Machine Learning Approach to Integrating Heterogeneous Team State and Process Measurement Items

Robert K. McCormack, Kara Orvis
Aptima, Inc.
Woburn, MA
rmccormack@aptima.com, korvis@aptima.com

Arwen H. DeCostanza
Army Research Lab
Aberdeen, MD
arwen.h.decostanza.civ@mail.mil

ABSTRACT

Data driven approaches to measurement can provide unique insights into team states and processes critical to effective training that are often difficult to achieve through traditional observational or survey-based methodologies. The Rational Approach to Developing Systems Based Measures (RADSM; Orvis, Duchon, & DeCostanza, 2013) is one process that has been successfully applied to identify and extract data features from sources such as communications (e.g., email, chat, face-to-face), systems interactions, and physiological sensors. The extracted features are often diverse in how they are extracted and represented (e.g., network-based metrics, content analysis features, temporal dynamics) and can number in the hundreds to thousands of items. This presents a daunting challenge in integrating these heterogeneous items and making sense of their relationships to the overall team state or process measurement and the impact on training. In this paper, we describe an approach utilizing machine learning (ML) models to integrate and better understand heterogeneous measurement items. ML models can integrate features derived from a variety of data through multiple algorithmic means into a single score or set of scores. When objective performance outcomes are available (e.g., from observations or systems data), supervised ML models (e.g., Classification Trees) are utilized to learn the relationships between the data features and the outcomes. In addition, we demonstrate how ML methods can highlight the relative impact of individual features on the overall team states or processes. This data fusion research provides insight into advanced approaches for constructing training metrics from multiple sources of systems-based data. These analytic methodologies enable trainers to gain a better understanding of team states and processes through utilization of data sources and measurement items that have proven difficult to make sense of previously.

ABOUT THE AUTHORS

Robert McCormack is a Principal Mathematician and Lead for the Organizational Performance Assessment Capability at Aptima. He has expertise in the areas of computational linguistics (NLP), machine learning, epidemiological modeling, and human sociocultural modeling and analysis. Dr. McCormack received a Ph.D. and M.S. in Mathematics from Texas Tech University, and a B.A. in Mathematics and Computer Science from Austin College.

Kara Orvis is a Principal Scientist and the Performance Assessment and Augmentation Division Director at Aptima. She has over 15 years of experience conducting military R&D with much of that experience being specific to the Army. Dr. Orvis holds a Ph.D. and M.A. in Industrial-Organizational Psychology from George Mason University and a B.A. in Psychology from Ohio Wesleyan University.

Arwen DeCostanza is Chief (A) of the Real-World Soldier Quantification Branch at the Army Research Laboratory (ARL), Human Research and Engineering Directorate. Dr. DeCostanza is an expert in the areas of team and organizational performance, leading a program of research focused on enhancing performance in complex, networked teams. Prior to ARL, Dr. DeCostanza was a research psychologist for the U.S. Army Research Institute. Dr. DeCostanza received her Ph.D. in Industrial and Organizational Psychology from The George Washington University and her B.A. in Psychology from Wake Forest University.

Machine Learning Approach to Integrating Heterogeneous Team State and Process Measurement Items

Robert K. McCormack, Kara Orvis
Aptima, Inc.
Woburn, MA
rmccormack@aptima.com, korvis@aptima.com

Arwen H. DeCostanza
Army Research Lab
Aberdeen, MD
arwen.h.decostanza.civ@mail.mil

INTRODUCTION

Effective team training relies on the ability of trainers and observers to accurately measure team states and processes in order to provide constructive and accurate feedback to trainees. Traditionally, this is achieved through structured observation, surveys and self-reports. Emergent team states such as cohesion, for example, are still measured primarily through cross-sectional self- or observer-based reporting. In non-laboratory settings, such as dynamic military environments for training mission command, though, these approaches may not be sufficient in capturing the full extent of behaviors and indicators of the team's state. This is because there may be dozens to hundreds of individuals (such as in Army brigade and battalion staffs) performing a variety of tasks across several locations using a variety of tools to communicate and collaborate. There are often not enough observers and trainers to cover each individual participant and individual interactions (such as an e-mail) are often difficult to directly observe. Furthermore, methods that rely on observers or self-report are inherently limited in that they do not afford continuous measurement of the team, are often obtrusive and impractical in non-laboratory training environments, and are subject to rater bias.

Within the context of the U.S. Army, a variety of team constructs have been identified as critical to training mission command. In this effort, we selected three constructs to focus measurement efforts on– shared situation awareness (SSA), shared interpretation of commander's intent (SICI), and cohesion. These three team states are aligned with the Army's mission command doctrine (U.S. Army, 2012b). Specifically, principles of effective mission command include: (1) creating shared understanding of the operational environment (situation awareness), (2) providing clear commander's intent that establishes shared mental models of the operation's purpose, key tasks, and the desired outcome, and (3) building cohesive teams through mutual trust.

