Proceedings of the 2013 Winter Simulation Conference
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds.

PARETO OPTIMIZATION AND TRADEOFF ANALYSIS APPLIED TO META-LEARNING OF MULTIPLE SIMULATION CRITERIA

Ofer M. Shir
Shahar Chen
David Amid
David Boaz
Ateret Anaby-Tavor

IBM Research
Haifa University Campus, Mt. Carmel, 31905, ISRAEL

Dmitry Moor

IBM Systems and Technology Group
Science and Technology Center, Moscow, RUSSIA

ABSTRACT

Simulation performance may be evaluated according to multiple quality measures that are in competition and their simultaneous consideration poses a conflict. In the current study we propose a practical framework for investigating such simulation performance criteria, exploring the inherent conflicts amongst them and identifying the best available tradeoffs, based upon multiobjective Pareto optimization. This approach necessitates the rigorous derivation of performance criteria to serve as objective functions and undergo vector optimization. We demonstrate the effectiveness of our proposed approach by applying it to a specific Artificial Neural Networks (ANN) simulation, with multiple stochastic quality measures. We formulate performance criteria of this use-case, pose an optimization problem, and solve it by means of a simulation-based Pareto approach. Upon attainment of the underlying Pareto Frontier, we analyze it and prescribe preference-dependent configurations for the optimal simulation training.

1 INTRODUCTION

Many simulations pose challenging computational tasks, but furthermore, simulations of an open-ended nature do not possess a known final solution. That is, the target of the learning/training is known, but its optimal or ultimate form is generally unknown. A common case is the family of ANN-based simulations. Generally speaking, the scope of the current study is simulations that perform learning tasks, and particularly ANN simulations that perform open-ended learning. From the Operational Research (OR) perspective, explicit objective functions for these simulations cannot be derived, but rather, quality measures may serve as reinforcement feedback. It is often the case that open-ended simulations possess more than a single quality attribute, as the evaluation of the learning success-rate may be subject to different perspectives. In these scenarios, these quality measures are often conflicting – i.e., making some progress in a given direction may be concordant with some quality measures and at the same time discordant with others. The current study is primarily targeted at treating the meta-learning of simulations with competing quality measures by proposing the Pareto multiobjective optimization framework for conflict exploration and

978-1-4799-2076-1/13/$31.00 ©2013 IEEE

tradeoff analysis. Practically speaking, the obtained Pareto Frontier (a rigorous definition will follow) relates to the best attainable parameter configurations for the learning outcome of the highest quality. Our proposed approach to treat such quality measures/attributes accounts for white- and black-box scenarios, and utilizes accordingly simulation-based optimization. In order to demonstrate our approach we consider, as a case-study, an ANN-simulation-based visualization technique, whose primary learning problem is mapping high-dimensional datasets onto 2-dimensional projections. Explicitly, we investigate the quality measures of Self-Organizing Maps for Multi-Objective Pareto Frontiers (SOMMOS: Chen et al. 2013), a semantically- and algorithmically-enhanced variant of the Self-Organizing Map algorithm (SOM: Kohonen 2001). The issue of identifying and analyzing performance criteria of computational tasks is raised in several papers. Caruana and Niculescu-Mizil (2004) described several criteria that can be used to evaluate the performance of supervised learning and introduced a new metric that combined several existing metrics into a single one. There are also multiple approaches used in the literature that aim at automated algorithms tuning, i.e., selection of some defining parameters that would yield the best-performing algorithm for solving a particular problem with respect to any performance criterion. For instance, investigating performance of Evolutionary Algorithms for solving OR problems is presented by Oltean (2005). Moreover, Kadioglu et al. (2010) and Hutter et al. (2009) describe some methods for automated algorithm configuration. In a study by Pölzlbauer (2004), some existing quality measures for Self-Organizing Maps are described. More quality measures for SOM may be found in a study by Kaski and Lagus (1996). In contrast to these works, our approach is aimed not only at identifying the set of parameters that would yield the best-performing algorithm with respect to any single criterion (e.g., computation time) but rather at obtaining the set of the most-preferred solutions with respect to multiple criteria. This added perspective would enable to further investigate the nature of the derived performance criteria of our use-case and only then to select the final solution. The contribution of the current study lies in the following:

- We propose a novel OR perspective on open-ended simulation challenges, which typically do not possess an objective function to undergo traditional optimization.
- We explicitly derive analytical performance criteria for a specific use-case from the domain of neural simulations, and illustrate the effectiveness of our proposed framework.
- We demonstrate how analyzing multiobjective optimization results has the potential to cater a better understanding of the simulation challenge, and at the same time to achieve fine-tuning of the primary simulation parameters.

2 META-LEARNING OF MULTIPLE SIMULATION CRITERIA

Here, we are concerned with the complex problem of meta-learning in simulations, possessing multiple quality measures, that upon training are to serve as players in decision making challenges. We consider it meta-learning because the simulations themselves are concerned with learning tasks. Each raw quality measure of a simulation outcome constitutes the performance criterion of the learning problem associated with the simulation, which is likely to be affected by a set of possibly overlapping learning parameters. Consequently, any attempt to manually set the parameters according to a specific performance criterion is likely to degrade the other criteria. To address this problem, we propose to apply simulation-based multiobjective optimization, where each performance criterion is considered as an individual objective function and the various learning parameters of the simulation are treated as decision variables as in conventional meta-learning frameworks. The result of the optimization is the entire spectrum of optimal performance criteria and their associated parameter configurations, constituting the Pareto Frontier and the Pareto Optimal Set, respectively. Figure 1 summarizes the proposed approach. In general, the authors acknowledge that the proposed approach and its aftermath are dependent upon the selection as well as the explicit definition of the actual performance criteria.

