Multi-branches Genetic Programming as a Tool for Function Approximation

Katya Rodríguez-Vázquez and Carlos Oliver-Morales

IIMAS-UNAM, Circuito Escolar, CU, 04510 Coyoacón, Mexico City, MEXICO
katya@uxdea4.iimas.unam.mx, oliver_carlos_99@yahoo.com

Abstract. This work presents a performance analysis of a Multi-Branches Genetic Programming (MBGP) approach applied in symbolic regression (e.g. function approximation) problems. Genetic Programming (GP) has been previously applied to this kind of regression. However, one of the main drawbacks of GP is the fact that individuals tend to grow in size through the evolution process without a significant improvement in individual performance. In Multi-Branches Genetic Programming (MBGP), an individual is composed of several branches, each branch can evolve a part of individual solution, and final solution is composed of the integration of these partial solutions.

1 Introduction

In function approximation problems based on GP approaches, two relevant aspects are considered. On the one hand, evolved functions must be accurate approximations; on the other hand, solutions complexity must be also kept as simple as possible without incurring in a deterioration in accuracy. This paper presents an alternative GP encoding based on multiple branches, showing the evolution of accurate and simple solution.

The multi-branches (MB) representation work consists of four parts: a root node, *N* branches, *N+1* coefficients and an output. The number of coefficients is *N+1*, a coefficient for each branch plus the constant term. Expression of section 2 provides details of multi-branches representation. In this example, rooted-node has been defined as the addition operation (ADD). The first branch only consists of $1/n^2$, while the second branch is defined as log(2n). Coefficients for each branch are also expressed. The last coefficient refers to constant term into polynomial expression.

2 Experiments

The harmonic number defined in equation (1) is used in order to approximate the well-known function expansion given in equation (2), an asymptotic expansion where γ is the Euler's constant ($\gamma \approx 0.57722$).

$$H_n \equiv \sum_{i=1}^{n} \frac{1}{i} \qquad (1)$$

K. Deb et al. (Eds.): GECCO 2004, LNCS 3103, pp. 719–721, 2004.
© Springer-Verlag Berlin Heidelberg 2004

$$H_n = \gamma + \ln(n) + \frac{1}{2n} - \frac{1}{12\,n^2} + \frac{1}{120\,n^4} + O(\frac{1}{n^6}) \quad (2)$$

where n = 1, 2, …, 50 (the first 50 harmonic numbers defined as fitness cases). The function set used in this problem is F = { +, -, *, %, -x, plog, psqrt, cos}, where %, plog and psqrt are the protected division, the protected natural logarithm and the protected square root, respectively. Then, the aim of this experiment is to rediscover terms of an already known approximation. In this case, the following approximation emerged,

```
(ADD
   (divd (divd n n) (.* n n))
   (log_p (+ n n))
   (divd (ngt n) (.* n n))
   -0.072444
   2.302
   -0.49414
   -0.1147)
)
```

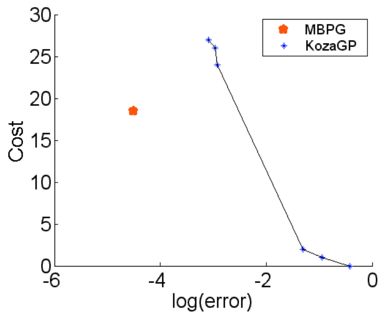

Fig. 1. Harmonic number, [1, 50].

Translating this MBGP model, the function expansion is expressed as,

$$H_n = -0.072444\frac{1}{n^2} + 2.302\ln(2n) - 0.49414\frac{-n}{n^2} - 0.1147$$

ordering this equation,

$$H_n = [\ln(2) - 0.1147] + \ln(n) + \frac{1}{2.0237n} - \frac{1}{13.8n^2} + 1.302\ln(2n)$$

$$= 0.5784 + \ln(n) + \frac{1}{2.0237n} - \frac{1}{13.8n^2} + 1.302\ln(2n)$$

It is observed that MBGP approximated the first four terms given in equation (3), showing an approximation error of $3.18742\text{x}10^{-5}$. Figure 1 shows the Padé approximations and GP Pareto front (Streeter and Becker, 2003) as well as the solution with smallest error generated by means of MBGP.

3 Conclusions

The multi-branches representation for genetic programming was proved to be powerful. It has been tested on a function approximation problem and results showed to be promising. It was also observed that complexity tends to be reduced by using this representation. Further studies will focus on both the flexibility of this representation in diverse domains and the effects and control of introns.

Acknowledgements. Authors would like to thank the financial support of CONACyT, México, under the project 40602-A and PAPPIT-UNAM under the project EN100201.

Reference

STREETER M. AND L.A. BECKER (2003) Automated Discovery of Numerical Approximation Formulae via Genetic Programming. *Genetic Programming and Evolvable Machines*, **4**(3), pp. 255-286.