Using Sequence Package Analysis as a New Natural Language Understanding Method for Mining Government Recordings of Terror Suspects

Amy Neustein

Abstract. Three years after 9/11, the Justice Department made the astounding revelation that more than 120,000 hours of potentially valuable terrorism-related recordings had yet to be transcribed. Clearly, the government's efforts to obtain such recordings have continued. Yet there is no evidence that the contents of the recorded calls have been analyzed any more efficiently. Perhaps analysis by conventional means would be of limited value in any event. After all, terror suspects tend to avoid words that might alarm intelligence agents, thus "outsmarting" conventional mining programs, which heavily rely on word-spotting techniques. One solution is the application of a new natural language understanding method, known as Sequence Package Analysis, which can transcend the limitations of basic parsing methods by mapping out the generic conversational sequence patterns found in the dialog. The purpose of this paper is show how this new method can efficiently mine a large volume of government recordings of the conversations of terror suspects – with the goal of reducing the backlog of unanalyzed calls.

1 Introduction

In December 2005, officials at the National Security Agency anonymously leaked to the press that, since the September 11th attacks, "the volume of information harvested from telecommunication data and voice networks, without court-approved warrants, is much larger than the White House has acknowledged" [1]. Ironically, a year earlier, the *New York Times* gave front-page coverage to an astounding report issued by the Justice Department's inspector general. The report revealed that "more than 120,000 hours of potentially valuable terrorism-related recordings have not yet been translated…[and] that the F.B.I. still lacked the capacity to translate all the terrorism-related material from wiretaps…" The report conceded that "the influx of new material has outpaced the Bureau's resources." Among the reasons given by the inspector general for this embarrassing backlog was the "shortage of qualified linguists and problems in the bureau's computer systems…[and] management and efficiency problems that dogged the bureau even before September 11th" [2]. There is no reason to believe that these problems have been solved, despite the government's obvious determination to gather still more data.

Indeed, it should be asked whether there may be another unchanged reason for the discrepancy between data collection and analysis: namely, that many government

Neustein A. (2006).
Using Sequence Package Analysis as a New Natural Language Understanding Method for Mining Government Recordings of Terror Suspects.
In *Proceedings of the 3rd International Workshop on Natural Language Understanding and Cognitive Science*, pages 101-108
DOI: 10.5220/0002473201010108

translators and linguists are skeptical about finding important clues to terror-related activities in recordings of conversations with terror suspects. Such skepticism, after all, is at least partly justified. Most audio data mining programs that parse recordings in search of “keywords” can be stymied by speakers who deliberately avoid the use of keywords – names of persons, locations, landmarks or references to times and calendar dates – that might serve as “red flags” to anyone listening in on the call. As a result, clever terrorists can outsmart a conventional mining program that relies on word-spotting techniques in parsing recorded dialog.

Against this background, some members of the intelligence community have noted the benefit of exploring newer and more efficient data mining methods. In the wake of 9/11, the National Law Enforcement Technology Center, a special program within the National Institute of Justice’s Office of Science and Technology that provides information as a service to law enforcement and forensic science practitioners, devoted part of one of its weekly newsletters to a new AI-based natural language understanding method (one which has been successfully peer reviewed), calling it “a new voice technology tool” to “help law enforcement better weed through wire-tapped conversations to learn of possible terrorist plots” [3].

This method, known as Sequence Package Analysis (or SPA), was developed and formulated by the author as a possible remedy for the common shortcomings of conventional word-spotting data mining programs [4, 5].

One of the main virtues of an SPA-driven mining program is its ability to point out to the human intelligence officer or agent (even in real time) those precise portions of the terror suspects’ conversations that require particularly close (human) analytic inspection, thus sparing the agent the need to listen to or comb through a transcript of the entire call. Another advantage of this method is that it allows the “discovery” of a whole new set of keywords, such as names of persons and places, which could not have been anticipated when the speech application vocabulary was designed.

