AUTOMATING CONVOY TRAINING ASSESSMENT TO IMPROVE SOLDIER PERFORMANCE

Noelle LaVoie
Parallel Consulting

Peter Foltz, Mark Rosenstein, Rob Oberbreckling
Pearson Knowledge Technologies

Ralph Chatham
ARPA Consulting

Joe Psotka
U.S. Army Research Institute

Monitoring teams of decision-makers in complex military environments requires effective tracking of individual Soldier and team performance. For performance measures to be meaningful in training applications, they must be both accurate and timely. Current measures of team performance often use global performance measures, such as mission completion, which do not provide sufficient detail for supporting Soldier performance and are temporally delayed. An untapped source of more timely and diagnostic information lies in ongoing communications among Soldiers operating as a team,. Communications reflect cognitive and task states, knowledge, situation awareness, vigilance, and patterns of information flow. With appropriate analyses, the communication data can be tied back to both the team's and each individual's performance, abilities and knowledge.

We developed an automated toolset that analyzed communication,, modeled performance and provided team performance metrics and indications of critical incidents during training. Based on the toolset, applications were developed that apply the metrics and models to support After Action Reviews (AARs). This toolset was developed at the National Training Center for use with STX lane convoy training as well as with the Mission Support Training Facility at Fort Lewis for use with DARWARS Ambush! simulated convoy training.

Providing feedback on team performance requires an effective set of performance metrics that can be associated with communication streams. In both the Ambush! and NTC convoy training contexts, evaluation occurred as part of the AAR process, so it was important that the performance measures were drawn from the same task contexts, and developed in conjunction with SMEs with extensive experience working with convoy training. We worked closely with SMEs to develop five scales that captured the important dimensions of performance based on a mission essential task list (METL), including command and control, situational understanding, use of standard operating procedures (SOPs), and battle drills. For each rating scale training events were scored by SMEs on a scale that ranged from untrained to practiced to trained. Seven SMEs rated the audio collected from Fort Lewis and NTC on these scales. SMEs were also asked to distinguish between critical events, defined as events that change the scope of battle, the commander's plan or disrupt the operational tempo, and other training events in the communication. Interrater reliabilities ranged from .38 to .66 using Intra Class Correlations for

single items, and exact agreements varied between 24% and 50% suggesting excellent reliability more than sufficient for performance modeling.

The toolset was built to automatically take in audio communication, convert it with automated speech recognition to text and then analyze the text to generate predictive performance models. The models were built and tested using a machine-learning approach which automatically learns to associate SME ratings of performance to critical indicators in the communication stream. The goals of this modeling were two-fold: assess team performance during convoy training and identify critical events during training all based on segments of audio communication. Team performance models were based on variables drawn from the text of the communications, such as semantic content, as well as variables drawn directly from the audio features of the communication, such as pitch and power used to predict the presence of stress in speech. The model's predictions were correlated with the SME ratings between .36 and .43. Critical event modeling was conducted using a spectrum method utilizing discrete time windows where the size of the window, and step size between windows, were optimized to predict critical events from the communication data. This approach was combined using a support vector machine to classify the data into categories of probability that a given time window includes a critical event. Using this approach, over 80% of the critical events were detected, with an acceptably low false alarm rate. The effectiveness of the performance models and critical event detection capabilities mean that team performance during convoy operations can be automatically assessed with an acceptable level of accuracy to support training, and potentially operational missions.

The predictive models of team performance for convoy training have several potential applications, including supporting After Action Reviews (AARs), effectively increasing the amount of training that a single Observer/Controller (O/C) can oversee. The AAR support tool's interface makes it easy to spot performance weaknesses at a glance and then drill down to understand these weaknesses by listening to sample radio communications. The tool also enables commanders and O/Cs to create a custom AAR by selecting events of interest and associated radio communication and adding their own comments.

The general approach used here further translates well to other military applications requiring monitoring and assessment of teams. It allows embedded near-real-time analysis and modeling of real (complex) communication data for networked teams. The combined toolset automatically models objective and subjective metrics of team performance and can generate its predictions within seconds. Because the models are automatically derived, the approach does not require large up front task analyses and instead capitalizes off of learning to model team performance in the same manner as SMEs.