The relatively recent utilization of mission command systems and communication tools within military training environments has led to an exponential rise in the amount of system and communication-based data available. This wealth of data presents an opportunity to extend and increase the accuracy of team measurement through unobtrusive means, but also presents a challenge in determining how to effectively and validly utilize the data. The sheer quantity of data (tens of thousands of individual messages may be sent across a variety of communication channels) means that manual analysis methods are tedious and labor-intensive at best, if not altogether impossible. New technological advances from a variety of computer and data science disciplines provide promise that there are tractable solutions to deriving meaningful measurements from this data, but require careful and principled application.

In this paper, we provide initial promising results from the application of machine learning (ML) methods to communication-based data for the purposes of automated measurement of team states to enable effective training and coaching. Our goals are to demonstrate that these approaches can (1) discover meaningful patterns (in the sense that they can effectively correlate derived features from the data to traditional survey-based team measurement) in the communications data, and (2) identify which data features are most impactful on overall team state. We first describe the training environment and the data that was collected for this study. We then elaborate on extraction methods that help derive potentially informative features from the data. After introducing the variety and purpose of ML methods, we then describe how one particular method (Classification Trees) was developed and applied to the data. Finally, we discuss the potential of these methods to have a high impact on team training and evaluation.

Training Environment and Data

In this study, we collected data at an Army brigade-level training exercise led by the Mission Command Training Program (MCTP). The MCTP trains battalion and higher units in Army Doctrine, Mission Command, and Operations Processes during highly-realistic week-long exercises. These events help the units prepare for full spectrum operations in decisive action and stability scenarios. They mirror the high-tempo continuous nature of both the planning and execution of operations. During Mission Command and "throughout the operations process, commanders encourage continuous collaboration and dialogue among commanders, staffs, and unified action partners to create shared understanding and facilitate unity of effort" (US Army, 2012a, p. 5). Mission Command therefore requires effective teamwork across space and cyberspace, over time, and in every echelon. Aspects of good teamwork include: high levels of unit cohesion to help units withstand the demands of combat (TRADOC, 2010a, p. 21), mutual trust that flows through the chain of command (US Army, 2012b, pp. 2–2), clear awareness of commander's intent so subordinates can exercise proper initiative in unexpected situations (US Army, 2012b, pp. 2–4), and accurate and timely situational awareness which enables mission command (TRADOC, 2010b, p. 40). In the end, good teamwork relies on good communication since information "needs to flow up and down the chain of command as well as laterally to adjacent units and organizations" (US Army, 2012b, pp. 2–86). When issues occur during training or operations, they are often blamed, after the fact, on poor communication.

Observer-Coach-Trainers (OCTs) from the MCTP are charged with conducting these exercises, observing actions of individuals and the team, supporting the commander's training goals, and running mid and final after action reviews (AARs) for the training unit. To effectively train Mission Command, OCTs need to be aware of the full range of teamwork and communications occurring within unit. While they can observe some face-to-face interactions, most communications are hidden from view via radio channels, email, chat, voice over IP (VOIP), and face-to-face interactions occurring in different locations. In this study, we focused on three particular communication-based data sources: email, XMPP (Extensible Messaging and Presence Protocol) chat, and face-to-face communications as measured through Sociometric Badges.

During an MCTP exercise, units operate on a closed network and run an internal Microsoft Exchange Server for e-mail services. We enable journaling on the server (a feature that sends a copy of all e-mail traffic to a single dedicated e-mail address) and collect all messages in a central mailbox. E-mail messages are characterized by the sender address, receiver(s) address(es), the receiver status (if they were on the to, cc, or bcc line), the subject text, the body text, and the time sent. In addition to e-mail, the units utilize a chat server and clients for real-time tactical coordination. Most commonly, they run an XMPP (Extensible Messaging and Presence Protocol) server such as OpenFire on the internal network along with TransVerse chat clients. Because XMPP is open-source, we can programmatically collect data through an API (Application Programming Interface). Unlike e-mail, where messages are sent directly between individuals, chat messages are posted to chat rooms. We can infer that individuals received (or were likely to receive) the message based on their presence status in the chat room. Table 1 includes the relevant features of the email and chat data.

