Benchmarking Many-Objective Topology and Weight Evolution of Neural Networks: A Study with NEWS/D

Adham Salih
School of Mechanical Engineering
Tel-Aviv University
Tel-Aviv, Israel
adhamsalih@mail.tau.ac.il

Amiram Moshaiov
School of Mechanical Engineering
Tel-Aviv University
Tel-Aviv, Israel
moshaiov@tauex.tau.ac.il

***Abstract*—This study aims to provide procedures and benchmark problems to test the optimization capabilities of algorithms for Many-Objective Topology and Weight Evolution of Artificial Neural Networks (MaO-TWEANNs). In particular, the proposed benchmarks are based on a combination of continuous functions that have commonly been used to test single-objective optimization algorithms. In addition, this paper applies the proposed procedures and benchmark problems to evaluate NEWS/D, which is such an algorithm that has recently been introduced. The results of this study validate the optimization capabilities of NEWS/D for MaO-TWEANN problems with continuous outputs.**

Keywords— Artificial Neural Networks, Many-objective Optimization, Decomposition Approach, Neuro-evolution, Topology and Weight Evolution, Neural Architecture Search.

I. Introduction

This study follows a recent development of a many-objective TWEANN (Topology and Weight Evolution of Artificial Neural Network) algorithm. The applied algorithm, which has been termed Neuro-Evolution of Weights and Structures by Decomposition (NEWS/D), searches for Pareto-optimal networks and their associated Pareto-front. The significance of using Pareto optimization for the design of Artificial Neural-Networks (ANNs) has become increasingly apparent from various studies (e.g., [1], [2], [3]). However, most such studies involve only a small number of objectives. NEWS/D opens up the possibility to solve TWEANN problems with more than three objectives, hence, it is categorized as a many-objective evolutionary algorithm for TWEANN.

NEWS/D is designed based on two major principles. First, it is founded on the decomposition approach [4] that has been tailored for TWEANN. The decomposition idea is a "divide and conquers" approach that decomposes the problem into a set of single-objective sub-problems, which are simultaneously solved. Second, inspired by [5], special care is taken in NEWS/D to ensure that there is no competition between novice and mature topologies during the evolution.

Commonly, single- and multi-objective TWEANN algorithms have been tested either with real-life problems, such as classification problems [6] or with academic problems such as robot control problems in simple simulated environments [7] and [8], and double pole balancing problems [5] and [9]. However, there are no such agreed benchmark problems with many objectives. To circumvent this lack of benchmarks, the introductory paper to NEWS/D [10] includes testing procedures using multi-class classification datasets ranging from three to seven objectives. However, that procedure includes only samples with discrete outputs. In contrast, here NEWS/D is tested by a set of synthetic benchmark problems which involves continuous outputs. Hence, the main contribution of this paper is to suggest and demonstrate a validation methodology for MaO-TWEANN algorithms as related to their ability to cope with problems with continuous outputs, such as in continuous control problems. For example, a possible application concerns robot control problems in which the controllers will be simultaneously adapted to many environments and many initial conditions, i.e., one objective will be associated with each of the environments and with each of the initial conditions.

The proposed approach to construct synthetic benchmark problems allows testing the algorithm with respect to a mixture of different mappings. It should be noted that this paper does not aim to provide instructions on what kind of a mixture should be used. Rather, it provides the benchmark construction approach and examples on such benchmarks and their application concerning NEWS/D.

The rest of this paper is organized as follows. Section II outlines the background material for this study. Section III presents the methodology that is used in the numerical study, while Section IV provides the numerical results and their analyses. Finally, Section V summarizes and concludes this study.

