Speech Act Profiling: A Probabilistic Method for Analyzing Persistent Conversations and Their Participants*

Douglas P. Twitchell
Department of MIS
University of Arizona
Tucson, AZ 85721
dtwitchell@cmi.arizona.edu

Jay F. Nunamaker, Jr.
Department of MIS
University of Arizona
Tucson, AZ 85721
jnunamaker@cmi.arizona.edu

Abstract

The increase in persistent conversations in the form of chat and instant messaging (IM) has presented new opportunities for researchers. This paper describes a method for evaluating and visualizing persistent conversations by creating a speech act profile for conversation participants using speech act theory and concepts from fuzzy logic. This method can be used either to score a participant based on possible intentions or to create a visual map of those intentions. Transcripts from the Switchboard corpus, which have been marked up with speech act labels according to a SWBD-DAMSL tag set of 42 tags, are used to train language models and a modified hidden Markov model (HMM) to obtain probabilities for each speech act type for a given sentence. Rather than choosing the speech act with the maximum probability and assigning it to the sentence, the probabilities are aggregated for each conversation participant creating a set of speech act profiles, which can be visualized as a radar graphs. Several example profiles are shown along with possible interpretations. The profiles can be used as an overall picture of a conversation, and may be useful in various analyses of persistent conversations including information retrieval, deception detection, and online technical support monitoring.

1. Introduction

With the increasing use of computer-mediated communication (CMC) tools such as chat, instant messaging, and e-mail, persistent conversations are becoming more common and can be useful for studying human behavior. However, automated tools for studying human behavior in persistent conversations are rare. This paper proposes an analytical tool based on speech act theory, corpus linguistics, and concepts from fuzzy logic for analyzing and visualizing behavior in synchronous persistent conversations such as chat and instant messaging. The method proposed is called speech act profiling and is a combination of fuzzy typing and speech act modeling or classification. It provides a way for profiling conversations and their participants using fuzzy typing to retain all of the information in the speech act classification model.

Speech act profiling is intended for synchronous or almost synchronous CMC such as chat and instant messaging with an aim at making searching through large amounts of conversational data easier than searching by keywords and reading whole conversations. Also presented is a visualization scheme for conversations that emphasizes the nature of each participant's conversation as a whole and allows the analyst to get an overall view of the interaction between the participants. The method is based on the work of Stolcke et. al [20] on dialog act modeling and Subasic and Huettner's work on fuzzy typing [21]. Dialog act modeling utilizes n-gram language modeling and hidden Markov models to classify conversational utterances into 42 dialog act categories. Using the concept of fuzzy typing, speech act profiling takes the probabilities (not the classifications) created by the combination of the language model and the hidden Markov model and sums them for the entire conversation. The resulting probabilities are an estimate of the number of each of the dialog acts uttered by the participants during the conversation. The probabilities for each participant can be separated and displayed on a radial graph, which is organized according to Searle's [17] taxonomy of speech acts, giving the analyst an overall view of the conversation.

There are several reasons for attempting automatic clas-

*Portions of this research were supported by funding from the U. S. Air Force Office of Scientific Research under the U. S. Department of Defense University Research Initiative (Grant #F49620-01-1-0394). The views, opinions, and/or findings in this report are those of the authors and should not be construed as an official Department of Defense position, policy, or decision.

sification of speech acts and creating a speech act profile. First, deception, an act intended to create a false impression in a hearer, can be made up of a number of speech acts each with its own intention. For example, because deceivers are loath to be caught in their deception, they will often put on a submissive front. The use of fewer statements and more agreements could indicate submissiveness. Second, information retrieval could benefit from the automated classification of speech acts. Large databases of chat data from sources such as online games or virtual meetings could be searched for angry or agitated interactions in a content independent manner. When combined with the ability to search for specific users, the speech act profile could be used to search for specific users in specific interactions. For example, many technology companies offer real-time online technical support in the form of chat interactions. A manager might want to look for conversations in which a support technician is behaving in a particular way. Or, the same manager might want to analyze the profile of technicians who are performing well for patterns of conversational interaction and behavior without reading thousands of conversations. Recently, the National Association of Securities Dealers (NASD) issued a Notice to Members that clarifies the responsibility of securities firms to store instant messaging conversations for three years [13]. As the use of instant messaging in the finance industry becomes more commonplace, the need for software to filter through the mountains of resultant data will increase. If a broker were to come under suspicion by firm managers, speech act profiling could be used to quickly sift through the broker's conversations to look for conversations that don't fit his or her normal pattern. Finally, with some decline in accuracy, several conversations could be monitored simultaneously in real time. These illustrative examples represent potential application areas where speech act profiling could be applied. Of course, in each area empirical evidence would be required to demonstrate the method's effectiveness.

