Analytical Support for Rapid Initial Assessment

Charles Twardy[*], Ed Wright[†,*], Kathryn Laskey[*], Tod Levitt[*], and Kellen Leister[*]

[*]George Mason University
4400 University Drive
Fairfax, VA 22030-4400
[ctwardy, klaskey, tlevitt, kleister]@gmu.edu

[†]Online Star, Inc
2515 Red Cedar Drive
Bowie, MD 20721
ewright@onlinestarinc.com

***Abstract*—This paper presents a methodology for rapidly assessing the likely effect of proposed initiatives. We have applied it to initiatives and metrics for the counter-IED fight, but the method is general. We build a probabilistic model that provides an explicit executable representation of the initiative's likely impact. The model is used to provide a consistent, explicit explanation to decision makers of the likely impact of the initiative. Sensitivity analysis on the model provides analytic information to support development of informative test plans. The method is designed to help when tight time constraints preclude or limit traditional test and evaluation methods, as is the case for initial decisions by the Joint IED Defeat Organization.**

Keywords-agile procurement; rapid assessment; Bayesian networks; test and evaluation

I. INTRODUCTION

In traditional military procurement, there is an extensive period of testing and evaluation before any new system is fielded. While effective, the traditional procurement approach can take years to field a new system. In some cases, we lack the time for testing. In others, expensive testing reveals results which in retrospect should have been obvious. We have developed a general structured method which in days to weeks can build a probabilistic model of the likely impact of the initiative. Such a model can help focus testing, or aid decisions that must be made before testing is complete. We have applied the model to counter-IED initiatives considered by the Joint IED Defeat Organization (JIEDDO).

Improvised explosive devices (IEDs) have become a weapon of choice in asymmetric warfare. Until recently they were responsible for the majority of casualties in Iraq, and recently they have become the leading cause of casualties in Afghanistan. JIEDDO's mission is to defeat IEDs as weapons of strategic influence. In particular, JIEDDO is expected to field new counter-IED (C-IED) initiatives much more rapidly than the traditional Department of Defense procurement process. Therefore, it cannot wait for the results of extensive testing: rapid funding (and re-funding) decisions must be made with limited information. The initiatives involve diverse technologies across a wide spectrum of potential C-IED applications, and are fielded in multiple theaters. Initiatives arrive for consideration on a frequent but irregular basis.

JIEDDO has developed a streamlined acquisition process – the Joint IED Capability Approval and Acquisition Management Process (JCAAMP). Key features include the irregular arrival of new initiatives, and very rapid turnaround. Our method is not specific to JCAAMP, but but these features compel an agile method such as ours. JIEDDO must rapidly assess the value to warfighters and decide within weeks whether or not to fund an initiative. Thus, measures of effectiveness for new initiatives must be developed very rapidly. Ideally, these measures should be comparable across different initiatives. The method should identify parameters for further data collection. Then when additional test, operational or field data are collected, it should be possible to update these measures and metrics based on the new information.

It follows that our initiative assessment methodology must provide an analyst with a way to rapidly:

- Formulate analytic measures or metrics for each initiative that are comparable across initiatives;
- Generate an explicit analytical representation of the explanation for how the initiative will work;
- Predict the qualitative impact of the initiative on consistent and comparable measures or metrics;
- Use data when available to estimate those measures or metrics for new initiatives;
- Identify parameters for which additional testing would have significant payoff; and
- Update those same measures and metrics based on new test, field or operations data.

In this paper we present such an initiative assessment

The research reported in this paper was supported by the Joint Improvised Explosive Device Defeat Organization (JIEDDO) under the Office of Naval Research (ONR) US Government contract number N00014-09-C-0419. The views, claims and opinions of the authors expressed in this paper do not necessarily state or reflect those of the United States Government or any of its agencies, and shall not be represented as such.

methodology and demonstrate its application to a case study.

II. Modeling Approach

We created a structured modeling method for initiative analysis, using Bayesian networks (BNs). BNs provide an intuitive graphical representation of causal dependencies, and propagate uncertainties through the model. The method has two top-level steps:

1. Identify the relevant Measures of Performance (MOPs). Using a common, consistent set of MOPs allows comparison of diverse initiatives. [1]
2. Model the dependence of the MOPs on system and environmental variables. The model is high-level, often reflecting only qualitative assessments of the influences.