II. Background

A. Evolutionary Multi-objective Optimization

Consider the following vector optimization problem:

$$\min_{\boldsymbol{x}\in X} \quad [F_1(\boldsymbol{x}), \ldots, F_M(\boldsymbol{x})] \tag{1}$$

2021 IEEE Symposium Series on Computational Intelligence (SSCI) | 978-1-7281-9048-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/SSCI50451.2021.9659924

where M is the number of objective-functions to be minimized, and $X \subseteq \mathbb{R}^d$ is the set of feasible solutions. According to the Pareto-optimality approach, solving this problem provides the Pareto-optimal-set of non-dominated solutions and their associated performance-vectors, which constitute the Pareto-front. The following provides some definitions of the Pareto optimality concept. A solution $\boldsymbol{a}$ with a performance vector $\boldsymbol{F}(\boldsymbol{a}) = [F_1(\boldsymbol{a}), \dots, F_M(\boldsymbol{a})]$ is said to dominate a solution $\boldsymbol{b}$ with a performance vector $\boldsymbol{F}(\boldsymbol{b}) = [F_1(\boldsymbol{b}), \dots, F_M(\boldsymbol{b})]$, denoted by $\boldsymbol{a} \preccurlyeq \boldsymbol{b}$, if the following condition is met:

$$\forall i \in \{1, \dots, M\}, F_i(\boldsymbol{a}) \leq F_i(\boldsymbol{b}) \;\&\; \exists j \in \{1, \dots, M\}, F_j(\boldsymbol{a}) < F_j(\boldsymbol{b}) \quad (2)$$

For a given optimization problem, the Pareto-optimal-set is the set of all non-dominated feasible solutions of the problem. Namely, a solution $\boldsymbol{a} \in X$ is called Pareto-optimal with respect to X if and only if there is no solution $\boldsymbol{b} \in X$ for which $\boldsymbol{b} \preccurlyeq \boldsymbol{a}$. Solving (1) means finding the Pareto-optimal-set and the associated Pareto-front of the problem.

Commonly, there is no analytical solution to the above problem. Hence, numerical techniques are needed to get at least a good approximation for the Pareto-optimal-set which has performances as close as possible to the true Pareto-front. Over the last two decades, a set of reliable multi-objective evolutionary algorithms have been developed for solving the problem in (1).

B. NEWS/D – Algorithm Description

This section provides an overview of an algorithm, which has been recently introduced by the authors in [10]. NEWS/D is a many-objective algorithm for the topology and weight evolution of artificial neural networks.

Algorithm 1: NEWS/D	
Input	N, K, MAX_{NT}, gen_{TM} and Stopping Criterion
Output	Archive
Initialize	Population, Archive and Topology Record
	Perform Sub-problem Generation Procedure
	While stopping criteria is not met:
	Perform Score Assignment Procedure
	Perform Elite Set Procedure
	Perform Protected Set Procedure
	Perform Selection Procedure
	Perform Reproduction Procedure
	Perform Population Update Procedure
	Update the Archive
	$gen \leftarrow gen + 1$
	End While

The following summarizes the required inputs for NEWS/D. The first input is N, which is the number of individuals in the population. The second input is K, which is the number of the decomposed sub-problems. The third input is MAX_{NT}, which defines the maximal number of evolving topologies in each generation. It should be noted that the condition $N > K + MAX_{NT}$ must be satisfied when the above inputs are selected. This will ensure diversity during the selection procedure. The fourth input is gen_{TM}, which defines the number of generations in which topology is considered as a novice. After reaching this number, a topology is considered mature. The output is an external archive of non-dominated solutions, which is updated during the evolutionary process.

In addition to the standard ideas of evolutionary algorithms and the decomposition approach, the design of NEWS/D is based on four special working principles, which aim to handle the MaO-TWEANN needs. The first principle is inspired by [5], where special care is taken to ensure that no competition between novice and mature topologies occurs. In NEWS/D, "fair aged-based" competition is applied. This principle aims to allow novice topologies to mature, namely, to adjust their weights before becoming subject to selection. This is implemented within the "Protected Set Procedure" and the "Population Update Procedure". The second principle is the sub-problem elitist selection. For each sub-problem, the individual with the best score in the sub-problem is automatically selected for the "Reproduction Procedure". This is implemented in the "Elite Set Procedure" and the "Selection Procedure". The third principle is to promote exploration, and in particular the exploration of topologies, by diversifying the individuals that could take a part in the selection within each sub-problem. Namely, rather than using neighborhood sub-problems as commonly done in decomposition-based MaOEAs, in NEWS/D the entire population is used in the "Selection Procedure" regardless of the involved sub-problem. In each selection step, a sub-problem and two individuals are randomly picked out of the entire population. The individual with the higher performance with respect to the considered sub-problem is selected for the "Reproduction Procedure". The fourth principle is the inclusion of input parameters, such as MAX_{NT} and gen_{TM}, which allows balancing between the topology search and the weight search.

