In today’s competitive market, where organizations fight each other over every customer, having a well designed website capable of attracting and retaining customers is very important. By utilizing the vast medium of the Internet, organizations may experience competitive capabilities such as increased profitability, new markets, better customer service and effective communication. This paper introduces an example of how Data Envelopment Analysis (DEA) can be used to analyse website design and how the optimal website design can be identified, given multiple design criteria, such as download time and visualization.

Key words: Performance, Data Envelopment Analysis, DEA

History:

1. Introduction

Over the years the Internet have provided organizations with a vast array of competitive capabilities through electronic commerce, such as increased profitability, reaching new markets, improving customer service, increasing distribution speed and more effective communication (Asllani and Lari (2007)). Achieving these benefits requires among other things, a website that is able to retain and attract customers to the site and for these customers to either purchase products or services from the organization. This ability relies heavily on how the website is designed in terms of web-objects, such as banners, splash screens etc. In Asllani and Lari (2007) the authors develops a "genetic algorithm" (GA) capable of identifying the best website design given the download time, the visualization score and the increase in potential sales, of a specific sequence of web-objects. Building on the setup and the results found using the GA approach, this paper develops a "Data Envelopment Analysis" (DEA) approach capable of identifying the optimal website design, given a multitude of design criteria.

This paper unfolds as follows: section 2 presents a review of the current research in the area of performance evaluation (using DEA and GA), problems encountered when applying DEA to data sets containing a large amount of "Decision Making Units" (DMUs) and identifying the "best" solution. Section 3 presents a short introduction to 3 different approaches useful when analyzing multi criteria decision making problems. Describing our model and how to apply the model to performance evaluating of websites is done in section 4. The paper concludes with a discussion of the method and its application possibilities in section 5 and section 6.

2. Literature Review

3. Multi Criteria Decision Making Models

When facing a problem that requires choosing between two or more alternatives, we usually define certain properties and selection criteria, that are useful in making the decision. An example of such a problem is the acquisition of a new car, where the following could be some of the selection criteria:
• Milage
• Price
• Space
• Number of seats

Evaluating the available cars according to the above criteria will lead to a decision of whether or not to make the purchase. In the following paragraphs, methods useful when dealing with multi-criteria problems, as the one outlined above, will be described.

3.1. Multi Criteria Optimization

One method useful when dealing with a decision problem as the one outlined above, is Multi Criteria Optimization or MCO, first defined by Steuer (1986) as a mean to solve multiple objective programming problems (MOPP):

\[
\begin{align*}
\text{max } & \{ f_1(x) = z_1 \} \\
\text{max } & \{ f_2(x) = z_2 \} \\
\vdots \\
\text{max } & \{ f_k(x) = z_k \} \\
\text{s.t. } & x \in S
\end{align*}
\] (1)

In the above equation "S" referees to the feasible region in decision space and is defined by

\[ S = \{ x \in \mathbb{R}^n \mid Ax = b, x \leq 0, b \in \mathbb{R}^m \} \]

where "x" is a point in the decision space.

From the MOPP we notice the similarities to the more common single objective program, the major difference being the stack of objective functions, shown in equation 1.

As mentioned in Steuer (1986) the ideal way to solve a MOPP would be to first asses the decision maker’s utility function "U" and then solve the following mathematical programming problem:

\[
\begin{align*}
\text{max } & \{ U(z_1, z_2, \ldots, z_k) \} \\
\text{s.t. } & f_i(x) = z_i, \ 1 \leq i \leq k \\
& x \in S
\end{align*}
\] (2)

The only problem with the above representation of the mathematical problem is that it might not be possible to represent the decision maker’s utility function mathematically. This means that we have to solve the problem stated in equation 1.

3.2. Analytical Hierarchy Process

Another method useful when dealing with the decision problem illustrated in the beginning of this section, is the "Analytical Hierarchy Process" or AHP, described by Saaty and Vargas (2000) as a basic approach to decision making, designed to cope with both the rational and the intuitive to select the best from a number of alternatives evaluated with respect to several criteria. The use of AHP in a decision making scenario, requires a structural representation of the overall goal, criteria and selection alternatives, as illustrated below.

