Rapid Initiative Assessment for Counter IED Investment

Charles Twardy[1], Ed Wright[2], Kathryn Laskey[1], Tod Levitt[1], and Kellen Leister[1]

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

There is a need to rapidly assess the impact of new technology initiatives on the Counter Improvised Explosive Device battle in Iraq and Afghanistan. The immediate challenge is the need for rapid decisions, and a lack of engineering test data to support the assessment. The rapid assessment methodology exploits available information to 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 on the likely impact of the initiative. Sensitivity analysis on the model provides analytic information to support development of informative test plans.

1 INTRODUCTION

The mission of the Joint Improvised Explosive Device (IED) Defeat Organization (JIEDDO) is to defeat IEDs as weapons of strategic influence. In support of this mission, JIEDDO has put in place a process to field new counter-IED initiatives much more rapidly than the traditional Department of Defense procurement process. 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. If JIEDDO is to meet its charge for rapid response to IED threat, it cannot wait for the results of extensive testing. Instead, 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 J9 Division[1] has called the problem of deciding which initiatives to fund "Portfolio Selection." Across a period of weeks or months, there is a set of initiatives in various stages of consideration for initial or continued funding. The funded initiatives and funding levels constitute a "portfolio" selected from all other alternatives of funding/non-funding and possible amounts to fund for each initiative.

There are many possible analytic Portfolio Selection formulations, but implementing any of them requires an implicit or explicit ranking of alternative portfolios. A ranking, in turn, requires a way to measure, or at least to place bounds on, the value of a portfolio. Measuring the value of a portfolio requires in turn that one or more measures of value be associated with individual initiatives in the portfolio. These measures must be comparable across initiatives, so that the portfolio selection process can compare portfolios containing different sets of initiatives.

It is not obvious that measures can be developed that are comparable across the range of initiatives JIEDDO must consider. For example, how should the value of a new jamming mechanism be compared against of the value a new unmanned surveillance platform, or a newly deployed military-intelligence team? Plausible measures such as casualty-avoidance potential are so high-level and context-specific that they are of little use in evaluating the performance of a specific initiative in its intended context. For example, the casualty avoidance potential of a jammer depends critically on the context in which it is employed: if no radio-controlled IEDs are encountered, or if any encountered IEDs are disabled without the necessity of jamming, then even a highly effective jammer will not reduce casualties. We seek conditional measures: given that the jammer encountered an IED that is susceptible to jamming, what is the chance the jammer prevented detonation?

Furthermore, measures for new initiatives must be developed very rapidly, because the time to consider initiatives before making funding decisions must be as short as possible. The method should identify parameters for further data collection. Then when additional test, operational or field data is collected, it should be possible

[1] J9 is Operations Research & Systems Analysis (ORSA).

to update these measures and metrics based on the new information.

It follows that a JIEDDO 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 and therefore can be used as inputs to the Portfolio Selection analysis;
- 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 methodology and demonstrate its application to a case study.

2 MODELING APPROACH

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

1. Identify the relevant Measures of Performance (MOPs). MOPs have been identified by JIEDDO for important classes of C-IED initiatives. Using a common, consistent set of MOPs allows comparison of diverse initiatives.
2. Model the dependence of the MOPs on system and environmental variables. A Bayesian Network (BN) model is developed to represent the influence of important system and environmental variables on the MOPs. By necessity, the model is general, in many cases reflecting only qualitative assessments of the influences.

Because engineering test results are not available at initial assessment, the modeling approach must exploit knowledge in whatever form it is available. Typically, knowledge comes from Subject Mater Experts (SME) at JIEDDO and elsewhere, 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 must also:

- Enable identification of information collection priorities for future testing.
- Integrate with a portfolio management process, which will optimize investment in a set of C-IED initiatives.
- Provide a consistent, repeatable, and extensible model. The BN methodology provides an explicit model, integrating all available knowledge that can be extended when additional information becomes available form testing.

3 METHODOLOGY

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

3.1 RAPID ASSESSMENT METHODOLOGY

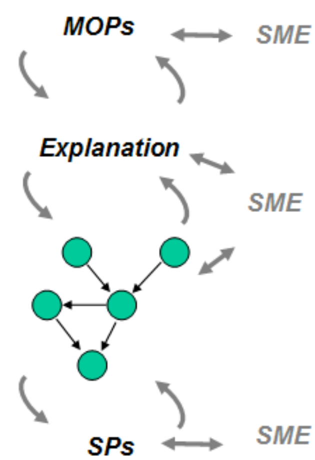

Figure 1. The Rapid Assessment Process.

Although the methodology is described below as a 5-step process, in practice it is a non-linear process involving continuous feedback and frequent interaction with SMEs, as depicted in Figure 1.

