Tools for Automated Analysis of Networked Verbal Communication

Peter W. Foltz, Noelle Lavoie, Robert J. Oberbreckling and Mark B. Rosenstein

ABSTRACT

One of the richest sources of information about performance in networked teams is their communication data. Nevertheless, analysis of communication requires tools that can assess both the content and patterns of information flowing in the network. This paper describes research and development of a set of tools for the automatic analysis of verbal communication. The tools use language technologies to analyze the content of communication, thereby permitting characterization of the topics and quality of information being transmitted. The tool can be used for automatically deriving accurate team performance metrics from communication in networked teams as well as providing visualization tools to provide teams and commanders better situational awareness. The toolset has the potential for providing near-real-time assessment of team performance including measures of situation awareness, knowledge gaps, workload, and detection of critical incidents. It can be used for tracking teams' behavior and cognitive states, determining appropriate feedback, and for automating After Action Reviews.

INTRODUCTION

As the military incorporates greater network-centric operations and technology, it becomes increasingly important to be able to monitor and assess performance of teams. However there are numerous challenges regarding how to effectively identify, track, analyze, and report on team performance in real-time in such complex operational environments. For example, current methods of assessing team and group performance rely largely on global outcome metrics, which often lack information rich enough to diagnose failures, detect critical incidents, or suggest improvements for the teams or for their collaborative aids. Thus, while there has been an increase in the availability of networked information, there needs to be a concomitant increase in the availability of tools that can leverage off of the networked data to monitor, support and enhance team performance.

Networked teams provide a rich source of information about their performance through their verbal communication. The communication data includes information both about the structure of the network and the content of information flowing on the network. The structure of the network can indicate such things as team member roles, paths of information flow and levels of connectedness within and across teams. The content of the information communicated provides a rich indication of what information team members know, what they tell others, and what their current situation is. in order to be able to assess that team's performance. Thus, communication data can provide information about team cognitive states, knowledge, errors, information sharing, coordination, leadership, stress, workload, intent, and situational status. Indeed, within distributed training, trainers and subject matter experts typically must rely on listening to a team's communication. Nevertheless, in order to exploit the communication data, technologies need to be available that can assess both the content and patterns of the verbal information flowing in the network and convert the analyses into results that would be usable by teams and commanders.

This paper provides an overview of ongoing research and development of a set of tools for the automatic analysis of verbal communication. The tools use language technologies to analyze the content of communication, thereby permitting characterization of the topics and quality of information being transmitted. These tools can both provide metrics of team performance as well as be integrated into applications to provide automated aids for training and operational feedback and monitoring. Finally the paper describe how these tools can be incorporated into to visualization tools designed to analyze the content and patterns of communication streams in order to provide teams and commanders with improved situational awareness and tests in a recent multi-national exercise.

VERBAL COMMUNICATION ANALYSIS

The overall goal of automated verbal communication analysis is apply a set of computational modeling approaches to verbal communication in order to convert the networked communication into useful characterizations of performance. These characterizations could include metrics of team performance, feedback to commanders, or alerts about critical incidents related to performance. In order to do this analysis, there are several components needed. The first component is that there has to be available sources of verbal communication. Second, there must be some pre-existing measures of performance with which to associate the communication to performance. Finally, there must be a set of computational approaches to apply to the communication in order to perform the analysis. These computational approaches include computational linguistic methods to analyze communication, machine-learning techniques to associate communication to performance measures, and finally cognitive and task modeling techniques.

By applying the computational approaches to the communication, we have a complete communication analysis pipeline. Figure 1 shows a representation of the communication analysis pipeline. By combining the tools in the pipeline, we are able to convert spoken and written communication directly into performance metrics which can then be incorporated into visualization tools to provide commanders and soldiers with applications such as, automated AARs and debriefing, near-realtime alerts of critical incidents, feedback to commanders of teams performing poorly, graphic representations of type and quality of information flowing among a team. We outline the approach to this communication analysis below.

