

MAP-Elites for Genetic Programming-Based Ensemble Learning: An Interactive Approach

Abstract

Evolutionary ensemble learning is an emerging research area, and designing an appropriate quality-diversity optimization algorithm to obtain a set of effective and complementary base learners is important. However, how to maintain such a set of learners remains an open issue. This paper proposes using cosine similarity-based dimensionality reduction methods to maintain a set of effective and complementary base learners within the MAP-Elites framework for evolutionary ensemble learning. Additionally, this paper proposes a reference point synthesis strategy to address the issue of individuals being unevenly distributed in semantic space. The experimental results show that the ensemble model induced by the cosine similarity-based dimensionality reduction method outperforms the models induced by the other seven dimensionality reduction methods in both interactive examples and large-scale experiments. Moreover, reference points are shown to be helpful in improving the algorithm's effectiveness. The main contribution of this paper is to provide an interactive approach to explore the methods and results, which is detailed in the full paper presented in IEEE Xplore.

I. Introduction

Evolutionary ensemble learning has emerged as a significant area in both evolutionary computation and machine learning domains [1], [2]. One of the

key advantages of evolutionary ensemble learning algorithms is their ability to generate a diverse set of high-quality base learners in a single run within the framework of quality-diversity (QD) optimization algorithms. QD optimization algorithms, such as the multi-dimensional archive of phenotypic elites (MAP-Elites) [3], have shown promising performance in evolutionary ensemble learning. In MAP-Elites, an important question that determines which individuals survive is how to define niches, where the dimensionality reduction method plays a crucial role. This paper investigates the impact of dimensionality reduction methods on MAP-Elites in the context of evolutionary ensemble learning. The main contribution of this paper is to provide an interactive approach to exploring the key aspects of using MAP-Elites for evolutionary ensemble learning, offering an enhanced understanding of MAP-Elites-based evolutionary ensemble learning and its performance on machine learning tasks.

II. Evolutionary Ensemble Learning

Evolutionary ensemble learning uses an evolutionary algorithm to generate base learners for building an ensemble model. It can be genetic programming (GP) [1], learning classifier systems (LCS) [4], or neural networks (NN) [5]. For evolutionary ensemble learning, it is important to evolve a set of high-quality and diverse base learners. For this purpose, bootstrap sampling strategy [6], niching strategy [7], and random decision trees [1] can be used.

Recently, MAP-Elites, a quality-diversity optimization algorithm, has shown good performance in evolutionary ensemble learning [7]. In MAP-Elites, it is important to define a behavior space where we can map individuals into and characterize the possible behaviors of high-quality individuals. The existing methods for designing such a space include using domain knowledge [8] and auto-encoders [7]. This paper focuses on studying how to induce a good behavior space in MAP-Elites for obtaining a good ensemble model.

III. Algorithm

A. Algorithm Overview

Using the MAP-Elites framework, the idea of evolutionary ensemble learning is to dynamically maintain a set of GP elites in a discretized behavior space. These GP elites form an ensemble model, where each contains m GP trees for feature construction and a machine learning model for prediction. Specifically, m GP trees represent m features constructed based on the original training data. The machine learning model is then trained on these features. The final prediction of the ensemble model is the average of all machine learning model predictions. In this paper, linear regression is used as the machine learning model due to its simplicity, efficiency, and strong predictive performance.

B. Solution Initialization and Evaluation

In the solution initialization stage, m GP trees in each individual are randomly initialized using the ramped half-and-half method [1], [2]. The solution evaluation stage then transforms the training data into a new feature space using the GP trees.

TABLE I Statistical comparison of test mean square error for different dimensionality reduction techniques (x/y/z mean that a method in a row is significantly better than, similar to, or worse than the method in the column on x/y/z datasets).

	PCA	KPCA(RBF)	KPCA(POLY)	t-SNE	Beta-VAE	Isomap	SpectralEmbedding
KPCA(COSINE)	44/62/0	61/44/1	73/33/0	58/48/0	70/36/0	63/43/0	64/42/0
KPCA(COSINE)+Reference Point	61/45/0	71/34/1	74/32/0	67/39/0	73/33/0	70/36/0	71/35/0

Based on the transformed feature space, a linear model is trained to make predictions. To ensure that the constructed features generalize well to unseen data, the model makes predictions on the training data by using cross-validation. The predictions made by the models are referred to as semantics in GP literature, and the target labels are known as target semantics. The semantics of all GP individuals together form a semantic space, where the objective of GP is to discover an individual that can output target semantics based on given inputs.

