Automated Techniques for Measuring Team Communication

Peter W. Foltz & Melanie J. Martin
New Mexico State University
pfoltz@nmsu.edu

At the operational warfighting level, Objective Force units as part of joint teams will conduct operational maneuvers from strategic distances
(U.S. Army White Paper: Concepts for Objective Force)

The Problem

Team communication data can provide a rich data set for analyzing team performance. However, hand coding for content can take a very long time and can be subjective.

Goal

Predict team performance based on automated analyses of communication among team members.

Approach

•Use Latent Semantic Analysis, a computational model of language and semantic memory, to code/measure and content of communication team tasks.
•Evaluate effectiveness of these techniques for predicting team performance and decision making.
•Evaluate the combination of these techniques with other methods of cognitive modeling for improving measurement performance.

The CERTT lab

•Research laboratory to develop, apply, and, evaluate measures of team cognition.
• Hardware and software support synthetic team tasks
•Can be configured to simulate co-located or distributed task environments
•Current Configuration: Uninhabited Air Vehicle Control

•An automated measurement and recording system captures team behaviors:
- audio & video streams
- task performance indices
- communication flow
- computer events

•Post processing routines summarize the data.
•Experimenters observe team behavior remotely via:
- audio & video monitors
- performance indices
- shared displays

•Experimenters record observations using time-stamped annotation software.
•Participants complete on-line measures that elicit task and team knowledge:
- factual tests
- structured interviews
- concept ratings (individual & team levels)

Latent Semantic Analysis (LSA) is a:

- Psychological Theory
  - •A theory of the acquisition, induction, and representation of knowledge
  - •How people learn the meanings of words
- Model
  - A mathematical system for computational modeling of cognitive processes
- Tool
  - An Artificial Intelligence (Machine Learning) system for matching words/texts at a semantic level

➢ LSA learns the relationships between text documents and their constituent words (terms) when trained on very large numbers of background texts (thousands to millions)

➢ LSA learns how to group documents and terms that are similar in a "Semantic Space"

➢ Documents can be domain knowledge, writing samples, e-mail files, course materials, personnel records, etc.

➢ LSA judgments of similarity agree well with human judgments

Measuring Team Communication with LSA

•A good team

__PLO__: DEMPC, this is PLO, what would be the next target after SEN-1
DEMPC: Your next target is RSTE, repeat RSTE, with an effective radius of 5
PLO: Effective radius is 5?
DEMPC: Yes ma'am
PLO: Roger that.
DEMPC: AVO, uh, right next to STR is MSTR and that has an altitude rule of a minimum of 2000 and a maximum of 5000, so if you want to do it for both, go right ahead.
PLO: AVO this is PLO, for this site I need you to be above 3000
AVO: OK, right now, PLO, we're right at 3150 so that should be OK for both, right?
PLO: Sounds good to me.
DEMPC: sounds good to me.
AVO: Alright.

A poor team

DEMPC: AV, AVO?
AVO: This is AVO.
DEMPC: Ok, what's going on? Why are we off course?
AVO: We're en, We're enroute to OAK.
DEMPC: Yeah, but we're way off. You are going around it.
AVO: Around it?
DEMPC: Well, it looks like it is coming down but
AVO: Well, I'm adjusting right now the course deviation.
DEMPC: You should have done that a long time ago, AVO, come on.
AVO: We're only like 10 10 degrees off though, it shouldn't be that much.
DEMPC:But see, that ROZ exit, that's a ROZ exit, we have to hit it at a radius of 2.5
AVO: 2.5 OK
DEMPC: OK
AVO: But according to my map we're doing good.
DEMPC: OK
PLO: Uh, where it says waypoint, the two, shouldn't it already say SST?

How do we automatically determine what makes a good/poor team?
•Analyze semantic components of communication
•Use these semantic components to predict knowledge and communications skills and predict team performance

Experiments

The Data
67 Transcripts from 11 teams, 7 missions
- • XML tagged
- •~2700 minutes of spoken dialogue
- •20,545 separate utterances (turns)
- •232,000 words (660 k bytes of text)
- •Logs of the speaker, listener, and duration of each communication from each participant

Experiment 1: Predicting team performance from dialogues as a whole
Goal: Predict objective individual and team performance measures based on transcript of team communication.

Approach:
- •Match team dialogue patterns and content against database of prior dialogues.
- •Assign a predicted team score based on similarity of a transcript to other team transcripts with known performance measures.

Result: Correlation between LSA-derived measures of communication to Team Performance measures r=.76 (p<.01)

Implication: We can automatically predict how well a team is performing based on analysis of their communication

Experiment 2: Automatic Tagging of Transcripts
Goal: Tag utterances from transcripts for types of dialogues

Frequency and sequences of tags can predict team performance (e.g., Bowers, Jentsch, Salas, & Braun, 1998)

Codes: *Uncertainty, action, acknowledgements, planning, factual, non-task related …*

Approach:
- •For each utterance, find the most semantically similar utterances that have already been tagged. Assign a probability of tags to that utterance.
- •2893 separate utterances coded by two human coders and automatically by computer.

Results
- •Human-human reliability: 0.71 (c-value of agreement)
- •LSA-Human reliability: 0.59
- •LSA+Syntax - Human reliability: 0.63
- •LSA-based confidence measure of tagging can help improve performance to be even closer to that of human performance

Implications:

We can automatically tag transcripts and use results to predict team performance.

Tagging by computer takes seconds compared to minutes to hours by human coders

Performance is not quite at human-human reliability, but can provide an acceptable level of accuracy to provide fast predictions.

Conclusions

•Objective Force and FCS put greater emphasis on distributed team communication.

• Monitoring and assessing team performance will become more critical in these distributed contexts.

•Discourse level is best for obtaining diagnostic information for training, design, and selection.

•Semantic and statistical analyses of team dialogues can reveal the effectiveness of a team.

•Permits automatic analyses of the content of team dialogues.

•This approach can be applied to any domain in which there is team dialogue.

Current Extensions

Creation of new laboratory for studying team communication in a simulated SASO context underway at NMSU using DDD

Development of a web-based communication analysis system for use by ARL CTA member underway

Can use the approach to determine what makes a good team
- –Match team dialogue patterns and content against database of prior dialogues.
- –Predict individual and team performance.
- –Detect "unusual" events.
- –Automatically code utterances for types of dialogues.
- –Matching individuals, skills,training material to teams.

Applications of LSA

Individual/Team Assessment
- –Evaluating teams and individuals through measurement of communication
- –Identifying critical events
- –For training teams
  - •Suggest when trainer needs to intervene
  - •Suggest ways to improve team communication
- –Matching individuals, skills, teams, and training material

Additional Applications
- Integrated into existing cognitive models (ACT-r)
- •Automated essay scoring/ knowledge assessment embedded within training (scaffolding)
- •Automated tools for aiding collaborative learning environments
- •Within and Cross-Language filtering/retrieval/topic-detection systems
- •Lessons Learned systems

Acknowledgements

This research was completed in collaboration with the CERTT laboratory including, Nancy Cooke, Steven Shope, Preston Kiekel, and Jamie Gorman.