SENSITIVITY ANALYSIS OF BAYESIAN NETWORKS IN COTS-BASED SOFTWARE DEVELOPMENT

Yousef Ibrahim Daradkeh

ABSTRACT
The process of developing software applications by integrating one or more Commercial Off-The-Shelf (COTS) components has received much attention lately because it provides potential benefits including shortening the development time, reducing effort and shrinking budgets as well as improving the quality of the final product. However, COTS-based development (hereafter CBD) in particular the evaluation and selection of COTS components, which is an essential activity in CBD, is not a trivial task and associated with various challenges. One of the most critical challenges is uncertainty inherent to COTS-related information and their vendors. Ignoring the uncertainty challenge negatively influences the quality of COTS selection decisions. In this paper, a bayesian-based evaluation model is extended to allow the allocation of various weights to evaluation criteria. We also investigate the impact of using various weights on the belief about the satisfaction level for various COTS candidates. Furthermore, the paper shows how the analytic hierarchy process (hereafter AHP) is used along with the model to rank various candidates. A digital library system is selected as an example to illustrate how the model along with AHP help decision makers to select the most promising COTS candidate.

KEYWORDS
Sensitivity Analysis Control, Bayesian Belief Networks, COTS-Based Software Development.

1. INTRODUCTION
CBD is defined as the development of new software systems by integrating one or more Commercial Off-The-Shelf (e.g. COTS) products [1]. The use of CBD offers several potential advantages such as shrinking budgets, lowering the development time and effort, improving the quality of the final software product, and offering functionality to stakeholders which might not have been requested initially but is still beneficial [2]. However, CBD is a non-trivial and risky task and associated with various challenges as uncertainty which is inherent to COTS-related information [3, 4]. Generally, uncertainty is an important issue required to be addressed in software engineering. It was stated in [5] that “Uncertainty is inherent and inevitable in software development processes and products.”

There are many definitions of uncertainty. One of these definitions is that, “Uncertainty is a general concept that reflects our lack of sureness about something or someone, ranging from just short of complete sureness to an almost complete lack of conviction about an outcome” [6].

In decision-making problems such as COTS selection, it is obviously preferred to have complete, consistent, and accurate information to make a good decision. Unfortunately, in the real situations the available information, on which decision makers rely to make decisions, is less than perfect. There may be missing, unknown, or ambiguous information that are of significant importance [7]. This is also the case of COTS selection. The uncertainty, the lack of being sure that we have all the information we need to select the best COTS candidate, is often a challenge and possibly influences the quality of the COTS selection decision. More details about various uncertainty forms related to CBD can be found in [8].

In this paper, the Bayesian belief network (hereafter BBN) based COTS evaluation model proposed in [8] is extended to investigate the impact of allocating various weights to evaluation criteria on the belief about the satisfaction level for various COTS candidates. Moreover, the paper shows how the analytic hierarchy process (hereafter AHP) is used within the model to rank various candidates. Therefore, the model has two outcomes; the first, named the satisfaction value created by AHP, represents how much each candidate satisfies the evaluation criteria and the second represents the probability that the candidate’s satisfaction is high and created by the Bayesian-based evaluation model during the process of evidence propagation and belief updating. The latter outcome represents our confident regarding the satisfaction level of various candidates. The model outcomes are used to select the best one among the COTS candidates. The overall model is a step towards conducting the COTS evaluation and selection process while the uncertainty is explicitly represented and managed.

The remainder of the paper is organized as follows: Section 2 presents an overview of related literature. Section 3 discusses the proposed model and explains how it can be used for evaluating various COTS candidates while explicitly representing uncertainty. In section 4, a
digital library example is introduced to demonstrate how the model can be used in a case study and to investigate the impact of changing the belief about the satisfaction of weighted evaluation criteria on the belief about the overall satisfaction level of COTS candidates. Furthermore, the outcomes of the usage of the model are presented and discussed. Finally, section 4 summarizes our conclusions.

2. RELATED WORK

There are already several methods for COTS evaluation and selection [9-15]. Reviewing these research efforts reveals that uncertainty is inherent and inevitable in CBD because the evaluation and selection process is cut short due to limited resources in terms of time, budget, and workforce allocated to the CBD project and various assumptions. Furthermore, uncertainty management is essential to improve the accuracy and quality of COTS selection decisions and ignoring it may lead to sub-optimal selection, low quality of the final solution, and stakeholders’ dissatisfaction [16]. However, the current COTS selection methods either partially addressed or did not address at all the uncertainty challenge and none of these selection methods consider uncertainty in a comprehensive manner. The following section discusses some of these research efforts that attempt to address the uncertainty and manage its possible consequences.