Within a dynamic environment such as the MCTP exercises, many of the communications occur in ad-hoc face-to-face conversations. In order to record these communications, we employ a sensor known as a Sociometric badge (Olguin and Pentland, 2008). The badges are approximately the size of a deck of playing cards and are worn on a lanyard around the neck or are clipped to the front of the individual's shirt. The badges contain a suite of sensors in a single enclosure. The badges provide the ability to capture face-to-face interaction time using an infrared (IR) sensor that can detect when two people wearing badges are facing each other within a 30-degree cone and a one meter distance. In addition, radio sensors are used to record interactions via a Bluetooth handshake protocol and RSSI (Received Signal Strength Indicator) value to provide a coarse indication of relative distance between badges at 30-second intervals. Sociometric badges are also capable of recognizing common daily human activities (such as sitting, standing, walking, and running) using a 3-axis accelerometer. In addition, an omnidirectional microphone is used for extracting speech features in real time to capture non-linguistic social signals such as interest, excitement, the amount of influence each person has on another in a social interaction, and unconscious back-and-forth interjections. Table 1 summarizes all of these data features.

Table 1. Data Sources and Features Utilized

Source	E-Mail	XMPP Chat	Sociometric Badges			
			Infrared Ping	**Bluetooth Ping**	**Accelerometer**	**Speech Characteristics**
Features	Sender address	Sender chat handle	Sender Badge ID	Sender Badge ID	Badge ID	Badge ID
	Receiver(s) Address(es)	Chat room name	Receiver Badge ID	Receiver Badge ID	Time recorded	Time recorded
	Receiver status (to, cc, bcc)	Chat handles present in each chat room	Time sent	Time sent	X-axis energy	Speech amplitude
	Subject text	Chat text	Time received	Time received	Y-axis energy	Speech signals
	Body text	Time sent		RSSI signal strength	Z-axis energy	
	Time sent					

Self-Report Survey Items

While the ultimate goal of utilizing ML models is to supplement (or perhaps eventually replace) traditional survey methods for assessing team states, we need to first garner an understanding of how the objective data we collect from the above communication sources relates to the team state constructs of interest: SSA, SICI, and cohesion. Each of these constructs was measured at two points during the training exercise. The first was after the mission planning phase and the second was after the mission execution phase. We measured these constructs utilizing self-report surveys which are briefly described here.

Shared Situation Awareness (SSA). SSA was measured using four items developed by the research team and used in previous unpublished studies. Specifically, the items assessed the team's ability to identify mission-critical cues, understand what was going on, predict what was about to occur next, and know how to best achieve the team's goals. These items align with the Endsley model of SA (Endsley, 1995), which has three hierarchical phases or levels. Level 1 SA (perception - identify) is the most basic and involves the perception of the elements in the environment within a volume of time and space. Level 2 SA (comprehension - understand) has a processing component and includes both the individual's comprehension of the current situation and an understanding of significant elements as they relate to their task-specific goals. Level 3 SA (projection - predict) is the highest level of SA in her model and it goes beyond simply perceiving and understanding elements in the environment, but involves projecting the future actions of elements in the environment. The items were each rated using a four-point scale, ranging from very (easy/well/aware) to very (difficult/poorly/unaware), and reverse coded for ease of interpretation. The self-report surveys collected ratings of SSA for both the planning and action (mission) phases of the exercise. The self-report responses to the four items were aggregated to allow for a planning phase SSA score and an execution phase SSA score for each individual.

Shared Interpretation of Commander's Intent (SICI). For the self-report surveys, three measures were developed to capture both the explicit and implicit nature of Commander's Intent. The first measure, the Explicit SICI measure, was intended to measure the degree of shared understanding regarding Commander's Intent as based on the published Commander's Intent statement. The measure consisted of five multiple choice items, each with one correct answer. A total score was equal to the number of questions correct, ranging from 0 to 5. The second measure, the Values-based Implicit SICI measure was intended to capture the shared understanding of the commander's basic values. The measure included 14 pairs of competing values and participants were asked to note which of two values the commander had a preference for. Participants received one point for each selection that matched the commander's, ranging from 0 to 14 (for both planning and mission phases). The third measure, the scenario-based Implicit SICI measure, was intended to capture the degree to which a participant could anticipate the commander's preferential course of action for an ambiguous situation. For this measure participants were provided with five unfamiliar scenarios. Each scenario had two equally attractive or unattractive courses of action. The participants were asked to select the course of action that their brigade commander would prefer. The participants received a point for each correct response (as indicated by the brigade commander), with scores ranging from 0 to 5 (for both planning and mission phases). The values-based and scenario-based questions and responses were discussed with the brigade commander prior to the exercise.

Cohesion. Cohesion was assessed via a 10-item scale adapted and combined from scales used in prior research (e.g., Mullen & Copper, 1994). Three items were used to tap into the social cohesion of the team (e.g., *This team is a close one*); four items were used to target task cohesion (e.g., *The team is united in trying to reach its goals for performance*); and three items were included to assess group pride (e.g., *Solders are proud to be part of this team*). The items were rated on a five-point Likert scale ranging from strongly disagree (1) to strongly agree (5). The items for each cohesion dimension (e.g., task cohesion) were aggregated to allow for a single task, social and pride score for each individual during each mission phase (planning and execution).

