ISSN 0254-2757

QUÆSTIONES INFORMATICÆ

Volume 6 • Number 3

November 1988

T McDonald	A Proposed Computer Network for Researchers	95
T H C Smith	Finding a Cheap Matching	100
P J S Bruwer	Ranking Information System Problems in a User Environment	104
S W Postma N C K Phillips	The Parallel Conditional	109
D G Kourie R J van den Heever	Experiences in CSP Trace Generation	113
G de V de Kock	Die Meting van Sukses van Naampassingsalgoritmes in 'n Genealogiese Databasis	119
R Short	Learning the First Step in Requirements Specification	123
E C Anderssen S H von Solms	Frame Clipping of Polygons	129

The official journal of the Computer Society of South Africa and of the South African Institute of Computer Scientists

Die amptelike vaktydskrif van die Rekenaarvereniging van Suid-Afrika en van die Suid-Afrikaanse Instituut van Rekenaarwetenskaplikes

QUÆSTIONES INFORMATICÆ

The official journal of the Computer Society of South Africa and of the South African Institute of Computer Scientists

Die amptelike vaktydskrif van die Rekenaarvereniging van Suid-Afrika en van die Suid-Afrikaanse Instituut van Rekenaarwetenskaplikes

Subscriptions

The annual subscription is

	SA	US	UK
Individuals	R20	$7	£5
Institutions	R30	$14	£10

to be sent to:
Computer Society of South Africa
Box 1714 Halfway House 1685

Quæstiones Informaticæ is prepared by the Computer Science Department of the University of the Witwatersrand and printed by Printed Matter, for the Computer Society of South Africa and the South African Institute of Computer Scientists.

Ranking Information System Problems in a User Environment

P J S Bruwer
Department of Computer Science, Potchefstroom University for CHE, Potchefstroom, 2520

Abstract

In a previous research project (Bruwer [1]) a methodology to identify information system problems from a user point of view, was presented. This methodology was also illustrated by means of empirical data collected from the management of a very large organisation. In this methodology a number of statistical techniques such as regression (stepwise and multiple) and operational research techniques (linear programming) were used in the analysis of the data. With these methods it was possible to find optimal solutions for certain decision variables contributing to the level of success of the information system.
It was also suggested in the article that problem areas which were identified could be prioritise by making use of differences between optimum values of decision variables and their corresponding measured mean values. In this article, however, a methodology is presented where not only these differences but also other computed values give a better criteria for the management of an organisation to identify and prioritise problems in the overall information system function.

Opsomming

In 'n vorige navorsingsprojek (Bruwer [1]) is 'n metodologie om inligtingstelselprobleme in 'n organisasie te identifiseer voorgestel, bespreek en aan die hand van empiriese data geïllustreer. In die artikel word daar van verskillende tegnieke o.a. regressie- en operasionele navorsingsmetodes gebruik gemaak om sodoende optimale waardes vir beslissingsveranderlikes te verkry wat die grootste bydrae lewer om die suksesvlak van die inligtingstelsel te verhoog.
In die genoemde artikel word daar aangedui dat die probleemareas geprioritiseer word deur die grootste verskille tussen die optimale waardes van die beslissingsveranderlikes en die gemete of waargenome waardes daarvan as maatstaf te gebruik. In hierdie artikel sal aangetoon word dat dié waardes tesame met ander berekende waardes 'n beter maatstaf daar stel vir die bestuur van 'n organisasie om knelpunte in die totale inligtingstelsel te prioritiseer.

Received June 1988, Accepted August 1988

1. Introduction

The background and philosophy of the methodology is described in full detail and illustrated with examples in Bruwer and Hattingh [2].

In short, data is collected from user management of a large organisation in South Africa. A questionnaire was developed with 34 information system attributes to be evaluated on a 7-point scale. The aspects were fully described in Bruwer [1]. Managerial users of the information system have to evaluate the importance of the attributes as well as the performance of the attributes in the organisation. In a number of previous research projects the 7-point scale was also used successfully as a measuring instrument (Alloway [3], Bruwer [4], [5], [6].