A. *Communication data*

For analysis purposes, communication data can include most kinds of verbal communication among team members. Typed communication (e.g., chat, email or instant messages) can be automatically formatted for input into the analysis tools. Audio communication can include the capture of any kind of spoken data, including use of Voice over IP systems, radios, and phones.

Because a majority of military communication is typically spoken, automatic speech recognition systems (ASR) can be applied for converting speech to text for input into the communication analysis system. The communication analysis technologies have been tested for the analysis of ASR input for a number of datasets of verbal communication (see [1]). The results indicated that even with typical ASR systems degrading word recognition by 40%, the model prediction performance degraded less than 10%. Thus, the approach appears to be quite robust to typical ASR errors. Nevertheless, specialized speech models must be developed for each new domain to optimize ASR performance.

B. Performance metrics

In order to provide feedback on team performance, the toolset learns to associate team perform-

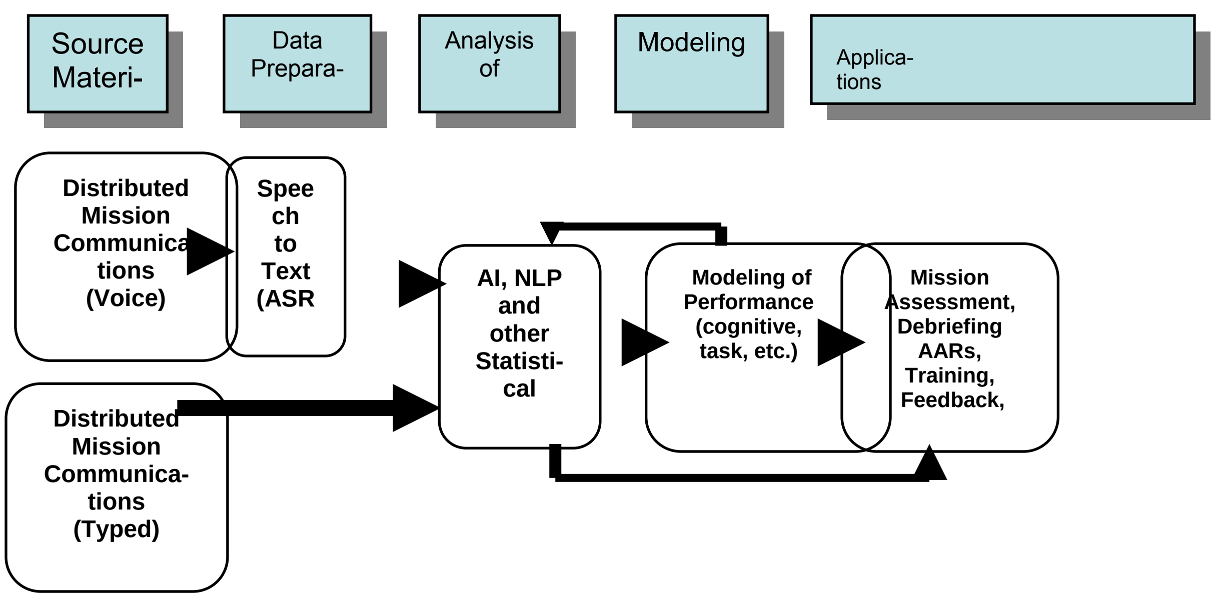

Figure 1. The Communication Analysis pipeline

ance metrics with the communication streams from those teams. Thus, the system typically requires one or more metrics of team performance. There are a wide range of issues in determining appropriate metrics for measuring team performance [2]. For example, metrics need to be associated with key outcomes or processes related to the team's tasks, they should able to be used in a manner to be able to provide feedback on deficiencies for individuals and/or teams, and they need to have a degree of reliability so that experts can agree on both the value of the metric and on how it should be scored for different teams [3].

Objective measures of performance can be used as metrics, indicating specific aspects performed by the teams. These measures can include such aspects as kills, deviations from optimal solution paths, objectives completed, and ACE reports. Alternatively, subjective measures of performance can be used as metrics. These can include, Subject Matter Experts (SME) ratings of such aspects as leadership, management of engagement, following doctrine, communication quality, situation awareness. Additionally SME provided information from AARs or identification of specific critical incidents, failures or errors can be used to measure performance. Nevertheless, all metrics will have varying levels of reliability as well as validity. For new metrics, it is often advisable to use ratings from more than one SME in order to determine reliability.