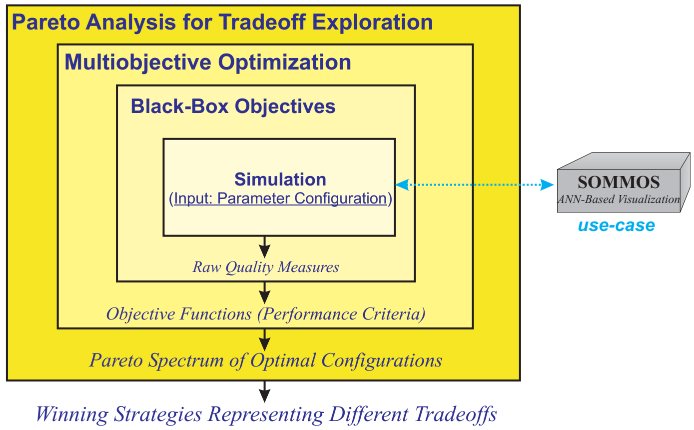

Figure 1: Five layers of the meta-learning framework, subject to a pipeline flow, and the targeted use-case.

2.1 Pareto Multiobjective Optimization

Multiobjective optimization aims at simultaneously treating a number of conflicting objectives, and thereby revealing the entire trade-off surface amongst the objectives. This framework is in contrast to traditional optimization approaches that consider multiobjective problems by posing a weighted sum of its objectives and employ singleobjective solvers to obtain an individual solution point. These traditional methods do not provide a complete view of the conflict amongst the objectives (they do not reveal the Pareto Frontier), and moreover, necessitate *a priori* prescription of the objectives weighing. Other mathematical limitations are described by Das and Dennis (1997). Let a vector of objective functions in $\mathbb{R}^m$, $\vec{f}(\vec{x}) = (f_1(\vec{x}), \ldots, f_m(\vec{x}))^T$, be subject to *minimization*, and let a partial order be defined in the following manner: given any $\vec{f}^{(1)} \in \mathbb{R}^m$ and $\vec{f}^{(2)} \in \mathbb{R}^m$, we state that $\vec{f}^{(1)}$ strictly **Pareto** dominates $\vec{f}^{(2)}$, denoted $\vec{f}^{(1)} \prec \vec{f}^{(2)}$, if and only if $\forall i \in \{1, \ldots m\} : f_i^{(1)} \leq f_i^{(2)} \wedge \exists i \in \{1, \ldots, m\} : f_i^{(1)} < f_i^{(2)}$. In addition to the strict domination $\prec$, we define $\vec{f}^{(1)} \preceq \vec{f}^{(2)} \iff \vec{f}^{(1)} \prec \vec{f}^{(2)} \vee \vec{f}^{(1)} = \vec{f}^{(2)}$. The individual Pareto-ranking of a given candidate solution is defined as the number of other solutions dominating it. The crucial claim is that for any compact subset of $\mathbb{R}^m$, there exists a non-empty set of minimal elements with respect to the partial order $\preceq$ (see, e.g., Ehrgott 2005). **Non-dominated points** are then defined as the set of minimal elements with respect to the partial order $\preceq$, and by definition their Pareto-ranking is zero. The goal of Pareto optimization is thus to obtain the *non-dominated set* and its pre-image in the design space, the so-called *Pareto optimal set*, also referred to as the *efficient set*. Finally, the Pareto Frontier is defined as the set of all points in the objective space that correspond to the solutions in the Pareto-optimal set. The computational problem of attaining the Pareto Frontier of a multiobjective optimization problem (Papadimitriou and Yannakakis 2000) can be either treated by means of algorithms utilizing mathematical programming solvers – for instance, the so-called Diversity Maximization Approach by Masin and Bukchin (2008), employing, e.g., CPLEX (2009) – or approximated by population-based heuristics (see, e.g., Beume, Naujoks, and Emmerich 2007).

2.2 Simulation-Based and Black-Box Optimization

Unlike traditional Operational Research modeling, which targets the explicit analytical forms of the objective functions and their associated constraints, modern treatment of real-world models may consider them as white- or black-boxes. Moreover, the advent of existing heuristics allows efficient optimization of such black-box models. In essence, simulations may also be treated as black-boxes and optimized accordingly, i.e., their characteristic parameters may be effectively tuned to yield the optimal behavior. Here, due to the nature of the simulated performance criteria, we choose to realize this option with Evolutionary Multiobjective Algorithms, as will be specified.

2.3 Pareto Analysis and Tradeoff Exploration

The optimization process outputs the Pareto Frontier and the Pareto Set, representing the optimal parameter configurations for the simulation. The solution points on the Pareto Frontier are mathematically indifferent with respect to each other, and thus the selection phase, entitled Multi-Criterion Decision Making (MCDM: Köksalan, Wallenius, and Zionts 2011), is subjectively driven by the human decision maker. This process involves exploration of the Frontier, and eventually, the challenge in selecting a solution is to account for gains and losses while adhering to personal preferences. Various supporting methodologies do exist – e.g., consider AHP (Saaty 1980) or ELECTRE (Roy 1991) – yet, in the current study we shall demonstrate the effectiveness of manual exploration and selection from the attained Frontier.

3 SOMMOS AS A SIMULATOR

We choose to specifically address an ANN simulation problem from the realm of visualization, namely, the generation of mappings from high-dimensional datasets onto 2-dimensional projections, which relies on unsupervised neural competitive learning. The defining parameters of the neural learning are to constitute the decision variables of the optimization process, whose objective functions are to be derived from the visualization quality measures of accuracy, semantic orientation, etc. The ultimate goal would be to select specific configurations from the attained Pareto Frontier of the meta-learning problem in order to simulate/generate high-quality maps. In practice, the meta-learning problem we investigate is regarding the quality measures of SOMMOS (Chen et al. 2013), a semantically- and algorithmically-enhanced variant of SOM (Kohonen 2001). These specific maps visualize Pareto Frontiers of given MCDM problems, and thus, they are to play a role in supporting the human decision making process. The careful reader should distinguish between the 2 different utilizations of Pareto Frontiers at 2 layers of the current work: one, as input to SOMMOS, representing an external decision problem, and the other, output of the general proposed framework. The current problem is especially important to the OR community since it describes a novel visualization for the display of Pareto Frontiers. The means to solve this problem, as will be shown here, is simulation-based multiobjective optimization.

