Proceedings of US Army Research Laboratories Collaborative Technology Alliances Symposium,
April 29-May 1, 2003, University of Maryland Conference Center, College Park, MD.

FROM SIMULATION TO INSIGHTS: EXPERIMENTS IN THE USE OF A MULTI-CRITERIAL VIEWER TO DEVELOP UNDERSTANDING OF THE COA SPACE

Richard Kaste, Janet O'May, and Eric Heilman
U.S. Army Research Laboratory
Aberdeen Proving Ground, MD

and

B. Chandrasekaran and John Josephson
The Ohio State University
Columbus, OH

ABSTRACT

Building on advances in modeling and simulation, the U.S. Army Research Laboratory (ARL) has developed a capability for simulating detailed Courses of Action (COAs). The Ohio State University (OSU) has developed a multi-criterial decision technology known as the Seeker-Filter-Viewer. In this paper, we report on initial results of a collaborative effort between researchers at ARL and OSU in experimenting with the decision tool for mining ARL combat simulation data to gain battle-planning insights. The capability of simulating detailed COAs opens up possibilities of mining collected data for insights. Decision support systems could assist commanders in examining simulation data for relationships between COA structure and various battle objectives. The synergy of data mining tools, high performance computing, and advanced high-resolution combat simulation has potential to lead battle planners to new insights for imminent combat, translating to improved battlefield assessments and expedient modification of COAs.

INTRODUCTION

Advances in simulation and data mining are proving relevant to providing battlespace decision support. High performance computing, improved modeling techniques, and new decision support techniques have driven these advances. Combat simulations now generate behaviors at increasingly finer scales. Data mining provides a mechanism for uncovering patterns in increasingly larger data sets, such as those generated by modern simulations. The capability of simulating detailed courses of action (COAs) opens up the possibility of mining collected data for insights. Decision support environments could assist commanders in examining simulation data for relationships between COA structure and various battle objectives. The synergy of data mining tools, high performance computing, and high resolution simulation has the potential to lead battle planners to new insights for expedient modification of COAs.

The Battlespace Decision Support Team (BDST) of the U.S. Army Research Laboratory (ARL) and the Laboratory for Artificial Intelligence Research (LAIR) at The Ohio State University (OSU) are exploring the applicability of combat simulation to command and control. Our endeavors have centered on COA evaluation and metrics for the planning/re-planning process of modern combat. Completely automated COA generation is difficult due to the complexities of combat instances and the elusiveness of capturing expertise in a computational framework. Building on advances in modeling and simulation, BDST has developed a capability for simulating detailed COAs. LAIR has developed a multi-criterial decision technology, the Seeker-Filter-Viewer (SFV) [Josephson et al., 1998]. The collaborating researchers have developed an approach based on multiple criteria decision-making to enhance future battle planning.

COAs are executed in a battlespace in which responses are stochastic: behaviors on both sides are subject to a variety of apparently random events. Given a tractable analytical model of the whole system, one might be able to derive outcome expectations and variances as functions of plan parameters. However, such models do not exist for realistic battles. An alternative is to utilize simulations such as One Semi-Automated Forces (OneSAF), as was used in this study. Several types of insights might be then be sought. Here are some examples:

Prepared through collaborative participation in the Advanced Decision Architectures Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0009. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

(1) To explore the effectiveness of a new doctrine, one might create a database covering a variety of instances of engagements, terrains, and COAs. Some engagements would have the new doctrine, and some not. Each would be simulated many times to produce a statistically valid data set. Then one could examine the set to see if the new doctrine makes a difference, and if so, explore conditions under which it does.
(2) For a given terrain and assumptions about the enemy, one might wish to explore a range of COAs. This would require generating a variety of friendly COAs, and for each, assuming a range of enemy COAs. After simulating each combination and examining the data, the best set of COAs might be identified.
(3) For given friendly and enemy COAs, one might wish to know predictability of various outcomes. A specific insight one might seek is which events or subgoals have predictable effects. For example, one might wish to know how sensitive an outcome is to certain platforms being incapacitated early in the battle.