FEATURE EXTRACTION

While possible to directly apply some ML approaches to the raw data described above, results are often greatly improved through a preliminary step known as feature extraction. Feature extraction describes the process of building derived values from the raw data with the intent of reducing the dimensionality of the data, eliminating redundancies, and perhaps most importantly, providing human interpretable results from the ML methods. For example, if we were applying ML methods to find patterns in a corpus of images, we could directly utilize each pixel as input and potentially find repeating patterns. However, it might be more useful to first apply image recognition technologies to identify objects (e.g., a person or a tree) present in the images and then apply the ML methods on those objects. The results would then be at the level of human interpretable objects (versus pixels) and would almost certainly be more meaningful.

Similarly in our study, the goal was to take the raw communications data, which consists of messages sent between individuals and groups, and derive a set of meaningful features that represent behaviors likely to be related to team states. In order to identify these behaviors and how they manifest within the data, we utilized a previously-developed process known as the Rational Approach to Developing Systems Based Measures (RADSM; Orvis, Duchon, & DeCostanza, 2013). The RADSM process relies on the underlying assumption that as team members interact with each other and take actions (such as sending an email, executing a command in a program, or having a face-to-face meeting) they leave behind "trace data" that can provide windows into the team's state and processes. In order to produce measures that are both relevant to the constructs of interest and reasonable to implement given the available data, RADSM employs a bi-directional approach. As a top-down methodology, team state and process constructs are broken down into behavioral indicators based on the theoretical underpinnings for the constructs and the context of the data collection environment. From a bottom-up point of view, the raw data available to collect in the environment is analyzed to determine the most reasonable set of data features to focus on the maximize information gain on the construct of interest. Utilizing both approaches allows us to arrive at a set of measurement items that are both reasonably realistic to produce from the available data and relevant to the overall measurement goals.

To illustrate this process, we present a breakdown of the cohesion team state into subcomponents and indicators in Table 2. These represent the top-down half of the RADSM process; that is, first breaking down cohesion into subcomponents based on previous research and literature and then further identifying behavioral indicators that embody these subcomponents within the context of the training environment.

From the bottom-up side, we start with the data and data features described previously and identify which of the indicators in Table 2 can be extracted from those sources. For example, one way to measure the indicator "All team members engage in face-to-face interactions" is to utilize the Sociometric badge IR data and calculate the density of the team network. Likewise, "Team members use positive, uplifting language when talking with one another" can be measured through Dialogue Act (DA) analysis of e-mail messages between team members. While these data measurements might not provide the full depiction of the behavioral indicators, they do provide a piece of the picture, and taken in conjunction with other measurements may provide the needed insights into the overall team construct. We mentioned two measurement features: density and DA analysis. These represent network-based and text-based analysis techniques for extracting data features respectively. We now describe these and similar approaches.

Table 2. Subcomponents of Cohesion and Example Indicators.

Cohesion Subcomponent Description	Indicators
Task cohesion reflects the extent to which team members "remain united in the pursuit of its goals and objectives" (Carron, 1982; p. 124; see also Tekleab, Quigley, & Tesluk, 2009)	• All team members are on task – using task-related language and engaging in work-related tasks. • Team members use language that emphasizes "teamness" and suggests unity in decisions, outcomes, etc. • (All) team members contribute to task discussions, participate in meetings and face-to-face conversations, email, document creation, etc. • Team members agree and are certain when planning and acting out their tasks.
Cohesion Subcomponent Description	**Indicators**
Social cohesion, or interpersonal or group, cohesion reflects the extent of "the bonding together of members of a unit in such a way as to sustain their will and commitment to each other..." (Johns et al., 1984; p. ix). Also reflects the extent of team members' "attraction to or liking of the group" (Evans & Jarvis, 1980), "attraction and commitment to their team, [and] team members..." (LePine et al., 2008); or "being united, 'sticking together' and forming social bonds..." (Carron, Widmeyer, & Brawley, 1985).	• Team members use positive, uplifting language when talking with one another. • Team members do not have (or have limited) conflicts and do not use negative language when talking with one another. • Team members do not form sub-groups unless required by the mission or the organization, and even then, there continues to be links between sub-groups. • Team members interact outside of the required interactions in the mission environment.

