Conversation Clusters: Human-Computer Dialog for Topic Extraction

Tony Bergstrom
University of Illinois
Urbana, IL 61820
abergst2@cs.uiuc.edu

Karrie Karahalios
University of Illinois
Urbana, IL 61820
kkarahal@cs.uiuc.edu

CHI 2008, April 5 – April 10, 2008, Florence, Italy
ACM 978-1-60558-012-8/08/04.

Abstract
In this paper, we look at projects leveraging human knowledge and understanding in computer systems for extracting conversational topics. Tasks like speech recognition are difficult for computers, but simple for people engaged in conversation. This task is a cornerstone of transcription, speech summarization, and topic recognition. We propose using tabletop interaction to enhance the computer's basic categorization. We then describe how the results of tabletop interaction help to create meaningful archival visualizations and personal reflecting tools.

Keywords
Clustering, Human understanding, Topic extraction.

ACM Classification Keywords
H.5.3 [Group and Organization Interfaces]: Collaborative computing

Motivation
Language has long been one of the most efficient and often used forms of communication between people. However, computational linguistics has a much more difficult task in deciphering the meaning and nuances of human speech [1]. Computers must be able to understand references to objects, previous

conversation, past experiences, and cultural norms. Natural language processing techniques are limited to understanding precisely structured sentences and focus on identifying statistical patterns in language. This can be highly effective in specified domains, but altering the conversation's style, topic, or context from that of the training set can drastically reduce automated comprehension [11]. Even over a fairly homogenous group, conversation varies drastically with age, gender, and context [2]. Words can change meanings, underlying assumptions of knowledge change, and sentence structure can adapt to a speaker's whim. A fairly high degree of computer comprehension can be garnered from newspaper articles, academic papers, and other well-structured compositions, but the fluidity and irregularity of everyday conversation is much more difficult.

We hope to engender speech based tabletop interfaces to bridge the language barrier. Participants of a conversation already have an understanding of the verbal exchange, but this understanding is not easily transferable. While engaged in a conversation, people must be aware of our tabletop visualization and the computer must be reconfigurable at any point. Previous work demonstrates simple visualizations are interpretable [7][8]. With interactive tabletop interfaces, we allow individuals to correct and change the depiction.

While we can rely on humans for interpretation, we rely on computers for memory and synthesis. Computers have a more extensive and sharable memory whereas aural conversation is an ephemeral experience for people. We hope to extend the persistency of conversation to allow individuals to better understand how they change and conversations evolve over time, to more easily access archive audio, and to better understand large quantities of interactive data.

Scenario

Robert rushed to the computer science building Friday morning. His research group was meeting to brainstorm new projects. Robert hurried to the meeting room, knowing he missed a good deal of the discussion. Glancing at the visualization on the table, he sees conversation has been hedging social visualizations, social networks, and mobile technology. He immediately had some ideas for creating presence with mobile devices to pursue, maybe a social network tie in would work well.

As the group continues discussion of visualizations, they begin to discuss using visualizations for awareness. Still thinking about how social networks might play into his mobile device ideas, Robert suggests his own ideas with mobile devices. As he speaks, his words "awareness," "GPS," and then "context" begin to appear.

Jen, another member of the group, had been staring at the table while she listened to project ideas. She saw that words some of Robert's words had inappropriately been grouped with the social network discussion. She touched the surface, moved the offending words to the correct cluster, and offered her own ideas to expand Robert's.

After a week of further discussion and research, Robert was satisfied that he was on to a solid project. He had arranged a meeting with a professor in the Information Science school across campus for a final bit of critique

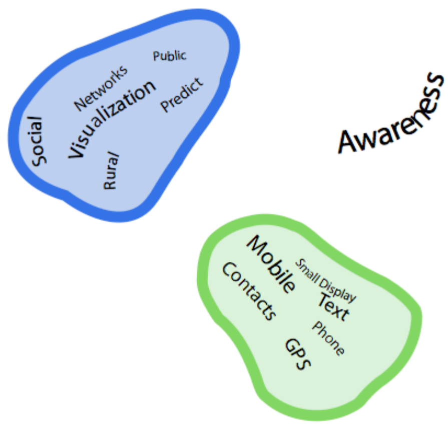


figure 1. This is a rendering of the tabletop visualization of topic clustering. Established and relevant topics appear grouped together and highlighted by color. New potential keywords appear from the edge of the table and either merge into the pre-established clusters or remain solo. Solo words can fade away or anchor a new cluster of related words.

before truly beginning this work. Of course, Robert wanted to provide a page describing the high points of his project proposal. Gathering thoughts, he vaguely remembered excellent points made at the brainstorm session. He pulling up the archived discussions and quickly searched over the intertwined threads of conversation. The discussion on mobile awareness jumped out as a think green thread overlaying the blue social networks thread. They would later combine into a single thread, but Robert remembered Jen having made some excellent points early on that he wanted to be sure to include. Zooming in, he found Jen's critique and began writing up his proposal.