2 Methodology

What distinguishes Sequence Package Analysis, or SPA, from conventional audio mining programs is that for SPA the primary analytical focus is the *unit of interaction* in its entirety – the “sequence package” – whereas conventional mining programs generally focus on single or multiple lexical items, such as a “content word” (e.g., “attacking”) or its corresponding “content term root” (e.g., “attack”).

Sequence packages involve different phases of dialog and conversational activities, such as call openings and closings, complaints, and the making of plans or arrangements. Reduced to algorithms, many sequence packages are naturally transferable from one contextual domain to another, which means that many of the same sequence package structures found in the conversations of terror suspects also appear in call center dialogs between customers and call center agents.

The sequence package consists of a series of related turns and turn construction units (that is, the syntactically bounded parts of the turn at the completion of which the speaker may yield to the other speaker) that are discretely packaged as a sequence of conversational interaction [6]. By relying on the sequence package as the primary unit of analysis, rather than on an individual word or word combination, an SPA-driven mining program parses the conversation for its relevant sequences, which consist of clearly defined sets of sequence packages. Given that dialog itself is more or less a blend of sequences folding into one another, rather than a string of isolated keywords, a mining program driven by SPA can better accommodate how people really talk, especially in those instances when speakers deliberately avoid the use of certain words that can alarm intelligence agents. Thus, because SPA is not restricted to the matching of keywords, it can work more flexibly with speaker input – which naturally becomes more convoluted and elliptical in a guarded, secretive conversation.

The way SPA adjusts to speech that is less than "perfect" is to offer a set of algorithms that can work with, rather than be hindered by, the ambiguities, ellipses, idioms, metaphors, colloquialisms, and the many other facets of natural language dialog. Ironically, SPA mines conversations to find the very sort of dialog data that would have been discarded (or simply ignored) by most speech systems as unwieldy talk or talk that is far too amorphous to grasp. And while some of these discarded data (such as the occurrence of inter-sentential connectives, or slight variations in inter- and intra-utterance spacing) might appear relatively unimportant to a mining program, these data can be very significant in properly interpreting natural language dialog, including the conversations of terror suspects.

It is no easy task to map out the conversational sequence patterns of natural language dialog. To do this, SPA draws from the field of conversation analysis as its methodological basis. What conversation analysis provides is a rigorous, empirically-based method of recording and transcribing verbal interactions by using highly refined transcription signals to identify both verbal components and paralinguistic features, such as stress, pauses, gaps, overlaps and changes in intra-utterance spacing [7].

Conversation analysis breaks down natural language communication into its primary units of analysis: sequences and turns within sequences (rather than isolated sentences or utterances). In this way, conversation analysts have studied interactive dialog for over 35 years as a socially organized activity. In essence, the conversation analyst can be distinguished from the linguist by the fact that the linguist focuses on grammatical discourse structure, while the conversation analyst focuses on social action [8]. And by focusing on social action, rather than on grammatical discourse structure solely, the SPA method can be readily applied to a myriad of other languages, including Arabic and Farsi, because "*all* forms of interactive dialog, regardless of their underlying grammatical discourse structures, are ultimately defined by their *social* architecture" [9].

3 Design

There are two ways that an SPA-driven mining program can work. First, it can serve as an "add on" layer for conventional data mining programs, including those built on vector-based models, which assign n-grams and bi-grams and hold spaces in between words and word phrases accordingly. If SPA functions as an "add on" layer, the "global weighting" to be applied for the next layer of analysis need no longer be limited to content words or their term roots; rather, it can now also encompass sequence package material. To accomplish this, SPA uses Statistical Language Modeling (SLM) – the standardized method for matching speech input to the speech application vocabularies – but instead of generating candidate words and word phrases for the speech input, SPA generates *candidate* sequence packages. Thus, using the same method of weighting possibilities used for candidate words and word phrases, SPA detects the range of possible sequence packages present at each stage of the conversational sequence, the totality of which makes up the dialog.