2. Persistent Conversation

By its nature, CMC is persistent. That is, once uttered, CMC utterances can be (and in the case of securities brokers are required by law to be) saved to some storage medium. Once saved, they can then be, "searched, browsed, replayed, annotated, visualized, restructured, and recontextualized" [6] Chat rooms and instant messaging systems are continually passing millions of messages, and many of the resulting conversations are persisted indefinitely. This large corpus of messages provides a potentially valuable resource for researchers of human language, communication, and behavior. However, the continually-growing corpus represents a challenge to researchers by its mere size, which provides statistical power, but also a lengthy and costly process of coding the behavior in the conversations. Because of the immensity of coding large corpora, an automated process would be useful.

CMC, especially text-based CMC, has a number of qualities that make it unique from face-to-face (FTF) conversations. First, unlike FTF, participants in text-based CMC may create utterances simultaneously, removing the necessity of strict synchronous turn-taking. Though turn-taking is not necessary, it is often still practiced in chat and instant messaging since participants must read and formulate responses to each other [11, 5]. Second, typing messages can be cognitively more difficult than speaking. This difficulty motivates participants to use abbreviations for common phrases (such as ROTFL for "roll on the floor laughing") and shortened spellings of words ("r u" instead of "are you" and "4" instead of "for"). The fact that the conversation is persisted and the fact that throughout the conversation participants can refer back to previous utterances reduces the cognitive and short-term memory load on the participants. Finally, the knowledge that their communication may persist forever may cause participants to be more cautious in their language.

Despite these differences, CMC still shares the underpinnings of everyday language with FTF communication. Participants are still likely to conform to most grammar and spelling rules and even some punctuation rules. Furthermore, though abbreviations are used for many words, the words themselves still mostly have the same meaning and function that they do in FTF. English CMC and FTF communication are still likely to conform to most of the same word usage and grammar rules and norms. Indeed, it could be said that synchronous CMC such as IM and chat are more like FTF than the more commonly used email. Researchers should be able to use many of the same methods to study CMC as they do the verbal aspects of FTF communication.

2.1. Speech Act Theory

Speech act theory was first introduced by Austin [3] in the 1960s. Austin's method for analyzing speech acts splits any utterance into three different acts: the locutionary act, the illocutionary act, and the perlocutionary act. The locutionary act is the act of uttering a sentence with a particular meaning. The illocutionary act is what the speaker *does* when uttering a sentence (e.g. promising, requesting, declaring, etc.), and the perlocutionary act is the creation (intentional or not) of some effect in the hearer [7]. Illocutionary acts, the speech acts we are interested in and the acts that express intent, are said to each have an illocutionary force f and a propositional content p with the entire act labeled as $f(p)$. While force and content are related, this study focuses on the force or intent of an utterance. Illocutionary acts are sometimes referred to as speech acts, but may also be called dialogue acts specifically when referring to acts in dialogue rather than discourse. This paper refers to these acts mostly as speech acts but refers to the same acts when using the terms illocutionary act or dialogue act.

2.2. Speech Act Potential

Alston [2] suggests that every utterance has a *speech act potential*. That is, for every utterance, there is a set of possible speech acts that the utterance could realize, including acts that are non-literal and/or indirect. Some of the potential acts are derived from the semantic meaning of the utterance. For example, someone hearing the sentence "I promise to be there" knows that the speaker was promising because the speaker said, "I promise." Other acts are derived from the sentence structure. "Are you going to be there?" is a question because the auxiliary verb has been moved to the front and a question mark has been added to the end. However, note that the sentence, "Are you going to be there?" could be an order rather than a question if it were uttered by an army general to a private in the right intonation. Thus, this sentence has at least these two speech act potentials and maybe more. Speech act profiling captures the idea of illocutionary act potential by retaining all potential speech acts and their probabilities.

2.3. Speech Act Typing

In recent years researchers in the computational linguistics community have made an effort to automate speech act classification and, in doing so, have created a speech act classification and tagging system called DAMSL (Dialogue Act Markup in Several Layers) [1]. Even though speech acts are found in both dialogue and discourse, DAMSL is used primarily for marking up dialogue. In a large effort Stolcke et al. [20] marked up the 198,000 utterance Switchboard corpus of spontaneous human-to-human telephone dialogue using a slightly modified DAMSL tag set of 42 shallow dialogue acts called SWBD-DAMSL (shown in Table 1). A full description of each act and how the acts were labeled can be found in [8].