3.1 SOMMOS: Technique's Description

The SOM algorithm constitutes a popular method for visualization of large high-dimensional datasets. By means of unsupervised competitive learning, a network of prototype vectors is formed, and upon employing a 2-dimensional regular *rectangular* or *hexagonal* network model, visualization is directly attained as data items can be mapped according to the simulated spatial structure (Vesanto 1999). The actual training phase constitutes the percolation of dataset values to the neural network based upon its topology. Due to approximate topology-preserving properties of this method, the resulting visual displays can be intuitively interpreted by the user. From its origin, SOM inventors concluded that the learning process is open in nature and cannot be validated. Optimization-wise, SOM has been practically addressed as a learning problem with a primary objective of minimizing its quantization error (Kohonen 2001). In this paper, we examine SOMMOS (Chen et al. 2013), an enhanced SOM variant especially designed for visualizing Pareto Frontiers to facilitate effective multi-criterion decision making. Given an m-dimensional Pareto Frontier, SOMMOS generates a symmetric *m-gon* whose vertices represent the optima of the m objective functions and participate in the training phase of the dataset. The anchoring of these extreme solution points necessitated algorithmic enhancement to the original SOM, as outlined in what follows with reference to the pseudo-code presented in Figure 2[LEFT]. Data is first to be normalized within $[0,1]^m$ by considering minimally- and maximally-attainable objective values (the invoked routine is entitled `normalizeFrontier()`). Let the map be denoted by $\mathscr{M}$. During the training phase, every so-called anchoring-epoch, T_{anchor}, a given vertex-neuron, $\vec{a}_\ell \in \{\vec{a}_\ell\}_{\ell=1}^m \subset \mathscr{M}$, percolates its value to the surrounding neurons based on the learning functions. Consequently, the closer a neuron is to a vertex, the stronger it learns the appropriate anchor value. However, as in each update-step data points keep percolating their values, the anchors should also `learnSOMMOS()`

input Dataset: Pareto Frontier $\mathcal{F} = \left\{\vec{f}_k\right\}_{k=1}^{|\mathcal{F}|}, \ \vec{f}_k \in \mathbb{R}^m$

`normalizeFrontier()`
`initSOMMOS()`
for $t = 0 \dots T_{\text{period}}$ **do**
  $\mathcal{S} \longleftarrow$`randomSampling`$(\mathcal{F})$
  `trainSOMMOS`$(\mathcal{M}, \mathcal{S}, h_F(), \alpha_F(t))$
  if $(\bmod(t, T_{\text{anchor}}) == 0)$ **then**
    $\{\vec{a}_\ell\}_{\ell=1}^m \subset \mathcal{M} : \ \vec{a}_\ell \longleftarrow \mathcal{I}_{:,\ell}$ {anchoring: explicit fixing at the vertices}
    `trainSOMMOS`$(\mathcal{M}, \mathcal{A}, h_A(), \alpha_A(t))$
  end if
end for
return $\mathcal{M}$

`initSOMMOS()`

set map size: N
set **anchoring dataset** by means of the $m \times m$ identity $\mathcal{I}$: $\mathcal{A} = \{\mathcal{I}_{:,\ell}\}_{\ell=1}^m$
generate an m-gon layout as the neural map: $\mathcal{M} = \left\{\vec{u}^{(j)}\right\}_{j=1}^N, \quad \vec{u}^{(j)} \in \mathbb{R}^m$
randomly initialize neurons: $\vec{u}^{(j)} \longleftarrow \mathcal{U}(0,1), \quad j = 1 \dots N$
identify and reset the m-gon's **vertices**: $\{\vec{a}_\ell\}_{\ell=1}^m \subset \mathcal{M} : \ \vec{a}_\ell \longleftarrow \mathcal{I}_{:,\ell}$
initialize learning-rate functions: $\alpha_F(t), \ \alpha_A(t)$
initialize neighbourhood kernels: $h_F(), \ h_A()$
set training period and anchoring epoch: $T_{\text{period}}, \ T_{\text{anchor}}$

`trainSOMMOS`(map $\mathcal{M}$, data $\mathcal{X}$, kernel $h()$, learning-rate α)

for each point in the dataset $\vec{x}_k \in \mathcal{X}$ **do**
  for each neuron $\vec{u}^{(j)} \in \mathcal{M}$ **do**
    $\vec{u}^{(j)} \longleftarrow \vec{u}^{(j)} + \alpha \cdot h(\vec{x}_k, \mathcal{M}) \cdot \left(\vec{x}_k - \vec{u}^{(j)}\right)$
  end for
end for

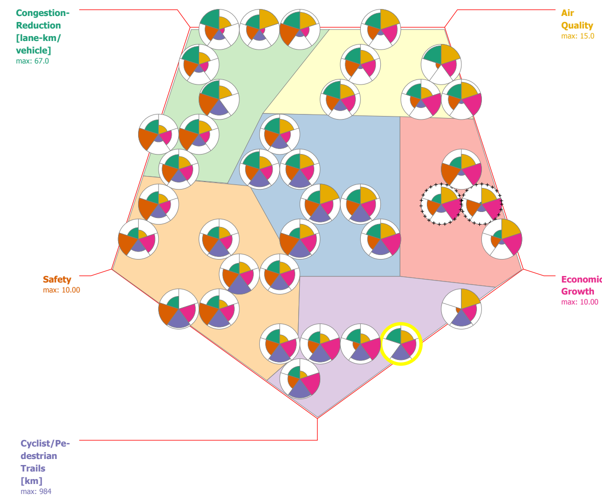

Figure 2: **[LEFT]** Pseudo-code summarizing SOMMOS. **[RIGHT]** A 5-dimensional Pareto Frontier of a Transportation Asset Management problem, visualized by SOMMOS. Each anchor represents an objective. Each Polar Area Bar chart represents a solution. The sector filling represents the values of that solution per a specific objective. K-Means (Lloyd 1982) was applied to cluster the solutions, reflected as the background color. It is evident that due to the anchor training, solutions with high values per the corresponding objective will be closer to that anchor.

repeatedly train the ANN as an additional independent dataset $\mathscr{A}$, which exclusively represents the vertices; the actual vertex-units are also explicitly reset. Essentially, the learning procedure comprises two competing processes with different learning-rates. While the learning-rate function for the input dataset $\mathscr{F}$, noted as $\alpha_F(t)$, declines over time towards convergence, the anchoring learning-rate function, $\alpha_A(t)$, linearly declines from 1.0 to $\alpha_F(t)$ for neurons up to a radius r_A from the vertex, and otherwise it reads $\alpha_F(t)$ for neurons positioned farther. The overall learning of the m-gon ANN is summarized as `learnSOMMOS()` in Figure 2[LEFT], where the Pareto Frontier dataset is normalized and the anchoring is carried out using the identity matrix. A SOMMOS map for a 5-dimensional problem is presented in Figure 2[RIGHT].