Network Features

For a given set of communications between individuals, we can characterize the data as a network graph – a set of vertices (individuals) and edges (messages). This representation affords us a number of descriptive analyses that can inform the overall team measurements. When constructing a network from a data source, there are several considerations that must be taken into account. Networks can be either directed or undirected. In directed graphs, edges between vertices can have a direction associated with them – *from* one vertex and *to* another vertex. Undirected graphs, on the other hand have no associated directionality on edges. For some data sources, such as e-mail, a directed graph may be most informative as there are clear senders and receivers of messages. For other data, such as face-to-face infrared communications, directionality may not be appropriate as the interaction is symmetrical. Network graphs may also have edges that are either weighted or unweighted. Weighted graphs have a numerical value associated with the edges which could represent, for example, the number of communications occurring between the respective vertices. Unweighted graphs have binary edges – either an edge exists or it does not. Choosing the right representation for the network graph depends on both the data source under consideration as well as the questions being asked by the analysis.

Once a network representation has been created a number of analyses can be performed on the graph. We will discuss a few common approaches but note that there are a multitude of descriptive analyses that can be performed depending on the questions being asked.

- ***Distance*** between two vertices in a network graph refers to the number of edges on the shortest path between them. If the edges are weighted, the distance can be computed as the sum of those weights. In a military setting, the distance between and individual and the unit commander can provide strong indications of interpretation of commander's intent, for example.
- ***Density*** is the proportion of edges that exist compared to the total number of possible edges. For an undirected network with N number of vertices and M number of edges, the density is given by:

$$\frac{2M}{N(N-1)} \quad (1)$$

Density provides us with an overall measure of how well connected a team is. This can provide indications of, for example, cohesion as discussed previously in this paper.

- ***Centrality*** refers to a number of approaches that provide an indication of the relative importance of a vertex in the network. ***Degree Centrality*** is simply the number of edges connected to the vertex in question. High degree centrality would indicate that the vertex communicates directly with a large number of other people. ***Closeness Centrality*** is computed as the average length of the shortest paths from the vertex in question to all other vertices in the network. This metric provides an indication of how far reaching the influence of the individual is. ***Betweenness Centrality*** is defined as the number of shortest paths in the network that pass through the vertex in question. High betweenness centrality is an indication of the brokering power of the individual – that is, if individuals need to go through a person to communicate with others in the network, that person would score highly on betweenness centrality.

Text Features

The content of communications can provide rich insights into both what the team is discussing, but also how they are communicating with each other. Natural language processing (NLP) approaches offer a wide variety of techniques for analyzing free text, such as found in e-mail or chat. These vary from low-level grammatical parsing of sentences to high-level summarization of documents. We utilize two methods for text-based feature extraction: topic analysis and dialogue analysis.

Topic analysis encompasses a set of approaches for generating descriptive topics for a set of documents. These topics provide a general sense of what is being discussed in each document. For example, topics may describe an e-mail as being about civilian elections or an IED attack. We utilize a technique called Latent Dirichlet Allocation (LDA; Blei, Ng, Jordan, 2003) to derive the topics. This is a statistical technique based on a bag-of-words model of the document corpus. These topics can be used as a basis for a variety of team-related measurement. For example, a component of *Task Cohesion* is "All team members are on task – using task-related language and engaging in work-related tasks." Analysis of the topics used by each team member in relation to their prescribed duties can provide an indication of this component.

Dialogue analysis, on the other hand, aims to measure *how* individuals are talking to each other. For example, are individuals asking lots of questions or are they expressing frustration? We utilize a tool called the General Architecture for Text Engineering or GATE (Cunningham 2011; Cunningham et. al., 2013) to extract dialogue acts from the text. Utilizing a number of gazetteers, or word lists, we identify dialogue acts such as frustration, politeness, uncertainty, appreciation, clarification, or criticism within the message text. From these extracted features we can relate the messages to the theoretical team constructs. For example, a component of Task Cohesion is "Team members agree and are certain when planning and acting out their tasks." In this case we would identify the proportion of team emails and chat messages that contain certainty and acknowledgement dialogue acts.

Even with the dimensionality reduction and focused feature extraction afforded through the RADSM process, the space of measurement items can still be quite large and daunting to make sense of manually. As an example, let's assume that through the RADSM process we hypothesize that network density during planning meetings is indicative of cohesion. For this single feature we must consider a multitude of variables such as: across which time periods it is measured; the data sources from which it is computed; the directionality of messages in the network; whether it is measured at the level of the team, division, or other organizational structure; and so on. If we assume there are 2 phases during the data collection, 4 data sources (e-mail, chat, face-to-face, and all combined), 3 directions to compute (sent messages only, received only, or both sent and received), and 2 organizational contexts (computation of density within the team and within entire organization) this gives us 48 total density measurements. This could even be expanded further by considering only messages with positive sentiment in the text or only messages that mention are particular topic. The number of possible measurements just for network density grows exponentially even with the data reduction provided through the RADSM processes.