The focus of this paper is on problems of the third type. A challenge in mining simulation data for insights is that the data might not contain information at the level of abstraction of the independent variables of interest. For example, the BDST simulation tracks individual vehicles, and has no record of subgoals (e.g., "more than half of enemy vehicles destroyed"). However, insights needed by the commander are often in terms of these intermediate abstractions. For example, if it turns out control of hill22 is highly correlated with a desirable outcome, the commander might change the COA to reflect this insight and resimulate to confirm. Or, if during the battle, control of hill22 is proving difficult with the assigned resources, the commander might explore how to reallocate.

Since insights desired are in terms of intermediate abstractions, an initial goal of our research was to identify a set of such abstractions and methods of computing them. We are utilizing ARL's Killer/Victim Scoreboard capability, which provides information on direct fire combat (who shot what, at whom, when, where, with what effect) and logistical states of battlefield entities (ammunition, fuel, damage) [Heilman & O'May, 2002]. Abundant computing resources are required to support such an ambitious data collection and analysis effort. Identifying abstractions is itself an interesting research topic, one that requires substantial experimentation over a period of time.

NATURE OF A PROPOSED DECISION AID

SFV can provide two related, but distinct, types of decision assistance in the COA problem. In one mode, for a given COA, SFV can help a commander understand the COA space from simulations covering a range of statistical possibilities. In the second, given a number of alternate COAs, SFV can help the commander visualize trade-offs between alternatives along various pairs of potentially conflicting criteria. The long term vision of the collaborating BDST and LAIR researchers is to integrate these two modalities, but in this paper our initial focus is on the use of the architecture as a data mining aid to generate insights about the COA space for a specific COA.

"Data mining" is analyzing data to identify interesting patterns, especially those suggesting hypotheses about causal relationships. Typically, data mining algorithms look for correlations: high correlation between two variables suggests a direct causal relation or the existence of another variable having a causal relation with both. Sometimes the two variables may display opposite correlations in different regions of the space. In such cases an algorithm not sensitive to local variation might conclude the variables have little correlation, thus missing an important insight. In identifying potential candidates for correlation, some systems use automated clustering algorithms, which, again, may miss the fine structure. On the other hand, humans have powerful visual pattern recognition capabilities. Thus, it is useful to have interactive systems, where the human utilizes a visual interface to spot patterns and form hypotheses, which may then be accepted or rigorously verified via additional tests.

Let us consider some examples of the kinds of insights that might be useful. It might be hypothesized that if Red is heavily damaged early on, then Blue should win. If this is indeed the case, then the simulation data should show high correlation between the degree of early Red damage and Blue success. A more interesting situation is when the correlation exists but is limited to a sub-region. Or, on further analysis, one might observe this correlation holds only provided Blue has sufficient reserves. We can check the sensitivity of performance measures to various parameters: if for each value of the parameter the measure covers its full range, we infer that measure is not especially sensitive to this parameter. Some art on the part of the user is required to determine items of interest. Negative results are also useful: we can observe whether measures expected to be directly or inversely correlated do in fact behave that way.

For running an initial set of experiments, we developed a set of example questions to ask about the simulation results for one COA:
(1) How much of the success is determined by which intermediate objectives? E.g., if a stated objective is taken (missed), then success probability goes up (down)

significantly. How do intermediate objectives correlate with other measures of performance?
(2) How much of success is random?
(3) How much is determined by mobility (e.g., average speed)?
(4) How much by conserving ammunition?
(5) How much by synchronization (e.g., arrival times)?
(6) How much by lethality (e.g., damage inflicted)?
(7) How much by victory in the first exchange?
(8) How much by local concentration of forces? Of fires?
(9) Is it an advantage to be the first to shoot? To hit?
(10) How correlated are various measures of success?

To use the decision aid to answer these (or similar) kinds of questions, we need to know:
(1) Measures of performance (such as friendly and enemy attrition, relative final strengths, ammunition and fuel expenditures, and win/loss status) that can be used to evaluate a COA.
(2) Ways of computing the performance measures from the simulation data.
(3) Methods of abstracting from the simulation events to events at appropriate levels of abstraction (such as "first to shoot," or "victory in the first exchange").

INTERMEDIATE ABSTRACTIONS

For the long haul, abstractions corresponding to tactical sub-goals would be useful; but identifying these requires significant subject matter expertise as well several cycles of experimentation in mining the data. For the first few rounds of experiments, we are considering abstractions that can be computed relatively quickly, either from the raw data or from other abstractions already computed.