In their method of speech act typing, Stolcke et al. [20] use the words and their order in sentences as evidence for creating a n-gram language model. This model assumes that the dialogue act of a given utterance is dependent not only on the words contained in the sentence but also the dialogue act classification of the previous two sentences. This idea is expressed in Equation 1 where s_t represents a specific dialog act at time t, and o_t represents the words in the utterance at time t.

$$P(s_t|s_{t-1}, s_{t-2}, o_t) \tag{1}$$

To account for word sequences and speech act sequences that may never be encountered in the training set, Katz back-off [9] and Witten-Bell discounting [23] are used as smoothing algorithms. With the speech act labels as node labels in the HMM, the Forward-Backward algorithm for HMMs is used to obtain the most likely sequence of dialogue acts in the set of utterances. The sequence of dialogue acts is matched to the sequence of utterances to classify each utterance as a dialogue act. With words alone, this method can achieve a classification accuracy of 71% compared to a baseline accuracy of 35% (35% of the testing corpus was of a single dialogue act category—STATEMENT) and a human accuracy of 84% [20].

2.4. Speech Act Categorization

Because of the unwieldy nature of 42 separate speech acts, it is useful to group the speech acts into speech act categories. Searle [17] provides a taxonomy of speech acts that is suited to this purpose. In his taxonomy Searle categorizes speech acts according to what he calls their illocutionary point or purpose, their direction of fit to the world, their expressed psychological state, and their propositional content. The five major categories introduced by Searle include:

Assertives: The speaker commits herself to something being true. *The sky is blue.*

Directives: The speaker attempts to get the hearer to do something. *Clean your room!*

Commissives: The speaker commits to some future course of action. *I will do it.*

Expressives: The speaker expresses some psychological state. *I'm sorry.*

Declarations: The speaker brings about a different state of the world. *The meeting is adjourned.*

The SWBD-DAMSL tag set can be mapped to Searle's taxonomy. For example, all of the question tags such as YES-NO-QUESTION and OPEN-QUESTION can be placed under directives since the speaker is attempting to get the hearer to provide the speaker with information, and OPINION, APOLOGY, and THANKING can be categorized as expressives since they each express some psychological state. BACKCHANNEL, AGREEMENT, and ACKNOWLEDGEMENT are also expressives; they express the psychological states of desiring the other participant to continue and agreement with them. The remaining tags are similarly categorized. It is unclear where some tags such as OTHER and NON-VERBAL should be mapped. These tags were placed in a category called *other* that does not exist in Searle's taxonomy, but is acceptable for the purposes of this study. The entire mapping is shown in Table 2.

Table 1. The 42 SWBD-DAMSL dialogue acts (Adapted from [8]).

Tag Name	Tag	Example
STATEMENT-NON-OPINION	sd	Me, I'm in the legal department.
ACKNOWLEDGE (BACKCHANNEL)	b	Uh-huh.
STATEMENT-OPINION	sv	I think it's great.
AGREE/ACCEPT	aa	That's exactly it.
ABANDONED, TURN-EXIT OR UNINTERPRETABLE	%	So,-
APPRECIATION	ba	I can imagine.
YES-NO-QUESTION	qy	Do you have to have any special training?
NON-VERBAL	x	[Laughter], [Throat-clearing]
YES ANSWERS	ny	Yes.
CONVENTIONAL-CLOSING	fc	Well, it's been nice talking to you.
WH-QUESTION	qw	Well, how old are you?
NO ANSWERS	nn	No.
RESPONSE ACKNOWLEDGEMENT	bk	Oh, okay.
HEDGE	h	I don't know if I'm making any sense or not.
DECLARATIVE YES-NO-QUESTION	qyCd	So you can afford to get a house?
OTHER	other	Well give me a break, you know.
BACKCHANNEL IN QUESTION FORM	bh	Is that right?
QUOTATION	Cq	You can't be pregnant and have cats.
SUMMARIZE/REFORMULATE	bf	Oh, you mean you switched schools for the kids.
AFFIRMATIVE NON-YES ANSWERS	na	It is.
ACTION-DIRECTIVE	ad	Why don't you go first
COLLABORATIVE COMPLETION	C2	Who aren't contributing.
REPEAT-PHRASE	bCm	Oh, fajitas
OPEN-QUESTION	qo	How about you?
RHETORICAL-QUESTIONS	qh	Who would steal a newspaper?
HOLD BEFORE ANSWER/AGREEMENT	Ch	I'm drawing a blank.
REJECT	ar	Well, no
NEGATIVE NON-NO ANSWERS	ng	Uh, not a whole lot.
SIGNAL-NON-UNDERSTANDING	br	Excuse me?
OTHER ANSWERS	no	I don't know
CONVENTIONAL-OPENING	fp	How are you?
OR-CLAUSE	qrr	or is it more of a company?
DISPREFERRED ANSWERS	arp	Well, not so much that.
3RD-PARTY-TALK	t3	My goodness, Diane, get down from there.
OFFERS, OPTIONS COMMITS	commits	I'll have to check that out
SELF-TALK	t1	What's the word I'm looking for
DOWNPLAYER	bd	That's all right.
MAYBE/ACCEPT-PART	aap	Something like that
TAG-QUESTION	Cg	Right?
DECLARATIVE WH-QUESTION	qwCd	You are what kind of buff?
APOLOGY	fa	I'm sorry.
THANKING	ft	Hey thanks a lot