Topic Extraction

Summarizing speech automatically is a challenging task. Firstly, people do not speak the same way they write, speech allows for much more variation and flexibility in language rules. Additionally, algorithms to recognize speech only transcribe about 85% of the spoken words accurately in average circumstances [3]. This error-laden transcript must then be summarized by another process that requires a computer to choose the most salient aspects of a conversation [4]. Challenges are only compounded when dealing with multiple speakers and atypical accented speech. But, why recreate a skill when it is already present in the participants of the conversation?

Human Solutions

While it's difficult for a computer to correctly follow a conversation or to make contextual assumptions of a situation, human participants are exceptionally gifted at these tasks. Harnessing the conversational knowledge of humans, interfaces need only show the right information to allow people to do the bulk of the creative or associative work. Work like Peek-a-boom and other works by Louis von Ahn show that leveraging human processing cycles is a viable approach for classification [5]. By using games that people enjoy, Peek-a-boom encourages people to label objects in thousands of images.

Other sites such as flickr and facebook have extensive databases of annotated photographs. People and things are labeled explicitly or tagged. However, the respective companies did not hire people to do the work; they relied on the individual users of their system. Given the opportunity and motivation, the user base proved to be a valuable resource of human knowledge. New projects, such as TagMaps are only just being developed to use this data [6].

Our own prior work with Conversation Clock and Conversation Votes relies on the interpretive abilities of people [7][8]. Both of these visualizations demonstrate the interaction history of groups over the course of a conversation. They provide a summary of monitored inputs via microphones, but the context of these conversations is left for the participants to decipher. Participants reported a heightened awareness of the conversation, while retaining a sense of natural interaction.

Adapting Interfaces for Context

In our developing work, we are seeking to tie more contextual cues into conversation visualization. Using tools for speech recognition, our work will use the error laden transcript of conversation to create a transcript. From the generated transcript, we intend to extract key words, based upon relevance and confidence, indicative of topic and discussion. Rather than simply computing a topic from this, we use people for context and error correction.

Keyword Extraction

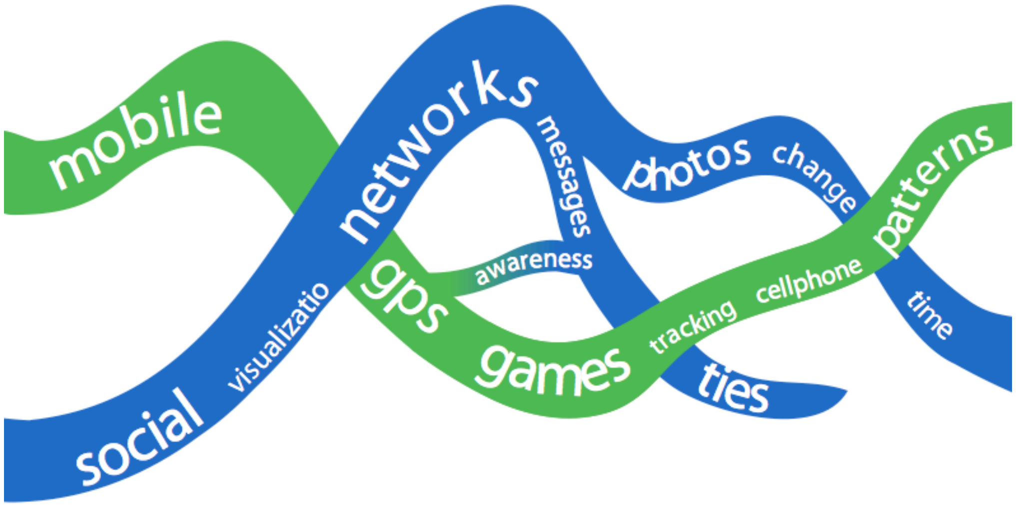
During a conversation, participants will see some of the words spoken. These words are chosen to be indicative of the conversation at that moment and are projected onto the interactive table. Using machine learning, these selections will be chosen from the audio stream and clustered into topics as in Figure 1.

To generate our initial topics and word clusters, we first use a relatedness metric. This metric provides an indication of how often these words appear together. The words campus, university, school, and other synonyms are highly related while words books, project, research, draft, and professor indicate a weaker relation. A basic relationship metric can be generated, using Term Frequency - Inverse Document Frequency weights, by counting the co-occurrence of words in a pre arranged corpus and in past conversation.

