From Generating to Mining: Automatically Scripting Conversations Using Existing Online Sources

Nathan Nichols, Lisa Gandy, and Kristian Hammond

Intelligent Information Laboratory
Electrical Engineering and Computer Science, Northwestern University
2133 Sheridan Rd., Evanston, IL 60208
ndnichols@cs.northestern.edu, lisagandy@u.northwestern.edu, hammond@cs.northwestern.edu

Abstract

Hearing people argue opposing sides of an issue can be a useful way to understand the topic; however, these debates or conversations often don't exist. Unfortunately, generating interesting natural language conversations is a difficult problem and typically requires a deep model of both a domain and its language. Fortunately, there is a huge amount of interesting text, written both by professional writers and amateurs, already available on the web. In this paper, we describe a system that builds compelling conversations between two characters—not by generating wholly new natural language, but by gathering, assembling, and processing existing online textual content. Our initial system authors conversations between two simulated movie reviewers, in a style similar to "Siskel and Ebert." Using various online repositories, the system searches for a variety of facts and opinions about a given film. The system then uses this mined data to choose between various conversational templates and construct the dialogue. Using this information, the system is able to generate natural-sounding, colorful conversations and arguments without a deep representation of the subject being discussed.

Generating conversations

Natural language generation systems typically begin with high-level communication goals, break the goals down into subgoals that can be atomically expressed in speech, plan the sentences to achieve these communication goals, and finally embody these subgoals in natural language (Reiter, 1994.) This approach requires an elaborate language model and a rich representation of the system's domain. For some domains, this representation may be readily available; the inherently quantitative nature of stock trading, for example, would make a basic system that generated a natural language summary of the day's trading reasonably straightforward to implement.

A representation for the domain of movie reviews, however, would need to contain detailed knowledge of the specific movie being reviewed, information about the actors, and general knowledge of film. This amount of knowledge and representation would be extremely difficult to create even once, and it would need to be updated regularly as new films are released.

Fortunately, there are alternative methods to generating simulated conversation between two characters that require very little explicit representation. Instead of creating completely new sentences to achieve communicative goals, the system can instead mine for *existing utterances* that already express the desired communicative intent. Of course, determining the communicative intent of an utterance can also be a very challenging problem; this is why we first apply our technique to the movie review domain, where relatively unambiguous statements are common. Once we have shown our approach succeeds on movie reviews, we will expand it to more varied domains.

Preparing a script

The system begins with the name of a movie, which it uses to lookup information on the film from various websites. The Internet Movie Database is used to create explicit but light-weight representations of the movie and the actors in it. More importantly, the system mines movie review aggregation sites, like Metacritic and Rotten Tomatoes, to generate a list of labelled review snippets. These snippets are short excerpts taken from a full review; because the snippets are chosen by human editors, they are consistently on-point and indicative of the full review. The reviews are also paired with the explicit score the reviewer gave the film. The review snippets and scores pulled from these sites are a mixture of professional critiques and amateur reviews submitted by users.

The system uses the review snippets' associated numerical score to choose between templates that contain the basic framework of the conversation. We have built three templates: one is used when critics disagree about the movie, one when critics all liked the movie, and one when critics all disliked the movie. Because small sections of each template are randomized, and so much of the eventual conversations are incorporated from the external review snippets, the system is able to generate a large number of reviews without sounding repetitive.

The system performs the final preparation step by constructing a collection of "review sets." Each review set

is a collection of reviews about a certain aspect of the movie. For example, a review that says, "Heath Ledger is terrific as the Joker in The Dark Knight," would be a member of the "Heath Ledger," "Joker," and "The Dark Knight" review sets.

Scripting a review

Once the gathering and preparation steps are completed, the system actually generates a conversation by filling in the chosen template. Some of the slots are trivial to fill, but most are more complex and require various operations on the review sets. For example, a conversational template for when the reviewers disagree on the film may incorporate logic like "Anchor A says a positive review about the star of the film, Anchor B disagrees and says a negative review about the star." Another template may have logic such as "Anchor A says a positive review of the movie, Anchor B agrees and reads a positive review that has few words in common with the review A just read" (so the anchors don't repeat each other.)

