Automatic Causal Explanation Analysis for Combined Arms Training AAR

Randy Jensen, David Yu Chen
Stottler Henke Associates, Inc.
San Mateo, CA
Jensen@StottlerHenke.com, Chen@StottlerHenke.com

Margaret Nolan
Jacobs Sverdrup
Orlando, FL
NolanM@castupgrade.com

ABSTRACT

One of the challenges in developing intelligent, automated after action review (AAR) capabilities for simulation based training systems is the identification of causal explanations for significant events or performance errors detected during an exercise. Automated evaluation methods which use only the raw data from observable simulation events could yield limited training benefits, compared to intelligent evaluations that go one step further by identifying causal linkages with the preceding actions of the training participants - where, when, how, and by whom decisions were made and executed. This concept is being applied in the development of automated AAR tools for the Marine Corps' Combined Arms Command and Control Trainer Upgrade System (CACCTUS) program, which provides a distributed training environment for USMC command, control and coordination in combined arms operations. This paper describes a developmental approach and set of methodologies not only for detecting errors or training points in exercises conducted in this architecture, but also specifically for providing causal explanations. The methodology requires definition of a catalog of the training points to be detected and explained, in order to establish the indexing structure for all rules that can be triggered to determine root causes. This structure is applied in the implementation of a ruleset to capture the logic for linking significant training points to data from the command and execution chain. Since radio and voice communications are an integral part of this chain, a natural language processing capability is required to detect and parse key spoken transmissions and to apply reasoning to establish causal relationships with detected simulation events. Traceable data from distributed C4I tools and direct simulation commands also serve as additional inputs to the explanation rules. By incorporating causal explanations into an automatically generated AAR, the result is meaningful feedback that the training audience can directly apply.

ABOUT THE AUTHORS

Randy Jensen is a project manager at Stottler Henke Associates, Inc. He has developed numerous intelligent tutoring systems for Stottler Henke, as well as authoring tools, simulation controls, and assessment logic routines. He is currently leading projects to develop automated after action review for Marine Corps combined arms training and scenario authoring tools for an Intelligent Structured Trainer integrated with an Army Future Combat Systems training testbed. He holds a B.S. with honors in symbolic systems from Stanford University.

Margaret Nolan is a Senior Training Analyst with Jacobs Sverdrup. She has over 20 years of experience in various phases of the analysis, design, development, and evaluation of training systems. She is currently the Training Lead on the Marine Corp's Command and Control Trainer Upgrade System (CACCTUS) program. She holds an M.S. in Industrial/Organizational Psychology from the University of Central Florida.

David Yu Chen is a lead software engineer at Stottler Henke Associates, Inc., currently working on the after action review component of the Marine Corps Command and Control Trainer Upgrade System (CACCTUS). He holds an M.S. in Computer Science from Stanford University.

Automatic Causal Explanation Analysis for Combined Arms Training AAR

Randy Jensen, David Yu Chen
Stottler Henke Associates, Inc.
San Mateo, CA
Jensen@StottlerHenke.com, Chen@StottlerHenke.com

Margaret Nolan
Jacobs Sverdrup
Orlando, FL
NolanM@castupgrade.com

INTRODUCTION

One of the challenges in developing intelligent, automated after action review (AAR) capabilities for simulation based training systems is the identification of causal explanations for significant events or performance errors detected during an exercise. Automated evaluation methods which use only the raw data from observable simulation events could yield limited training benefits, compared to intelligent evaluations that go one step further by identifying causal linkages with the preceding actions of the training participants - where, when, how, and by whom decisions were made and executed. This concept is being applied in the development of automated AAR tools for the Marine Corps' Combined Arms Command and Control Trainer Upgrade System (CACCTUS) program, which provides a distributed training environment for USMC command, control and coordination in combined arms operations. This paper describes a developmental approach and set of methodologies not only for detecting errors or training points in exercises conducted in this architecture, but also specifically for providing causal explanations. The methodology requires definition of a catalog of the training points to be detected and explained, in order to establish the indexing structure for all rules that can be triggered to determine root causes. This structure is applied in the implementation of a ruleset to capture the logic for linking significant training points to data from the command and execution chain. Since radio and voice communications are an integral part of this chain, a natural language processing capability is required to detect and parse key spoken transmissions and to apply reasoning to establish causal relationships with detected simulation events. Traceable data from distributed C4I tools and direct simulation commands also serve as additional inputs to the explanation rules. By incorporating causal explanations into an automatically generated AAR, the result is meaningful feedback that the training audience can directly apply.