Using each element defined in the "Hierarchy" shown in figure 1, the decision-maker performs pairwise comparisons of the criteria to the overall goal, followed by a comparison of each alternative to each of the criteria. These pairwise comparisons are performed using a "fundamental scale" illustrated in table 1.
Figure 1  Structural representation of the decision problem

<table>
<thead>
<tr>
<th>Intensity of</th>
<th>Definition</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Equal importance</td>
<td>Two activities/elements contribution to the overall decision is equal</td>
</tr>
<tr>
<td>2</td>
<td>Weak</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Moderate importance</td>
<td>A slight favor of one activity/element over another</td>
</tr>
<tr>
<td>4</td>
<td>Moderate plus</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Strong importance</td>
<td>A strong favor of one activity/element over another</td>
</tr>
<tr>
<td>6</td>
<td>Strong plus</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Very strong or demonstrated importance</td>
<td>There exists a very strong preference for one activity/element compared to another</td>
</tr>
<tr>
<td>8</td>
<td>Very, very strong</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Extreme importance</td>
<td>One activity/element is favored compared to another</td>
</tr>
</tbody>
</table>


When performing the pairwise comparisons it is assumed that there exists a reciprocal relationship between the two activities/elements compared. Hence if element "i" compared to element "j" is assigned one of the above numbers, the reciprocal value is assigned the case where element "j" is compared to element "i". One example of this relationship is illustrated in table 2.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Milage</td>
<td>1/2</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>B: Price</td>
<td>1/4</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C: Space</td>
<td>1/2</td>
<td>1/3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>D: # of seats</td>
<td>1/2</td>
<td>1/4</td>
<td>1/5</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2  Car purchase.
Table 2 is known as the matrix of pairwise comparisons "A" given by:

\[
A = \begin{pmatrix}
     w_1/w_1 & w_1/w_2 & \ldots \\
     w_2/w_1 & w_2/w_2 & \ldots \\
     \vdots & \vdots & \vdots \\
     w_n/w_1 & w_n/w_2 & \ldots
\end{pmatrix}
\]

Using the different vectors of priorities "a_ji" found in the matrix A, it is possible to estimate if the judgements/weights assigned to each element of the hierarchy is consistent. If say "a_{ij}" represents the importance of element i over j and "a_{jk}" the importance of element j over k, then the importance of element i over k ("a_{ik}"") has to be equal to \(a_{ij} \cdot a_{jk} = a_{ik}\) in order for the judgements/weights to be consistent.

3.3. Data Envelopment Analysis

Measuring the performance of organizations such as business firms, government agencies, hospitals, educational institutions etc. can be accomplished via the method "Data Envelopment Analysis or DEA (Cooper et al. (1999))"

Each of the organizations or "Decision Making Units (DMU)" is evaluated using virtual inputs and outputs determined using:

\[
\text{Virtual Input} = v_1x_{1o} + v_2x_{2o} + \ldots + v_mx_{mo} \quad (3)
\]

\[
\text{Virtual Output} = u_1y_{1o} + u_2y_{2o} + \ldots + u_my_{mo} \quad (4)
\]

where 

\[
v_i = \text{input weights}
\]

\[
u_i = \text{output weights}
\]

Where the different weights for input and output is unknown and needs to be determined so as to maximize the ratio:

\[
\frac{\text{Virtual Output}}{\text{Virtual Input}} \quad (5)
\]

There exists 2 fundamental models in DEA:

- The CCR model: This is one of the most basic DEA models, first developed by Charnes, Cooper and Rhodes in 1978. The main assumption in this model, is the fact that the model deals with constant return to scale.
- The BCC model: This model extends the basic CCR model, by relaxing the assumption of constant returns to scale. The model was, developed by Banker, Charnes og Cooper in 1984.

Each of these methods are explained in detail below.

