Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

An Effective Measure to Identify Meaningful Concepts in Engineering Design Optimization

Felix Lanfermann, Sebastian Schmitt, Stefan Menzel

2020

Preprint:

This is an accepted article published in IEEE Symposium Series on Computational Intelligence (SSCI). The final authenticated version is available online at: https://doi.org/10.1109/SSCI47803.2020.9308484 Copyright 2020 IEEE

An Effective Measure to Identify Meaningful Concepts in Engineering Design Optimization

Felix Lanfermann
Honda Research Institute Europe
Offenbach, Germany
felix.lanfermann@honda-ri.de

Sebastian Schmitt
Honda Research Institute Europe
Offenbach, Germany
sebastian.schmitt@honda-ri.de

Stefan Menzel
Honda Research Institute Europe
Offenbach, Germany
stefan.menzel@honda-ri.de

***Abstract*—Identifying similar solutions during an engineering design process and organizing the design data set into several concepts has substantial benefits. A concept is an abstract representation of design solutions that share comparable properties and behavior. Inspecting such concepts facilitates an increase of knowledge about the structure of the design problem. Concepts also allow for the selection of archetypal representatives which can be used as prototypes for further processing. Each prototype represents a different part of the design domain and can, for example, be effectively used to initialize the starting population of a design optimization leading to meaningful variations towards increased design performance. However, identifying meaningful concepts in a large engineering design data set and objectively quantifying the quality of the identified set of concepts is a challenging task. Existing measures to evaluate concepts of design solutions exhibit substantial drawbacks as they do not consider the simultaneous existence and interactions of multiple concepts on one design data set thoroughly. Therefore, we propose a new measure for objectively quantifying the quality of a set of concepts which explicitly takes the overlap and sizes of the concepts into account. We show the benefits of our measure by comparing it to state-of-the art measures in an automated optimization-based concept identification approach for a real-world inspired engineering design data set.**

***Index Terms*—Design Optimization, Design Concepts, Data Mining, Concept Analysis, Prototype Theory**

I. Introduction

In the context of an engineering design process, a multitude of different design candidates are created and analyzed. Much insight can be generated if reasonable patterns from the typically very large design data sets and the design problem at hand can be uncovered. The identification of different meaningful design concepts constitutes a very useful aspect in that regard. A design concept constitutes an abstract representation of a set of designs sharing similar properties in terms of their design parameters along with comparable performance measures and behavior [1]–[3]. As suggested by philosophical investigations [4] and Prototype Theory, design concepts can provide valuable insight into the design problem [5], and the most typical examples within the concept can be a good basis to start looking for suitable variations. Such archetypes share a substantial amount of features with neighboring design candidates and can be used as prototypes in further processing, for example, as initial starting points for subsequent design optimization studies under changed environmental conditions, or case-based reasoning approaches [6]. Prototypes from different design concepts characterize different parts of the search space and therefore create improvement potential in multiple directions.

978-1-7281-2547-3/20/$31.00 ©2020 IEEE

In a given engineering design data set, the designs are characterized in multiple description spaces. The most common description spaces are the space of design parameters which are used to create the design, e.g. parameters of the parametric CAD representation, and the objective space which comprises all considered performance values. Depending on the actual design problem, additional description spaces might be useful, such as statistical or geometrical features derived from the design, or flow field properties of a given design in case of fluid dynamic assessment. Identifying concepts amounts to finding different groups of designs which share similar properties in all description spaces simultaneously. So the designs from one concept should have similar design parameters as well as similar performance values. For finding meaningful concepts, an automated framework is highly desirable due to the large sizes of engineering design data sets and the complexity of the representations in the different descriptions spaces, which effectively prohibit a manual concept identification.

Technically, concept identification can be viewed as a type of clustering problem. A design is represented by a vector comprising its characterizations in all description spaces, e.g. the design parameters along with the corresponding performance values. A concept amounts to a set of corresponding clusters, one in each description space. Considering the clustering of the space of complete design vectors, the clustering must be preserved in all sub-spaces defined by the individual description spaces, i.e. several specific projections of the full space. This renders the direct application of typical clustering approaches such as k-means [7] or k-means++ [8] difficult. Additionally, the identification of a concept with a cluster in a technical sense is also problematic. It is not expected that every design can be associated to one and only one concept. There might be designs which do not belong to any concept as well as designs which might be associated with two or more concepts as they are in the transition region between the concepts. Classical clustering algorithms do not allow for such overlap of clusters or unclassified design candidates. Fuzzy c-means [9] and Gaussian mixture models [10] allow for less strict separations, but they are also not well-suited

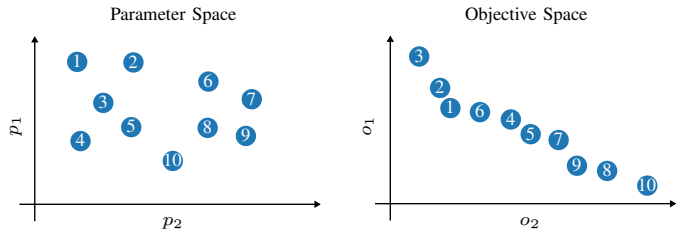


Fig. 1. Fictional example data set: 10 numbered sample design solutions (blue circles) are illustrated in 2D parameter and 2D objective space.

to account for the special structure of multiple descriptions spaces. Approaches from the field of subspace clustering [11] are tailored to discover clusters in high dimensional sparse data, mostly by identifying clusters within lower dimensional projections of the initial data space. Unfortunately, these lower dimensional projections cannot be predetermined by the user and thus are also not well-suited for concept identification.