TRAINING OBJECTIVES

There is a concept of "threads" in team/staff training, defined as a sequence of events – decisions, actions, reactions, etc. – that lead to a defining moment or an outcome. In many cases, the precipitating events have occurred long before and are seemingly unconnected. Or, in some instances, the critical point was the *lack of* an event or the performance of some step, decision, or an outcome. It may not be immediately apparent to instructors or to the exercising force exactly where and when a breakdown might have occurred, especially in unit training where many actions are performed simultaneously by different roles.

The capture of the perceived operational picture – what the humans-in-the-loop thought happened and/or is happening – and their concomitant communications (voice or data) can go a long way in providing valuable input for an AAR. The capability to capture the linkages between and among events and then generate a causal explanation for an event adds an intelligence component to the AAR which goes beyond outcome based summaries.

Causal analysis in this application differs from other approaches which aim to automatically generate explanations for the actions of semi-automated forces (SAF) based on the intrinsic behaviors of the SAF themselves. (Lane et al, 2005) describe an explainable artificial intelligence technology for SAF explainability in training. Similarly, the goal with this application is not necessarily to statistically prove causal linkages for a particular outcome via multiple regression or correlational analyses, per se, as is the application of causal analysis in other domains; but rather, to provide insight and emphasis on decision-making performance within the context of the larger performance of a task or mission. The generated causal explanations, then, can be an important asset to both the instructors who prepare AAR, and also the training audience who are subsequently briefed and further probed about their role in specific decisions or procedural actions that led to an error or a specific, unwanted outcome.

Three categories of errors are distinguished for analysis purposes. *Procedural* errors are skill-based and can typically be clearly defined. *Cognitive* errors are knowledge-based, and pertain to individuals' understanding of general concepts, as well as their awareness and modeling of the situation during execution. The *Cognitive* errors could result from the application of "bad" or incomplete knowledge, or the misapplication of "good" knowledge. *Unintended* errors are simply that – mistakes traceable to human error, despite the skills and knowledge of the individual rather than resulting from them. Given the inherent complexity in combined arms scenario training, the taxonomy of dependencies, influencing factors, and culture of team performance dimensions present a range of challenges in defining and detecting causality in these forms.

CHALLENGES IN CAUSAL ANALYSIS

Endeavoring to apply causal analysis as an intelligent component in Marine Corps combined arms training AAR is challenging in many ways. While it's prudent to start with rules, based on doctrine, for *pro forma* reasoning, the methodology must account for individual differences and a degree of allowable variability. Varying sets of unit-specific standard operating procedures (SOPs), and situation-specific rules of engagement, as well as instinctual, contextual decision-making are justifiable and realistically prevalent in real world battlefields. The Marine Corps values flexibility, authority, and responsibility, whereby the decision-making is carried out at responsibility levels from Lance Corporal all the way to the General, and the "most effective decision might be the least predictable one."

Because varying SOPs are acceptable in the training setting, one of the challenges in automated analysis of Exercising Force (EXFOR) actions, decisions, and outcomes is the task of constructing a set of rules for tracing causes from a varying set of inputs. EXFOR performance can be detected and identified, and then used in a tool to further investigate performance in the conduct of a deductive AAR. Nevertheless, in addition to defining an initial ruleset, subsequent challenges are to determine the relevant data to be collected, data sources, and how data should be collected. Questions to be posed and/or factors to be considered may include:

- What is being evaluated? Performance and outcome are multi-attributed and multivariable. (e.g., empirical outcomes, cognitive processes, behavioral procedures)

- Who is responsible for the decisions that directly affect outcomes? (e.g., authoritative source)

- What are the available sources for data/input? (Observations, anecdotes, judgments, simulation traffic, voice communications, C4I messages and representations, etc.)

