Geometric Semantic Genetic Programming for Biomedical Applications: A State of the Art Upgrade

Leonardo Vanneschi,
Mauro Castelli
and Ivo Gonçalves
NOVA IMS, Universidade Nova de Lisboa,
1070-312 Lisboa, Portugal
Email:
lvanneschi@novaims.unl.pt
mcastelli@novaims.unl.pt
igoncalves@novaims.unl.pt

Luca Manzoni
DISCo, University of Milano Bicocca,
20126 Milano, Italy
Email:
luca.manzoni@disco.unimib.it

Sara Silva
BioISI – Biosystems &
Integrative Sciences Institute,
Departamento de Informática,
Faculdade de Ciências,
Universidade de Lisboa,
1749-016 Lisboa, Portugal;
CISUC,
Department of Informatics Engineering,
University of Coimbra, Portugal
Email:
sara@fc.ul.pt

***Abstract*—Geometric semantic genetic programming is a hot topic in evolutionary computation and recently it has been used with success on several problems from Biology and Medicine. Given the young age of geometric semantic genetic programming, in the last few years theoretical research, aimed at improving the method, and applicative research proceeded rapidly and in parallel. As a result, the current state of the art is confused and presents some "holes". For instance, some recent improvements of geometric semantic genetic programming have never been applied to some popular biomedical applications. The objective of this paper is to fill this gap. We consider the biomedical applications that have more frequently been used by genetic programming researchers in the last few years and we systematically test, in a consistent way, using the same parameter settings and configurations, all the most popular existing variants of geometric semantic genetic programming on all those applications. Analysing all these results, we obtain a much more homogeneous and clearer picture of the state of the art, that allows us to draw stronger conclusions.**

I. Introduction

The definition of new computational methods aimed at integrating semantic awareness in Genetic Programming (GP) [1] is a hot topic in the field of evolutionary computation (refer to [2] for a recent survey). The focus of this paper is on the study of genetic operators, called Geometric Semantic Operators (GSOs) [3], that directly explore the space of the underlying semantics of the programs. GP that uses GSOs, instead of traditional crossover and mutation, is known to the GP community as Geometric Semantic GP (GSGP) [4]. Several recent works have shown that GSGP is able to produce good-quality results over different domains and, in particular, they are able to outperform standard GP as well as other state-of-the-art machine learning techniques [5], [6], [7], [8], [9]. Among the different applicative domains that have been explored, GSGP has produced competitive results in Biological applications [10], [11] and in the Medical domain [12], [6]. More specifically, great importance was given in the GP community to the study of the following problems:

i. Predition of pharmacokinetic parameters of potential new drugs, in the drug discovery and development process. In particular, a noteworthy relevance was given in the GP community to the prediction of human oral bioavailability [13], plasma protein binding levels (PPB) [14] and LD50, one of the most used measures to quantify the toxicity of a molecular compound [14].
ii. Prediction of the severeness of Parkinson's disease using a unified rating scale assessment [6].
iii. Prediction of relative positions of CT Slices [12], [15].
iv. Prediction of the 3D structure of proteins [10]

With the advances of the study of semantics-based method in GP, the original GSOs have been developed to improve their performance. Under this perspective, particular interest has been raised in the last few years by the possibility of integrating, inside geometric semantic mutation, a local search optimizer (the idea was first presented in [16]). This basically gave raise to three, quite popular and competitive GP variants:

a. GSGP [4], where standard GSOs have been applied, as defined in [3];
b. GSGP-LS, where geometric semantic mutation is improved by the integration of a fast an powerful local search method, as discussed in [16];
c. GSGP-HYBRID, which uses the local search optimizer to improve mutation only in the first k generations (where k is a parameter of the algorithm) and then continues with standard GSGP. In other words, GSGP-HYBRID (also presented in [16]) works like GSGP-LS in the first k generations, and like GSGP in the remaining part of the

978-1-5090-4601-0/17/$31.00

evolution.

Problems i.–iv. mentioned above have been tackled in the literature using the computational methods a., b. and c., as specified in Table I. The objective of this paper is extending

	GSGP	GSGP-LS	GSGP-HYBRID
Bioavailability	[17], [16]	[16]	[16]
LD50	[17], [16]	[16]	[16]
PPB	[17], [16]	[16]	[16]
Parkinson	[16], [6]	[16]	[16]
CT Slices	[12]	*novel*	*novel*
3D Protein Structure	[10]	*novel*	*novel*

Table I
PREVIOUS STUDIES APPLYING THE COMPUTATIONAL METHODS REPORTED IN THE UPPER LINE TO THE PROBLEMS REPORTED IN THE LEFTMOST COLUMN. THE TERM *novel* MEANS THAT THE METHOD IS APPLIED TO THE PROBLEM FOR THE FIRST TIME IN THIS PAPER.

the studies mentioned in the table, giving for the first time a global view of the state of the art of semantics-based methods for biological and medical applications. Also, we have the goal of upgrading the state of the art, by completing the picture with those experiments (tests of some computational methods for some problems) that have never been performed before (indicated with the keyword *novel* in the table). To make the study as comprehensive as possible, 6 different datasets characterized by different numbers of features and instances (observations) have been taken into account.