The review sets for a movie are all linked, so if a certain review is used to satisfy one requirement, it won't be used again to satisfy another. The review sets are also arranged in a small hierarchy, with the review set for the movie itself at the root, and more specific review sets further down the branches. With this organization, if the template calls for a negative review of an actor, but there are none, the system can instead insert a negative review of the movie itself and the conversation still sounds natural.

Below is an example of a script generated entirely by the system. This particular script incorporates seven different reviews and seven pieces of information from IMDB. The two anchors, Zack and Zooey, are reviewing the film "Valkyrie" starring Tom Cruise.

> Zack: Tom Cruise stars with Carice van Houten and Kenneth Branagh in the film Valkyrie. In this PG-13 rated drama, based on actual events, a plot to assassinate Hitler is unfurled during the height of WWII. I have to tell you, I loved this movie. The movie works like a clock. A few minor quibbles aside (the casting of Hitler, for instance), Valkyrie is a highly intelligent and deeply engrossing historical drama and, frame for frame, the year's most suspenseful nail-biter. Singer has a masterful touch with composition, creating tension simply by the way he places his actors around a room. And even though we know how it turned out, the assassination plot remains a gripping tale.
>
> Zooey: No way! Bryan Singer's long-awaited account of the near-miss assassination of Adolf Hitler by a ring of rebel German army officers on July 20, 1944, has visual splendor galore, but is a cold work lacking in the requisite tension and suspense. What you miss in both Defiance and Valkyrie is inner conflict. Their protagonists have not an instant of self-doubt. They're figures in historical pageants, not characters in a drama.
>
> Zack: Aww, come on. You had to love Tom Cruise. Mr. Cruise's performance turns out to be brisk and reasonably plausible, though unexceptional, while the production as a whole succeeds as an elaborate procedural, impressively staged in historical locations. Tom Cruise is perfectly satisfactory, if not electrifying, in the leading role.
>
> Zooey: No. If there are Nazis fighting other Nazis in a movie and it's still boring, something's gone wrong. Valkyrie has a coterie of problems, and represents a whole new front in Tom Cruise's public relations war, but first and foremost there's the tedium.

Even without a deep representation of the film, or any real concept of movie reviewing in general, the system is still able to assemble very plausible-sounding and interesting conversations. The script generated by the system is eventually used to drive a movie review segment in the News at Seven system (Nichols & Hammond, 2008).

Future work

First, we want to expand where the system searches for the review snippets; this includes mining an entire review (instead of just the editor-chosen snippets) and going directly to the blogosphere to find reviews people are posting to their personal blogs. Not using editor-chosen snippets from professional reviewers introduces a number of difficulties. The system will need to be able to extract useful, representative snippets from reviews, as well as assign ratings to reviews that people don't explicitly rate themselves. Fortunately, these are familiar information extraction problems and there are a number of techniques we can bring to bear.

Second, we hope to expand the domains for which the system can generate conversations. Politics and political debates are a natural target for a system that can generate conversations and arguments, as are product reviews.

Finally, we hope to incorporate social networking to create personalized conversations. The movie review dynamic, for example, could incorporate friends' opinions on the film pulled from Facebook and Twitter.

Acknowledgements

We are grateful to the National Science Foundation for sponsoring this work under grant number 0535231.

References

Nichols, N. and Hammond, K. 2008. "Machine-Generated Multimedia Content." In the Proceedings of the Second International Conference on Advances in Computer-Human Interactions.

Reiter, E. "Has a consensus NL generation architecture appeared, and is it psycholinguistically plausible?" In Proceedings of the 7th International Workshop on Natural Language Generation, pages 163-170, Maine, 1994.