- Data can be misleading if assumptions are applied uniformly without contextual clues. What are the relative and weighted values of the data?

- What is the temporal integration of collected data? (e.g., continuous, intermittent, irregular, defined epochs, event-based)

- For input data coming from human coordination, such as spoken radio communications, how reliable is the existing technology for conclusively processing the input source? When confidence levels are less than 100%, how does this impact the analysis logic? How can contextual information be used to augment the analytical logic?

- How will the data be aggregated or combined for the results? How to construct event explanations which follow an intuitive path matching the decision-making and execution paths.

Measures are either complete or incomplete. The AAR component is predicated on a set of defined semantics of causality built to establish relationships between a chain of facts and events (decisions-outcomes), along a timeline. Ideally the results can be used not only to provide feedback to the training audience in AAR, but also to stimulate further after action dialog with the humans-in-the-loop to review why a specific decision was made or an action taken or not taken.

OVERVIEW OF DOMAIN AND APPLICATION

CACCTUS will support a live, virtual, and constructive training environment that facilitates the interaction between and among levels and echelons of command and control agencies normally found in the tactical environment in the conduct of real-time combined arms fire support operations. Training exercise scenarios allow for the practice of communication and coordination, tactics, techniques, and procedures in support of specific training goals and objectives of the EXFOR.

Simulation based training exercises take place in Combined Arms Staff Trainer (CAST) facilities, and may involve 100 or more participants at various stations in the facility, carrying out their respective operational responsibilities. At the conclusion of an exercise, an AAR debrief is prepared either by instructors who monitored the execution throughout, or by commanders within the EXFOR unit taking part in the training. Traditionally the CAST facilities have made use of sand tables, but have more recently begun adopting a computer based simulation model, with three dimensional visualization tools to support the task areas for forward observers and fire support teams. As a result, the opportunity for automated analysis of simulation events and states, as well as analysis of EXFOR coordination activities, has become a reality due to the availability of digital data. The purpose of automated intelligent analysis capabilities is to inform and facilitate the AAR process for instructors or commanders overseeing the training. The upgraded CAST facility will support a full range of simulated combined arms operations, with the same decision-making and execution chain as in the operational setting.

Safely coordinating indirect fires in consonance with maneuver is an integral focus in CACCTUS combined arms training objectives. An example of a significant event, or friction point, that may occur during an exercise is described below and represented in the following illustration.

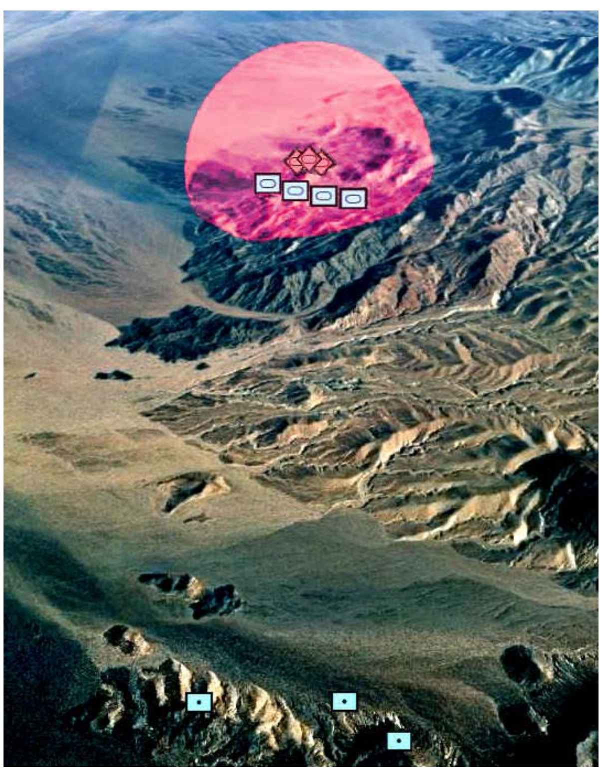

Figure 1. Artillery MSD conflict

In Figure 1, friendly artillery units in the foreground are engaging an enemy target in the distance with an active fire mission. In this example, under the applicable Rules of Engagement (ROEs) for the exercise, the minimum safe distance (MSD) for artillery fire is 1000m, which extends as a radius in all directions from the detonation or target point. This MSD is depicted as a hemisphere in Figure 1, and in this example it encompasses the positions of a friendly tank platoon. The fact that these friendly tanks are inside the artillery MSD while the fire mission is active is considered a conflict, as the situation presents a risk of fratricide which goes beyond the safety parameters established for this training exercise.

