Automatic Modeling of Team Discourse with Latent Semantic Analysis

MC-02

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

The Problem

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

The Goal

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

Our Approach

•Use Latent Semantic Analysis, a computational model of language and semantic memory, to annotate the content of communication team tasks.

•Evaluate effectiveness of this technique for predicting team performance and decision making.

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) (Landauer, Foltz, & Laham, 1998)

➢ 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 (Landauer, Foltz, & Laham, 1998)

The Data

•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 (Kiekel, et al. 2002)

The Corpus

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: 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)

Human tagging takes approximately 1.5 hours per transcript, computer, with current algorithm, takes approximately 5 minutes.

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

Approach:

• For each utterance, find the most semantically similar utterances that have already been tagged. Assign a probability of tags to that utterance.

• Incorporate syntactic features to improve tagging accuracy.

• 2507 separate utterances coded by two human coders and automatically by computer.

Results:

TAGGER-AGREEMENT	C-VALUE	KAPPA
Human-Human	0.70	0.62
LSA-Human	0.59	0.48
LSA + Syntax-Human	0.63	0.53

• C-value of agreement (Schvaneveldt, 1990)

• Kappa statistic corrects for chance agreement (Cohen, 1960)

• Failure analysis indicates that computer has difficulty making distinctions between tags which are hard for humans to distinguish (e.g. action vs. fact)

Implications:

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

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

Simple syntactic features, such as increasing the probability that an utterance should be tagged as "uncertainty" if a "?" is present, can increase tagging reliability over LSA alone.

Experiment 2: Speech Recognition

Goal: Tag dialogue utterances in real time.

Approach: Rong and Rudnicky at CMU ran an experiment to determine baseline speech recognition accuracy on one of the transcripts in our corpus. We ran our LSA + Syntax algorithm to predict the tags on his output.

Results: Word error by speech recognition system was 38.9%. Loss of tagging accuracy, measured by C-value, was 14.9%.

Implications: Preliminary results indicate that noise introduced by current speech recognition technology may be mitigated by LSA's ability to detect semantic similarity.

Experiment 3: Team Performance

Goal: Use coded transcripts to predict team performance

Approach: Compute correlations between team performance score and tag frequencies in team-at-mission transcript.

Results: Preliminary results confirm the findings of Bowers, et al., that frequency of certain types of utterances correlate with team performance. We found correlations for tags predicted by computer:

TAG	PEARSON CORRELATION	SIG. 2-TAILED
Acknowledgement	0.335	0.006
Fact	0.320	0.008
Response	-0.321	0.008
Uncertainty	-0.460	0.000

Future work:

• Perform a similar experiment for utterances of individual team members.

• Explore patterns of sequences of tags as a predictor of performance.

Conclusions.

• Monitoring and assessing team performance is critical in distributed contexts, such as Unmanned Air Vehicle flight.

•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.

References

Bowers, C. A., Jentsch, F., Salas, E., & Braun, C. C. (1998). Analyzing communication sequences for team training needs assessment. *Human Factors, 40,* 672-679.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 34-46.

Kiekel, P. A., Cooke, N. J., Foltz, P. W., Gorman, J., & Martin, M. (2002). Some Promising Results of Communication-Based Automatic Measures of Team Cognition. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting.*

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

Schvaneveldt, R. W. (1990). *Pathfinder associative networks: Studies in knowledge organization.* Norwood, NJ: Ablex.

Acknowledgements

This research was completed in collaboration with the CERTT laboratory including, Nancy Cooke, Steven Shope, Preston Kiekel, and Jamie Gorman. This work was supported by Office of Naval Research and Army Research Laboratory grants.