3.2 SOMMOS Performance Criteria: Rigorous Formulation

As suggested in Section 2.2, fine-tuning a simulation would require careful evaluation of its quality metrics. For readability and clarity, this section will formulate the quality metrics in our case study – SOMMOS. In order to evaluate the quality of SOMMOS generation, we would like to rigorously quantify the targeted measures of interest and propose respective analysis criteria. We consider four performance criteria, two of which have been widely utilized to evaluate SOM resolution and topology preservation properties, while the other two are introduced here particularly for visualizing Pareto Frontier datasets. These additional criteria stem in the semantic enhancement to SOM that resulted in SOMMOS. This section will rigorously define each of these four performance criteria. The input dataset constitutes a Pareto Frontier $\mathscr{F}$ involving m objectives. Given an input data vector $\vec{f}_k \in \mathscr{F} \subseteq \mathbb{R}^m$, its best-matching unit (BMU) in the map $\vec{u}_{1:c}^{(k)} \in \mathscr{M} \subseteq \mathbb{R}^m$ is the unit that minimizes the Euclidean distance $\|\vec{f}_k - \vec{u}_i\|$: $c^{(k)} = \arg\min_{i \in \mathscr{M}} \left\{ \|\vec{f}_k - \vec{u}_i\| \right\}$. The second-best-matching unit, denoted as $\vec{u}_{2:c}^{(k)}$, is defined respectively. We then consider the so-called *mean quantization error*, representing the learning quality of the attained ANN, which is defined as the mean over $\|\vec{f}_k - \vec{u}_{1:c}^{(k)}\|$ (Kohonen et al. 1996):

$$QE = \mathbf{E}_{j \in \mathscr{F}} \left[\|\vec{f}_j - \vec{u}_{1:c}^{(j)}\| \right] \tag{1}$$

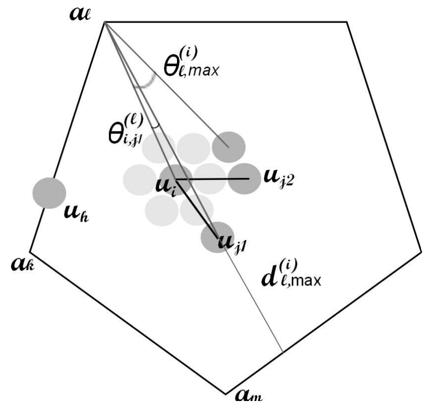

Figure 3: Illustration of the local orientation error computation, representing the ℓ^{th} objective values by a gray-scale subject to maximization.

Let $\delta()$ denote the adjacency indicator of the ANN, i.e., $\delta(\vec{u}_i, \vec{u}_j) = \begin{cases} 1 & \text{if } \vec{u}_i, \vec{u}_j \text{ are adjacent ANN units} \\ 0 & \text{otherwise} \end{cases}$.
Next, we consider the *topographic error*, defined as the proportion of the input data vectors whose first-best and second-best matching units are not adjacent within the ANN:

$$TE = \frac{\sum_{j \in \mathscr{F}} \left[1 - \delta\left(\vec{u}^{(j)}_{1:c}, \vec{u}^{(j)}_{2:c}\right)\right]}{|\mathscr{F}|} \tag{2}$$

We now define the so-called measure of *local orientation error*, which considers each of the anchor units and accounts for ranked-based ordering violations with respect to it. In other words, local orientation violation with regard to the ℓ^{th} vertex (anchor) unit, $\vec{a}_\ell \in \mathscr{M}$, occurs if a unit $\vec{u}_i$ is placed closer to $\vec{a}_\ell$ than $\vec{u}_j$, but its ℓ^{th} objective function is inferior. By constructing this measure, we would also like to apply larger error values to units that are closer to the anchor, since these neurons are the foremost interesting when concentrating on the corresponding objective function during the decision making process. Generally speaking, by minimizing this error measure we would like to enable exploration phases as typically encountered in human decision making: following an examination of a certain solution point on the Frontier, the decision maker may be interested in rovering around its neighborhood, under the assumption that the local orientation is concordant with the objective functions. This implies that the local orientation error should read larger values for violations occurring between neighboring neurons in comparison to violations between neurons that are located far from each other. Moreover, when considering the relative positions of these two neurons, care should be given not only to the distance between them, but rather also to their orientation with respect to the anchor. See Figure 3 for an illustration: the error measure for the neurons $\langle \vec{u}_i, \vec{u}_{j1} \rangle$ is larger than the error measure of the neurons $\langle \vec{u}_i, \vec{u}_{j2} \rangle$ because of their different relative positions with respect to $\vec{a}_\ell$. Let $d^{(i)}_{\ell,\max}$ denote the maximally attainable distance within the ANN boundaries from the ℓ^{th} anchor in the direction of the vector $\vec{u}_i - \vec{a}_\ell$, and let $\theta^{(\ell)}_{i,j1}$ denote the angle between $\vec{u}_i$ and $\vec{u}_{j1}$. Given $\langle \vec{u}_i, \vec{u}_{j1} \rangle$, we calculate $\theta^{(i)}_{\ell,\max}$ as the maximal angle between $\vec{u}_i$ and all the neurons $\vec{u}_\wp$ that satisfy $(d(\vec{u}_\wp, \vec{a}_\ell) > d(\vec{u}_i, \vec{a}_\ell)) \wedge (d(\vec{u}_\wp, \vec{u}_i) \leq d(\vec{u}_i, \vec{u}_{j1}))$. The local orientation error is then defined as the summation over all local violations, V_ℓ, with respect to all anchors, weighted by the neighborhood kernel $h()$:

$$OE = \sum_{\ell=1}^{m} \sum_i \sum_j h\left(d(\vec{a}_\ell, \vec{u}_i)\right) \cdot V_\ell\left(\vec{u}_i, \vec{u}_j\right) \tag{3}$$