The basic steps to be taken and computations to be done in this methodology follows in the next paragraphs.

The Model

Consider the model

$$Y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_k X_{ik} + \varepsilon_i$$

where ε_i are independent random variables with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$ for all i.

For this model:

a) Values $X_{i1}, X_{i2}, ..., X_{ik}$ are specified by the researcher,

b) X_{ij} is the i-th value of the j-th independent variable,

and

c) $\beta_0, \beta_1, ..., \beta_k$ are unknown parameters to be estimated.

Several methods to estimate the parameters exist. The estimates of the parameters β_j (j = 0,1,...,k) will be denoted by $b_0, b_1, \ldots, b_k$ are to be determined from N observations $(Y_i, X_{i1}, \ldots, X_{ik})$, $i = 1,2,\ldots,N$.

One of the best known methods is the method of least squares where $b_0, b_1, \ldots, b_k$ are determined in such a way that

$$\sum_{i=1}^{N} \Delta_i^2 \text{ is minimised}$$

where $\Delta_i = Y_i - b_0 - b_1X_{i1} - \ldots - b_kX_{ik}$.

Stepwise or multiple linear regression methods can now be used to obtain a less complicated model. Although we may be quite satisfied with this model, some questions like the following may still exist:

How do we interpret the regression function?

What is the influence of a specific "independent" variable?

How do we use the estimated function when we are interested in optimisation of the dependent variable?

In one of the questions in the questionnaire the overall success of the information systems has to be evaluated. This variable is used as the dependent variable and all 34 other "performance" variables (attributes) as independent variables in either a stepwise linear regression or multiple linear regression computer program. The result is as less complicated model with, say k "independent" variables.

Because of the above raised questions and the possibility of interdependency between all or some of the decision variables in the model a restricted linear regression method was developed. With the use of this method it is possible to study the influence of a decision variable on the dependent variable. This methodology is discussed in the next paragraph.

The Area of Experience of the Restricted Linear Regression Method

In this paragraph we take a look at the area of experience measured by the existence of data points in this area. For this purpose we consider the convex set obtained by looking at all possible convex combinations of the observed data points. This is enclosed by the convex hull of the data points.

Suppose we have data points

$$V_i = (X_{i1}, X_{i1}, \ldots, X_{ik}) \text{ for } i = 1,2,\ldots,N.$$

The convex hull could now be represented as the following set:

$$C = \{Z/Z \in E^k \text{ and } Z = \sum_{i=1}^{N} \lambda_i V_i \text{ with } \lambda_i \geq 0 \text{ and } \sum_{i=1}^{N} \lambda_i = 1\}, \text{ and where } E^k \equiv \text{ k-dimensional}$$

Euclidian space.

We are now interested in the behaviour of the regression function in this area C.

The Influence of an "Independent" Variable

A researcher is usually interested in the influence of a specific independent variable upon the dependent variable. In this particular case we are interested in the influence of the aspects regarding the information system which are used as "independent" variables in the regression model upon the dependent variable (overall success of the information system).

When no interdependence exists between decision variables, we need only take notice of the regression coefficient of the specific variable. In such a situation it could be possible theoretically to obtain for example the maximum of the dependent variable $\hat{Y}$ by setting the values of the decision variables with positive regression coefficients as high as possible and those with negative regression coefficients as low as possible. The problem that arises with this approach is the fact that it could be impossible physically to implement the proposed combination of levels of the variables. In this discussion we assume that we only want to propose levels of variables that fall within the area of experience of the experimenter.

That is, levels of the decision variables that fall in the convex hull C.

Suppose the regression function obtained in this case is the following:

$$S = b_0 + b_1X_1 + \ldots + b_kX_k$$

We want to determine the influence of a specific decision variable X_p on the dependent variable S. In the first place we could determine the range of observations on X_p. Suppose the minimum value observed is K' and the maximum K".

If we let $X_p = q$ where $q \in (K', K")$ we now want to determine the values of the remaining decision variables so that S is maximum or minimum. (Because a 7-point scale was used for the collection of the data K' = 1 and K" = 7).