As an "add on" layer, SPA can take the output of a speech engine and provide a deeper level of analysis of the terror suspects' dialog by interpolating sequence package information into the output stream. By marking sequence package boundaries and specifying package properties, the SPA-enhanced mining program gives the software downstream the contextual indicia – the precise location points in the flow of interactive dialog which signify the different conversational activities and phases of the dialog – needed to interpret the rest of the data stream reliably.

Another advantage of this approach is that demarcating the circumscribed boundaries and properties of sequence packages helps resolve anaphoric connectivity issues. Anaphors pose a knotty problem for natural language systems, particularly when anaphors, such as pronouns, cannot be understood as referring back either to their antecedents or as variables that are bound by their antecedents [10, 11]. SPA can begin to address such anaphoric connectivity problems by first drawing the boundaries that circumscribe the sequence packages, and then connecting each anaphor *only* with the referent that is contained within the tight boundaries of the sequence package. This way, only those referents enclosed within the sequence package can be related to the anaphoric word or word phrase, thus insuring that what remains outside the sequence package will not be mistakenly designated as the referent for the anaphor.

Second, SPA might be used as a wholly integrated system rather than as an "add on" layer to conventional data mining programs. In such a case, data mining programs would use sequence package grammars rather than content words as their starting point. Such a use would allow the building of an entire vocabulary of appropriate content words, and their corresponding root terms, without necessarily having to have an *a priori* knowledge of such words. Using this same heuristic approach, a data mining program would seek to discover, in addition to content words and their term roots, new or related sets of sequence packages that demonstrate the patterned way humans engage in interactive dialog.

But regardless of whether SPA is built into a system as an "add on" layer of intelligence or in the alternative as a wholly integrated system, it can be argued that SPA, for the most part, can enhance the scalability of data mining programs. This is so because SPA can help to streamline the corpus of data required to build a statistical language model, by focusing on commonly occurring sequence packages that are generic to a large population of speakers, and thereby eliminate the need to construct elaborate speech application vocabularies, in anticipation of every possible word to be used by a speaker.

4 Demonstration

Here is a hypothetical example of a conversation between two terror suspects taking place in Brooklyn shortly after 9/11. Although the dialog is a hypothetical construction, the sequence patterns contained in the dialog example below are themselves empirically derived from the analysis of actual conversations [12].

In the example below, Speaker "A" is trying to inform Speaker "B" about an important meeting to take place at a new location, which is right at the foot of the Brooklyn Bridge. However, Speaker "A" is confronted with two difficulties: First, he must make a concerted effort to avoid any direct reference to Brooklyn Bridge – a known heavily surveilled location for terrorist activities – because it could arouse the suspicions of an intelligence agent who might be listening in on the call.

Second, Speaker "A" must try to maintain an air of nonchalance, refraining from making any prefatory remarks to the other speaker that could convey a sense of "urgency" that might arouse suspicion in a third party listening in on the call. As part of this air of nonchalance, the speaker must also prevent any sudden changes in prosody (vocal stress patterns) that could draw the attention of a third party.

Yet, in spite of these constraining conditions placed upon Speaker "A," he must try to accomplish the work at hand of unequivocally conveying to Speaker "B" where to meet – making sure he understands the directives, lest the plans be spoiled. Here is how the speaker might accomplish this delicate task:

Speaker "A": Come to the intersection near River Cafe? (the question mark shows an upward intonation) 0.2-0.5 second pause

Speaker "B": 1.6 second pause

Speaker "A": You know the busy street with the big traffic light?

Speaker "B": River Café, yeah.

