

Self-Organizing Maps for Multi-Objective Pareto Frontiers

Shahar Chen, David Amid*, Ofer M. Shir,
Lior Limonad, David Boaz, Ateret Anaby-Tavor
IBM Research, Haifa Lab,
Mt. Carmel, Israel

Tobias Schreck†
University of Konstanz
Konstanz, Germany

ABSTRACT

Decision makers often need to take into account multiple conflicting objectives when selecting a solution for their problem. This can result in a potentially large number of candidate solutions to be considered. Visualizing a Pareto Frontier, the optimal set of solutions to a multi-objective problem, is considered a difficult task when the problem at hand spans more than three objective functions. We introduce a novel visual-interactive approach to facilitate coping with multi-objective problems. We propose a characterization of the Pareto Frontier data and the tasks decision makers face as they reach their decisions. Following a comprehensive analysis of the design alternatives, we show how a semantically-enhanced Self-Organizing Map, can be utilized to meet the identified tasks. We argue that our newly proposed design provides both consistent orientation of the 2D mapping as well as an appropriate visual representation of individual solutions. We then demonstrate its applicability with two real-world multi-objective case studies. We conclude with a preliminary empirical evaluation and a qualitative usefulness assessment.

Index Terms: [Information Systems]: Information Systems Applications—Decision Support Systems; [Human-Centered Computing]: Visualization—Visualization Design and Evaluation Methods; [Computing Methodologies]: Machine Learning—Machine Learning Approaches Neural Networks

1 INTRODUCTION

In real-world problems, decision makers must often consider multiple conflicting objectives and a large solution space with many candidate alternatives. The multi-criterion decision-making process commonly involves two mathematical spaces: (a) the **design space**, comprising the defining variables of the candidate solutions, and (b) the **objective space**, constituting the mapping of each candidate solution to the multiple objective functions values. The latter is the space where *optimality* is defined, *tradeoffs* are explored, and *decisions* are typically reached. A long-studied topic in the area of Multi-Criterion Decision Making (MCDM) is how to assist decision makers reach better decisions in a more efficient way [22]. The MCDM research community has identified two major challenges: (a) *reducing the number of options* by means of an optimization process, yielding a smaller set of *optimal* solutions; and (b) *effectively visualizing* these solutions to facilitate selection in a way that best satisfies users' subjective/self-determined criteria. This paper addresses the topic of visualizing the set of multi-objective optimal points, also known as the Pareto Frontier.

The task of visualizing the Pareto Frontier is based upon multivariate visualization, and is generally considered a difficult problem for more than three objective functions [23]. In this design study paper, we follow Munzner's [32, Chapter 27] four levels of visualization validation. Hence, we start with the definition of the data

and tasks of the multi-criterion decision-making problem domain. We continue with the design requirements for an effective visual encoding, and a design based on a modification of the well-known self-organizing map (SOM) algorithm [20]. Our solution comprises the introduction of semantically meaningful anchor points, which enable the consistent orientation of the resulting map. In addition, it creates a clear and comprehensible visual representation of individual solutions to facilitate user exploration of the Pareto Frontier. The outcome offers a concise overview of the Pareto Frontier that fosters the efficient building of a mental model by the user, eventually leading to an improved decision process.

The remainder of this paper is organized as follows. In Section 2, we present related work. In Section 3, we carefully describe the multi-criterion decision-making domain. In Section 4, we define design requirements for effective visual encodings of Pareto Frontier data and discuss our design choices. In Section 5, we present the Self-Organizing Maps for Multi-Objective Pareto Frontiers (SOMMOS) visualization design, followed by the description of the algorithm technique in Section 6. In Section 7, we present the application of our newly proposed approach to two real-world problems. We continue in Section 8 with both a comparative evaluation and initial user feedback. In Section 9, we discuss the limitations and drawbacks of our approach. Finally, Section 10 concludes this study and outlines directions for future work.

2 RELATED WORK

We review here existing work on visualizing Pareto Frontier data, with emphasis on methods based on Self-Organizing Maps.

2.1 Multivariate Visualization of Pareto Frontiers

Korhonen and Wallenius [23] categorize the suitability of MCDM visualization techniques based on the cardinality of the result set, i.e., visualizing a single solution, a finite solution set, or an infinite set of solutions. Selecting a visualization technique depends upon the formulation of the decision-maker's preferences, which can be either given *a priori*, *progressively*, or *a posteriori* [17]. In *a priori* techniques, decision makers express their preferences prior to the search (optimization) procedure, e.g., identify relevant regions of interest in the objective space. In *progressive* techniques, the decision-makers' preferences are integrated within the search procedure. Finally, in *a posteriori* techniques, the search is first conducted, and it is followed by a multi-criterion decision making process. In this paper we focus on *a posteriori* techniques for attaining the decision maker preferences. The key advantage of *a posteriori* techniques is that decision makers are offered the opportunity to explore the Pareto Frontier prior to applying their personal preferences. Decision makers are often unsure of their preferences at the beginning of the process and may be informed and influenced later on by the entire spectrum of options. The key disadvantage to *a posteriori* techniques is that they require much more computational resources to generate the entire Pareto Frontier. In addition, the result of *a posteriori* techniques may contain a significant number of solutions, leading to a harder decision-making challenge. According to Lotov [26], methods that are more suited for visualizing a single solution or a small number of solutions (e.g., radar charts or Chernoff faces [6]) are less effective for visualizing the

*e-mail: davida@il.ibm.com

†email: tobias.schreck@uni-konstanz.de

Pareto Frontier as a whole, as compared to methods for visualizing large multivariate data sets (e.g., principal component analysis (PCA) [19] or multidimensional scaling (MDS) [7]).

2.2 SOM: Self-Organizing Maps

For this paper, we enhanced a multivariate data visualization methodology for large data sets, with a focus on the self-organizing map (SOM) visualization technique [20], a popular method for visualizing large high-dimensional datasets. Using a competitive learning algorithm, a network of prototype vectors is formed. Typically, upon employing a regular two-dimensional *rectangular* or *hexagonal* network model, visualization is straightforward, as data items can be mapped and visualized according to a spatial structure [35]. Due to approximate topology-preserving properties of the SOM method, the resulting visual displays can be intuitively interpreted by the user. To date, this method has been successfully applied in a large variety of real-world domains. Moreover, several variants extend the basic SOM method to provide customized network shapes or data layouts. In [28], the baseline network is modeled to reflect the data domain being visualized. In [31], an approach was proposed for interactive specification of the global network layout by manually assigned reference prototypes. The well-known SOM package, SOMPAK [21], includes an option to pre-assign data vectors to network positions, allowing indirect control over the SOM layout to be achieved. To some extent, this approach is similar to the RadViz approach [15], in which high-dimensional data is mapped onto a two-dimensional space by a concentric spring-based layout, where dimensions exert force on points. The approach allows to assign relative influence of dimensions between different points, based on their spatial position. RadViz has also been generalized by the Dimensional Anchor framework [16], which constructs a whole family of mappings based on a small set of structural parameters.