MACHINE LEARNING APPROACHES

Given the arbitrarily large set of possible features that we want to consider in the measurement of team states and processes, machine learning methods offer a promising approach to making sense of the complex feature space and its relation to the outcome variables. Our goals in applying machine learning are twofold. First, we wish to discover patterns within the feature set itself, such as correlations between different features or the relative importance of each feature. Second, we wish to learn which combinations of features can provide accurate estimates and predictions of team states and processes.

At the highest level we can divide machine learning approaches into two categories: supervised and unsupervised methods. In supervised machine learning tasks, the system is provided with example data inputs and their corresponding outputs. Its goal is to learn a mapping from inputs to outputs that minimizes the error in the training examples while providing high predictive accuracy for novel inputs. Examples of supervised learning approaches include backpropagation neural networks, support vector machines, and decision trees. In unsupervised machine learning methods, no exemplar outputs are provided for the inputs. The goal is to find inherent structure within the input data itself. Examples of unsupervised methods include k-means clustering, latent variable analysis, and self-organizing maps. In this effort, we chose to develop a supervised machine learning model known as classification trees (Breiman, et. al., 1984), which are a particular case of decision trees.

Decision Tree Models

Decision trees are an approach to modeling data that utilizes a graph-like tree structure to represent decision points and outcomes. The aim is to arrive at an answer to a question after traversing through the tree. As an example, consider Figure 1 which provides a simple decision tree for the question "Should I water the lawn today?" The yellow leaf nodes are the overall outcome answers (yes or no, in this case) and the white internal nodes are the decision points. Starting from the top and working down, the tree is traversed by answering each decision node until a leaf is reached. In this example, we would first determine if it is currently raining. If "yes" is the answer, then we reach a leaf node with the final answer of "no" for the question. If, on the other hand, "no" is the answer, we would then continue on to the next node which asks whether or not the forecast calls for rain. Again, if the answer is "no", the final decision point asks how long it has been since the last time the lawn was watered. If it has been less than 48 hours, the final answer is "no". If it has been greater than 48 hours, then the answer is "yes".

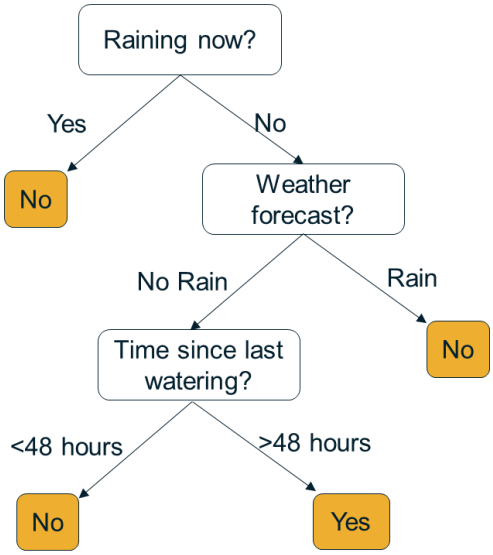

Figure 1. Example of a decision tree for answering the question "Should I water the lawn today?"

While this example is quite simple, arbitrarily complex decision trees can be created to answer a variety of questions and to model a variety of data. However, it becomes intractably difficult to develop a decision tree manually when the number of independent variables under consideration becomes large. In this case, we can employ a technique called decision tree learning which aims to automatically derive a decision tree based on a given set of training data for the independent variables and the corresponding dependent variable. When the dependent variable can take on one of a finite set of values (such as on a Likert Scale), the model is known as a classification tree. If the dependent variable can take on any value in a continuous range, it is known as a regression tree. There are a number of algorithmic approaches to decision tree learning. In general, the process operates recursively by selecting the variables that best split the dependent variables into different classes. Because of this, it can generally be assumed that nodes that fall higher in the tree have a larger overall impact on the outcome variables. For the sake of brevity we omit further algorithmic discussion here but point to (Loh, 2011) as a suggested overview of several of these approaches.

APPLICATION OF CLASSIFICATION TREES TO TEAM TRAINING MEASUREMENT

We utilize decision tree learning, and specifically classification trees, in this effort to develop a predictive model connecting the large set of features derived through RADSM process with the self-report survey items from the MCTP exercise. The ultimate goal of this model is to provide timely and accurate team state assessments through automated data analysis as would normally be derived from surveys.

For our initial model, we considered three features for each person derived from the Sociometric badges: his/her degree centrality, his/her betweenness centrality, and his/her distance within the network to the leader. We chose these three as the basis for the model after consideration of the theoretical constructs (SSA, SICI, cohesion) from the RADSM process. For each of these network measures, a host of parameters were considered. First, we considered the organizational context in which the measurement is taken: their immediate cell, their larger team, their even larger unit, or the overall organization. We also took into account if the measurement was taken during the planning phase of the exercise, the action phase, or the overall exercise. We looked at both unweighted networks and networks weighted by number of messages. We considered both infrared and Bluetooth signals from the sensors. Finally, we looked at networks that were either undirected, directed by messages sent, or directed by messages received. Combinatorically, this gives us 432 different input variables to consider. These are summarized in Table 3.