Table 2. Categories of SWBD-DAMSL tags based on Searle's taxonomy [17]

Assertives	Expressives	Directives
Statement	Opinion	Yes-No-Question
Yes Answers	Abandoned/Uninterpretable	Wh-Question
No Answers	Backchannel/Acknowledge	Declarative Yes-No-Question
Quotation	Response Acknowledgement	Backchannel-Question
Affirmative non-yes answers	Signal-non-understanding	Summarize/Reformulate
Collaborative Completion	Agreement/Accept	Action-Directive
Rhetorical-Questions	Appreciation	Open-Question
Negative non-no answers	Conventional-closing	Tag-Question
Other Answers	Hedge	Declarative Wh-Question
Or-clause	Hold before answer/agreement	
Dispreferred answers	Reject	Other
	Conventional-opening	Other
Commissives	Downplayer	Third-Party Talk
Offers, Options, & Commits	Maybe/Accept-part	NonVerbal
	Apology	
Declarations[1]	Thanking	
	Repeat-Phrase	

[1]None of the 42 dialogue acts could be described as a declaration

2.5. Other Related Work

Concepts of fuzzy logic and speech act theory have intersected once before. Kim, et. al. [10] introduce a method of using fuzzy logic and neural networks to classify speech acts. Their method was more successful than trigram models over sentence features when a small training set was used. There has yet to be a comparison between their method and Stolcke et. al.'s [20] method, which combines trigram *language* (not sentence feature) models and a HMM. The HMM in the combined method, however, provides a fairly simple mechanism for extracting probabilities, which is useful for the speech act profile.

3. Methodology

3.1. Fuzzy Typing

One attribute of the approach to speech act classification used by Stolcke et. al. [20] is that with each utterance the classification is either all right or all wrong. That is, when the method classifies incorrectly, the correct classification is discarded, even if the correct classification would have been the second choice. For example, if dialog act modeling classifies a particular utterance as a Statement when in actuality it is an Opinion, the user may never know if Opinion was the classifier's second choice (For example, the probability of the utterance being a Statement is .43 and the probability of it being an Opinion is .40). The user is only informed that the classifier has chosen Statement. One approach to this problem is the use of n-best lists [16, 15], which are lists of the n top possible choices for a classification problem. The n-best lists can be used by humans or a second stage classifier with more evidence to choose the final classification. Another approach is fuzzy typing proposed by Subasic and Huettner [21] who address the problem of multiple potential types for a single item in their research on multiple potential meanings for affect (emotion) words. In their system, rather than attempting to completely disambiguate an affect word, they use fuzzy logic to retain all of the possible meanings (or at least the most likely ones) and their probabilities of occurring. While they still don't know the correct sense of the word they were trying to disambiguate, those times when they would have incorrectly disambiguated the word, they retain the correct information. This method allows the authors to create a fairly accurate affect profile of a document, which they graph to create a visual affect profile.

The same method can be applied to speech act classification. As described in Section 2.2, a single sentence token may have multiple illocutionary act types. Since it is difficult for classifiers to decide which act to assign ambiguous sentences, speech act profiling instead captures all potential act types and their probabilities of occurring.

3.2. Speech Act Profiling

Speech act profiling includes two major steps: (1) Determining speech action potential probabilities; and (2) Creating a speech act profile.

3.3. Determining Speech Act Potential Probabilities

Two possible ways of calculating the probabilities associated with each potential illocutionary act include: (1) using knowledge of the language and (2) counting speech acts in a marked up corpus. In their method of fuzzy semantic typing, Subasic and Huettner [21] employ a linguist to create a fuzzy lexicon which contains all of the words of interest along with the probabilities of the words belonging to a set of affect categories. The method presented in this paper, however, follows the corpus-based approach used by Stolcke et al [20].

Unlike the Stolcke et al.'s method [20], which sought the one most likely classification for an utterance, speech act profiling seeks all of the potential classifications and their probabilities. These probabilities could be obtained by counting the number of times the utterance of interest occurs as each speech act in a large tagged corpus such as the Switchboard corpus; however, since it would be unlikely for a given sentence to be repeated enough times in a corpus to produce a reliable probability distribution, an n-gram language model [12] is built for each of the speech acts. The n-gram language model obtains the probability of a given sentence by chopping it up into small sequences of words or n-grams. For example, the sentence *Are you going to be there?* could be represented as the following set of bigrams (2-grams):

Are you

you going

going to

to be

be there

there ?