2.3 Existing Pareto Frontier Visualization Techniques

Parameter Space visualization in which users need to relate multivariate input- and output spaces to each other is an important yet difficult problem. To date, several studies have proposed using visualization to help users relate input and output spaces in interactive systems. Berger et al [3] propose an interactive approach, based on 2-dimensional projections, which allows the user to navigate multivariate input spaces while observing potential output behavior. Variations along selected dimensions of the input space are allowed, and the resulting sensitivity in the output space is shown by a covered area in the projected output space. SOM has been employed within the context of multi-objective optimization [29], particularly for *design space* exploration by projecting the m -dimensional Pareto frontier onto a two-dimensional map and analyzing its relationship with the decision variables. During this process, the SOM map is displayed m times, and each display is colored with respect to the value of a certain objective. The map is then colored once again per decision variable, to visualize the effect of the decision variables on each objective. While this might be effective for exploring the relationships between the design- and the objective spaces, it has deficiencies, such as the inability to identify a specific solution or traverse over the Frontier in an oriented manner. Another use of multiple displays is Interactive Decision Maps [25], in which Lotov et al. employ scatter-plot matrices to display bi-objective slices of the Frontier. Each slice is a scatter-plot of two objectives. The points are colored according to the value of the third objective. All other $m - 3$ objectives are set to a certain value. Using sliders, the user can change the values of the $m - 3$ objectives and explore the Frontier. Andrienko et al. [13] enhanced this approach by using utility bar chart as means to both filter the visible options and compare amongst them. To date, the Parallel Coordinates method [18] is the prevalent technique used to

visualize Pareto frontiers [2]. The major advantage in using this technique is the ability to easily compare between solutions across all objectives. However, as the number of solutions increases, the display becomes cluttered and unclear, making it difficult to gain insight and identify patterns in the entire frontier without filtering the solutions. Further details on techniques for Pareto Frontier visualizations can be found in a review by Korhonen et al. [23].

3 PROBLEM DOMAIN

Our first contribution of this design study is a characterization of the problem domain. This characterization includes a description of the Pareto Frontier data and the tasks decision makers face as they identify the solution that best meets their criteria.

3.1 Pareto Frontier Data

The rigorous definition of optimal solutions in multi-objective optimization, also referred to as Pareto optimality [10], necessitates the formulation of a partial order within the objective space called *Pareto domination*. Pareto domination states that one objective vector dominates another if all its values are at least as good as the other's, but it is strictly better in at least one of the objectives. The aim of Pareto optimization is thus to obtain the non-dominated set for the given objective functions, called the **Pareto Frontier**, and its pre-image in the design space, called the **Pareto optimal set**. The Pareto Frontier data constitutes multivariate data in which each variable is mapped onto an objective. The variable is quantitative and sequential, as each objective value ranges between minimally- and maximally-attainable values. In principal, Pareto frontiers are expected to possess a large number of solutions (at least a couple dozen). The dimensions of the Pareto Frontier data depend on the number of objectives and can vary from 2 to an infinite number. In practice, problems consisting of more than 8 objectives are rare and considered very challenging, both at the optimization and the human-cognitive levels.

3.2 Tasks

Multi-objective optimization is typically used in a variety of domains, ranging from physics and chemistry to finance, consumer markets, and portfolio management. Decision makers from these domains have different skill-sets and different expectations from a multi-objective visualization tool. In this paper, we focus on the common denominator. The decision maker's first and foremost important goal – to identify the solution that best matches their preferences. We did not focus on tasks, common in the scientific community, such as statistical analysis, whose goal is to facilitate a deeper understanding of the alternatives at hand. We believe these tasks are secondary to the overall goal of decision making. In Table 1, we describe the tasks decision makers face in their quest for a decision. In each task, we list the questions the decision maker seeks to answer and the reasons they ask these questions. The tasks were identified based on discussions held with analysts and managers from a variety of industries.

4 DESIGN DECISIONS

The second contribution of this design study is a set of requirements that can help generate an effective visual encoding of Pareto Frontier data. Our first design decision was to focus the proposed encoding and design choices on the need to fulfil the first three tasks: Exploration, Inspection and Scoping. The fourth task, Comparison, usually occurs after the first three tasks and is already addressed by existing methods of Pareto Frontier visualization such as parallel coordinates [2].

In Pareto Frontier data, each dimension is meaningful from a semantic point of view, since it represents an objective function, and all objective functions are considered to be equally important in these stages. Hence, our second design decision was that a visual

Table 1: Tasks

Task Name	Question	Reason
Exploration	What is the nature of the conflict between the objectives – soft conflict or hard conflict [11]? How diverse are the solutions?	Decision makers seek to gain insight from viewing the overall picture, e.g., one informative view that can help them comprehend the entire Pareto frontier. Insights may be gained by identifying areas of interest, i.e., a set of "similar" solutions, where similarity can be defined using a distance metric (e.g., Manhattan, Euclidean, Mahalanobis)
Inspection	What are the values of each objective? What is the tradeoff between them?	Decision makers seek to easily identify and differentiate a solution from other solutions. Furthermore, they wish to examine the tradeoff between the objectives values in a specific solution (i.e., the magnitude of each objective with respect to other objective values)
Scoping	Can the search be narrowed down to a subset of solutions?	Decision makers seek to apply their subjective preferences, constraints, or analytical methods to facilitate the selection process [27].
Comparison	What are the differences between multiple selected solutions?	As decision makers' attention traverses from exploring the problem to examining the specific solutions, they seek to understand the tradeoffs of selecting one solution over another.

encoding must maintain *Objectivity*. The *Objectivity* requirement states that a visualization should equally represent the Pareto Frontier dimensions to allow non-biased interpretation by the user, unless explicitly requested otherwise.