Under normal procedures, the execution of a fire mission as in this example requires the coordination of several distinct roles. In the case of an artillery fire mission, there is typically a requesting observer who sends a call for fire, an approval authority, and a fire direction controller within the artillery battery. Within a CAST facility, different individuals at different stations may perform the responsibilities of these different roles, with each having the potential to contribute errors. As a result, the objective of a causal explanation analysis capability within this domain is to analyze the actions of each EXFOR participant which may have precipitated the detected event, and provide this analysis in the after action debrief.

EXFOR participants may coordinate with each other in the CAST facility using radio communications, messaging on shared software applications such as C4I tools, and verbal or handwritten communications for those in proximity to each other. Some of these forms of coordination are not available to the system for analysis, but for those that are, the forms of coordination can be evaluated to determine if they were contributing causal factors in detected events.

For the purposes of illustration, the artillery MSD conflict serves as a suitable example for describing the range of different categories of coordination errors which could be identified in a causal explanation analysis.

Procedural example: the artillery fire direction controller acts without clearance from the proper authority, or with clearance from an improper authority, and executes a fire mission that results in a conflict.

Cognitive example: the approval authority has a poor understanding of the battlefield situation or the applicable safety parameters, and gives clearance for a fire mission that results in a conflict.

Unintended example: an EXFOR operator erroneously enters an inappropriate target grid location in software, resulting in an unintended conflict situation.

Each of these examples of causal explanations can be automatically derived when the supporting data is available and suitable for analysis. In practice, more than one causal explanation may apply, as human errors are often compounded and propagated in team-based operations.

TECHNICAL APPROACH

Within the CACCTUS architecture, an application called AAIRS (After Action Intelligent Review System) is responsible for performing the causal explanation analysis for AAR. There are four main functional areas within AAIRS which contribute to the causal explanation methodology.

1. Event Detection. Detection of significant training events from execution, for which causal explanation analysis will be performed.
2. Explanation Ruleset. Representation of rules linking events to possible explanations.
3. Rule-driven Data Analysis. Automated methods for investigating execution data to support or eliminate possible explanations in specific cases.
4. Collection and Presentation. Automated tools to collect and present explanations in the context of an AAR debrief.

Event Detection

For the purposes of Event Detection in AAIRS, knowledge engineering work with the USMC combined arms operations domain resulted in a catalog of training points to be identified in after action review in accordance with what is deemed significant from a training requirements perspective. These training points correspond to detectable events that are generally simulation states or situations which can be identified and which reflect an undesirable outcome. The example presented earlier, with an artillery fire mission constituting a battlespace geometry conflict during execution, would be a typical example of a significant event.

Event Detection functionality is a combination of data collection and analysis with respect to pre-defined criteria. The data collection requires both periodic (entity states and position updates) and non-periodic (mission parameters and trigger events) information from the simulation, as well as similar usage data from any other devices used by the EXFOR such as C4I tools which further reflect the intentions of the operators. An additional category of data collection is the exercise specific data such as the applicable safety parameters in the artillery example above. Event focused analysis is required in order to process the raw data from the simulation and other sources, and accurately detect when significant events occur. In the case of an artillery MSD violation, the analysis involves a variety of routines for detecting geometrical intersections, typically for a position or route line segment with a three dimensional battlespace geometry shape such as a hemisphere around a target or detonation point.

