Multi-Layered Evolutionary System suitable to Symbolic Model Regression

Tomas Brandejsky

Abstract—Presented paper deals with description of multilayered evolutionary system optimized to symbolic regression of difficult systems like e.g. deterministic chaos ones. This paper starts with brief introduction of previous work and ideas. Then the structure of Genetic Programming Algorithm – Evolutionary Strategy hybrid system is described and analyzed including such problems as suitability to parallel implementation, optimal set of building blocks, or initial population generating rules. GPA-ES system combines GPA to model development with ES used for model parameter identification. Such hybrid system eliminates many weaknesses of standard GPA. The paper ends with examples of GPA-ES application.

Keywords—Genetic Algorithm, Genetic Programming Algorithm, Evolutionary Strategy, Symbolic Regression, Model Development.

I. INTRODUCTION

T HIS document is denoted to presentation of multilayered evolutionary system combining Genetic Programming Algorithm (GPA) and Evolutionary Strategy (ES) designed to symbolic regression of model with reduced risk of so(called blowing (production of overcomplicated solutions), see [1]. Presented work continues in effort to increase structural sensitivity of GPAs; it means attempt to produce solutions which are not overcomplicated, which are near to analytical model describing of original data, etc.

Standard GPA (see e.g. [2]-[4]) finds both structure and its parameters (e.g. equation and magnitudes of its constants). Big problem is caused by the fact, that wrong structure with perfectly fitted parameters might give better fitness function value than good structure with poorly estimated parameters. As it was published in [5] and [6], it is possible to find many situations, when imprecise structure with better parameters is evaluated as more suitable than structure with better evolution potential or even perfect structure, but with wrong parameters. For example, the following list of pairs (x, f(x)) (1) is described by (2) and symbolic regression by GPA gives the best solutions (3) and (4).

Non-looking that (3) gives worse fitness function value 0.98, it has proper structure. Equation (4) gives smaller sum of error squares equal to 0.34, but it is linear function and thus it is harder to transform it into good form, see Fig. 1.

Because it is needed to measure structural potential of each solution precisely, it is needed to eliminate influence of random perturbation of parameters. GPA uses single fitness function but evolves both structures and their parameters. Such situation does not allow to separate perfect structures reliably. In this paper, GPA ES hybrid system uses GPA to develop structures, but ES (or Genetic Algorithm) system is applied in each cycle for each individual to optimize parameters, especially constants, in algebraic relations. GPA develops only structural operators, it does not care about constant terminals. The application of specific evolutionary system for parameters decreases error of their determination, but increases computational complexity of the algorithm, as will be discusses latter.

In the previous works, many attempts to increase ability of GPA to solve difficult problems were formed. The work [7] studied possibility to evolve qualitative model by GPA and then to precisiate it by GA into standard differential equations. This way is difficult due to inherited ambiguity in qualitative model and its transformation. There are also problems with measurement of distance of qualitative behaviors, because space of qualitative behaviors is ordered, but without scale.

\[
((0,0),(0.5,0.2125),(1,0.85),(1.5,1.9125))
\]

\[
y = 0.85x^2
\]

\[
y = 1.25x^2
\]

\[
y = 1.25x
\]

Fig. 1 masking of structure by wrongly estimated parameters

(Manuscript received Jul 30, 2011. This work was supported by the research project of MŠMT ČR No 684070043 „Improvement of methods of design and employment of transportation networks from optimization viewpoint”.

T. Brandejsky is with the Czech Technical University, Prague, 110 00 Czech Republic (corresponding author to provide phone: (+420) 224-359-530; fax: (+420) 224-359-534; e-mail: brandejsky@fd.cvut.cz).
Also, precisiation of qualitative model is a special case of hierarchical evolutionary system building hierarchy of models on different abstraction levels by related hierarchy of evolutionary algorithms, as it was published in [8], [9].

II. HYBRID GPA-ES SYSTEM

Structure of hybrid GPA-ES system is outlined at Fig. 2. Fig. 3 then illustrates relationship between individuals in GPA population and vector of ES populations (one population for each GPA individual) in the GPA-ES system. While Fig. 2 outlines control flow of hybrid algorithm, Fig. 3 is denoted to fundamental data structure of this system.