Because engineering test results are not available at initial assessment, the modeling approach must exploit knowledge in other forms. Typically, knowledge comes from Subject Mater Experts (SME), from requirements documents from the field, from the contractor who is proposing the initiative, and from experience with previous initiatives. This information is assembled into a BN to predict the likely impact of the initiative.

The impact assessment methodology should also:

- Prioritize future information collection.
- Integrate with portfolio management, to optimize investment in a set of initiatives.
- Provide a consistent, repeatable, and extensible model.

III. Methodology

This section discusses the rapid assessment methodology and the sensitivity analysis metrics that we use.

A. *Rapid assessment methodology*

Our methodology has 6 steps, similar to the knowledge engineering methodologies described in [1] and [2]. (In reality, there is continuous feedback and frequent interaction with SMEs, as depicted in Figure 1.) The steps are:

1. Identify MOPs:

- Create a variable for each MOP.
- Specify clear operational definitions for each variable.
- Determine the state space for each variable.
- Identify primary indicators of the MOPs. Connect each MOP directly to a variable that toggles the initiative on and off (or switches among alternatives). Estimate the MOPs using the model. Assess the model to rank the MOPs according to need for refinement.

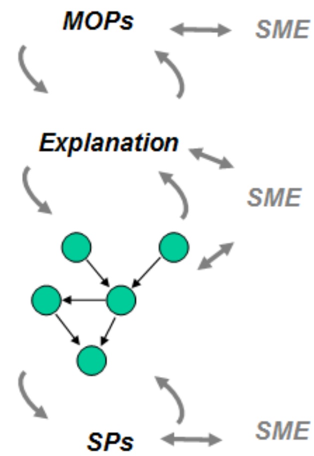

Figure 1. The Rapid Assessment Process.

2. Generate an Explanation of how the initiative is expected to affect the MOPs.
3. Implement the explanation as a probabilistic model. This step is a loop where we repeat until satisfied or out of time:

- Select most important variable to refine (the target).
- Refine definitions and state space for the variable. (For example, transform a qualitative “high, med, low” variable to a quantitative one.)
- Identify the “first-order” causes and effects of the target.
    - Identify the primary causes of the target.
    - Identify any additional key indicators (typically effects that are easier to measure than the target itself)
    - Create variables for the causes and indicators
    - Specify clear operational definitions for each variable
    - Determine the state space for each variable
- Determine the dependence relationships among the variables. Estimate local distributions.
    - Determine the structural assumptions for the local probability distributions.
    - Determine the values of any free parameters.

[1] In our case, MOPs have been identified by JIEDDO for important classes of C-IED initiatives.

- Select various combinations of causes (parents) and indicators (children), and check that results are in line with expectations, or justified. Modify and recheck as required.
- Document assumptions & limitations. Quantify uncertainties and bound errors, if possible. Determine what you most need to know next.
- Evaluate the model.
    - Internally, by team review
    - Internally, via sensitivity analyses
    - Externally, by consulting with SMEs

4. Execute & Analyze the Model to Assess Performance – there are several possibilities, including:
    - Set input variables for use cases, and evaluate the predicted effect on various MOPs for those cases.
    - Embed the model in a simulation-style scenario, and "roll up" performance for the whole scenario.
    - Systematically vary the value of a variable or parameter, and graph the effect on MOPs. This overlaps with the next step.
5. Determine the Sensitive Parameters (SPs). Create final ranked list of SPs for each MOP. Use both subjective judgment from the model-building and formal methods such as those described under "Sensitivity Analysis," below.
6. Report Findings

B. Sensitivity Analysis

We employ four kinds of formal sensitivity analysis:

1. Global sensitivity to findings: Mutual Information;
2. Local sensitivity to findings: Link Strength;
3. Sensitivity to particular parameters (probabilities) in the model;
4. Change plots (*dx*/*dy*) for sensitivity to particular parameters identified in previous steps, and of practical interest (e.g. because we can test or control them)..

We describe these below.