Comparative Evaluation Process (CEP) is a systematic and repeatable process used for evaluating and selecting COTS products. CEP realizes the importance of tackling uncertainty by assigning a credibility level to various data sources used during the evaluation process [17]. In [18], a Bayesian belief network-based approach has been proposed to certify the reliability of COTS software systems. [19] discusses how uncertainty related to COTS license cost can be managed and how possible risks associated with it can be mitigated. [20] realizes the importance of doing the COTS assessment while uncertainty is explicitly represented. It has adapted a model developed in [21] and used it for assessing a pair of COTS products considering the fault logs that might be available for the COTS candidates being assessed. However, none of these attempts can be considered as a complete solution that covers the evaluation/selection process taking into account various forms of uncertainty. This is because their focus was only on a specific evaluation perspective (e.g. reliability, license cost) not on the overall evaluation and selection process.

3. BAYESIAN-BASED EVALUATION MODEL OVERVIEW

The model uses AHP to rank various COTS candidates and Bayesian belief network to represent the uncertainty in terms of belief. The AHP is selected because it is a well-known decision making technique used by many existing COTS selection methods. AHP is also based on conducting pair-wise comparison which is more accurate than assigning an absolute value to each candidate. Furthermore, the AHP decision matrix involves a lot of redundancy in pair-wise comparison that enables the consistency check so as to reduce possible judgment errors. However BBN is selected because it is a well-known technique for solving problems that involve reasoning under uncertainty and it provides a graphically and mathematically sound technique for explicit representation of uncertainty. Figure 1 shows an abstract view for the model. Considering the causal relationship between various elements (i.e. COTS candidates, various evaluation criteria, and attributes) in COTS evaluation and selection problem, the model has three levels
- The upper level where COTS candidates nodes exist.
- The intermediate level where evaluation criteria nodes exist.
- The lowest level where attribute nodes exist.

Figure 1: Model abstract view

More details about the model and how the model can be used to represent the uncertainty can be found in [8]. In this paper, this model is extended to enable the allocation of various weights to the evaluation criteria. Allocating various weights is a way to represent the preferences of stakeholders towards the evaluation criteria. This means, weights reflect which evaluation criterion is more important than others from stakeholders’ perspective. Furthermore, the use of various weights enables decision makers to avoid the selection of a candidate that only satisfies the less important evaluation criteria. It is essential to consider the weights while the belief is updated and during the estimation of the satisfaction value for each candidate.

Regarding the belief update and how the weight is considered, assume that we have:

1. A set of COTS Candidates
   \[ C = \{ C_k, k = 1, 2, \ldots, l \} \]
2. A set of Evaluation Criteria
   \[ E = \{ E_j, j = 1, 2, \ldots, n \} \]
3. A set of attributes associated with each evaluation criterion.
   \[ A = \{ A_{i,j} \text{, where } i = 1, 2, \ldots, m"attribute-number" \} \]
   \[ \& j = 1, 2, \ldots, n"Criterion-number" \]
4. \( Bel(X_i) \): refers to the belief that the satisfaction level of the variable \( X_i \) is high (e.g. COTS candidate).
5. Once a new evidence about the satisfaction of the attributes is discovered, the message passing algorithm is used to update the belief in other
nodes (i.e., evaluation criteria and the candidates) according to the following equation [22]:

\[
Bel(X_i) = \alpha \lambda(X_i) \pi(X_i) \rightarrow (1)
\]

Where
\[\alpha: \text{a normalizing factor.}
\]
\[\pi: \text{refers to a message received from any of the } X_i \text{’s parent nodes. This message reflects the belief’s change in parent nodes.}
\]
\[\lambda: \text{refers to a message received from any of the } X_i \text{’s child nodes. This message reflects the belief’s change in child nodes.}
\]

6. In the COTS evaluation and selection problem, the forward propagation (i.e., propagate the evidence from attributes nodes towards the COTS candidate nodes) is only performed. So, \( \lambda \) is initially set to 1 and is not changed during the evaluation process. However, \( \pi \) messages are estimated as follows considering different weights allocated to the criteria.

\[
\pi(x_i) = \sum_{u_1, \ldots, u_n} P(x_i | u_1, \ldots, u_n) \prod_{j=1}^n \pi(u_j) w(u_j) \rightarrow (2)
\]

The above equation is a modified version of the equation appeared in [22]. Obviously, the new term \( w(u_j) \), representing the \( u_j \)’s weight that reflects its relative importance, is the extension introduced to the equation used to estimate \( \pi \) messages. For example, if \( x_i \) is a COTS candidate, \( u_i \) represents \( x_i \)’s parents (e.g., evaluation criteria), then \( w(u_i) \) represents weights allocated to these criteria.

