# Feature Selection Using Geometric Semantic Genetic Programming

G. H. Rosa, J. P. Papa
Department of Computing
São Paulo State University
Bauru, São Paulo, Brazil 17033-360
[gustavo.rosa,papa]@fc.unesp.br

L. P. Papa
São Paulo Southwestern College
Avaré, São Paulo, Brazil 18707-150
lucienepapa@yahoo.com.br

## ABSTRACT

Feature selection concerns the task of finding the subset of features that are most relevant to some specific problem in the context of machine learning. During the last years, the problem of feature selection has been modeled as an optimization task, where the idea is to find the subset of features that maximize some fitness function, which can be a given classifier's accuracy or even some measure concerning the samples' separability in the feature space, for instance. In this paper, we introduced Geometric Semantic Genetic Programming (GSGP) in the context of feature selection, and we experimentally showed it can work properly with both conic and non-conic fitness landscapes.

## CCS CONCEPTS

•**Computing methodologies → Genetic programming;**

## KEYWORDS

Feature selection, Geometric Semantic Genetic Programming

**ACM Reference format:**
G. H. Rosa, J. P. Papa and L. P. Papa. 2017. Feature Selection Using Geometric Semantic Genetic Programming. In *Proceedings of GECCO '17 Companion, Berlin, Germany, July 15-19, 2017,* 2 pages.
DOI: http://dx.doi.org/10.1145/3067695.3076020

## 1 INTRODUCTION

Machine learning techniques have been the forerunner of several advances in Computer Science and application-driven areas, which range from medical image understanding to video summarization, just to name a few. However, even the most accurate approaches may have their performance degraded due to the high dimensionality of the datasets. In this context, *feature selection* arises to mitigate that problem by selecting the subset of the most representative features.

A Binary Flower Pollination Algorithm was presented for feature selection purposes and compared against Particle Swarm Optimization (PSO), Harmony Search and Firefly Algorithm [11]. Evolutionary-oriented optimization techniques have been also used to find out the most representative features. Yang and Honavar [14], for instance, used Genetic Algorithms together with Neural Networks for feature selection purposes. Genetic Programming (GP) [5] was also employed for the very same purpose, either representing classifiers instanced with different subsets of features [6, 10] or using a two-stage approach [3].

Moreover, Geometric Semantic Genetic Programming (GSGP) [7] has been employed to a number of problems very recently, such as electoral redistricting problem [2] and real-life applications [12]. However, as far as we are concerned, GSGP has never been considered in the context of feature selection up to date, which turns out to be the main contribution of this paper. One strong point of geometric semantic operators concerns their ability in inducing unimodal fitness landscapes on some problems where one knows the matching between the input and the output data.

## 2 METHODOLOGY

The proposed approach aims at selecting the set of features that minimizes the classification error of some supervised classifier over the validation set (i.e. the so-called wrapper approach). Although one can use any supervised pattern recognition technique, we opted to use the Optimum-Path Forest (OPF) classifier [8, 9], since it is parameterless and fast for training. This procedure is hereinafter called "Experiment A". However, "Experiment A" does not guarantee a unimodal fitness landscape, since the fitness function is not based on the Hamming distance. The main idea of "Experiment B" is to find the best subset of features as the one that maximizes the OPF accuracy over a validation set. The best subset is considered among all possible subsets, say $2^n$, where $n$ stands for the number of features. Finally, the best subset is taken as our gold standard, and the fitness function now aims at minimizing the Hamming distance between the current solution and that gold standard.

### 2.1 Datasets

Table 1 describes the datasets used in this work.

| | # Training Set | # Testing Set | # Features |
|---|---|---|---|
| Letter [4] | 15,000 | 5,000 | 16 |
| Pendigits [1] | 7,494 | 3,498 | 16 |
| Segment [4] | 1,155 | 1155 | 19 |
| Vehicle [4] | 423 | 423 | 18 |

**Table 1: Datasets considered in the work.**


*GECCO '17 Companion, Berlin, Germany*

DOI: http://dx.doi.org/10.1145/3067695.3076020

## 2.2 Experimental Setup

In this work, we compared Geometric Semantic Genetic Programming against three approaches for feature selection purposes, say that: Bat Algorithm (BA), GP and PSO. In order to provide a statistical analysis by means of Wilcoxon signed-rank test [13], we conducted a $k$-fold cross-validation with 15 runnings for both experiments (Experiment "A" and Experiment "B"). We employed 15 agents over 25 iterations for convergence considering all techniques and experiments. In regard to the source-code, we used the optimization library LibOPT[1], and the development library LibDEV[2]. Concerning the OPF classifier, we used the LibOPF library[3].

## 3 EXPERIMENTS

In this section, we present the results concerning the experiment that holds the assumption the fitness landscape is unimodal. In regard to this experiment, we need to find out the gold standard by means of an exhaustive search over $2^n$ possibilities, where $n$ stands for the number of features. As aforementioned in Section 2, the idea is to find out the subset of features that minimizes the Hamming distance with respect to the gold standard. Table 2 presents the average Hamming distance concerning the aforementioned datasets and the techniques used. The best results are in bold according to Wilcoxon statistical test. In this situation, the technique is better when the distance is smaller.

| | BA | GP | GSGP | PSO |
|---|---|---|---|---|
| Letter | **0.71** | 2.02 | 1.87 | **0.74** |
| Pendigits | **2.90** | 4.22 | 4.20 | **2.91** |
| Segment | **2.50** | 3.34 | 3.47 | **2.66** |
| Vehicle | **2.07** | 2.95 | 2.88 | **1.99** |

**Table 2: Average Hamming distance considering the datasets used.**

One can observe that BA and PSO obtained the best results, followed by GSGP and GP. However, an interesting point concerns a direct comparison between GSGP and naïve GP, given the first one obtained slightly better results.