Our third design decision was to determine which type of multivariate visualization to use. According to Korhonen et al. [23], one of two visualization approaches may be employed to represent multivariate data: (1) depicting each solution as an object by addressing its different attributes (also known as axis reconfiguration techniques [12]); or (2) reducing dimensionality, i.e., by projecting the multivariate data onto a two-dimensional plane.

Despite the differences between the two approaches, they are often utilized to complement each other. The former approach is potentially more informative, since no information reduction is necessarily carried out, as occurs in the latter approach. The latter approach is more intuitive for navigation and exploration. This is since displaying points (representing solution vectors) on a plane, while analyzing the relationships amongst them, is a particularly convenient representation for the human eye, which uses position and Euclidean distance for elementary analysis. Nevertheless, some dimensionality reduction methods such as dimension slicing, cutting, and hierarchies should be used with caution, as they do not equally consider all dimensions. As described in the *Exploration* task, we expect decision makers to employ our visualization to identify an area of interest and eventually a solution. This can be achieved through the facilitation of easy navigation through the Pareto Frontier. To meet this task, we define additional two requirements for an effective visual representation of a Pareto Frontier – *Adjacency* and *Orientation*.

Adjacency is defined as the facility to generate pairwise neighboring relations between objective vectors that satisfy the following – close proximity in the visualization of any two objective vector representations will take place if and only if the two objective vectors are similar. SOM and MDS are examples of visualizations that aspire (though not guarantee) the adjacency criterion, whereas linear

projection may only satisfy it in a limited manner.

In *Orientation*, we distinguish between two orientation types as follows – (1) *global orientation*: at a high-level of visualization, the ability to identify correlation between any two depicted *areas* of solutions (i.e., dissimilarities, trends, etc.); versus (2) *local orientation*: at the individual solution level, it should be clear which alternative solution can be selected, should an increase (or decrease) in a certain objective be desired. Note that SOM and MDS both lack orientation, whereas linear projection possesses good global orientation and Chernoff faces offer good local orientation.

In light of the greater support for the *adjacency* requirement, we chose to focus on a dimensionality reduction technique and enhance it with *orientation* as necessary.

5 THE SOMMOS VISUAL DESIGN

Our third contribution is the design of a Self-Organizing Map for visualizing multi-objective Pareto frontiers (SOMMOS). We designed SOMMOS' visual encoding and interactions based on the problem domain and design considerations previously mentioned. SOMMOS visualization comprises three pillars: a map layout, a visual representation of individual solutions, and interactive capabilities. The map layout addresses the Exploration task by enabling the user to draw insights from its structure. Each solution on the map is encoded using glyphs to meet the Inspection task. Interactive capabilities enable the decision maker to fixate on a narrow subset of solutions (the Scoping task).

Layout Visualizing a Pareto Frontier requires attributing equal importance to each of the objectives. This requires equal representation of the visualization dimensions to allow equal interpretation of the dimensions by the user. In SOMMOS, the m objectives are represented by the corners of the map. This can be accomplished by using a symmetric m -gon, a polygon with m edges of equal length – and positioning the objectives at all m vertices. Thus, given an m -dimensional objective space **subject to maximization**, we construct m vectors, where the i^{th} vector, $1 \leq i \leq m$, maximizes the value of the i^{th} objective and minimizes the values of the remaining $m - 1$ objectives. For instance, in Figure 3, the Pareto Frontier data is a set of triplets and as such has a triangular layout. This figure is thoroughly explained in Section 7.1. As described in Section 6, setting these vectors yields a map with orientation, e.g. the closer the solution is to a corner of an objective, the greater its value is in that objective. An initial discussion with two advisory users showed that users first examine the solution nearest to each anchor to determine what is maximally attainable in each objective. Then users tend to begin their exploration from a certain point based on their preferences, and browse through the solutions with a general tendency towards the center. When asked, they explained they have identified an area of interest and then began to examine the tradeoffs between the objective values at each solution until they reached a certain perceived threshold they were not willing to compromise. As a result of this feedback, we added the depiction of the maximally attainable numeric value at the anchor to eliminate this redundant initial check of the boundaries. To enhance the ability to gain insight when viewing the overall picture, an additional clustering layer is added, on top of SOM inherent clustering. This division of the map into areas enables the decision maker to easily identify an area of interest. For instance, if the decision makers seek a meet-in-the-middle solution, they will focus on a cluster at the center of the map. By choosing a cluster, it is guaranteed that similar solutions will be selected for detailed examination as opposed to when exercising manual selection. The clusters shown in Figure 3 were obtained by processing the Pareto Frontier data with the K-Means algorithm [24]. Each cluster is assigned a color following the Colorbrewer's qualitative color scale [5]. The value of K , i.e., the desired number of clusters given as input to K-Means, should be derived from the number of objectives. Here, the following rule



Figure 1: Demonstrating the *proposed* radius formulation (Eq. 1) by depicting a specific solution from the TAM example (Section 7) compared to the common radial bar charts formulae ('sqrt', 'linear').

of thumb was used: $K \in \{m+1, m+2\}$. Clearly, a rigorous experimental tuning of this parameter should be conducted.

Visual Representation of Individual Solutions on the Map Display Visualizing the Pareto Frontier as a whole must include the representation of individual solutions in a manner that enables the easy interpretation of their value. Therefore, we augmented the plain SOM using large-scale glyphs to provide full-data extraction and easy exploration of the tradeoffs between the objectives. Each objective is assigned a color following the Colorbrewer's qualitative color scale. We chose radial bar charts (Polar Area Chart) as the glyphs due to their economical use of space, as opposed to profile glyphs, such as bar charts, in which the occupied space depends on the number of objectives. Pie charts were ruled out because of their implicit denotation of the ratio between the objectives, rather than the magnitude of the objectives, as presented in radial bar charts. Following this decision, it was important to devise a formulation such that changes in the magnitude are perceptible. In radial bar charts the magnitude of each sector depends on the radius value. Unfortunately, changes towards the range boundaries are unnoticeable for the typical radius formulation (linear, square root). As can be seen in Figure 1, setting the radius proportionally to the area, e.g. square root of the objective's normalized value, results in good distinction in the regime of minimal values, while it is practically impossible to note differences in the regime of maximal values. On the other hand, setting the radius to be equal to the normalized value itself (e.g. a linear function) would only result in good distinction in the regime of maximal values. To accommodate user's perception of variations across the entire range and in both regimes, we devised the following formulation for the radial bar chart radius, given the normalized objective value v :

$$r(v) = \begin{cases} 0.6 \cdot \sqrt{2 \cdot v} & \text{if } v < 0.5 \\ 0.6 + ((v - 0.5)/2) & \text{if } 0.5 \leq v < 0.7 \\ 1 - 0.3 \cdot \sqrt{(1-v)/0.3} & \text{if } v \geq 0.7 \end{cases} \quad (1)$$