The paper is organized as follows: Section II presents the definition and the main properties of the GSOs presented in [3] and the variant proposed in [16]. Section III describes the datasets that have been considered in the study, highlighting their main characteristics. Section IV discusses the experimental settings and the obtained results. Finally, Section V concludes the paper and summarizes the main contributions of this work.

II. GEOMETRIC SEMANTIC GENETIC PROGRAMMING

In Section II-A, we present the definition of GSOs for symbolic regression problems (the problem domain that is studied in this paper), as first introduced in [3] and in Section II-B, we summarize the work of [16] about the inclusion of a local search technique within the geometric semantic mutation operator.

A. Geometric Semantic Genetic Operators

Let $\mathbf{X} = \{\overrightarrow{x_1}, \overrightarrow{x_2}, ..., \overrightarrow{x_n}\}$ be the set of input data (training instances, observations or fitness cases) of a symbolic regression problem, and $\overrightarrow{t} = [t_1, t_2, ..., t_n]$ the vector of the respective expected output or target values (in other words, for each $i = 1, 2, ..., n$, t_i is the expected output corresponding to input $\overrightarrow{x_i}$). A GP individual (or program) P can be seen as a function that, for each input vector $\overrightarrow{x_i}$ returns the scalar value $P(\overrightarrow{x_i})$. Following [3], we call *semantics* of P the vector $\overrightarrow{s_P} = [P(\overrightarrow{x_1}), P(\overrightarrow{x_2}), ..., P(\overrightarrow{x_n})]$. This vector can be represented as a point in a n-dimensional space, that we call *semantic space*. Remark that the target vector $\overrightarrow{t}$ itself is a point in the semantic space and, in general, it does *not* correspond to the origin of the Cartesian system (except for the very particular and rare case in which the expected output is equal to zero for each observation in the training set).

The objective of GSOs is to define modifications on the syntax of GP individuals that have a precise effect on their semantics. More in particular, as schematically shown in Figure 1, GSOs are:

- Geometric semantic crossover. This operator generates only one offspring, whose semantics stands in the line joining the semantics of the two parents in the semantic space.
- Geometric semantic mutation. With this operator, by mutating an individual i, we obtain another individual j such that the semantics of j stands inside a ball of a given predetermined radius, centered in the semantics of i.

One of the reasons why GSOs are becoming so popular in the GP community is related to the fact that GSOs induce an unimodal error surface (on training data) for any supervised learning problem, where fitness is calculated using an error measure between outputs and targets. In other words, using GSOs the error surface on training data is guaranteed to not have any locally suboptimal solution, for instance, for any regression or classification problem, independently on how big and how complex data are (reference [4] contains a detailed explanation of the reason why the error surface is unimodal when GSOs are used).

The definitions of the GSOs are, as given in [3], respectively:

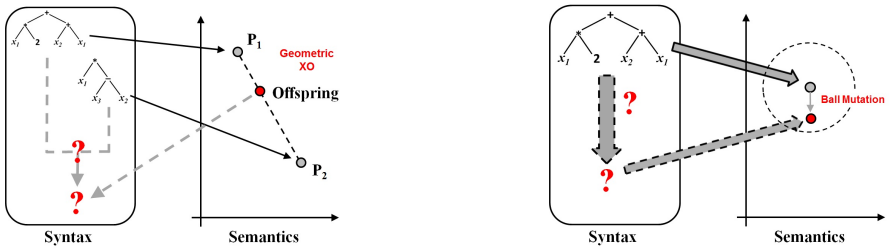
Geometric Semantic Crossover (GSC). Given two parent functions $T_1, T_2 : \mathbb{R}^n \to \mathbb{R}$, the geometric semantic crossover returns the real function $T_{XO} = (T_1 \cdot T_R) + ((1 - T_R) \cdot T_2)$, where T_R is a random real function whose output values range in the interval $[0, 1]$.

Geometric Semantic Mutation (GSM). Given a parent function $T : \mathbb{R}^n \to \mathbb{R}$, the geometric semantic mutation with mutation step ms returns the real function $T_M = T + ms \cdot (T_{R1} - T_{R2})$, where T_{R1} and T_{R2} are random real functions.

Figure 1 shows a graphical representation of the mapping between the syntactic and semantic space given by geometric semantic operators. Using these operators, the semantics of the offspring is completely defined by the semantics of the parents: the semantics of an offspring produced by GSC will lie on the segment between the semantics of both parents (geometric crossover), while GSM defines a mutation such that the semantics of the offspring lies within the ball of radius ms that surrounds the semantics of the parent (geometric mutation).

