Visual Structured Summaries of Human Conversations

Giuseppe Carenini and Gabriel Murray
University of British Columbia
Vancouver, BC, Canada
carenini@cs.ubc.ca

ABSTRACT

This paper presents an interactive interface to create visually structured summaries of human conversations via ontology mapping. We have built highly accurate classifiers for mapping the sentences of a conversation in an ontology, which includes nodes for the Dialog Acts (DA) properties such as decision and subjective, along with nodes for the conversation participants. In contrast with previous work, our classifiers do not rely on features specific to any particular conversational modality. We are currently developing an interactive interface that allows the user to generate visual structured summaries by searching a conversation for sentences according to the ontology mapping. Our first prototype comprises two panels. The right panel displays the ontology, while the left panel of the our prototype displays the whole conversation, where sentences are temporally ordered. Given the information displayed in the two panels, the user can generate visual, structured summaries by selecting nodes in the ontology. As a result, the sentences that were mapped in the selected nodes will be highlighted. Our initial prototype builds on a component of the GATE system, which was originally developed as a tool for text annotation.

INTRODUCTION

Our lives are increasingly comprised of multimodal conversations with others. We email for business and personal purposes, attend meetings in person and remotely, chat online, and participate in blog or forum discussions. It is clear that automatic summarization can be of benefit in dealing with this overwhelming amount of interactional information. Automatic meeting abstracts would allow us to prepare for an upcoming meeting or review the decisions of a previous group. Email summaries would aid corporate memory and provide efficient indices into large mail folders.

The dominant approach to the challenge of automatic summarization has been *extraction*, where informative sentences in a document are identified and concatenated to form a condensed version of the original document. Extractive summarization has been popular at least in part because it is a binary classification task that lends itself well to machine learning techniques, and does not require a natural language generation component. There is evidence that human abstractors at times use sentences from the source documents nearly verbatim in their own summaries, justifying this approach to some extent [9]. Extrinsic evaluations have also shown that, while extractive summaries may be less coherent than human abstracts, users still find them to be valuable tools for browsing documents [7, 10, 13].

However, these same evaluations also indicate that concise abstracts are generally preferred by users and lead to higher objective task scores. The limitation of a cut-and-paste summary is that the end-user does not know *why* the selected sentences are important; this can often only be discerned by exploring the context in which each sentence originally appeared. One possible improvement is to create *structured extract summaries* that represent an increased level of abstraction, where selected sentences are grouped according to the entities they mention as well as to phenomena such as *decisions*, *action items* and *subjectivity*, thereby giving the user more information on why the sentences are being highlighted. For example, the sentence *Let's go with a simple chip* is about a *simple chip* and represents both a decision and the expression of a positive subjective statement.

While much attention in recent years has been paid to (unstructured) extractive summarization of human conversations, including meetings [5], emails [17, 2], telephone conversations [21] and internet relay chats [20], in this paper we present a novel approach to generating visual, structured summaries of human conversations. In our approach sentences are first mapped to nodes in a conversation ontology. Then, the user can search the conversation through an interactive visualization that effectively display both the ontology and the conversation, and allows the user to search the conversation based on the ontology mapping.

The mapping of sentences to the ontology is performed by first identifying all the entities referred to in the conversation, and then by utilizing classifiers relating to a variety of sentence-level phenomena such as *decisions*, *action items* and *subjective sentences*. We achieve high classification accuracy by using a very large feature set integrating conversation structure, lexical patterns, part-of-speech (POS) tags and character n-grams.

Once the mapping is created the user can generate visual, structured summaries by searching a conversation for sen-

CHI 2009, April 3 - 9, 2009, Boston, MA, USA.

tences that convey information about nodes in the ontology. These sentences are highlighted in the context of the whole conversation. For instance, if a user wanted to highlight all the sentences in an email thread expressing *decisions* on the *remote control* made by the *project manager*, she could achieve that by simply selecting the corresponding nodes in the ontology.

In this paper, we first describe the process of mapping sentences to a conversation ontology and then we present our interface to generate visual structured summaries.

ONTOLOGY MAPPING

Our approach relies on a simple conversation ontology. The ontology is written in OWL/RDF and contains two core upper-level classes: Participant and Entity. When additional information is available about participant roles in a given domain, Participant subclasses such as ProjectManager can be utilized. The ontology also contains six properties that express relations between the participants and the entities. For example, the following snippet of the ontology indicates that *hasActionItem* is a relationship between a meeting participant (the property domain) and a discussed entity (the property range).

```
<owl:ObjectProperty rdf:ID="hasActionItem">
  <rdfs:domain rdf:resource="#Participant"/>
  <rdfs:range rdf:resource="#Entity"/>
</owl:ObjectProperty>
```

Similar properties exist for decisions, actions, problems, positive subjective sentences, negative subjective sentences and general extractive sentences (important sentences that may not match the other categories), all connecting conversation participants and entities. The goal is to populate the ontology with participant and entity instances from a given conversation and determine their relationships. This involves identifying the important entities and classifying the sentences in which they occur as being decision sentences, action item sentences, etc.

