Visualizing links between sentences in asynchronous conversations: a new feature for ConVis

CPSC 547 2015W1 project proposal
Jordon Johnson (jordon@cs.ubc.ca)

1. Domain

The domain is Natural Language Processing (NLP), a subdomain of Artificial Intelligence (AI). More specifically, this project deals with the study of asynchronous conversations (threads of emails, forum posts, or blog comments). Various NLP techniques, such as topic modeling, sentiment analysis or discourse parsing, can be applied to these conversations in order to extract meaning from them. Many aspects of such conversations can be visualized using UBC's ConVis [Hoque and Carenini 14] (see Figure 1).
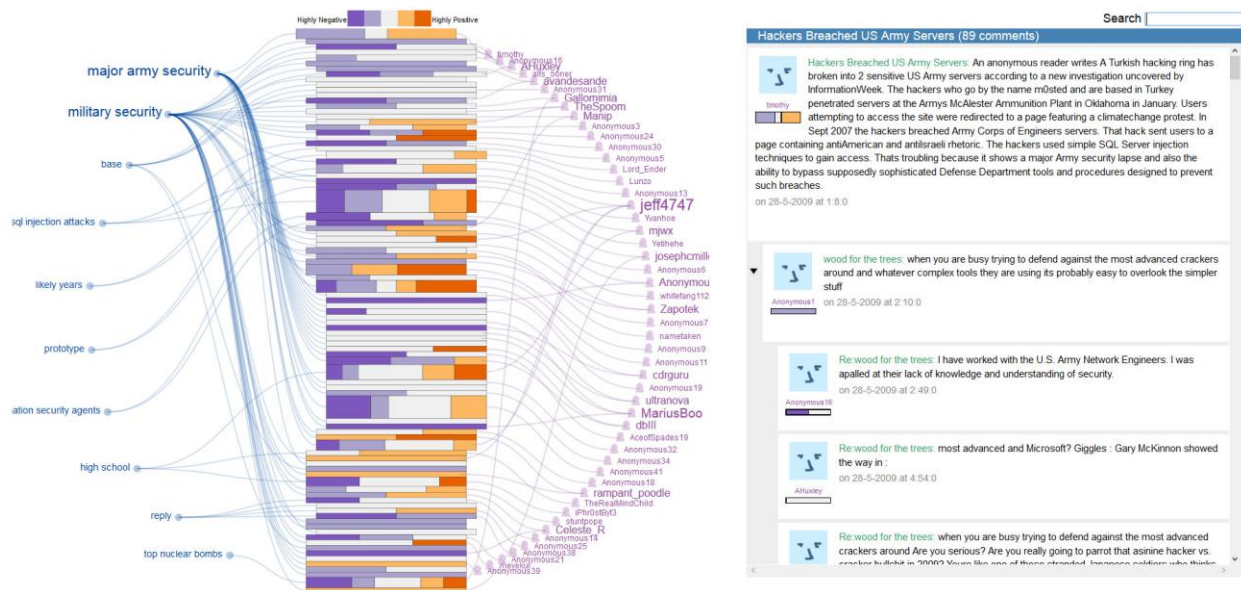

Figure 1. The default UI of ConVis. Image source: http://www.cs.ubc.ca/cs-research/lci/research-groups/natural-language-processing/image/convis/3.jpg

2. Task

This project relates to the endeavour to accurately find "links", where a given sentence relates to an earlier sentence in the conversation; such links may prove useful in tasks such as conversation classification and summarization. Current work involves establishing a reliable "gold standard" and evaluating candidate link-generating algorithms. Future work may involve finding patterns in the link attributes as they form "chains" and relating those patterns to different conversation types.