Explanation Ruleset

The Explanation Ruleset encompasses all the domain specific logic for tracing causal roots for errors and events. To generalize, the causal explanation rules represent the decision points associated with an operational event, and the range of possible system input points which can reveal the EXFOR decision-making as it occurred. The complete set of significant events to be detected as training points are hierarchically organized into categories which share common sets of possible explanations. This is analogous to a principle hierarchy from the area of Intelligent Tutoring Systems. For each category of detected event, the set of possible explanations and the methods for investigating them is represented in this hierarchy. Explanations fall into groups corresponding to the three categories of errors described above: *Procedural, Cognitive,* and *Unintended*. In the example *Procedural* explanations for the artillery MSD conflict above, the corresponding rules define the following:

- Proper approval authority
- Person responsible for execution
- Proper approval procedures, along with a mechanism for reviewing procedures performed for correctness
- Possible improper procedures to detect

In this example, the proper approval authority is the liaison officer with the Fire Support Coordination Center. The person responsible for execution is the

Fire Direction Officer. The proper approval procedures are a sequence of radio communications between the authority and executor on a specified net, with certain expected keywords to be detected in the spoken transmissions. The possible improper procedures include situations like the execution person listening to the wrong net, or an improper authority giving approval on the right net. The rules in particular define which radio nets are appropriate to review, and how to determine which transmissions may specifically apply for the analysis with respect to a specific causal explanation. In other cases these rules also define other data to process, beyond the radio communications data, which may support or eliminate a given explanation.

Rule-driven Data Analysis

In accordance with the Explanation Ruleset, Data Analysis involves the review of data from execution to prune off any possible causal explanations that can be disproven. In addition to all available data for the detected events themselves, the primary available data sources for causal analysis are spoken radio communications and digital messages sent on C4I tools. Speech recognition on communications, consisting of keyword spotting and contextual analysis, is a major component of the analysis. This requires not only the capability to parse specific keywords out of individual transmissions, but also to apply logic about which role is speaking, on which radio net, with which listeners. For example, if an artillery battery fire direction coordinator calls in to ask, "Is that mission approved?" it would be a mistake to parse the keyword "approved" from this transmission and determine that the mission was approved. Not only does the transmission itself not constitute an approval, it is spoken by a person who is not an approval authority. Similarly, the language processing portion of the analysis must be able to consider higher level aspects of communications procedures that go beyond individual transmissions, such as copied back messages and messages intended for multiple listeners on a given radio net.

Collection and Presentation

The Collection and Presentation function collects the results from Event Detection and the Rules-driven Data Analysis and provides tools for assembling these materials in an after action debrief.

CAPABILITIES DEVELOPED

Figure 2 below shows how the current causal explanation analysis capabilities are structured within the overall CACCTUS architecture, in the existing AAIRS implementation.