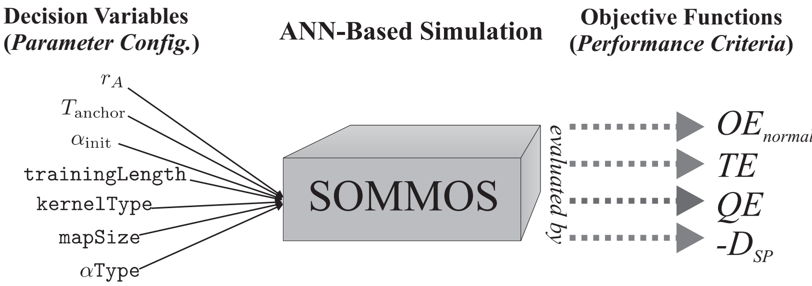

Figure 4: Model summary: the decision variables and their boundaries comprise the continuous variables $r_A \in [1,200]$ and $\alpha_{\text{init}} \in [0.01, 1.0]$, an integer variable $T_{\text{anchor}} \in \{5,6,\dots,1000\}$, and categorical variables `mapSize`, α`Type`, `kernelType`, `trainingLength` (see Kohonen 2001 and Chen et al. 2013).

We choose to normalize the local orientation error measure with respect to the equivalent measure of a random SOMMOS instance (i.e., whose neurons are randomly initialized and do not undergo learning), which is denoted by OE_{rand}:

$$OE_{\text{norm}} = \frac{OE}{OE_{\text{rand}}}, \tag{4}$$

Evidently, minimizing the local orientation error measure does not guarantee SOMMOS maps with proper global orientation characteristics. Consider, for instance, a large SOMMOS map where the units are grouped around its center, possessing good local orientation, but overall resulting in poor usability since the map is cluttered and the surface is poorly utilized. This scenario provides us with the motivation to introduce global orientation as another quality measure of SOMMOS. Such characteristics of global orientation may be accomplished by minimizing the local orientation error while simultaneously maximizing the global spatial diversity of all the units. Especially, we adopt the so-called Solow-Polasky diversity measure (see Solow and Polasky 1994 and Ulrich, Bader, and Thiele 2010). Given the pairwise distances between the units, $d(\vec{u}_i, \vec{u}_j)$, let $\Psi = (\psi_{ij}) \in \mathbb{R}^{|\mathscr{M}|\times|\mathscr{M}|}$ be constructed with matrix elements $\psi_{ij} = \exp(-\gamma \cdot d(\vec{u}_i, \vec{u}_j))$, then the Solow-Polasky measure is defined as:

$$D_{SP} = \vec{1}^T \Psi^{-1} \vec{1}, \tag{5}$$

i.e., the summation over all the elements of Ψ^{-1}; γ is a domain-specific normalization factor. The Solow-Polasky diversity measure, which originates in biology, strives to quantify the number of existing species within the given spatial domain. Thus, the larger this scalar, the more diverse the map is, and following the aforementioned rationale, better global orientation is obtained. Equivalently, minimizing $-D_{SP}$ will achieve the same goal.

4 EXPERIMENT DESIGN

In our ANN-based use-case, the simulation possesses multiple quality metrics, as described in Section 3.2, whose simultaneous optimization poses a quad-criteria problem formulated as follows:

$$\boxed{\begin{array}{l} f_1 = OE_{\text{norm}} \longrightarrow \min \\ f_2 = TE \longrightarrow \min \\ f_3 = QE \longrightarrow \min \\ f_4 = -D_{SP} \longrightarrow \min \end{array}} \tag{6}$$

We consider a 7-dimensional decision (design) space corresponding to the defining mechanism of SOMMOS, considering otherwise the conventional setting alternatives for SOM (Kohonen 2001). Figure 4 summarizes this model, and the reader is also referred to Chen et al. (2013) for more details.

4.1 Method: Evolutionary Multiobjective Algorithms

The method we employ to tackle the aforementioned simulation-based multiobjective optimization problem belongs to the family of Evolutionary Algorithms (EAs: Bäck 1996). EAs are powerful stochastic global search methods gleaned from the model of organic evolution, which have been for several decades successful in treating high-dimensional optimization problems. They especially excel in scenarios where quality evaluation provided by computer-based simulation constitutes the objective function, or in black-box evaluations, such as in *experimental optimization* (Shir et al. 2012). Their broad success in this domain is primarily attributed to two factors – first, the fact that they constitute direct search methods, i.e., do not require derivatives determination, and second, their inherent robustness to noise. In the current study we are especially interested in evolutionary *multiobjective* optimization algorithms (EMOA) – which have undergone considerable development in the last two decades (Zitzler, Deb, and Thiele 2000) – to constitute the optimizers in the current framework (Figure 1). In particular, we employ the SMS-EMOA heuristic (Beume, Naujoks, and Emmerich 2007) as the multiobjective procedure with a mixed-integer evolution strategy (MIES) as its solving engine. Our implementation and parameter settings followed Reehuis and Bäck (2010).