C. Computational modeling tools

The verbal communication data is converted to a computational representation which include measures of the content (what team members are talking about), quality (how well team members seem to know what they are talking about) and fluency (how well team members are talking about it). This process uses a combination of computational linguistic and machine learning techniques that analyze semantic, syntactic and statistical features of the communication stream.

The primary underlying technology used in this analysis is a method for mimicking human understanding of the meaning of natural language called. Latent Semantic Analysis (LSA) (see [4] for an overview of the technology). LSA is first automatically trained on a body of text containing knowledge of a domain, for example a set of training manuals, and/or domain relevant verbal communication. After such training, LSA is able to measure the degree of similarity of meaning of two communication utterances in a way that closely mimics human judgments. This capability can be used to understand the spoken interactions much in the same way a subject matter expert can compare the performance of one team or individual to others. The techniques has been widely used in other machine understanding technologies including commercial search engines, automated scoring of essay exams, and methods for modeling human language acquisition.

The results from the LSA analysis is combined with other computational language technologies which include techniques to measure syntactic complexity, patterns of interaction and coherence among team members, and statistical features of individual and team language (see [5] for examples of typical language analyses). The computational representation of the team language features are then used with machine-learning technology to predict the team performance metrics. In a sense, the overall method learns which features of team communication are associated with different metrics of team performance and then can predict scores for those metrics for any new set of communication data.

D. Performance metric prediction with the communication analysis toolkit

Tests of the toolkit's use for communication analysis have shown great promise. Tests are performed by training the system on one set of communication data and then testing its prediction performance on a new data set. This ensures that the models built will generalize to new communication. Using range of different types of military communication data, the toolkit is able to provide accurate predictions of the overall team performance and individual team metrics, it makes reliable judgments of the type of statements each team member is making, and it can predict team performance problems based on the patterns of communication among team members [6, 7].

Using human and ASR transcriptions of team missions in a UAV environment, in Air Force simulators of F-16 missions, and in Navy Tactical Decision-Making Under Stress (TADMUS) exercises, the tools predicted both objective team performance scores and SME ratings of performance at very high levels of reliability (correlations typically range from r=0.5 to r=0.9 over 20 tasks). It should be noted that the agreement between the toolkit's predictions and SMEs is typically within the range of one SME to another. In addition, the tools are able characterize the type of communication for individual utterances, (e.g., planning, stating facts, acknowledging) [8].

Applications of the Communication Analysis Toolkit

A number of applications have been developed to further test the performance and validate the use of the toolkit in live situations. Below we describe three applications using the method to monitor and assess learning in online discussion environments, providing realtime analyses and visualizations of multi-national Stability and Support Operation exercises, and providing automated team performance metrics and detection of critical incidents in convoy operations.

A. Knowledge Post

In large networked organizations, it is difficult track performance in distributed exercises. Knowledge Post was designed for monitoring, moderating and assessing asynchronous collaborative learning and planning, The tools within Knowledge Post have been tested in a series of studies at the U.S. Army War College and the U.S. Air Force Academy [9,10, 11]. The application consists of an off-the-shelf threaded discussion group that has been substantially augmented with Latent Semantic Analysis based functionality to evaluate and support individual and team contributions.

Currently Knowledge Post supports the ability:

- To automatically notify the instructor when the discussion goes off track.
- To enhance the overall quality of the discussion and consequent learning level of the participants
- To have expert comments or library articles interjected into the discussion in appropriate places by automatically monitoring the discussion board activity.
- To find material in the discussion or electronic library that is similar in meaning to a given posting.
- To have contributions automatically summarized.

