

# Solution Concepts for the Optimal Selection of Software Components

*Salah Merad and Rogério de Lemos*  
Department of Computer Science  
University of Newcastle upon Tyne, NE1 7RU, UK  
{salah.merad, r.delemos}@newcastle.ac.uk

## Abstract

We consider the problem of the optimal selection of software components from a library of components with respect to some non-functional attributes. Two potential solutions are presented, one based on utility theory, and the other based on a game theoretic formulation of the problem.

## 1. Introduction

In this position paper, we consider the problem of selection from a library of several components which have the same functionality, but differ in their non-functional attributes (NF-attributes). Examples of NF-attributes are reliability, performance, availability, which together specify the quality of service provided by a component, and cost. In the selection process, a trade-off between cost and quality of service has to be made.

A similar multi-criteria decision problem was addressed by Kontio /Kontio 96/ for which he recommended using the Analytic Hierarchy Process developed by Saaty /Saaty 90/. We propose two solution procedures: the first one is based on the construction of a combined utility function that quantifies the user's preference patterns over the available components, and the second one is the result of a game theoretic formulation of the problem. In the latter, a group or subset of components that have the same functionality is to be selected, and we propose an index which is a measure of the trade-off between the overall quality of service to be provided by the group and the total cost of the software to be built. We briefly discuss the conditions under which each of the solutions is applicable and the criteria which need to be considered to evaluate the merits of each solution procedure.

## 2. Optimal selection of components

Given a user's non-functional requirements (NF-requirements), the problem is to select the "best" component or subset of components from a library of components with respect to their NF-attributes.

### 2.1. The model

Let  $A = \{A_1, A_2, \dots, A_M\}$  be the set of components which satisfy the user's requirements. Let  $X_1, X_2, \dots, X_N$  be  $N$  NF-attributes that specify the quality of a component. Let  $\alpha_{mn}$  be the level of attribute  $X_n$  in component  $A_m$ ,  $1 \leq m \leq M$  and  $1 \leq n \leq N$ . Then, vector

$\mathbf{a}_m = (a_{m1}, a_{m2}, \dots, a_{mN})$  represents the profile of component  $A_m$ .

We distinguish two types of NF-requirements, strict and flexible. Without loss of generality, we define these terms for an attribute  $X_n$  in which the user's preference is for high levels over low levels. A NF-requirement for attribute  $X_n$  is strict when the level of the attribute in a component has to be equal or higher than a critical level  $cv_n$ , and it is flexible when the level can be lower than  $cv_n$ , though the user aspires for higher levels. The user's strength of preferences over the levels of attribute  $X_n$  are summarised in a utility function  $u_n(x_n; cv_n)$  whose expression is found empirically. The user's NF-requirements are specified by the vector  $\mathbf{cv} = (cv_1, cv_2, \dots, cv_N)$  and the individual utility functions  $u_n(x_n; cv_n)$ ,  $1 \leq n \leq N$ .

Let  $c_m$  be the cost of component  $A_m$ , and  $c^*$  the total amount of money available for expenditure on components. Without loss of generality, we assume that  $c_m \leq c^*$  for all  $1 \leq m \leq M$ .

## 2.2 Solution procedures

This is a multi-attribute decision problem for which there exists a large literature /Fishburn 70, Keeney 76/. For this problem, there is no obvious solution concept because there are many attributes, some of which are conflicting. We consider two solutions: one is an extension of the Von Neumann one-dimensional utility function to several dimensions, and the other is the result of a game theoretic formulation of the problem.

### 2.2.1. Combined utility function

It is shown that, under some independence assumptions between the attributes, it is possible to represent the user's preferences between alternatives with a combined utility function (CUF)  $U(x_1, x_2, \dots, x_N)$  which has a simple expression. For example, under the assumption of preference independence, in which the user's preference pattern over a subset of attributes is independent of the levels of the complementary subset, the CUF has the additive form