Automating Convoy Training Assessment to Improve Soldier Performance

Noelle LaVoie, Parallel Consulting
Peter Foltz, Mark Rosenstein, Pearson Knowledge Technologies
Rob Oberbreckling, Perceptive Research
Ralph Chatham, ARPA Consulting
Joseph Psotka, U.S. Army Research Institute

Abstract

Monitoring teams of decision-makers in complex military environments requires effective tracking of individual Soldier and team performance. An untapped source of timely and diagnostic performance information lies in ongoing communications among Soldiers operating as a team. The DARCAAT program developed and tested a toolset for automating team assessment and near real-time alarms. The toolset uses Automated Speech Recognition and Statistical Natural Language-based techniques for embedding automatic, continuous, and cumulative analysis of team communication in training and operational environments. This toolset was developed at the National Training Center for use with STX lane convoy training as well as with the Mission Support Training Facility at Fort Lewis for use with DARWARS Ambush! simulated convoy training. Based on the toolset, applications were developed that apply the metrics and models to support After Action Reviews (AARs) and real-time alarms.

Background

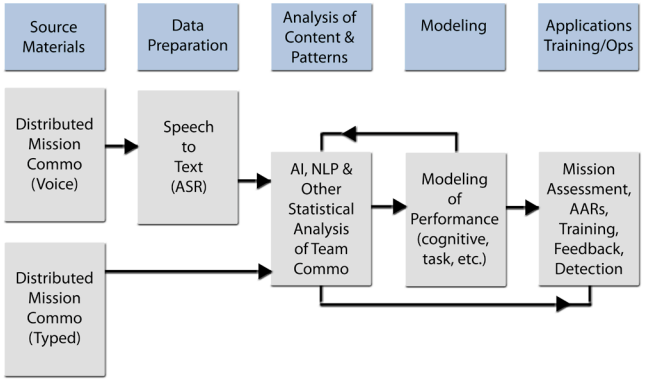
Performance measures need to be timely and acurate.
Limitations of current methods:

- Labor intensive: Review of data and communications can take months or years
- Lack of specificity: Global measures lack detail of what went wrong
- Domain specific: Models often lack generality

The Solution
Automated Communication Analysis

- Commo is a rich source of performance indicators, including workload, C2, and SA
- Commo is critical for expert and trainers to identify strengths & weaknesses
- New technology allows automated analysis and provides instant feedback

Communication Analysis Pipeline



Data Collection: Convoy Training

Collected communication from over 300 missions.

National Training Center STX Lanes

- 6-8 hour training sessions
- Ambush, IED, complex ambush, VBIED

Fort Lewis DARWARS Ambush!

- 20-60 minute missions
- Events similar to NTC STX lanes

Metric Development

Goal: Reliable, interpretable performance metrics. We worked closely with SMEs and O/Cs to identify performance metrics based on the Army's Mission Essential Task List (METL). Models were trained using ratings from seven SMEs.

Metrics

- Command & control
- Situation understanding
- Adherence to Standard Operating Procedures
- Battle drills
- Overall team performance

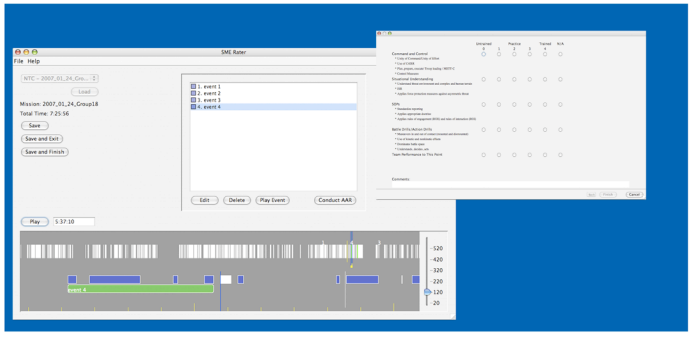
Scale

- Untrained
- Practiced
- Trained

SMEs also rated critical events and provided AARs with sustains & improves for each mission.

SME Reliability

Intraclass correlations: r=.38 to .66
Exact agreement: 24% to 50%
Adjacent agreement: 74% to 96%



Predictive Models

Goal: Automated methods that accurately predict SME ratings of critical events and performance metrics.

Critical Event Detection

Approach: Spectrum method with discrete moving time windows.
Outcome: Over 80% of critical events were correctly classified.

Performance Assessment

Approach: Combined content analysis with speech stress indicators to select best set of variables.
Outcome: Model predictions for each metric correlated with SME ratings: r=.36 to .43.
For entire missions, model predictions agreed with SMEs: r=.70 to .81.

Conclusions

- Near real-time analysis and modeling of actual, complex data for networked teams.
- Accurate prediction of objective and subjective metrics of performance and critical events.
- Models are automatically derived.
- Can be integrated into systems to monitor and provide feedback.

AAR Tool

An automated AAR tool was developed to support O/Cs to provide training and feedback to teams.

PEARSON

Peter Foltz
(303) 545-9092
pfoltz@pearsonkt.com

parallel consulting

Noelle LaVoie
(303) 817-7436
lavoie@parallel-consulting.com