In addition, NEWS/D uses the Tchebycheff (TCH) decomposition approach for the sub-problem scores and NEAT genetic encoding.

III. Methods

This section is organized as follows. First, it presents the devised Many-Objective Synthetic Benchmark Problems (MaO-SBPs). Second, it describes S-NEAT, which is based on the well-known NEAT algorithm. S-NEAT is to be used in the following section to compare the optimization capabilities of NEWS/D. Finally, the comparisons and validation methodologies are outlined. While resembling regression problems, it should be emphasized that the benchmark problems and the validation study are about network optimization and not about generalization!

A. The Synthetic Benchmark Problems

Fig. 1 presents a scheme for the network error calculation, $\boldsymbol{e}$, given an input sample, $\boldsymbol{In} \in [\mathbf{0},\mathbf{1}]^{\boldsymbol{d}}$. The ANN maps the input vector into an output vector $\widetilde{\boldsymbol{y}} \in [\mathbf{0},\mathbf{1}]^{\boldsymbol{M}}$. In this study, the network aims at mapping several benchmark functions, as detailed in Tables I and II. The j^{th} component of the output vector aims to provide the mapped value of the j^{th} benchmark function. It is noted that the benchmark functions have different domains and ranges; hence, two special transformations are used, which are denoted as $\boldsymbol{N_1(In)}$ and $\boldsymbol{N_2(f)}$. The role of the former transformation is to map the network input vector, $\boldsymbol{In}$, into the relevant domain of the function. The role of the latter transformation is to map the values of the vector function $\boldsymbol{F(x)}$ into a normalized output vector $\boldsymbol{y}$.

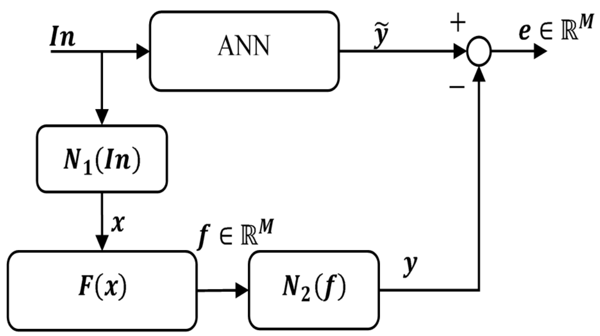

Fig. 1. MaO-SBP Error Calculation Scheme

These transformations, which are presented in equations (3) and (4), allow the same network to be used for the entire set of functions regardless of their different domains and ranges.

$$x_i = N_1(In_i) = x_i^{(L)} + \left(x_i^{(U)} - x_i^{(L)}\right) \cdot In_i \tag{3}$$

$$y_j = N_2(f_j) = \frac{f_j - f_j^{(L)}}{f_j^{(U)} - f_j^{(L)}} \tag{4}$$

In these equations, $x_i^{(L)}$ and $x_i^{(U)}$ are the lower and the upper limits of the i^{th} component of the input vector, respectively. Similarly, $f_j^{(L)}$ and $f_j^{(U)}$ are the minimal and the maximal value of the j^{th} benchmark function. Table I summarizes the details of the seven (scalar) functions that are used to generate MaO-SBPs. These functions are commonly used as benchmark problems for single-objective optimization (e.g., [11]). Yet, here they are combined into benchmark problems for MaO-TWEANN algorithms. It should be noted that in contrast to global optimization problems, where the task of the algorithm is to found the global minima, here the task is to find a mapping model that can simultaneously maps several of these functions.