Table 3. Parameters for selection of independent variables to be included in the model

Algorithm	Organizational Context	Exercise Phase	Weighted / Unweighted	Data Source	Directionality
Degree centrality	Cell	Planning	Weighted	Infrared	Sent
Betweenness centrality	Team	Action	Unweighted	Bluetooth	Received
Distance to the leader	Unit	Entire			Undirected
	Overall				

For each of the survey items (SSA, SICI, and Cohesion) a classification tree was trained on these input variables. The result is a predictive model that allows us to assess these survey items based solely on the interaction data input into the model. We present a partial view of the classification tree developed for SICI in Figure 3. Because of the large number of variables, it is impossible to depict both the overall span of the tree as well as each individual node in the same picture. Only those variables near the top of the tree (and therefore which are most discriminatory) are shown for clarity here.

As stated previously, the higher a node falls in the tree, the greater the overall impact it holds on the outcome variable. In this case, we find the weighted network distance to the leader of sent Bluetooth messages as measured for the overall organization across the entire exercise is the most impactful on SICI. At face value, this makes sense as your distance to the commander is likely to be related to how well you interpret his or her intent.

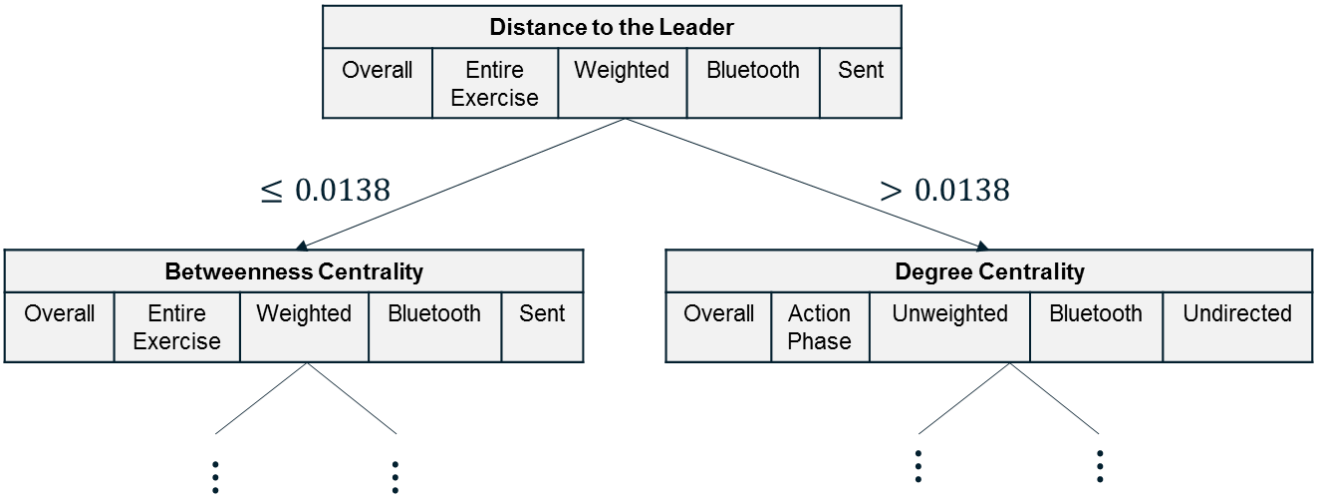

Figure 2. First split in the SICI classification tree

IMPACT ON TEAM TRAINING

The ultimate goal of the MCTP is to provide actionable and timely feedback to warfighters to ensure that they understand and integrate those lessons into their ongoing mission training. Current approaches to team assessment based on observations and surveys are slow and may miss critical aspects of the team's actions and state. The machine learning model described here provides a basis for improving training through real-time, less intrusive, quantitative assessments.

We believe that there are several advantages to this model over current training standards. While OCT observations of the training exercise are based on a high-level of experience and expertise, this approach is ultimately insufficient to provide a complete assessment. At the levels of brigade and higher unit training, there are hundreds of individuals interacting in a dynamic and distributed environment. It is simply not possible for observers to cover each warfighter role continuously given both time and space constraints. Furthermore, individual actions and team interactions are not always observable due to the variety of tools and communication media used by the warfighters. Even when OCTs are observing critical training events, they miss subtle interactions occurring through e-mail, chat, or side conversations. And indeed, given the subjective nature of observer-based assessment, rater bias is an ever-present concern. Self-report surveys are another tool used to gather data for trainee assessment, but they too are subject to issues of bias, timeliness, and incompleteness.

