Mining Chat Conversations: The Next Frontier

Dr. Sowmya Ramachandran[1], Randy Jensen[1], Oscar Bascara[1], Tamitha Carpenter[1], Todd Denning[2], Lt Shaun Sucillon[3]

[1]Stottler Henke Associates Inc.
[2]AFRL/RHA
[3]AFRL
{sowmya, jensen, bascara, tamitha}@stottlerhenke.com
todd.denning.ctr@nellis.af.mil
shaun.sucillon@wpafb.af.mil

Abstract

Analyzing chat traffic has important applications for both the military and the civilian world. This poster will report on an effort to automatically separate chat messages into topic threads.

Introduction

A significant portion of human knowledge is built and shared via conversations and dialogues. Key human activities like decision-making are mediated by conversations. It is important to be able to analyze this mode of communication so we can develop intelligent tools that will further augment our collective intelligence.

Previous related research involving multi-party dialog analysis has included work to characterize spoken interactions in multi-party meetings, social structures, and collaborative work environments. While CALO (Tur 2008; Zimmermann 2006), has a broader goal of interpreting dialogs to extract meeting minutes, the aim of our research is to separate the dialogs into different threads based on their topic. Furthermore, we focus on chat-based conversations whereas CALO analyzes spoken discussions. . Herring (2006) describes VisualDTA, a tool designed to generate a visualization of a chat conversation that has been manually coded. Our data cannot be manually coded prior to analysis.

In this poster, we will report on techniques for automatically identifying topic threads in chat-based conversations. This work is in support of the research at the Air Force Research Lab. The objective is to improve team training outcomes by developing exercise visualization and debriefing tools. As a targeted training domain, the Training Research Exercise (T-REX) conducted by the Air Force provides a controlled research environment to investigate team performance dynamics in an air and space operations center. Most of the team communication in a T-REX exercise occurs via chat. *Intelligent Diagnostic Assistant* (IDA) is a chat visualization and analysis tool to support team after-action review following a training exercise. Based on a requirement analysis, we have determined that classification of chat data according to missions (topics) is an important capability for IDA.

The problem we are addressing is:

Given: A database of chatlogs from a T-REX training session and other data logged/generated during training,

Produce: For each chat message, identify the mission to which it refers.

Chat data in this domain is fraught with abbreviations and typographical errors that present interesting challenges. Furthermore, while topics are conceptually distinct, there can be significant overlap between them in terms of references to objects, assets, tactics, etc. This make the topic identification task a challenge even for human experts.

IDA first starts out with an untagged set of chat messages sorted in a chronological order. It incrementally tags the messages with associated topics as described below. It is possible for a message to be associated with multiple topics. IDA performs multiple passes through the data to recognize associations.

- In its first pass, IDA uses unique topic identifiers to classify chat messages. Each mission or topic in this domain typical refers to a target that has a unique identifier given during the exercise. Players use these identifiers a small fraction of the time in their messages and this helps with message identification.

- In its second pass, IDA performs classification based on keywords that are correlated with the topics but are not typically unique.

- In the third pass, IDA uses temporal pattern heuristics to handle the remaining untagged message. IDA looks for a pattern of turn-by-turn interaction between two people in the same room (e.g. A says something to B and 3 minutes later B says something to A). Making an assumption of dialog coherence, IDA assigns a high degree of confidence that such conversation dyads refer to the same topic thread.

We are currently exploring alternative approaches to the second pass. Our initial approach used mission-specific keywords are specified by Subject-Matter Experts (SMEs). However, our objective is to completely automate this step.

The domain provides a related data source that can be usefully exploited. All trainees use a database system called Joint Automated Deep Operations Coordination System (JADOCS) to record critical information about the various missions, such as target intelligence, operational orders etc. A very common practice is to copy over messages from chat streams to the JADOCS database (DB) as annotations. This results in a set of chat messages stored in JADOCS with definite mission associations that can be mined to learn mission-specific identifiers. It must be noted that this set is a small percentage of the chat messages generated during an exercise.

We modified the analysis algorithm to use Naïve Bayes classifiers that were trained on known mission-message associations. To handle multiple classes, we have used a one-against-all approach, where each class has a dedicated classifier trained on a data set where the positive examples are labeled messages belonging to that class, and negative examples are labeled messages belonging to the rest of the classes. Unlabelled chat messages are classified by passing each message to each of the classifiers. A message is labeled with a topic/mission if the corresponding classifier assigns it a high probability (i.e. higher than a parameterized threshold).

Table 1 shows the results of classification of chat messages by IDA from five exercise databases. The data sets are from actual T-REX sessions. Each data set has an average of 800 chat messages. The numbers of topics in each data set range from 7 to 19. All accuracies reported in this paper are averages of the precision, recall, and F-score measures for each mission. The data sets were hand labeled with message-mission associations by an SME to provide a standard of comparison to measure accuracy.

Table 1 shows the result of applying the baseline algorithm *without* any SME-provided keywords compared with those from the classifiers trained on JADOCs data to classify messages during the second pass of the classification process. The Naïve Bayes approach leads to improvements in the overall classification accuracy. The increase in the number of false positives is more than compensated by an increase in the number of true positives that are identified.

The results indicate that a weak statistical approach is viable for this problem. Further research is necessary to understand the limits of such techniques for this domain and, if necessary, find add other domain-specific features to enhance classification accuracy.

References

Cakir, M; Xhafa, F.; Zhou, N.; and Stahl, G. 2005. Thread-based analysis of patterns of collaborative interaction in chat. Paper presented at the Conference of Artificial Intelligence for Education, Amsterdam, Netherlands.

Herring, S. C.; and Kurtz, A. J. 2006. Visualizing dynamic topic analysis. In *Proceedings of CHI 2006*. New York: ACM Press.

Langley, P. 1995. *Elements of Machine Learning. Morgan Kaufman Series in Machine Learning*. Morgan Kaufman.

Zimmermann, M.; Liu, Y.; Shriberg, E.; and Stolcke, A. 2006. Joint Segmentation and Classification of Dialog Acts in Multiparty Meetings. In Proceedings of IEEE ICASSP, Toulouse, France.

Acknowledgements

The research reported in this paper is funded under a contract with the Air Force Research Laboratory, Wright-Patterson AFB. We are grateful for this support.

	No keywords, No automated classification			No keywords, With automated classification		
Data Set	**Precision**	**Recall**	**F-Score**	**Precision**	**Recall**	**F-Score**
TREX Dataset 1	0.70	0.34	0.40	**0.76**	**0.57**	**0.58**
TREX Dataset 2	0.71	0.41	0.48	0.51	**0.69**	0.48
TREX Dataset 3	0.81	0.19	0.26	0.60	**0.45**	**0.46**
TREX Dataset 4	0.73	0.32	0.41	0.66	**0.48**	**0.51**
TREX Dataset 5	0.68	0.38	0.48	0.59	**0.54**	**0.54**
Average	**0.73**	**0.33**	**0.40**	**0.64**	**0.51**	**0.50**

Table 1. Classification Accuracy without Naïve Bayes Classifiers