

ePub^{WU} Institutional Repository

Edward Bernroider and Nikolaus Obwegeser and Volker Stix

Analysis of Heuristic Validity, Efficiency and Applicability of the Profile Distance Method for Implementation in Decision Support Systems

Article (Accepted for Publication)
(Refereed)

Original Citation:

Bernroider, Edward and Obwegeser, Nikolaus and Stix, Volker (2011) Analysis of Heuristic Validity, Efficiency and Applicability of the Profile Distance Method for Implementation in Decision Support Systems. *Computers and Operations Research*, 38 (5). pp. 816-823. ISSN 0305-0548

This version is available at: <http://epub.wu.ac.at/3038/>

Available in ePub^{WU}: March 2011

ePub^{WU}, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.

This document is the version accepted for publication and — in case of peer review — incorporates referee comments. There are minor differences between this and the publisher version which could however affect a citation.

Analysis of Heuristic Validity, Efficiency and Applicability of the Profile Distance Method for Implementation in Decision Support Systems

Edward Bernroider^a, Nikolaus Obwegeser^{*,b}, Volker Stix^b

^a*Institute for Management Information Systems, WU - Vienna University of Economics and Business, Augasse 2-6, 1090 Wien, Austria*

^b*Institute for Information Business, WU - Vienna University of Economics and Business, Augasse 2-6, 1090 Wien, Austria*

Abstract

This article seeks to enhance acceptance of the profile distance method (PDM) in decision support systems. The PDM is a multiple attributive based decision making as well as a multiple method approach to support complex decision making and uses a heuristic to avoid computationally complex global optimization. We elaborate on the usability of the method and question the heuristic used. We present a bi-section algorithm, which efficiently supports the discovery of transition profiles needed in a user-friendly and practical application of the method. Additionally, we provide empirical evidence showing that the proposed heuristic is efficient and delivers results within 5% of the global optimizer for a wide range of data sets.

Key words: Profile Distance Method, Decision Support Systems, Multi Criteria Decision Making, Heuristics, Usability

1. Introduction

The Profile Distance Method (PDM) introduced by Bernroider and Stix [1] is a combined method for multiple attribute decision making (MADM). It combines the advantages of two well known approaches individually applied in decision making: the utility ranking method (URM) and data envelopment analysis (DEA) [2, 3]. While DEA has a wide field of applications not only in the area of selection problems, URM is an approach specifically designed for decision making [4]. DEA was applied as a single method approach to evaluate the impact of information technology (IT) on multiple stages [5] as well as part of multi-method selection approaches such as DEA to screen alternatives followed by MADM [6], DEA to engage self evaluations and compensate for a lack of preference information followed by MADM based cross evaluations [7], or DEA in combination with goal-programming models [8]. Combined models of Analytic Hierarchy Processes (AHP) and DEA have been extensively proposed, e.g., by [9, 10, 11]. Multi-method approaches have become popular especially in the context of complex IT transformations such as Enterprise Resource Planning (ERP) systems, which is the original application of the PDM. Recent ERP contributions combine artificial neural networks and the analytic network process approach [12] to better support assessments from multiple experts or develop frameworks supported with multiple methods to achieve organisational fit [13, 14], which is also a main goal of the PDM.

Contrary to the optimization methods mentioned, the PDM additionally takes advantage of the inner structure of the multi-dimensional optimizers by calculating distances between a desired system profile (gained e.g. from the additive multi-attribute utility model) and the optimal individual alternative profile calculated by means of the DEA derived optimization process. It utilizes the concept of organizational fit to support organizational decision making in the context of IT transformations. This is achieved by exploring the distance based on attribute weights to a desired product or company profile. This distance provides additional selection information and together with efficiency scores and DEA multipliers help to discriminate between alternatives. Insufficient discrimination has been identified as

*Corresponding author

Email addresses: edward.bernroider@wu.ac.at (Edward Bernroider), nikolaus.obwegeser@wu.ac.at (Nikolaus Obwegeser), volker.stix@wu.ac.at (Volker Stix)

problem hindering its practicability for decision making [15, 16, 17]. The PDM provides the decision maker with a fade-parameter to emphasize the respective importance of the desired profile, enabling the user to fade between the pure utility ranking and data envelopment method. This feature enhances not only discrimination but also enables further uses for example to validate diverse ranking outcomes in decision making [18], which is a critical factor in MADM. The results can be visualized to support the decision maker in justifying and communicating the model outcomes [19].

This article aims at making methods in operations research more accessible to decision makers while providing assurance about the validity of results. More specifically, we focus on two objectives: validity of the heuristic proposed for PDM and applicability of the method as a decision support tool in practice. According to IT acceptance literature based on the well accepted Technology Acceptance Model (TAM) users formulate a positive attitude toward the technology when they perceive it to be useful and easy to use [20]. The resulting behavioral intention of users has been validated in numerous empirical studies as a recent meta study shows [21]. We improve PDM practicability by targeting the mentioned fade operator, which is used to control PDM optimization outcomes. The user is searching for profiles with transition points, where profile distances change as well as according profile structures. Without support the user will face difficulties in finding these transition profiles due to the infinite space of possibilities between a known lower and an unknown upper boundary. This problem prevents a practical computer software implementation of the PDM and constitutes a limitation of the original publication. We therefore propose and test a divide and conquer algorithm in a PDM implementation to identify the upper boundary and generate the list of all profile distance transitions based on a given precision. By calculating the boundaries and transitions of the underlying decision space we allow for the practical use of the PDM in actual decision problems and thereby build a bridge between the theoretical model and practical usage as a computer aided decision support system (DSS). We exemplify the usage of a DSS with PDM methodology in a software prototype and a preliminary empirical validation of its usability. In its original specification, the PDM suggests a heuristic and relaxation of the PDM linear program in order to ensure scalability and applicability. However, there is limited assurance on the validity and efficiency of the suggested heuristic. We therefore also aim to test and validate the suggested PDM heuristic, a pre-condition for implementation in decision support systems. All suggestions are implemented in a computer software application that hides complexities while enabling the decision maker to use and exploit the properties of the underlying method.