Although, in this example, both speakers avoided any reference to the "Brooklyn Bridge" as well as any reference to the importance of getting these directives straight, an SPA-driven mining program could have detected their intent. To do this, it would have mapped out the following six-part sequence package for making arrangements, paying particularly close attention to the spacing of inter utterance and intra utterance pauses that are found in the dialog:

Speaker "A"

1) A noun referent ("River Cafe") with an upward intonation: "Come to the intersection near River Cafe?"
2) A brief pause, giving the listener the opportunity to show recognition or in the alternative, ask for clarification: 0.2-0.5 seconds

Speaker "B"

3) A long pause by the listener which indicates his lack of understanding or possible confusion: 1.6 seconds

Speaker "A"

4) A clarification of the noun referent ("River Cafe"): "You know the busy street with the big traffic light"

Speaker "B"

5) A repetition of the noun referent, which had been the source of the recognition trouble: "River Cafe"
6) A "recognition marker" immediately after the repeat of the noun referent, which had been the source of the recognition trouble: "yeah"

5 Analysis

In this example, an SPA-driven mining program would have uncovered the term "River Café" upon its search of the dialog for sequence package templates that form the most likely match for the sequences found in the speech engine's output stream. Here's how:

First, the mining program would look for a noun referent marked by an upward intonation followed by a brief pause. Second, the program would identify the deviations from the norm in inter-utterance spacing – *i.e.*, wherever the gap between speaker "A" and speaker "B" exceeds what conversation analysts call the "tolerance interval" (p. 170) [13], an interval "which marks the acceptable length of absent talk in conversational interaction" (p. 144) [14]. The consensus among conversation analysts is that silences exceeding 1.2 seconds signal trouble in the dialog. In this example, we have an inter-utterance pause lasting 1.6 seconds, which would be noted by the mining program.

Third, the program would look for a clarification of the noun referent that caused the recognition trouble displayed by the other speaker. Since the clarification attempt is constructed as an anaphor ("…the busy street with the big traffic light"), the program must search solely within the boundaries of the sequence package to link the anaphor correctly to its antecedent referent. In so doing, the program would locate the prior utterance which begins the sequence package. At that point in the dialog, the speaker raises his inflection when identifying a new meeting place, pausing slightly to give the other speaker the chance for feedback ("Come to the intersection near River Café?" 0.2-0.5).

It should be noted that in the example given above, the program's decision to link the anaphoric expression to its antecedent referent in the prior utterance is not governed by grammatical rules, which might dictate the linking of anaphors to their antecedent referents in the immediately preceding sentence. Sequence package configurations work differently, recognizing the patterned regularities of talk as a socially organized activity. In light of such regularities, anaphoric connectivity may in fact deviate from strict grammatical rules – as in the case of an enraged speaker who fails to identify the subject or object of his ire until *after* several speaking turns of "venting" which have been punctuated by anaphoric expressions.

The last part of this sequence package template indicates that the trouble, which provoked a long silence and a subsequent clarification, has been resolved. The speaker's repetition of the noun referent that had been the source of the trouble ("River Café"), followed by a recognition marker ("Yeah"), ends the sequence and, in so doing, ends the phase of the dialog in which arrangements to meet are made.

A mining program that uses SPA to uncover critical information about suspects' activities (such as their meeting places) would now have the option of adding "River Café" to the speech application's vocabulary as an important word to look out for in the future because of its close proximity to Brooklyn Bridge. In short, an SPA mining program would work in two phases: first, it would generate candidate sequence packages for the speech input found in the speech engine's downstream; second, it would extract from these sequences packages "new" references to persons or places so that they can be properly added to the speech application vocabulary. In this way, one can empirically design an application vocabulary that better matches the reference terms (names and locations) that suspects actually use, when discussing terrorism-related activities, than a vocabulary that is derived from a list of "keywords" that one thinks they will use.

The six-part sequence package analyzed above consists of a concatenation of utterance components that are *commonly* found in dialog, whether or not the conversation revolves around the activities of terror suspects. A mining program can expect to see this linguistic pattern with some degree of predictability when one speaker in the course of making arrangements introduces a new term (such as a name of a person or a place) to another speaker – and where the "uninformed" speaker seeks, for whatever reason, to minimize his "ignorance" of the new term, by allowing the conversation to continue without stopping first to "topicalize" his lack of recognition of the new term ("Oh, I had not heard of River Café before now!"). This shows that the algorithmic design of sequence packages, including those that underlie the conversational activity of "making arrangements," is generic enough to be detected not only in conversations of terror suspects but across other domains.