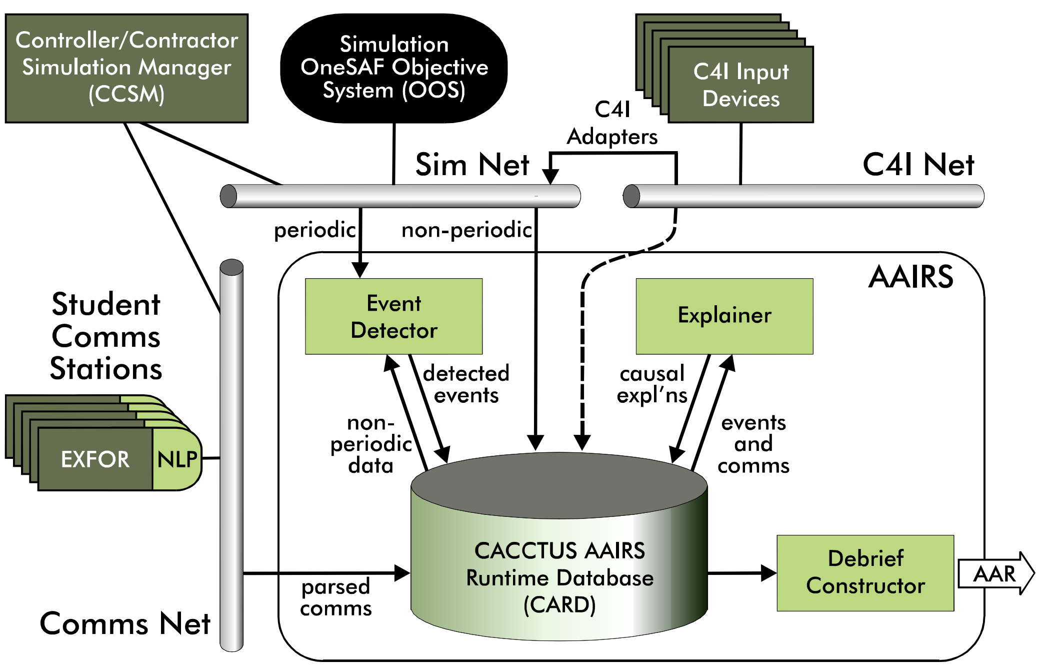

Figure 2. AAIRS Architecture

The Simulation Net, Communications Net, and C4I Net are the three data networks in the CACCTUS architecture which contain the flow of data that is consumed in the causal explanation analysis. The Simulation Net contains all periodic and non-periodic data from the Simulation (OOS), as well as from the control interfaces used by EXFOR during an exercise. The primary function of the Comms Net is to provide radio connectivity to EXFOR for their interaction and coordination during the exercise. Every participant has a communications Student Station with a variety of available voice nets. A secondary function on each Student Station, integrated for the causal explanation purposes, is a developmental Natural Language Processing (NLP) application which parses each spoken transmission for words and phrases relevant to causal analyses. The NLP on the Student Stations is based on a tool called GISTER, initially developed at the Naval Air Systems Command Training Systems Division (NAVAIR TSD). The C4I Net contains the flow of messages on all networked C4I Input Devices, which send command and control inputs to OOS via a set of C4I Adapters.

One of the cornerstones of the AAIRS architecture is the CACCTUS AAIRS Runtime Database (CARD). The CARD serves as a central, persistent repository for all data collected during an exercise. Periodic data from OOS (regular entity position and state updates) is not collected in the CARD; rather, it is evaluated by the Event Detector as the exercise proceeds, but not retained in the CARD. Non-periodic data, such as the specific orders given by EXFOR and operators in the Contractor / Controller Simulation Manager (CCSM), and mission parameter information from OOS, is collected in the CARD. Messages from C4I Input Devices provide an additional planned source of non-periodic data to be integrated with the CARD. The parsed communications from the Student Stations on the Comms Net are also stored in the CARD for retrieval in subsequent causal explanation analyses.

The concept of an abstracted runtime database in the form of the CARD is incorporated into the system design specifically in order to facilitate the development cycle and also make the system modular and scalable. By having a common data structure specification defined for the CARD to support the required analysis functions for AAIRS, the development of the analysis can proceed in parallel with the implementation of the data collection methods in all the components that will ultimately be responsible for writing to the CARD. Furthermore, this contributes to the overall modularity of the system design, as the analysis logic is independent of the specific data collection code in any other component, and in fact, other components may be substituted in the architecture with the only requirement that their collection methods comply with the data structures in the CARD.

The Event Detector functions with a combination of the periodic data stream direct from OOS via the Simulation Net, and the non-periodic data collected in the CARD. In the artillery MSD example above, the conflict event would be detected by the following steps. Once the artillery fire mission begins, the CARD receives a non-periodic mission start event from OOS, and pushes a notification of this event out to the Event Detector. The Event Detector then consults the CARD for the mission parameters – the start time, firing unit, origination location, target location, intended duration, ordnance or projectile, and trajectory information. These parameters are already present in the CARD, as they are collected when first issued in either the CCSM or an OOS fire support editor. Once the Event Detector is notified that the artillery mission is active, it analyzes periodic data from OOS regarding entity positions, to determine if there is a conflict. The tank platoon positioned inside the artillery MSD area is thus noted, and the Event Detector generates an event for this conflict, which is stored in the CARD. This in turn triggers the Explainer to begin its analysis task to determine the causal roots for this conflict.