For this purpose we have to solve the following linear program:

$$\begin{matrix}\text{Max} \\ \text{Min}\end{matrix} \; S = b_0 + b_1X_1 + \ldots + b_kX_k$$

subject to the following constraints:

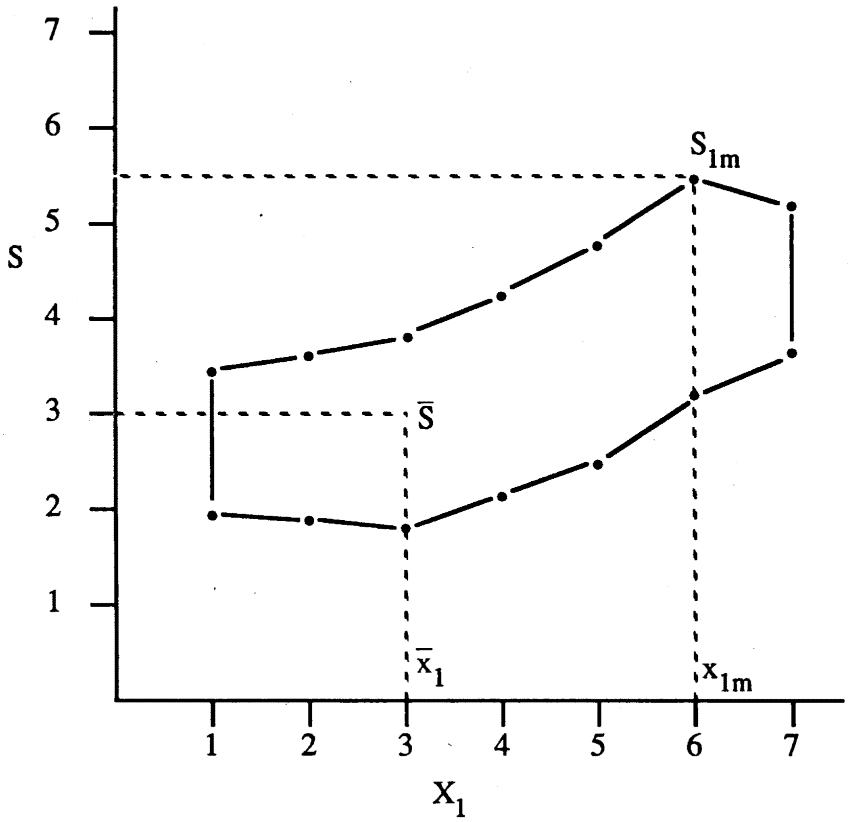

Figure 1 Graphic Representation of the Optimum Levels of S

$$\sum_{i=1}^{N} \lambda_i X_{ij} = X_j \geq 0 \text{ for } j = 1 \ldots k$$

$$\sum_{i=1}^{N} \lambda_i = 1, \lambda_i \geq 0 \text{ for } i = 1 \ldots N$$

i=1

$$X_p = q$$

The solution of the linear programming problem then yields the maximum (minimum) of S as well as the levels of the decision variables $X_1, X_2, \ldots, X_k$, where this optimal level is reached.

By doing this, a clear indication of the influence of the decision variable X_p on the dependent variable S is obtained (see figure 1) as well as the optimum solutions of the other decision variables where this optimum level is reached.

Prioritising

We are interested in a priority list of the decision variables $X_1, X_2, \ldots, X_k$ in terms of their importance regarding their influence on S.

Because of the fact that in this case K' = 1 and K" = 7 the LP will be solved by setting $X_1 = 1$ and maximise and minimise S. In the same way we will maximise and minimise S for $X_1 = 2,3,\ldots,K"$ (K" = 7). This optimum levels of S could be represented graphically as follows:

Suppose the maximum optimum value of S is S_{1m} as shown in the graph. This optimum level of S is reached when $x_1 = x_{1m}$ (6 in the example).

Suppose the optimum solution for the other decision variables at this point are x_{2m} — x_{km} and the measured mean values of x_1, x_2 — x_k are $\bar{x}_1, \bar{x}_2$ — $\bar{x}_k$.