$$U(x_1, \dots, x_N; cv_1, \dots, cv_N) = \sum_{n=1}^N k_n u_n(x_n; cv_n),$$

where  $k_1, k_2, \dots, k_N$  are scaling coefficients through which the user's value trade-offs between the attributes are evaluated empirically. Note that both the individual utility functions and the CUF are scaled from 0 to 1. The hardest task is the evaluation of the scaling coefficients which is done through what are called indifference experiments. We illustrate such experiments with two attributes  $X_1$  and  $X_2$  in which the lowest levels are  $x_1^L$  and  $x_2^L$ , respectively. Let  $x_1$  be a possible level for attribute  $X_1$ . The user is then asked to find the level  $x_2$  for attribute  $X_2$  such that he/she is indifferent between the two components which have profiles  $\{x_1^L, x_2\}$  and  $\{x_1, x_2^L\}$ , respectively. The profiles  $\{x_1^L, x_2\}$  and  $\{x_1, x_2^L\}$  are thus equivalent and yield the same utility. It is easy to verify that, by combining the equality of the utilities with the form of the utility function, we obtain a linear equation. For  $N$  attributes, we need  $N$  linearly independent equations, and hence at least  $N$  indifference experiments need to be conducted. Because the evaluation process is subjective, it often yields inconsistencies which are difficult to eliminate

even after repeating the process many times. Note that the cost of components is one of the attributes, and hence the trade-off between cost and quality is incorporated in the CUF.

Kontio considered a similar problem and investigated two solution methods, the Weighted Scoring Method (WSM) and the Analytic Hierarchy Process (AHP) /Kontio 96/. He recommended the latter method because it is theoretically sound, it constructs the user's utility function for each attribute and it has a strong empirical validation. It is true that the AHP is superior to the WSM, but it has some shortcomings. The AHP is a method which yields an additive aggregate utility function which has the same form as the combined utility function under preference independence given above. In this aggregate utility function, the scaling coefficients are replaced by the weightings of the attributes, and the scaled individual utilities of the CUF are replaced by utilities evaluated by pairwise comparisons of the alternatives for each attribute separately. These utilities are in fact weights which add to 1, and which represent ratio scale preferences between the alternatives. Hence, although the AHP can include the cost of the components in the aggregate utility, it does not incorporate the user's value trade-offs; it is a heuristic which approximates the user's true preferences under the preference independence assumption. The AHP has an advantage in that there are simple statistical tests to check for consistency in the evaluation of the weighting coefficients and the individual utilities, but the power of these tests decreases rapidly with the number of alternatives and attributes.

### 2.2.2. A game theoretic solution

As mentioned above, the combined utility function requires the tedious task of the evaluation of scaling coefficients. Moreover, the NF-requirements for some attributes may be determined by the environment and hence uncertain (e.g., memory available to run an application at a given time). For these reasons, we need to define another solution concept which is appropriate for dealing with the uncertainties associated with the environment of a component based software system.

Note that although there may be uncertainty regarding the NF-requirements of some attributes, we assume that the user can quantify their relative likelihood with a completely specified probability distribution.

In this solution, we assume that, subject to the user's purchasing power, a group of different components all fulfilling the functional requirements is to be selected according to some criterion. The main issue here is how to evaluate a group of components. To distinguish between the quality of service provided by a group of components and their total cost, the cost attribute is not included among the  $N$  attributes.

Let  $\Omega$  be the set of subsets of  $A$  whose total cost does not exceed  $c^*$ . Let  $S$  be an element from  $\Omega$  containing the  $K$  components  $A_{m_1}, A_{m_2}, \dots, A_{m_K}$ , and  $C_S$  its cost. Note that  $C_S$  can be more than the total of its constituent components, because we need to add the cost of putting them together.

In any subset, there may be more than one component which satisfies the NF-requirements. If we try to optimise simultaneously all the attributes, then it can be shown that this selection problem can be formulated as a bargaining game whose solution achieves a compromise between the preference patterns of all the attributes /Merad 99/. The solution is obtained by maximising a function, called the Nash product, over part of the boundary of a well defined convex set - see /Nash 1950, Luce 57/. It is a randomised procedure in which the components of the subset are selected with some probabilities during run-time. The Nash product is

$$F(v_1, v_2, \dots, v_N) = \prod_{n=1}^N v_n^{\omega_n},$$

where  $v_n$  represents the expected utility for attribute  $X_n$ , and  $\omega_n$  is the weight allocated to attribute  $X_n$ , with  $0 \leq \omega_n \leq 1$  and  $\sum_{n=1}^N \omega_n = 1$ . These weights are evaluated empirically.

