Figure 1: A visual interface for browsing and summarizing conversations with 4 integrated views - a) the Ontology View (left), b) the Transcript View (middle), c) the Entity View (bottom), and d) the Summary View (right)

1 INTRODUCTION AND RELATED WORK

In our daily lives, we have conversations with others in many different modalities like meetings, emails, chats, blogs etc. The Web has significantly increased the volume and the complexity of the conversational data generated through our day to day communication. Automatic summarization can be of benefit in dealing with this overwhelming amount of interactional information by providing quick access. However, due to its highly informal nature, the multimodal diverse data has proved to be a challenge from Natural Language Processing (NLP) point of view; rendering text processing using syntactic, lexicographic and semantic rules often insufficient. Here we present an interactive interface (see Fig. 1) that takes advantage of human perception in conjunction with automatically extracted conversation knowledge to generate a brief, focused overview of a conversation.

The dominant approach to summarization has been extractive, where the summary is generated by selecting and concatenating the most informative sentences from the source document(s). Extractive summarization has been popular as it is a simple binary classification task using machine learning techniques. The alternative is abstractive summarization, where the summary is generated by extracting and aggregating information from the conversation. This approach requires a Natural Language Generation (NLG) component and is preferred by users for coherency but the resultant summary sentences are often too generic and lack the details of the original conversation sentences. The objective of our interface is to provide a Visual Structured Summary that gives the user more control over choosing the topics she wants to appear in the concise resultant overview generated through interactive exploration. Thus the summary generated can be either extractive or abstractive focused summary.

Several projects in the past have investigated tools for browsing meetings. The Ferret Meeting Browser [9] provides the ability to quickly find and play back a combination of available audio, video, transcript and projected display segments from a meeting side by side for comparison and inspection synchronously. The Meeting Miner [2] aids browsing multimodal meetings through recordings of online text and speech collaborative meetings using timeline navigators of content of edits. In addition, it can retrieve a set of speech turns spread throughout the conversation focused on particular keywords that can be selected from a list of automatically generated keywords and topics. For a more comprehensive list of relevant meeting browsers and document collection visualization tools please refer to [8]. Our interface differs from these other browsers because its purpose is to explore and summarize multi-modal conversations in general.

We use an ontology containing nodes for speakers, dialogue acts (DA), and a list of entities referred to in the conversation to provide entry points to the conversation. This approach requires a Natural Language Generation (NLG) component and is preferred by users for coherency but the resultant summary sentences are often too generic and lack the details of the original conversation sentences. The objective of our interface is to provide a Visual Structured Summary that gives the user more

control over choosing the topics she wants to appear in the concise resultant overview generated through interactive exploration. Thus the summary generated can be either extractive or abstractive focused summary.

Several projects in the past have investigated tools for browsing meetings. The Ferret Meeting Browser [9] provides the ability to quickly find and play back a combination of available audio, video, transcript and projected display segments from a meeting side by side for comparison and inspection synchronously. The Meeting Miner [2] aids browsing multimodal meetings through recordings of online text and speech collaborative meetings using timeline navigators of content of edits. In addition, it can retrieve a set of speech turns spread throughout the conversation focused on particular keywords that can be selected from a list of automatically generated keywords and topics. For a more comprehensive list of relevant meeting browsers and document collection visualization tools please refer to [8]. Our interface differs from these other browsers because its purpose is to explore and summarize multi-modal conversations in general.

We use an ontology containing nodes for speakers, dialogue acts (DA), and a list of entities referred to in the conversation to provide entry points to the conversation. This approach requires a Natural Language Generation (NLG) component and is preferred by users for coherency but the resultant summary sentences are often too generic and lack the details of the original conversation sentences. The objective of our interface is to provide a Visual Structured Summary that gives the user more

control over choosing the topics she wants to appear in the concise resultant overview generated through interactive exploration. Thus the summary generated can be either extractive or abstractive focused summary.