The probability of these n-grams is multiplied together to obtain the probability of the entire sentence. Despite this technique, there are still n-grams that are not found in the training set even when the training set is very large. To overcome this problem, Katz back-off smoothing [9] with Witten-Bell discounting [23] are used. These techniques reserve some probability for future unseen n-grams.

The probabilities for a sentence calculated by the language model are given to a HMM as symbol emission probabilities, which determines the final classification. However, speech act profiling is not concerned with a single classification only; instead it requires a set of possible classifications and the probabilities of those classifications occurring. For a given utterance, a list of probabilities can be derived from the HMM using the Forward-Backward algorithm. A HMM is represented by three probability matrices: the state transition probabilities, A; the symbol emission probabilities, B; and the initial state probabilities, Π. Since the states in the HMM represent the speech acts, the state transition probability a_{ij} represents the probability of on speech act following another speech act and is obtained from a trigram model of the speech acts themselves. Each symbol emission probability b_{ijo_t} is the probability of a given sentence being a certain speech act. As explained earlier, these probabilities are derived by calculating the probability of a given sentence being in each of the speech act language models. Finally, the initial state probability π_i represent the probability of each speech act i being the first speech act in a conversation. Given a sequence of utterances, the Forward procedure calculates the forward probability $\alpha_i(t)$ of each utterance at time t being in state i (where N is the number of states). The forward probabilities are combined with the backward probabilities $\beta_i(t)$ to create a more accurate probability $\gamma_i(t)$ that the HMM is in state i at time (or utterance) t as show in Equation 2. [14].

$$\gamma_i(t) = \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^{N} \alpha_j(t)\beta_j(t)} \tag{2}$$

For a given utterance t there is a $\gamma_i(t)$ for each state i, therefore the Γ matrix has a row for each utterance, and a column for each potential speech act represented by the states. The full Γ matrix is used to create a speech act profile.

At this point, instead of using the Γ matrix, a n-best list of possible dialog acts could be generated using a Viterbi decoder [16]. Several algorithms for generating n-best hypotheses lists exist, both exact and approximate [15]. Stolcke et. al., however, found in their empirical comparisons that forward-backward decoding was consistently slightly more accurate. That the greater accuracy of the forward-backward algorithm over Viterbi n-best hypotheses lists holds for speech act profiling is a question that should be addressed empirically. The Γ matrix from the forward backward algorithm was used for all of the examples introduced in this paper,

3.4. Creating a Speech Act Profile

A speech act profile for a set of utterances includes all of the potential speech acts found in those utterances along with their probabilities. A useful way to group the probabilities is by speaker. To create profiles based on individuals, we first calculate the average of each speech act for each speaker. Let $\gamma_i^A(t)$ represent probability of the speech act of type i spoken by Speaker A at time t, and let $\delta^A(t)$ be a binary variable that is 1 when A is the speaker at time t and 0 otherwise. The averages (avg) for Speaker A are computed as follows:

$$avg(\gamma_i^A) = \frac{\sum_{t=1}^{T} \gamma_i^A(t)\delta^A(t)}{\sum_{i=1}^{N}\sum_{t=1}^{T} \gamma_i^A(t)\delta^A(t)} \qquad 1 \leq i \leq N \tag{3}$$

We calculate the average for each speaker, then we subtract the average occurrence of each speech act in the training set to obtain a measure of how divergent (dvg) the current speaker is from the average speaker in the training set. So, for Speaker A:

$$dvg(\gamma_i^A) = avg(\gamma_i^A) - \frac{|i_{training}|}{\sum_{i=1}^{N} |i_{training}|} \qquad 1 \leq i \leq N \tag{4}$$

With the speech act divergences and the categories of dialogue acts defined, a visual representation of the speech act profile can be created. Like Subasic and Huettner's [21] visualization scheme, the speech act profile visualization consists of a radar graph with all of the potential dialogue acts and their divergences as shown in Figure 1. The probabilities for each speech act have been normalized to 0 according to Equation 4 to allow for easy comparison against the overall distribution of dialogue acts in the Switchboard corpus.

4. Examples and Discussion

The graph in Figure 1 shows a profile of a single conversation between two participants in the Switchboard corpus. The abbreviations around the outside of the graph refer to the speech act tags found in Table 1, and these tags are grouped according to Searle's taxonomy as shown in Table 2. The profile shows that Speaker A used many more potential STATEMENTS (sd) and more OPINIONS (sv) than were used in the training set. Speaker B, on the other hand, had fewer STATEMENTS (sd) and a lot more backchannels (b). This profile indicates that Speaker A dominated the conversation with statements while Speaker B simply acknowledged. Reading the conversation confirms that Speaker A was dominant even though Speaker B produced more utterances (92 vs. 118).