Presentation of a full-up set of abstractions collected in the basic OneSAF runs is beyond the scope of this paper, and work on BDST analytical mining of voluminous combat simulation data has been reported elsewhere [e.g., Bodt et al., 2002]. However, here is one mid-level set considered for the current effort:
(1) Number of Red vehicles damaged by Eastern Blue firers (scenario involved Eastern and Western avenues of approach)
(2) Number of Red damaged by Western Blue firers
(3) Number of Eastern Blue vehicles damaged by Red
(4) Number of Western Blue damaged by Red
(5) Percentage of Blue ammunition wasted on ineffective Red (firepower-killed) targets
(6) Percentage of Red ammunition wasted on ineffective Blue (firepower/mobility-killed) targets
(7) Percentage of Blue ammunition effective
(8) Percentage of Red ammunition effective
(9) "Eric score" (BDST system rating position and strength)
(10) Whether Blue won
All abstractions were to be extracted at the end of the conflict. This proposal provides a mixture of types of measures (1-8) for SFV analysis, as well as two measures of goodness (9,10) for "objective" comparison. We would consider multiple hits only if additional damage resulted (e.g., if an already firepower-killed target were catastrophically-killed) and make no upfront judgment about tactical utility (e.g., value of early kills).

For the examples discussed later in this paper, BDST created a data set of 228 runs, with fields including:
(1) Run number
(2) Number of Red damaged by Western Blue at .25, .5, .75, 1 of battle completion (based on the number of hits)
(3) Red damaged by Eastern Blue at each period
(4) Western Blue damaged by Red at each period
(5) Eastern Blue damaged at each period
(6) Percentage of Red effective ammunition (number of damage-causing rounds divided by total rounds used)
(7) Percentage of Blue effective ammunition
(8) Eric score
(9) Win or lose (win declared if 4 of the original 13 Blue vehicles arrived at the objective undamaged)

THE SFV ARCHITECTURE

To support multiple criteria decision-making, LAIR has developed the SFV architecture. Among the objectives for SFV in the Advanced Decision Architectures Collaborative Technology Alliance are to enhance the technology and evaluate its utility for a variety of Army decision-making tasks. This multi-year effort will analyze usefulness of the technology, and extensions of it, for a range of Army tasks, build systems for several, and perform experiments to evaluate performance.

The SFV has three synergistic components:
(1) Seeker: supervises the generation and evaluation of many different alternatives along several different criteria.
(2) Filter: removes alternatives that, given a user's values, are dominated by superior alternatives. (Alternative A dominates alternative B if A is better than or equal to B in every criterion and is strictly better than B in at least one criterion.) The survivors comprise the Pareto subset wherein no element dominates another, and the only way to improve with respect to one criterion is to take a loss with respect to another. The Filter is effective, discarding a high percentage of alternatives, and can accommodate uncertainties in values of the evaluation criteria.
(3) Viewer: enables viewing decision alternatives (or a set

of data) in a variety of displays. The Viewer can be used to perform visual trade-offs of the Pareto elements and select a small subset for further exploration. It can also be used to simply view the data along different criteria, COA specifications, intermediate events, etc., to examine relations. The former mode is useful when narrowing the choice, the latter when the task is to develop insights.

USE OF THE VIEWER

For the current data mining application, the Viewer is the most useful architectural element, especially given that BDST already has tools to simulate the COA. In this section, we describe the Viewer in more detail. As mentioned, the Viewer is a graphical interface to view a set of data along multiple dimensions. The data can be seen along multiple pairs of criteria, along selected individual criteria, as histograms, and in tables sorted in various ways. The displays are cross-linked and interactive, enabling the user to explore selected regions in any of the displays and visually evaluate how the selected alternative fares along the dimensions in other displays.

When the Viewer is used for decision-making, the data correspond to alternatives, and the dimensions are evaluations of the alternatives along various criteria of interest. In this mode, the Viewer provides a visual picture of the trade-offs. The user can narrow the choices naturally, going from one diagram to another as needed, guided by his visual recognition abilities to identify interesting subsets. Selections can be retracted at will, and saved as choice sets for later examination. Choice sets can also be viewed as tables and sorted lists. The computer and the human do what each is best at: the human makes the ultimate value judgments, while the computer provides for acquiring, scoring, tracking, and routine comparisons.