In order to objectively asses the quality of a set of identified concepts for a given engineering design problem, a suitable measure is required. In the field of concept identification, several measures have been proposed and examined [12]. However, those proposed measures do not consider the cases of multiple concepts and their interaction, i.e. their overlap and size distribution, thoroughly. We propose a novel measure which balances the number of data samples in each concept and the overlap between concepts. The measure can thus be used in a computational optimization for automatically determining representative design concepts. We show the efficiency of this measure on an illustrative example as well as on a real-world inspired vehicle design optimization.

The remainder of the paper is structured as follows: Section II summarizes existing measures to define and evaluate concepts within description spaces, followed by section III which introduces a new measure that allows to assess the distribution of multiple concepts within multiple spaces. In section IV, a series of experiments on a data set containing car rear designs is described and discussed, which compare the aforementioned measures. In section V we discuss the adaptability of the measure towards desired concept sizes. The paper is concluded in section VI with a brief summary of the findings and discussion on future research.

II. Data-driven Concept Evaluation Measures

In the following, we review existing measures from literature which have been proposed to find and numerically evaluate concept candidates from data sets. We focus on only two description spaces where a design consists of pairs of parameters and corresponding performance values (figure 1).

Several objective measures to define concepts as groups of individuals with common characteristics and to evaluate their relevance (e.g. concept consistency) have been developed. In contrast to subjective measures, they transparently assign numerical values independent on human input. In the following, we review selected measures from state of the art.

For a consistent layout and easier understanding, we directly follow the notation of [2] where the description of design concepts is interpreted as a formulation of rules. A rule can be written in the abstract form of *if A (antecedent), then C (consequent)*, or in the compact notation $A \rightarrow C$.

When applying this notation to traditional design optimization [13], [14], the antecedent A relates to the representation of solutions in one description space, for example the parameter space as description of the design space. The antecedent would therefore describe the shape of the design. Accordingly, the consequent C would describe the solutions in another space that is dependent on the first space, here the objective space of the optimization task, e.g. the physical behavior of each design shape. The rule notation could then be loosely translated to: If the design solutions looked like A, then they will behave like C.

A. Coverage, Support, Confidence

Many of the relevant objective measures to evaluate concepts are summarized and thoroughly investigated in [12]. The basic measures *coverage*, *support* and *confidence* form the foundation of the contemporary measures. With regard to the introduced rule based formulation, *coverage* is defined as

$$\mathrm{cov}(A \rightarrow C, \mathbf{D}) = P(A), \tag{1}$$

where $\mathbf{D}$ is the data set of designs under investigation and the probability $P(A)$ quantifies the likelihood that a design meets all conditions in A. If a sample of the data set was to be drawn, coverage would hence be the probability to end up with a solution out of the group that is defined by A. Note that coverage only refers to the probability in one space, the parameter space in our optimization scenario.

Both description spaces are, however, relevant to the *support*, which is defined as

$$\mathrm{supp}(A \rightarrow C, \mathbf{D}) = P(AC), \tag{2}$$

where the probability $P(AC)$ quantifies the likelihood that a design meets all conditions in A and C. Using the same analogy, support is the probability to draw a sample solution from a data set that belongs to a group which is defined by A in the first space and another group that is defined by C in the second space. Therefore, the design shape of the sample solution is described by A, the physical behavior by C.

The *confidence* of a concept is defined by the relation between these two measures as

$$\mathrm{conf}(A \rightarrow C, \mathbf{D}) = \frac{\mathrm{supp}(A \rightarrow C, \mathbf{D})}{\mathrm{cov}(A \rightarrow C, \mathbf{D})} = P(C \mid A) = \frac{P(AC)}{P(A)}. \tag{3}$$

The relation describes how many samples of a particular group in one space can also be found within a specific second group in another description space. It can therefore be interpreted as an indicator for the consistency of a concept. For design optimization tasks this can be interpreted in the following way: If a group of design solutions is defined by geometrical constraints, high confidence means that geometrically similar samples, i.e. with numerically similar parameters, also lead to similar behavior in the objective space.

B. Interestingness and Significance

The confidence measure alone can, nevertheless, produce counter-intuitive evaluation results for specific situations, i.e. produce rules that do not make sense [12]. Hence, [12] introduced a measure that also considers the inverse confidence $\mathrm{conf}(C \rightarrow A, \mathbf{D})$. They interpret the correlation between antecedent and consequence as represented by confidence and inverse confidence as an indicator for both, the interestingness (I) and significance (S) of a pattern. The so called *IS* of a concept is defined as the geometric mean of both confidence values,

$$\begin{aligned} \mathrm{IS}(A \rightarrow C, \mathbf{D}) &= \sqrt{\mathrm{conf}(A \rightarrow C, \mathbf{D}) \cdot \mathrm{conf}(C \rightarrow A, \mathbf{D})} \\ &= \sqrt{\frac{P(AC)}{P(A)} \cdot \frac{P(CA)}{P(C)}}. \end{aligned} \tag{4}$$

There are further extensions and modifications to the above mentioned measures. As one reasonable extension, a Utility measure is introduced in [2], combining IS with the hypervolume of solutions in the performance space. In the remainder of this work, we focus on the effect that multiple different concepts have on each other and how this can be represented in a measure.