4.2 Pre-Experimental Planning

Both SOMMOS and the SMS-EMOA were implemented and run in MATLAB. SOMMOS is applied here to the visualization of a 5-dimensional Pareto Frontier from the domain of Transportation Asset Management (TAM). In TAM, when building a portfolio of projects/initiatives, taking cost-effective decisions regarding resource allocation in order to preserve, maintain, or improve transportation infrastructure (roads, bridges, or buildings) is crucial (see, e.g., AASHTO 2002). Here, the objectives were formulated as maximizing congestion reduction, maximizing pedestrian and cyclist trails, maximizing safety, maximizing economic growth, and maximizing air quality across the transportation network. A single SOMMOS generation approximately takes 1sec on a machine with Intel $i7$ CPU featuring 4 1.60 GHz processors. Due to the stochastic nature of SOMMOS, we invoked 20 simulations per evaluation of a candidate configuration, and applied averaging to yield the objective function values. Toward the end of solving the 4-objectives optimization problem of Eq. 6, here is our planned *modus operandi*: (1) individually applying the MIES solver to each one of the objective functions as 4 singleobjective independent optimization problems (employing a population of 15 parents and 30 offspring with non-elitist selection); (2) computing the Pareto Frontier of this problem (utilizing the steady-state SMS-EMOA with a population size of 50); (3) investigating the results obtained in those calculations.

5 EXPERIMENTATION AND RESULTS

We describe here our practical observations of the optimization methodology applied to Eq. 6.

5.1 Preliminary: Optimization Aftermath

We first consider the individual treatment of the 4 objective functions. By doing so we aim at understanding the nature of each particular objective function and at identifying some conflicts in the early stages of the analysis. It is evident that all four minimizers dramatically differ. For instance, we briefly compare the minimizers of OE and QE: in order to obtain small OE values, training neurons with values that boost the orientation quality of the ANN (i.e., anchoring) should be carried out more frequently (small T_{anchor} values). At the same time, such effective training may be achieved also when it is propagated to neurons located farther from the proximity of the anchor unit, e.g., by setting large r_A values and strengthening the anchor's impact on the ANN. QE, on the other hand, is concerned with the degree to which the dataset fits the trained neurons. Therefore, a frequent anchoring operation may compromise the desired degree of fit, since anchors constitute inorganic elements with respect to the dataset; minimizing QE should be intuitively

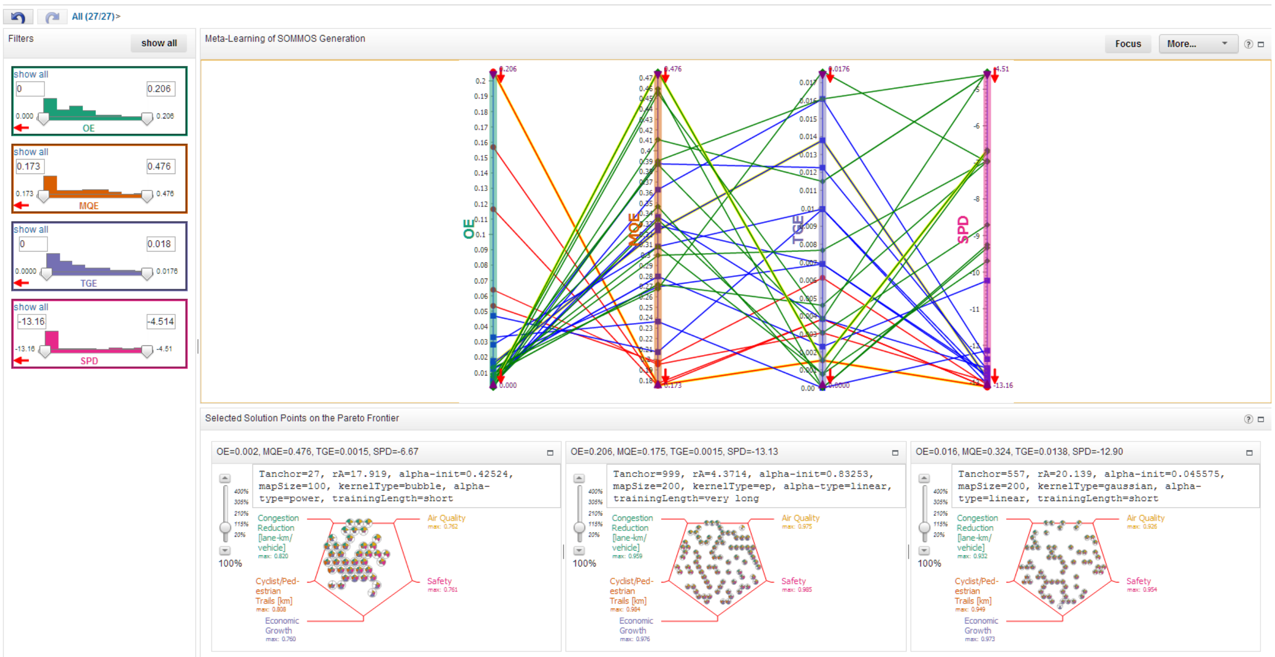

Figure 5: The Pareto Frontier for Eq. 6 is depicted in Parallel Coordinates at the top. Three specific solutions (highlighted in yellow at the top visualization) have their SOMMOS realizations depicted at the bottom. These realizations are prescribed by the Pareto Optimal Set information.