B. GSOs and Local Search

While GSGP has produced competitive results [18], [19], its original formulation has several drawbacks. In particular, the GSC operator is not useful when the semantics of the parents do not surround the target semantics. This issue has been investigated in [20], where a method for overcoming the aforementioned problem has been proposed. Basically, authors



(a) (b)

Figure 1. Geometric semantic crossover (plot (a)) (respectively geometric semantic mutation (plot (b))) performs a transformation on the syntax of the individual that corresponds to geometric crossover (respectively geometric mutation) on the semantic space. In this figure, the irrealistic case of a bidimensional semantic space is considered, for simplicity.

suggested an initialization method that guarantees that the semantics of the individuals in the initial population forms a convex hull that contains the target semantics. On the other hand, the GSM operator can sometimes produce offspring with a worse fitness than the parent, which is unnecessary since the target semantics are known. Moreover, the offspring produced by GSOs will always be larger than the parents; this means that program growth cannot be eliminated. In this work we follow the proposal outlined in [16], where a local search approach is integrated into GSM, an operator we will refer to as GSM-LS. This operator exploits the fact that the geometric mutation defines a linear combination of the parent K program with random programs, expressed as:

$$K_M = \alpha_0 + \alpha_1 \cdot K + \alpha_2 \cdot (K_{R1} - K_{R2}) \qquad (1)$$

where $\alpha_i \in \mathbb{R}$; notice that α_2 replaces the mutation step parameter ms used in the definition of GSM. This in fact defines a basic multivariate linear regression problem, which can be solved, for example, by Ordinary Least Square regression (OLS). In this case we have n linear equations given by the number of fitness cases, and only three unknowns (the α_s). This gives an overdetermined multivariate linear fitting problem, which can be solved through singular value decomposition (SVD). Hence, the GSM-LS operator produces the best linear fit based on the target semantics of the program. The use of this operator has produced good performance over different domains [21], [12], sometimes outperforming GSGP and several other regression techniques.

III. Data

This section presents the datasets that have been studied in this work. The main characteristics of the datasets, covering the biological and the medical domains, are summarized in Table II, by reporting the number of features and the number of instances for all of them. As it is possible to see, the considered problems, besides being of interest for the biological and medical community, present a non-negligible difference between each other with respect to the number of variables as well as for the number of instances. This gives us the possibility of testing the functioning of the studied computational methods on problems that have rather diverse characteristics.

The datasets are the following:

a) Bioavaialbility:: human oral bioavailability (indicated with %F from now on) is the parameter that measures the percentage of initial drug dose that effectively reaches the systemic blood circulation. This parameter is particularly relevant for pharmaceutical industries, because the oral assumption is usually the preferred way for supplying drugs to patients and because it is a representative measure of the quantity of active principle that effectively can actuate its biological effect. The dataset consists of 241 variables and 360 instances. For additional information the reader is referred to [13].

b) Toxicity (LD50):: LD50 is one of the most used parameters to measure the toxicity of a drug. More precisely, LD50 refers to the amount of compound required to kill 50% of the cavies. It is usually expressed as the number of milligrams of drug related to one kilogram of mass of cavies (mg/kg). Depending on the specific organism (rat, mice, dog, monkey and rabbit usually) and on the precise way of supplying (intravenous, subcutaneous, intraperitoneal, oral generally) chosen, it is possible to define a wide spectrum of LD50 experimental protocols. We consider the LD50 measured using rats as model organisms and supplying the compound orally. The dataset consists of 626 variables and 234 instances. For additional information the reader is referred to [14].

c) Protein Plasm Binding Level:: this value (PPB from now on) corresponds to the percentage of the initial drug dose which binds plasma proteins. This measure is fundamental, both because blood circulation is the major vehicle of drug distribution into human body and because only free (unbound) drugs permeate the cellular membranes and reach the targets. The dataset consists of 626 variables and 234 instances. For additional information the reader is referred to [14].

d) Parkinson:: this dataset contains the data related to 52 subjects with idiopathic Parkinson disease (PD). A subject was diagnosed with PD if he had at least two of the following: rest tremor, bradykinesia (slow movement) or rigidity, without evidence of other forms of parkinsonism. The study was supervised by six US medical centers: Georgia Institute of Technology (7 subjects), National Institutes of Health (10 subjects), Oregon Health and Science University (14 subjects), Rush University Medical Center (11 subjects), Southern Illinois University (6 subjects) and University of California Los

Angeles (4 subjects). The selected subjects had at least 20 valid study sessions during the trial period. For this dataset the target is related to the Unified Parkinsons Disease Rating Scale (UPDRS), the most used scale for tracking Parkinsons disease symptom progression. This scale reflects the presence and severity of symptoms, expressing them in a range from 0 to 176, with 0 representing a healthy state and 176 total disability. The UPDRS contains three sections: (1) Mentation, Behavior and Mood, (2) Activities of daily living and (3) Motor. The motor section of the UPDRS encompasses tasks such as speech, facial expression, tremor and rigidity and expresses the severity of the related symptoms in a range from 0 to 108, where 0 represents a symptom free state and 108 denotes severe motor impairment [6]. The dataset contains 19 variables and 6000 instances.