TABLE I: THE BENCHMARK FUNCTIONS

#	Name	$f(\boldsymbol{x})$	$\boldsymbol{x}$ limits	$f(\boldsymbol{x})$ limits
1	Sphere	$f(x) = \sum_{i=1}^{d} x_i^2$	$[-5.12,5.12]$	$[0,263)$
2	Rosenbrock	$f(x) = \sum_{i=1}^{d-1} (100(x_{i+1} + x_i^2)^2 + (x_i - 1)^2)$	$[-5,10]$	$[0,11 \cdot 10^6)$
3	Quartic	$f(x) = \sum_{i=1}^{d} i x_i^4$	$[-1.28,1.28]$	$[0,148)$
4	Step	$f(x) = \sum_{i=1}^{d} (x_i + 0.5)^2$	$[-100,100]$	$[0,1.01 \cdot 10^5)$
5	Schwefel	$f(x) = \sum_{i=1}^{d} \lvert x_i \rvert + \prod_{i=1}^{d} \lvert x_i \rvert$	$[-10,10]$	$[0, 10^{10})$
6	SumSquare	$f(x) = \sum_{i=1}^{d} i x_i^2$	$[-10,10]$	$[0,5.5 \cdot 10^3)$
7	Rastrigin	$f(x) = \sum_{i=1}^{d} (x_i^2 - 10\cos(2\pi x_i) + 10)$	$[-5.12,5.12]$	$[0,463)$

In this study, d is set to ten (i.e., ten decision variables) for all the benchmark functions. The seven functions of Table I are used to generate ten MaO-SBPs as detailed in Table II. As seen in Table II, the number of objectives, M, in each of the MaO-SBPs is between three to seven.

TABLE II: MAO-SBPS

MaO-SBP	Applied Functions	#Obj (M)
1	2,4,6	3
2	1,2,3,4	4
3	1,2,3,7	4
4	1,3,5,7	4
5	4,5,6,7	4
6	1,2,3,4,5	5
7	3,4,5,6,7	5
8	1,2,3,4,5,6	6
9	2,3,4,5,6,7	6
10	1,2,3,4,5,6,7	7

B. S-NEAT Algorithm

Algorithm 2 presents the pseudo-code of Sequential-NEAT (S-NEAT). It is based on [5], where further information on NEAT can be found.

Algorithm 2: S-NEAT	
Input	N, K Parameters of the Stopping Criterion
Output	Pareto-Set Approximation (PSA)
Initialize	$PSA = \emptyset$
1	Perform Sub-problem Generation Procedure
2	**For** each sub-problem $k \in \{1,2,..,K\}$:
3	**Initialize** $P_{gen=0}$, $Best_k$, and Species Record
4	**While** stopping criteria is not met:
5	Perform **NEAT** for the k^{th} sub-problem
6	Save the best individual from **5** as $Best_k$
7	$PSA[k] \leftarrow Best_k$
8	**End While**
9	$k \leftarrow k + 1$

10 **End For**

C. Types of Algorithm Comparisons and Validation

For each MaO-SBP, an optimization set of samples is randomly created from the participating functions (as given in Table II). For each MaO-SBP, NEWS/D and S-NEAT are used to optimize a population of ANNs according to:

$$\min_{ANN \in X} \boldsymbol{Cost}\,(ANN) \tag{5}$$

where $\boldsymbol{Cost} \in \mathbb{R}^{\boldsymbol{M}}$, X is the feasible set of the considered ANNs and the j^{th} component of this cost vector is calculated as follows:

$$Cost_j = \frac{1}{2N_s}\sum_{n=1}^{N_s} {e_j}_n^2 \tag{6}$$

where n is the sample index and N_s is the number of samples in the optimization set.

The study in this paper aims to determine if NEWS/D can find a better approximation of the Pareto front as compared with S-NEAT. For each MaO-SBP, four types of comparisons are carried out. The first two types, A and B, are based on errors which are calculated using the optimization-set, whereas, for the other two types of comparisons, C and D, the errors are calculated using random samples that are not included in the optimization-set, which is referred to as the testing-set. It is further emphasized that, while using the term testing-set, the procedure is not about generalization, but rather about optimization.

For each of the optimization-set and the testing-set, a reference front is generated by collecting all non-dominated solutions from all the runs of both algorithms. For comparisons A and C, the reference front is used to compare the algorithms based on the Inverted-Generational-Distance (IGD) [12]. Then, for comparisons B and D, the algorithms are compared based on the availability of the ideal points of the MaO-SBPs. For these optimization problems, the true ideal point is at the origin of the objective space (as seen from equation 6). For comparisons B and D, the following Ideal-Point Measure (IPM) is used:

$$IPM = \frac{1}{M}\sum_{i=1}^{M} z_i \tag{7}$$

where z_i is the normalized value of the i^{th} component of the ideal-point, as obtained by running the algorithm.