3.1 MODEL USAGE

Figure 2 shows the steps applied during the model usage given that a set of COTS candidates have been already selected (i.e., 26 COTS candidates were identified and only four COTS candidates “COTS2, COTS3, COTS7, and COTS13” were selected to be used in the example). More details about the identification and selection process of COTS candidates can be found in [8]. The following sections discuss these steps briefly.

1. Weights Estimation

In this step which is performed before the actual usage of the model, weights are estimated and assigned to the evaluation criteria. The estimation process starts with developing a pair-wise comparison matrix. The matrix cells represent the relative importance of each evaluation criterion with respect to other criteria. Saaty’s scale is used to represent the relative importance values [23]. Then, the matrix is used to estimate the weights and a consistency check is performed to be sure the pair-wise judgments are consistent. If the judgments are inconsistent, the evaluation team and stakeholders will review the pair-wise comparison matrix and modify the values in the matrix to reduce the inconsistency. More details about how weights are estimated and how the consistency check is done can be found in [23].

2. COTS Information Review/Analysis

The information provided by COTS vendors through their websites, product documentations, product prototypes, etc. will be reviewed and analyzed by the evaluation team. The collected/reviewed/analyzed information will be used to update the belief in various nodes in the model and to rank various candidates using AHP technique.

3. Belief Updating

When the evaluation team discovers new evidence (e.g., during trying a prototype for COTS candidates, reading their documentation, or getting feedback from other users, etc.) for any of the attributes, the evidence propagation starts to update the belief in the nodes which have a relationship with those attributes according to equations 1 and 2.

4. COTS Ranking using AHP

AHP technique is used to rank various COTS candidates according to how much each candidate satisfies the evaluation criteria. The ranking process starts also with developing a pair-wise comparison matrix for each criterion. Each row contains a set of values representing the relative satisfaction values of a candidate with respect to other candidates. We suggest using the following steps to identify the relative satisfaction values.

- Based on the information provided by vendors, a set of absolute values (e.g., a value out of 100) measuring how much the candidates satisfy the criteria are assigned. Assume that \( A(C_i) \) refers to the absolute value of the candidate \( C_i \).
- Estimate the difference between the candidates \( C_i \) and \( C_j \) as follows:

\[
\text{Diff} = A(C_i) - A(C_j)
\]

- Based on how much is the difference, the relative satisfaction value is identified. Table 1 is an example that can be used to determine the relative importance values.

Once the pair-wise comparison matrix is developed, it will be used to estimate the overall relative satisfaction values for the candidates with respect to the criterion. After that, a consistency check is performed to ensure the consistency of judgments. The above process is repeated for each criterion. Finally, the satisfaction value...
measuring the overall performance of the candidate with respect to all criteria is estimated as follows:

\[ S_k = \sum_{j=1,...,n} W(E_j)S_{k,j} \rightarrow (3) \]

Where:
- \( S_k \): the satisfaction value for the candidate \( C_k \) (\( k=1,...,l \)).
- \( W(E_j) \): the weight for the criterion \( E_j \) (\( j=1,...,n \)).
- \( S_{k,j} \): the overall relative satisfaction value for the candidate \( C_k \) with respect to the criterion \( E_j \).

<table>
<thead>
<tr>
<th>Difference (X)</th>
<th>Assigned Value</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>Both candidates satisfy the criterion equally.</td>
</tr>
<tr>
<td>5&lt;=X&lt;10</td>
<td>3</td>
<td>( C_i ) satisfies the criterion somewhat more (i.e. weakly) than ( C_j ).</td>
</tr>
<tr>
<td>15&lt;X&lt;=25</td>
<td>5</td>
<td>( C_i ) satisfies the criterion much more (i.e. strongly) than ( C_j ).</td>
</tr>
<tr>
<td>30&lt;=X&lt;40</td>
<td>7</td>
<td>( C_i ) satisfies the criterion very much more (i.e. very strongly) than ( C_j ).</td>
</tr>
<tr>
<td>50&lt;X</td>
<td>9</td>
<td>( C_i ) satisfies the criterion absolutely more than ( C_j ).</td>
</tr>
</tbody>
</table>

Table 1: The use of difference between absolute values to identify the relative satisfaction

5. COTS Selection Decision

Considering the results produced by the belief updating process and the use of AHP, one of the candidates will be selected.

