Analysis of C2 and "C2-Lite" Micro-message Communications

Andrew Duchon, Robert McCormack, Brian Riordan, Charlotte Shabarekh, Shawn Weil, Ian Yohai

Aptima, Inc.
12 Gill Street Suite 1400
Woburn, MA 01801
{aduchon,sweil,rmccormack,cshabarekh,briordan,iyohai}@aptima.com

Abstract

Microtext can be considered one form of micro-messaging which can include non-digital communications which have been digitized. We present our experiences and some results analyzing micro-messages in both military command and control (C2) and more general "C2-Lite" domains in order to predict other information such as the performance of a team, the topics of discussion, or the spread of ideas. In both situations, the data can vary by the breadth of discussion, the number of participants, and the looseness of protocol. In all cases, organizing the data, combining temporal or network analysis with increasingly statistical content analyses is critical. The underlying similarities imply that analyses created for one domain can be applied to the others.

Introduction

Microtext media (Ellen, 2011), such as SMS, IM, Twitter, and text chat, have in common that they use short strings for immediate communication or broadcast. Microtext can be construed as one form of *micro-messaging* (e.g., Milstein, et al., 2008) which we extend here to include any of a number of other modalities (e.g., telephone calls, face-to-face interaction) used for short, immediate and (potentially) persistent message passing among coordinating agents. In this paper, we describe several recent attempts to study micro-messaging in military and related organizational contexts.

For the purposes of this paper, we segment our work into two broad contexts: Military C2 and Organizational "C2-Lite." C2 is defined as "the exercise of authority and direction by a properly designated commander over assigned and attached forces in the accomplishment of the mission" (Joint Chiefs of Staff, 2010). How best to implement C2 in the information age is still being debated as decisions get pushed further out to the "edge" (Alberts & Hayes, 2006) where the actions of individual warfighters can have far-ranging effects. It requires communication, information sharing, and the issuance and execution of orders, which increasingly occur as micro-messages.

However, when the military is using micro-messaging for interacting with non-governmental organizations (as in Haiti) or with local populations (e.g., intelligence gathering in Iraq), or when the organization itself is self-forming via micro-messages (as in Egypt), models of traditional C2 break down. The information sharing and influence for coordination that still occur, often via micro-messages, we call "C2-Lite."

Like others analyzing the flow and content of micro-messaging networks (e.g., Asur & Huberman, 2010; Bollen, et al., 2009; Goel, et al., 2010), our goal is not to provide any kind of "search" capability or even try to learn much about any individual message. Rather, the goal is to gather relevant messages, organize them, and extract some other kind of useful information from them, such as how well a team is performing or what people are talking about and when. However, micro-messages do not exist in a vacuum; they are contextually oriented and may be part of a larger network of communications which includes e-mail, telephone and other media, including "macro-text." Given this, we have found that natural language processing of the microtext must be paired with temporal or network analysis of the context. To demonstrate this process, we first describe some of our analyses of three unique datasets from military C2 training exercises with micro-messages of increasing size, diversity, and need for context to understand. We then describe some of our ongoing analyses of three datasets from a variety of C2-Lite situations following a similar pattern.

Micro-messaging for C2

Military C2 has come a long way from Morse code and semaphore. As networked personal computers became ubiquitous, their use in operations centers, aviation control

NOTE: The views expressed are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, or any other governmental organization.

rooms, and aboard ships became commonplace. Microtext media, such as text chat, has become the predominant communication channel in many areas. Given this use, microtext may be a ready source for analyzing the dynamics of these organizations.

Radio Brevity Codes

Our first work on micro-messaging examined radio communications from a Distributed Mission Operations (DMO) platform consisting of four high-fidelity F-16 simulators and one high-fidelity Airborne Warning and Control System (AWACS) simulator (Schreiber & Bennett, 2006). The goal of this work was to build a Support Vector Machine (SVM) using features from the communications to predict the team's performance as judged by observers with a structured set of metrics (MacMillan, et al., in press), using only information from transcriptions of the radio communications. These communications typically follow the Air Force's 3-1 communication standards (AFDC, 2006) which are designed to use very few words to convey as much information as possible.

Unlike the micro-messages we discuss in the rest of the paper, the grammar in these communications is close to being a finite state machine, making normalization and feature extraction with simple regular expressions reasonably successful. However, in common with these other types, the terms here reference a large body of knowledge and context (e.g., the location of the "bullseye"). Because communication is in broadcast form (like chat), there is typically addressing to call attention to the intended recipient. But because it is an audio channel, the sender of the message is also typically added.