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.

2. Generate an Explanation.

3. Implement the explanation as a probabilistic model: repeat until satisfied or out of time:

- Select most important variable to refine. Make that 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.
- Assess the target variable
- Select various combinations of causes (parents) and indicators (children). Instantiate variables. Assess whether results for the target are in line with expectations, or at least justified. Modify and recheck as required.
- Select various states of the target and evaluate distributions of parents and children, to ensure they are also 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 the SMEs

4. Assess the model

- Perform global sensitivity analysis and consistency checking to evaluate model adequacy.
    i. Mutual information tables;
    ii. Link strength graphs;
    iii. $\frac{dx}{dy}$ plots for select parameters identified in previous steps, and of practical interest (e.g. because we can test or control them).

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:

- Mutual information table;
- Link strength graphs;
- $\frac{dx}{dy}$ plots for select parameters.

3.2 SENSITIVITY ANALYSIS

We employ four main kinds of sensitivity analysis:

1. Global sensitivity to findings: Mutual Information;
2. Local sensitivity to findings: Link Strength;
3. Sensitivity to CPT parameters;
4. $\frac{dx}{dy}$ plots for sensitivity to CPD[2] parameters.

We describe these below.

3.2.1 Mutual Information

Mutual information measures the information gained about one variable by learning another. Let X be a factor of interest, and let Y be a MOP or other variable of interest. Let $MI(X,Y)$ be the mutual information between X and Y, and let $H(\cdot)$ be the entropy in a variable. 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 might 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.

---

[2]Conditional Probability Distribution – the local tables or functions "inside" a node in a Bayesian network or similar graphical probability model.

$\mathbf{C_{XY}}$ 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.

3.2.2 Link Strength

MI is defined between any two nodes, or indeed sets of nodes. However, it tends to decrease with the number of links between X and Y, because we usually lose certainty with each step. For example, although the chain $X \xrightarrow{.9} A \xrightarrow{.9} B \xrightarrow{.9} C \xrightarrow{.9} D \xrightarrow{.9} Y$ has strong links at each stage, the aggregate influence $X \cdots \xrightarrow{.6} \cdots Y$ is not very strong. Conversely, as Ebert-Uphoff notes, high mutual information does not reveal which of multiple paths carries the weight; indeed, some paths may be quite weak. A link strength measure allows us to examine the individual influence of each arc of interest.

Ebert-Uphoff Link Strength:

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 (LS_T) and blind average link strength (LS_B).

LS_T is the mutual information between X and Y, conditional on $\mathbf{Z}$, the set of all the other parents of Y.

$$LS_T(X \to Y) = MI(X, Y|\mathbf{Z}) = H(X|\mathbf{Z}) - H(X|Y, \mathbf{Z})$$

LS_T averages over X and Z using the actual joint distribution. In contrast, LS_B assumes "that X,$\mathbf{Z}$ are independent and all uniformly distributed," which gives us a simplified version of MI that we can calculate solely by inspecting conditional probability table for Y, without performing any inference at all. As Ebert-Uphoff notes, this purely local measure is often quite useful, such as when evaluating an expert-specified CPT.

<u>Cut Link Strength</u>:

Another approach is to compare $P(Y|x)$ with and without the link $X \to Y$. This was the "gold standard" that Nicholson & Jitnah (1998) used to evaluate their (link-strength-like) approximate inference. But we can afford to use the gold standard itself.

When cutting the arc $X \to Y$, we marginalize over X, which leaves unchanged the marginal distribution for Y. However, if the arc was not completely superfluous, the new $P(Y|x)$ will differ from the old for at least some $x \in X$.

To control for possible back paths like $X \leftarrow W \to Y X$, we use an intervention operator "||" rather than a regular conditioning operator "|". (An intervention operator, often called "do(x)", blocks backwards inference, effectively cutting the links into X.) Let $P(Y||x)$ be the resulting distribution in the original graph, and let $P'(Y||x)$ be the same in the new graph, with $X \to Y X$. The link strength is the expected distance between these two distributions:

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

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 have 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).

Following Ebert-Uphoff (2007), a True measure weights each x by its marginal probability, while a Blind measure assigns equal weight to all x.

Ebert-Uphoff wrote his scripts for the Matlab-based Bayes Net Toolbox (BNT) (Murphy 2007) and Intel's Probabilistic Network Library (BNT's C++ offspring) (Intel 2005). We implemented our variant in Quiddity*Script (IET, 2007). It would be relatively easy to do the same for Netica (Norsys 2008). Like Ebert-Uphoff, we rely on Graphviz (AT&T, 2008) for the actual graph drawing. Figure 2 shows an example.