Our current definition of entity is simple. The entities in a conversation are noun phrases with mid-range document frequency. This is similar to the definition of concept as defined by Xie et al. [19], where n-grams are weighted by *tf.idf* scores, except that we use noun phrases rather than any n-grams. We use mid-range document frequency instead of *idf* [4], where the entities occur in between 10% and 90% of the documents in the collection. We do not currently attempt coreference resolution for entities; recent work has investigated coreference resolution for multi-party dialogues [11, 6], but the challenge of resolution on such noisy data is highlighted by low accuracy (e.g. F-measure of 21.21) compared with using well-formed text (e.g. monologues).

We map sentences to our ontology's object properties by building numerous supervised classifiers trained on labeled decision sentences, action sentences, etc. A general extractive classifier is also trained on sentences simply labeled as important. After predicting these sentence-level properties, we consider a participant to be linked to an entity if the participant mentioned the entity in a sentence in which one of these properties is predicted. We give a specific example of the ontology mapping using this excerpt from the AMI corpus [3]:

1. A: And you two are going to work together on a *prototype* using *modelling clay*.
2. A: You'll get *specific instructions* from your *personal coach*.
3. C: Cool.
4. A: Um did we decide on a *chip*?
5. A: Let's go with a *simple chip*.

Example entities are italicized. Sentences 1 and 2 are classified as action items. Sentence 3 is classified as positive-subjective, but because it contains no entities, no $< participant, relation, entity >$ triple can be added to the ontology. Sentence 4 is classified as a decision sentence, and Sentence 5 is both a decision sentence and a positive-subjective sentence (because the participant is advocating a particular position). The ontology is populated by adding all of the sentence entities as instances of the Entity class, all of the participants as instances of the Participant class, and adding $< participant, relation, entity >$ triples for Sentences 1, 2, 4 and 5. For example, Sentence 5 results in the following two triples being added to the ontology:

```
<ProjectManager rdf:ID="participant-A">
<hasDecision rdf:resource="#simple-chip"/>
</ProjectManager>

<ProjectManager rdf:ID="participant-A">
<hasPos rdf:resource="#simple-chip"/>
</ProjectManager>
```

We have tested our classifiers both on meeting and email data, the AMI [3] and BC3 [18] corpus respectively. On meetings, we achieve remarkable performances, with classification AUROCs ranging from .93 to .77, depending on the classification task. On emails, results are slightly lower, but still potentially useful, with classification AUROCs ranging from .75 to .66. For a detailed discussion of the results see [12] [1].

A key feature of our mapping approach is that it only relies on generic conversational features and can therefore be applied to a multi-modal conversation, for instance a conversation that spans both an email thread and a meeting. Noticeably, our classifiers achieve similar results to [8], [15, 14], [16], who perform these classification tasks by relying on meeting-specific or email-specific features (e.g., prosody for meetings).

GENERATING VISUAL STRUCTURED SUMMARIES

We are developing an interactive interface that allows the user to generate visual structured summaries by searching a conversation for sentences according to the ontology mapping. Figure 1 shows our first prototype for such an interface. The right panel displays the ontology which includes, at the time of writing, nodes for the Dialog Acts (DA) properties such as decision and subjective, along with nodes for

[1] If this paper is not be accepted to NAACL, a draft version can be requested to the authors.