We trained the SVM on 504 engagements and tested on 115, assessing performance based on the correlation of predicted the actual overall observer-based score. To add features to the model related to the flow of communications (Kiekel, et al., 2003), we created a "word" representing the speaker order. For example, if the order of speakers was: AWACS, Viper1, Viper2, Viper1; then added to the text of the last utterance would be the words V1, V2_V1, V1_V2_V1, and A_V1_V2_V1. This information typically improved performance by 50%. Another common theme in micro-message analysis is creating pseudo-documents, and in this case we created pseudo-documents from overlapping sequences of messages, and using the average scores from multiple blocks also improved performance. In the end, we were able to predict an observer's overall score of a engagement with a correlation of $r^2 = 0.45$.

Chat for Dynamic Targeting

The Air Force is progressing towards similar 3-1 communication standards for Internet Relay Chat (IRC), especially when used for Dynamic Targeting (ALSA, 2009), e.g., when the Air and Space Operations Center (AOC) must prosecute time-sensitive, emerging targets often with less than 30 minutes to coordinate intelligence, aircraft and personnel. Similar to our work in the DMO, we wanted to analyze the communications in Dynamic Targeting (DT) in order to predict observer-based measures of performance in a DT exercise (Duchon & Jackson, 2010).

We assessed a "dialogue act analysis" (an assessment of commands, questions, acknowledgements, etc.) for matching observer-based scores of individual targets. Because up to a dozen targets could be active at once, the first task was to disentangle the threads of communications about the various targets that might be active at one time.

The problem of micro-message "threading" is an active area of research looking at micro-messaging (Elsner & Charniak, 2008; Wang & Oard, 2009; Wang, et al., 2008). These clustering techniques have had some success at grouping forum messages with free-ranging topics. However, in the "tactical chat" of DT, the target, or "mission," can be used as the organizing principle. In fact, with DT chat data, a subset (but less than a quarter) of the messages related to a particular mission will explicitly mention the mission ID number (e.g., "JA0013"). As a result, we can use *semi-supervised clustering* (Kulis, et al., 2009) to take advantage of these explicit mentions in order to perform the threading function and extract all the messages about a particular mission/target.

This technique was applied to a "gold standard" of message threads created by a subject matter expert. Besides the mission IDs, a number of temporal and textual features were investigated. The grammar is more freeform, like civilian chat, so normalization and feature extraction were more difficult. However, the constraining mission information and IRC protocols which emphasize the use of addressing, enabled us to achieve 80% precision and recall. Given this threading, we applied a simple Dialogue Act Analysis along the lines of Webb et al. (2005), but hand-made, on messages about individual targets to build a multiple regression model against the observer-based performance measures. While only 12 missions were available, the results suggest it may be possible to build a real-time system that could assess DT performance mission-by-mission (Duchon & Jackson, 2010).

Multi-modal Communications for Division Operations

While the DMO has five operators, and a DT cell has about 12, an Army Division Operations Center (DOC) can have scores of operators who are in turn collaborating with potentially hundreds of others. Relating communications in this environment to assess performance on C2 brings a new set of challenges, the first of which is the matter of obtaining, organizing and synchronizing communications from multiple systems, files formats, and time zones.

For example, we are working on communications data from an Army exercise that included a division headquarters with about 100 warfighters and two brigade headquarters (one multinational) with about 40 warfighters each. In the course of two weeks, which simulated just 48 hours in "game time," there were 2800 chat messages, 3200 phone calls, and nearly 6,000 unique emails. In addition, we deployed MIT Media Lab's Sociometric badges (Olguín et al., 2009) to capture face-to-face interactions of which there were over 600 recorded (and estimated 1600 more) interactions among just the 27 members with badges in the DOC. To deal with the diversity of data, we have developed the *CommsDB* which is a database schema which aims to accommodate the vagaries of all these types of communications, while emphasizing their similarities

We are investigating measures of performance in this domain as well, in particular, how Commander's Intent (CI) is transmitted throughout an organization. CI is defined as "a leader's personal expression of an operation's end state, along with guidance on how to achieve that state" (Lewis, et al., 2000).