The difference $v_1 = x_{1m} - \bar{x}_1$, gives an indication how much the attribute x_1 must be improved to reach it's optimum level. The same apply to all the other decision variables and a table of differences can now be constructed.

$$v_1 = x_{1m} - \bar{x}_1$$
$$v_2 = x_{2m} - \bar{x}_2$$
$$\vdots$$
$$vk = x_{km} - \bar{x}_k$$

Table 1
Differences between optimum levels of decision variables and their measured, mean values when S reaches the maximum optimum level.

As far we have calculated the differences of the measured mean values of the decision variables from the desired optimum levels.

Another important aspect to be considered in the attempt to prioritise these decision variables or attributes in order to know which of the factors are the most important in their contribution to the success of the information system, is the relative importance of the factors.

In the questionnaire the managerial users of the information system in the organisation have to evaluate the importance of the attributes in order to ensure that the overall information system should be effective and successful. The importance ratings of the attributes (x_1 — x_k) appearing in the restricted linear regression model can now be calculated from the measured values of importance.

These ratings (say I_{x1}, I_{x2} — I_{xk}) are used as an importance factor or weight and by multiplying the differences v_1 — v_k for each attribute by this weight the result is a list of priority factors indicating the priority of each attribute.

Attribute	Difference	Importance Factor	Priority Factor
x_1	v_1	I_{x1}	$p_{x1} = v_1 \times I_{x1}$
x_2	v_2	I_{x2}	$p_{x2} = v_2 \times I_{x2}$
.	.	.	.
.	.	.	.
.	.	.	.
x_k	v_k	I_{xk}	$p_{xk} = v_k \times I_{xk}$

Table 2
Priority factors for each attribute

The highest value of p_{xi} (i=1,2 k) indicates the attribute which should receive attention first. The second highest, second, etc. Negative priority factors means that the performance of the corresponding attributes are sufficient to ensure a successful information system.

Example

This example is taken from Bruwer [1]. There were seven decision variables in the restricted linear regression model namely DOWN, QUALIT, VOL, LEAD, ACCUR, ANAL and SECUR. The meaning of these abbreviations for the attributes is as follows:

DOWN – a low percentage of hardware and system downtime.
QUALIT – quality and competence of systems analysts employed by I/S department.
VOL – volume of output information provided by the system.
LEAD – short lead time required for new systems development.
ACCUR – accuracy of output information.
ANAL – user-oriented systems analysts who know user operations.
SECUR – data security and privacy.

In the model DOWN was restricted to the values 1,2 — 7 and the optimum values of SUCCESS and the decision variables calculated. The mean values of the decision variables and their importance ratings were calculated as well. Table 3 gives the variables mean values, the optimum values where the maximum optimum levels of success is reached, the differences v_1, v_2 — v_7, the importance ratings and the priority factors.

Attribute	Mean value	Opti. value	Diff.	Importance factor	Prior. factor
DOWN	4,18	5,00	0,82	6,23	5,20
QUALIT	4,78	5,80	1,02	5,90	6,02
VOL	4,60	5,40	0,80	5,45	4,36
LEAD	3,31	5,18	1,87	5,50	10,29
ACCUR	5,18	6,59	1,41	6,60	9,31
ANAL	4,43	6,59	2,16	5,85	12,64
SECUR	4,92	6,58	1,67	6,50	10,86

Table 3
Priority factors for the attributes

Using the priority factors the following priority list of the attributes can be constructed:

Priority	Attribute	Meaning
1	ANAL	user orientated analysts who know user operations
2	SECUR	data security and privacy
3	LEAD	short lead time required for new systems development
4	ACCUR	accuracy of output information
5	QUALIT	quality of systems analysts
6	DOWN	percentage of hardware and systems downtime
7	VOL	volume of output information.

Table 4

Although this priority list does not differs much from the one in Bruwer [1] it gives a more reliable priority of the attributes when the importance of the attributes is also taken into account.