The profile in Figure 2 shows a different conversation with different participants. Here Speaker B is questioning Speaker A as indicated by the greater than normal number of WH-QUESTIONS (qw) and YES-NO-QUESTIONS (qy) by Speaker B and the greater than normal number of STATEMENTS (sd) by Speaker A. In this conversation, Speaker B is directing the conversation, and Speaker A is merely replying in an interview-like manner. Table 3 is an excerpt of the conversation profiled in Figure 2 (33 of the 311 utterances in the conversation are shown). In this excerpt, Speaker B is asking numerous questions and Speaker A is responding to the questions with statements. Other portions of the same conversation follow a similar pattern.

The profile in Figure 3, however, is more difficult to interpret. In this profile, the two participants seem to have the same number of almost every type of speech act indicating a balanced conversation, which could be explained by the tendency for participants in conversations to mirror

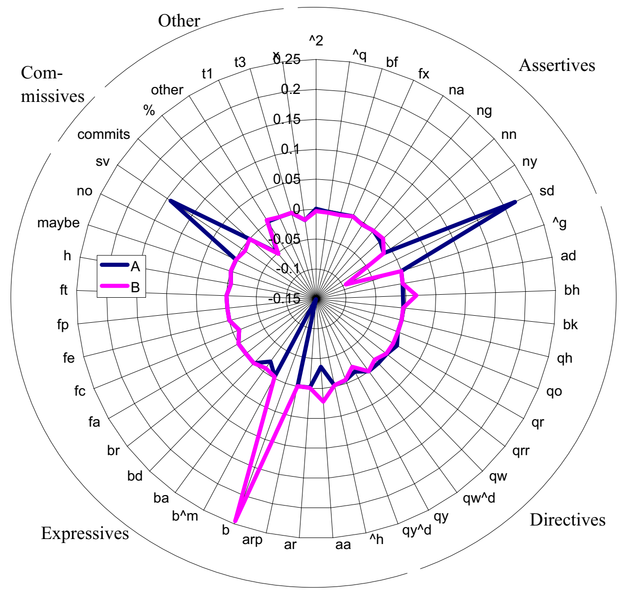

Figure 1. Sample fuzzy speech act profile showing domination

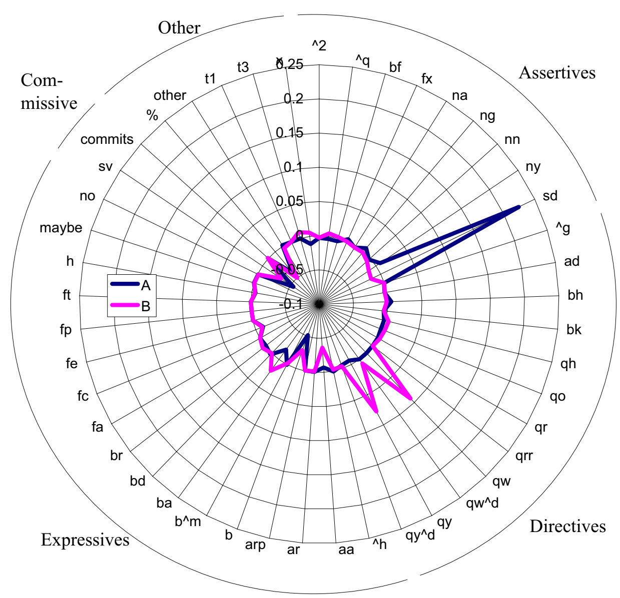

Figure 2. Sample fuzzy speech act profile showing interview-like behavior

Table 3. Excerpt of conversation represented by the speech act profile in Figure 2

Speaker	Act	Utterance
B	qw	Where are you going to school?
A	sd	N C State.
B	qw	What's that?
A	sd	Uh, North Carolina State.
B	qy	So you're on Spring Break?
A	ng	Not yet.
A	sd	Ours don't start until, uh, next week.
B	qw	So where are you?
A	br	Where am I?
B	ny	Yeah.
A	qw	What do you mean, where?
A	bk	Oh,
A	sd	in Raleigh.
B	bk	Oh, okay.
A	%	Little, -
A	sd	the burgeoning metropolis of Raleigh.
A	sd	What a dead place [laughter].
B	qy	Is it? [laughter].
A	ny	Yeah.
A	sd	[laughter] I've got a nice little business at home,
A	sd	[talking] and I sit around and tinker with that most of the time.
B	qw	So what are you getting your degree in?
A	sd	[Talking] Um, uh, human factors.
B	qw	And what do you do with it?
A	sd	Not a thing [laughter].
B	qw	Well, what is it?
A	sd	It's really looking at systems and design systems and seeing how people interact with them.
B	bf	So it's sociology.
A	sd	It's more psychology and engineering.
A	%	Uh,
A	sd	My master's is in industrial engineering.
B	qy	So you're working on your doctorate?
A	ny	Yeah.

or reciprocate their partner. There are several reasons that this type of behavior commonly occurs, including biological, social, and strategic motivations [4].