Measurements of shared interpretation of CI (SICI) are typically achieved through surveys or expert observation. Our goal was to exploit the information in communication networks to unobtrusively measure SICI. We are currently investigating the shortest path or "communication distance" between individuals and the commander as a proxy for how "in touch" each is with the commander's vision. Initial results suggest that distance to the commander in the face-to-face network is very related to SICI, but distance in other communication networks is not. If these results hold up under further analysis and in future exercises, it suggests that despite improvements in digital communications, physical interactions are still most important. Thus, real-time monitoring via the badges could help determine the most efficient person for the commander to speak to in order to increase SICI.

Micro-messaging for C2-Lite

More and more situations are developing where micro-messaging, while not being used directly for C2, certainly has many implications for military C2, constituting what we call "C2-Lite." These are situations where there are elements of C2 but the organization does not have command of the situation and cannot control the behaviors of the individuals involved.

In intelligence gathering, C2 stands for "*Circulation and Comprehension*." That is, what is most required is that information be shared quickly and widely, and disparate pieces of information be organized for better comprehension of the state of the world. Microtext can often appear in the semi-structured documents used for recording short observations of entities and events.

The US Military is also becoming more directly involved in non-military operations such as disaster relief and humanitarian assistance missions such as the 2010 Haiti Earthquake. While the military might desire complete C2, it must in practice work with non-governmental organizations and even individuals for relief efforts in order to *coordinate* (NGO gathers bottled water, military airdrops it) and *cooperate* (NGO provides medical assistance, military provides protection).

Finally, recent events in the Middle East have shown that when a proto-organization desires real C2, micro-messaging can be used for C2-Lite to both *enCourage* participation and *Coordinate* activities to be more effective.

All three forms of C2-Lite raise similar issues to the micro-messages in C2: who is talking about what, what is related to what, how can one see the C2 information that is embedded in a sea of hundreds, thousands, even millions of non-C2 messages? We discuss below some work we are beginning to undertake on these three areas of C2-Lite and how we are dealing with this issues.

Intelligence C2-Lite: Circulation and Comprehension

A central challenge in C2 is managing and integrating multiple information sources for the extraction of actionable intelligence. These data are often generated quickly by soldiers in the field, resembling tweets in their terseness, inconsistency, incompleteness, and lack of fixed terminology. A single entity may be referred to by name, by biometrics or other physical attributes, by the activities it performs, or by the other with which entities it associates. Even when referenced by name, there is great variability in spelling and the use of nicknames or aliases. This inconsistent reporting on entities makes "connecting the dots" challenging, because the analyst must be able to resolve disparate references to the same entity across documents. This is a pervasive challenge that appears across INT types and across domains.

The Empire Challenge 2010 (EC10) unclassified dataset embodies the inconsistencies of entity reporting described above and also exhibits sparse and contradictory observations. EC10 was a two week live action military exercise hosted by Joint Forces Commanders (JFCOM) at

Ft. Huachuca in Sierra Vista, AZ in August, 2010 during which multiple audio, imagery and human sensors were deployed to evaluate Intelligence, Surveillance and Reconnaissance (ISR) technologies. The dataset includes SALUTE reports (standing for Size, Activity, Location, Unit, Time, Equipment) which consist of unstructured text fields. The activity fields in particular contain short, one to three sentence descriptions of entities and actions, which resemble tweets. While almost all of the activity fields in the EC10 dataset were focused on entity behaviors, like the DT chat data, just 12% actually referenced an entity by name, while 32% reported generic entities (i.e. “a man”) and 33% reported entities by their physical attributes (i.e. “a mustached man carrying a rifle”). The challenge was to identify whether the generic entities and entities described by attributes were the same or related to the named entities.

To do this entity resolution and latent relationship discovery, we employed techniques from statistical relational learning, specifically the Infinite Relational Model (IRM; Kemp, et al., 2006). The IRM is a nonparametric Bayesian model that infers a system of structured categories and relationships that best fit a set of observed data. It simultaneously clusters attributes and relationships to infer latent relations between entities and between an entity and a vector of attributes. It naturally handles missing and sparse data, making it ideal for the SALUTE and other intelligence reports.

Two experiments were run using entity attributes and relations extracted from the SALUTE reports. The first experiment grouped together references to the same entity using a vector of physical appearance attributes. This produced four clusters, one of which almost exclusively contained references to the same key entity, yielding an F-Measure of 44%. In the second experiment, we employed a two-step clustering approach, augmenting the data by labeling references to the key entity identified in the first experiment, allowing us to extract more information from the sparse relationship matrix. By using the partially labeled relation data as additional inputs to the IRM, the F-Measure increased to 75%. These early results demonstrate that entities can be resolved and latent relationships can be detected in noisy, short micro-messages. The approach is also generalizable to work across any type of microtext from any intel form and domain.