3. Digital Library Example

The digital Library system is selected as an example to demonstrate how the model is used to evaluate and rank various COTS candidates. Moreover, the example is used to investigate the impact of introducing weights on the belief updating and the satisfaction values. The digital library system is selected because it is a well-known system with rich functionality, and basic requirements sets are easily available and have been used in various other papers. In addition some other COTS selection methods [20] used it so we can easily compare our results with their results.

Three people (i.e. two have experience in library science and automation and the third has experience in software development) participated in the preparation of the case study and the usage of the model. The process consists of the following steps. Figure 3 shows the Bayesian-based evaluation model for the digital library system.

1. Weights Estimation

The participants start with constructing the pair-wise comparison matrix shown in table 2. The values in the table represent the relative importance of the criteria from the perspective of the participants. These values may change from project to project. This means, other participants may judge that the relative importance of catalogue search is much more important than the performance especially if the number of users accessed the system is small and the server configuration as well as Internet speed is high. In other projects, the three criteria may be equally important.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Catalogue Search</th>
<th>Performance</th>
<th>Vendor Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalogue Search</td>
<td>1</td>
<td>1/3</td>
<td>3</td>
</tr>
<tr>
<td>Performance</td>
<td>3</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Vendor Experience</td>
<td>1/3</td>
<td>1/7</td>
<td>1</td>
</tr>
<tr>
<td>Column Total</td>
<td>4.3333</td>
<td>1.4762</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 2: Pair-wise comparison between criteria

The following approximate method is used to estimate the weights [23]:
- Add the values in each column.
- Divide each entry in each column by the total or that column to obtain a normalized matrix.
- Adding the values in each row of the normalized matrix then dividing the results by the number of evaluation criteria.

The values 0.24, 0.67, and 0.09 are the estimated weights for catalogue search, performance, and vendor experience respectively. This means the performance criterion is the most important criterion followed by the catalogue search and finally the vendor experience. Once the weights are estimated, a consistency check is performed to check the consistency of participants’ judgments. By applying the consistency method in [23], the estimated consistency ratio is 0.007 which is less than 0.1. This means the judgments passed Saaty’s consistency condition and they are consistent.

2. COTS information review/analysis

The participants reviewed and analyzed the information about various candidates and started the process of introducing new evidences, belief updating and using AHP technique to rank the candidates. The following sections present the results of the belief update and the ranking of candidates produced by the AHP.

3. Belief updating

In this step, the participants introduced and propagated evidences that either increase or decrease the belief in various model nodes. Two cases are considered to investigate the impact of assigning different weights to the evaluation criteria.

- **Using similar weights**
  Similar weights are assigned to the criteria. This means, all of the criteria are equally important to the participants. Figure 4 shows the results.

- **Using different weights**
  Different weights means the criteria are not equally important to the participants. The weights, estimated in the weights estimation step, are assigned to the criteria and considered during belief updating. Figure 5 shows the
the results. Comparing the results of the two cases reveals the following points:

- **For COTS₂**, the probability that its satisfaction level is high increased 15% because COTS₂ satisfies the criteria with the highest weights (i.e. performance and catalogue search) more than other candidates.
- **For COTS₁₃**, the probability that its satisfaction level is high increased 3% because COTS₁₃ satisfies the catalogue search somehow more than COTS₃ and COTS₇.
- **For COTS₃ and COTS₇**, the probability that their satisfaction is high decreased 20% and 7% respectively because their performance with respect to the high weighted criteria was low.

Based on the above results, we conclude that satisfying the high-weighted criteria play a key role in increasing the belief that the satisfaction level is high and vice versa.

### 4. COTS ranking using AHP

AHP is used as follows to rank various candidates.

#### Constructing pair-wise comparison matrices

For each criterion, a pair-wise comparison which shows how much each candidate satisfies the criterion with respect to other candidates (i.e. relative satisfaction) is constructed. As mentioned to determine the relative satisfaction values, absolute values are assigned based on the collected information. Table 3 shows the absolute values for various candidates with respect to catalogue search.

<table>
<thead>
<tr>
<th>Table 3: absolute satisfaction values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Catalogue Search</strong></td>
</tr>
<tr>
<td>COTS₂</td>
</tr>
<tr>
<td>COTS₃</td>
</tr>
<tr>
<td>COTS₇</td>
</tr>
<tr>
<td>COTS₁₃</td>
</tr>
</tbody>
</table>

After that the differences between these values are estimated as shown in table 4.

