A New Natural Language Method for Performing Efficient Mining of Government Tape-Recordings of the Conversations of Terror Suspects

Amy Neustein, Ph.D.

Linguistic Technology Systems
1055 River Road, Suite 1013
Edgewater, New Jersey 07020
U.S.A.
lingtec@banet.net

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 have yet to be transcribed. Why has there been such a lapse in homeland security? Perhaps linguists performing transcriptions for the Department are skeptical about finding any solid "clues" about terror activities in this elusive mass of recorded conversations of terror suspects – and if they are, it is for good reason. After all, suspects tend to avoid words that might alarm intelligence agents listening in on the call. By doing so, they "outsmart" conventional mining programs, which heavily rely on word-spotting techniques in parsing recorded dialog. One solution to this problem 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 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.

Keywords: homeland security, data mining, word-spotting, natural language understanding, sequence package analysis

1. Introduction

Shortly after the third anniversary of the September 11 attacks, 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" [1].

These were doubtless contributing factors. Still, it should be asked whether there may be another reason as well: namely, that many government 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.

In recent years, however, some members of the intelligence community have shown an interest in a new AI-based natural language understanding method (which has been successfully peer reviewed) as "a new voice technology tool" to "help law enforcement better weed through wire-tapped conversations to learn of possible terrorist plots" [2]. This method, known as Sequence Package Analysis (or SPA), was developed and formulated by the author as a remedy for the common shortcomings of conventional word-spotting data mining programs [3].

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. 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.

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").

The sequence package consists of a series of related turns and turn construction units (lexically bounded parts of turns at which point the turn may conceivably, but not necessarily, be yielded to the next speaker) that are discretely packaged as a sequence of conversational interaction [4]. By relying on the sequence package itself as the primary unit of analysis, rather than an individual word or word combination, a mining program can potentially accommodate better to 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 bound 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.

It is no easy task to map out the conversational sequence patterns of natural language dialog [5]. 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 [6].

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 [7]. 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" [8].

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 that are 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.

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.

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 [9].

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. 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": 1.2 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, 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:

1.2 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”

This six-part sequence package consists of a concatenation of utterance components that are g*enerically* found in dialog when one speaker introduces a new term (such as a name of a person or a place) to another speaker – and where the “uninformed” speaker seeks 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!”)

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. Designers of such a mining program would then have the option of adding “River Café” to the speech application’s vocabulary as an important word to look out for because of its close proximity to the Brooklyn Bridge. In short, an SPA mining program would work in two phases: First, it would search for generic dialog material appearing in the form of clearly delineated sequence package structures; and 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.

5. Conclusion

SPA technology brings to data mining a new method of parsing dialog, which allows 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.

References

[1] Eric Lichtblau, “F.B.I. Said to Lag on Translations of Terror Tapes,” *New York Times*, September 28, 2004, pp. 1, 22.

[2] “Linguistics Expert Predicts Voice Technology Will Play Pivotal Role in Spotting Terrorists,” *JustNet-Law Enforcement and Corrections Technology News Summary,* October 18, 2001.

[3] Amy Neustein, “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*. Porto, Portugal: University of Portugal, April 2004, pp. 64-74.

[4] Amy Neustein, “Using Sequence Package Analysis to Improve Natural Language Understanding,” *International Journal of Speech Technology, 4*(1), 2001, pp. 31-44.

[5] Amy Neustein, “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] J.M. Atkinson and J. Heritage, Transcript notation. In J.M. Atkinson and J. Heritage (Eds.), *Structures of Social*

Action: Studies in Conversation Analysis. Cambridge University Press, Cambridge, 1984, pp. ix-xvi.

[7] P. McIlvenny, P. and 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*. Roskilde, Denmark: Department of Languages and Culture, Roskilde University, Roskilde, Denmark, 1992, pp. 263-277.

[8] Amy Neustein, "Sequence Package Analysis: A New Global Standard for Processing Natural Language Input?" Globalization Insider, vol. X111(1,2), February 18, 2004.

[9] Harvey Sacks and Emanuel A. Schegloff, "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, pp. 15-21.