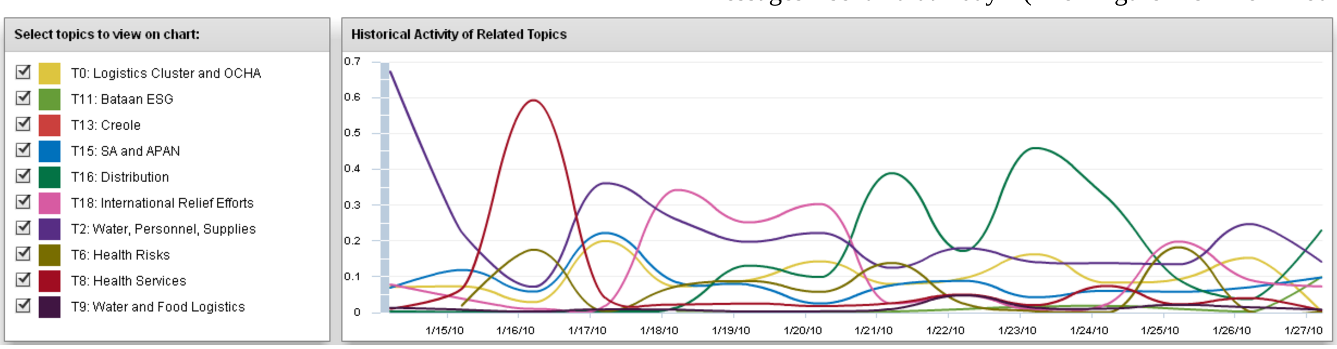
Military-NGO-Civilian C2-Lite: Coordination and Cooperation

Moving from the purely military domain, we are examining micro-messages between the military and non-governmental organizations (NGOs). To this end, United States Southern Command (SOUTHCOM), among others, is using the collaboration site, All Partners Access Network (APAN: community.apan.org) to coordinate its efforts with non-governmental organizations (e.g., the Haiti Children Project) during Humanitarian Assistance/Disaster Relief operations such as those occurring after the 2010 Haiti Earthquake. A massive international relief effort was begun almost immediately after the disaster, and so was an APAN group with forums, chat and file sharing capabilities.

We received the functional database behind the website and imported the relevant data into the CommsDB. Between January 12 and June 3, there were about 5400 forum messages sent to over 30 different forums. One of the goals in this work was to understand what was being discussed, where it was being discussed and who might be the best person to discuss something with.

To capture the “what,” we applied probabilistic topic modeling to extract topics from the set of messages. In particular, we implemented a fast, efficient, and parallelized version of Latent Dirichlet Allocation (LDA; Blei, et al., 2003) by combining insights from (Newman, et al., 2009) for parallelization methods and (Yao, et al., 2009) for sparse representations. LDA characterizes each document by its “gist” or a small set of topics that the document is about. A topic is a distribution of words, with the most probable words acting as a summary of a document which is directly interpretable by users.

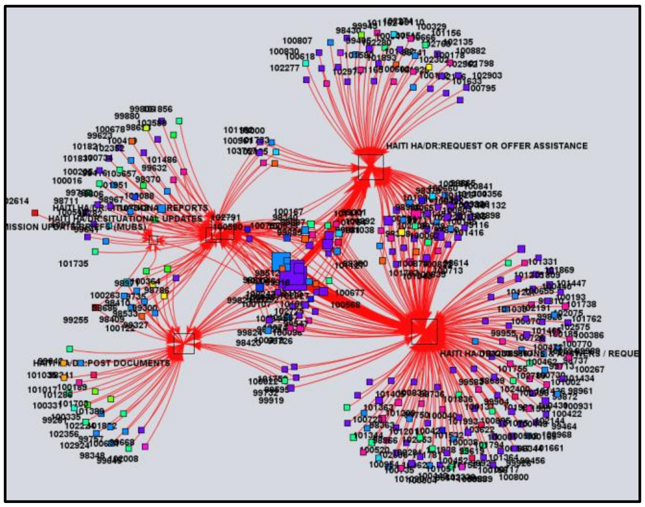
For the APAN data, the micro-messages were like those in other domains, in that the message itself was often sent merely to draw attention to referenced material, in this case a document. Thus, we represented each message as the concatenation of the message text, tags applied by the author, and text from attachments. A 20-topic model was created and each day characterized by the topics in the messages sent that day (The figure is from our



LinkingTimes application showing topics over the first two weeks.)