The remainder of this article is structured as follows. In the next section we shortly present the Profile Distance Method in order to subsequently introduce an algorithm for finding all profiles in Section 3. Section 4 uses several data sets to validate the efficiency of the original optimization heuristic and in Section 5 we summarize the findings and identify further research issues.

2. The profile distance method

In this section we shortly introduce the PDM for a better understanding of the results presented in the next section. For more details we refer to [1]. The original fractional CCR-DEA model (introduced by Charnes, Cooper and Rhodes in [3]) optimizes the weighted output per weighted input, where the weights are the variables. PDM uses this original CCR-DEA model, which translates the fractional program into a linear program (LP). In the following we have n decision making units (DMUs) each with m input attributes represented through the $m \times n$ matrix X and s output attributes stored in the $s \times n$ matrix Y . The vectors v and u are the weight vectors for input- and output-attributes, respectively. We have for each DMU a different LP which can lead to a different optimal solution. The parameter k selects the DMU for which the optimization should be performed.

$$h_k = \max_{u,v} \sum_{r=1}^s u_r y_{rk} \quad (1)$$

subject to:

$$\begin{aligned} \sum_{i=1}^m v_i x_{ij} - \sum_{r=1}^s u_r y_{rj} &\geq 0 \quad \text{for } j = 1, \dots, n \\ \sum_{i=1}^m v_i x_{ik} &= 1 \\ u_i, v_j &\geq 0 \quad \text{for all } i, j \end{aligned}$$

The PDM replaces the objective (1) with

$$h_k = \max_{u,v} \sum_{r=1}^s u_r y_{rk} - \alpha f(u, w) \quad (2)$$

where f is a penalty function measuring distance between the weight vector u and the desired profile w . Only output weight-profiles are considered and such a desired profile w has to be expressed in relative figures (measuring the relative importance of weights rather than the absolute ones).

Here α is a fade-factor controlling the mentioned tradeoff between pure DEA ($\alpha = 0$) and pure URM ($\alpha \rightarrow \infty$). Hence, the PDM allows the decision maker to change the model outcome by fading between both techniques, thereby exploring the organizational fit of the current alternative under evaluation. The upper boundary of α , however, is data dependent and therefore unknown.

2.1. Heuristic for PDM solutions within a linear program

The PDM does not represent a LP. Bernroider and Stix proposed a heuristic LP solution by choosing the 1-norm, the sum of absolute distances, as the distance measure, i.e. $f(u, w) = \sum_{i=1}^s |u_i - w_i u_1|$ and by relaxing the problem which results in

$$h_k = \max_{u,v} \sum_{r=1}^s u_r y_{rk} - \alpha \sum_{i=1}^s d_i (u_i - w_i u_1)$$

subject to:

$$\begin{aligned} \sum_{i=1}^m v_i x_{ij} - \sum_{r=1}^s u_r y_{rj} &\geq 0 \quad \text{for } j = 1, \dots, n \\ \sum_{i=1}^m v_i x_{ik} &= 1 \\ \alpha d_i (u_i - w_i u_1) &\geq 0 \quad \text{for } i = 1, \dots, s \\ u_i, v_j &\geq 0 \quad \text{for all } i, j \end{aligned} \quad (3)$$

with

$$d_i = \begin{cases} 1, & \hat{w}_i > w_i \\ -1, & \text{else} \end{cases}$$

where \hat{w} is derived in phase one, solving the pure DEA problem with $\alpha = 0$. From that solution the relative weight vector \hat{w} can be extracted [1].

The additional constraints in (3) ensure that the distance measure f is always positive and thus the absolute value sums up to a positive penalty. The relaxation of the problem can be seen to allow the first weight-profile \hat{w} (estimated through DEA) to approximate the desired weight-profile w only from one direction with respect to each attribute. Clearly this heuristic does not have to find the global solution of the original objective (2).