<table>
<thead>
<tr>
<th>Table 4: Differences between absolute values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Catalogue Search</strong></td>
</tr>
<tr>
<td>COTS₂</td>
</tr>
<tr>
<td>COTS₃</td>
</tr>
<tr>
<td>COTS₇</td>
</tr>
<tr>
<td>COTS₁₃</td>
</tr>
</tbody>
</table>

**Table 5: Relative satisfaction with respect to the catalogue search**
Using the approximate method used previously in estimating the weights, the relative satisfaction values, shown in table 6, are calculated.

<table>
<thead>
<tr>
<th>Catalogue Search</th>
<th>Relative Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>COTS₂</td>
<td>42.97</td>
</tr>
<tr>
<td>COTS₃</td>
<td>8.77</td>
</tr>
<tr>
<td>COTS₇</td>
<td>19.99</td>
</tr>
<tr>
<td>COTS₁₃</td>
<td>28.27</td>
</tr>
</tbody>
</table>

Table 6: Relative performance for various candidates with respect to the catalogue search

By performing the consistency check, the consistency ratio is 0.03 which is less than 0.1. So, the relative satisfaction judgments are consistent. By repeating the same process for the performance and vendor experience, the results, shown in tables 7 and 8, are estimated.

<table>
<thead>
<tr>
<th>Performance</th>
<th>Relative Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>COTS₂</td>
<td>45.01</td>
</tr>
<tr>
<td>COTS₃</td>
<td>7.72</td>
</tr>
<tr>
<td>COTS₇</td>
<td>15.45</td>
</tr>
<tr>
<td>COTS₁₃</td>
<td>31.82</td>
</tr>
</tbody>
</table>

Table 7: Relative performance for various candidates with respect to the performance

<table>
<thead>
<tr>
<th>Vendor Experience</th>
<th>Relative Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>COTS₂</td>
<td>11.32</td>
</tr>
<tr>
<td>COTS₃</td>
<td>56.64</td>
</tr>
<tr>
<td>COTS₇</td>
<td>24.49</td>
</tr>
<tr>
<td>COTS₁₃</td>
<td>7.55</td>
</tr>
</tbody>
</table>

Table 8: Relative performance for various candidates with respect to the vendor experience

Using equation 3, the overall performance for the candidates with respect to all criteria is estimated and shown in table 9 and figure 6. For example, the overall performance for COTS₂ is estimated as follows:

\[
(42.97 \times 0.24) + (45.01 \times 0.67) + (11.32 \times 0.09) = 41.49
\]

<table>
<thead>
<tr>
<th>Overall Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>COTS₂</td>
</tr>
<tr>
<td>COTS₃</td>
</tr>
<tr>
<td>COTS₇</td>
</tr>
<tr>
<td>COTS₁₃</td>
</tr>
</tbody>
</table>

Table 9: The overall performance for various candidates with respect to all criteria

Table 10 summarizes the results produced in belief updating and COTS ranking using AHP.

<table>
<thead>
<tr>
<th></th>
<th>Belief Updating</th>
<th>AHP Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>COTS₂</td>
<td>0.89</td>
<td>41.49</td>
</tr>
<tr>
<td>COTS₃</td>
<td>0.51</td>
<td>12.38</td>
</tr>
<tr>
<td>COTS₇</td>
<td>0.60</td>
<td>17.35</td>
</tr>
<tr>
<td>COTS₁₃</td>
<td>0.76</td>
<td>28.78</td>
</tr>
</tbody>
</table>

Table 10: Model outcomes

4. CONCLUSION

In this paper we extended our proposed model to include the assignment of different weights to the evaluation criteria. The model is also used along with AHP to evaluate and rank various COTS candidates and to select the best candidate. The use of the extended model along with AHP is illustrated with the help of a digital library example. Using the results presented, we can conclude that:

1. Assigning weights to the evaluation criteria is more realistic than supposing that the criteria have the same weight and are equally important to stakeholders.
2. The overall performance of the candidate is influenced by how much it satisfies the high-weighted criteria.
3. The proposed model helps in assessing and ranking various candidates using two parameters. The first one, created by the Bayesian model, represents how much the evaluation team is confident regarding the satisfaction level of each candidate (i.e. probability that satisfaction level is high). However the second, created by AHP, represents how much each candidate satisfies the criteria (i.e. the overall performance).

REFERENCES