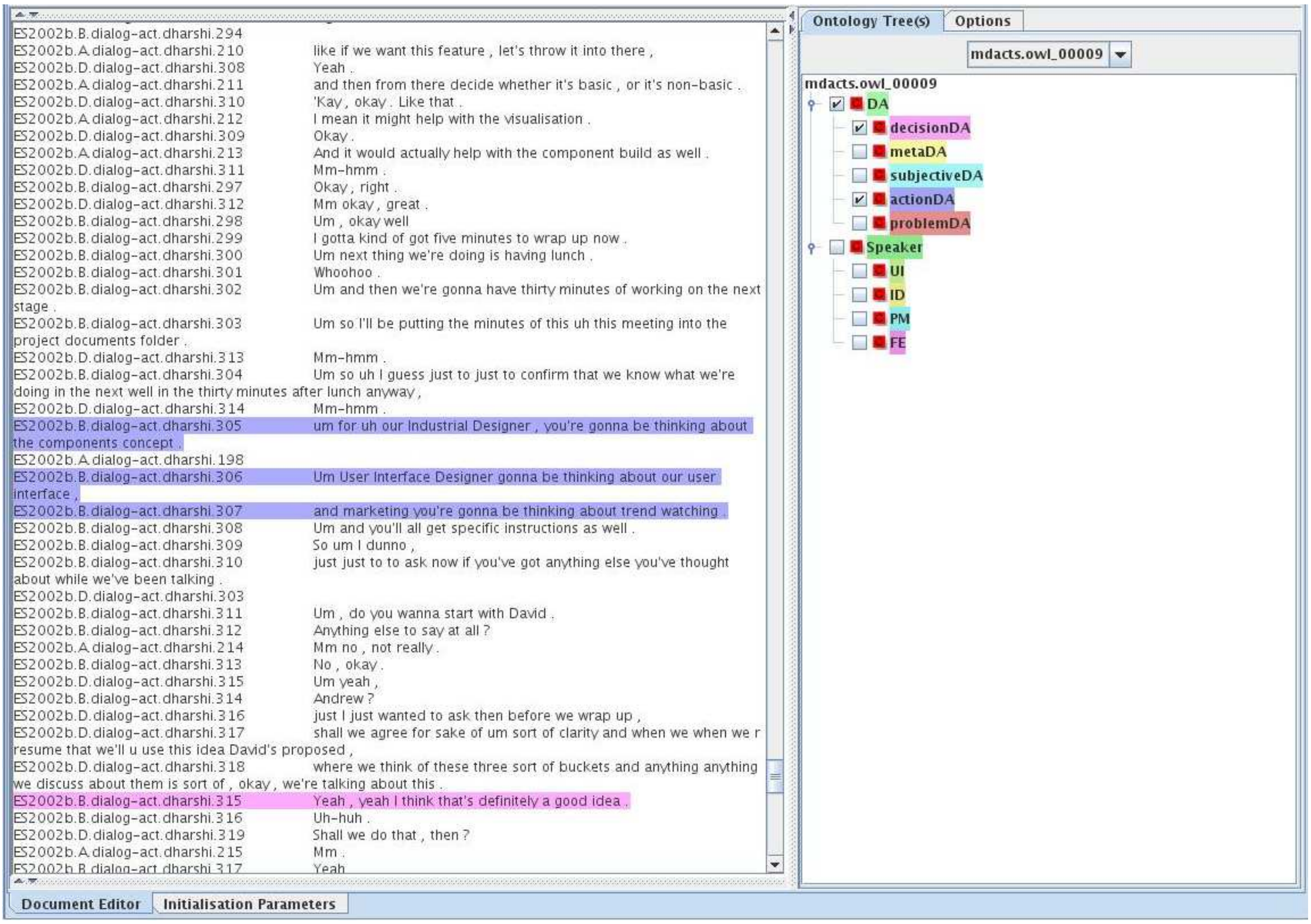

Figure 1. Screenshot of our interface for creating visual, structured summaries of human conversation

the conversation participants (Speaker in the figure) [2]. The left panel of the interface displays the whole conversation, where sentences are temporally ordered. Given the information displayed in the two panels, the user can generate visual, structured summaries by selecting nodes in the ontology. As a result, the sentences that were mapped in the selected nodes will be highlighted.

For instance, the left panel in Figure 1 displays a sample meeting from the AMI corpus whose sentences have been classified and mapped in the conversation ontology. In the example, since the user has selected the nodes *decision* and *action* in the ontology, the sentences mapped in those nodes are highlighted in the context of the whole conversation. In the current interface each node is associated with a different color and a sentence mapped into multiple selected nodes is colored as the "intersection" of the corresponding colors. This solution is not satisfactory and we are investigating more effective techniques to visually convey this information.

[2]We are currently adding to the interface nodes for all the entities extracted from the conversation (as described in the previous section).

Our initial prototype builds on a component of the GATE system [1], which was originally developed as a tool for text annotation.

CONCLUSIONS AND FUTURE WORK

This paper presents an interactive interface to create visually structured summaries of human conversations via ontology mapping. So far, we have built highly accurate classifiers for the mapping phase, that, in contrast with previous work, do not rely on features specific of any particular conversational modality. We have also implemented a first prototype of the interface that display both the ontology and the conversation, and allows the user to search the conversation based on the ontology mapping.