3.3.1. CCR Model In the CCR model, measuring the efficiency/performance of each DMU is done once, requiring \(n\) optimizations, one for each of the DMUs under evaluation. The weights that maximize the ratio shown in equation 5 is determined using a "fractional programming problem" as shown below:
\[
\max \theta = \frac{u_1 y_{1o} + u_2 y_{2o} + \ldots + u_s y_{so}}{v_1 x_{1o} + v_2 x_{2o} + \ldots + v_m x_{mo}} \\
\text{subject to} \\
u_1 y_{1o} + u_2 y_{2o} + \ldots + u_s y_{so} \geq 1 \\
v_1 x_{1o} + v_2 x_{2o} + \ldots + v_m x_{mo} \\
v_1, v_2, \ldots, v_m \geq 0 \\
u_1, u_2, \ldots, u_s \geq 0
\] (6)

As shown in Cooper et al. (1999) this fractional programming problem, can be transformed into a linear program yielding the following:

\[
\max \theta = \mu_1 y_{1o} + \mu_2 y_{2o} + \ldots + \mu_s y_{so} \\
\text{subject to} \\
v_1 x_{1o} + v_2 x_{2o} + \ldots + v_m x_{mo} = 1 \\
\mu_1 y_{1j} + \mu_2 y_{2j} + \ldots + \mu_s y_{sj} \leq 1 \\
v_1, v_2, \ldots, v_m \geq 0 \\
\mu_1, \mu_2, \ldots, \mu_s \geq 0
\] (7)

From this linear program the optimal weights for the DMU under evaluation, DMU\textsubscript{o}, can be determined.

By using a single input and single output scenario, the use of the CCR model can be illustrated. In this scenario the ratio in 5, is treated as a measure of how efficient the DMU under evaluation is compared to other DMUs.

In the following example 6 different car dealers, labeled A to F, is evaluated according to the "Sales/Employees" ratio.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Single Input and Single Output Case.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car dealer</td>
<td>A</td>
</tr>
<tr>
<td>Employees</td>
<td>2</td>
</tr>
<tr>
<td>Sale</td>
<td>1</td>
</tr>
<tr>
<td>Sale/Employees</td>
<td>0.5</td>
</tr>
</tbody>
</table>

The bottom line of table 3 shows the sales per employee, which is a measure of the efficiency of the dealership being evaluated. Representing the data from table 3 in a two-dimensional plot, with sales on the vertical axis and number of employees on the horizontal axis.

The "efficient frontier" shown in figure 2 represents the highest possible slope connecting at least one of the points to the origin. Using the CCR model we seek to determine how DMUs located under this frontier can be improved and move up to the frontier. For example, car dealership "A" can be improved by either reducing the number of employees(A2) or by increasing the sale(A1).

Any point on the line segment, between "A1 and A2" offers a chance to improve the performance of dealership "A". This fact leads to the question of how much to reduce the staff and/or by how much the sale needs to be increased in order for dealership "A" to be termed efficient. Answering this question can be achieved by solving the linear programming shown in 7.
3.3.2. BCC Model  In the CCR model, described above, the key assumption was the fact that each of the DMUs under evaluation, exhibited constant returns to scale. In the BCC model the efficient frontier spans a "convex hull" of the existing DMUs and the frontier has piecewise linear and concave characteristics which leads to variable returns to scale (see Cooper et al. (1999)).

Returning to the single input and single output example used in the description of the CCR model, we see that the efficiency frontier now corresponds to the lines connecting A, B and C.