The machine learning model and similar approaches can provide a means for filling in the gaps left by current training approaches. We envision the model to serve not as a replacement for the OCTs, but rather another tool they can use to integrate their expertise with a more complete objective assessment of the team. The classification tree approach provides a means for ongoing assessment based on a variety of data (e-mail, chat, face-to-face communications) that is currently underutilized (or, often, completely unutilized) by the training team. Through the quantitative assessments of team constructs, such as cohesion, SSA, and SICI, the OCTs can provide trainees with a more comprehensive assessment. As the output of these values from the model changes throughout the exercise, OCTs can investigate the underlying reasons for these changes and enable learning opportunities for the participants. Because the classification tree model exposes the underlying data and variables used in the assessments, OCTs can easily investigate the reasons for the model outputs. For example, if the model predicts a drop in SSA, the OCT may see that this is due to a drop in overall team density and suggest that trainees increase communications during critical events.

CONCLUSIONS AND FURTHER RESEARCH

We believe that machine learning approaches to team training can provide a powerful tool to the arsenal of trainers. The ability of the models to integrate across a variety of data and analyses enables more complete, timely, and objective data for OCTs to utilize during training exercises. Our initial results show a great deal of promise that classification tree models will provide utility for trainers, but further work is needed to show the validity of the results.

To avoid overlearning of the model, we will also need to determine how applicable a tree trained on one training event will be to future events and identify additional contextual variables that may need to be considered. In addition, the model does not currently predict team performance, but rater provides team state assessments to OCTs for their evaluations. In future iterations of the model, it may be possible to provide objective performance assessments, in part, through the use of this model. Our next goal is to integrate data from several additional training exercises to determine which constructs are generalizable across different participants and scenarios.

ACKNOWLEDGEMENTS

This work was partially supported under U.S. Army contract W911QX-12-C-0026. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army or the U.S. Government.

REFERENCES

Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). *Classification and Regression Trees*. (The Wadsworth Statistics/Probability Series). Belmont, CA: Wadsworth.

Carron, A.V. (1982). Cohesiveness in sport groups: interpretations and considerations. *Journal of Sport Psychology, 4*, 123 – 138.

Carron, A.V., Widmeyer, W.N., & Brawley, L.R. (1985). The development of an instrument to assess cohesion in sport teams: the group environment questionnaire. *Journal of Sport Psychology, 7*, 244 – 266.

Cunningham, H., Tablan, V., Roberts, A., & Bontcheva, K. (2013). Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics. PLoS Comput Biol 9(2): e1002854. doi:10.1371/journal.pcbi.1002854

Cunningham, H., et al. (2011). Text Processing with GATE (Version 6). University of Sheffield Department of Computer Science. 15 April 2011. ISBN 0956599311.

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Hum. Factors 37 (1)*, 32 – 64.

Evans, C.R., & Jarvis, P.A. (1980). Group cohesion: A review and reevaluation. *Small Group Behavior, 11*, 359–370.

Johns, J.H., Bickel, M.D., Blades, A.C., Creel, J.B., Gatling, W.S., Hinkle, J.M., Kindred, J.D., & Stocks, S.E. (1984). *Cohesion in the U.S. military.* Washington, DC: National Defense University Press.

LePine, J.A., Piccolo, R.F., Jackson, C.L., Mathieu, J.E., & Saul, J.R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology, 61*, 273-307.

Loh, W.Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 1, 14–23.

Mullen, B., & Copper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin, 115:* 210-227.

Olguin, D.O., & Pentland, A. (2008). Social sensors for automatic data collection. *14th Americas Conference on Information Systems, Toronto,* 1-10.

Orvis, K., Duchon, A., & DeCostanza, A. (2013). Developing Big Data Based Performance Measures: A Rational Approach. *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), vol. 2013*, Orlando, FL.

Tekleab, A.G., Quigley, N.R., & Tesluk, P.E. (2009). A longitudinal study of team conflict, conflict management, cohesion, and team Effectiveness. *Group & Organization Management, 34*, 170-205.

TRADOC. (2010a). TRADOC Pam 525-3-1: The Army Operating Concept. Fort Monroe, VA: Training and Doctrine Command.

TRADOC. (2010b). TRADOC Pam 525-3-3: The United States Army Functional Concept for Mission Command 2016-2028. Fort Monroe, VA: Training and Doctrine Command.

US Army. (2012a). Army Doctrine Publication (ADP) 5-0: The Operations Process. Fort Leavenworth, KS: United States Army Combined Arms Center.

US Army. (2012b). Army Doctrine Reference Publication (ADRP) 6-0: Mission Command. Fort Leavenworth, KS: United States Army Combined Arms Center.