Let  $\Delta_S^*(\mathbf{cv}) = (c_{m_1}^*(\mathbf{cv}), c_{m_2}^*(\mathbf{cv}), \dots, c_{m_k}^*(\mathbf{cv}))$  be the resulting optimal selection strategy within the subset during run-time under NF-requirements  $\mathbf{cv}$ . The elements of vector  $\Delta_S^*(\mathbf{cv})$  are the probabilities with which the respective components are selected. We propose to characterise the subset  $S$  by the index

$$I_S(\mathbf{cv}) = \left( \prod_{n=1}^N \left( E_{\Delta_S^*(\mathbf{cv})}(x_n) \right)^{\omega_n} \right) / C_S,$$

where  $E_{\Delta_S^*(\mathbf{cv})}(x_n)$  is the expected level of attribute  $X_n$  with respect to probability distribution  $\Delta_S^*(\mathbf{cv})$ .

If the subset contains only one component, or only one component in the subset satisfies the NF-requirements, then  $E_{\Delta_S^*(\mathbf{cv})}(x_n)$  must be replaced by the level of the component for attribute  $X_n$ . If no component in the subset is feasible under requirements, then the index takes value 0. The index under all possible requirements, denoted by  $I_S$ , is then the average index under all possible requirements, that is

$$I_S = E(I_S(\mathbf{cv})),$$

where  $E$  is an expectation with respect to the distributions of the unknown requirements. If there are no attributes for which the NF-requirements are unknown, then the index is  $I_S(\mathbf{cv})$ .

Note that this index is to be computed during the design phase, and the subset with the highest index in the one that should be purchased.

*Example:* Suppose that there are two attributes specifying the quality of service provided by a component, each of weight 0.5:

$X_1$  represents the reliability of a component, and it is measured in average failures during a given length of time;

$X_2$  represents performance, and it is measured in the time taken to execute a number of tasks.

The following table gives the profiles of 3 components, together with the user's critical values and utilities which appear between brackets in bold:

	Reliability $X_1$	Performance $X_2$
<b>Critical Values Components</b>	1	0.8
$A_1$	0.85 <b>(1)</b>	0.80 <b>(0)</b>
$A_2$	0.90 <b>(0.5)</b>	0.75 <b>(0.5)</b>
$A_3$	0.95 <b>(0)</b>	0.70 <b>(1)</b>

Consider the two subsets  $S_1 = \{A_2\}$  and  $S_2 = \{A_1, A_3\}$ . In the latter subset, the Nash product is maximised by choosing each of components  $A_1$  and  $A_3$  with probability 0.5 /Merad 99/. The expected level of reliability is 0.9 and the expected level of performance is 0.75. The indices of the two sunsets are:  $I_{S_1}(1,0.8) = (0.9 * 0.75)^{1/2} / C_{S_1}$ , and  $I_{S_2}(1,0.8) = (0.9 * 0.75)^{1/2} / C_{S_2}$ . If the cost of component  $A_2$  is less than the total cost of subset  $S_2$ , then component  $A_2$  is preferred.

### 3. Conclusion and future work

We have proposed two solution concepts to the optimal selection of components under scarce resources, and there are others simpler but ad-hoc procedures /Merad 99/. The solution concept that should be adopted depends on the information the user can provide, the practicality of the computations, especially during run-time, and most importantly the trade-offs between the computational effort and the gain in utility over simpler but cruder methods. But the best way of carrying out such an evaluation is a properly designed experiment in a real situation using real decision makers /Chankong 83/.

### References

- /Chankong 83/ V. Chankong and Y.Y. Haines. *Multiobjective Decision Making: Theory and Methodology*. North-Holland, 1983.
- /Fishburn 70/ P.C. Fishburn. *Utility Theory for Decision Making*. Wiley, New York, 1970.
- /Keeney 76/ R.L. Keeney and H. Raiffa: *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Wiley, New York, 1976.
- /Kontio 96/ J. Kontio. "A Case Study in Applying a Systematic Method for COTS Selection". *Proceedings of the 18<sup>th</sup> International Conference on Software Engineering*. Berlin, Germany, March 25-30, 1996. Los Alamitos, CA: IEEE Computer Society Press, 1996. pp. 201-209.

/Merad 99/ S. Merad, R. de Lemos and T. Anderson. *Dynamic Selection of Software Components in the Face of Changing Requirements*. Technical Report No 664, Department of Computing Science, University of Newcastle, UK, 1999.

/Nash 50/ J.F. Nash: The Bargaining Game. *Econometrica*, 18, 1950. pp. 155-162.

/Saaty 90/ T. L. Saaty. *The Analytic Hierarchy Process*. McGraw-Hill, New York, 1990.

/Von Neumann 47/ J. Von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Wiley, 2<sup>nd</sup> edition, 1947.