Several projects in the past have investigated tools for browsing meetings. The Ferret Meeting Browser [9] provides the ability to quickly find and play back a combination of available audio, video, transcript and projected display segments from a meeting side by side for comparison and inspection synchronously. The Meeting Miner [2] aids browsing multimodal meetings through recordings of online text and speech collaborative meetings using timeline navigators of content of edits. In addition, it can retrieve a set of speech turns spread throughout the conversation focused on particular keywords that can be selected from a list of automatically generated keywords and topics. For a more comprehensive list of relevant meeting browsers and document collection visualization tools please refer to [8]. Our interface differs from these other browsers because its purpose is to explore and summarize multi-modal conversations in general.

We use an ontology containing nodes for speakers, dialogue acts (DA), and a list of entities referred to in the conversation to provide entry points to the conversation. This approach requires a Natural Language Generation (NLG) component and is preferred by users for coherency but the resultant summary sentences are often too generic and lack the details of the original conversation sentences. The objective of our interface is to provide a Visual Structured Summary that gives the user more

control over choosing the topics she wants to appear in the concise resultant overview generated through interactive exploration. Thus the summary generated can be either extractive or abstractive focused summary.

Several projects in the past have investigated tools for browsing meetings. The Ferret Meeting Browser [9] provides the ability to quickly find and play back a combination of available audio, video, transcript and projected display segments from a meeting side by side for comparison and inspection synchronously. The Meeting Miner [2] aids browsing multimodal meetings through recordings of online text and speech collaborative meetings using timeline navigators of content of edits. In addition, it can retrieve a set of speech turns spread throughout the conversation focused on particular keywords that can be selected from a list of automatically generated keywords and topics. For a more comprehensive list of relevant meeting browsers and document collection visualization tools please refer to [8]. Our interface differs from these other browsers because its purpose is to explore and summarize multi-modal conversations in general.

We use an ontology containing nodes for speakers, dialogue acts (DA), and a list of entities referred to in the conversation to provide entry points to the conversation. This approach requires a Natural Language Generation (NLG) component and is preferred by users for coherency but the resultant summary sentences are often too generic and lack the details of the original conversation sentences. The objective of our interface is to provide a Visual Structured Summary that gives the user more

control over choosing the topics she wants to appear in the concise resultant overview generated through interactive exploration. Thus the summary generated can be either extractive or abstractive focused summary.

Several projects in the past have investigated tools for browsing meetings. The Ferret Meeting Browser [9] provides the ability to quickly find and play back a combination of available audio, video, transcript and projected display segments from a meeting side by side for comparison and inspection synchronously. The Meeting Miner [2] aids browsing multimodal meetings through recordings of online text and speech collaborative meetings using timeline navigators of content of edits. In addition, it can retrieve a set of speech turns spread throughout the conversation focused on particular keywords that can be selected from a list of automatically generated keywords and topics. For a more comprehensive list of relevant meeting browsers and document collection visualization tools please refer to [8]. Our interface differs from these other browsers because its purpose is to explore and summarize multi-modal conversations in general.

We use an ontology containing nodes for speakers, dialogue acts (DA), and a list of entities referred to in the conversation to provide entry points to the conversation. This approach requires a Natural Language Generation (NLG) component and is preferred by users for coherency but the resultant summary sentences are often too generic and lack the details of the original conversation sentences. The objective of our interface is to provide a Visual Structured Summary that gives the user more

control over choosing the topics she wants to appear in the concise resultant overview generated through interactive exploration. Thus the summary generated can be either extractive or abstractive focused summary.

Several projects in the past have investigated tools for browsing meetings. The Ferret Meeting Browser [9] provides the ability to quickly find and play back a combination of available audio, video, transcript and projected display segments from a meeting side by side for comparison and inspection synchronously. The Meeting Miner [2] aids browsing multimodal meetings through recordings of online text and speech collaborative meetings using timeline navigators of content of edits. In addition, it can retrieve a set of speech turns spread throughout the conversation focused on particular keywords that can be selected from a list of automatically generated keywords and topics. For a more comprehensive list of relevant meeting browsers and document collection visualization tools please refer to [8]. Our interface differs from these other browsers because its purpose is to explore and summarize multi-modal conversations in general.