Finally, to gain a better perception of the magnitude, a contour was added by overlaying the radial bar chart on a blank equiangular pie chart with a radius of 1.0. The glyphs were initially designed to enable addressing the *Inspection* task. In other words, rather than using simple marks, the use of glyphs can assist the decision maker in understanding the tradeoffs between objective values for the specific solution at hand. However, our advisory users noted that they mainly used the glyphs to understand the nature of the conflict. They explained that they could instantly identify global trends, and that they were able to spot an area of solutions that maximized one of the objectives at the expense of others. In another case, they found the glyphs helpful for understanding the nature of the compromise of the area at the center of the polygon. Furthermore, they were able to identify those cases in which the objectives correlated. Next, they noted that when reviewing a specific solution they found the tool-tip (see, e.g., Figure 3), which explicitly states the numeric values of each objective, far more convenient.

5.1 Interactive Visualization

As part of the design of an interactive visualization environment, we used Shneiderman's visual information-seeking mantra [33]: "Overview first, zoom and filter, then details-on-demand". User

```

learnSOMMOS()
input Dataset: Pareto Frontier  $\mathcal{F} = \{\vec{f}_k\}_{k=1}^{|\mathcal{F}|}$ ,  $\vec{f}_k \in \mathbb{R}^m$ 
1: normalizeFrontier()
2: initSOMMOS()
3: for  $t = 0 \dots T_{\text{period}}$  do
4:    $\mathcal{S} \leftarrow \text{randomSampling}(\mathcal{F})$ 
5:   trainSOMMOS( $\mathcal{M}, \mathcal{S}, h_F(), \alpha_F(t)$ )
6:   if (mod( $t, T_{\text{anchor}}$ ) == 0) then
7:      $\{\vec{a}_\ell\}_{\ell=1}^m \subset \mathcal{M} : \vec{a}_\ell \leftarrow \mathcal{I}_{:, \ell}$  {anchoring: explicit fixing at the vertices}
8:     trainSOMMOS( $\mathcal{M}, \mathcal{A}, h_A(), \alpha_A(t)$ )
9:   end if
10: end for
11: return  $\mathcal{M}$ 

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

trainSOMMOS(map  $\mathcal{M}$ , data  $\mathcal{X}$ , kernel  $h()$ , learning-rate  $\alpha$ )
1: for each point in the dataset  $\vec{x}_k \in \mathcal{X}$  do
2:   for each neuron  $\vec{u}^{(j)} \in \mathcal{M}$  do
3:      $\vec{u}^{(j)} \leftarrow \vec{u}^{(j)} + \alpha \cdot h(\vec{x}_k, \mathcal{M}) \cdot (\vec{x}_k - \vec{u}^{(j)})$ 
4:   end for
5: end for

```

Figure 2: Pseudo-code summarizing SOMMOS.

exploration starts with an overview of the map. After using the anchors as reference points, users are able to judiciously explore the map. As can be seen in Figure 3, at any given point, users can use the filter sliders to denote their preferences. The solutions that do not meet their preferences are then grayed out. As users fixate on an area of interest, they can zoom in and view the glyphs more clearly. Using pan, users can navigate through the map at a resolution that meets their needs. When hovering over a specific solution, a tool tip pops up to provide more details and enable certain interactions. As can be seen in Figure 3, once decision makers have reviewed the details of a given solution, they can filter it out to avoid reviewing it again, select it (resulting in the glyph being highlighted), or mark it for further review by adding it to the basket. Finally, the user can select the nearest neighboring solutions to examine them in more detail, using the existing map or additional views.

6 SEMANTICALLY-ENHANCED SOM GENERATION

Our fourth contribution is the modification of the SOM algorithm to semantically represent the optimum of an individual objective function as an anchor point within the map. From an analytical perspective, the SOM neural training phase equally considers each of the objectives, yet from a visualization perspective, SOM does not explicitly represent the individual extremes (optima) of the objective functions. The Self-Organizing Maps for Multi-Objective Pareto Frontiers (SOMMOS) introduces an explicit representation of the dimensions using a so-called *SOM anchor* as a fixed point for each objective. To obtain symmetry between the objectives, the objective anchors must be set on the perimeter, and consequently, we choose to employ a symmetric m -gon – i.e., a polygon with m edges of equal length – and to position the anchors at all m vertices. A possible way to implement an m -gon layout is to overlay the map over the default layout and discard the neurons that are left out – ensuring that dataset points are associated only with neurons within the m -gon. As described in Section 4, orientation is one of two requirements that visual representations of a Pareto Frontier must meet to fulfil the Exploration task. Orientation reduces the likelihood of the occurrence of distant similar neighborhoods, leading to a reduction in the topographic deformation. The use of

anchors during the training phase of SOM results in a boosting of orientation [21], and it is therefore the core of the proposed procedure to be described in what follows with reference to the pseudo-code in Figure 2. During the training phase, every epoch denoted as T_{anchor} , a given vertex, $\vec{a}_\ell \in \{\vec{a}_\ell\}_{\ell=1}^m \subset \mathcal{M}$, percolates its value to its surrounding neurons based on the learning functions (this is termed "anchoring"). Therefore, 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 repeatedly train the map as an additional independent dataset \mathcal{A} , which exclusively represents the vertices; the actual vertices are also explicitly reset. Essentially, the learning procedure comprises two processes that require different learning-rates. Since the learning-rate function for the input dataset \mathcal{F} , noted as $\alpha_F(t)$, declines over time (to stabilize the network and achieve convergence), a different learning-rate function for the anchoring dataset \mathcal{A} is required. We derived an anchoring learning-rate function, $\alpha_A(t)$, which 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 neural map is summarized as `learnSOMMOS()` in Figure 2, where the Pareto Frontier dataset is normalized and the anchoring is thus carried out using the identity matrix. In terms of computational effort, SOMMOS operates in a similar manner to the classical SOM algorithm, except for the additional anchor training, which results in a performance factor of $1 + \frac{m}{|\mathcal{F}| \cdot T_{\text{anchor}}}$ with respect to SOM.