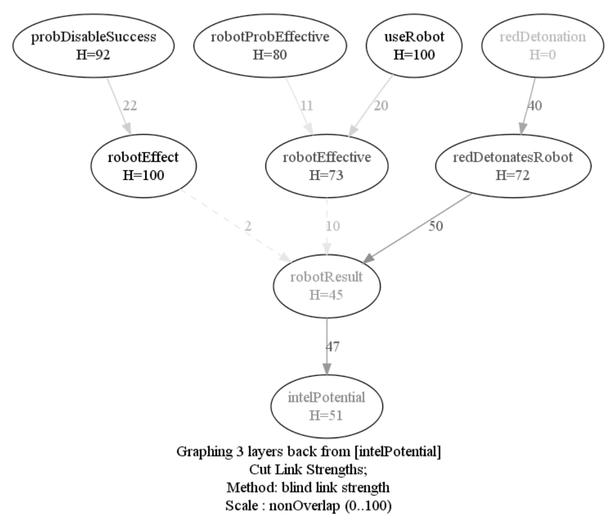

Figure 2: Example Link Strength Graph for Intelligence Potential.

3.2.3 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.

4 EXAMPLE

This section provides an example applying the rapid assessment methodology to a C-IED initiative. We modeled a generic explosive ordnance disposal (EOD) robot such as the one shown in Figure 3. For our purposes, specific characteristics of the robot are unimportant. Rather, we were concerned with broad capabilities. Any EOD robot provides a capability to remotely neutralize an IED, either by disabling it or detonating it. Further, we assume that if the robot is unavailable or unsuccessful, an EOD soldier will neutralize the IED.

Figure 3. Explosive Ordnance Disposal Robot

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

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.

4.1 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 4. MOPs for the EOD Robot.

4.2 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.

4.3 IMPLEMENT THE EXPLANATION AS A PROBABILISTIC MODEL.

Our explanation can be transformed rather directly to a structural model, or graph, as shown in Figure 5. 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)?

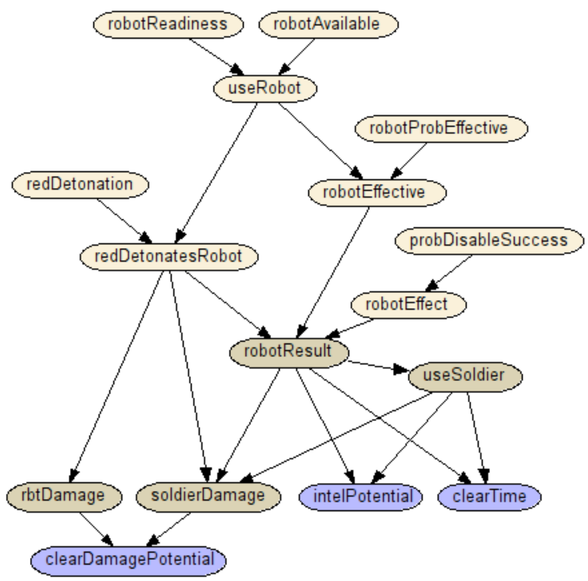

Figure 5. The Robot Explanation Model.

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 by necessity be qualitative assessments of the influence that that variables have on each other.

4.4 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.

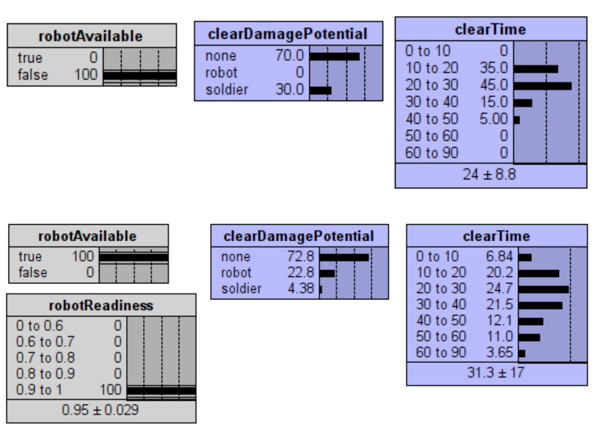

Figure 6. Model Results Showing Impact of Robot Availability on Damage Potential and Clear Time.

If the robot is not available, then a soldier is at risk while disabling the IED (Figure 6). The distribution reflects our assumptions.

If a robot is available and it is working correctly, it can be used to attempt to remotely disable or detonate an IED.

We can see that this lowers the risk, but takes more time. This distribution reflects the consequences of our explanation and assumptions. Although the three-decimal-place estimate of a 6.84% probability of disposal in under 10 minutes is not to be taken seriously, it is believable that the robot increases the time, roughly as shown. It is also believable that the actual time has a wide distribution, itself an average of the distributions for various specific settings of various unobserved ancestor variables (and 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.