Given an asynchronous conversation annotated with links, an infoVis solution should:
- Summarize the entire set of links for the conversation

- Edit links (create new links; modify attributes of, annotate, or delete existing links)
- Discover features in the graph formed by the links (specifically, discover attribute patterns in "chains" of links)

3. Dataset

The dataset consists of ten articles from The Guardian and their associated comments, provided as text files with sentence indices for convenience and XML tags indicating conversational structure. Each file has been annotated by humans to include links between sentences in the comments and related sentences in the article or ancestor comments. These annotations are intended to be a gold standard (though some examination and revision may be warranted).

Each link contains the following relevant attributes:

com_sen, art_sen	Ordered variables representing indices of sentences involved in a given link. For simplicity I will use these attributes to refer to the relevant sentences themselves. Thus, we can describe a link as "com_sen is related to art_sen."
argument	Diverging variable indicating agreement of com_sen to art_sen. May take the values "in_favour", "impartial", or "against".
sentiment	Diverging variable indicating sentiment of com_sen towards art_sen. May take the values "positive", "neutral", or "negative".

Additional sets of links may be generated for these ten files by candidate linking algorithms during the course of the proposed project.

4. Personal expertise

I worked on the problem of determining links between sentences in asynchronous conversations during the summer of 2015 under an NSERC USRA grant. The research was exploratory and dealt largely with searching for relationships between the gold standard links provided and derivable attributes from topic modeling and discourse parsing.

I have no experience in implementing graphical interfaces.

5. Proposed solution

A toggle is added to ConVis that allows a user to switch between the display of commenters ("commenter view") and a new "link view"; the names of commenters and curves connecting them to various posts in the conversation are replaced with the links in the dataset.

The sentences and links can be thought of as a graph where the sentences are nodes, and the links are the edges. In this case, however, there is also a temporal relationship between the sentences, and so using the existing sequential representation of posts in ConVis is reasonable. Thus, the attributes of each link (see Section 3) are encoded as follows:

art_sen and **com_sen** are encoded as a curve from the post containing *com_sen* upwards to the post containing *art_sen*. Since ConVis encodes entire posts as individual glyphs, and there may be multiple sentences in a given post, the connection points of the links to a given post are ordered by vertical position, with the highest connection point associated with the first sentence in the post that has a link. When a link is selected, its curve is highlighted, and the sentences associated with *art_sen* and *com_sen* are displayed in the full-text view on the right.

argument and **sentiment** may be encoded within the link curves in one of two ways: as a bivariate colour map; or (more likely) with *sentiment* encoded using colour (for consistency with the rest of ConVis) and *argument* encoded using shape (solid, dash-dotted, and dotted as *in_favour*, *impartial*, and *against*).

Annotation controls are found in the full text view on the right. They include *art_sen* and *com_sen* in text fields as indices, and *argument* and *sentiment* in drop-down lists as their string values.