The utility of each of the aforementioned functions has been empirically evaluated with Army officers participating in planner exercises at the U. S. Army War College, and with cadets at the U. S. Air Force Academy

Among the findings of the studies were: the superiority of learning in a Knowledge Post environment over a face-to-face discussion with significant improved quality of discussion $F(2, 113) = 9.5$, $p < .001$; the usefulness of having a software agent automatically alert moderators when groups and individuals are floundering or drifting; increasing the solution quality of the group by interjecting expert comments automatically; and the usefulness to the participants of the Knowledge Post searching and summarizing features (see [9,10]).

TeamViz

TeamViz is a set of visualization tools and enhancements built on the Knowledge Post. toolset TeamViz was developed and run live during a U.S.-Singapore exercise designed to evaluate collaboration among joint, interagency, and multinational forces conducting combat and stability operations [12]. The system automatically analyzed the content and patterns of information flow of the networked communication and provided automated summarizations of the ongoing communications as well as network visualization tools to help improve situation awareness of team members. Analyses showed that the technology could track the flow of commander's intent among the team members as well as detect the effects of information injects on performance within the coalition task force and brigades who participated in the exercise. Singapore Officers used TeamViz in realtime to monitor the communication streams and inform commanders of important information flowing in the network as well as perceived information bottlenecks. Overall, the TeamViz technologies permit knowledge management of large amounts of communication as well as improve cognitive interoperability in distributed operations.

B. Competence Assessment and Alarms for Teams

Convoy operations require effective coordination among a number of vehicles and other elements, while maintaining security and accomplishing specific goals. Nevertheless, in training for convoy operations, it is difficult to monitor and provide feedback to team members in this complex environment. The DARPA Automated Competence Assessment and Alarms for Teams (DARCAAT) program was designed to do automated performance assessment and provide alarms for live and virtual convoy operations training. In currently ongoing research and development, we have collected communication data and SME-based performance measurements and developed specialized tools to assess performance in convoy operations.

Two sources of data were used to develop and validate the toolset, one from teams in a virtual environment and one from teams in live training environments. For the virtual environment, communication data was collected from the Fort Lewis MSTF, PKT which uses the DARWARS Ambush! virtual environment convoy training. In the environment, up to 50 soldiers are able to jointly practice, battle drill training and leader/team development during convoy operations. Live training data was collected at the National Training Center, Fort Irwin from convoy lane STX training. Then, in collaboration with NTC observer controllers, SMEs rated team performance on a number of scales (Battle Drills, Following SOP, Situation Awareness, Command and Control, and overall team performance). The system is currently undergoing development and testing, however preliminary results indicate that the DARCAAT system is able to accurately match SME ratings of team performance as well as detect critical events (e.g., "training moments" or performance alarms) in teams. The developed toolset can then be used to provide automated performance assessment for the Observer-Controllers, permitting efficient automation of After Action Reviews.

CONCLUSIONS

Communication is the glue that holds networked teams together. The verbal content of a team's communication provides a window into its performance and cognitive states of the individuals and the team as a whole. The approach described in this paper can convert the communication into specific metrics of performance thereby permitting a better picture of the state of networked teams at any point.

The toolkit allows the analysis and modeling of both objective and subjective metrics of performance and is able to work with large amounts of communication data. Indeed, because of its machine-learning foundation, it works best with more data. The toolkit can automatically extract measures of performance by modeling how SMEs have rated such communication in related situations as well as modeling objective performance measures. Further, because the methods used are automatic and do not rely on any hand-coded of models, they allows performance models to be developed without the large amount of efforts typically involved in standard task-analysis or cognitive modeling approaches. Nevertheless, the approach can still be integrated with traditional assessment methods to develop objective and descriptive models of distributed team performance.

CHALLENGES/NEW DIRECTIONS

There remain a number of challenges to incorporating automated analysis of the content of communication into full-scale operational and training venues. While the results described in this paper use teams from 3 to about 70 soldiers, it is important to determine the challenges for scaling to large scale operations. In addition, there a number of additional technologies that can be included to improve and help generalization in performance. These include better modeling of network structures, incorporation of additional modalities of information (e.g., event and action information), improved computational modeling tools, and leveraging off of additional advances in how to measure performance in complex networked environments.