IV. Experimental Study

A. Experimental Setup

Table III summarizes the parameters that were applied in running NEWS/D. In the current study, the stopping criterion of NEWS/D is the total number of generations, which is denoted as gen_{max}. For each MaO-SBP, 6000 vectors were randomly generated for the input vectors ($\boldsymbol{In}$). These vectors were divided into two sets, 5000 for the optimization set and 1000 for the testing set.

TABLE III: NEWS/D Run-Parameters

Parameter	Value
Num. of Generations gen_{max}	$40M$
Population size N	$40(M-1)$
Number of sub-problems K	$\lceil N/1.5 \rceil$
Number of protected generations gen_{TM}	$\lceil gen_{max}/20 \rceil$
Maximum number of topologies MAX_{NT}	$\lceil N/10 \rceil$
Topology crossover probability	0.8
Add node probability	0.2
Add connection probability	0.1
Weight Mutation (WM) mechanism	Polynomial
WM probability, distribution index	(0.2,15)
Weight Crossover (WC) mechanism	SBX
WC probability, distribution index	(0.8,20)

Table IV summarizes the parameters that are taken for the S-NEAT algorithm (see Section II.B) per each sub-problem. As evident from Tables III and IV, the compared algorithms run for the same number of evaluations, which makes the comparisons fair.

TABLE IV: S-NEAT Run Parameters

Parameter	Value
Num. of Generations	$\lceil gen_{max}/\sqrt{K} \rceil$
Population size	$\lceil N/\sqrt{K} \rceil$
Crossover probability	0.2
Add node probability	0.2
Add connection probability	0.7
Weight mutation probability	0.8
Gene reenable probability	0.25
Selection pressure (ranking and stochastic universal sampling)	1.5
Speciation Parameters	$C_1 = 1, C_2 = 1$ $, C_3 = 0.4$
Speciation Threshold	3

Given the stochastic nature of evolutionary algorithms, 31 runs were conducted for each MaO-SBP to obtain statistical data. For the statistical inference, the Wilcoxon rank-sum test with a significance level of 5% is used for all comparisons.

B. Experimental Results

This section presents the numerical results of the study. Table V provides the results of comparisons that belong to type B. It lists the medians, and the standard deviations of the

IPM, as obtained by the algorithms using the optimization samples.

TABLE V: STATISTICAL COMPARISON – IPM, OPTIMIZATION-SET

Case	NEWS/D		S-NEAT		p-value
	Median	STD	Median	STD	
1	**2.42e-2**	4.18e-3	3.42e-2	2.75e-3	4.0e-11
2	**2.73e-2**	3.74e-3	4.17e-2	3.31e-3	1.5e-11
3	**2.92e-2**	3.98e-3	4.47e-2	4.19e-3	3.0e-11
4	**2.66e-2**	4.45e-3	4.07e-2	3.62e-3	2.2e-11
5	**2.96e-2**	5.10e-3	4.73e-2	3.86e-3	2.5e-11
6	**2.56e-2**	3.62e-3	4.28e-2	3.12e-3	1.4e-11
7	**3.75e-2**	3.98e-3	5.59e-2	4.18e-3	1.5e-11
8	**4.18e-2**	4.56e-3	5.90e-2	3.41e-3	1.5e-11
9	**4.16e-2**	4.87e-3	6.07e-2	5.13e-3	6.4e-11
10	**6.79e-2**	3.79e-3	8.33e-2	6.42e-3	7.8e-11

The obtained low error values confirm that both algorithms are suitable for the tested MaO-SBPs. However, the p-values in Table V indicate that NEWS/D generates a closer ideal point to the true one.