6.1 Determining the Objectives Arrangement

Spatially arranging the objectives on the map is a crucial step to attaining an effective visualization, as seen in other anchor-based techniques, such as Radviz and Barycentric [1, 9]. To achieve good arrangement of the objectives, while at the same time maintaining good map orientation and the adjacency property, we propose the following methodology. The rationale behind this approach is to arrange the objectives such that correlative pairs are assigned to relatively close anchors, while anti-correlative pairs are assigned to distant ones. Towards this end, we use Pearson's product moment coefficients to calculate the correlation between each pair of objectives, noted as $\rho_{i,j}$. In addition, we compute the Euclidean distance between every two anchors, noted as $D_{i,j}$. We then define the optimal arrangement of objectives as the optimal objective-to-anchor assignment, π^* , such that $\rho_{i,j}$ is anti-correlative to $D_{\pi(i),\pi(j)}$ for all pairs i, j . The proposed technique iterates over all $\frac{(m-1)!}{2}$ possible arrangements, calculates Pearson's product moment coefficient between a $\binom{n}{2}$ -long pairwise correlation vector to its corresponding Euclidean distance vector – and selects the permutation with the minimal correlation value.

7 INSTANTIATION: REAL-WORLD PROBLEMS

To demonstrate the wide applicability of the Self-Organizing Maps for Multi-Objective Pareto Frontiers (SOMMOS), we consider application scenarios from two different domains: Project Portfolio Management (PPM) and Asset Management. These two real-world examples differ in their optimization modeling, the number of objectives, and the size of their Pareto frontiers. The utilized parameters described here were derived following empirical trial-and-error, and will undergo parameter tuning in future work.

7.1 The Project Portfolio Management Problem

We consider a multi-objective PPM optimization problem. PPM explores the complete set of projects a company is involved in to decide which projects should be included in the firm's portfolio. Interdependencies between projects are also accounted for to verify that no contradicting or overlapping projects are carried out. Achieving an optimal portfolio mix requires a company to clearly articulate the portfolio's goals and objectives. Although objectives may vary

across organizations, the most prominent ones usually seek to maximize financial metrics such as revenue, net present value, and return on investment, while minimizing the company's risk [30].

Practical Observation We consider a three-dimensional PPM problem, where the objectives are defined as maximizing the risk reduction, maximizing the revenue, and maximizing the cost reduction. Although revenue and cost may be merged into a single objective, it is practically preferred to keep them separate, since organizations cannot typically spend a large budget even if they are expected to gain larger amounts in the future. As an optimization solver, we employ the mixed-integer version of the so-called SMS-EMOA [4]. Figure 3 displays the map attained by the solver, which was further processed using k-means clustering ($K = 4$). The map was generated with 50 neurons and follows these parametric values (for more details, see Section 6): $T_{\text{anchor}} = 251$, $r_A = 2.5$, $\alpha_F(t = 0) = 1$. As decision makers explore this map, they can spot the nature of the conflict between the objectives. By focusing on the top right cluster, one can see that the *cost reduction* and *risk reduction* objectives are in a hard conflict, i.e., much is lost in cost reduction as you maximize risk reduction. By willing to lose in cost reduction while still focusing on revenue, it can easily be observed that the bottom-right cluster contains solutions that maximize revenue and even possess moderate risk reduction. Finally, it is clearly evident that the bottom-left cluster reflects a good compromise (soft conflict) between *cost reduction* and *revenue*.

7.2 The Asset Management Problem

We also consider a multi-objective asset management optimization problem. Similarly to PPM, asset management explores the complete set of actions required in a designated time frame to maintain the overall condition of the assets. Asset management is considered a wide field that spans different domains such as transportation fleets, transportation infrastructure, sewage networks, real estate, and even power plants.

Practical Observation We consider a 5-dimensional transportation asset management (TAM) problem. In transportation asset management, making cost-effective decisions about allocating resources to preserve, maintain, or improve transportation infrastructure (roads, bridges, and buildings) is crucial. In the presented investment plan, the multiple objectives were defined as maximizing congestion reduction, maximizing pedestrian and cyclist trails, maximizing safety, maximizing economic growth, and maximizing air quality across the transportation network. As in the PPM problem, we employ the SMS-EMOA as the optimization solver, and utilize SOMMOS to visualize the attained Pareto surface (depicted in Figure 4). The map was generated with 86 neurons and following these parametric values: $T_{\text{anchor}} = 431$, $r_A = 4$, $\alpha_F(t = 0) = 1$. It is evident in the map that neighboring objectives, such as air quality and economic growth, are in soft conflict, as their segments in the polar area glyphs can be simultaneously colored. In contrast, objectives that are distant from each other, such as economic growth and congestion reduction, are in a harder conflict. For example, when congestion reduction has a dominant colored segment, economic growth is nearly empty, and vice versa, reflecting the conflict.

8 EVALUATION

8.1 Initial User Feedback

We organized an exploratory field study to test our visualization. Overall thirty Information-Systems Engineering undergraduate senior students were presented with our system and asked to use it to decide on two problems. The first problem, was in a domain with which they have strong familiarity – deciding on an apartment to lease given three objectives: minimum distance from the university, minimum rent, and minimum number of roommates. The second problem domain for which the students were less familiar with