The original design goal was to assist the decision maker in selecting from alternatives. Although the Viewer served this purpose well, LAIR found that designers were using it to obtain insights about the design space, specifically how performance criteria correlated with design specifications and how these correlations changed across the space. Designers treated design candidates, specifications, and performance among criteria as a database that they could mine for insights; and they treated the Viewer as an interface for visually identifying hypotheses. LAIR further explored this use of the Viewer in a study that demonstrated how the Viewer could be used to generate hypotheses about relationships between force variables and simulated battle outcomes [Chandrasekaran et al., 2000].

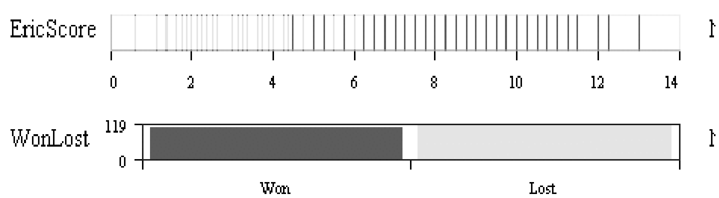
Moving toward discussion of the current study now, we have for each simulation run the values of a number of intermediate abstractions for both forces at various times. Events of interest (e.g., attrition) can be viewed as abstractions of specified types computed from the simulation data. Outcomes of interest refer to evaluation measures for how well a COA performed (e.g., win/loss information). We can then state the goal of data mining as identifying sensitivities of various desirable or undesirable outcomes of interest to various intermediate events, including achievement or non-achievement of intermediate objectives, and to various assumptions.

The analyst may call up several types of display, in each of which all runs are represented. He might select a region in which two abstractions seem correlated. The points there then become a different color, and these candidates appear in the new color in all other displays. He might notice that in all the runs in the selected group, certain other abstractions take high values. He can select two regions in different diagrams and see how the candidates in their intersection appear in other diagrams. He can select a small number of candidates and ask to see their values in table form.

EMERGING RESULTS

We now illustrate how the Viewer is used to form and explore hypotheses about the significance of various abstractions for various outcomes. As mentioned, the battle involved East and West avenues of approach and was divided into four approximately equal periods. These periods were based on the number of direct fire hits delivered, and as such represent the main part of the battle. For each period, and for East and West, a number of abstraction parameters for Blue and Red were computed. For example, *NumRedDamagedW1* is the number of Red platforms damaged at the end of the first period on the West. Some were further abstracted into overall platform losses for each side for West and East combined. Another type of abstraction is ammunition effectiveness, *PerRedEffAmmo* and *PerBlueEffAmmo. EricScore,* an abstraction of positional strength measured towards the end, was hypothesized to be predictive of the final outcome. Finally, of course, we considered the outcome itself: *WonLost* indicates if Blue won. These data were computed for 228 separate simulation runs.

The current Java-based version of the Viewer has a simple interface by which the analyst can specify what abstractions he wishes to see in what type of display. The next figure is an example: EricScore is seen as an analyzer "spectrum," each run represented by a line corresponding to its score; and WonLost is a histogram, showing Blue won about half the time. The figure also indicates an important Viewer capability: the interface allows selection of a subset within a region. Upon selection of the set of Blue wins, instantly all such runs changed to a color chosen by the analyst, here shown as the darker gray. All runs corresponding to an EricScore of 6 and above appear darker, corresponding to a Blue win, and almost all with lower scores correspond to a loss. These observations confirm the hypothesis that EricScore is highly predictive of final outcome.