III. A Novel Measure for Concept Evaluation

The above mentioned measures generally provide good options to quantify the quality of a design concept for many cases. However, there are situations where these measures may lead to non-intuitive or undesired evaluations. In particular when looking at multiple concepts at once, the confidence and IS measure do not take the overlap between concepts or their sizes into account. In design optimization, the identification of strongly overlapping concepts would not be in favor of the designer, because they would characterize the same associations between parameter and objective space and no further insight is gained. Moreover, very small or very large groups of samples represent trivial associations. For example, a "concept" consisting of only one sample would be labeled highly interesting, as it is very consistent, however, a single sample can hardly qualify as a concept under our definition.

A. Restricted Interestingness

As a remedy for concept overlap, we propose a modification of the IS measure, which, instead of using the full support, only utilizes a restricted support by calculating the probability to draw a sample that satisfies three conditions: a) Being part of a specific group in one space, b) also being part of a specific group in the second space, and c) not being part of another group in the first space. The first two conditions describe the regular support, the third decreases the probability if concept candidates overlap. We introduce an index notation for the restricted support and the proposed new measures in order to distinguish between different concept candidates i. The restricted support as described above is then formulated as $\mathrm{supp}_{r,i}(A \rightarrow C, \mathbf{D}) = P(A_i C_i \backslash A_{\bar{i}})$, where A_i refers to the antecedent for concept candidate i, C_i to the consequence of candidate i and $A_{\bar{i}}$ denotes the antecedent of all candidates but i.

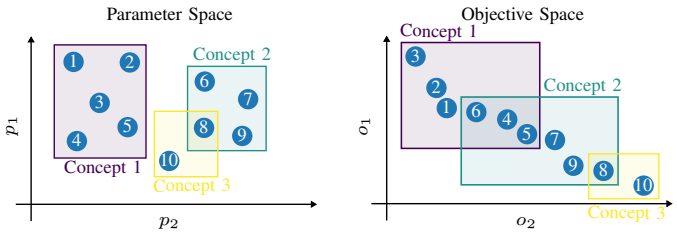


Fig. 2. Example distribution of samples and concepts: The 10 sample solutions (blue circles) of a design data set are grouped into 3 concepts (purple, green and yellow rectangles) in parameter and objective space.

Figure 2 shows an example distribution of 10 design samples which are split into 3 concept candidate groups (1:purple, 2:green, 3:yellow). The yellow concept consists of samples 8 and 10 in both spaces. Therefore, the support for the yellow concept is $P(A_3 C_3) = 2/10$. The restricted support only takes sample 10 into account, because sample 8 is also part of the green concept, hence $P(A_3 C_3 \backslash A_{\bar{3}}) = 1/10$. With the inverse restricted support defined analogously, the modification of the IS measure then becomes

$$\mathrm{RI}_i = \sqrt{\frac{P(A_i C_i \backslash A_{\bar{i}})}{P(A_i)} \cdot \frac{P(C_i A_i \backslash C_{\bar{i}})}{P(C_i)}}. \tag{5}$$

B. Size Aware Restricted Interestingness

One problem still remains when using the RI measure to evaluate concepts. As described above, the relative size of a concept candidate plays an important role for design optimization. Both, very small and very large concept candidates may not be relevant for the designer. Very small concepts often do not represent valid clusterings of the problem and very large concepts tend to be trivial.

We propose to use a smooth scaling function which is used to set a corridor for the desired size of the concepts,

$$f_s(x) = \begin{cases} \sqrt{1 - \left(\frac{x-p}{p}\right)^2}, & \text{if } x < p, \\ 1, & \text{if } p < x < 1-p, \\ \sqrt{1 - \left(\frac{x-1+p}{p}\right)^2}, & \text{if } x > 1-p, \end{cases} \tag{6}$$

where x is the probability to draw a sample from one specific group in one description space $P(A)$ or $P(C)$. The parameter p can be used to favor a desired size of concept distribution (figure 3). For example, for $p = 0.2$, the scaling function $f(x)$ equals 1 if x lies between 0.2 and 0.8. For values of x below 0.2 and above 0.8, the function smoothly decreases down to 0. The size aware restricted interestingness measure uses the scaling function twice, for $P(A_i)$ and $P(C_i)$, and is defined as

$$\mathrm{SARI}_i = \mathrm{RI}_i \cdot f_s(P(A_i)) \cdot f_s(P(C_i)). \tag{7}$$

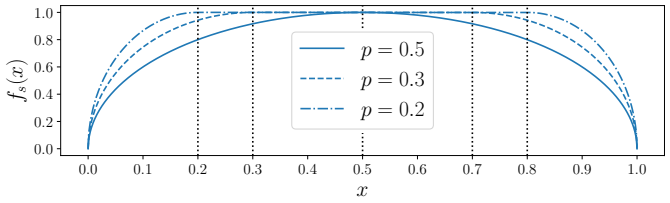


Fig. 3. Scaling function $f_s(x)$: The function is depicted for three different values of p (blue lines). The dotted lines indicate the boundaries with respect to p between where the function equals one.