6. Scenario of use: editing/deleting an existing link

The user already has ConVis open, and she switches the toggle from commenter view to link view. The names of commenters and associated curves are replaced with curves encoding links, the full-text view is cleared of text, and the annotation controls appear (Fig. 2a). The user clicks on one of the curves; it is highlighted, the linked sentences appear in the full-text view, and the link attributes are shown in the annotation controls (Fig. 2b). If desired, she may change any of the attributes, and the relevant displayed curves and sentences in the full-text view update automatically. She may also choose to delete the link by clicking the appropriate button. Clicking on whitespace in the main view will de-select the link, returning the vis to the state shown in Fig. 2a.
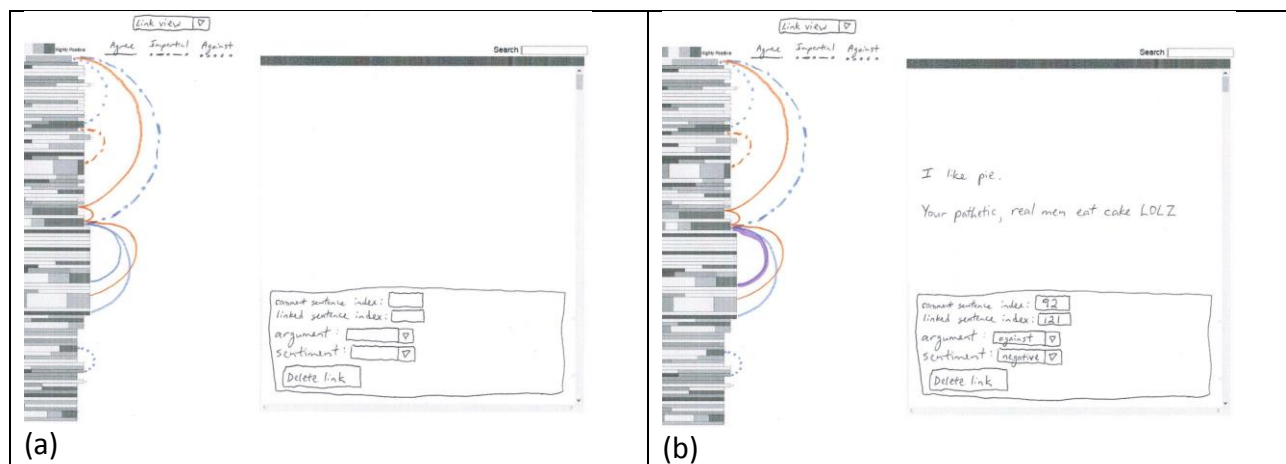

(a)   (b)

Figure 2. Depictions of the link view to be added to ConVis. (a) Upon selecting the link view, the links are displayed as curves, the full-text view is cleared, and the annotation controls appear. Clicking on whitespace in the link view area returns the vis to this state. (b) Selecting a link highlights it and populates the full-text view with the appropriate sentences and link attributes. The link attributes can be changed using the annotation controls.

7. Implementation approach

The proposed visualization will extend the functionality of the existing ConVis software; as such, it will be implemented chiefly using D3, JavaScript and PHP. While development will take place on the Linux platform, these languages should ensure cross-platform usability.

Schedule

Milestone (non-developmental tasks italicized)	Completion date
Ensure dataset is readable by ConVis	Nov 11
Create toggle between existing commenter view and link view	Nov 14
Submit project update	Nov 23
Draw links when link view is active	Nov 25
Modify ConVis full-text view to display the two sentences in a selected link when link view is active	Dec 2
Add annotation controls to full-text view when link view is active	Dec 9
Presentation	Dec 15
Submit paper	Dec 17

8. Related work

As mentioned earlier, this work extends the functionality of ConVis, a state-of-the-art tool for visualizing asynchronous conversations that encodes conversation topics, post sentiment, and user activity. ConVis does not yet support the visualization of links between individual sentences.

The OnForumS project [Kabadjov et al. 15] is an endeavour to combine summarization, argument mining and sentiment mining into a single shared task, of which a portion is creating algorithms to find links between sentences.

The Constellation system [Munzner 00, Munzner et al. 99; cited in Munzner 2014] shares a domain (NLP) and high-level task (evaluating candidate algorithms) with this project; however, there may prove to be a long-term use for the proposed extension to ConVis if patterns in link chains can be found and characterized.

The TIARA system [Liu et al. 12] provides a visualization of topics extracted from large collections of documents using streamgraphs; however, its scope is not suited to relationships between individual sentences.

9. References

Enamul Hoque and Giuseppe Carenini, ConVis: A Visual Text Analytic System for Exploring Blog Conversations, Journal of Computer Graphics Forum (Proceedings of the EuroVis), 2014.

Mijail Kabadjov et al., OnForumS: A Shared Task on On-line Forum Summarisation, FIRE 2015 (to appear).

S. Liu et al., TIARA: Interactive, Topic-Based Visual Text Summarization and Analysis, ACM 2014.

Tamara Munzner, Visualization Analysis and Design. CRC Press, 2014.