e) CT Slices:: one of the most common techniques in radiology is the computerized tomography (CT) scan. Automatically determining the relative position of a single CT slice within the human body can be very useful. It can allow for an efficient retrieval of slices from the same body region taken in other volume scans and provide useful information to the non-expert user [12]. Each CT image is described by 385 features. The first feature is the ID of the patient; features 2-241 are related to the histogram describing bone structures; features 242-385 are related to the histogram describing air inclusions. The last feature is the target variable that is the relative location of the image on the axial axis. Values are in the range $[0; 180]$ where 0 denotes the top of the head and 180 the soles of the feet. The dataset contains 4000 instances, a subset of the original dataset described in [15].

f) 3D Protein Structure:: as reported in [10], this dataset is related to the physiochemical properties of proteins. All proteins are polymers composed of the same building blocks, the amino acids, which are covalently joined together by amide links, known as peptide bonds. They differ only in the number, the nature, and the sequential order of their constituent amino acids. To understand the functional diversity of proteins, it is important to appreciate the physicochemical properties of the different amino acids, even though the properties of a protein molecule are hugely more complex than the sum of the properties of its different constituent amino acids. It is then possible to determine the three-dimensional structures that these linked building blocks can acquire and to analyze the biological properties of the corresponding polymers. In this dataset the target variable relates to the size of the residues considering the protein tertiary structure data. The dataset is available at the UCI machine learning repository (http://archive.ics.uci.edu/ml/) and it consists of 9 variables and 45730 instances.

DATASET	VARIABLES	INSTANCES
BIOAVAILABILITY	241	360
TOXICITY	626	234
PROTEIN PLASMA BINDING LEVEL	626	131
PARKINSON	19	6000
CT SLICES	386	4000
3D PROTEIN STRUCTURE	9	45730

Table II
CONSIDERED DATASETS WITH NUMBER OF INDEPENDENT VARIABLES AND INSTANCES.

IV. EXPERIMENTAL STUDY

This section describes the experimental settings and the results achieved by the considered semantics-based GP systems. In particular, Section IV-A presents the experimental settings in order to insure complete replicability of the presented experiments, while Section IV-B discusses the performance of the GP systems by considering training and test fitness.

A. Experimental Settings

The experimental study considers the following semantics-based systems: GSGP (where the GSOs introduced in [3] are used in the evolutionary process), GSGP-HYBRID (where the GSM-LS operator is used in the first k generations, where k is a parameter of the algorithm) and GSGP-LS (where the GSM-LS operator has been used during the whole GP evolution). It is worth pointing out that the results obtained by standard GP [1] are not reported in this paper, because standard GP is consistently outperformed by the three studied methods on all the considered problems. The motivation for considering GSGP-LS and GSGP-HYBRID is related to the fact that the GSM-LS operator, being a very powerful and fast optimizer of training data, can lead to overfitting. Hence, we want to analyze how the use of local search with the GSM operator affects the performance of the best solution in terms of generalization (i.e., performance on unseen instances), both using the local search for the whole evolution and interrupting it earlier in the run. A preliminary experimental study has shown that $k = 10$ is a reasonable number of generations for using GSM-LS in GSGP-HYBRID for the applications considered in this paper. So, 10 is the value used for parameter k in all the experiments involving GSGP-HYBRID. For all the considered test problems 30 independent runs of each studied system have been executed. In each problem, the data was split into a training and a test set, where the former contains 70% of the data samples selected randomly with uniform distribution, while the latter contains the remaining 30% of the observations. The datasets were randomly partitioned before each run, in such a way that each one of the 30 independent executions will have a different training/test partition. For each generation of each studies GP variant, the best individual on the training set has been considered, and its fitness on the training and test set stored. For simplicity, in the continuation, we will refer to the former as training fitness and to the latter as test fitness. While training fitness will also be reported for completeness, the most interesting results are clearly the ones concerning test fitness, which give us an idea of the prediction accuracy, generalization ability and robustness of the models evolved by the studied computational methods.

Table III summarizes the parameters that have been used in our experiments, for all the studied algorithms. In each one of the presented experiments, fitness was the mean absolute

Parameter	Value
Number of Generations	1000
Population Size	100
Initialization	Ramped Half and Half
Max. Initial Depth	6
Crossover Rate	0.6
Mutation Rate	0.4
Function Set	+, -, *, //
Terminal Set	Input Variables and random constants
Selection	Tournament of size 4
Elitism	Best individual survives
Max. Tree Depth	None
Mutation Step	Random in [0;1]
k (GSGP-HYBRID)	10

Table III
EXPERIMENTAL SETTINGS. k IS THE NUMBER OF GENERATIONS IN WHICH LOCAL SEARCH IS APPLIED IN GSGP-HYBRID, BEFORE THE ALGORITHM BECOMES IDENTICAL TO GSGP.

error (MAE) between calculated values (outputs) and known targets. The definition of MAE is:

$$MAE(T) = \frac{1}{N} \sum_{i \in Q} |t_i - y_i| \quad (2)$$

where T is a GP individual, $y_i = T(\mathbf{x}_i)$ is the output of T on input data (observation) $\mathbf{x}_i$ and t_i is the target value corresponding to observation $\mathbf{x}_i$. N denotes the number of samples in the training (or test) set, and Q contains the indices of that set.