For example, the SARI for the yellow concept of figure 2 is calculated as

$$\mathrm{SARI}_3 = \sqrt{\frac{1/10}{2/10} \cdot \frac{1/10}{2/10}} \cdot 0.917 \cdot 0.917 = 0.420. \quad (8)$$

C. Comparison of the Measures

To demonstrate how differently the measures evaluate concepts, a comparison was conducted. The investigated fictional exemplary design data set $\mathbf{D}$ contains 10 samples which are grouped into 3 concept candidates in the parameter space and objective space of the set (figure 2). For all three concepts, 1 (purple), 2 (green) and 3 (yellow), the measures coverage, support, confidence, IS, RI and SARI were calculated (table I).

While the $\mathrm{conf}(A \to C, \mathbf{D})$ equals 1 for all concepts, IS also takes into account the inverse $\mathrm{conf}(C \to A, \mathbf{D})$ and is therefore able to quantify the smaller consistency of the concepts in the objective space. The yellow concept candidate shows the highest (actually: ideal) IS value, since all samples lie within both concept regions in the different spaces.

In contrast to the IS, the RI measure finds a different order of concept quality. Since all candidates overlap with others in at least one space, all values are lower than the corresponding IS. The biggest difference can be seen for the yellow candidate, because half of the samples can also be found in other candidates.

The SARI measure was calculated for $p = 1/3$ and therefore decreases the RI value for all candidates that are smaller than 1/3 of the overall number of samples. This applies only to the yellow candidate. It can be seen that the presented measures evaluate the same candidate configuration very differently. We argue that the SARI measure leads to the most desirable order, and also the absolute values of the measure reflect the naive expectation of the identified concepts.

IV. Design Optimization Experiments

In the following we will demonstrate our proposed measure in a real-world inspired design optimization application to identify optimal distributions of concepts.

A. Description of the Investigated Data Set

For the design scenario, realistic data were created, where each sample is a two-dimensional profile that resembles the rear of a car (figure 4). The design parameters of each sample consist of two parameters specifying the distance d of a wing structure from the rear basis and the rotation angle r of the wing. The object files of the design candidates were created using Blender[1] and their aerodynamic behavior was evaluated with OpenFoam[2]. As performance values, the lift and drag coefficients were calculated during the post processing of the simulations defined as $C_{\mathrm{lift}} = \frac{2F_{\mathrm{lift}}}{\rho\nu^2 A_{\mathrm{ref}}}$ and $C_{\mathrm{drag}} = \frac{2F_{\mathrm{drag}}}{\rho\nu^2 A_{\mathrm{ref}}}$, where $\rho = 1.2\,\mathrm{kg\,m^{-3}}$ is the fluid density of air, $A_{\mathrm{ref}} = 1\,\mathrm{m^2}$ the reference surface area and $\nu = 10\,\mathrm{m\,s^{-1}}$ the chosen flow speed of the simulation. The data set consists of 1000 random samples that are represented in a two-dimensional parameter space (d and r) and a two-dimensional objective space (C_{lift} and C_{drag}).

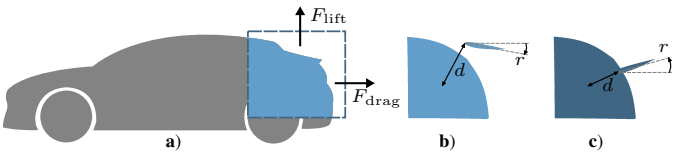


Fig. 4. Application background and samples of the data set: a) The samples resemble the rear of a car, the two forces of interest are F_{lift} and F_{drag}. The samples are defined by distance d and rotation r of the wing structure. Samples with either b) detached or c) attached wing structures are feasible.

B. K-means Clustering Tests

To illustrate the problem further and demonstrate why regular approaches such as k-means clustering are not feasible for this application, we carried out a preliminary experiment in which we try to identify concepts using a standard k-means++ implementation from scikit-learn[3]. We let the algorithm split the samples of the 4-dimensional feature space (d, r, C_{lift} and C_{drag}) into 3 clusters and projected them into the two 2-dimensional spaces that we are interested in (figure 5). In contrast to the 4-dimensional feature space, the clusters are not clearly separated in the projected two-dimensional parameter and objective space. In the parameter space, an overlap of the green and yellow clusters can be observed, and in particular in the objective space, all three clusters overlap strongly. Suitable design concepts cannot be identified, because there is no clear relation between the properties of the samples (described by the parameter space) and their physical behavior (described by the objective space). For example, even the cluster centers (marked with a star in figure 5) of the green and purple cluster lead to very similar objective values despite coming from very different regions in the parameter space.

If we allow a "limited scope"[4] for the concept candidates, we are still not able to produce distinct concepts with the k-means algorithm (figure 5). With this, we explicitly confirm the view expressed in the introduction, that classical clustering algorithms, like k-means cannot be used to identify meaningful concepts. Therefore, a measure such as the proposed SARI, which is tailored to solve this specific problem, is needed.

[1] www.blender.org
[2] www.openfoam.com
[3] scikit-learn.org
[4] Only the samples that lie within a defined distance around the cluster centers are considered to be part of the respective concept candidates.