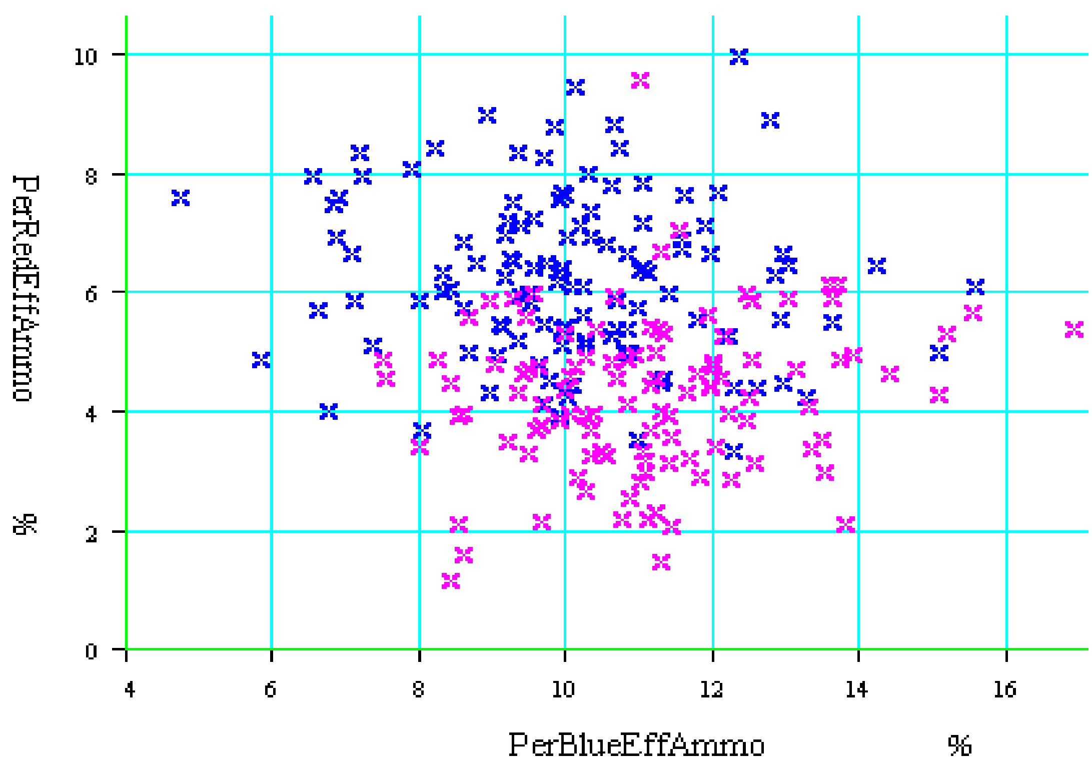


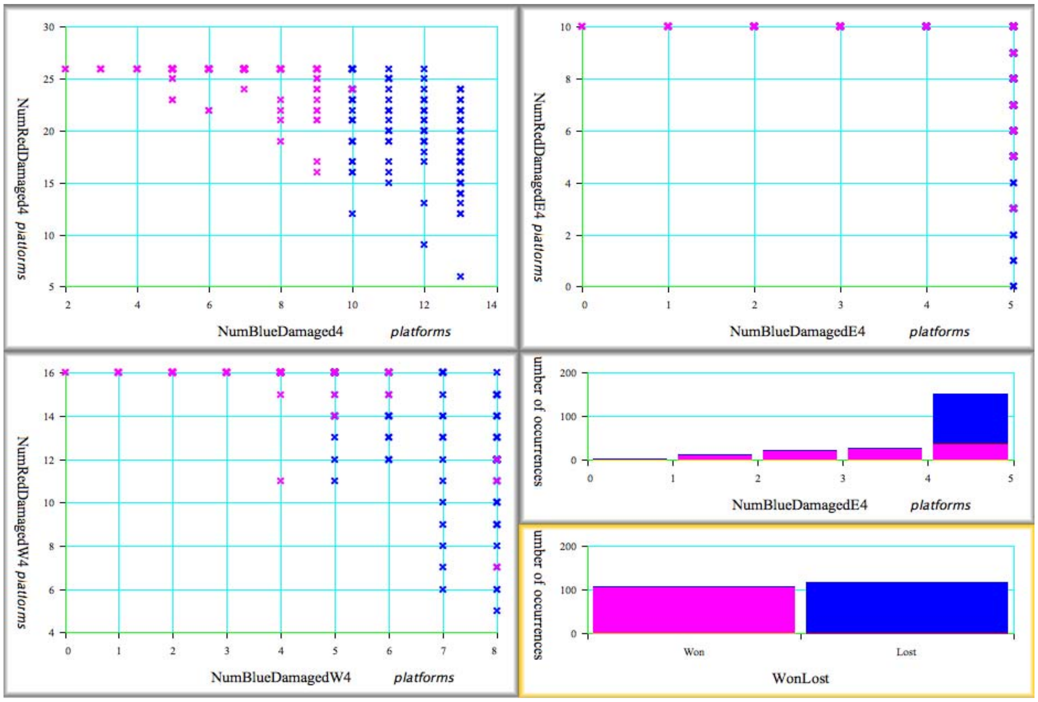
The analyst will obviously have additional hypotheses and puzzles. For example, in the next figure he has called up a display of ammunition effectiveness for Blue and Red, and whether the run corresponds to Blue win or loss is indicated as before. Almost all runs in the bottom (top) half correspond to Blue win (loss), suggesting Blue effectiveness is not as strong a predictor as Red effectiveness. The Viewer visually highlights the rather significant difference in the predictive power of the two similar measures. The analyst can proceed to generate hypotheses and see displays to confirm or refute them.

