Arbitrarily Close Alignments in the Error Space: a Geometric Semantic Genetic Programming Approach

Ivo Gonçalves
CISUC, Department of Informatics Engineering,
University of Coimbra,
3030-290 Coimbra, Portugal
icpg@dei.uc.pt

Sara Silva
BioISI - Biosystems & Integrative Sciences Institute,
Faculty of Sciences,
University of Lisbon, Campo Grande, 1749-016 Lisbon,
Portugal
sara@fc.ul.pt

Carlos M. Fonseca
CISUC, Department of Informatics Engineering,
University of Coimbra,
3030-290 Coimbra, Portugal
cmfonsec@dei.uc.pt

Mauro Castelli
NOVA IMS, Universidade Nova de Lisboa, 1070-312
Lisbon, Portugal
mcastelli@novaims.unl.pt

ABSTRACT

This paper shows how arbitrarily close alignments in the error space can be achieved by Genetic Programming. The consequences for the generalization ability of the resulting individuals are explored.

Keywords

Geometric Semantic Genetic Programming;
Error Space Alignment; Generalization; Overfitting

1. SEMANTIC AND ERROR SPACE

Geometric Semantic Genetic Programming (GSGP) [3] is a form of Genetic Programming (GP) in which its variation operators induce a cone landscape over any supervised learning problem. These operators have generated considerable interest within the research community. Recently, the geometric semantic mutation was even extended to the real of feedforward Neural Networks [2]. In this context semantics refers to the outputs of a GP individual over a set of data instances. In line with the work being conducted in semantics in GP, a recent work by Ruberto et al. [4] explored the possibility of reaching optimal individuals in terms of training data by searching for aligned or coplanar individuals in the error space. Following and adapting from Ruberto et al. [4], let $\mathbf{X} = \{\overrightarrow{x_1}, \overrightarrow{x_2}, ..., \overrightarrow{x_n}\}$ be the set of input data of a supervised learning problem, and $\overrightarrow{t} = [t_1, t_2, ..., t_n]$ the vector of the respective target values. In the supervised learning context, a GP individual is essentially a function that computes an output for each input instance $\overrightarrow{x_i}$. These outputs define the semantics of an individual. Let $I(\overrightarrow{x_i})$ be the output of an individual I to a given input instance $\overrightarrow{x_i}$. Consequently, the semantics of an individual can be defined as the vector $\overrightarrow{s_I} = [I(\overrightarrow{x_1}), I(\overrightarrow{x_2}), ..., I(\overrightarrow{x_n})]$. The error vector of an individual can be constructed from the semantics vector by taking into account the target vector: $\overrightarrow{e_I} = \overrightarrow{s_I} - \overrightarrow{t}$. Two GP individuals A and B are optimally aligned in the error space if a scalar k exists such that: $\overrightarrow{e_A} = k \cdot \overrightarrow{e_B}$. Ruberto et al. [4] showed that from these two aligned individuals it is possible to analytically construct an optimal individual in terms of training data. However, in the experiments conducted by these authors they were unable to find aligned or coplanar individuals. Consequently, the question of how to effectively find these regularities in the error space is still unanswered. An important related question is that of generalization to unseen data. It is still unclear how would the individuals created from the aligned or coplanar individuals behave in terms of generalization. Would these individuals generalize well or would they overfit?

2. ARBITRARILY CLOSE ALIGNMENTS

The approach proposed in this paper is based on the fact that a target semantics that guarantees an optimal alignment can be easily defined. Consequently, the search can be modified to find individuals that produce this new target semantics, as opposed to searching directly for individuals that fit the original targets. By choosing a given individual and a given error vector proportionality constant (k), it is immediately defined what is the semantics of the other individual needed to reach an optimal alignment. Concretely, given any individual A and any error vector proportionality constant k, an individual B is optimally aligned with A if it produces by the following error vector: $\overrightarrow{e_B} = \overrightarrow{e_A}/k$, which consequently derives from the following semantic vector: $\overrightarrow{s_B} = \overrightarrow{e_B} + \overrightarrow{t}$. Knowing now the definition of $\overrightarrow{s_B}$, the search can be redefined to find an individual that produces this semantics. By achieving this, an optimal alignment is found and, consequently, an optimal individual in terms of

GECCO'16 Companion July 20-24, 2016, Denver, CO, USA

ACM ISBN 978-1-4503-4323-7/16/07.
DOI: http://dx.doi.org/10.1145/2908961.2908988

training data can be constructed. It is important to emphasize that this redefinition of the target semantics does not change the nature of the semantic space. In particular, by using the geometric semantic operators, the search process can still take advantage of the unimodality of the semantic space. Since this space has no local optima, and given that the geometric semantic mutation is an effective operator [1], the search can be focused around the current best individual. In other words, a geometric semantic hill climber is an effective and efficient strategy for conducting the search. Recently it has been shown that using a geometric semantic hill climber leads to faster convergence when compared to a traditional population-based approach [1]. The geometric semantic crossover was also shown to be inefficient in comparison with the geometric semantic mutation [1]. For these reasons, the approach proposed in this paper uses a geometric semantic hill climber to explore the search space. This approach will be referred to as ACA-SSHC: Arbitrarily Close Alignments (ACA) - Semantic Stochastic Hill Climber (SSHC) [3]. Starting with a given k and any initial individual (A), ACA-SSHC computes the new target semantics ($\overrightarrow{s_B}$), and it continually applies the geometric semantic mutation until a desired alignment is found. In this work the stopping criterion is based on the training error of the individual that can be constructed analytically from the current best alignment found.