TABLE I
EVALUATION RESULTS FOR EACH CONCEPT CANDIDATE SHOWN IN FIGURE 2

	Cov		Supp	Conf		IS	RI	SARI
	$A \to C$	$C \to A$		$A \to C$	$C \to A$			
Concept 1	5/10	6/10	5/10	1	0.83	0.91	0.71	0.71
Concept 2	4/10	6/10	4/10	1	0.67	0.82	0.5	0.5
Concept 3	2/10	2/10	2/10	1	1	1	0.5	0.42

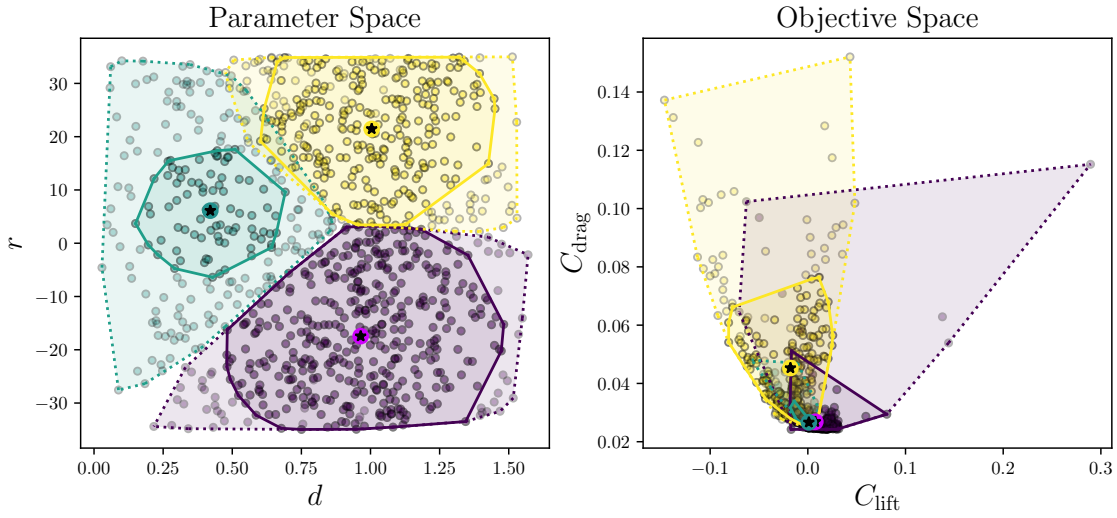


Fig. 5. K-means clustering of the car rear data set: The objects in the 4-dimensional feature space (d, r, C_{lift} and C_{drag}) are separated into 3 clusters and projected back to the parameter and objective space. The dotted lines denote the convex hulls of the clusters in the two spaces and resemble the concept candidates. The stars mark the cluster centers. The solid lines encompass only the samples that lie within 35% of the maximum inner cluster distance from the cluster center. For both cases, at least in the objective space, the candidates overlap strongly.

C. Concept Identification Tests

All the discussed measures quantify the quality of any given concept candidates. In order to identify the most desirable distribution of concepts as favored by each measure, it is sensible to use them directly as an evaluation function for an optimization. Therefore, we carried out a series of independent optimization runs with the IS, RI and SARI measure, always using an (8, 24) CMA-ES [15] for 3000 generations. To realize a simple implementation, we chose to optimize 3 concept candidates with rectangular boundaries for each of the measures. Hence, the parameter space of the optimization consists of the 24 boundary elements that define the candidate regions in both data set spaces. The single objective of the optimization is the equally weighted sum of the measures for each concept candidate. Differently initialized optimizations showed in principle the same qualitative behavior and we therefore only discuss one typical example for each case in the following.

Since IS evaluates each concept candidate individually and does not consider concept overlap, as expected all three concept candidates overlap highly and contain almost the entire data set in both spaces each (figure 6, table II), while having almost converged to optimal IS measures for all three concepts (table III). Due to the overlap, it can not be clearly differentiated between the three concept candidates and we argue that the requirements to qualify as a design concept are thus not met by any of the candidates.

The RI measure on the other hand produces candidates that are distinguishable in both spaces and hardly overlap (figure 7). However, the size of the candidates is extremely unbalanced (table II). The green concept contains only one sample, having converged to the optimal RI value of 1 (table III).

The optimization with the SARI measure leads to distinguishable and more equally distributed concepts (figure 8, table II). For the scaling function f_s, the parameter $p = 1/3$ was chosen in order to provide an incentive to the optimization to split the concept roughly equally. Concept 1 and 2 are slightly punished by the scaling function with an overall factor of 0.94 and 0.88, respectively. The SARI measure is generally lower for all concepts compared to the RI experiment, though we argue that the identified candidates are more reasonable and more likely to qualify as design concepts. As a reference, the SARI values for the k-means clustering approach (with limited scope) can also be obtained from table II.

In addition to identifying sensible design concepts and increasing domain knowledge, another important impetus is the discovery of archetypal concept representatives. For the concepts discovered with the SARI measure, the geometric