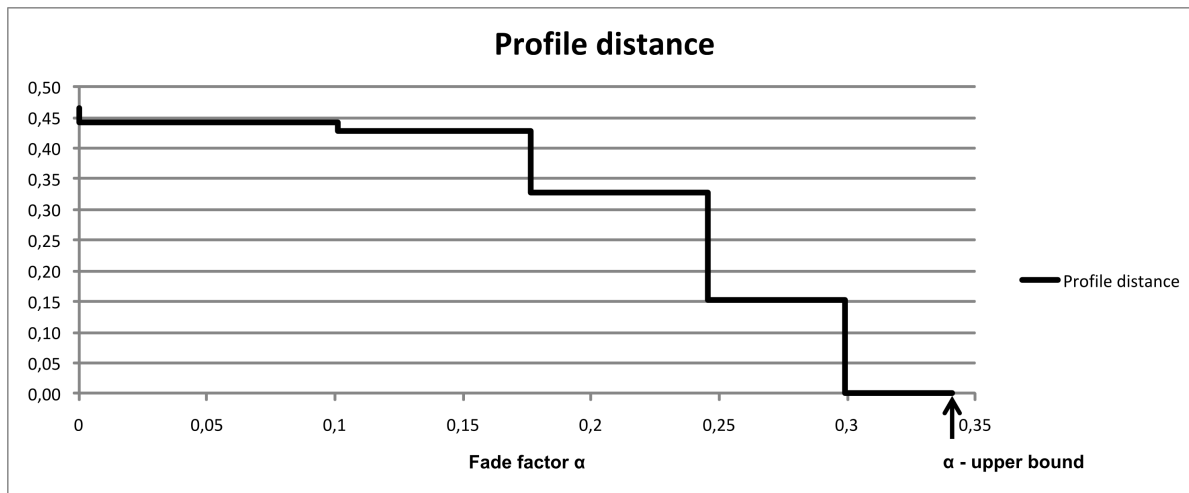


Figure 1: Profile distance distribution with respect to increasing α . Number of discontinuities (=transitions) and their position are unknown.

3. Discovering all profiles

The PDM model contains this fade factor α allowing the decision maker to control the tradeoff between DEA and URM. The lower strict bound of α is obvious and given through pure DEA ($\alpha = 0$), the upper bound, however, when the profile distance f reaches zero and pure URM is reached, depends on the given data. Between these bounds there are unknown many optimal profiles, given the respective α (see Figure 1 for an example). The profile distance f decreases monotone with increasing α . The DM is virtually allowed to test the strength of the alternatives structural composition by increasing the α value. Since the saltus-distribution of the profile distance function is not known, we firstly addressed the issue of exploring the decision space in an analytical manner to find the data specific discontinuities. This step is necessary in order to apply the PDM in practice, since leaving the calculation necessary to find the transitions to the DM would hinder the applicability substantially. We solved this issue by applying a two phase bi-section algorithm approach as presented below. This approach assures results with scalable accuracy while keeping the calculation time and complexity at a minimum. This is required for realtime rendering of the problem space, providing the decision maker ad-hoc feedback on applied variations.

3.1. Bi-section algorithm for discovering all profile distance transitions

The algorithm is divided into two phases. In the first phase, the unknown upper bound U is searched using exponential expansion. In the second phase the whole interval is searched using a divide and conquer principal to perform an exhaustive search. Both algorithms perform in logarithmic time complexity. For a simpler notation we define $F(x)$ to be the distance $f(u^*, w)$ of an optimizer u^* and the desired profile w using the fade-value $\alpha = x$, or in other words the profile distance using $\alpha = x$. Note, that for each call of F one LP needs to be solved.

Phase 1: Estimation of the upper bound. The fade factor alpha, intended to aid the user in varying the profile distance, is initially set to 0. This represents the lower bound of the fade factor, at that point the optimization output of the PDM is consistent with the DEA optimization.

The upper bound of α is where the profile distance becomes zero, that is where it totally matches the user defined desired profile. Due to the nature of the following bi-section algorithm, a coarse upper bound estimation is already satisfactory. The upper bound is found by multiplying the fade factor by 2, with an initial value of 1, until the profile distance reaches 0 (see Algorithm 1). The upper bound is not strict, it will be, however, after using the bi-section algorithm which follows next. There is one optimization to be done within the loop, so the algorithm's time complexity is $O(\log U)$, depending only on the magnitude of the upper bound. Since the upper bound is not large and thus not really problem dependent, the complexity can be considered as $O(1)$.

Algorithm 1:**Input:** Problem description**Output:** Upper bound U $U = 1$ while ($F(U) > 0$) $U = U * 2$ return U

Phase 2: Discovering all transitions. The bi-section method is a divide and conquer algorithm that enables us to generate the list of all profile distance transitions within a given precision ϵ . The implemented method uses a problem queue, where each problem is defined by an interval $[L, U)$ defining L the lower and U the upper bound to be examined. In the course of the calculation, the profile distance in the middle $M = (U + L)/2$ of each interval is used to proceed in one of the following options: If a transition is observed between L and M (i.e. $F(L) > F(M)$), a new problem with the respective interval $[L, M)$ is inserted into the problem queue. The same procedure is applied to the interval $[M, U)$. If no transition is present (i.e. $F(L) = F(U)$), the algorithm continues with the next problem enqueued, if any, or stops otherwise. This approach is pursued until the initially defined significance ϵ is reached (see Algorithm 2 for more detail).

Algorithm 2:**Input:** Lower Bound L , Upper Bound U **Output:** List of Transitions T Problem Queue $P = \phi$ Transition List $T = \phi$ insert $[L, U)$ into P while ($P \neq \phi$) { remove problem $[L, U)$ from P if ($L - U > \epsilon$) if ($F(L) - F((L + U)/2) > \epsilon$) insert $[L, (L + U)/2)$ into P if ($F((L + U)/2) - F(U) > \epsilon$) insert $[(L + U)/2, U)$ into P

else

 insert L into T

}

return T

We conducted our laboratory research on tested and stable code from a recent Java based PDM implementation acquired from the original authors of the method [19].

4. Validating the PDM heuristics

4.1. Exhaustive data simulation

As mentioned before the linear optimization model contains a problem relaxation designed to ensure a positive penalty for the objective value and a considerable improvement as far as calculation time is concerned. There is, however, no guarantee that the global, or at least a good result is achieved.