Evaluating the efficiency/performance of the different DMUs, can be done by solving the following “input oriented” linear program:

$$\min \theta_B$$

subject to

$$\theta_b x_o - X \lambda \leq 0$$
$$Y \lambda \leq y_0$$
$$e \lambda = 1$$
3.4. Comparison

As mentioned in the beginning of this chapter the methods described in the above paragraphs are all useful when dealing with multi criteria problems. However they differ on a number of areas:

- In AHP the weights assigned to each of the criteria and alternatives are determined by the decision maker and the best alternative is selected based on the overall score of each alternative. In DEA the weights are calculated based on the data.
- In AHP the number of decision alternatives that can be compared is limited. This is due to the fact that, a very complex decision problem would require an infeasible amount of pairwise comparisons. Hence the mathematical representations of the decision problem used in DEA and MCO might be more feasible when the decision problem is very complex.
- Using AHP will help the decision makers find the solution that best suits their goal and their understanding of the problem. However identifying, which of the decision alternatives, that performs the best in terms of converting inputs into outputs is not possible neither is the determination of the sources of inefficiency in terms of excessive use of particular resources. This however is possible in DEA, where each of the alternatives are measured according to an efficient frontier on which only the efficient alternatives are located.

4. Website Performance Evaluation

4.1. Introduction

Over the years the Internet has emerged as a promising medium for organizations to connect with their customers/users and thousands of organizations have developed websites that provide company information, online stores and much more. Since these web-sites are the "face" of the organizations, understanding the websites "performance" in terms of its ability to retain or "keep" the customers at the website is very important.

An example of how to evaluate the performance of a website is presented in Asllani and Lari (2007) where the authors present a genetic algorithm (GA) that is capable of evaluating the arrangements of different web-objects (such as banners, images, splash screens etc.) such that the
design of the website is optimized. The performance of the website is measured, via performance indicators such as (a) download time, (b) degree of visualization and (c) increase in potential sales. Using these performance indicators the authors evaluates the website using the below fitness function:

$$F = \left( \frac{\sum_{k=1}^{m} D(k)}{\max_{k=1}^{m} D(k)} \right) \cdot w_1$$

$$+ \left( 1 - \frac{\sum_{k=1}^{m} V(k)}{\max_{k=1}^{m} V(k)} \right) \cdot w_2$$

$$+ \left( 1 - \frac{\sum_{k=1}^{m} P(k)}{\max_{k=1}^{m} P(k)} \right) \cdot w_3$$

where

- $D(k)$ is the download time of the kth web-object.
- $V(k)$ is the visualization score of the kth web-object.
- $P(k)$ is the probability of the product or the service represented by the kth web-object will be sold if followed by the (k-1)th web-object.
- $w_1, w_2, w_3$ is the respective weights assigned to each component of the fitness function
- $m$ is the number of web-objects in the sequence.

By using this GA the authors believe that web-design teams will be able to create adaptive web-sites, recognize product-line patterns that will increase sales, retain customers and increase visualization (see Asllani and Lari (2007) page 1768).

While reading the paper by Asllani and Lari (2007) one realizes that the authors solves a multi criteria optimization problem using a GA, a rather stange approach given the vast literature on MCDM and the different implementations of MCDM in a vast range of application areas. However the use of GA in the Asllani and Lari (2007) paper is motivated by GAs ability to find a near optimal solution quickly (early convergence) and the GAs ease of use.

So why use other techniques? Why consider applying complex methods like DEA or AHP to evaluate the best website design?

Answering these questions requires a more "critical" evaluation of the GA approach:

- Problem complexity. In GA, evaluating the performance of an object, is based on the value of the fitness function. However if the problem has a high degree of complexity, calculating this fitness function, could be infeasible with regards to the computation time.
- Local optima instead of global optima.
- Dynamic datasets. Early convergence might not be valid in the long run.
- Other optimization methods may find better solutions.

Given the above issues/limitations of the GA approach and the fact that the paper by Asllani and Lari (2007) sets up a multi criteria decision making problem we find it interesting to test, whether DEA or AHP could be used to evaluate the performance of websites and if it would be able to generate a "better" solution than the one found using the GA approach.
4.2. MCDM and website evaluation - Comments

Evaluating the performance of web-sites using MCDM methods, such as the ones described in section 3, requires a careful selection of the input and output variables. If the "wrong" input/output combinations are selected, it might skew the efficiency measure such that websites that perform the best are termed inefficient and vice versa. Even more important is the "level of analysis" or the DMUs/selection alternatives in the model, which requires careful considerations.