associated with suppressing the anchoring operation. We begin by investigating the depicted Frontier, as shown in the Parallel Coordinates visualization at the top of Figure 5. Each axis denotes a performance criterion and each line denotes an optimal solution for all four performance criteria. K-Means clustering (Lloyd 1982) was applied to this Frontier, and 3 clusters are colored accordingly. A birds-eye view at the Frontier reveals that OE is in a hard conflict with all other 3 objectives, which are in soft conflicts amongst themselves. This is mainly because training on the objectives' maximal values (i.e., the anchors) interferes with the training of the data itself. This interference is so severe that the known SOM conflict between QE and TE (Pölzlbauer 2004) appears as a soft conflict when compared to the conflict between OE and these regular SOM quality measurements. This view also contains 3 actual SOMMOS maps, each corresponding to a solution within one of the clusters on the Frontier; these were generated by prescribing the optimal decision variables provided by the optimization process (Pareto set) into the SOMMOS algorithm. The cluster colored in green corresponds to a subset of maps with low OE but low diversity and high QE (its representative SOMMOS map is depicted in the left). When analyzing the generated maps, we found that the high QE values reflect a misrepresentation of the data. Also, having a low diversity hinders the decision maker's ability to spot an area of particular interest. On the positive side, it is very easy for a user to explore the space since there is a clear sense of direction, i.e., it is easy to associate the solutions with a direction towards an objective. The next cluster is colored in red (its representative SOMMOS map is depicted in the middle). Unlike the green cluster, the current subset of maps is characterized by very small values of QE, highly diverse maps, somewhat low TE but high values of OE. These are maps that accurately represent the data and are well-diverse. However, the high values of OE are reflected in the inability to practically rover from one solution to another when exploring the Frontier. The last cluster, colored in blue, represents solutions of *"meeting-in-the-middle"* in terms of OE, QE as well as diversity (its representative SOMMOS map is depicted in the right). Evidently, maps in this cluster possess high diversity and fine OE values. However, some of the solutions in this cluster, specifically those with high QE, constitute problematic representations of the data.

5.2 Analysis and Discussion

In what follows, we shall demonstrate the final step in the proposed framework – how analyzing multiobjective optimization results may cater a better understanding of the learning challenge, and at the same time achieve fine-tuning of the primary simulation parameters. While the tradeoff analysis contributes to our understanding of the learning performance criteria, the end goal is to determine which solutions on the Frontier constitute preferable balance joints amongst these performance criteria. This final selection phase is subjective, and in what follows we demonstrate a decision-making flow that adheres to a certain set of arguments regarding the aforementioned SOMMOS example. The following preferences are considered in this example. First, a SOMMOS map should possess local orientation error as small as possible. This prerequisite is essential for any user that utilizes the map, since maps possessing high local orientation error may hinder the exploration process. However, taking into account only this first argument will potentially lead to maps with the maximally attainable quantization error within the cluster (due to the inherent conflict) – which is impractical to utilize, since each neuron is likely to be mapped onto many data points on the Frontier. Therefore, we decide to compare maps generated with parameters that correspond to low orientation error and medium-to-low quantization error. It was apparent that solutions within the green cluster are inadequate, leading to an interesting deduction that the diversity criterion does not play a significant role in this selection process as all maps in the blue and red clusters contain high diversity. Upon filtering out all the solutions from the green cluster, it becomes evident that the low orientation error is represented by the blue cluster and the low quantization error is represented by the red cluster. We further narrow down our focus on the solutions that are in the joint between those clusters. Here, given two solutions with similar degrees of quantization error and local orientation error, we would like to account for both the topographic error and the diversity measure – and advise to select the solution point with better values in the latter pair of criteria. Overall, following the insights gained from the tradeoff analysis, and given subjective preferences of the decision maker that we prescribed for this example, a procedure to select an optimal parameter configuration has been described. It should be noted that an additional subjective criterion upon which the decision maker is likely to select a solution from the Frontier would be the *visual aesthetic* aspect of the SOMMOS map, an aspect which is left for future work.

6 CONCLUSIONS

In this paper we proposed the employment of simulation-based multiobjective optimization as a computational framework to address conflicts amongst multiple performance criteria of simulations (Figure 1). In order to illustrate the effectiveness of this framework, we analyzed a specific ANN-simulation-based visualization technique with multiple competing performance criteria, that does not possess a known best outcome. We formulated a quad-criteria optimization problem and employed multiobjective solvers to attain its Pareto Frontier. We further showed how to explore the available tradeoffs amongst the prescribed performance criteria and how to gain insights concerning the inherent conflicts. Evidently, this approach also allowed us to locate desirable areas within the Pareto Frontier that are more likely to meet expectations regarding the simulation task. Altogether, our mechanism for analyzing multiobjective optimization results led to a better understanding of the simulation challenge and fine-tuned the primary simulation parameters.

REFERENCES

AASHTO 2002. “American Association of State Highway and Transportation Officials: Transportation Asset Management Guide”. http://downloads.transportation.org/AMGuide.pdf.

Bäck, T. 1996. *Evolutionary Algorithms in Theory and Practice*. New York, NY, USA: Oxford University Press.

Beume, N., B. Naujoks, and M. Emmerich. 2007. “SMS-EMOA: Multiobjective Selection based on Dominated Hypervolume”. *European Journal of Operational Research* 181 (3): 1653–1669.

Caruana, R., and A. Niculescu-Mizil. 2004. “Data mining in metric space: an empirical analysis of supervised learning performance criteria”. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD ’04, 69–78. New York, NY, USA: ACM.

Chen, S., D. Amid, O. M. Shir, D. Boaz, T. Schreck, L. Limonad, and A. Anaby-Tavor. 2013. “Self-Organizing Maps for Multi-Objective Pareto Frontiers”. In *Proceedings of IEEE Pacific Visualization*: IEEE Press.

CPLEX 2009. “IBM-ILOG”. www.ibm.com/software/integration/optimization/cplex-optimizer/.

Das, I., and J. Dennis. 1997. “A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multucriteria optimization problems”. *Structural Optimization* 14:63–69.

Ehrgott, M. 2005. *Multicriteria Optimization*. second ed. Berlin: Springer.

Hutter, F., H. H. Hoos, K. Leyton-Brown, and T. Sttzle. 2009. “ParamILS: An Automatic Algorithm Configuration Framework.”. *J. Artif. Intell. Res. (JAIR)* 36:267–306.

Kadioglu, S., Y. Malitsky, M. Sellmann, and K. Tierney. 2010. “ISAC - Instance-Specific Algorithm Configuration.”. In *ECAI*, edited by H. Coelho, R. Studer, and M. Wooldridge, Volume 215 of *Frontiers in Artificial Intelligence and Applications*, 751–756: IOS Press.

Kaski, S., and K. Lagus. 1996. “Comparing Self-Organizing Maps”. In *Proceedings of the Sixth International Conference on Artificial Neural Networks: ICANN-96*, edited by C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, and B. Sendhoff, Lecture Notes in Computer Science 1112, 806–814. Berlin; New York: Springer-Verlag.