We use an ontology containing nodes for speakers, dialogue acts (DA), and a list of entities referred to in the conversation to provide entry points to the conversation. This approach requires a Natural Language Generation (NLG) component and is preferred by users for coherency but the resultant summary sentences are often too generic and lack the details of the original conversation sentences. The objective of our interface is to provide a Visual Structured Summary that gives the user more

control over choosing the topics she wants to appear in the concise resultant overview generated through interactive exploration. Thus the summary generated can be either extractive or abstractive focused summary.

Several projects in the past have investigated tools for browsing meetings. The Ferret Meeting Browser [9] provides the ability to quickly find and play back a combination of available audio, video, transcript and projected display segments from a meeting side by side for comparison and inspection synchronously. The Meeting Miner [2] aids browsing multimodal meetings through recordings of online text and speech collaborative meetings using timeline navigators of content of edits. In addition, it can retrieve a set of speech turns spread throughout the conversation focused on particular keywords that can be selected from a list of automatically generated keywords and topics. For a more comprehensive list of relevant meeting browsers and document collection visualization tools please refer to [8]. Our interface differs from these other browsers because its purpose is to explore and summarize multi-modal conversations in general.

We use an ontology containing nodes for speakers, dialogue acts (DA), and a list of entities referred to in the conversation to provide entry points to the conversation. This approach requires a Natural Language Generation (NLG) component and is preferred by users for coherency but the resultant summary sentences are often too generic and lack the details of the original conversation sentences. The objective of our interface is to provide a Visual Structured Summary that gives the user more

control over choosing the topics she wants to appear in the concise resultant overview generated through interactive exploration. Thus the summary generated can be either extractive or abstractive focused summary.

Several projects in the past have investigated tools for browsing meetings. The Ferret Meeting Browser [9] provides the ability to quickly find and play back a combination of available audio, video, transcript and projected display segments from a meeting side by side for comparison and inspection synchronously. The Meeting Miner [2] aids browsing multimodal meetings through recordings of online text and speech collaborative meetings using timeline navigators of content of edits. In addition, it can retrieve a set of speech turns spread throughout the conversation focused on particular keywords that can be selected from a list of automatically generated keywords and topics. For a more comprehensive list of relevant meeting browsers and document collection visualization tools please refer to [8]. Our interface differs from these other browsers because its purpose is to explore and summarize multi-modal conversations in general.

We use an ontology containing nodes for speakers, dialogue acts (DA), and a list of entities referred to in the conversation to provide entry points to the conversation. This approach requires a Natural Language Generation (NLG) component and is preferred by users for coherency but the resultant summary sentences are often too generic and lack the details of the original conversation sentences. The objective of our interface is to provide a Visual Structured Summary that gives the user more

control over choosing the topics she wants to appear in the concise resultant overview generated through interactive exploration. Thus the summary generated can be either extractive or abstractive focused summary.

Several projects in the past have investigated tools for browsing meetings. The Ferret Meeting Browser [9] provides the ability to quickly find and play back a combination of available audio, video, transcript and projected display segments from a meeting side by side for comparison and inspection synchronously. The Meeting Miner [2] aids browsing multimodal meetings through recordings of online text and speech collaborative meetings using timeline navigators of content of edits. In addition, it can retrieve a set of speech turns spread throughout the conversation focused on particular keywords that can be selected from a list of automatically generated keywords and topics. For a more comprehensive list of relevant meeting browsers and document collection visualization tools please refer to [8]. Our interface differs from these other browsers because its purpose is to explore and summarize multi-modal conversations in general.

