Dialogue Acts

Bin Zhang

February 1, 2011
Outline

A. Stolcke et al. 2000, Dialogue act modeling for automatic tagging and recognition of conversational speech

D. Davidov et al. 2010, Semi-supervised recognition of sarcastic sentences in Twitter and Amazon
Dialogue Act

- It is a specialized speech act
- Typical dialogue acts
  - Statement
  - Question
  - Backchannel
  - Agreement
  - Disagreement
  - Apology
### Examples of Dialogue Acts

**Table 1**
Fragment of a labeled conversation (from the Switchboard corpus).

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Dialogue Act</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>YES-NO-QUESTION</td>
<td>So do you go to college right now?</td>
</tr>
<tr>
<td>A</td>
<td>ABANDONED</td>
<td>Are yo-,</td>
</tr>
<tr>
<td>B</td>
<td>YES-ANSWER</td>
<td>Yeah,</td>
</tr>
<tr>
<td>B</td>
<td>STATEMENT</td>
<td>it’s my last year [laughter].</td>
</tr>
<tr>
<td>A</td>
<td>DECLARATIVE-QUESTION</td>
<td>You’re a, so you’re a senior now.</td>
</tr>
<tr>
<td>B</td>
<td>YES-ANSWER</td>
<td>Yeah,</td>
</tr>
<tr>
<td>B</td>
<td>STATEMENT</td>
<td>I’m working on my projects trying to graduate [laughter].</td>
</tr>
<tr>
<td>A</td>
<td>APPRECIATION</td>
<td>Oh, good for you.</td>
</tr>
<tr>
<td>B</td>
<td>BACKCHANNEL</td>
<td>Yeah.</td>
</tr>
<tr>
<td>A</td>
<td>APPRECIATION</td>
<td>That’s great,</td>
</tr>
<tr>
<td>A</td>
<td>YES-NO-QUESTION</td>
<td>um, is, is N C University is that, uh, State,</td>
</tr>
<tr>
<td>B</td>
<td>STATEMENT</td>
<td>N C State.</td>
</tr>
<tr>
<td>A</td>
<td>SIGNAL-NON-UNDERSTANDING</td>
<td>What did you say?</td>
</tr>
<tr>
<td>B</td>
<td>STATEMENT</td>
<td>N C State.</td>
</tr>
</tbody>
</table>
Dialogue Act Labeling

- Unlabeled dialogue data from Switchboard corpus
  - Contains conversational telephone speech
  - Speech is segmented into utterances
  - Each utterance has a dialogue act label
Unlabled dialogue data from Switchboard corpus
  ▶ Contains conversational telephone speech
  ▶ Speech is segmented into utterances
  ▶ Each utterance has a dialogue act label

Tag set
  ▶ Dialogue act markup in several layers (DAMSL)
  ▶ SWBD-DAMSL tag set: 50 tags original, reduced to 42 tags for better annotator consistency
Unlabeled dialogue data from Switchboard corpus
- Contains conversational telephone speech
- Speech is segmented into utterances
- Each utterance has a dialogue act label

Tag set
- Dialogue act markup in several layers (DAMSL)
- SWBD-DAMSL tag set: 50 tags original, reduced to 42 tags for better annotator consistency

1115 conversations (205000 utterances) annotated by 8 linguistic graduate students in 3 months, $\kappa = 0.80$ (excellent agreement)
Common Dialogue Act Types

**Statements**  descriptive, narrative, or personal statements

**Opinions**  often include such hedges as *I think, I believe, it seems, and I mean*

**Questions**  yes-no questions, declarative questions, WH questions

**Backchannels**  short utterances that play discourse-structuring roles, e.g., indicating that the speaker should go on talking

**Abandoned utterances**  those that the speaker breaks off without finishing, and are followed by a restart

**Turn exits**  similar to abandoned utterances, but with speaker change

**Answers**  yes answers and no answers

**Agreements**  mark the degree to which a speaker accepts some previous proposal, plan, opinion, or statement
Sequence Modeling of Dialogue Acts