The automated analysis of communication can be used in a range of applications. These can include applications to monitor teams, give feedback, provide visualizations of information flow, alert commanders for poor performance, as well as be integrated into adaptable training systems which can adjust training based on performance of the team. Finally the overall approach helps understand the role of communication in complex networks. Results from analyses of teams in real-world situations can help understand both how communication affects team performance as well as how performance is reflected through communication. These are critical goals to achieve if we are to better understand performance in modern networked environments (e.g., [12]).

ACKNOWLEDGEMENT

This work was supported in part by grants and contracts from DARPA, ARI, ARL, ONR, and AFRL. The authors acknowledge the contributions of David Wroblewski, Brent Halsey, Tim McCandless, Kyle Habermehl, and Manju Putcha in development and data analyses for the projects described.

REFERENCES

Foltz, P. W., Laham, R. D. & Derr, M. (2003). Automated Speech Recognition for Modeling Team Performance. In *Proceedings of the 47th Annual Human Factors and Ergonomic Society Meeting*.

Brannick, M.T. Salas, E. & Prince, C. (1997) *Team performance assessment and measurement: Theory, methods, and applications*. Mahwah, NJ: LEA.

C.R. Paris; E. Salas; J.A. Cannon- Bowers (2001). Teamwork in multi-person systems: a review and analysis. *Ergonomics* (43:8) 1052-1075.

Landauer, T. K, Foltz, P. W. & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes*, *25*(2&3), 259-284.

Jurafsky, J. & Martin, J. (200) Speech and Language Processing: An Introduction to Natural Language

Processing, Computational Linguistics, and Speech Recognition, New York, Prentice Hall.
Foltz, P. W. (2005). Tools for Enhancing Team Performance through Automated Modeling of the Content of Team Discourse. In *Proceedings of HCI International,* 2005.
Gorman, J. C., Foltz, P. W. Kiekel, P. A., Martin, M. A. & Cooke, N. J. (2003) Evaluation of Latent Semantic Analysis-based measures of communications content. *In Proceedings of the 47th Annual Human Factors and Ergonomic Society Meeting.*
Foltz, P. W., Martin, M. A., Abdelali, A., Rosenstein, M. B. & Oberbreckling, R. J. (2006). Automated Team Discourse Modeling: Test of Performance and Generalization. In *Proceedings of the 28th Annual Cognitive Science Conference.*
LaVoie, N., Psotka, J., Lochbaum, K. E., & Krupnick, C. (2004). Automated Tools for Distance Learning. Paper presented at the New Learning Technologies Conference. February 18-20, 2004. Orlando, FL
LaVoie, N, Streeter, L., Lochbaum, K. Boyce, L., Krupnick, C., and Psotka, J. Automating Expertise in Collaborative Learning Environments. International Journal of Computer-supported Collaborative Learning. Submitted.
Lochbaum, K., Streeter, L., & Psotka, J. (2002). Exploiting technology to harness the power of peers. Interservice/Industry Training, Simulation and Education Conference, Orlando, FL, December 2-5, 2002.
Pierce, L., Sutton, J., Foltz, P. W., Lavoie, N., Scott-Nash, S., Lauper, U. (2006). Technologies for Augmented Collaboration. *Proceedings of CCRTS,* San Diego, CA, July.
Committee on Network Science for Future Army Applications, (2006). National Research Council. ISBN: 0-309-65388-6

AUTHORS

All authors are with Pearson Knowledge Technologies. 4940 Pearl East Circle, Suite 200, Boulder, CO, 80305. Peter W Foltz is the corresponding author:, phone: 303-545-9092, email: pfoltz@pearsonkt.com

QUOTES WORTH PONDERING

Language is not merely a set of
unrelated sounds, clauses, rules
and meanings;
it is a total coherent system of these
integrating with each other,
and with behavior, context,
universe of discourse
and observer perspective.

— Kenneth L. Pike

The properties of a complex
information system are rarely
independent of the processes by which
it has been produced.

The methods used in such system
development processes always
embed social perspectives on values;
on the power structure of the
organization carrying out the process;
on how to treat conflicts; and so on.

— Kristen Nygaard,
Preface to Claudio Ciborra's
The Labyrinths of Information