As we are continuously tracking topics, we must presuppose they change. To account for this we break conversations into overlapping windows of time. Each window must be evaluated and compared to the previously established topics. Relevant topic words are chosen for each window. Using the relatedness metric, we first try to fit the newly chosen words into established topics. If a threshold level of relatedness is not met, the word will start its own topic if more related words are generated, otherwise it will fade away.

These clusters represent overarching threads of conversation related to recent speech. Each cluster will have its own classifier to identify potential new members of the visual cluster. As new keywords emerge, an existing topic cluster will absorb them if a base level of confidence is met. Otherwise, keywords remain free from existing clusters as their own entity.

As words related to each of these clusters cease to be added, their size in the visualization will be decreased. Visually, this signals a topic’s decreasing significance in conversation. At the same time, the keywords added to the visualization feed the growth of new topic clusters.



Human Interaction

Given a perfect topic recognition agent, there should be relatively little interaction. Human interaction takes place to ensure the chosen words are salient and relevant. If our algorithm chooses a word and places it in the wrong topic, it can be removed or placed in a different topic. We use a touch sensitive table, the Mitsubishi DiamondTouch surface, with an ceiling mounted projector to facilitate this interaction. Simple interactions such as striking a word to remove it or dragging a word to relocate it can be detected during the conversation. The machine learners associated with clustering words into topics, can adjust to accommodate the move. Because of the interaction, the clusters will contain more appropriate groups of salient words and better describing the conversation.

Applications

Topic extraction provides contextual information for conversation. Our previous work did not incorporate context forcing the viewer to retain context indefinitely to get a full picture. We believe topic extraction and tracking provides a better facility for archiving and reviewing past audio recordings (figure 2).

Consider the case of a journalist, historian, or investigator: people who must extensively interview others in order to synthesize their own work. Large volumes of aural data are left to review. Though the individual might have a general idea and extensive notes to provide understanding, topic tracking these conversations provides meaningful visual access points into the recording. An exceptional full visual of conversation would provide a random access entry point into otherwise serial data.

As a self-reflective tool, topic tracking could provide annotated archives of personal interaction. Tools such as iRemember have been shown to allow individuals to access specific events as a memory tool [9]. A topic tracking tool would allow one to see what they talked about and how it changes over time.

A final application would tie topic tracking to a physical location. A coffee house, for example, has a wide variety of individuals that stop in and chat over the course of a day or week. They talk and discuss relevant events of their own day. Tracking topics in the location could provide an idea of local events and concerns in the community while still providing anonymity. This awareness can be demonstrated with explicit interaction as in [10], but we believe implicit interaction creates a more complete picture.

Discussion

The impact of this visualization lies in the ability to annotate live conversations. Over the course of a day, a week, a year or more, the visualization can provide an index into the large volume of aural information we hear. Extensions of this tool would allow the exploration into how a person's conversation changes over time, showing the rise and fall of themes. As a personal reflection tool, this would hopefully enable people to better understand their own interaction.

Additionally, the tool could be used to annotate live feeds of audio, such as a debate, interview, or live broadcast for the later benefit by others. The appropriately chosen groups of keywords offer meaningful specific indexes into serial streams of audio. Appropriate keywords and topics make a conversation more indexable, searchable, and computer interpretable.

References

[1] Allen, J. Natural language understanding (2nd ed.). Benjamin-Cummings publishing Co., Inc., Redwood City, CA, USA (1995)

[2] Tannen, D. Teachers’ classroom strategies should recognize that men and women use language differently. The Chronicle of Higher Education (1991).

[3] Munteanu, C., Baecker, R., Penn, G., Toms, E., James, D.: The effect of speech recognition accuracy rates on the usefulness and usability of webcast archives. In Proc. of SIGCHI 2006.

[4] Zechner, K. Summarization of spoken language - challenges, methods, and prospects. Speech Technology Expert. eZine Issue 6 (2002).

[5] von Ahn, L., Liu, R., Blum, M. Peekaboom: a game for locating objects in images. In Proc. of SIGCHI 2006.

[6] Yahoo! Tagmaps http://tagmaps.research.yahoo.com

[7] Bergstrom, T., Karahalios, K. Seeing more: Visualizing audio cues. In Interact 2007.

[8] Bergstrom, T., Karahalios, K. Conversation Votes: enabling anonymous cues. In Extended Abstracts of CHI 2007.

[9] Vemuri, S., Schmandt, C., Bender, W. iRemember: APersonal Long-term Memory Prosthesis. In Proc. of CARPE 2006.

[10] Churchhill, F., Nelson, L. Interactive Community Bulletin Boards as Conversation Hubs and Sites for Playful Visual Repartee. HICSS 2007.

[11] Rosenfeld, R. Two decades of statistical language modeling: where do we go from here? Proc of IEEE Vol 88 iss 8. Aug 2000.