There are numerous potential applications of speech act profiles. If the conversations above were between technical support personnel and customers, a manager might want to take a closer look at the conversation in Figure 1 and use the conversations in Figure 2 and Figure 3 as examples. The manager could use the fuzzy speech act profiles as an efficient method of finding positive and negative examples of technical support encounters. In the financial services industry, where IM is enjoying increasing use, speech act profiles could be combined with data mining to find patterns of conversations that lead to higher sales or deceptive practices. With some loss of accuracy (because backward probabilities would be unavailable), a manager could monitor several conversations in real time without having to read each one. When a conversation's profile seemed out of the ordinary, the manager could investigate further. Finally, speech act profiling and keyword search techniques could be combined to help law enforcement agencies increase their abiltiy and coverage in identifying potential sexual predators in chat rooms.

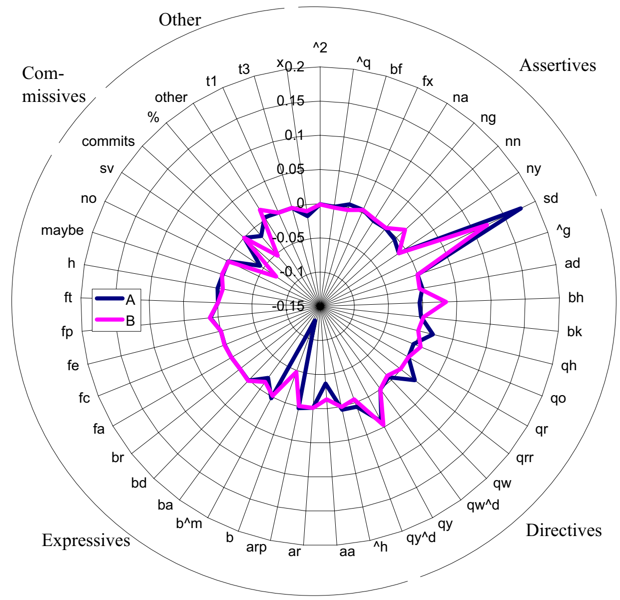

Figure 3. Sample fuzzy speech act profile showing reciporical behavior

5. Conclusion

The examples provided in this paper are based on the SwitchBoard corpus of telephone conversations, but the method presented should also be useful in synchronous and almost synchronous CMC such as chat or instant messaging. Although it would be desirable to train and test the model on a CMC corpus, we could find no CMC corpus of comparable size and type. While spoken telephone conversations and written conversations share some similarities, they also contain many differences. For example, the Switchboard corpus contains a high percentage of backchannel utterances such as *uh-huh*, which aren't as likely to be contained in written conversation. Such differences will likely introduce error when training the speech act typing method on the spoken Switchboard corpus and using for profiling CMC such as IM and chat. Despite the increase in error, CMC and spoken language are likely similar enough for speech act typing to be useful. Moreover, the dialog act modeling, on which speech act profiling is based, obtains much of its classification power from the words that appear in the utterances rather than the structure of those utterances, and many of the "key" words for the dialog acts are the same for spoken language and CMC. Therefore, it would be fruitful to use speech act typing trained on the large Switchboard corpus to profile CMC, and such studies are currently being undertaken by the authors.

More recent work in dialog act modeling has introduced prosodic (duration, pauses, pitch, energy, etc) features as further evidence for classifying speech acts in dialogs [18, 22]. These prosodic features, which were useful in disambiguating between dialog act categories such as BACKCHANNEL and AGREEMENT, are unavailable in text-based CMC. Because the audio channel is unavailable, participants in CMC may attempt to compensate by transmitting the information that would have been conveyed via the audio channel through the text channel. For example, when speaking a declarative questions such as *You went to Paris?*, a speaker will use rising intonation to indicate to the hearer that the utterance is a question. In CMC, the sender of the same message would likely just use a question mark, something that is unavailable in the audio channel. Speech act profiling of CMC looses the additional evidence provided by prosodic features, but gains other evidence in the form of punctuation and compensation by the sender of CMC messages.

Studies should also be done to ensure that speech act profiling is valid. Such studies could employ human coders to give their impression of a conversation and compare that impression to the results returned by a speech act profile. Further studies could test the usefulness of speech act profiling in fields such as information retrieval and deception detection.