B. Results

We first of all discuss the performance achieved on training data. As reported in Figure 2, the results obtained on training data are similar across all the studied problems. More in detail, GSGP-LS is the best performer, followed by GSGP-HYBRID and GSGP. These results are in line with our expectation and can be explained by considering the features of the three systems. In particular, GSGP-LS exploits the properties of GSM, coupling the operator with a local searcher. This allows GSM-LS to produce the best linear fit based on the target semantics of the programs and, as a side effect, to speed up the convergence of the search process towards the optimal individual. Hence, the excellent performance of GSGP-LS on training data was expected and also the differences with the other systems can be easily understood: GSGP-HYBRID uses the GSM-LS operator only in the initial generations of the search process, while GSGP only uses the standard GSM operator, with no local search optimizer. While the results achieved on the training data corroborate the hypothesis about the beneficial effect of integrating a local search operator within GSM, it is more important to understand what are the effects of using this operator over unseen instances. In fact, even if GSM-LS speeds up optimization on training data, it is fundamental to evaluate the generalization ability of the obtained models.

Results on test data are reported in Figure 3. As it is possible to note, test fitness presents a behaviour that is very different over the benchmarks taken into account. In particular, there are problems where the good performance achieved by GSGP-LS on the training does not correspond to a good performance on unseen data. Examples of this situation are the Bioavailabilty problem, the 3D Protein Structure problem and the Toxicity problem. In all of these problems, GSGP-LS is clearly affected by overfitting and it is the method with the worst performance on the test set among the studied ons. It is interesting to see that, in two of these problems (Bioavailability and 3D Protein Structure), GSGP-HYBRID is able to preserve the good performance achieved on the training data, still outperforming GSGP. These results allow us to draw some conclusions about the use of the GSM-LS operator: considering the results obtained on both training and test instances, it seems clear that using a local search operator within GSM is beneficial, providing that the local search is applied only in the initial stage of the search process. This results in good performance on both training and test data for all the studied problems. Continuing the analysis of the results achieved on the test data, it is interesting to see that, on the remaining benchmark problems (i.e. Parkinson, PPB and CT slices), GSGP-LS is the best performer, showing not only good performance on the training set, but also presenting a good generalization ability. However, it is appropriate to remark that GSGP-HYBRID is able to produce robust models, that are able to handle unseen data, also for these three further problems.

Our final interpretation of the presented results is that the GSM-LS operator should be used carefully: there is a certain stage in the evolutionary process in which the use of GSM-LS degrades the performance on unseen data, causing overfitting. Hence, in order to prevent overfitting, it is fundamental to interrupt the usage of this operator at a given point in the evolution. A deeper study of the number of generations in which GSM-LS can be used without causing overfitting is one of the subjects of our current research.

C. Statistical Validation

To analyze the statistical significance of the results presented so far, a set of tests has been performed on the median errors. The Wilcoxon rank-sum test for pairwise data comparison has been used under the alternative hypothesis that the values from the first sample are smaller or equal than the values of the second sample with probability grater than 0.5. A value of $\alpha = 0.05$ has been used and, considering the presence of more than two samples, a Bonferroni correction for this value has been applied. The p-values obtained are reported in Table IV. In this table, a value smaller than the corrected value of α at the intersection of row i and column j means that technique i produces better results than technique j in a statistically significant way. As it is possible to see, the three semantics-based systems produce results that are statistically different over all the benchmark problems taken into account.

V. CONCLUSIONS AND FUTURE WORK

Geometric semantic genetic programming (GSGP), i.e. genetic programming (GP) that uses geometric semantic genetic operators, has been widely investigated in the last few years.