1) Mutual Information: The mutual information between X and Y is the amount of uncertainty in Y that we eliminate by knowing X (and vice versa). Information is measured in bits, and is formally equivalent to the number of well-chosen yes/no questions we would need to determine the actual value of the variable. For example, let X be the unknown outcome of a fair coin toss. Because a single question ("Was it heads?") will tell us the answer, there is 1 bit of information in the variable. Now suppose that X is the height of a person. Let Y be their sex. Learning their sex will reduce our uncertainty about their height. The mutual information measures that change in certainty. By how many yes/no questions have we reduced the uncertainty? Formally, it's just the difference in information. Let MI(X,Y) be the mutual information between X and Y, and let H(·) be the information in a variable, and let "|" mean "given". Then:

$$MI(X,Y)=H(X)-H(X|Y)$$

Mutual information is an absolute measure whose scale varies with the number of states of the variables – we need more questions to determine the outcome of a die roll than a coin toss. Therefore, we consider three variants, all on a scale from zero to 1.

Scaled MI uses a scale in which 1 is the MI of a uniform distribution on Y. This is useful for tracking progress in learning Y, such as in a sensor-tasking system, since it provides a stable reference.

Normalized MI uses a scale in which 1 is the highest MI in the current set **X** of potential measurements. This presents each potential variable to observe as a proportion of the best one.

C_{XY} uses a scale in which 1 is the current $H(Y)$. It represents the proportion of uncertainty reduced, so that 1 means that X fully determines Y.

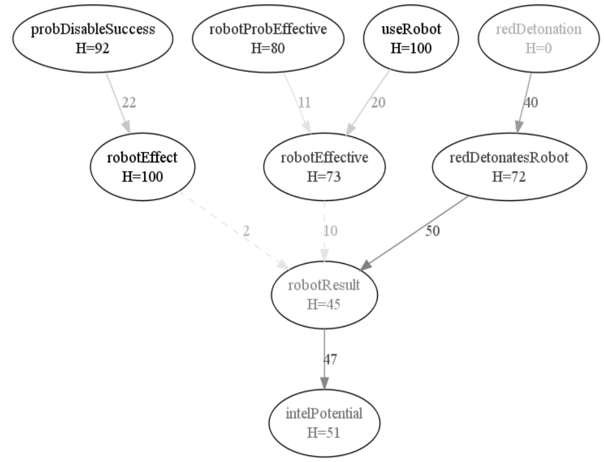

Figure 2: Example Link Strength Graph for Intelligence Potential.

2) Link Strength: MI is defined between any two nodes, or sets of nodes. However, if there are multiple paths, it might be that one carries most of the influence. A link strength measure allows us to examine the individual influence of each arc.

Ebert-Uphoff defined several measures of link strength based on Mutual Information, drawing on the earlier work by Nicholson & Jitnah. The two most important are true average link strength (*LST*) and blind average link strength (*LSB*). Both are based on the mutual information between *X* and *Y*, conditional on **Z**, the set of all the other parents of *Y*. Conditioning on the other parents isolates the influence of X alone, mimicking an intervention. *LSB* makes simplifying assumptions and can be calculated without performing any inference at all. We take a slightly different approach.

Cut link strength compares $P(Y|x)$ with and without the link $X \rightarrow Y$. This was the "gold standard" that Nicholson & Jitnah [3] used to evaluate their (link-strength-like) approximate inference. But we can afford to use the gold standard itself.

When cutting the arc $X \rightarrow Y$, we average over *X*. This operation does not change the overall (marginal) distribution on *Y*. However, unless the arc was spurious, the new $P(Y|x)$ will differ from the old for at least some $x \in X$. To isolate the influence of *X* alone, we use an intervention operator (denoted "||") rather than a regular conditioning operator ("|"). It has much the same effect as fixing all the other parents at all values, but is more efficient. In symbols, let $P(Y||x)$ be the distribution in the original graph, and let $Q(Y||x)$ be the distribution in the new graph, with $X \rightarrow Y$ cut. The link strength is the expected distance between these two distributions:

$$\sum_{x \in X} P(x) \times \text{Distance}[P(Y||x), Q(Y||x)]$$

We considered two Distance functions, Kullback-Leibler divergence (KLD) and NonOverlap. Although KLD is the closest to MI, it is highly nonlinear and hard to interpret. Therefore we used 1-Overlap:

$$1 - \text{Overlap}(p, q) = 1 - \sum_{x} \min(p_x, q_x)$$

1-Overlap is a true distance measure ranging from 0 (identical) to 1 (no overlap).