To validate the process heuristics and their quality we simulated ten different PDM-supported decision making scenarios. We used secondary data to support the diverse model applications including data from OECD's statistical online database [22] including decision problems on public health, OECD member country profiles, general and area specific higher education graduate statistics, population profiles by sex and age, patent statistics, GDP (Gross Domestic Product) statistics of non-member and member countries, and water usage statistics. We also explored an ERP selection case study drawn from the original PDM publication [1] as well as on an evaluation of University departments [23]. All data selected does match the necessary conditions of being output oriented and can be defined as a selection

problem. The simulation was primarily applied (i) to test the formal validity of the proposed heuristics compared to a global optimization and (ii) to consider the topic of multiple optimization vectors and their relevance to the underlying decision problem. Therefore, a further interpretation of the calculated results in terms of solving an economic decision problem is neither focus nor part of this research.

Details of the data shown will be discussed in the remainder of this article. Considering the limitations in the weight distribution included in the PDM model, which restricts the optimized weights to stay within a ratio of at most 1:10 (taken from the original PDM publication), we downscaled all input data sets into comparable magnitudes to avoid numerical problems. The PDM relies on the condition that the weight optimization is bounded so that attributes are not optimized to infinity. The ratio limitation does not influence the ranking of alternatives but is crucial to help the DM in understanding the structural differences of the alternatives, which is a core feature of the PDM. Results obtained with this process are opposed to exhaustive calculation of all possible combinations of directions yielding the global optimization result. We compare these results in the following section with our algorithm and relate to the validity of the proposed relaxation used in the PDM.

In total, 171 DMUs of 10 different data sets have been simulated and analyzed using the formerly described bi-section algorithm, thereby extracting a list of almost 1500 transitions in the respective profile distance distribution.

The following two show-cases have been selected to present step-by-step the validation procedure and the applied measurements. The underlying data sets and DMUs have been chosen to represent the lower and upper boundaries of the measured data in terms of deviation between heuristics and global optimization objective, respectively. The results of our calculation are presented both as tables and figures to support better understanding of the underlying problem. The show-cases are described by the following indicators which are listed in the respective tables:

α - the value of the fade factor directly restricting the 'freedom' of the linear program

Obj-H - the objective value (comprising profile distance and efficiency) for the heuristic calculation

Obj-G - the objective value for the global calculation under the same conditions (alpha)

Obj (%) - ratio between Obj-Heur / Obj-Glob

G-O - the number of optimization directions reaching the global optimum

D-O - the number of actually different optimizers reaching the global optimum

PD-H - the profile distance value using the heuristic optimization

PD-G - the profile distance value using the global optimization

PD (%) - PD-Heur / PD-Glob

Eff-H - the DEA efficiency value using the heuristic optimization

Eff-G - the DEA efficiency value using the global optimization

Eff (%) - ratio between Eff-Heur / Eff-Glob

We believe this combination of characteristics is adequate to efficiently elaborate on the different approaches and expose the shortcomings and benefits of both heuristic and global optimization, as given in the following sections. The show-cases have been chosen to effectively demonstrate the differences and similarities in results of global and heuristic optimization in both best and worst case scenarios.

4.1.1. Show-case 1 (low deviance)

As can be seen in Table 1, the objective values reached by heuristics in respect to those reached by global optimization are almost identical for the respective data set. This means that throughout all 12 transitions that have been found by the bi-section algorithm for this DMU, the variance in terms of objective is of a maximum significance of 1E-10. Figure 2 gives a clear picture of the unambiguous results. In other words, given the numeric accuracy, the numbers calculated using the proposed heuristics match with the global results for the complete range. As depicted in (2), the objective value is composed of efficiency value minus α times profile distance value. The slight deviance in profile distance and efficiency does therefore not result in a deviance in objective value since only its composition changes.

Table 1: DMU 2 of dataset 4 (low-deviance example) - transitions

α	Obj-H	Obj-G	Obj (%)	G-O	D-O	PD-H	PD-G	PD (%)	Eff-H	Eff-G	Eff (%)
0,0000	0,98039	0,98039	100,00	64	1	0,71121	0,71121	100,00	0,98039	0,98039	100,00
0,00918	0,97386	0,97386	100,00	2	1	0,57979	0,57979	100,00	0,97918	0,97918	100,00
0,07067	0,93821	0,93821	100,00	4	1	0,57854	0,57854	100,00	0,97909	0,97909	100,00
0,07684	0,93464	0,93464	100,00	4	1	0,54226	0,54226	100,00	0,97631	0,97631	100,00
0,08482	0,93031	0,93031	100,00	2	1	0,48311	0,48311	100,00	0,97129	0,97129	100,00
0,17972	0,88447	0,88447	100,00	2	1	0,41486	0,48311	85,87	0,95902	0,97129	98,74
0,52824	0,73988	0,73988	100,00	2	1	0,34237	0,34237	100,00	0,92073	0,92073	100,00
0,53322	0,73817	0,73817	100,00	2	1	0,27227	0,27227	100,00	0,88335	0,88335	100,00
0,61468	0,71600	0,71600	100,00	4	1	0,14065	0,14065	100,00	0,80245	0,80245	100,00
0,67156	0,70800	0,70800	100,00	16	2	0,10133	0,10133	100,00	0,77604	0,77604	100,00
0,72731	0,70235	0,70235	100,00	32	2	0,05891	0,05891	100,00	0,74519	0,74519	100,00
2,06540	0,62352	0,62352	100,00	64	2	0,00000	0,00000	100,00	0,62352	0,62352	100,00
Average			100,00					98,82			99,89