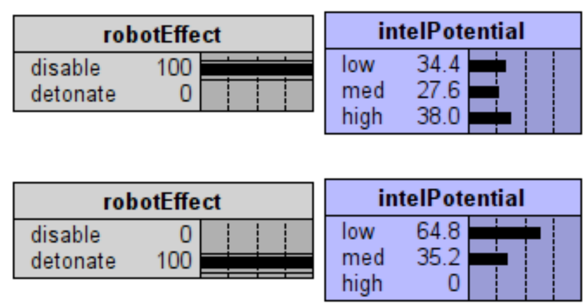

Figure 7. Model Results Showing Impact of Robot on Intelligence Collection.

4.5 DETERMINE THE SENSITIVE PARAMETERS (SPS)

An attractive feature of an executable model is the ease of performing 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`, and excluding uninteresting variables such as deterministic Boolean children of continuous "auxiliary" variables that represent the true parameters of interest.

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.

ClearTime	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 gives a ranking, and we can look at the mutual information values themselves, but those represent average effects. Figure 8 shows how MOPs change as we move a variable through its range. The figure shows that the effect of probability that Red detonates the IED against the robot is quite strong.

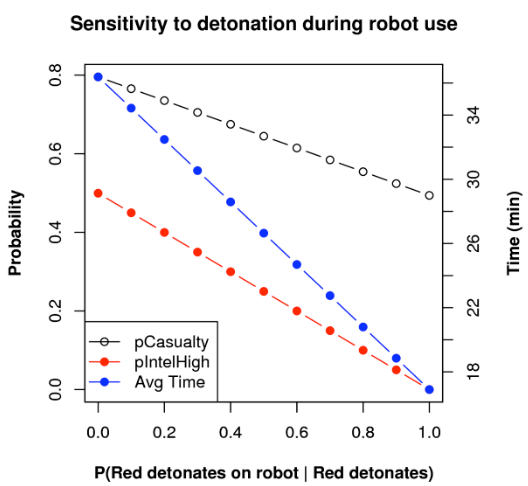

Figure 8. Sensitivity Analysis showing the influence of Red tactics: a command detonation of the IED on the robot.

The size of the effect stems in part from considering the whole range of reliabilities, from 0 to 1. This is equivalent to comparing no robot to a perfectly reliable robot. However, given that caveat, having a robot in this scenario makes a big difference. The most important result is that probability of a casualty drops from 80% to below 60%. However, our chance of getting "High" intelligence drops from 50% to 0, which entirely reflects our scenario and assumptions: as the robot is more reliable, we are more likely to use it. If Red detonates on the robot, that means we lose our chance for gathering intelligence. More dramatic, but far less interesting, we see that 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, which can easily take an hour.

5 CONCLUSION

The rapid initiative evaluation methodology provides a structured approach for assessing initiatives even when there is little formal test data to support evaluation. The modeling approach uses relevant MOEs which provide a consistent framework for evaluation. The graphical structure of the BN supports clear communication to decision makers about the influences and interactions of relevant system and environmental variables. Populating the BN with local probability distributions, even when they are informed only by qualitative expert knowledge, makes the model executable. The executable model supports what-if analysis or alternative scenarios that can be used to asses the likely impact of the initiative, and supports sensitivity analysis that can be used to identify the important system variables to be evaluated during formal testing.

Acknowledgements

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.

References

AMSAA. RECCE II initiative evaluation plan. Technical report, AMSAA, August 2008. (FOUO).

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

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`.

Atul Gawande. The checklist. *The New Yorker*, December 2007. December 10.

Tim Hudson. Knowledge elicitation interview, 18 December 2008. (FOUO)

Inc. IET. 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.

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

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

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.

Archer M. Macy, USN RDML. Counter-IED response in the GWOT: Technology transition for the current war, October 2006. Slides from talk at the 11th Expeditionary Warfare Conference, Panama City, FL, 23 Oct.

Paul E. Meehl. *Clinical vs. Statistical Prediction*. University of Minnesota Press, 1954. Reprinted in 1996.

Paul E. Meehl. Causes and effects of my disturbing little book. *Journal of Personality Assessment*, 50(3):370–375, 1986.

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

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

A. 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.

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

Perry et al. Minimizing the threat from improvised explosive devices in iraq. Technical report, RAND, 2007.

Noah Shachtman. The Baghdad bomb squad. *Wired*, 13(11), November 2005.

Craig Stroudt. RECCE I evaluation data, 2007. (SECRET).

Defense Acquisition University. Key performance parameters. Website https://acc.dau.mil/CommunityBrowser.aspx?id=28953. Visited February, 2009.