C. Archive Maintenance

The archive maintenance step is crucial for selecting individuals to form the final ensemble model. In this paper, MAP-Elites is used to select a set of high-quality and complementary solutions, which involves four stages:

- ❑ Reference Point Synthesis: For a learning dataset (X, Y) of a supervised learning task, where the target label Y is known, the semantics of ideal individuals Φ can be synthesized by $(1 - \alpha) \times \Phi(X) + \alpha \times Y$. α is a hyperparameter set to 0.1 and 1.1 in this paper, which corresponds to the ideal points of high-quality individuals and symmetrical high-quality individuals, respectively.
- ❑ Dimensionality Reduction: The ideal semantic space is high-dimensional, making it challenging to define a niche in such a space. To address this issue, MAP-Elites employs a dimensionality reduction method, transforming the high-dimensional space into a lower-dimensional space that can be discretized more easily. Many methods can be used for dimensionality reduction in MAP-Elites. In this work, kernel principal component analysis (KPCA) with a cosine kernel is adopted because it demonstrated superior performance in the experiments, as discussed in Section IV.

- ❑ Space Discretization: Based on the reduced space, MAP-Elites discretizes the semantic space into a $k \times k$ grid, where k is a hyperparameter that determines the granularity of MAP-Elites. Each grid cell represents a niche, containing individuals with similar behaviors.
- ❑ Elites Selection: Finally, the best individual in each grid cell is preserved based on the discrete behavior space.

D. Solution Selection and Generation

Once obtained a set of diverse and high-quality individuals in the external archive A , promising individuals are selected from the external archive A using random selection. Based on the selected parent individuals, offspring are generated using random subtree crossover and mutation. For multi-tree GP, crossover and mutation are applied to randomly select GP trees.

IV. Experimental Results

The experimental results in Table I show the impact of dimensionality reduction methods on the test mean squared errors of ensemble models across 106 datasets. The results indicate that incorporating KPCA (COSINE) as the dimensionality reduction technique within MAP-Elites leads to a significant improvement in the performance of ensemble models. Specifically, KPCA (COSINE) outperforms PCA in MAP-Elites on 44 datasets and does not perform worse on any dataset. Furthermore, when synthesizing reference points, KPCA (COSINE) exhibits even better performance, surpassing PCA on 61 datasets and not being outperformed on any dataset.

V. Conclusion

This paper provides an interactive approach to understanding how to use MAP-Elites for evolutionary ensemble learning, as well as eight dimensionality reduction methods for automatically inducing a behavior space based on semantics of GP. The experimental results from interactive examples and large-scale

experiments show that the dimensionality reduction method significantly impacts the predictive performance of the ensemble model within the MAP-Elites framework, with cosine-kernel-based PCA outperforming other methods.

While this paper focuses on MAP-Elites, the idea of using cosine similarity for defining a behavior space could potentially be extended to other QD optimization algorithms for evolutionary ensemble learning. In the future, it would be interesting to investigate the impact of different distance metrics in other QD optimization algorithms.

Acknowledgment

This work was supported in part by the Marsden Fund of New Zealand Government under Contracts VUW1913, VUW1914, and VUW2016, in part by the Science for Technological Innovation Challenge (SfTI) fund under Grant E3603/2903, in part by MBIE Data Science SSIF Fund under Contract RTVU1914, in part by Huayin Medical under Grant E3791/4165, and in part by MBIE Endeavor Research Programme under Contracts C11X2001 and UOCX2104.

References

- [1] H. Zhang, A. Zhou, and H. Zhang, "An evolutionary forest for regression," *IEEE Trans. Evol. Comput.*, vol. 26, no. 4, pp. 735–749, Aug. 2022.
- [2] H. Zhang, A. Zhou, Q. Chen, B. Xue, and M. Zhang, "SR-Forest: A genetic programming based heterogeneous ensemble learning method," *IEEE Trans. Evol. Comput.*, early access, Feb. 07, 2023, doi: 10.1109/TEVC.2023.3243172.
- [3] J.-B. Mouret and J. Clune, "Illuminating search spaces by mapping elites," 2015, *arXiv:1504.04909*.
- [4] H. H. Dam, H. A. Abbas, C. Lokan, and X. Yao, "Neural-based learning classifier systems," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 1, pp. 26–39, Jan. 2008.
- [5] Y. Liu, X. Yao, and T. Higuchi, "Evolutionary ensembles with negative correlation learning," *IEEE Trans. Evol. Comput.*, vol. 4, no. 4, pp. 380–387, Nov. 2000.
- [6] M. Virgolin, "Genetic programming is naturally suited to evolve bagging ensembles," in *Proc. Genet. Evol. Comput. Conf.*, 2021, pp. 830–839.
- [7] K. Nickerson, A. Kolokolova, and T. Hu, "Creating diverse ensembles for classification with genetic programming and neuro-map-elites," in *Proc. Eur. Conf. Genet. Program.*, 2022, pp. 212–227.
- [8] E. Dolson, A. Lalejini, and C. Ofria, "Exploring genetic programming systems with map-elites," in *Genetic Programming Theory and Practice XVI*. Cham, Switzerland: Springer, 2019, pp. 1–16.