One of the earliest major topics (T2) in the forums concerned *Water*, *Personnel and Supplies*, basically all of the "stuff" that needed to get to the island. Soon after, this was replaced by discussion of *Distribution* (T16), that is, how to move all this "stuff" to the victims.

We can also use topics to characterize users by the types of messages they send. Below is a network representation (in our CIFTS application) of users and forums with users colored by their dominant topic. While a core set of users engage many of the forums, most users sent exactly one message, suggesting no real interaction.



The exploratory data analysis we have done so far we plan to continue for further investigation and product development. For example, do these temporal topic patterns repeat across HA/DR efforts? Could premature discussions be kindly tabled for more appropriate conditions? Could questions be brought to the attention of those most capable of answering them? Could overload be reduced by pushing information only to those most likely to be interested?

Civilian C2-Lite: Coordination and 'Couragement

The recent wave of protests and rebellions across North Africa and the Middle East highlight the quickness in which sentiment and action can spread. Dubbed "The Facebook Revolution," micro-messaging services, such as Facebook and Twitter, played an essential role in the coordination of protests and the encouragement of protesters. Understanding the spatiotemporal dynamics of information on these websites can reveal how groups informally coordinate and how ideas and sentiment spread from region to region. This, in turn, can lead to better strategies for coordinating or encouraging the spread of information.

Analyzing these micro-messages affords unique challenges in that tweets in this context are in multiple languages with no protocol and most information exists in links to web-pages are even voicemail. Much of the research on Twitter has been traditional social-network analyses, focusing on the "follower" networks. Recent work, though, has started to tackle the problems of content analysis by adapting topic modeling techniques (in particular LDA) to characterize microtext conversations (Ramage, et al., 2010; Ritter, et al., 2010). We are interested, in particular, in adapting a variant of LDA, called Dirichlet Multinomial Regression (Mimno & McCallum, 2008) to the analysis of Twitter. DMR combines text data and document metadata in a single model framework allowing inference over both topics and arbitrary features. Thus, Twitter metadata such as user locations, hash tags, and external links can naturally be modeled along with the text content.

Characterizing the content of micro-messages by topic as we did with the APAN data, may provide a basis for modeling the dynamics of information over time and across spatial networks. We are presently adapting concepts form epidemiology to model the spread of information. At a fundamental level, the dynamics of disease spread and information spread share many of the same characteristics. We have seen promising results in the application of population-level epidemiological models to topics extracted from news and blogs (McCormack & Salter, 2010). Work is also underway to analyze the spread of sentiment about topics. Through the use of DMR and other content analysis models, we believe these techniques can provide a fruitful way of characterizing the spatial and temporal dynamics of information spread in large micro-messaging data sets. However, as with the Army Division data, these communications networks would ideally be integrated with face-to-face networks (e.g., of those in Tahrir Square) to get a proper understanding of events.

Conclusions

Analysis of micro-messages in these C2 and C2-Lite contexts demand similar approaches despite the different domains. First, one needs an overall organizing principle to apply to the content of the messages; we have used targets, entities, and statistical topics. Second, one needs some representation of the flow of messages over time or through the network; we have used sequencing, timing, addressing, centrality and epidemiological models. However, as the data move from controlled and self-contained language protocols for use by a few individuals, to uncontrolled, multi-lingual, referential data used by millions, the certainty of any analysis drops, though the amount of data perhaps increases in turn. This suggests that more statistical methods can be applied, but that detailed grammatical analysis is likely to be frustrating; thus our increasing use of bag-of-feature models. In any case, the similarities in micro-messages means that we are able to apply analyses from one domain to another. For example, we might look at the spread of ideas in Command and Control, or we might be able to asses the performance of an organization in C2-Lite.

Acknowledgments

This work mentioned here was supported by L3 (PO-DS11223), AFOSR (FA9550-07-C-0134), ONR (N00014-10-M-0453, N00014-08-C-0188, N00014-09-M-0189), and ARI (W91WAW-09-C-0058). The views expressed are solely those of the authors.

References

AFDC. (2006). Air Force Tactics, Techniques, and Procedures 3-1.1 Operational Brevity Words, Definitions, and Communication Standards. Maxwell AFB, AL: Air Force Doctrine Center.