Kohonen, T. 2001. *Self-Organizing Maps*. 3rd ed. Springer.

Kohonen, T., J. Hynninen, J. Kangas, and J. Laaksonen. 1996. “SOM PAK: The Self-Organizing Map Program Package”. Technical Report A31, Helsinki University of Technology.

Köksalan, M., J. Wallenius, and S. Zionts. 2011. *Multiple Criteria Decision Making: From Early History to the 21st Century*. World Scientific.

Lloyd, S. 1982, mar. “Least squares quantization in PCM”. *Information Theory, IEEE Transactions on* 28 (2): 129 – 137.

Masin, M., and Y. Bukchin. 2008. “Diversity maximization approach for multiobjective optimization”. *Operations Research* 56 (2): 411–424.

Oltean, M. 2005, Fall. “Evolving Evolutionary Algorithms Using Linear Genetic Programming”. *Evolutionary Computation* 13 (3): 387–410.

Papadimitriou, C. H., and M. Yannakakis. 2000. “On the Approximability of Trade-offs and Optimal Access of Web Sources”. In *FOCS*, 86–92.

Pölzlbauer, G. 2004, June 24–27. “Survey and Comparison of Quality Measures for Self-Organizing Maps”. In *Proceedings of the Fifth Workshop on Data Analysis (WDA’04)*, edited by J. Paralič, G. Pölzlbauer, and A. Rauber, 67–82. Sliezsky dom, Vysoké Tatry, Slovakia: Elfa Academic Press.

Pölzlbauer, G. 2004. “Survey and comparison of quality measures for self-organizing maps”. In *Fifth Workshop on Data Analysis (WDA 2004)*, 67–82.

Reehuis, E., and T. Bäck. 2010. “Mixed-integer evolution strategy using multiobjective selection applied to warehouse design optimization”. In *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, GECCO ’10, 1187–1194. New York, NY, USA: ACM.

Roy, B. 1991. “The Outranking Approach and the Foundations of ELECTRE Methods”. *Theory and Decision* 31:49–73.

Saaty, T. L. 1980. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. McGraw-Hill.

Shir, O., J. Roslund, Z. Leghtas, and H. Rabitz. 2012. “Quantum control experiments as a testbed for evolutionary multi-objective algorithms”. *Genetic Programming and Evolvable Machines* 13:445–491.

Solow, A., and S. Polasky. 1994. “Measuring biological diversity”. *Environmental and Ecological Statistics* 1:95–103.

Ulrich, T., J. Bader, and L. Thiele. 2010. "Defining and Optimizing Indicator-Based Diversity Measures in Multiobjective Search". In *PPSN-XI*, 707–717.

Vesanto, J. 1999. "SOM-based data visualization methods". *Intelligent Data Analysis* 3 (2): 111–126.

Zitzler, E., K. Deb, and L. Thiele. 2000. "Comparison of Multiobjective Evolutionary Algorithms: Empirical Results". *Evolutionary Computation* 8 (2): 173–195.

AUTHOR BIOGRAPHIES

OFER SHIR is currently a Research Scientist at IBM Haifa Research Labs. His topics of interest include multi-criterion decision making, computational intelligence, experimental optimization, and quantum control. Prior to joining IBM Research in 2010 as a staff member, he has been a Postdoctoral Research Associate in the Department of Chemistry at Princeton University, USA. He holds a B.Sc. in Physics and Computer Science from the Hebrew University of Jerusalem, Israel, and both M.Sc. and Ph.D. in Computer Science from Leiden University, The Netherlands. His email address is ofersh@il.ibm.com.

DMITRY MOOR is a Software Engineer at IBM Systems and Technology Group, currently working on design and development of system management and control software for zSeries machines. He joined IBM Lab in Moscow in 2011 as a Software Engineer Intern in the SDAT team where he was working on the development of z/OS components. At the same time he carried out a research project in multiobjective optimization in Bauman University, and he continues to study this topic with the Decision Analytics group in IBM Haifa Research Labs. He has a B.Sc. and an M.Sc. in Computer Science from the Bauman Moscow State Technical University. His email address is dmitry.moor@ru.ibm.com.

SHAHAR CHEN is a PhD student in the Department of Computer Science at the Technion – Israel Institute of Technology. He is also an intern at IBM Haifa Research Labs, working on the design, analysis and implementation of algorithms, ranging from highly theoretical to the most practical of efforts. His main research interests are in combinatorial optimization, approximation algorithms, online algorithms, and other areas of theoretical Computer Science. His email address is shaharch@cs.technion.ac.il.

DAVID AMID is a technical leader in the Decision Analytics group at IBM Haifa Research Labs. His main focus is the exploration of future tooling for multi-criteria decision support. In the past he explored the interplay between office tools and traditional modeling tools. His background research interests mainly include requirements and service engineering, and the development of distributed computing systems. He has a B.A. in Computer Science and an M.Sc. in Information Management Engineering from the Technion – Israel Institute of Technology. He is a certified PMI Project Management Professional and was recently nominated an IBM Master Inventor. His email address is davida@il.ibm.com.

DAVID BOAZ is a research staff member in the Decision Analytics group at IBM Haifa Research Labs. His research focuses on new visualization techniques for facilitating better decision-making. He holds a B.Sc. in Industrial Engineering and an M.Sc. in Information Systems Engineering from the Ben-Gurion University in Israel. His email address is davidbo@il.ibm.com.

ATERET ANABY-TAVOR is the manager of Decision Analytics group at IBM Haifa Research Labs. Her main responsibility focuses on the development of techniques and tools for Decision Analytics, mainly in the domain of Strategic Planning. These tools lay on the foundation of Simulation, Optimization, and Visual Analytics. Prior to this assignment her research was focused on the future of business modeling tools levering the Business Analysts and Business Architects work and teamwork. She holds a B.Sc. in Industrial Engineering and Management and an M.Sc. in Information Management Engineering (Cum Laude) from the Technion – Israel Institute of Technology. Her email address is atereta@il.ibm.com.