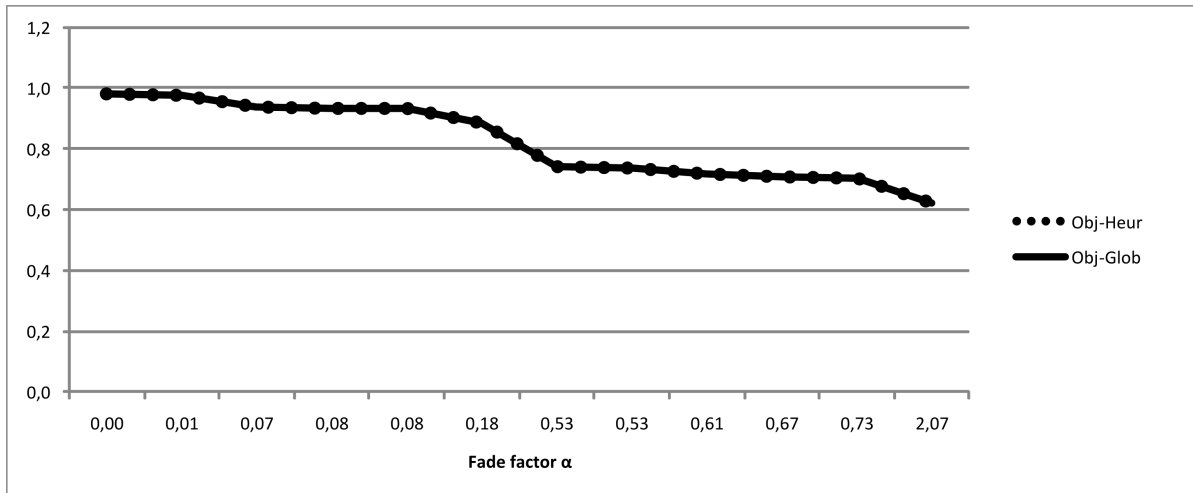


Figure 2: DMU 2 of dataset 4 - LP total objective

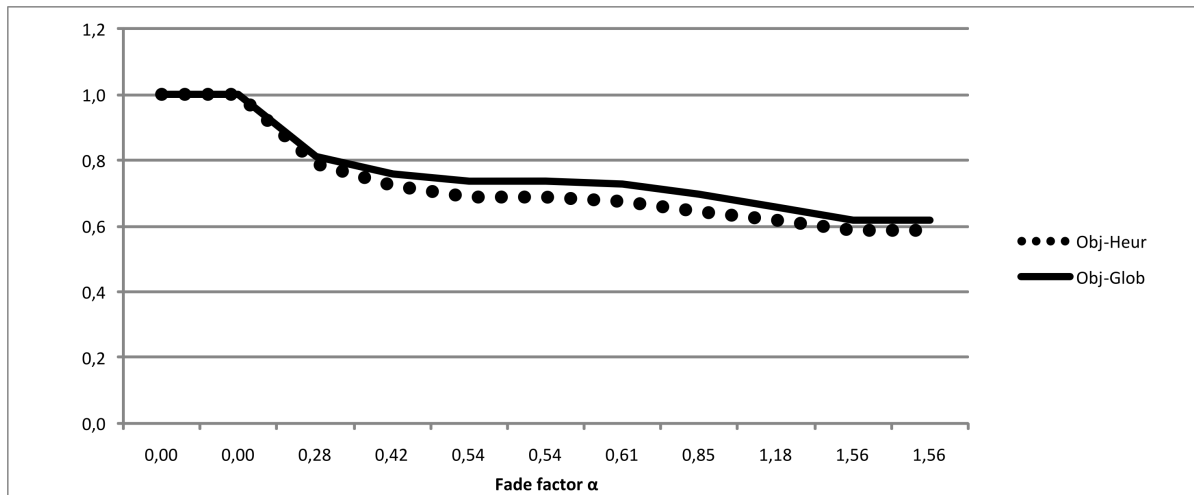


Figure 3: DMU 5 of dataset 2 - LP total objective

4.1.2. Show-case 2 (high deviance)

In opposition to 4.1.1, this showcase aims at demonstrating the data set-dmu combination found that show the least compliance between heuristics and global calculation. As can be seen in Table 2 and Figure 3, the highest deviance found within all optimized data combinations still shows a rather high compliance in relative observation of at least 95%. In this “worst case” scenario we can clearly discover that not only the composition of the objective value but also the sum of it changes. Although the example shows the worst average case for all 171 DMUs calculated, Figure 3 still depicts the only minor differences in terms of objective value between heuristics and global optimization.