From practical viewpoint, it is useful to parallelize rather outer (GPA) loop than inner ES loop of the presented GPA-ES algorithm. Parallelization of inner loop is efficient only when regressed data set and found regression functions are extremely complex and thus evaluation of ES takes a lot of computational time. This is the reason why only parallelization of the GPA loop will be discussed.

From many parallelization schemas, OpenMP library has been chosen, especially due to its efficiency on single processor machines with multicore processors or common Symmetric MultiProcessor machines (SMP). Implementation then reduces to parallelization of evaluation loop in GPA, which reduces to single directive and including of the OpenMP library header file, because OpenMP implicitly supports C and Fortran languages, but only C/C++ is applicable on GPA implementation. It is also possible to use e.g. Microsoft PPL library which is much simpler than OpenMP because only cycle parallelization is supported. It is possible to use special languages like Chapel or X10 developed within High Performance Computing Initiative, but the efficiency of these languages on small machines is not great and many features of these languages remain unused, see [12]-[14].

When the evaluation loop of GPA is parallelized, each ES population of parameters is optimized in parallel with others. Thus, for each individual of the GPA population, special copy of the ES is created and executed. Because no communication between them is required, execution runs extremely efficiently without necessity of any synchronization.

Computational complexity of GPA-ES is expressed as (5),

\[ O(GPAES) \leq p q n m \log m + p q n \log n \]  

(5)

Where

- \( n \) is number of GPA individuals
- \( m \) is number of ES individuals
l is complexity of structures created by GPA
k is average number of constants in GPA genes, where
\[ k = 2^{l-2} \]
p is number of GPA populations
q is number of ES populations

This complexity estimation of GPA-ES system described by Fig. 2 presumes that computational complexities of building blocks are following:
- GPA population initialization complexity is given by (6)
\[ O(GPA\text{Init}) = n2^l \]
- GPA population evaluation complexity
\[ O(GPA\text{Evaluation}) = 2^l n \]
- Ending complexity
\[ O(GPA\text{End}) = 1 \]
- Evolutionary operators with population ordering
\[ O(GPA\text{Eval}) = n \log n + n \cdot 2^l + kn + n \]
- ES population initialization
\[ O(ES\text{Init}) = m \cdot 2^l \]
- ES population evaluation
\[ O(ES\text{Eval}) = mk \]
- ES Termination condition
\[ O(ES\text{Term}) = 1 \]
- Population ordering and intelligent crossover
\[ O(ES\text{Sort}) = m \log m + mk \]

In case of the hardware threads and parallelization of evaluation cycle of GPA only, the relationship 5 changes to
\[ O(GPA\text{ES}) \approx \frac{pq m \log m + pq \log n}{\eta} \]

IV. OPTIMAL POPULATION SIZES

Efficiency of the algorithm depends on sizes of populations. These dimensions influence computational complexity (5) or (14), as it is discussed in the previous chapter, but it also influences average error of slave ES and consequently also of whole GPA-ES. Unfortunately, while the computational time of single GPA-ES cycle increases with sizes of populations, average error decreases and thus the number of GPA-ES system cycles decreases too. Because number of constants (and it is function of structure complexity) play the significant role in (5) and (14), the number of ES individuals in ES population depends on expected complexity of resulting structure created by master GPA. Some examples of these magnitudes will be discussed in GPA-ES application chapter.

V. SUITABLE SET OF BUILDING BLOCKS

GPA-ES system is influenced also by many other parameters, especially by predefined set of building blocks like operators and functions in case of symbolic regression of data set. It is well known; that the use of simply interchangeable functions as e.g. sin and cos strongly decreases efficiency of GPA algorithm and it is better to use one of them. In contrary, even bigger problem occurs, when suitable function or operator is missing. Initial structures in population influence efficiency of evolutionary process too.