Ebert-Uphoff wrote his scripts for the Matlab-based Bayes Net Toolbox (BNT) [5] and Intel's Probabilistic Network Library (BNT's C++ offspring) [6]. We implemented our variant in Quiddity*Script [7]. It would be relatively easy to do the same for Netica [8]. Like Ebert-Uphoff, we rely on Graphviz [9] for the actual graph drawing. Figure 2 shows an example.

3) *Sensitivity to CPT Parameters*

If *y* is continuous, then by definition, $\text{Sensitivity}(y|x,e) = \frac{\partial p(y|e)}{\partial x}$, which gives the slope along *x* of $p(y|e)$ near the current value of *x*. For example, *x* may be a particular probability in a CPT, such as *P*(tuberculosis=true | xray=true). There are efficient methods to calculate $\frac{\partial p(y|e)}{\partial x}$ using only 3 inference propagations, after which querying for that slope at any *x* is constant time. However, even without that, we can just vary the parameter over its range, and plot the effect on the MOPs of interest.

IV. Example

In this section we apply the rapid assessment methodology to a generic explosive ordnance disposal (EOD) robot. Any EOD robot provides a capability to remotely neutralize an IED, either by disabling it or detonating it. We assume that if the robot is unavailable or unsuccessful, an EOD soldier will neutralize the IED.

To develop the model, we executed the five steps of the assessment methodology:

1. Identify relevant MOPs.
2. Generate an Explanation of how the initiative is expected to affect MOPs.
3. Implement the explanation as a probabilistic model.
4. Execute & analyze model to assess performance
5. Determine the sensitive parameters (SPs) to help prioritize information collection.
6. Report findings.

C. *Identify MOPs*

Figure 4 shows the MOPs deemed relevant, and the assumptions and considerations to use in the model. Note that the robot does not affect detection, so there are no Detection MOPs.

MOP	Assumptions and Considerations
Time	Robot may take longer than an EOD soldier If the robot is unsuccessful, we still must use a soldier
P(neutralize *by robot*)	Distinguish *disable* from *destroy*
Casualties or Damage per Attack	• Replace with generalized, qualitative P(damage) • If Red detonates the IED during robot neutralization, soldiers are not exposed. The robot may be damaged or lost. • If the robot is unavailable, or fails, then a soldier will be at risk. • If the IED is not spotted, robot has no effect on damage / casualties.
P(collecting valuable Intelligence)	• If Blue *disables* the IED, it can be examined for forensic intelligence. • If Blue *detonates* it, there may be some intelligence collected before the detonation. • If Red detonates it, there is little intelligence gained.

Figure 3. MOPs for the EOD Robot.

D. Generate an Explanation

The explanation describes the influences of important system and environmental variables on the MOPs. In this explanation, we assume that an IED is present and has been successfully detected.

- If a robot is available and working correctly, it can be used to attempt to disable or detonate an IED.
- If there is a Red detonation during neutralization, Blue soldiers are not exposed. The robot may be damaged or destroyed.
- If the robot is not available or not successful, a soldier will be at risk while disabling the IED
- If the robot succeeds in disabling the IED, we can gather forensic intelligence.
- Little intelligence can be collected if the robot detonates the IED.
- Using the robot may take longer than using an EOD soldier.
- If unsuccessful, a soldier must still disable the IED.

E. Implement the explanation as a probabilistic model.

Our explanation can be transformed directly to a structural model, or graph, as shown in Figure 4. For example, the top three nodes allow us to express that we will only use the robot if it is available (on this RCT) and ready. The MOP `clearTime` depends on the `robotResult`: was the robot used, did Red detonate the IED against the robot, or did it work (and if so, did it disable or destroy the IED)?