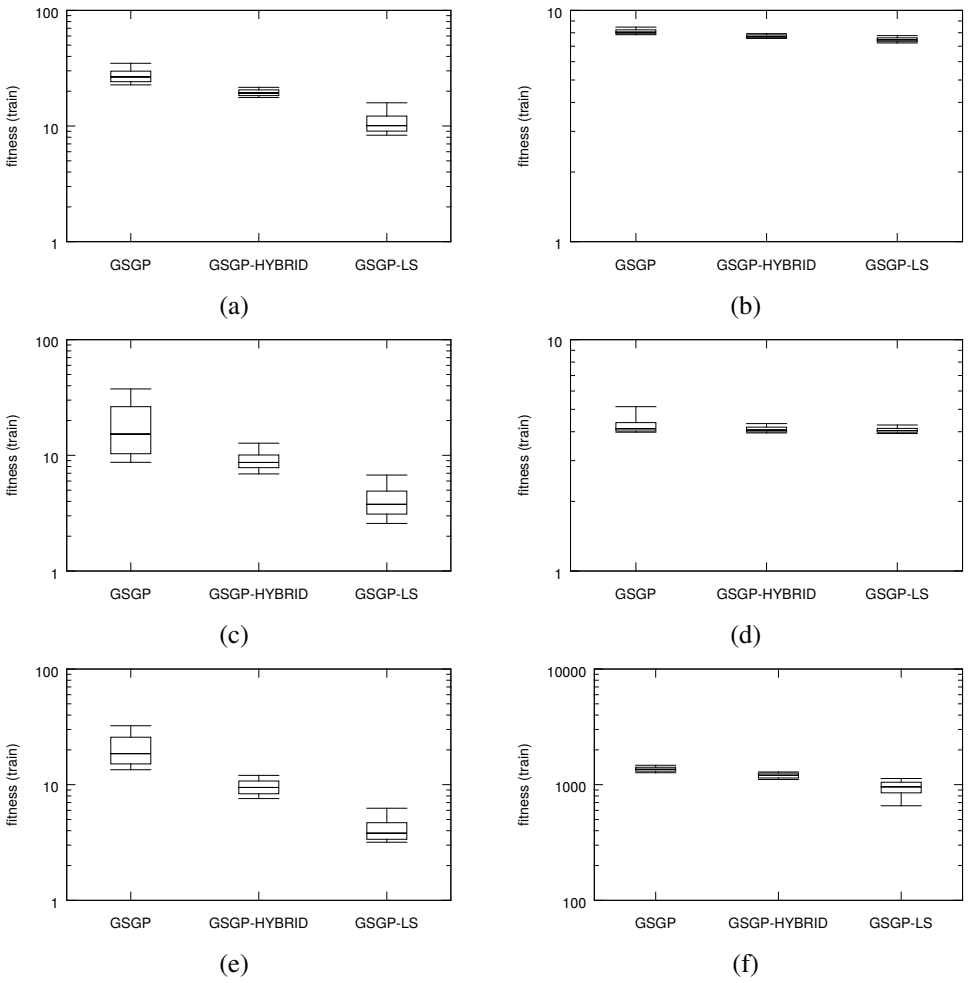


Figure 2. Boxplots of mean absolute error for training instances at the end of the evolution for the following datasets: (a) BIOAVAILABILITY (b) PARKINSON (c) PROTEIN PLASMA BINDING LEVEL (d) 3D PROTEIN STRUCTURE (e) CT SLICES (f) TOXICITY. On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, and the whiskers extend to the most extreme data points not considered outliers.

Those studies have lead to several improvements of GSGP, among which particularly promising are GSGP-LS, that combines GSGP with local search, and GSGP-HYBRID, which integrates GSGP and GSGP-LS.

While these three variants (GSGP, GSGP-LS and GSGP-HYBRID) were developed, several biological and medical applications were used as case studies to validate them. In particular, six applications received particular attention from GP researchers: the prediction of three different pharmacokinetic parameters (human oral bioavailability, plasma protein binding level and toxicity), the prediction of the severity of Parkinson's disease, the prediction of relative positions of CT slices and the prediction of the 3D structure of proteins.

By revising the state of the art, "crossing" these three computational methods with these six applications, we realized that some further work was in demand, which motivated us to write this paper. In particular, the state of the art picture was confused, since different experimental settings were used in different contributions. Furthermore, information was lacking since two of these methods (GSGP-LS and GSGP-HYBRID) had never been used on two of these applications (prediction of relative positions of CT slices and prediction of the 3D structure of proteins).

In this paper, we have reconsidered these three algorithms and these six problems and we have repeated all the "crossed" experiments, also performing for the first time the ones that were missing. Contrarily to the current state of the art, the experiments conducted here are now consistent in the sense that the same experimental settings are used for all the studied computational methods and test problems. As a results, we now have a clearer picture concerning the use of semantics-based GP methods for this kind of applications. In particular, we are able to draw the following general conclusion: combining a local search strategy with geometric semantic mutation is always beneficial on training data. On the other hand, on test data we can clearly recognize two phases in the evolutionary process: the first part of the evolution in which error on test data decreases, and a later phase in which the use of local search causes overfitting. Our experiments clearly show that if we were able to find the point on which GSGP-LS