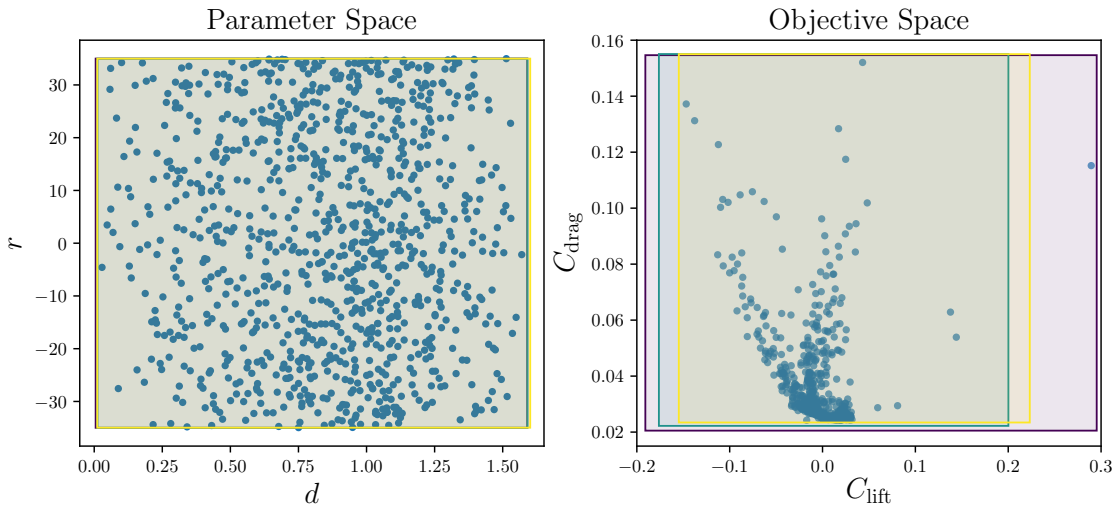


Fig. 6. IS evaluation: All three concept candidates (purple, green and yellow rectangles) highly overlap and contain almost the entire data set in both spaces.

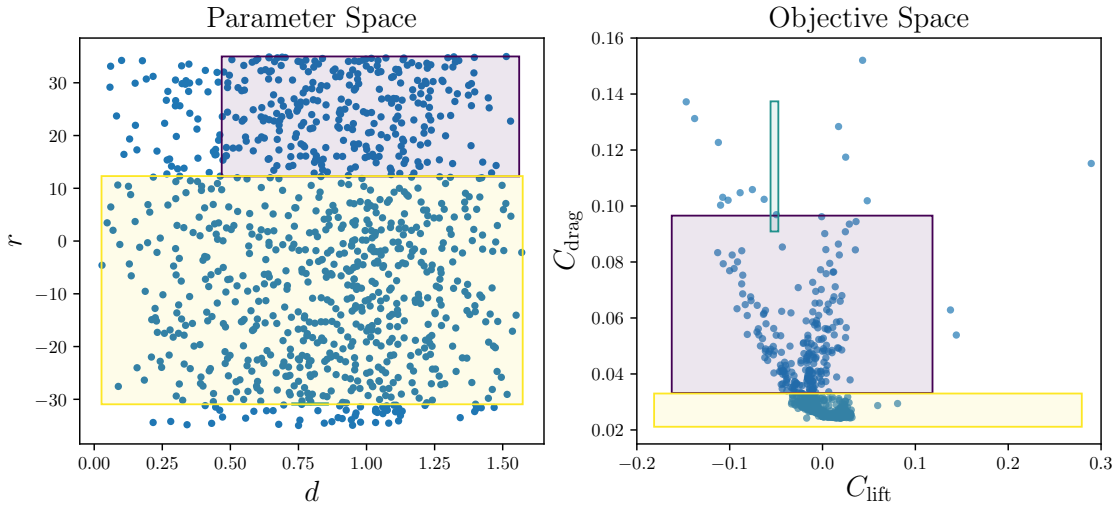


Fig. 7. RI evaluation: The three concept candidates (purple, green and yellow rectangles) hardly overlap, but contain very different amounts of samples. The green candidate for example only contains a single sample in both spaces and is hence hardly visible. In the parameter space it is located roughly at $d = 0.75$ and $r = -30$.

mean of the concept region boundaries in parameter space were chosen as example representatives (figure 8). The 3 samples show different properties and can intuitively be considered as archetypes of their concepts. Of course, other methods than the geometric mean could be used to select archetypes [16], [17].

In summary, the experiments show that if the IS measure was used to optimize the distribution of potential design concepts in a data set, the optimization may result in an undesired outcome. Since IS evaluates each candidate separately and does not take overlap into account, it is feasible for the optimization to join all samples in all spaces into one big "concept" (figure 6). The RI measure can rule out overlap, however, the resulting candidates might consist of very different amounts of samples, even as little as one sample per candidate (figure 7). Using the SARI measure, overlap can also be avoided and the number of samples per concept are likely to be similar (figure 8). The scores of the measures (table III) show why the concept candidates are distributed the way they are. In case of IS, all scores are nearly ideal (1), this shows that optimization with IS leads to the above described overlap behavior. In case of RI, the metric for concept 2 (green) is 1, so the optimization discovered an optimum. In case of SARI, all metric scores are lower than in both other experiments, however the result is more reasonable.