TABLE VI: STATISTICAL COMPARISON – IPM, TESTING-SET

Case	NEWS/D		S-NEAT		p-value
	Median	STD	Median	STD	
1	2.35e-4	1.01e-5	2.35e-4	1.35e-5	4.2e-01
2	4.60e-4	6.96e-6	4.59e-4	1.66e-5	2.2e-01
3	**3.75e-4**	2.52e-5	3.87e-4	4.62e-5	2.8e-02
4	**3.70e-4**	2.37e-5	3.97e-4	5.75e-5	1.0e-03
5	**3.63e-4**	2.63e-5	3.77e-4	3.80e-5	7.1e-03
6	4.54e-4	5.06e-6	4.58e-4	8.96e-6	1.3e-01
7	3.67e-4	1.53e-5	3.77e-4	2.33e-5	1.6e-01
8	4.64e-4	1.10e-5	4.64e-4	8.58e-6	8.3e-02
9	3.72e-4	1.66e-5	3.67e-4	2.09e-5	9.7e-01
10	**5.89e-4**	1.49e-5	6.15e-4	2.46e-5	9.6e-07

TABLE VII: STATISTICAL COMPARISON – IGD, OPTIMIZATION-SET

Case	NEWS/D		S-NEAT		p-value
	Median	STD	Median	STD	
1	**4.99e03**	1.25e-3	8.84e-3	9.60e-4	3.3e-11
2	**8.02e-3**	5.42e-4	1.32e-2	6.75e-4	1.4e-11
3	**9.32e-3**	6.19e-4	1.56e-2	1.04e-3	1.4e-11
4	**6.78e-3**	6.38e-4	1.24e-2	7.27e-4	1.4e-11
5	**6.94e-3**	7.66e-4	1.30e-2	6.85e-4	1.4e-11
6	**1.22e-2**	3.41e-4	1.98e-2	6.76e-4	1.4e-11
7	**1.19e-2**	3.27e-4	1.86e-2	5.29e-4	1.4e-11
8	**2.26e-2**	4.14e-4	3.10e-2	8.28e-4	1.4e-11
9	**2.15e-2**	1.27e-2	2.85e-2	5.14e-4	2.5e-10
10	**2.82e-2**	6.58e-4	4.15e-2	8.34e-4	1.4e-11

Table VI presents the associated results for the testing samples, i.e., the comparisons of type D. Based on the results of the statistical comparisons, NEWS/D is either better or as-good-as S-NEAT.

Tables VII and VIII present summaries of the medians and the standard deviations of the IGD measure as obtained by each of the algorithms, i.e., for comparisons of types A and C, respectively. Based on the results of Tables VII and VIII, it is evident that NEWS/D generates better-approximated fronts.

TABLE VIII: STATISTICAL COMPARISON – IGD, TESTING-SET

Case	NEWS/D		S-NEAT		p-value
	Median	STD	Median	STD	
1	**1.44e-2**	2.39e-3	1.63e-2	2.63e-3	2.4e-05
2	**2.45e-2**	2.00e-3	3.51e-2	3.11e-3	1.7e-11
3	**2.64e-2**	1.63e-3	3.64e-2	3.38e-3	1.7e-11
4	**2.31e-2**	1.96e-3	2.91e-2	3.34e-3	7.1e-11
5	**2.68e-2**	2.60e-3	3.54e-2	3.15e-3	2.0e-11
6	**3.13e-2**	1.51e-3	4.08e-2	3.72e-3	1.4e-11
7	**3.49e-2**	1.45e-3	4.38e-2	1.94e-3	1.4e-11
8	**5.52e-2**	1.76e-3	6.71e-2	2.52e-3	1.4e-11
9	**5.35e-2**	2.53e-3	6.27e-2	2.76e-3	5.8e-11
10	**7.10e-2**	2.24e-3	8.32e-2	2.63e-3	1.5e-11

Fig 2 provides a parallel coordinate illustration [13], which presents the objective trade-offs of the reference front (see Section III.C), as obtained for the tenth MaO-SBP of Table II. This problem involves all seven functions. The y-axis is the normalized cost performance, and the x-axis is the objective indices. The figure includes the cost performances of each of the solutions, which are associated with the reference front, that has the minimal cost performance in one of the objectives. As evident from these plots, most of the objectives are conflicting.