- A conversation can be considered as a sequence of utterances
- Dialogue acts of the utterances in a conversation are inter-dependent
Sequence Modeling of Dialogue Acts

- A conversation can be considered as a sequence of utterances
- Dialogue acts of the utterances in a conversation are inter-dependent
- Denote $U$ as the dialogue act sequence of the conversation, and $E$ as the evidence about the conversation (audio and/or text)

\[
U^* = \arg\max_U P(U|E)
\]
Sequence Modeling of Dialogue Acts

- A conversation can be considered as a sequence of utterances.
- Dialogue acts of the utterances in a conversation are inter-dependent.
- Denote $U$ as the dialogue act sequence of the conversation, and $E$ as the evidence about the conversation (audio and/or text).

$$U^* = \arg\max_U P(U|E)$$

- Using Bayes rule

$$U^* = \arg\max_U P(E|U)P(U)$$

$P(E|U)$ dialogue act likelihood
$P(U)$ discourse grammar
Dialogue Act HMM

- Markov assumption

\[ P(U_i | U_1, U_2, \ldots, U_{i-1}) = P(U_i | U_{i-k}, \ldots, U_{i-1}) \]

- Independence assumption

\[ P(E | U) = \prod_i P(E_i | U_i) \]
Dialogue Act HMM

- Markov assumption

\[ P(U_i|U_1, U_2, \ldots, U_{i-1}) = P(U_i|U_{i-k}, \ldots, U_{i-1}) \]

- Independence assumption

\[ P(E|U) = \prod_i P(E_i|U_i) \]

\[ \begin{align*}
E_1 & \quad \quad \quad E_i & \quad \quad \quad E_n \\
\uparrow & \quad \quad \quad \uparrow & \quad \quad \quad \uparrow \\
<\text{start}> & \quad \rightarrow U_1 & \quad \rightarrow \cdots & \rightarrow U_i & \quad \rightarrow \cdots & \rightarrow U_n & \rightarrow <\text{end}> 
\end{align*} \]

**Figure 1**
The discourse HMM as Bayes network.
Discourse Grammar

- As is the nature of dialogues, the dialogue acts of the utterances in a dialogue are highly related
Discourse Grammar

- As is the nature of dialogues, the dialogue acts of the utterances in a dialogue are highly related.
- Back-off dialogue act n-gram models as a simple and efficient discourse model.
As is the nature of dialogues, the dialogue acts of the utterances in a dialogue are highly related.

Back-off dialogue act n-gram models as a simple and efficient discourse model.

If speaker is known, speaker-dependent dialogue act n-gram models are better.
Combining Evidence

- Words ($W$)
- ASR acoustics ($A$)
- Prosodic features ($F$): pitch, duration, energy, etc., of the speech signal
Combining Evidence

- Words ($W$)
- ASR acoustics ($A$)
- Prosodic features ($F$): pitch, duration, energy, etc., of the speech signal

$$P(W, A, F|U)$$

<start> $\longrightarrow$ $U_1$ $\longrightarrow$ $\cdots$ $\longrightarrow$ $U_i$ $\longrightarrow$ $\cdots$ $\longrightarrow$ $U_n$ $\longrightarrow$ <end>

$A_1$ $\uparrow$ $A_i$ $\uparrow$ $A_n$ $\uparrow$

$W_1$ $\uparrow$ $W_i$ $\uparrow$ $W_n$ $\uparrow$

$F_1$ $\uparrow$ $F_i$ $\uparrow$ $F_n$ $\uparrow$

Figure 4
## Experimental Results

<table>
<thead>
<tr>
<th>Discourse Grammar</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prosody</td>
</tr>
<tr>
<td>None</td>
<td>38.9</td>
</tr>
<tr>
<td>Unigram</td>
<td>48.3</td>
</tr>
<tr>
<td>Bigram</td>
<td>49.7</td>
</tr>
</tbody>
</table>
Outline

A. Stolcke et al. 2000, Dialogue act modeling for automatic tagging and recognition of conversational speech

D. Davidov et al. 2010, Semi-supervised recognition of sarcastic sentences in Twitter and Amazon
Sarcasm – *the activity of saying or writing the opposite of what you mean, or of speaking in a way intended to make someone else feel stupid or show them that you are angry*
Example of Twitter

Listening to some groovy music while I wait for the moderator to start my conference call.