The Explainer component contains the Ruleset for causal explanations, and the analysis mechanisms for deriving conclusions. Upon notification that an event has been detected and written to the CARD, the Explainer immediately consults the Ruleset for the applicable explanations to test, and triggers the analysis methods associated with each rule. Any relevant speech recognition parses from preceding communications are delivered to the CARD as they occur, so in the analysis step these are already present, and can be reviewed for the causal roots they reveal. In most of the Explainer rules, the spoken transmissions provide some of the most direct evidence of EXFOR decision-making and intentions, so the task is to attempt to extract this from what can be processed automatically. The language processing function is based on a tool called GISTER, which was developed at NAVAIR TSD. The GISTER concept focuses on a speech recognition engine and a statistical language model constructed from transcripts of recorded audio communications from USMC combined arms operations training exercises. The objective with GISTER was not to attempt complete speech recognition on all radio transmissions, but rather to identify basic meaning by identifying the presence of specific keywords and/or events in individual transmissions and use this in reasoning about the intended content of the communications. In the current implementation, the first pass recognition focuses on keywords that reflect specific intentions of the EXFOR humans-in-the-loop appearing in critical coordination communications, including words like “approved”, “denied”, “cleared”, “cleared hot”, “abort”, and “check firing.” This is supplemented with contextual analysis for the roles and radio nets on which the transmissions occur, as well as evidence about dialogs taking place between roles, in the form of sequences of transmissions with specific expected keywords. All results from the Explainer are written back to the CARD and associated with the detected events populated there.

The Debrief Constructor component provides a toolset for instructors to filter and collect training points from the detected events from the exercise, along with the causal explanations generated for each event, and assemble them as desired into an AAR debrief. For the CACCTUS application, instructors have limited time for the debrief construction task, so it is critical for this tool to make the job simple and fast, including a one-click capability to automatically assemble the entire debrief based on existing default settings for the form and content of the debrief. Under the current implementation, the causal analysis outputs provided to instructors within the Debrief Constructor also include indexing information for retrieving specific relevant recorded audio from the radio communications during the exercise (by radio net, time, and speaker).

An additional function of detecting predicted conflicts is also under development with the existing AAIRS architecture, making combined use of the Event Detector and Explainer capabilities. For example, if an artillery fire mission is approved by the proper authority, but in a circumstance where it would result in a conflict situation if executed, then this can be detected in advance. Using the same logical rules that apply for looking back to determine if a proper approval was given for a fire mission after it was executed and after it caused a conflict condition, and the same event detection routine to detect the conflict condition in the battlespace geometry, the system can alert instructors to predicted conflicts before they happen. This helps instructors more closely follow the decision-making of EXFOR as the exercise unfolds. When a predicted conflict is not recognized and remedied by the EXFOR before it is allowed to take place, it will be detected again as an executed conflict event. The AAR does not debrief the EXFOR on predicted conflicts when they are remedied. But in cases where a predicted conflict becomes an actual conflict, the initial detection provides further context for the causal explanation analysis of the actual conflict event when it is later incorporated into the AAR.

LESSONS LEARNED AND FUTURE WORK

In the current implementation, the causal explanations handled by the Explainer are primarily in the *Procedural* category. Under the ongoing development plan, the scope of Explainer analyses is being expanded in conjunction with the evolving implementation of data collection routines for additional sources and EXFOR inputs during execution. For example, the evidence that supports or eliminates causal explanations in the *Cognitive* category frequently depends on the system’s access to data collected from C4I tools in use by EXFOR, where some form of a representation of the “perceived truth” is maintained. At a system architecture level, this data is not collected and provided to the CARD yet, and no reasoning is performed with this data as an input, either artificially provided or from the actual source. Due to the abstracted design of the CARD, the development of the analysis logic can proceed to include C4I messaging even before the methods to collect this data are implemented, if the CARD is artificially populated with test data. As digital data from operational C4I tools becomes available to the analysis logic, the result will be a significant enhancement to the causal explanation conclusions and the quality of feedback that can be collected for AAR. With access to the operating picture representations maintained by key EXFOR decision-making authorities, the Explainer will be able to compare this directly with the Simulation side “ground truth” and determine specifically where cognitive mismatches exist.