Alberts, D., & Hayes, R. (2006). Understanding Command and Control. Command and Control Research Program.

ALSA. (2009, March). IRC: Multi-Service Tactics, Techniques, and Procedures for Internet Relay Chat for Command and Control Operations (Coordinating Draft). Air Land Sea Application (ALSA) Center.

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. arXiv:1003.5699v1. Retrieved from http://arxiv.org/abs/1003.5699

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

Bollen, J., Pepe, A., & Mao, H. (2009). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011). Retrieved from http://arxiv.org/abs/0911.1583

Duchon, A., & Jackson, C. (2010). Chat analysis for after-action review. Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).

Elsner, M., & Charniak, M. (2008). You talking to me? A Corpus and Algorithm for Conversation Disentanglement. Proceedings of the 46th Annual Meeting on Association for Computational Linguistics (pp. 834-842). Columbus, OH, USA.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. Proceedings of the National Academy of Sciences, 107(41), 17486 -17490. doi:10.1073/pnas.1005962107

Joint Chiefs of Staff. (2010). Department of Defense Dictionary of Military and Associated Terms. Joint Chiefs of Staff.

Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., & Ueda, N. (2006). Learning systems of concepts with an infinite relational model. Proceedings of the 21st National Conference on Artificial Intelligence.

Kiekel, P. A., Gorman, J. C., & Cooke, N. J. (2003). Measuring speech flow of co-located and distributed command and control teams during a communication channel glitch. Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting (pp. 683-687).

Kulis, B., Basu, S., Dhillon, I., & Mooney, R. (2009). Semi-supervised graph clustering: a kernel approach. Machine Learning, 74(1), 1-22. doi:10.1007/s10994-008-5084-4

Lewis, J., Butler, C., Challans, T., Craig, D., & Smidt, J. (2000). United States Army Leadership Doctrine for the Twenty-First Century. The Human in Command, Exploring the Modern Military Experience. McCann & Pigeau (Eds.).

MacMillan, J., Entin, E., Morley, R., & Bennett, W. J. (in press). Measuring team performance in complex dynamic environments: The SPOTLITE method. Military Psychology.

McCormack, R. K., & Salter, W. J. (2010). An application of epidemiological modeling to information Diffusion. Proceedings of the International Conference on Social Computing, Behavioral Modeling, and Prediction (SBP 2010) (pp. 382-389).

Milstein, S., Chowdhury, A., Hochmuth, G., Lorica, B., & Magoulas, R. (2008). Twitter and the micro-messaging revolution: Communication, connections, and immediacy—140 characters at a time. O'Reilly Media.

Mimno, D., & McCallum, A. (2008). Topic Models Conditioned on Arbitrary Features with Dirichlet-multinomial Regression. Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI '08).

Newman, D., Asuncion, A., Smyth, P., & Welling, M. (2009). Distributed Algorithms for Topic Models. Journal of Machine Learning Research, 10, 1801–1828.

Olguín, D. O., Waber, B. N., Kim, T., Mohan, A., Ara, K., & Pentland, A. (2009). Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39(1), 43-55.

Ramage, D., Dumais, S., & Liebling, D. (2010). Characterizing Microblogs with Topic Models. ICWSM.

Ritter, A., Cherry, C., & Dolan, B. (2010). Unsupervised Modeling of Twitter Conversations. Proceedings of HLT-NAACL.

Schreiber, B. T., & Bennett, W. J. (2006). Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Summary Report (No. AFRL-HE-AZ-TR-2006-0015). Air Force Research Laboratory: Warfighter Readiness Research Division.

Wang, L., & Oard, D. (2009). Context-based Message Expansion for Disentanglement of Interleaved Text Conversations. The 2009 Annual Conference of the North American Chapter of the ACL (Vol. p., pp. 200-208).

Wang, L., Jia, Y., & Chen, Y. (2008). Conversation Extraction in Dynamic Text Message Stream. Journal of Computers, 3(10). doi:10.4304/jcp.3.10.86-93

Webb, N., Hepple, M., & Wilks, Y. (2005). Dialogue Act Classification Based on Intra-Utterance Features. Proceedings of the AAAI Workshop on Spoken Language Understanding.

Yao, L., Mimno, David, & McCallum, Andrew. (2009). Efficient Methods for Topic Model Inference on Streaming Document Collections. KDD '09: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.