6 Conclusion

SPA technology brings to data mining a new method of parsing dialog, one that examines conversation for its relevant sequences, consisting of clearly defined sets of sequence packages. By breaking up dialog into discrete sets of sequence packages – which often include linguistic data discarded by most mining programs – SPA-driven automated mining programs can help intelligence practitioners decipher the covert dialog of terror suspects, characteristically ambiguous and elliptical. This new method of natural language understanding can enhance efficient mining of important information that is all too often masked by terror suspects, who carefully avoid the use of names and locations, among other things. Perhaps with the availability of a more efficient method for mining terrorism-related calls, the F.B.I. will be able to reduce its enormous backlog of untranscribed and unanalyzed calls. This could only help to paint a more encouraging picture of our homeland security, which could stand as a model for intelligence operations throughout the world.

References

1. Lichtblau, E., Risen, J.: Spy Agency Mined Vast Data Trove, Officials Report. New York Times (December 24, 2005) A1, 20
2. Lichtblau, E.: F.B.I. Said to Lag on Translations of Terror Tapes. New York Times (September 28, 2004) A1, 22
3. JustNet.: Linguistics Expert Predicts Voice Technology Will Play Pivotal Role in Spotting Terrorists. JustNet-Law Enforcement and Corrections Technology News Summary (October 18, 2001)
4. Neustein, A.: Sequence Package Analysis: A New Natural Language Understanding Method for Performing Data Mining of Help-Line Calls and Doctor-Patient Interviews. Proceedings of the First International Workshop on Natural Language Understanding and Cognitive Science. University of Portugal, Porto, Portugal (April, 13, 2004) 64-74
5. Neustein, A.: Using a New Method of Natural Language Intelligence for Performing Wiretap Analysis. Policy Sciences Annual Institute. Yale Law School, New Haven, Conn. (October 23, 2004)
6. Neustein, A.: Using Sequence Package Analysis to Improve Natural Language Understanding. International Journal of Speech Technology 4 (1) (2001) 31-44
7. Atkinson, J.M., Heritage, J.: Transcript notation. In: Atkinson, J.M., Heritage, J. (eds.): Structures of Social Action: Studies in Conversation Analysis. Cambridge University Press, Cambridge (1984) ix-xvi
8. McIlvenny, P., Raudaskoski, P.: The mutual relevance of conversation analysis and linguistics: A discussion in reference to interactive discourse. In L. Heltoft and H. Haberland (eds.): Proceedings of the Thirteenth Scandinavian Conference on Linguistics. Department of Languages and Culture, Roskilde University, Roskilde, Denmark (1992) 263-277
9. Neustein, A.: Sequence Package Analysis: A New Global Standard for Processing Natural Language Input? Globalization Insider X111(1, 2) (February 18, 2004) 1-3
10. Asher, N.: A Typology for Attitude Verbs and their Anaphoric Properties. *Linguistics and Philosophy 10 (10)* (1987) 125-197
11. Edelberg, W.: A New Puzzle about Intentional Identity. *Journal of Philosophical Logic* 15 (1986) 1-25
12. Sacks, H., Schegloff, E.A.: Two Preferences in the Organization of Reference to Persons in Conversation and Their Interaction. In: G. Psathas, (ed.): Everyday Language: Studies in Ethnomethodology. Irvington Publishers, Inc, New York (1979) 15-21
13. Jefferson, G.: Notes on a possible metric for a “standard maximum” silence of approximately one second in conversation. In: Roger, D, Bull, P. (eds.): Conversation: An Interdisciplinary Perspective. Multilingual Matters, Clevedon and Philadelphia (1989) 166-196
14. Wooffitt, R., Fraser, N.M., Gilbert, N., McGlashan, S.: Humans, Computers and Wizards: Analysing Human (Simulated) Computer Interaction. Routledge, London (1997)