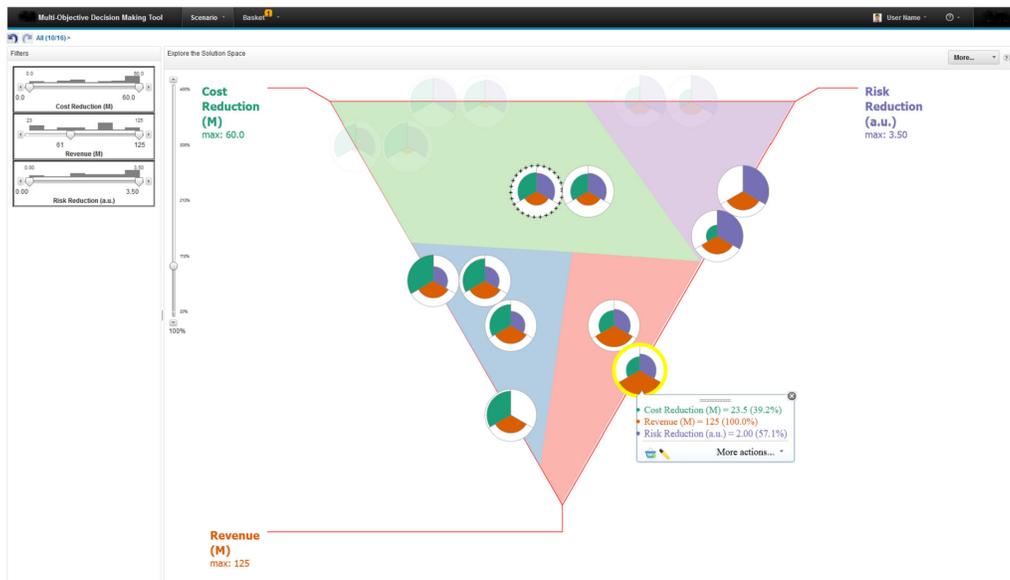


Figure 3: A **screen-shot** of the Self-Organizing Maps for Multi-Objective Pareto Frontiers (SOMMOS) as part of the PPM system. The decision maker decides she would like to gain in revenue at least \$60M, so she filtered out the solutions with low revenue. She then adds a solution to the basket (outlined in + signs) that she sees as a good compromise between all the objectives. Given that she does not want to compromise too much on risk but still would like to increase her revenue, she examines another solution within the cluster that maximizes revenue. She notices that by compromising on 0.5 in risk she can maximize her revenue and still have a decent cost reduction. She can now either add that solution to the basket and compare the two more rigorously or continue exploring other options.

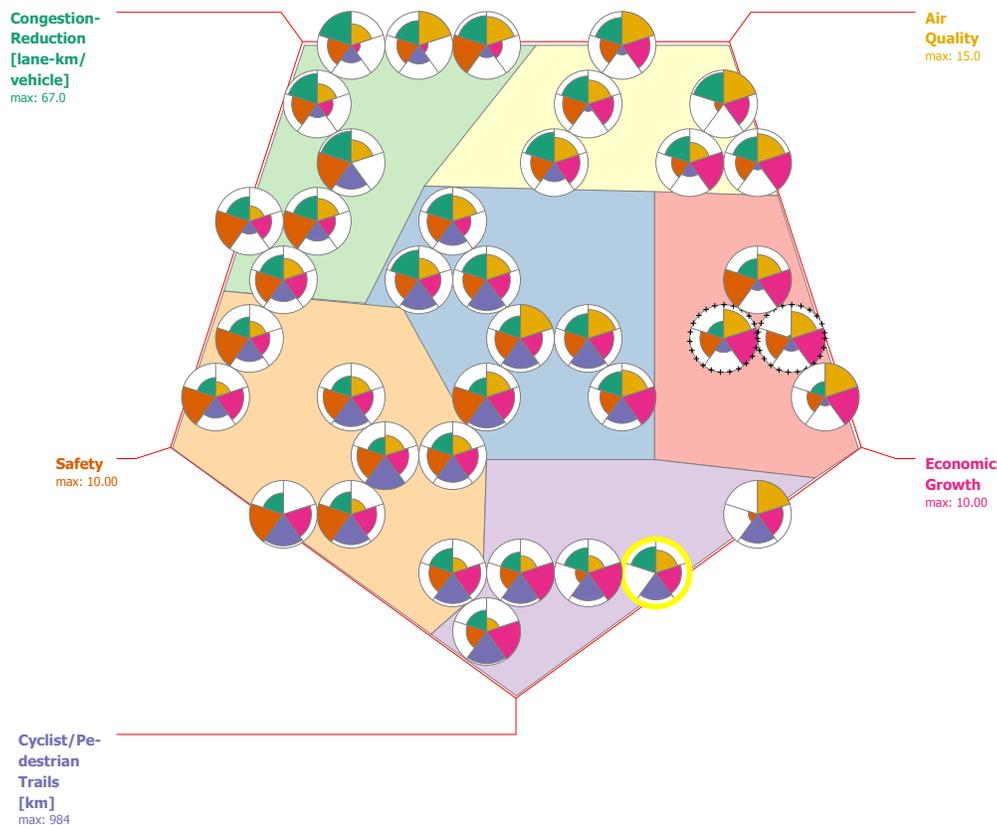


Figure 4: A five-objective Transportation Asset Management problem. Focusing on the yellow cluster, which aims at maximizing air quality, the decision maker can clearly identify that she can maximize both air quality and congestion reduction. She can compromise a bit on congestion reduction to gain higher economic growth. Unfortunately, there is no option for a decent compromise with cyclist/pedestrian trails without hindering the air quality. Note that both the glyph fill color and the order of the objectives contribute to the decision maker's ability to understand the above insights.

was the TAM problem presented in the previous section. Students were arbitrarily assigned into four groups, each of which was presented a different visualization technique (parallel coordinates, the Self-Organizing Maps for Multi-Objective Pareto Frontiers (SOMMOS), both SOMMOS and parallel coordinates, and a table). Note that advanced SOMMOS capabilities such as clustering were removed to match plain parallel coordinates. The pilot started with a short training session in the domain of multi-objective decision making, accompanied by a demonstration of the visualization in a separate problem of choosing a car by examining three objectives: fuel efficiency, speed and car safety. Students were then given a 10-question quiz to determine if they understood the concepts and use of the tool presented in the training. Prior to the experiment, a passing score of 70% was determined. Quiz results indicated that all participants understood well the visualization presented to them and how to operate it. For each problem, the students were requested to make a decision and record a short explanation. We then followed with a set of questions to determine their understanding of the nature of objective tradeoffs and their usage experience. As the amount of participants was small, we were more attentive to observing their actual work-in-progress and reading their short choice argumentations. Most argumentation indicated that for many students who had used SOMMOS, a global perspective was retained, considering both a wide range of alternatives, and several objective tradeoffs. This was expressed in statements such as the following typical response: *"I have tried maximizing all parameters [objectives] at minimum conflicts, while attaining threshold values for safety and air-pollution being important nowadays. Gradually, alternative scope has been narrowed down with filtering till a choice was reached."* Contrary to the above, the other groups demonstrated a need to rely on some a-priori objective ordering, demonstrated by responses of the following spirit: *"Distance was most critical in my decision consideration. Then, I filtered out [alternatives] according to maximal price threshold and number of roommates being not more than 2."* In terms of user interaction, it was also apparent that regardless of the concrete visualization type in use, the filter interaction usage was the most prominent one. Amongst the participants who had used SOMMOS it was interesting to note that in order to determine the level of conflict between objectives the students moved the filter back and forth resulting in animation of the gradual disappearance of colors from the map. This preliminary observation served as a trigger to the inception of a more conventional and extensive laboratory experiment.