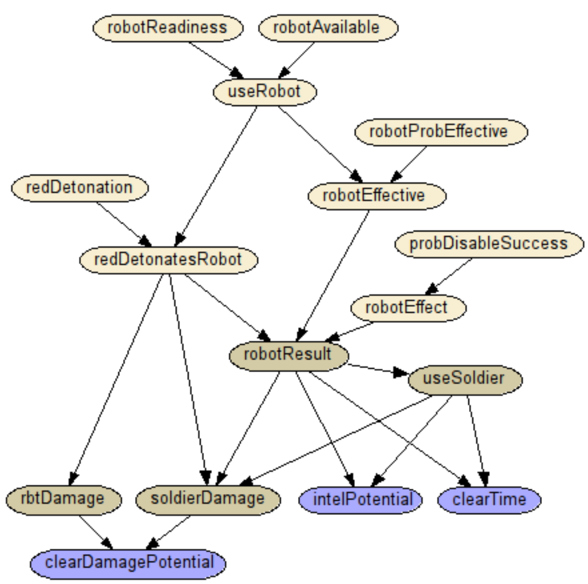

Figure 4. The Robot Explanation Model.

The next step is to specify the domain of each variable.

In practice, the domain evolves with the struture, as modeling choices are made. The model shown here is already the 6th revision. The revision incorporates feedback from modelers unfamiliar with Bayesian networks, to make it more intelligible.

Local probability distributions for each node are generated based on an available knowledge. Without engineering test data, they will necessarily be qualitative.

F. Execute & analyze model to assess performance

The quickest and most intuitive analysis is to interact with the model in a live session. The following screenshots are taken from the Netica GUI[2].

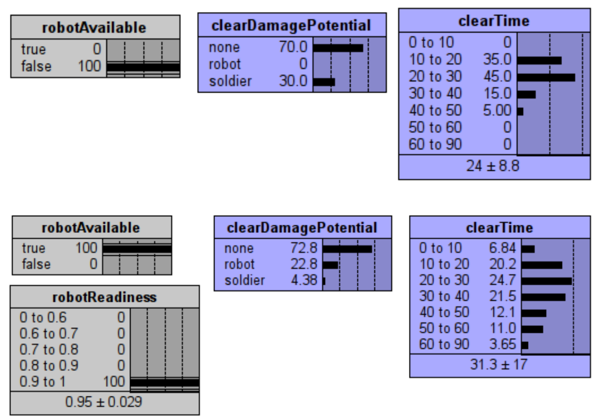

Figure 5. Model Results Showing Impact of Robot Availability on Damage Potential and Clear Time (minutes).

Figure 5 shows that if the robot is not available, then a soldier is at risk while disabling the IED (top), and that if a robot is available and it is working correctly, it takes most of the soldier's risk, and affects the clear time.

These distributions are the logical consequences of our explanation and assumptions. We should not believe the three-decimal-place estimate of a 6.84% probability of disposal in under 10 minutes. But given our assumptions, we should believe the robot increases the time, roughly as shown. The wide distribution on clear time averages over the distributions for various unobserved ancestor variables (conditioned on downstream evidence, if any).

Finally, our Intelligence MOP reflects our understanding that if the robot succeeds in disabling the IED, it can be examined for forensic intelligence. Less intelligence can be collected if the robot detonates the IED.

---

[2] The Netica GUI is cleaner and more compact than the Quiddity GUI.

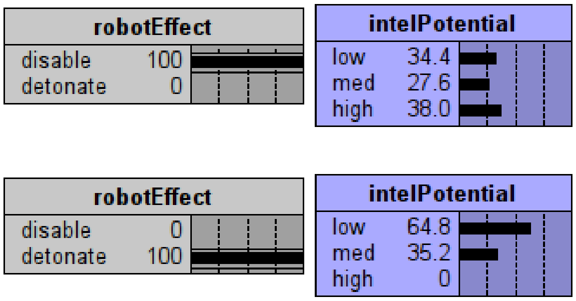

Figure 6. Model Results Showing Impact of Robot on Intelligence Collection.

G. *Determine the sensitive parameters (SPs)*

Executable models foster sensitivity analysis. In a Bayesian network, we look first at the mutual information between variables. The mutual information between X and Y is the amount of uncertainty in Y that we eliminate by knowing X (and vice versa). Table 1 shows the Top 5 most influential variables for each of the three main MOPs, assuming `robotAvailable` has value `true`.[3]

The key performance parameters (KPPs) here are δ = P(Red Detonates) and the conditional probability ρ = P(Red Detonates on Robot | Red Detonates). In our model, they dominate intrinsic parameters such as reliability (Readiness) and effectiveness. After all, the main reason for using the robot is to prevent casualties.

e	Intelligence	Damage
Red Detonates on Robot	Red Detonates on Robot	Red Detonation
Red Detonation	Red Detonation	Red Detonates on Robot
Readiness	P(Disable Success)	P(Effective)
P(Effective)	P(Effective)	Readiness
P(Disable Success)	Readiness	–

Table 1: Robot: Top 5 Sensitive Parameters by MOP. Assumes the robot is available, and excludes uninteresting nodes. Names are made into readable English.