In the final large figure on the next page, we see several displays on the screen. The analyst has chosen to see:

1) Top left: Total number of damaged Blue and Red platforms at time period 4
2) Top right: Number damaged at period 4 in the East
3) Bottom left: Number damaged at period 4 in the West
4) Bottom right: Won-Lost histograms and number of damaged Blue platforms at period 4

The analyst selects the Blue Wins on the histogram (the Wins are in red and the losses are in blue, a glitch to be corrected in future versions), and in the rest of the displays the runs take on the appropriate colors. The top left display is interesting: While low and high damages to Blue platforms (below and above about 9) correlate almost perfectly with Blue win and loss, respectively, Red damage is less predictive. Runs having relatively high damage to Red in period 4 still have a Blue loss. A similar result holds for correlation between Blue and Red damage in the West in period 4: low damage to Blue correlates with Blue win, without a similar correlation for Red damage. The top right display, of Red versus Blue damage in period 4 in the East, is even more interesting. If Blue keeps his loss to 4 and below, not only is he sure to win, he is also sure to inflict maximum damage on Red. Even when Blue's damage is 5 (the apparent maximum), Blue still has an excellent chance of winning if he can inflict reasonable damage, say 5 or above. This suggests that what happens in the East is highly predictive of final outcome: with reasonable attention paid to damaging Red, even while taking significant damage, Blue can win.





CONCLUSION

This paper has reported on our initial approach to the problem and provided motivation toward a solution by considering some initial abstractions. These experiments are best considered as examples of how the Viewer might be used to generate hypotheses, rather than as insights in themselves. It is, however, significant that even a simple set of intermediate abstractions can provide a rich set of hypotheses.

Although our research is in an early stage, progress is being made in the computation of intermediate abstractions and in exploratory viewing of the data at higher levels of abstraction. Our next experiments will focus on a more difficult set of abstractions, including events during the battle and achievement of intermediate goals. After exploring the solution of questions for one COA, we will expand our work toward mining results with different COAs. BDST is beginning to apply its earlier methodology of simulation data mining to excursions involving urban terrain, and the SFV technique should carry over as well. We intend to explore sensitivity analyses of the impact of modifying various criterion definitions in an effort to help the commander toward more reasonable mission success. We are also interested in development of a generic technology for compositional modeling, often useful for generating and evaluating alternatives. This collaborative research will continue over the next several years.

REFERENCES

Bodt, B., J. Forester, E. Heilman, R. Kaste, and J. O'May, "Pursuit of New Battlefield Metrics through Simulation and Statistical Modeling," 70th Military Operations Research Symposium, 2002

Chandrasekaran, B., J. Josephson, and N. Iyer, "Experiments with a Decision Support Architecture for Exploring Large Spaces of Decision Alternatives in Agent-Based Simulation Models," Report to Marine Corps Combat Development Command, 2000

Heilman, E. and J. O'May, "OneSAF Killer/Victim Scoreboard Capability," ARL TR-2829, 2002

Josephson, J., B. Chandrasekaran, M. Carroll, N. Iyer, B.Wasacz, G. Rizzoni, Q. Li, D. Erb, "An Architecture for Exploring Large Design Spaces," *Proc AAAI-98*

DISCLAIMER