We performed an additional experiment to evaluate GSGP over both non-unimodal and unimodal fitness landscapes. Table 3 presents a comparison between GSGP under "Experiment A" (GSGP-A) and GSGSP under "Experiment B" (GSGP-B), being the best techniques in bold according to Wilcoxon statistical test. In this experiment, we considered the best subset of features selected by both experiments to train and evaluate an OPF classifier in order to assess GSGP behavior under that different conditions, i.e., we would like to assess whether GSGP would benefit or not from unimodal fitness landscapes concerning the problem of feature selection. One can observe that both GSGP-based techniques obtained similar results in 2 out of 4 datasets, being GSGP over unimodal fitness the best one in all situations, which was expected, since we assume the operators are "really semantic".

## 4 CONCLUSIONS

In this paper, we tackled the problem of feature selection as an evolutionary optimization problem. We showed GSGP can obtain very much reasonable results in 4 public datasets without the guarantee one has unimodal fitness landscapes (GSGP-A). A second experiment (GSGP-B) used four datasets in order to obtain a gold standard to be used as the fitness function. In this case, we hold the assumption of using semantic operators. In 2 out of 4 datasets, GSGP-A obtained similar results to GSGP-B, being the latter one the best in all four datasets, as expected. We believe the results presented in this paper can make even broader the applications of GSGP. In regard to future works, we intend to compare GSGP with different meta-heuristic techniques, as well as to present a quaternion-based version of GSGP.

| | GSGP-A | GSGP-B |
|---|---|---|
| Letter | 93.37% | **97.01%** |
| Pendigits | **97.10%** | **97.74%** |
| Segment | **97.43%** | **97.53%** |
| Vehicle | 76.58% | **78.74%** |

**Table 3: GSGP comparison between "Experiment A" and "Experiment B".**

[1]https://github.com/jppbsi/LibOPT
[2]https://github.com/jppbsi/LibDEV
[3]https://github.com/jppbsi/LibOPF

## ACKNOWLEDGMENTS

This work was supported by FAPESP grants #2014/16250-9, #2014/12236-1 and #2015/25739-4, as well as CNPq grant #306166/2014-3.

## REFERENCES

[1] Fevzi Alimoglu and Ethem Alpaydin. 1996. Methods of combining multiple classifiers based on different representations for pen-based handwritten digit recognition. In *Proceedings of the Fifth Turkish Artificial Intelligence and Artificial Neural Networks Symposium (TAINN 96*. Citeseer.
[2] M. Castelli, R. Henriques, and L. Vanneschi. 2015. A geometric semantic genetic programming system for the electoral redistricting problem. *Neurocomputing* 154 (2015), 200–207.
[3] R. A. Davis, A. J. Charlton, S. Oehlschlager, and J. C. Wilson. 2006. Novel feature selection method for genetic programming using metabolomic 1H {NMR} data. *Chemometrics and Intelligent Laboratory Systems* 81, 1 (2006), 50–59.
[4] Chih-Wei Hsu and Chih-Jen Lin. 2002. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks* 13, 2 (2002), 415–425.
[5] J.R. Koza. 1992. *Genetic programming: on the programming of computers by means of natural selection*. The MIT Press, Cambridge, USA.
[6] J.-Y. Lin, H.-R. Ke, B.-C. Chien, and W.-P. Yang. 2008. Classifier Design with Feature Selection and Feature Extraction Using Layered Genetic Programming. *Expert Systems with Applications: An International Journal* 34, 2 (Feb. 2008), 1384–1393.
[7] A. Moraglio, K. Krawiec, and C. G. Johnson. 2012. *Geometric Semantic Genetic Programming*. Springer Berlin Heidelberg, Berlin, Heidelberg, 21–31.
[8] J. P. Papa, A. X. Falcão, V. H. C. Albuquerque, and J. M. R. S. Tavares. 2012. Efficient supervised optimum-path forest classification for large datasets. *Pattern Recognition* 45, 1 (2012), 512–520.
[9] J. P. Papa, A. X. Falcão, and C. T. N. Suzuki. 2009. Supervised pattern classification based on optimum-path forest. *International Journal of Imaging Systems and Technology* 19, 2 (2009), 120–131.
[10] Ramirez R. and Puiggros M. 2007. *A Genetic Programming Approach to Feature Selection and Classification of Instantaneous Cognitive States*. Springer Berlin Heidelberg, Berlin, Heidelberg, 311–319.
[11] D. Rodrigues, X.-S. Yang, A. N. Souza, and J. P. Papa. 2015. *Recent Advances in Swarm Intelligence and Evolutionary Computation*. Springer International Publishing, Cham, Chapter Binary Flower Pollination Algorithm and Its Application to Feature Selection, 85–100. DOI: http://dx.doi.org/10.1007/978-3-319-13826-8_5
[12] L. Vanneschi, S. Silva, M. Castelli, and L. Manzoni. 2014. *Geometric Semantic Genetic Programming for Real Life Applications*. Springer New York, New York, NY, 191–209.
[13] F. Wilcoxon. 1945. Individual Comparisons by Ranking Methods. *Biometrics Bulletin* 1, 6 (1945), 80–83.
[14] J. Yang and V. G. Honavar. 1998. Feature Subset Selection Using a Genetic Algorithm. *IEEE Intelligent Systems* 13, 2 (March 1998), 44–49.