- In AHP, the number of pairwise comparisons needed to select the best among the selection alternatives, are dependent on the number of criteria and the number of selection alternatives. If the level of analysis is to specific, eg. the sequence of websites, the placement of buttons, etc., there might be a huge number of alternatives and the number of pairwise comparisons will become infeasible.

- In DEA, the relative efficiency and inefficiency of the DMUs are modeled based on linear programs, which are solved for each of the DMUs independently. As for the AHP case working with a level of analysis that is too specific, might create a huge number of DMUs and an equivalent number of linear programs, which hampers the performance of the DEA model in terms of computation time.

Given the above considerations, it becomes clear that using MCDM to evaluate website performance, requires "tweaking" the selected method, such that it is capable of working with a large number of input/output combinations and a large number of DMUs/selection alternatives.

Evaluating the performance of selection alternatives/DMUs requires a method that is capable of comparing each alternatives ability to convert inputs into outputs. Given the comparison of the selected methods described in section 3.4 it is clear that DEA is the method, that has this ability and therefore this method is chosen as the preferred method when evaluating the performance of websites.

4.2.1. Why use DEA to evaluate website performance? The evaluation of hospital, school, university and organizational performance have been the focus of much of the existing literature regarding performance evaluation and DEA. This is due to the fact that the tasks of identifying DMUs and input/output variables in these cases are "fairly easy" compared to other scenarios, such as website evaluation where the DMUs can be harder to determine, since it all depends on the level of analysis, as described in the above section. So why use DEA to evaluate the performance of a website?

In GA the evaluation of website performance is based on a very strong a priori assumption regarding the size of the weights used to calculate the value of the fitness function. Thus in many cases the developed GA algorithm will find a local optimum or an arbitrary point instead of a global optimum. Therefore selecting a method that can perform the evaluation independently of any a priori knowledge regarding how the different input/output combinations are related (the weights used). From the description of the different MCDM methods we know that DEA (see section 3.3, page 4) is capable of evaluating the performance without any a priori knowledge regarding the weights assigned to the different input/output combinations. This means that the analysis performed using DEA is a more detailed/accurate evaluation of the website performance and provided a global optimum(s).

Also when using GA, a functional relationship between the selected inputs and outputs is required. However when working with IT-related products, such as websites, establishing this relationship can often be challenging. By using DEA we are able to perform the evaluation, without a function relationship, since DEA is a non-parametric method.
4.2.2. Caution flags when using DEA  Even though the above section indicates that using DEA will outperform the GA approach when performing performance evaluations, a few words of caution requires attention:

- As mentioned the weights used in the evaluation is determined by the data and thus they are “unknown” once the evaluation begins and calculated via the available data (the input/output combinations). This means that selecting the wrong data will greatly influence the efficiency scores of the evaluated websites and it may lead to inefficient observations defining the frontier.

- Selecting to many inputs/outputs while only analyzing a small number of data points/DMUs will lead to more combinations if inputs and outputs being termed efficient.

- When performing a performance evaluation using DEA, selecting the level of analysis is very important. A level of analysis that is too general can result in too many efficient combinations of inputs and outputs. However selecting a level that is too specific can results in an exponential number of selecting alternatives/DMUs that must be considered in the evaluation.

Theses words of caution means that using a traditional DEA model (CCR or BCC) when evaluating the performance of a website, might not be feasible. In the next 3 sections (4.2.3, 4.2.4 and 4.2.5) techniques useful when using DEA on data sets that are subject to the mentioned cautions, will be presented.