Table 2: DMU 5 of dataset 2 (high-deviance example) - transitions

α	Obj-H	Obj-G	Obj (%)	G-O	D-O	PD-H	PD-G	PD (%)	Eff-H	Eff-G	Eff (%)
0,00000	1,00000	1,00000	100,00	1024	1	1,23884	1,23884	100,00	1,00000	1,00000	100,00
0,00000	1,00000	1,00000	100,00	4	1	0,76890	0,76890	100,00	1,00000	1,00000	100,00
0,27654	0,78737	0,81153	97,02	2	1	0,45685	0,58631	77,92	0,91370	0,97367	93,84
0,41839	0,72256	0,75697	95,45	8	1	0,28750	0,21410	134,28	0,84285	0,84655	99,56
0,53800	0,68817	0,73684	93,40	8	1	0,19756	0,13631	144,93	0,79446	0,81018	98,06
0,53895	0,68799	0,73671	93,39	16	1	0,17452	0,13631	128,03	0,78204	0,81018	96,53
0,60629	0,67624	0,72807	92,88	32	1	0,12198	0,12702	96,03	0,75019	0,80508	93,18
0,85096	0,64639	0,69770	92,65	64	1	0,09350	0,12340	75,77	0,72596	0,80271	90,44
1,17595	0,61600	0,65759	93,68	64	1	0,07255	0,12340	58,79	0,70132	0,80271	87,37
1,56000	0,58814	0,62007	94,85	256	1	0,04052	0,06052	66,95	0,65134	0,71448	91,16
1,56325	0,58801	0,61988	94,86	256	1	0,00000	0,06052	0,00	0,58801	0,71448	82,30
Average			95,29					89,34			93,86

4.2. Multiple optimizers

By using our exhaustive global optimization method as reference we were able to see that most optimizations tend to have ambiguous global optima, which initiated further investigation about this issue and its impact on the PDM. It is common knowledge that optimization based decision support methods often lead to multiple optimal solutions. See Figure 4 for an example of multiple optimal solution vectors, differing in their inner structure but summing up to the same objective value (optimum). In cases based on the original DEA approach, this does not constitute an issue since the method only focuses on the optimum and not on the optimizers. The PDM additionally takes the optimizer’s characteristics into consideration and therefore is dependent on either unambiguous optimization results (that is, the linear optimization model is augmented to only allow one optimal solution) or sound discrimination within the group of optimal optimization vectors. The column “G-O” given in Table 1 and Table 2 shows the number of combinations in global optimization that reach the highest possible objective value. Thus, for fade-factor 0, all possible combinations, that is number of output variables to the power 2, reach

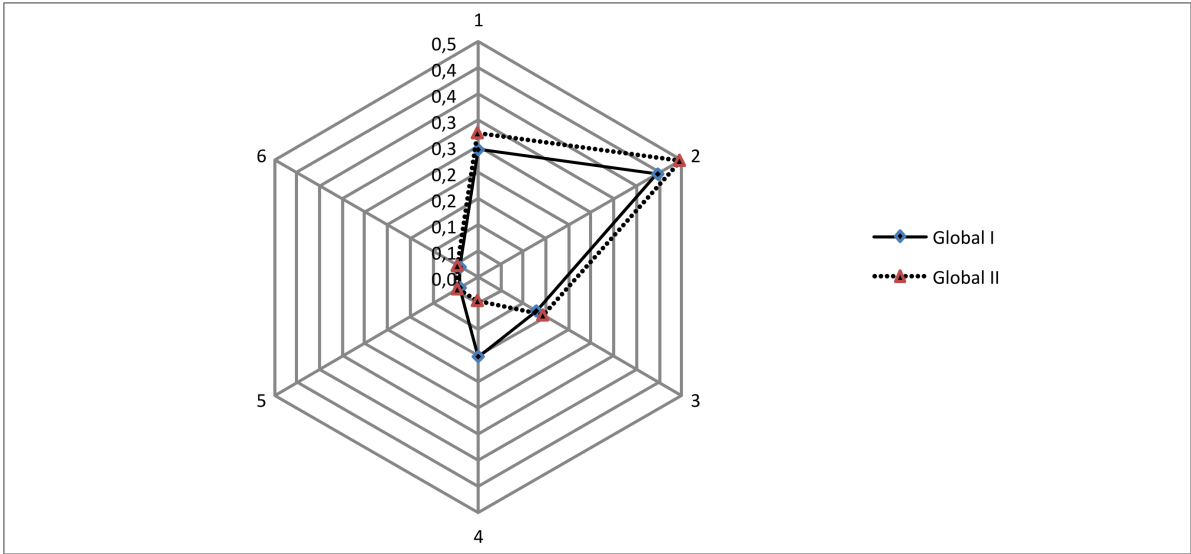


Figure 4: Example for multiple optimization vectors - DMU 2 of dataset 4

the maximum objective. In further examination, column “D-O” states the number of optimizers that (i) reach the maximum objective value and (ii) differ in their optimization vector. This situation can potentially lead to different interpretations and understandings of the PDM outcomes. In Table 1, the last three rows are bold to depict cases with more than 1 optimal optimization vector, as can be seen in column “D-O”.

4.3. Application of the PDM as a software prototype

We implemented a first prototype to test the practical application and usability of the PDM. We aimed at building a lightweight software tool that represents the decision support abilities of the PDM. Therefore, we chose to stick to the data drawn from the original PDM publication which describes a classic IS/IT investment problem (ERP software selection). The prototype was developed in Java SE to allow for future community development and/or open source integration. The software partly integrates existing open source projects, primarily used for linear optimization (LPSolve) and graphical representation (Graph2D). The setup of the prototype was designed as a Java Web Start application in order to easily deploy it and collect usage data. The prototype is available for download at [24].

