Multi-objective Genetic Programming for Symbolic Regression with the Adaptive Weighted Splines Representation

Christian Raymond, Qi Chen, Bing Xue, Mengjie Zhang
School of Engineering and Computer Science Victoria University of Wellington
Wellington, New Zealand
Christian.Raymond,Qi.Chen,Bing.Xue,Mengjie.Zhang@ecs.vuw.ac.nz

Abstract

Genetic Programming (GP) for symbolic regression often generates over-complex models, which overfit the training data and have poor generalization onto unseen data. One recent work investigated controlling model complexity by using a new GP representation called Adaptive Weighted Splines (AWS), which is a semi-structured representation that can control the model complexity explicitly. This work extends this previous work by incorporating a new parsimony pressure objective to further control the model complexity. Experimental results demonstrate that the new multi-objective GP method consistently obtains superior fronts and produces better generalizing models compared to single-objective GP with both the tree-based and AWS representation as well a multi-objective tree-based GP method with parsimony pressure.

Keywords

Genetic Programming, Symbolic Regression, Generalization, Model Complexity, Evolutionary Multi-objective Optimization

ACM Reference Format:
Christian Raymond, Qi Chen, Bing Xue, Mengjie Zhang. 2021. Multi-objective Genetic Programming for Symbolic Regression with the Adaptive Weighted Splines Representation. In *2021 Genetic and Evolutionary Computation Conference Companion (GECCO '21 Companion), July 10–14, 2021, Lille, France.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3449726.3459461

1 Introduction and Background

Genetic Programming (GP) for Symbolic Regression (GPSR) [4] aims to learn a mathematical function that best represents some unknown functional relationship between the given features X and the continuous target y. Compared to traditional regression methods GPSR is able to learn both the model structure and the model parameters simultaneously without any prior assumptions about the distribution of the data or the model structure. With its flexible representation ability and the symbolic nature of solutions, GPSR is typically good at learning complex underlying relationship from the data. However, the downside of the flexible representation is that it often learns over-complex models which are prone to overfitting, causing poor generalization onto the unseen data.

GECCO '21 Companion, July 10–14, 2021, Lille, France

ACM ISBN 978-1-4503-8351-6/21/07.
https://doi.org/10.1145/3449726.3459461

Recent work presented in [7] attempted to regulate this problem of poor generalization via introducing a new representation for GPSR called Adaptive Weighted Splines (AWS). Which has the benefit of more explicit control over the model complexity through the use of splines. When using GP-AWS, models are represented by an aggregation of p *feature splines*, which models each feature in the input space X. Each feature spline has three components: *a smoothing spline* S, *a primary coefficient* θ and *a secondary coefficient* β. The two coefficients θ and β simulate embedded feature selection and feature weighting. The p splines are linearly combined using a weighted summation operation to predict y [7].

Goals: This work aims to develop a new GP-AWS method which uses Evolutionary Multi-objective Optimization (EMO) techniques to minimize both the training error and the number of active features/parameters in the model. To investigate whether considering the model complexity as a separate objective in GPSR with the AWS representation can lead to the development of more parsimonious and better generalizing models.

2 The Proposed Method

Research into the AWS representation has thus far been treated as a single-objective optimization problem [7], where only the empirical error is considered. Historically a promising avenue of research to improving the generalization capabilities of GP models has been to minimize both the training error and some model complexity penalty simultaneously.

Accordingly a new multi-objective method for GPSR with the newly proposed AWS representation is developed, which is called Adaptive Weighted Splines with Parsimony Pressure (GP-AWS-PP). Instead of only minimizing the training Mean Squared Error (MSE) as shown in f_1, where y and $\hat{y}$ are the true and predicted output values respectively, and n is the number of instances, an additional objective is also minimized simultaneously which is the total number of active features used by the model, made explicit by the primary coefficients θ as shown in f_2.

$$f_1 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \quad (1) \qquad f_2 = \sum_{i=1}^{p}\theta_i\begin{cases}1, & \text{if } \theta_i \geq 0.5\\ 0, & \text{if } \theta_i < 0.5\end{cases} \quad (2)$$

The additional objective f_2 is expected to improve the generalization of the models generated by GP with the AWS representation. This is because minimizing the number of active/selected features can reduce the tendencies of a model to learn spurious patterns in features that are not relevant to the target y, or are redundant if the signal is captured by another feature [1, 8].

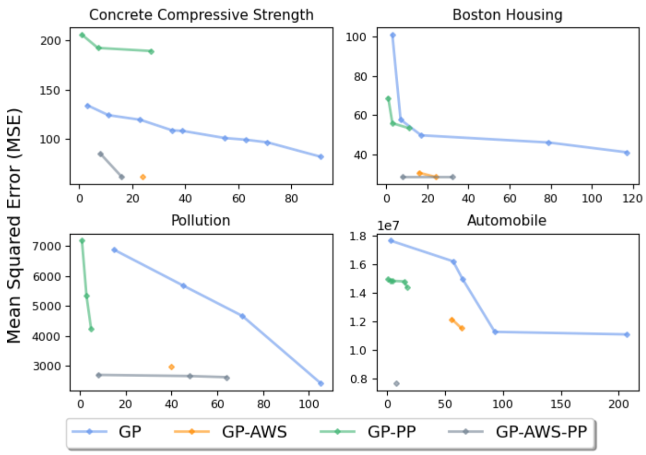

Figure 1: Best Non-dominated Fronts based on the *Testing* MSE and the Number of Parameters.