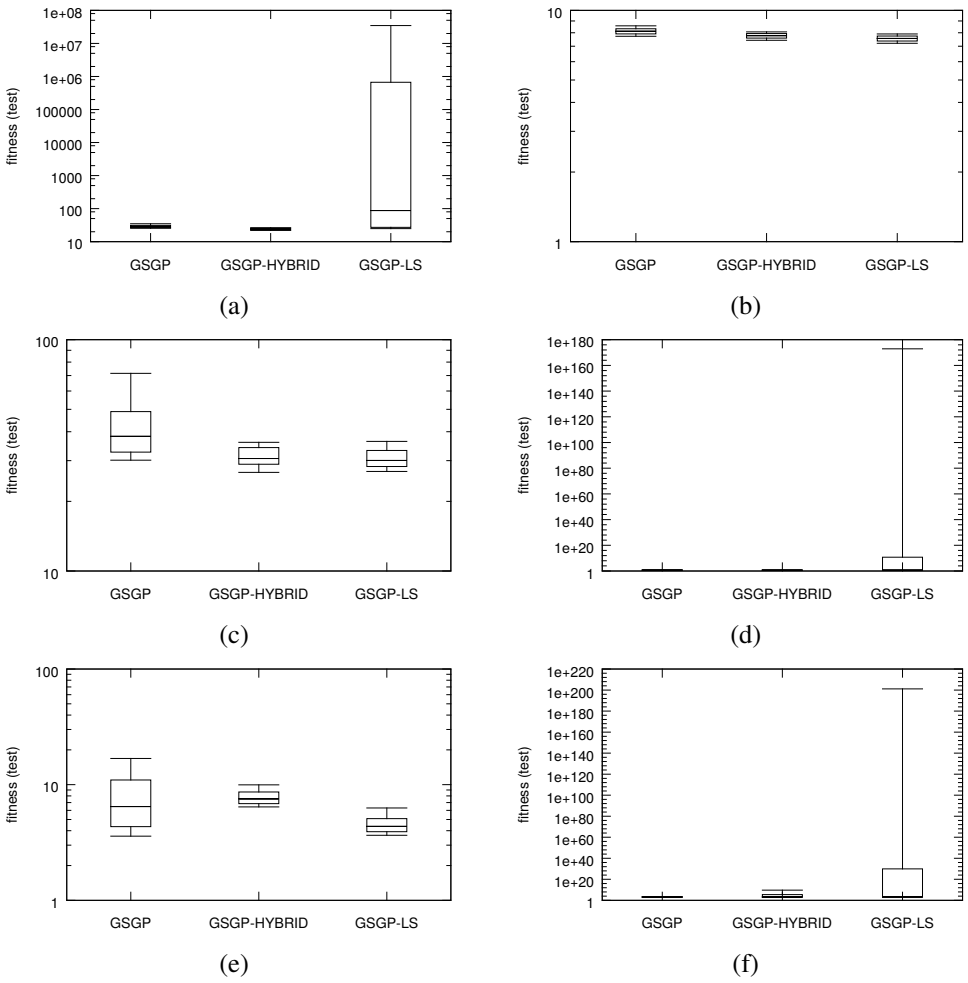


Figure 3. Boxplots of mean absolute error for test instances at the end of the evolution for the following datasets: (a) BIOAVAILABILITY (b) PARKINSON (c) PROTEIN PLASMA BINDING LEVEL (d) 3D PROTEIN STRUCTURE (e) CT SLICES (f) TOXICITY. On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, and the whiskers extend to the most extreme data points not considered outliers.

starts to overfit, and we interrupted the execution of the local search optimizer in that point (thus dynamically turning the algorithm into GSGP), we would be able to outperform the state of the art methods on all the studied applications. This conclusion paves the way to future research, aimed at finding the appropriate point in the evolution in which local search must be interrupted.

ACKNOWLEDGMENT

This work was partially funded by project PERSEIDS (PTDC/EMS-SIS/0642/2014) and BioISI RD unit, UID/MULTI/04046/2013, funded by FCT/MCTES/PIDDAC, Portugal and by CONACYT (Mexico) project FC-2015-2/944 "Aprendizaje evolutivo a gran escala".

REFERENCES

[1] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA, USA: MIT Press, 1992.

[2] L. Vanneschi, M. Castelli, and S. Silva, "A survey of semantic methods in genetic programming," *Genetic Programming and Evolvable Machines*, vol. 15, no. 2, pp. 195–214, 2014.

[3] A. Moraglio, K. Krawiec, and C. Johnson, "Geometric semantic genetic programming," in *Parallel Problem Solving from Nature - PPSN XII*, ser. Lecture Notes in Computer Science, C. Coello, V. Cutello, K. Deb, S. Forrest, G. Nicosia, and M. Pavone, Eds. Springer Berlin Heidelberg, 2012, vol. 7491, pp. 21–31.

[4] L. Vanneschi, *An Introduction to Geometric Semantic Genetic Programming*. Springer, 2017, pp. 3–42.

[5] L. Vanneschi, S. Silva, M. Castelli, and L. Manzoni, "Geometric semantic genetic programming for real life applications," in *Genetic Programming Theory and Practice XI*. Springer New York, 2014, pp. 191–209.

[6] M. Castelli, L. Vanneschi, and S. Silva, "Prediction of the unified parkinson's disease rating scale assessment using a genetic programming system with geometric semantic genetic operators," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4608 – 4616, 2014.

[7] M. Castelli, L. Vanneschi, and M. D. Felice, "Forecasting short-term electricity consumption using a semantics-based genetic programming framework: The south italy case," *Energy Economics*, vol. 47, pp. 37 – 41, 2015.

[8] M. Castelli, D. Castaldi, I. Giordani, S. Silva, L. Vanneschi, F. Archetti, and D. Maccagnola, "An efficient implementation of geometric semantic genetic programming for anticoagulation level prediction in pharmacogenetics," in *Progress in Artificial Intelligence*. Springer Berlin Heidelberg, 2013, pp. 78–89.

[9] M. Castelli, L. Vanneschi, and S. Silva, "Prediction of high performance concrete strength using genetic programming with geometric semantic

Table IV
$P-$VALUES RETURNED BY THE WILCOXON RANK-SUM TEST FOR PAIRWISE DATA COMPARISON.