We conducted a preliminary empirical investigation using the software prototype with the following research design. The research effort was split into three parts: scenario introduction, tool usage and questionnaire. Interviewees were randomly collected from a group of decision makers with various different backgrounds. They were provided with a short introduction to the decision problem and their role in the scenario. They were told to use the prototype, already prepared with the problem specific data (3 DMUs with 8 criteria each), to (i) get in touch with the decision space, (ii) analyse the advantages and disadvantages of each alternative using the PDM prototype and (iii) make an informed decision. The results of their considerations supported by the PDM prototype were finally collected by an online questionnaire. Our focus was on acceptance of the prototype, usability of the method, understanding of the decision process and agreement with the outcome of the process. We openly asked for improvements and/or desired additional features.

The result of the empirical validation showed that all respondents (100%) could use the prototype successfully. 75% of the respondents found the chosen technology, the appearance and the overall usability of the software good or very good. Especially the distribution as a Java Web Start application was found to be at least appropriate by almost 94%. Although the empirical validation conducted attracted only a rather small number of responses (n=32), we were able to obtain some valuable comments on further improvements and general acceptance of the PDM.

Table 3: Accumulated findings over all datasets

Dataset	# of DMUs	# of Outputs	# of Transitions	Calc.T-H	Calc.T-G	Obj (%)	PD (%)
1	21	11	12	2	6540	99,95	106,12
2	6	10	8	2	2339	99,30	100,76
3	20	6	8	0	91	100,00	99,77
4	12	6	8	1	79	100,00	100,92
5	29	5	5	1	44	100,00	96,32
6	25	7	9	1	231	99,97	102,43
7	22	7	10	1	226	99,96	102,14
8	30	9	8	1	1222	100,00	95,22
9	3	11	7	3	5357	100,00	105,22
10	3	8	5	1	394	99,95	100,58
Average				1,3	1652,3	99,91	100,95

5. Findings

Since each input data set consists of more than one DMU, findings for DMUs have been summed up. Figures concerning calculation performance, objective value distribution and profile distance distribution have been computed for both the global and our heuristic optimization as well as deviation factors. Table 3 gives an aggregated overview on the calculated characteristics for all datasets. The table comprises the following information:

Dataset - dataset number

of DMUs - number of DMUs in the respective dataset

of Outputs - number of output criteria for the respective dataset

of Transitions - mean number of transitions in the profile distance distribution

Calc.T-H (m.sec) - mean calculation time for one heuristic optimization

Calc.T-G (m.sec) - mean calculation time for one global optimization (milliseconds)

Obj (%) - mean ratio between objective value of heuristic optimization / objective value of global optimization

PD (%) - mean ratio between profile distance of heuristic optimization / profile distance of global optimization

Throughout the computation of the 10 different input data sets with variable numbers of output factors, ranging from 5 to 11, results discovered the strengths of the heuristic versus global optimization in use with the PDM. While the overall deviation of approximately 0.1% in average (5% max) of the heuristic objective in respect to the global objective performance can be neglected, major improvements can be shown by the significantly improved calculation performance of the heuristic optimization. Although the calculation times for global optimization mentioned above can be subject to substantial improvements when applying other alternative optimization methods, the calculation times for heuristic optimizations are expected to be superior due to their simplicity. Additionally they can be solved with available technologies such as with any LP solver.

During the simulation analysis we defined the following standard model to ensure valid results and comparability (shown in pseudo code):

```

foreach dataset : datasetlist{
  foreach dmu : dataset{
    calculate heuristic transitions using bi-section algorithm (transitionlist)
    foreach heuristic transition : heuristic transitionlist{
      calculate global optima
    }
  }
}

```

The desired profile was set up beforehand and was identical for all alternatives, varying in length due to the different number of outputs.

The average time consumed for a single calculation using the heuristic approach is below 1.5 milliseconds, whereas the same characteristic estimated for global optimization is at about 1652 milliseconds at average. Moreover, with increasing complexity of optimization problems the time consumed for calculation rises exponentially with the number of attributes for the global optimization. This is due to the fact that in order to get a global optimum, all possible directional approaches have to be tested against each other, which constructs a complexity of 2^n . In contrast, by using the heuristic approach the rather low level of linear complexity does not restrict scaling. Again, we know that the global solution is found using brute force. We emphasized this approach, however, since it delivers all global solutions as well.

6. Conclusion and further research

In this article we set out to improve practicability and to question the validity of heuristics used in the original introduction of the PDM [1]. Both aspects are important to support the need for integrating complex decision making methods in simple and user-friendly techniques and to ensure that computationally complex procedures can be implemented in feasible computer software applications. We identified the specific problem of finding the right setting for the fade operator to support practical use of the original PDM approach. To solve this problem, we proposed a divide and conquer algorithm that worked well with all data sets in the PDM implementation and enables the decision maker to efficiently find profile distance transitions and their associated profiles based on a given precision. The preliminary validation of the implemented software prototype provided first insights into user behaviour and general acceptance of the PDM as Decision Support System in practice.

Our second goal was to test the efficiency of a PDM heuristic used. The validation analysis supports the original idea of relaxing the PDM linear program in order to ensure scalability and out-of-the-lab applicability. In average the PDM model with the heuristic solution achieved an objective performance of 99.91%. It delivers results within 5% of the global optimizer for a wide range of data sets. The minor drawbacks in objective value performance and profile distance distribution seem to be more than compensated by major improvements in calculation performance and the low level of linear complexity ensuring scalability. The PDM in its current form can be used to tackle complex decision making scenarios without creating a computationally intense problem.