4.2.3. DEA and the Free Disposal Hull method  The free disposal hull or FDH method was first introduced in Deprins et al. (1984) and is seen as an alternative to the DEA method, in which the convexity assumption of the BCC DEA method is relaxed. FDH is used to (1) establish a best practice group amongst a set of observed units and to (2) identify the units that are inefficient when compared to the best practice group (see Benslimane et al. (2008)). The FDH efficiency is obtained as follows:

\[
\begin{align*}
\text{Min} & \quad \theta^k \\
\text{subject to} & \quad \theta^k x^k_i - \sum_{h=1}^{n} \gamma^k x^k_i \geq 0, \ i=1,...,I \\
& \quad \sum_{h=1}^{n} \gamma^k y^k_j \geq y^k_j, \ j=1,...,J \\
& \quad \theta^k \geq 0, \ h=1,...,n \\
& \quad \sum_{h=1}^{n} \gamma^h = 1 \\
& \quad \gamma^h \in \{0,1\}
\end{align*}
\]

Illustrating the differences between the FDH method and the more traditional DEA methods, the BCC and CCR methods, is presented in figure 5.
4.2.4. DEA and Assurance Regions  As mentioned in section 3.3 the main advantage of the classical DEA models (the CCR and the BCC models) is the fact that there is no need for a priori knowledge regarding the weights assigned to the input/output combinations. However when there is a large difference between the weights used to evaluate the input/output combinations this advantage might become a disadvantage, since identifying the "best" or "most efficient" DMU is difficult. Thus using DEA to locate the best possible/most efficient DMU from a range of alternatives, might not always be the obvious choice, especially if the assumption of variable returns to scale is valid (the BCC model). If this is the case, using DEA will create an "efficient frontier" consisting of convex combinations of all the efficient DMUs. This efficient frontier makes it difficult to select the "best" or "most efficient" DMU. Thus the DEA method has to be modified and an additional constraint that restrict the weights assigned to the input/output combinations, has to be introduced in the model.

In Thompson et al. (1986) the authors describes a case where DEA is utilized to determine the best location for a "high-energy physics lab" in Texas. In the paper the authors find that 5 out of the 6 sites/DMUs (South Dallas, North Houston, South Houston, East/Central Texas, West/Central Texas and West Texas) where termed "extreme efficient". However the task faced by the authors where to select the best location among the 6 available.

Given this task the authors developed a special case of the BCC model, in which an additional constraint was assigned to the value of the assigned weights.

4.2.5. Preprocessing DEA

4.3. DEA and Website Evaluation

As mentioned in the beginning of section 4, Asllani and Lari (2007) presents a GA, capable of evaluating the performance of website design in terms of download time, visualization score and increase in potential sales. In this section, we intend to develop a DEA approach suitable for the multi criteria setup defined by Asllani and Lari (2007). Thus we develop a DEA model with the following characteristics:

- **Ability to work with a high number of DMUs.** In the paper by Asllani and Lari (2007) the evaluation of the website was based on the arrangement of 5 different web-objects (banners, splash screens, etc.). Identifying the optimal arrangement of these web-objects meant finding a website design that minimized the fitness function. By using GA the authors where able to dynamically change the arrangement of the 5 web-objects until the value of the fitness function achieved an acceptable level (occurs after a predefined number of generations or the fitness function achieves a minimum value).
Applying DEA to this GA setting, means evaluating all the possible arrangements of the 5 different web-objects, since the arrangements of the web-objects defines the design of the website. This means that the number of DMUs in this case will be equal to 120 or 5! or in more general terms the factorial of the number of web-objects.

- **Finding the optimal solution.** The main objective of the paper by Asllani and Lari (2007) was to identify the optimal design/arrangement of the 5 different web-objects, however using DEA to select the ”optimal” or ”most” efficient design/arrangement can be a very difficult task (see section 4.2.4 for further details). We therefore need to modify the DEA approach, such that identifying the ”most” efficient design among all the efficient designs becomes possible.

The proposed DEA model has the following setup:

<table>
<thead>
<tr>
<th>DEA model</th>
<th>DMU</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>A preprocessed FDH model, in which assurance regions can be used to identify the optimal solution</td>
<td>The different website designs/arrangements of the web-objects</td>
<td>The download time and visualization score for each of the web-objects</td>
<td>Potential sales, def. as: The probability of selling a product or service represented by a given web-object when another web-object is visited</td>
</tr>
</tbody>
</table>

5. Discussion
6. Conclusion
7. Future Research Directions
Acknowledgments

References