Also terminals play significant role from the efficiency viewpoint. In case of linear systems, generating of terminals in the form (15) is better than to generating of constants and variables randomly connected by any operator or function. But also in case of non-linear systems structure of initial population influences efficiency, as it will be discussed in latter in the following chapter.

\[ const1 \cdot \text{var} + const2 \]

Systems might be described as functions of time or previous states. Problems occur when building block set of GPA allows both descriptions. In such case, there is risk that crossover operation will try to combine individual using one description with opposite one. That situation tends to decrease of GPA efficiency due to poor results of such operation, which will be rejected frequently.

VI. GPA-ES APPLICATION

Symbolic regression of Lorenz attractor from previously computed data set is a good test bed for many techniques applicable to model chaotic system, as it is studied e.g. in [15]. Lorenz attractor is described by set of three equations (16), where parameter magnitudes are chosen as (17) (there are known many combinations of parameters capable to produce chaotic behavior and chosen parameters produces behavior outlined at Fig. 4).

\[ x'[t] = \sigma (y[t] - x[t]), \]
\[ y'[t] = x[t] (\rho - z[t]) - y[t], \]
\[ z'[t] = x[t] y[t] - \beta z[t] \]

\[ \sigma = 16 \]
\[ \beta = 4 \]
\[ \rho = 45.91 \]

Hybrid systems as used GPA-ES work with multiple populations of genes. Presented algorithm works with population of genes representing tree-like description of regressed algebraic relationships. This population was evolved by GPA and it contained 400 genes. In each step of evolution of each gene, the population of 2000 genes representing parameters of its structure is created and optimized. Fitness function magnitude of the best parameter set then serves as magnitude of fitness function of evolved structure in GPA.
During experiment all three forms of terminals were tested – free form, const*var, and const1*var+const2. Because Lorenz attractor equations are nonlinear, free form was expected as winner. Results are summarized in the following Table 1:

<table>
<thead>
<tr>
<th>TERMINAL FORM</th>
<th>NUMBER OF ITERATIONS</th>
<th>QUALITY OF RESULT</th>
</tr>
</thead>
<tbody>
<tr>
<td>FREE FORM</td>
<td>x: 1</td>
<td>x: decomposed</td>
</tr>
<tr>
<td></td>
<td>y: 19</td>
<td>constants</td>
</tr>
<tr>
<td></td>
<td>z: 9</td>
<td>y: perfect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Z: decomposed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>constants</td>
</tr>
<tr>
<td>CONST*VAR</td>
<td>x: 1</td>
<td>x: perfect</td>
</tr>
<tr>
<td></td>
<td>y: 4</td>
<td>y: decomposed</td>
</tr>
<tr>
<td></td>
<td>z: 2</td>
<td>constants</td>
</tr>
<tr>
<td></td>
<td></td>
<td>z: perfect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>x: perfect</td>
</tr>
<tr>
<td>CON1*VAR+CON2</td>
<td>y: 35</td>
<td>y: decomposed</td>
</tr>
<tr>
<td></td>
<td>z: 4</td>
<td>constants and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>variables</td>
</tr>
<tr>
<td></td>
<td></td>
<td>z: decomposed</td>
</tr>
</tbody>
</table>

Inexpectedly, the best form of terminals from both viewpoints of convergence (number of iterations) and solution quality is the form const*var. It is given by the fact that form const1*var+const2 is not present in equations of Lorenz attractor (16), but the const*var form is many times present in them. Free form of terminals is less efficient too. Quality of results mentioned in Table 1 is in range perfect (solution obtained by evolutionary system is identical to original equation), with decomposed constants (e.g. on the place of constant 5 the result contains form 2+3) or with decomposed constants and variables (e.g. on the place of constant 5 there is structure var+2+3-var). No totally different and false structures has been obtained. In all experiments, the limit of preciseness has been reached. Preciseness was expressed as sum of difference squares on 599 samples (6 seconds of prediction). The limit of this sum was 10^3.

VII. CONCLUSION

Presented paper describes the Multi-Layered Evolutionary System suitable to symbolic model regression. Described structure of the evolutionary system decreases risk of wrongly identified parameters influence. Convergence of the system is fast and obtained results have acceptable quality. This fact allows efficient applying of this evolutionary system in such tasks as symbolic regression of deterministic chaos system dataset.

REFERENCES