In the near future we plan to complete the development of the prototype. First, we are currently extending the displayed ontology to also include the entities mentioned in the conversation. Second, we will study how to effectively highlight sentences that were mapped to multiple nodes in the ontology. Once the summarization interface is completed, we intend to perform an extrinsic evaluation, in a way similar to [7, 10, 13].

REFERENCES

1. K. Bontcheva, V. Tablan, D. Maynard, and H. Cunningham. Evolving GATE to Meet New Challenges in Language Engineering. *Natural Language Engineering*, 10, 2004.

2. G. Carenini, R. Ng, and X. Zhou. Summarizing email conversations with clue words. In *Proc. of ACM WWW 07, Banff, Canada*, 2007.

3. J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, I. McCowan, W. Post, D. Reidsma, and P. Wellner. The AMI meeting corpus: A pre-announcement. In *Proc. of MLMI 2005, Edinburgh, UK*, pages 28–39, 2005.

4. K. Church and W. Gale. Inverse document frequency IDF: A measure of deviation from poisson. In *Proc. of the Third Workshop on Very Large Corpora*, pages 121–130, 1995.

5. M. Galley. A skip-chain conditional random field for ranking meeting utterances by importance. In *Proc. of EMNLP 2006, Sydney, Australia*, pages 364–372, 2006.

6. S. Gupta, J. Niekrasz, M. Purver, and D. Jurafsky. Resolving ”*You*” in multi-party dialog. In *Proc. of SIGdial 2007, Antwerp, Belgium*, 2007.

7. L. He, E. Sanocki, A. Gupta, and J. Grudin. Auto-summarization of audio-video presentations. In *Proc. of ACM MULTIMEDIA ’99, Orlando, FL, USA*, pages 489–498, 1999.

8. P.-Y. Hsueh, J. Kilgour, J. Carletta, J. Moore, and S. Renals. Automatic decision detection in meeting speech. In *Proc. of MLMI 2007, Brno, Czech Republic*, 2007.

9. J. Kupiec, J. Pederson, and F. Chen. A trainable document summarizer. In *Proc. of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Seattle, Washington, USA*, pages 68–73, 1995.

10. K. McKeown, J. Hirschberg, M. Galley, and S. Maskey. From text to speech summarization. In *Proc. of ICASSP 2005, Philadelphia, USA*, pages 997–1000, 2005.

11. C. Muller. Resolving *It*, *This* and *That* in unrestricted multi-party dialog. In *Proc. of ACL 2007, Prague, Czech Republic*, 2007.

12. G. Murray and G. Carenini. Interpretation and transformation for abstracting conversations. In *Submitted to the 2010 North American ACL*, 2010.

13. G. Murray, T. Kleinbauer, P. Poller, S. Renals, T. Becker, and J. Kilgour. Extrinsic summarization evaluation: A decision audit task. In *Proc. of MLMI 2008, Utrecht, the Netherlands*, 2008.

14. G. Murray and S. Renals. Detecting action items in meetings. In *Proc. of MLMI 2008, Utrecht, the Netherlands*, 2008.

15. M. Purver, J. Dowding, J. Niekrasz, P. Ehlen, and S. Noorbaloochi. Detecting and summarizing action items in multi-party dialogue. In *Proc. of the 9th SIGdial Workshop on Discourse and Dialogue, Antwerp, Belgium*, 2007.

16. S. Raaijmakers, K. Truong, and T. Wilson. Multimodal subjectivity analysis of multiparty conversation. In *Proc. of EMNLP 2008, Honolulu, HI, USA*, 2008.

17. O. Rambow, L. Shrestha, J. Chen, and C. Lauridsen. Summarizing email threads. In *Proc. of HLT-NAACL 2004, Boston, USA*, 2004.

18. J. Ulrich, G. Murray, and G. Carenini. A publicly available annotated corpus for supervised email summarization. In *Proc. of AAAI EMAIL-2008 Workshop, Chicago, USA*, 2008.

19. S. Xie, B. Favre, D. Hakkani-Tür, and Y. Liu. Leveraging sentence weights in a concept-based optimization framework for extractive meeting summarization. In *Proc. of Interspeech 2009, Brighton, England*, 2009.

20. L. Zhou and E. Hovy. Digesting virtual ”geek” culture: The summarization of technical internet relay chats. In *Proc. of ACL 2005, Ann Arbor, MI, USA*, 2005.

21. X. Zhu and G. Penn. Summarization of spontaneous conversations. In *Proc. of Interspeech 2006, Pittsburgh, USA*, pages 1531–1534, 2006.