One of the main challenges we faced when trying to improve the methodology of the PDM was that multiple feasible solutions were found in various different configurations. Due to the fact that the PDM relies on the distribution of the profile, that is the relative difference in output variables, differing optimization vectors can result in ambiguous decision support. For this reason current research is addressing the possibility of generating a more meaningful discrimination within the group of feasible solutions, using the profile distance measure, namely the sum of absolute distances, as a second phase discriminator after the optimized objective value of the linear program. Future research will seek to enhance the decision process by improving the input range and quality. One stream of research considers multiple input variables. In its current specification input variables are trimmed to a single representation to fit the existing model, which restricts the method's areas of application. Another improvement would be to support the decision maker in finding an initial weighting profile. The PDM currently does not consider the establishment of a sound desired profile but refers to the usage of a method from the field of URM for this task. We are therefore reviewing different approaches of integrating pairwise comparison techniques such as utilized in prominent application oriented Multiple Criteria Decision Making methods such as MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique [25]) or the AHP (Analytical Hierarchical Process [26]).

References

- [1] E. Bernroider, V. Stix, Profile distance method: a multi-attribute decision making approach for information system investments, *Decision Support Systems* 42 (2) (2006) 988–998.
- [2] S. Zanakis, A. Solomon, N. Wishart, S. Dublish, Multi-attribute decision making: a simulation comparison of select methods, *European Journal of Operational Research* 107 (3) (1998) 507–529.

- [3] W. W. Cooper, L. M. Seiford, K. Tone, Data Envelopment Analysis, Kluwer Academic Publishers, London, 2000.
- [4] Z. Zangemeister, Nutzwertanalyse in der Systemtechnik Eine Methodik zur multidimensionalen Bewertung und Auswahl von Projektalternativen, Wittemann, Munich, 1976.
- [5] F. Y. Y. Chen, L. Liang, J. Zhu, Evaluation of information technology investment: a data envelopment analysis approach, *Computers and Operations Research* 33 (2006) 1368–1379.
- [6] M. Khouja, The use of data envelopment analysis for technology selection, *Computers & Industrial Engineering* 28 (1995) 123–132.
- [7] G. Mavrotas, P. Trifillis, Multicriteria decision analysis with minimum information: combining dea with mavt, *Computers and Operations Research* 33 (2006) 2083–2098.
- [8] H. Bal, H. H. Örkücü, S. Çelebioğlu, Improving the discrimination power and weights dispersion in the data envelopment analysis, *Computers and Operations Research* 37 (1) (2010) 99 – 107.
- [9] Z. Sinuany-Stern, A. Mehrez, Y. Hadad, An AHP/DEA methodology for ranking decision making units, *International Transactions in Operational Research* 7 (2000) 109–124.
- [10] T. Yang, C. Kuo, A hierarchical AHP/DEA methodology for the facilities layout design problem, *European Journal of Operational Research* 147 (1) (2003) 128–136.
- [11] R. Ramanathan, Data envelopment analysis for weight derivation and aggregation in the analytic hierarchy process, *Computers and Operations Research* 33 (5) (2006) 1289 – 1307.
- [12] S. B. HR Yazgan, K. Goztepe, An erp software selection process with using artificial neural network based on analytic network process approach, *Expert Systems with Applications* 36 (2009) 9214–9222.
- [13] H. Deng, J. Quan, A decision model to erp systems evaluation and selection, in: 6th International Conference on Information and Management Sciences, Lhasa, China, 2007, pp. 412–418.
- [14] E. Karsak, C. Ozogul, An integrated decision making approach for erp system selection, *Expert Systems with Applications* 36 (2009) 660–667.
- [15] P. Anderson, N. Peterson, A procedure for ranking efficient units in data envelopment analysis, *Management Science* 39 (1993) 1261–1264.
- [16] R. Allen, E. Thanassoulis, Improving envelopment in data envelopment analysis, *European Journal of Operational Research* 154 (2004) 363–379.
- [17] F. Liu, H. Hsuan-Peng, Ranking of units on the dea frontier with common weights, *Computers and Operations Research* 35 (2008) 1624–1637.
- [18] E. Bernroider, V. Stix, A method using weight restrictions in data envelopment analysis for ranking and validity issues in decision making, *Computers and Operations Research* 34 (9) (2007) 2637–2647.
- [19] E. Bernroider, N. Obwegeser, V. Stix, Introducing complex decision models to the decision maker with computer software - the profile distance method, *Journal of Systemics, Cybernetics and Informatics* 8 (3) (2010) 24–28.
- [20] F. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly* 13 (1989) 319–339.
- [21] Q. Ma, L. Liu, The technology acceptance model: A meta-analysis of empirical findings, *Journal of Organizational and End User Computing* 16 (1) (2004) 59–72.
- [22] OECD, Statistical data of member and non-member economies of the oecd.
URL <http://stats.oecd.org/>
- [23] J. Beasley, Comparing university departments, *Omega* 18 (2) (1990) 171–183.

- [24] E. Bernroider, N. Obwegeser, V. Stix, Profile distance method (pdm) prototype implementation.
URL <http://xmtim.ai.wu-wien.ac.at/web/pdmnet/pdmnet.jnlp>
- [25] C. A. B. e Costa, J.-C. Vansnick, MACBETH – an interactive path towards the construction of cardinal value functions, *International Transactions in Operational Research* 1 (4) (1994) 489 – 500.
- [26] T. Saaty, Decision making with the analytic hierarchy process, *International Journal of Services Sciences* 1 (1) (2008) 83–98.