Table 1 ranks the MOPs. We could look at the mutual information values themselves, but those represent average effects. Figure 7 plots three MOPs versus ρ, Red's tendency to target the robot. The effect is quite strong, in part because ρ=0 is equivalent to having a perfectly reliable robot, and ρ=1 is like having no robot.

That said, the probability of a causalty drops from 80% to 55%, and our chance of getting "High" intelligence drops from 50% to 0. This follows from our assumption that a robot is used if it is available and working. If Red destroys the robot, we lose our chance for gathering intelligence.

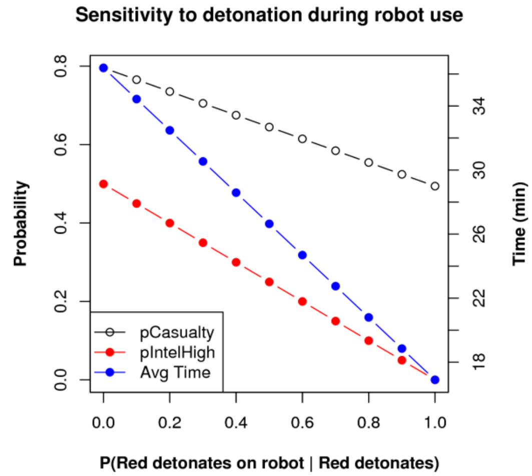

Figure 7. Sensitivity Analysis showing the influence of Red tactics: a command detonation of the IED on the robot.

More dramatic, but far less interesting, the average time drops in half, from 36 min to 16 min. This merely reflects the fact that once the IED detonates, we don't have to try to disable it anymore.

V. CONCLUSION

This rapid initiative evaluation methodology provides a structured approach for assessing initiatives even when there is little formal test data. We begin with relevant MOPs, explain how the initiative will affect the MOPs, and build a probabilistic model of that explanation. The graphical structure *shows* the influences. Adding local probability distributions makes the model executable. Executing the model provides average effects, what-if scenarios, and sensitivity analysis. These in turn can guide formal testing.

REFERENCES

[1] Kevin B. Korb and Ann E. Nicholson. *Bayesian Artificial Intelligence*. CRC Press, 2003.

[2] Kathryn B. Laskey and Suzanne B. Mahoney. Knowledge engineering for complex belief networks. In *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence (UAI'96)*. UAI, 1996. `http://ite.gmu.edu/%7Eklaskey/papers/neteng.pdf`.

[3] E. Nicholson and N. Jitnah. Using mutual information to determine relevance in bayesian networks. In *Proceedings of the 5th Pacific Rim International Conference on Artificial Intelligence*. Springer, 1998.

[4] Imme Ebert-Uphoff. Measuring connection strengths and link strengths in discrete Bayesian networks. Technical Report GT-IIC-07-01, Georgia Institute of Technology, 2007. `http://hdl.handle.net/1853/14331`.

[5] Kevin Murphy. Bayes Net Toolbox. Open Source Software, October 2007. `http://www.cs.ubc.ca/~murphyk/Software/BNT/bnt.html`.

[3] Excluding uninteresting variables such as deterministic indicators.

[6] Intel. Probabilistic network library. Open Source Software, 2005. http://sourceforge.net/projects/openpnl/.

[7] IET, Inc. Quiddity. Commercial Software, 2007. Official site: http://www.iet.webfactional.com/quiddity.html. Discussion group: http://groups.google.com/group/quiddity. Quiddity is now provided by OLS, Inc.

[8] Norsys Software Corp. Netica. Commercial Software, 2008. http://norsys.com/.

[9] AT&T Research. Graphviz. Open Source Software, 2008. http://www.graphviz.org/.

[10] U.S. Navy. Explosives ordinance disposal missions and equipment. Website, Visited March 2009. http://www.navy.com/about/navylife/onduty/eod/missionsandequipment/.