Moreover, the SARI-optimized concept boundaries show that the concepts are aligned with the observed correlation between the rotation parameter r of the wing structure and the drag coefficient C_{drag} in the objective space. The drag of the design element depends to a high extent on the rotation

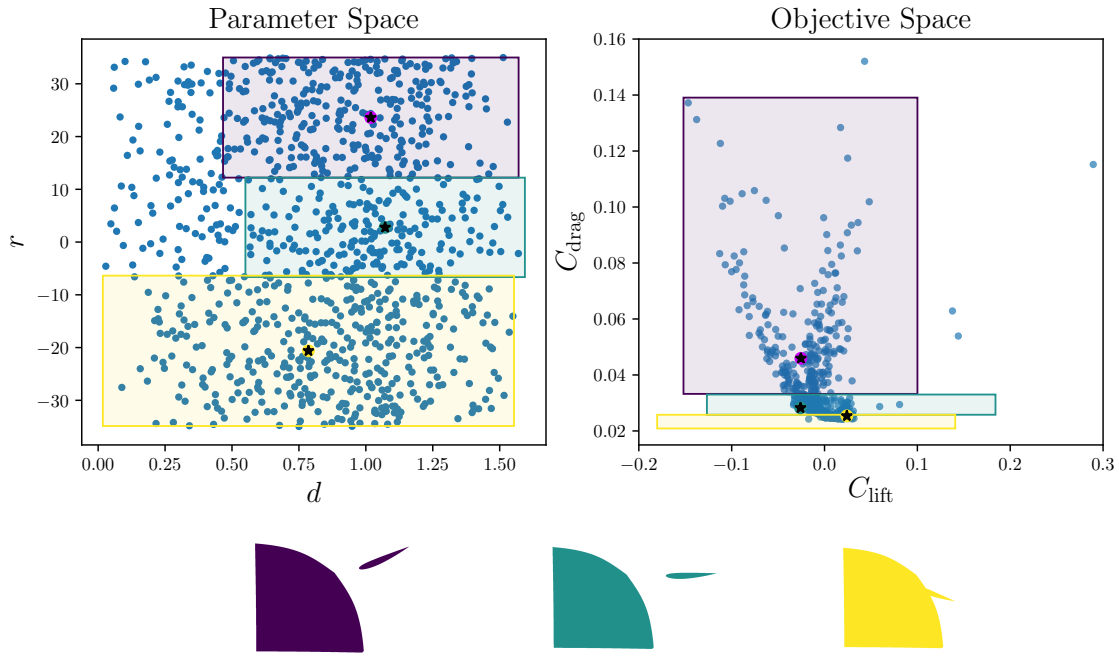


Fig. 8. SARI evaluation: The three concept candidates (purple, green and yellow rectangles) hardly overlap and consist of a more similar amount of samples compared to the RI experiment result (figure 7). For the scaling function, $p = 1/3$. The three archetypes are marked with a star in parameter and objective space and displayed below.

TABLE II
NUMBER OF SAMPLES PER CONCEPT CANDIDATE IN PARAMETER SPACE AND OBJECTIVE SPACE FOR EACH OF THE DIFFERENT MEASURES AFTER 3000 GENERATIONS OF THE OPTIMIZATION

	IS		RI		SARI	
	Parameter	Objective	Parameter	Objective	Parameter	Objective
Concept 1	1000	1000	287	218	288	231
Concept 2	1000	999	1	1	204	229
Concept 3	1000	999	610	764	401	535

TABLE III
IS, RI AND SARI SCORES OF THE BEST CONCEPT CANDIDATES AND THE SARI SCORE FOR K-MEANS CLUSTERING AS A REFERENCE.

	IS	RI	SARI	SARI (k-means)
Concept 1	1.0000	0.7476	0.7274	0.3835
Concept 2	0.9995	1.0000	0.4597	0.0
Concept 3	0.9995	0.8525	0.7071	0.2535

of the wing and the identified concepts uncover this structure, leading to an increase in knowledge about the problem.

V. INFLUENCE OF THE PARAMETER p IN THE SARI MEASURE

For the presented experiment, the parameter p of the scaling function f_s belonging to the SARI measure was set to the value $p = 1/3$. The intention behind this was to incentivise the optimization to divide the complete set of samples into the three roughly equally sized concept candidates.

In order to investigate how strongly this incentive is adopted, a series of additional experiments with the SARI measure was conducted. For six different configurations ($p = [0.0, 0.1, 0.2, 0.3, 0.4, 0.5]$), 10 optimizations with randomly initialized particle swarm optimization [18], each having 10 particles, were run for 1000 generations. Each run generates 3 concept candidates, thus for every fixed value of p, there are 30 possible candidates. In some situations, the optimization cannot escape local minima or plateaus within the given computational budget, resulting in candidates with very low SARI scores. 26 out of the overall 180 concept candidates contained less than five design samples in both spaces, which have been ignored for the statistical analysis.

The assumption that the variance of concept size will be smaller for higher values of p due to the nature of the scaling function can be confirmed (figure 9). The difference between the biggest and smallest candidates becomes smaller for higher values of p. The red areas illustrate the region of concept candidate sizes that are penalized by the scaling function. Another property of the SARI measure can be observed from

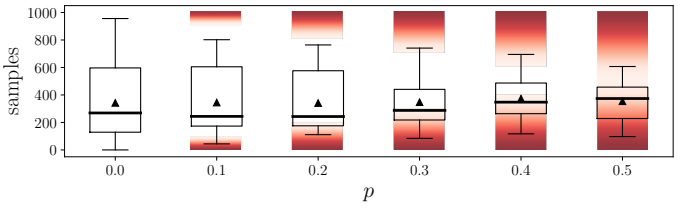


Fig. 9. Distribution of samples for variations of p: 10 optimization runs for each parameter value were conducted. The red areas represent the concept size ranges that are penalized by the scaling function f_s. The darker the color, the higher the penalty.

these results. For higher values of p, the scaling function pushes the size of each concept toward half the number of points, which would imply strong overlap of the three concept candidates which in turn is penalized by the SARI measure as well. As the overlap penalization is stronger than the scaling function penalization, this trade-off is resolved in favor of the concept sizes and the optimized concept sizes tend to vary around one third of the total number of samples.