BIOAVAILABILITY						
	TRAINING			TEST		
	GSGP	GSGP-HYBRID	GSGP-LS	GSGP	GSGP-HYBRID	GSGP-LS
GSGP	-	1	1	-	1	0
GSGP-HYBRID	0	-	1	0	-	0
GSGP-LS	0	0	-	1	1	-

PARKINSON						
	TRAINING			TEST		
	GSGP	GSGP-HYBRID	GSGP-LS	GSGP	GSGP-HYBRID	GSGP-LS
GSGP	-	1	1	-	1	1
GSGP-HYBRID	0	-	1	0	-	1
GSGP-LS	0	0	-	0	0	-

PROTEIN PLASMA BINDING LEVEL						
	TRAINING			TEST		
	GSGP	GSGP-HYBRID	GSGP-LS	GSGP	GSGP-HYBRID	GSGP-LS
GSGP	-	1	1	-	1	1
GSGP-HYBRID	0	-	1	0	-	0.989
GSGP-LS	0	0	-	0	0.011	-

3D PROTEIN STRUCTURE						
	TRAINING			TEST		
	GSGP	GSGP-HYBRID	GSGP-LS	GSGP	GSGP-HYBRID	GSGP-LS
GSGP	-	1	1	-	1	0
GSGP-HYBRID	0	-	1	0	-	0
GSGP-LS	0	0	-	1	1	-

CT SLICES						
	TRAINING			TEST		
	GSGP	GSGP-HYBRID	GSGP-LS	GSGP	GSGP-HYBRID	GSGP-LS
GSGP	-	1	1	-	0	1
GSGP-HYBRID	0	-	1	1	-	1
GSGP-LS	0	0	-	0	0	-

TOXICITY						
	TRAINING			TEST		
	GSGP	GSGP-HYBRID	GSGP-LS	GSGP	GSGP-HYBRID	GSGP-LS
GSGP	-	1	1	-	0	0
GSGP-HYBRID	0	-	1	1	-	0
GSGP-LS	0	0	-	1	1	-

genetic operators," *Expert Systems with Applications*, vol. 40, no. 17, pp. 6856 – 6862, 2013.

[10] M. Castelli, L. Vanneschi, L. Manzoni, and A. Popovi, "Semantic genetic programming for fast and accurate data knowledge discovery," *Swarm and Evolutionary Computation*, vol. 26, pp. 1 – 7, 2016.

[11] L. Vanneschi, S. Silva, M. Castelli, and L. Manzoni, *Geometric Semantic Genetic Programming for Real Life Applications*. New York, NY: Springer New York, 2014, pp. 191–209.

[12] M. Castelli, L. Trujillo, L. Vanneschi, and A. Popovi, "Prediction of relative position of CT slices using a computational intelligence system," *Applied Soft Computing*, vol. 46, pp. 537 – 542, 2016.

[13] F. Archetti, S. Lanzeni, E. Messina, and L. Vanneschi, "Genetic programming for human oral bioavailability of drugs," in *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation*, ser. GECCO '06. New York, NY, USA: ACM, 2006, pp. 255–262.

[14] ——, *Genetic Programming and Other Machine Learning Approaches to Predict Median Oral Lethal Dose (LD50) and Plasma Protein Binding Levels (%PPB) of Drugs*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 11–23.

[15] F. Graf, H.-P. Kriegel, S. Pølsterl, M. Schubert, and A. Cavallaro, "Position prediction in ct volume scans," in *Proceedings of the 28th International Conference on Machine Learning (ICML) Workshop on Learning for Global Challenges, Bellevue, Washington, WA*, 2011.

[16] M. Castelli, L. Trujillo, L. Vanneschi, S. Silva, E. Z-Flores, and P. Legrand, "Geometric semantic genetic programming with local search," in *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*, ser. GECCO '15, 2015, pp. 999–1006.

[17] L. Vanneschi, S. Silva, M. Castelli, and L. Manzoni, *Geometric Semantic Genetic Programming for Real Life Applications*. New York, NY: Springer New York, 2014, pp. 191–209. [Online]. Available: http://dx.doi.org/10.1007/978-1-4939-0375-7_11

[18] M. Castelli, L. Trujillo, L. Vanneschi, and A. Popovi, "Prediction of energy performance of residential buildings: A genetic programming approach," *Energy and Buildings*, vol. 102, pp. 67 – 74, 2015.

[19] L. Vanneschi, M. Castelli, E. Costa, A. Re, H. Vaz, V. Lobo, and P. Urbano, *Improving Maritime Awareness with Semantic Genetic Programming and Linear Scaling: Prediction of Vessels Position Based on AIS Data*, pp. 732–744.

[20] T. P. Pawlak and K. Krawiec, *Semantic Geometric Initialization*. Cham: Springer International Publishing, 2016, pp. 261–277.

[21] M. Castelli, L. Trujillo, and L. Vanneschi, "Energy consumption forecasting using semantic-based genetic programming with local search optimizer," *Computational intelligence and neuroscience*, vol. 2015, p. 57, 2015.