8.2 Comparative Evaluation

As with all new visualizations, a comparative evaluation is needed to determine the usefulness of the proposed visualization over existing solutions. In the case of SOMMOS, the introduction of anchors to the SOM layout creates a resemblance to other anchor-based visualizations such as RadViz. In Section 4, we defined two requirements needed to facilitate the easy exploration of the Pareto frontier - *Orientation* and *Adjacency*. However, while the RadViz visualization has good orientation, it does not aspire to maintain adjacency, defined as the facility to generate pairwise neighboring relations between objective vectors that provide close proximity in the visualization of any two objective vector representations, *if and only if* the two objective vectors are similar. Specifically, in the RadViz visualizations, different solutions in the objective space may be placed close in the visualization. For example, consider the following three solutions and their normalized values:

solutions	normalized values
(29, 250, 7, 10, 14)	(0.36, 0.24, 1.00, 0.70, 0.93)
(9, 212, 8, 10, 13)	(0.02, 0.20, 1.00, 0.80, 0.86)
(57, 850, 0, 8, 10)	(0.83, 0.86, 0.80, 0, 0.64)

As can be seen from their normalized values, the first two solutions are similar to each other with exception of the first dimension, while both considerably differ from the third solution. Indeed as can be seen in Figure 4, the two first solutions marked with a + sign on

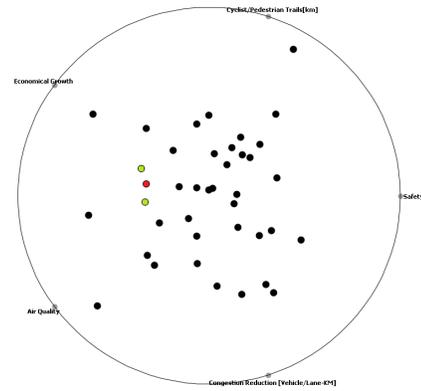


Figure 5: A RadViz application [8] of the transportation asset management problem. The colored points demonstrate a drawback of RadViz, as it does not support the adjacency requirement. Specifically, non-similar solutions can be placed close to each other.

their contour appear close to one another while the third solution, highlighted in yellow, is farther away. However, on the RadViz visualization in Figure 5, all three solutions are close to one another. Moreover, the solution that differs from the others, marked in red, is located between the two similar solutions. To avoid such circumstances, we chose to apply the self-organizing map algorithm that aspires to maintain adjacency, and augment it with orientation using anchors. Clearly, projecting high-dimensional solutions onto a two-dimensional visualization is imperfect, and no visualization can fulfil all requirements flawlessly. Indeed, in Section 9, we list some of the limitations of SOMMOS. However, we believe that small violations in some or all aspects is preferred over ignoring one of the requirements. Specifically, in SOMMOS we settle for a non-linear projection, with possible slight distortions of the objective space, to provide a satisfying solution for the Exploration task. Another quality metric for visualizations is how much of the available screen space is utilized to show data. Tufte's *data ink* ratio [34] measures data density and should be maximized. Screen-filling techniques such as TreeMaps utilize the whole display, giving it the potential to actually fill it to the maximum with information. Whether or not a given display canvas is utilized with meaningful information depends on the visualization design. In visualizing Pareto frontiers, space utilization is even more crucial to avoid overlaps when presenting a glyph per solution. In comparing Figures 4 and 5, it is clear that SOMMOS has a better utilization of space than RadViz, due to the benefits of employing non-linear projection as opposed to linear projection.

9 LIMITATIONS AND DRAWBACKS DISCUSSION

The virtues of the Self-Organizing Maps for Multi-Objective Pareto Frontiers (SOMMOS) lie in its ability to support both adjacency and orientation. However, these two requirements are in conflict and hence orientation errors may appear on maps. For example, in Figure 3, one would expect the solutions with maximum *revenue* to be nearest the *revenue* anchor. To resolve such issues, we are developing an orientation-error quality metric that we plan to use in conjunction with common SOM error indicators (for quantization and topographic errors) to assess the quality of the map. Another issue that arose during the design of SOMMOS was its ability to scale with respect to large Pareto frontiers (in particular, continuous Pareto frontiers). To address this situation, several approaches can be used. First, one can limit the map size and use SOM's inherent clustering property to associate multiple data points with a single neuron. A second approach is to employ tradeoff-based filtering methods such as the Smart Pareto Filter [27], which reduces

the density of solutions on the Frontier based on desired tradeoff patterns. Another open question concerns SOMMOS's scalability with respect to the number of objectives. As with all projection techniques, as the number of dimensions increases, the effectiveness of the visualization decreases. In SOMMOS, the number of objectives also affects the readability of the glyph, as the sector area becomes quite small. Based on our experience, when there are eight or more objectives, it becomes hard to comprehend the information on the glyph. However, in practice, problems in multi-objective decision making rarely have more than eight objectives, since humans cannot cognitively cope with such a large amount of objectives.

10 SUMMARY AND FUTURE WORK

The motivation underlying the Self-Organizing Maps for Multi-Objective Pareto Frontiers (SOMMOS) is to facilitate an efficient decision-making process. Using SOMMOS, decision makers can interactively identify the solution that best meets their criteria. The proposed semantically enhanced SOM generation technique provides a visualization of the entire Pareto Frontier. We augmented plain SOM with orientation and individual representation of solutions to meet the requirements of the multi-objective visualization problem. We used a symmetric m -gon layout of the map to allow equal interpretation of the dimensions by the user. We evaluated SOMMOS by demonstrating its applicability on two distinctly different case studies, presenting a preliminary empirical evaluation, and comparing it to known anchor-based visualization techniques. Future work is planned to extend the current evaluation with both quantitative and empirical evaluations. As mentioned, a possible avenue of future work would involve parametric study of SOMMOS's defining parameters to generate better maps. Finally, consideration of the *sensitivity* of a given solution with respect to its neighboring solutions can be relevant for the analysis. For example, in [14] local sensitivity for high-dimensional data points is computed by the distance of the local point's target value to the prediction of the target value obtained by regression analysis in its neighborhood. This approach can assess the degree of outlyingness or expectedness of local data. This idea could lead to the automatic suggestion of interesting local solutions for exploration by the user. Empirical evaluation is planned to examine the usefulness and the effectiveness of decision making while using this novel visualization. It is our strong belief that SOMMOS' core enhancements can be utilized beyond Pareto Frontiers into additional domains to visualize multivariate data requiring visual exploration based on adjacency, orientation, and objectivity.