VI. Conclusion and Discussion

We propose a new size aware restricted interestingness (SARI) measure, which quantifies the quality and distribution of concept candidates, for an automated design concept identification process. SARI explicitly penalizes the overlap between the concept candidates and incentivises desired concept sizes. The conducted experiments with an aerodynamic design data set demonstrate that the developed measure shows the intended behavior and is able to identify reasonable concepts when used as an objective in an optimization. These concepts lead to an increase in knowledge about the engineering design problem by illuminating correlations between parts of the parameter and objective space. Furthermore, from the discovered concepts, archetypal representatives with characteristic properties can be extracted. It is hence sensible to use the proposed SARI measure for tasks in which these benefits outweigh the larger implementation effort and high computational costs of an evolutionary optimization (in comparison to faster clustering techniques such as k-means).

The effect of the free parameter p of the SARI measure is examined in a second series of additional experiments. We discover that it correlates with the difference between maximal and minimal size of the concept candidates and can thus be used to "push" the concept sizes to desired average values, which can be different for each concept.

Future work will focus on the application of the SARI measure in more realistic engineering applications where the dimensionality of the parameters space as well as the objective space are much higher and the problems structure is much more complex. Additionally, the SARI measure is readily extendable to situations including more than two descriptions spaces. Promising cases to consider are the descriptions spaces of geometric features as for example used in geometric deep learning approaches [19], [20]. For more realistic applications, the current restriction of the candidate concepts to cover rectangular regions only in design and objective space is not very sustainable and allowing for more general shapes is much more appropriate.

We conclude that the SARI measure is a valuable extension to the set of existing concept evaluation measures with a promising potential for future applications. It incorporates multiple desired features during the concept identification process that are not considered by other measures and leads to intuitively meaningful concept distributions useful in engineering design processes, as well as being apt to further extensions in more complex applications.

References

[1] T. Adajian, "On the Prototype Theory of Concepts and the Definition of Art," *The Journal of Aesthetics and Art Criticism*, vol. 63, no. 3, pp. 231–236, 2005.

[2] L. Graening and B. Sendhoff, "Shape Mining: A Holistic Data Mining Approach for Engineering Design," *Advanced Engineering Informatics*, vol. 28, no. 2, pp. 166–185, 2014.

[3] S. Laurence and E. Margolis, "Concepts and Cognitive Science," in *Concepts: Core Readings*, 1999, pp. 3–81.

[4] L. Wittgenstein, *Philosophical Investigations*. Blackwell, 1953.

[5] E. Rosch, "Cognitive reference points," *Cognitive Psychology*, vol. 7, no. 4, pp. 532–547, 1975.

[6] J. Kolodner, *Case-based reasoning*. Morgan Kaufmann, 2014.

[7] J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281—-297, 1967.

[8] D. Arthur and S. Vassilvitskii, "k-means++: The Advantages of Careful Seeding," *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1027–1035, 2007.

[9] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Boston, MA: Springer US, 1981.

[10] C. E. Rasmussen, "The infinite Gaussian mixture model," *Advances in Neural Information Processing Systems*, no. 1, pp. 554–559, 2000.

[11] L. Parsons, E. Haque, and H. Liu, "Subspace clustering for high dimensional data," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 90–105, 2004.

[12] P. Tan and V. Kumar, "Interestingness Measures for Association Patterns: A Perspective," in *KDD Workshop on Postprocessing in Machine Learning and Data Mining*, 2000.

[13] J. Kmec and S. Schmitt, "Exploring the Fitness Landscape of a Realistic Turbofan Rotor Blade Optimization," in *Proceedings of the 6th International Conference on Engineering Optimization*, 2019, pp. 510–522.

[14] D. Sieger, S. Menzel, and M. Botsch, "A Comprehensive Comparison of Shape Deformation Methods in Evolutionary Design Optimization," in *Proceedings of the International Conference on Engineering Optimization*, 2012, pp. 1–5.

[15] N. Hansen and A. Ostermeier, "Completely Derandomized Self-Adaptation in Evolution Strategies," *Evolutionary Computation*, vol. 9, no. 2, pp. 159–195, 2001.

[16] N. Dommaraju, M. Bujny, S. Menzel, M. Olhofer, and F. Duddeck, "Identifying Topological Prototypes using Deep Point Cloud Autoencoder Networks," in *International Conference on Data Mining Workshops (ICDMW)*, 2019, pp. 761–768.

[17] A. Hagg, A. Asteroth, and T. Bäck, "Prototype Discovery Using Quality-Diversity," in *Parallel Problem Solving from Nature – PPSN XV*, 2018, pp. 500–511.

[18] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, 1995, pp. 1942–1948.

[19] T. Rios, T. Bäck, B. van Stein, B. Sendhoff, and S. Menzel, "On the efficiency of a point cloud autoencoder as a geometric representation for shape optimization," in *IEEE Symposium Series on Computational Intelligence*. IEEE, 2019.

[20] T. Rios, B. van Stein, S. Menzel, T. Bäck, B. Sendhoff, and P. Wollstadt, "Feature visualization for 3d point cloud autoencoders," in *International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020.