Conclusions

It is of utmost importance for the top management of any organisation to monitor and evaluate the information system function on a continuous basis. Because of the fact that it is users who must use and live with the information system, evaluation from the user point of view is absolute necessary. In most organisations problems in the information system area exists and not necessarily technical problems. The methodology discussed above not only allows the management of an organisation to identify information system problems but also serve as a tool to take the importance of the problem into account. With this information it is possible to decide which of the problems must first receive the necessary attention.

References

[1] P J S Bruwer, [1987], Strategic Planning models for information systems, *Quæstiones Informaticæ*, **5** (3), 44-50.

[2] P J S Bruwer and J M Hattingh, [1985], Constrained Regression Models for optimisation and forecasting, *ORION*, **1** (1), 2-15.

[3] R M Alloway, [1980], User Manager's system needs, CSIR Working Paper, No. 56 (Sloan School of Management, M.I.T.), May 1980.

[4] P J S Bruwer, [1983a], Faktore wat die werkverrigting van rekenaargebaseerde inligtingstelsels beinvloed, *S. Afr. J. Bus. Magmt.*, **14** (1), 6-10.

[5] P J S Bruwer, [1983b], Evaluating the performance of computer-based information systems using a restricted linear regression model, *Quaestiones Informaticae*, **2** (3), 1-6.

[6] P J S Bruwer, [1984], A descriptive model of success for computer-based information systems, *Inf. & Management*, **7**, 63-67 Amsterdam.

NOTES FOR CONTRIBUTORS

The purpose of the journal will be to publish original papers in any field of computing. Papers submitted may be research articles, review articles and exploratory articles of general interest to readers of the journal. The preferred languages of the journal will be the congress languages of IFIP although papers in other languages will not be precluded.

Manuscripts should be submitted in triplicate to:

Professor ~~J M Bishop~~ D. G. Kourie
Department of Computer Science
University of the Witwatersrand
Johannesburg
Wits
2050

Form of manuscript

Manuscripts should be in double-space typing on one side only of sheets of A4 size with wide margins.

The first page should include the article title (which should be brief), the author's name and affiliation and address. Each paper must be accompanied by an abstract less than 200 words which will be printed at the beginning of the paper, together with an appropriate key word list and a list of relevant Computing Review Categories.

Manuscripts may be provided on disc ~~using any Apple Macintosh package or~~ in ASCII format, once that a submitted paper has been accepted.

For authors wishing to provide camera-ready copy, a page specification is freely available on request from the Editor.

Tables and figures

Tables and figures should not be included in the text, although tables and figures should be referred to in the printed text. Tables should be typed on separate sheets and should be numbered consecutively and titled.

Figures should also be supplied on separate sheets, and each should be clearly identified on the back in pencil with the authors name and figure number. Original line drawings (not photocopies) should be submitted and should include all the relevant details. Photographs as illustrations should be avoided if possible. If this cannot be avoided, glossy bromide prints are required.

Charges

Symbols

Mathematical and other symbols may be either handwritten or typewritten. Greek letters and unusual symbols should be identified in the margin. Distinction should be made between capital and lower case letters; between the letter O and zero; between the letter I, the number one and prime; between K and kappa.

References

References should be listed at the end of the manuscript in alphabetic order of the author's name, and cited in the text in square brackets. Journal references should be arranged thus:

[1] E. Ashcroft and Z. Manna, [1972], The Translation of 'GOTO' Programs to 'WHILE' programs, *Proceedings of IFIP Congress 71,* North-Holland, Amsterdam, 250-255.

[2] C. Bohm and G. Jacopini, [1966], Flow Diagrams, Turing Machines and Languages with only Two Formation Rules, *Comm. ACM,* **9**, 366-371.

[3] S. Ginsburg, [1966], *Mathematical Theory of Context-free Languages*, McGraw Hill, New York.

Proofs

Proofs will be sent to the author to ensure that the papers have been correctly typeset and *not* for the addition of new material or major amendment to the texts. Excessive alterations may be disallowed. Corrected proofs must be returned to the production manager within three days to minimise the risk of the author's contribution having to be held over to a later issue.

Only original papers will be accepted, and copyright in published papers will be vested in the publisher.

Letters

A section of "Letters to the Editor" (each limited to about 500 words) will provide a forum for discussion ~~of recent problems~~.