REFERENCES

- [1] M. Ankerst, S. Berchtold, and D. A. Keim. Similarity clustering of dimensions for an enhanced visualization of multidimensional data. *INFOVIS*, pages 52–60, Washington, DC, USA, 1998. IEEE.
- [2] M. Bagajewicz and E. Cabrera. Pareto optimal solutions visualization techniques for multiobjective design and upgrade of instrumentation networks. *Industrial and Engineering Chemistry Research*, 42(21):5195–5203, 2003.
- [3] W. Berger, H. Piringer, P. Filzmoser, and E. Gröller. Uncertainty-aware exploration of continuous parameter spaces using multivariate prediction. *Computer Graphics Forum*, 30(3):911–920, 2011.
- [4] N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective Selection Based on Dominated Hypervolume. *European Journal of Operational Research*, 2007.
- [5] C. Brewer and M. Harrower. Colorbrewer. *On-line color brewing tool available from URL <http://www.colorbrewer.org>*, 2002.
- [6] H. Chernoff. The use of faces to represent points in k -dimensional space graphically. *J. Amer. Statist. Assoc.*, 68:361–368, 1973.
- [7] T. Cox and M. Cox. *Multidimensional Scaling*. Chapman-Hall, 2001.
- [8] T. Curk, J. Demar, Q. Xu, G. Leban, U. Petrovi, I. Bratko, G. Shaulsky, and B. Zupan. Microarray data mining with visual programming. *Bioinformatics*, 21:396–398, Feb. 2005.
- [9] L. Di Caro, V. Frias-Martinez, and E. Frias-Martinez. Analyzing the role of dimension arrangement for data visualization in *radviz*. *PAKDD'10*, pages 125–132, Berlin, 2010. Springer-Verlag.
- [10] M. Ehrgott. *Multicriteria Optimization 2nd ed.* Springer, Berlin, 2005.
- [11] M. Emmerich and A. Deutz. *Test Problems Based on Lamé Super-spheres*. EMO. Springer, 2007.
- [12] Y.-H. Fua, M. O. Ward, and E. A. Rundensteiner. Hierarchical parallel coordinates for exploration of large datasets. *VIS*, pages 43–50, Los Alamitos, CA, USA, 1999. IEEE.
- [13] P. J. Gennady Andrienko, Natalia Andrienko. Building spatial decision support tools for individuals and groups. *Journal of Decision Systems*, 12(2):193–208, 2003.
- [14] Z. Guo, M. Ward, E. Rundensteiner, and C. Ruiz. Pointwise local pattern exploration for sensitivity analysis. *VAST*. IEEE, 2011.
- [15] P. Hoffman, G. Grinstein, K. Marx, I. Grosse, and E. Stanley. Dna visual and analytic data mining. *VIS*, pages 437–ff., Los Alamitos, CA, USA, 1997. IEEE.
- [16] P. Hoffman, G. G. Grinstein, and D. Pinkney. Dimensional anchors: A graphic primitive for multidimensional multivariate information visualizations. In *Workshop on New Paradigms in Information Visualization and Manipulation*, pages 9–16, 1999.
- [17] J. Horn. *Multicriteria decision making and evolutionary computation*. Technical report, 1996.
- [18] A. Inselberg and B. Dimsdale. Parallel Coordinates: a Tool for Visualizing Multi-Dimensional Geometry. *VIS*, pages 361–378, Los Alamitos, CA, USA, 1990. IEEE.
- [19] I. Jolliffe. *Principal Components Analysis 3rd ed.* Springer, 2002.
- [20] T. Kohonen. *Self-Organizing Maps*. Springer, 3rd edition, 2001.
- [21] T. Kohonen, J. Hynninen, J. Kangas, and J. Laaksonen. SOM PAK: The Self-Organizing Map Program Package. Technical Report A31, Helsinki University of Technology, 1996.
- [22] M. Köksalan, J. Wallenius, and S. Zionts. *Multiple Criteria Decision Making: From Early History to the 21st Century*. World Scientific, 2011.
- [23] P. Korhonen and J. Wallenius. Visualization in the multiple objective decision-making framework. In *MultiObjective Optimization: Interactive and Evolutionary Approaches*, LNCS 5252, pages 195–212. Springer-Verlag, Berlin, 2008.
- [24] S. Lloyd. Least squares quantization in pcm. *Information Theory, IEEE Transactions on*, 28(2):129–137, mar 1982.
- [25] A. V. Lotov, V. A. Bushenkov, and G. K. Kamenev. *Interactive Decision Maps: Approximation and Visualization of Pareto Frontier*. Kluwer Academic Publishers, Norwell, MA, 2004.
- [26] A. V. Lotov and K. Miettinen. Visualizing the pareto frontier. In *MultiObjective Optimization: Interactive and Evolutionary Approaches*, LNCS 5252, pages 213–243. Springer-Verlag, Berlin, 2008.
- [27] C. Mattson, A. Mullur, and A. Messac. Smart Pareto Filter: Obtaining a Minimal Representation of Multiobjective Design Space. *Engineering Optimization*, 36(4):721–740, 2004.
- [28] R. Mayer, T. Lidy, and A. Rauber. The map of mozart. In *ISMIR*, pages 351–352, 2006.
- [29] S. Obayashi, S. Jeong, and K. Chiba. Multi-objective design exploration for aerodynamic configurations. In *35th AIAA Fluid Dynamics Conference and Exhibit*, 2005.
- [30] B. D. Reyck, Y. Grushka-Cockayne, M. Lockett, S. R. Calderini, M. Moura, and A. Sloper. The impact of project portfolio management on information technology projects. *International Journal of Project Management*, 23(7):524–537, 2005.
- [31] T. Schreck, J. Bernard, T. Tekušová, and J. Kohlhammer. Visual cluster analysis of trajectory data with interactive Kohonen maps. *Palgrave Macmillan Information Visualization*, 8:14–29, 2009.
- [32] P. Shirley, S. Marshner, M. Ashikhmin, M. Gleicher, N. Hoffman, G. Johnson, T. Munzner, E. Reinhard, K. Sung, W. B. Thompson, P. Willmesen, and B. Wyvill. A. K. Peters, MA, USA, 2009.
- [33] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages*, pages 336–343. IEEE, 1996.
- [34] E. Tufte. *Envisioning information*. Graphics Press, CT, USA, 1990.
- [35] J. Vesanto. SOM-based data visualization methods. *Intelligent Data Analysis*, 3(2):111–126, 1999.