 3]). In this paper, we have motivated and empirically shown that GP using a Pareto multi-optimization on the training set has a remarkably higher generalization ability than canonic or standard GP (besides counteracting bloat in a more efficient way and maintaining a higher diversity inside the population). Here is an informal motivation for this idea: in figure 1, we have plotted two simple hypothetical fitness functions and two simple hypothetical GP individuals with good fitness on the training set and bad generalization ability, if the sum of errors is considered as the sole evaluation criterium. Even though for points inside the training set the gray and black curves are very close (and thus fitness is good on the training set, if fitness is the sum of errors), outside the training set, they are very far from each other and they get farthest as we consider farthest points from the training set. This happens because the gray and black curves are *uncorrelated* and all the distances between the gray curves points and the black curves ones with the same abscissa inside the training set are different between each other. Thus, three optimization criteria have been used on the training set by our multi-optimization framework: sum of errors, statistical correlation between targets and outputs and variance of the pairwise distances between targets and outputs. Simulations have been executed on three

frequently used symbolic regression GP test problems. Experimental results are encouraging and should pave the way for the use of multi-optimization GP for generalization in real-life applications.
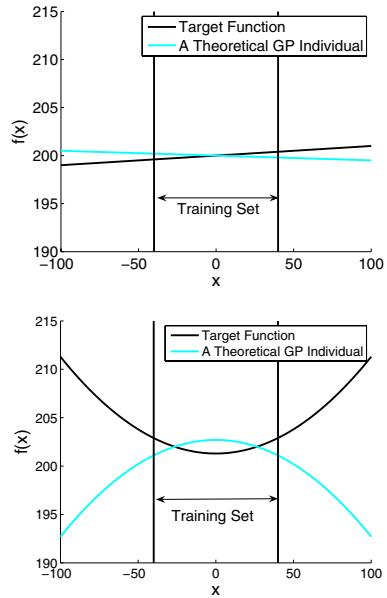

Figure 1: Two examples of hypothetical GP individuals (gray curves) that approximate two simple hypothetical target functions (black curves). If fitness on the training set is only the sum of errors, both of them have a "good" fitness inside the training set and a "bad" fitness outside it.

1. REFERENCES

[1] F. D. Francone, P. Nordin, and W. Banzhaf. Benchmarking the generalization capabilities of a compiling genetic programming system using sparse data sets. In J. R. Koza *et al.*, editor, *Genetic Programming: Proceedings of the first annual conference*, pages 72–80. MIT Press, Cambridge, 1996.

[2] C. Gagne, M. Schoenauer, M. Sebag, and M. Tomassini. Genetic programming for kernel-based learning with co-evolving subsets selection. In T. Runarsson et al., editor, *Parallel Problem Solving from Nature - PPSN IX*, volume 4193 of *Lecture Notes in Computer Science*, pages 1008–1017, Berlin, Heidelberg, New York, 2006. Springer.

[3] I. Kushchu. An evaluation of evolutionary generalization in genetic programming. *Artificial Intelligence Review*, 18(1):3–14, 2002.