The experimental parameters used are the same as in Gonçalves et al. [1]. The mutation used (DABM [1]) is a modified version of the original geometric semantic mutation that adds a weight to the parent, and optimally computes this weight and the weight equivalent to the mutation step. This mutation was shown to be very efficient [1]. A SSHC that searches directly for the original target is used as baseline for comparison against the proposed approach. For both approaches, the training error threshold used as the stopping criterion is 1.0 Root Mean Squared Error (RMSE). Notice that since the stopping criterion is based on the training error, the number of generations is variable. The error on unseen data is referred to as generalization error. The datasets used are the same as in Ruberto et al. [4]: PPB and LD50.

3. RESULTS AND CONCLUSIONS

Preliminary testing showed that the value of k does not have influence in the search outcome. For this reason, the following results are presented only for the specific case of $k = 1.1$. Starting with ACA-SSHC, we first present the results that show that this approach is effective in finding near-optimal alignments in the error space. The alignment of two individuals (A and B) can be computed by the absolute cosine similarity between their respective error vectors: $|\cos\theta| = |\frac{\overrightarrow{e_A} \times \overrightarrow{e_B}}{||\overrightarrow{e_A}|| \cdot ||\overrightarrow{e_B}||}|$. Where × represents the scalar product between two vectors, and $||.||$ represents the Euclidean norm of a vector. Two optimally aligned individuals will have an absolute cosine similarity of exactly 1, and two orthogonal individuals (i.e., with an angle of 90°) will have an absolute cosine similarity of exactly 0. Table 1 shows the median, average, and Standard Deviation (SD) of the absolute cosine similarity achieved in both datasets. These results show that the proposed approach is effective in finding near-optimal alignments.

Table 1: Absolute cosine similarity achieved

Dataset	**Absolute cosine similarity**		
	Median	**Average**	**SD**
PPB	0.99879	0.99868	0.00040
LD50	9.999996e-01	9.999996e-01	5.527597e-08

Table 2: Results for ACA-SSHC and SSHC

Dataset	**Method**	**Generalization error**		
		Median	**Average**	**SD**
PPB	ACA-SSHC	42.6	3.2e+07	1.5e+08
	SSHC	49.7	3.9e+09	2.1e+10
LD50	ACA-SSHC	5316.1	6.9e+49	3.8e+50
	SSHC	4457.4	3.8e+26	2.0e+27

Table 2 shows the generalization error achieved for ACA-SSHC and SSHC. Both approaches overfit the training data. In comparable experiments, the best median generalization error achieved is around 30 RMSE for PPB and around 2000 for LD50, e.g., [1]. This shows that searching for near-optimal alignments is risky in terms of generalization. However, it also shows that the issue of overfitting is not particular to an approach with near-optimal alignments, and that it is more clearly related to the very low training error threshold defined. In terms of the number of generations needed to reach the threshold (not shown), both approaches are relatively similar.

A closer look to the PPB results shows that it is possible, in the best runs, to reach near-optimal alignments while still achieving a competitive generalization (around or lower 30 RMSE). The best run achieves 27.7 RMSE. This indicates that a possible research venue is to identify some additional criteria that could help identify whether a given alignment/run is likely to be competitive in terms of generalization.

4. REFERENCES

[1] I. Gonçalves, S. Silva, and C. M. Fonseca. On the generalization ability of geometric semantic genetic programming. In *Genetic Programming*, pages 41–52. Springer, 2015.

[2] I. Gonçalves, S. Silva, and C. M. Fonseca. Semantic learning machine: A feedforward neural network construction algorithm inspired by geometric semantic genetic programming. In *Progress in Artificial Intelligence*, volume 9273 of *Lecture Notes in Computer Science*, pages 280–285. Springer International Publishing, 2015.

[3] A. Moraglio, K. Krawiec, and C. G. Johnson. Geometric semantic genetic programming. In *Parallel Problem Solving from Nature-PPSN XII*, pages 21–31. Springer, 2012.

[4] S. Ruberto, L. Vanneschi, M. Castelli, and S. Silva. Esagp–a semantic gp framework based on alignment in the error space. In *Genetic Programming*, pages 150–161. Springer, 2014.