3 Experiments and Results

Benchmark Methods: The proposed method is compared to 3 benchmark methods: GP with Parsimony Pressure (GP-PP) [5] a multi-objective GPSR method which minimizes both the training MSE as well as the the number of nodes, Genetic Programming (GP) [4] a standard single objective implementation of GPSR using the tree-based representation, and GP with AWS (GP-AWS) a single-objective GPSR method which uses the AWS representation [7]. The same hyper-parameters are used where possible to ensure the fairest possible comparison.

Benchmark Problems: The newly proposed GP-AWS-PP method as well as the 3 benchmark methods are evaluated on 4 real-world regression datasets taken from previous research [2, 6]. The following datasets *Concrete Compressive Strength*, *Boston Housing* and *Automobile* were procured from the UCI Machine Learning Repository. The remaining real-world dataset *Pollution* was taken from the CMU StatLib dataset archive.

Pareto Front Results: The final testing fronts extracted from 100 independent runs on each dataset are shown in Fig. 1. The fronts are computed by taking the union of 100 independent executions and taking the best MSE values (f_1) at each discrete increment with respect to the number of parameters (f_2) and removing all dominated solutions, thus giving the *best nondominated front*.

Analyzing the testing fronts it is observed that GP-AWS-PP generally achieves very promising generalization performance, which is notably better than both GP and GP-PP on all of the tested datasets. In contrast to the single-objective GP-AWS method which typically constructs highly complex solutions (in terms the number of parameters), GP-AWS-PP is able to develop simpler solutions using only a small fraction of the full feature set, consequently generalizes much better onto the testing set. This is especially noticeable on the both the Pollution and Automobile datasets.

Hypervolume Results: The mean ± standard deviation hypervolume values [3] over 100 independent executions and are shown in Table 1, comparing the performance of multi-objective methods GP-AWS-PP and GP-PP. Statistical significance testing is reported using the Mann-Whitney U-test (α = 0.01). A "+" indicates that

Table 1: Training and Testing Hypervolume Values based on the MSE and the Number of Parameters.

Dataset	Method	Training Avg ± Std	SS	Testing Avg ± Std	SS
Concrete	GP-PP	0.8116 ± 0.0181	-	0.8102 ± 0.0214	-
	GP-AWS-PP	0.9402 ± 0.0034	+	0.9353 ± 0.0042	+
Boston	GP-PP	0.9532 ± 0.0081	-	0.8576 ± 0.0228	-
	GP-AWS-PP	0.9819 ± 0.0004	+	0.9317 ± 0.0061	+
Pollution	GP-PP	0.9852 ± 0.0024	-	0.9831 ± 0.0037	-
	GP-AWS-PP	0.9917 ± 0.0000	+	0.9884 ± 0.0004	+
Automobile	GP-PP	0.8823 ± 0.0249	-	0.7766 ± 0.0996	-
	GP-AWS-PP	0.9374 ± 0.0077	+	0.8909 ± 0.0148	+

the method has significantly better performance compared to the opposing method, a "-" indicates that the method has significantly worse performance, and a "=" indicates no significant difference.

Analyzing the training hypervolume results it is observed that GP-AWS-PP performs statistically significantly better on all of the tested datasets when compared to GP-PP. Regarding the testing hypervolume results, the results remain largely consistent with the training results, continuing to show notably better performance on all datasets. These results reveals that in the majority of cases GP-AWS-PP is able to achieve much better hypervolume values compared to GP-PP. Furthermore, results also shown that GP-AWS-PP typically has a much higher mean and smaller standard deviation hypervolume values compared to GP-PP (with the exception of the Pollution dataset where results are somewhat close), suggesting that GP-AWS-PP consistently learns superior fronts compared to GP-PP on average.

4 Conclusion

This paper has conducted the first investigation into applying EMO techniques to the semi-structured AWS respresentation for GPSR. The experimental results conducted against the two tree-based GP methods, and the single objective GP-AWS method highlight the highly performant generalization capabilities of the newly proposed GP-AWS-PP method, which puts parsimony pressure on the number of active features to promote better generalizing models.

References

[1] Harith Al-Sahaf, Ying Bi, Qi Chen, Andrew Lensen, Yi Mei, Yanan Sun, Binh Tran, Bing Xue, and Mengjie Zhang. 2019. A survey on evolutionary machine learning. *Journal of the Royal Society of New Zealand* 49, 2 (2019), 205–228.
[2] Qi Chen, Bing Xue, and Mengjie Zhang. 2020. Improving symbolic regression based on correlation between residuals and variables. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*. 922–930.
[3] Carlos M Fonseca, Luís Paquete, and Manuel López-Ibánez. 2006. An improved dimension-sweep algorithm for the hypervolume indicator. In *2006 IEEE international conference on evolutionary computation*. IEEE, 1157–1163.
[4] John R Koza and John R Koza. 1992. *Genetic programming: on the programming of computers by means of natural selection*. Vol. 1. MIT press.
[5] Sean Luke and Liviu Panait. 2002. Fighting Bloat with Nonparametric Parsimony Pressure. In *International Conference on Parallel Problem Solving from Nature*. Springer, 411–421.
[6] Christian Raymond, Qi Chen, Bing Xue, and Mengjie Zhang. 2019. Genetic Programming with Rademacher Complexity for Symbolic Regression. In *2019 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2657–2664.
[7] Christian Raymond, Qi Chen, Bing Xue, and Mengjie Zhang. 2020. Adaptive weighted splines: a new representation to genetic programming for symbolic regression. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*. 1003–1011.
[8] Bing Xue, Mengjie Zhang, Will N Browne, and Xin Yao. 2015. A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation* 20, 4 (2015), 606–626.