Similarly with *Unintended* errors, causal explanation logic requires redundant sources of EXFOR input data that can be reliably compared with each other, in order to determine when errors such as a transposed digit or a typo have occurred. For example, consider the case when an operator mistypes a target grid location number (e.g., “123546” instead of “123456”) in the entry of a fire mission in the CCSM. When the parameters for this fire mission are collected and provided to the CARD, this is the only source of information regarding the “intended” grid location. As a result, this cannot be isolated as an error. But the spoken radio communications by the humans-in-the-loop on the Comms Net have the potential to provide the redundant information source for such cases. If a call for fire references a target grid number, and the fire mission is approved with the same grid number, and finally the entered execution instructions have a different grid number, then an *Unintended* entry error is likely to be the explanation. However, this requires highly reliable machine-based natural language processing, possibly enhanced with additional contextual information. If one or more digits in a spoken radio transmission containing a number reference is parsed inaccurately, then the parse itself serves less as a conclusive redundant source, and more as an informational tool for suggesting a possible explanation which can be validated by a human instructor reviewing recorded audio.

In this respect, the area of causal explanation analysis is not limited to binary positive or negative results with respect to individual candidate explanations. In fact, for the Marine Corps CACCTUS training application, the purpose is to serve as a tool for the human instructors in their process of constructing the AAR for EXFOR. The function of the Explainer for this application can be considered one of eliminating possible causal explanations, and providing information on the confidence level that can be conclusively reached for those explanations that remain, following an automated analysis. Under the operational concept in the CAST training, instructors may personally retrieve and review specific recorded communications highlighted by the automated analysis, to determine firsthand if a particular proposed explanation is supported. If so, instructors may elect to include the playback of these communicated transmissions in the after action debrief for EXFOR. In this use case, the automated causal explanation analysis adds considerable value to the AAR process, by serving a dual function. Its primary function is to produce conclusive results when possible and when supported by available data. But its secondary function is to provide information about any remaining possible causal explanations that could not be eliminated, and a toolset for directly referring back to available data from execution for confirmation and presentation.

There are several planned areas of work to continue the existing development path.

- Cascading explanations. An area of planned work involves the post-processing of related events and explanations. Due to the nature of combined arms operations, errors often cluster together around the same set of root causes. For example, a situational awareness deficiency on the part of an approval authority can impact several different events, and can also result in situations where an initial conflict becomes compounded when it goes unremedied. These kinds of conditions will be addressed by post-processing rules in the Explainer, which will help to consolidate related events and explanations for the purposes of AAR.

- Language processing. Planned work includes both experimentation with methods to improve performance of speech recognition for existing processing, and also expansion of the forms of analysis to include additional keywords and more areas of contextual evidence beyond the content of individual transmissions.

- C4I tool inputs. The inclusion of data from C4I tool messages and commands in the causal analysis will provide a significant additional source for reasoning about the decision-making of EXFOR. This is reflected in the message traffic on operational C4I tools, and also in the perceived truth representations maintained by EXFOR in these tools.

- Automatic audio retrieval for debrief. Planned work will include a step to further streamline the debrief construction and presentation steps by incorporating an automatic mechanism for retrieving and playing back recorded radio communications from within the Debrief Constructor toolset.

In addition to these specific areas, planned work also includes the expansion of the analytical functions of the causal explanation utilities with additional task areas and responsibilities within the range of training scenarios that can be executed at the CAST facilities; for example, to include logistics operations and urban operations.

REFERENCES

Connelly, E.M., “A Theory of Human Performance Assessment,” *Proceedings of the Human Factors Society, 31st Annual Meeting*, 1987.

Lane, H. C., M. Core, M. van Lent, S. Solomon, & D. Gomboc (2005). Explainable artificial intelligence for training and tutoring. *Proceedings of the 12th International Conference on Artificial Intelligence (AIED05)*, Amsterdam, The Netherlands.

Human Engineering Design Criteria for Military Systems, Equipment, and Facilities. MIL-STD-1472D, U.S. Department of Defense.