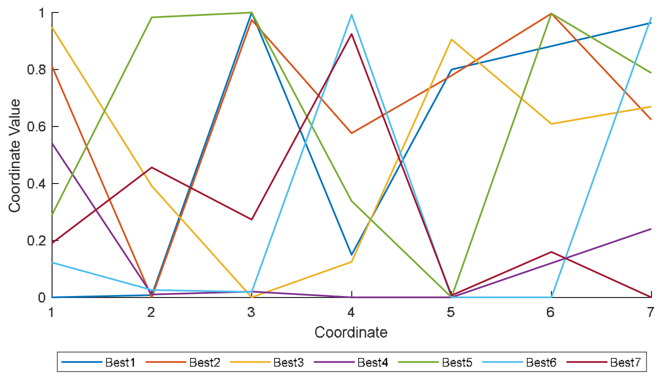

Fig. 2. Parallel coordinate representation MaO-SBP

V. Summary and Conclusions

This study deals with validation of the optimization capabilities of algorithms for Many-Objective Topology and Weight Evolution of Artificial Neural Networks (MaO-TWEANNs). First, Many-Objective Synthetic Benchmark Problems (MaO-SBPs) are introduced. Next, the proposed comparisons and validation methodologies are outlined. Finally, the proposed procedures and benchmark problems are applied to evaluate NEWS/D and compare this MaO-TWEANN algorithm with S-NEAT, which is a many-objective version of the single-objective well-known NEAT algorithm. This follows a previous study on NEWS/D that concentrated on its validation with respect to discrete outputs. The results of the current study show that the optimization capabilities of NEWS/D for MaO-TWEANN problems are not restricted to problems with discrete outputs, but that this algorithm is suitable also to solve problems with continuous outputs. The current validation study opens up the possibility to use NEWS/D for problems such as finding Pareto-optimal neuro-controllers that should cope with many objectives.

Future research may focus on using the proposed benchmarks validation procedures for the development of additional algorithms for MaO-TWEANNs.

Acknowledgment

The first author would like to acknowledge the generous scholarship that was provided by the Israeli Ministry of Science & Technology.

VI. References

[1] V. Trianni and M. López-Ibáñez, "Advantages of task-specific multi-objective optimisation in evolutionary robotics," *PLOS ONE,* vol. 10, no. 10, p. e0136406, 2015.

[2] Y. Jin, Multi-objective machine learning, vol. 16, Springer Science & Business Media, 2006.

[3] X. Zhou, A. Qin, M. Gong and K. C. Tan, "A survey on evolutionary construction of deep neural networks.," *IEEE Transactions on Evolutionary Computation,* 2021.

[4] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on evolutionary computation,* vol. 17, no. 6, pp. 712-731, 2007.

[5] K. O. Stanley and . R. Miikkulainen, "Evolving neural networks through augmenting topologies," *Evolutionary computation,* vol. 10, no. 2, pp. 99-127, 2002.

[6] A. Shenfield and S. Rostami, "Multi-objective evolution of artificial neural networks in multi-class medical diagnosis problems with class imbalance," in *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology*, 2017.

[7] O. Abramovich and A. Moshaiov, "Multi-objective topology and weight evolution of neuro-controllers," in *IEEE Congress on Evolutionary Computation (CEC)*, 2016.

[8] Showalter, I., Schwartz, H.M., "Neuromodulated multiobjective evolutionary neurocontrollers without speciation.," *Evol. Intel.,* 2020.

[9] S. Künzel and S. Meyer-Nieberg, "Evolving artificial neural networks for multi-objective tasks," in *Sim K., Kaufmann P. (eds) Applications of Evolutionary Computation. EvoApplications 2018. Lecture Notes in Computer Science, vol 10784.*, Springer, Cham, 2018.

[10] A. Salih and A. Moshaiov, "Modified decomposition framework and algorithm for many-objective topology and weight evolution of neural networks," in *2021 IEEE Congress on Evolutionary Computation, CEC 2021*, 2021.

[11] . J. Vesterstrom and R. Thomsen, "A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems," in *2004 congress on evolutionary computation*, 2004.

[12] C. A. C. Coello and C. C. Nareli , "Solving multiobjective optimization problems using an artificial immune system," *Genetic programming and evolvable machines,* vol. 6, no. 2, pp. 163-190, 2005.

[13] D. J. Walker, R. Everson and J. E. Fieldsend, "Visualizing mutually nondominating solution sets in many-objective optimization," *IEEE Transactions on Evolutionary Computation,* vol. 17, no. 2, pp. 165-184, 2012.