Speech act profiling creates new way to visualize and analyze persistent conversations. It is significant because it combines the idea of fuzzy typing with an established method of speech act classification providing a way to capture all of the information in the speech act model. The example profiles illustrated show how insight into conversations and their participants can be gained. Profiles such as these could be used in several interesting areas including intent-based information retrieval, deception detection, and online customer interactions. Speech act profiling as it is presented here is in its first stages and needs refinement, but still represents a potentially valuable automated tool for analyzing and visualizing persistent conversations.

Acknowledgements

The authors thank SRI International and Hyoungsoo Yoon for the allowing public use of the SRI Language Modeling Toolkit [19] and the HMMLib package [24], respectively. Both were used extensively in this study.

References

[1] J. Allen and M. Core. Draft of DAMSL: Dialog act markup in serveral layers. http://www.cs.rochester.edu/research/cisd/resources/damsl/revisedmanual/revisedmanual.html, 1997.

[2] W. P. Alston. *Illocutionary Acts and Sentence Meaning*. Cornell University Press, Ithaca, N.Y., 2000.

[3] J. L. Austin. Locutionary, illocutionary, perlocutionary. In *How to do things with words*. Harvard, Cambridge, MA, 1962.

[4] J. K. Burgoon, L. A. Stern, and L. Dillman. *Interpersonal Adaptation: Dyadic Interaction Patterns*. Cambridge University Press, Cambridge, UK, 1995.

[5] S. Condon and C. G. Čech. Profiling turns in interaction: Discourse structure and function. In *Proceedings of the thirty-fourth Hawaii International Conference on System Sciences (CD/ROM)*, Maui, Hawaii, 2001. IEEE Computer Society Press.

[6] T. Erickson. Persistant conversation: An introduction. *Journal of Computer Mediated Communication*, 4(4), 1999.

[7] D. Jurafsky and J. H. Martin. *Speech and Language Processing : An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall Series in Artificial Intelligence. Prentice Hall, Upper Saddle River, N.J., 2000.

[8] D. Jurafsky, E. Shriberg, and D. Biasca. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual, Draft 13, 1997.

[9] S. M. Katz. Estimation of the probabilities from sparse data for the language model component of a speech recognizer. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(3):400 – 401, 1987.

[10] H. Kim, J.-M. Cho, and J. Seo. Fuzzy trigram model for speech act analysis of utterances in dialogues. In *IEEE International Fuzzy Systems Conference*, volume II, pages 598–602, Seoul, Korea, 1999. IEEE.

[11] M. van der Wege and G. Clark. Turn-taking systems for computer-mediated communication. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, page 1077. Lawrence Erlbaum, 1997.

[12] C. D. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts, 2002.

[13] NASD. Instant messaging: Clarification for members regarding supervisory obligations and recordkeeping requirements for instant messaging. http://www.nasdr.com/pdf-text/0333ntm.pdf, 2003.

[14] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257 – 286, 1989.

[15] R. Schwartz and S. Austin. A comparison of several approximate algorithms for finding multiple (n-best) sentence hypotheses. In *International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 701–704, Toronto, Ontario, Canada, 1991. IEEE.

[16] R. Schwartz and Y.-L. Chow. The n-best algorithm: An efficient and exact proceedure for finding the n most likely sentence hypotheses. In *International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 81–84, Albuquerque, New Mexico, 1990. IEEE.

[17] J. R. Searle. A taxonomy of illocutionary acts. In *Expression and Meaning: Studies in the Theory of Speech Acts*, pages 1–29. Cambridge University Press, Cambridge, UK, 1979.

[18] E. Shriberg, R. Bates, P. Taylor, A. Stolcke, D. Jurafsky, K. Ries, N. Coccaro, R. Martin, M. Meteer, and C. Van Ess-Dykema. Can prosody aid the automatic classification of dialog acts in conversational speech. *Language and Speech*, 34(3-4):439–487, 1998.

[19] A. Stolcke. SRILM - an extensible language modeling toolkit. In *Proceedings of the International Conference on Spoken Language Processing*, Denver, Colorado, 2002.

[20] A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, C. Van Ess-Dykema, R. Martin, and M. Meteer. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–373, 2000.

[21] P. Subasic and A. Huettner. Affect analysis of text using fuzzy semantic typing. *IEEE Transactions on Fuzzy Systems*, 9(4):483–496, 2001.

[22] A. Venkataraman, L. Ferrer, A. Stolcke, and E. Shriberg. Training a prosody-based dialog act tagger from unlabeled data. In *International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 272–275, Hong Kong, 2003. IEEE.

[23] I. H. Witten and T. C. Bell. The zero-freqency problem: Estimating the probabilities of novel events in adaptive text compression. *IEEE Transactions on Information Theory*, 37(4):1085–1091, 1991.

[24] H. Yoon. Hmmlib. http://www.vilab.com/hmmlib/home.html, 1999.