We use an ontology containing nodes for speakers, dialogue acts (DA), and a list of entities referred to in the conversation to provide entry points to the conversation. This approach requires a Natural Language Generation (NLG) component and is preferred by users for coherency but the resultant summary sentences are often too generic and lack the details of the original conversation sentences. The objective of our interface is to provide a Visual Structured Summary that gives the user more
The above utterance is a decision type statement made by the ProjectManager at the meeting. The beginning time of utterance is used to temporally order the whole conversation and the unique identifier of the Utterance object is used to match the utterance with the actual sentence being said and thus any relevant entities.

Our interface consists of four integrated views - the Ontology View, the Entity View, the Transcript View and the Summary View (see Fig. 1). The Ontology View provides a structured way for the users to explore the relevant concepts in the conversation and their relations. It contains a tree hierarchy with core nodes Speaker and DA type. The Entity View is a textual collage of the list of entities referred to in the conversation represented in a tag cloud format. Both of these views show a count beside every concept, within parenthesis, to indicate the frequency with which these concepts appear during the conversation. Also, the labels are scaled according to the counts to make the more frequent ones stand out more due to larger font. The Transcript View is designed to allow the users to view the whole conversation as well as the mapping of each sentence using the criteria set from the Ontology View and the Entity View. It provides an easier way to assess the conversation overview on a particular information need without scrolling through the whole transcript i.e. it acts as a filtered view of the sentences. The summary is also linked to the transcript and can act as a navigation tool by clicking on the summary sentences. The extractive summary sentences (displayed in the top panel of the Summary View) are prefixed with the tag icons to indicate why they were selected. The abstractive summary sentences (in the bottom panel) are followed by a list of the transcript sentences that were aggregated to generate it. We now briefly describe how the abstractive summaries are generated; more details can be found in [5, 6]. The abstractive summary is generated by first combining utterances about pairs of participants and entities that repeatedly co-occur into messages. Then the most informative messages are selected by using an optimization function combining sentences/utterances and messages subject to 3 constraints (a length constraint and 2 constraints tying messages and sentences together). The messages relevant to interactively selected utterances are selected to be shown in the Summary View. An NLG component is used to generate a summary sentence corresponding to a message. The optimization function to sift more informative messages is maximizeield by using a optimization function combining sentences/utterances and messages subject to 3 constraints (a length constraint and 2 constraints tying messages and sentences together). The messages relevant to interactively selected utterances are selected to be shown in the Summary View. An NLG component is used to generate a summary sentence corresponding to a message. The optimization function to sift more informative messages is maximize

\[
\text{maximize} \left(1 - \lambda \right) \sum_i w_i s_i + \lambda \sum_j m_j u_j. \]

Here \(w_i\) is the sum of posterior probabilities for the ontology-mapping classifiers for sentence \(i\), \(s_i\) is a binary variable indicating whether sentence \(i\) is selected, \(u_j\) is the number of sentences contained by the message \(j\) and \(m_j\) is a binary variable indicating whether message \(j\) is selected.

3 Discussion and Future Directions

In the user study that we conducted to evaluate our interface, we found that different people take different approaches in navigating through a conversation and in trying to find informative sentences - some rely on the summary sentences (mainly extractive), some on basic keyword search, some on the markers etc. These results show the necessity for including all these components in our interface even though there is an overlap among their functionality.

Our current design has evolved through numerous iterations of small modifications applying HCI, NLP and InfoVis principles [3, 8]. Still, there is much scope for further improvement. We could reduce human cognition overload by dynamically adjusting the layout of the list of entities based on the keyword prefix the user types in the Search Box (placed on the top right corner of the Transcript View). In the future, we shall extend the interface to display non-linear asynchronous conversation threads like blogs and emails. This modification entails including a thread preview in the interface as in [7] and changes in the Transcript View. However, the classifiers for the DA types and the entity parsers are readily usable for other modes of conversational data. Recurring lexical patterns or phrases are common among individual posts in an email or blog thread or among the series of meetings in the AMI corpus. An overview of this type of repetitions as in [1] may be helpful to the users to satisfy complex query needs. There is a number of ways the extracted entities could be clustered based on their semantic relationships or collocation; such a grouping will also reduce load on human cognition.
REFERENCES