38 minutes ago from web

Abdur is sending me Gao Bagua YouTube videos. Suddenly, I fear him. about 1 hour ago from web

Up, out, thinking. about 3 hours ago from txt

Making a lot of lists. about 14 hours ago from web

Putting on a mic. about 24 hours ago from txt

I have $3.50 in my pocket. 07:39 AM September 04, 2008 from txt

Trimming the sails: http://tinyurl.com/6j3ee6 10:59 AM September 03, 2008 from web

A woman to Britt: "Are you wearing Hermes?" 09:03 AM September 03, 2008 from txt
Example of Amazon

Secrets of a Jewish Mother: Real Advice, Real Stories, Real Love
by Lisa Wexler
Edition: Hardcover
Price: $11.69
Availability: In Stock

3 of 50 people found the following review helpful:

⭐⭐⭐⭐⭐ I've been waiting to talk to my mom for a long time... April 15, 2010

Amazon Verified Purchase

This review is from: Secrets of a Jewish Mother: Real Advice, Real Stories, Real Love (Hardcover)

This book is exactly what every girl needs. I felt like I was hearing my own mother give me the advice I need. I will have this in my library forever.

I am disgusted by what other people wrote on this site. MEAN. They never even read the book. Scary how they believe everything they see. The book is amazing and they are just jealous. So sad. Good Luck Jill, Gloria and Lisa...I hope you make #1 on NYT List!

UPDATE:

I feel sorry for those who have spent hours writing mean ugly things about Jill and her family. They are so jealous. This book is perfect for so many people. They will see right through the haters who have nothing else to do all day.

Comments (12) | Permalink | Most recent comment: Apr 20, 2010 7:30 PM PDT

Reviewer's Tags: advice, dating, dating advice, finance, friendship, humor, marriage, marriage advice, money, personal transformations, real housewives

The Real Housewives of New York City Reunion: Watch What Happens
DVD
Price: $1.99
Availability: In Stock

0 of 2 people found the following review helpful:

⭐⭐⭐⭐⭐ Love it., June 22, 2008

This review is from: The Real Housewives of New York City (Video On Demand)

Who would 't love a show that has it all. Great characters, NYC and lots of stuff. Without these women..I wouldn't watch. I love Jill because she is the connector and an honest character. I don't think I will watch the show if at least Jill and Ramona and Bethenny don't come back. This is one of the few great shows on TV. It reminds me of the Dallas and Dynasty rivalry's but real life. I will not watch if they don't come back.

Comments (1) | Permalink | Most recent comment: Dec 3, 2008 6:23 PM PST

Reviewer's Tags: reality tv
Sarcastic Sentences in Two Genres

- In Twitter messages, sarcastic sentences appear in a wide range. Some sarcastic sentences are marked #sarcasm by the user.
Sarcastic Sentences in Two Genres

- In Twitter messages, sarcastic sentences appear in a wide range. Some sarcastic sentences are marked #sarcasm by the user.

- In Amazon product reviews, sarcastic sentences are usually with the negative reviews.
Examples of Sarcasm

- thank you Janet Jackson for yet another year of Super Bowl classic rock! (Twitter)
- *Hes with his other woman: XBox 360. Its 4:30 fool. Sure I can sleep through the gunfire* (Twitter)
- *Wow GPRS data speeds are blazing fast* (Twitter)
- *[I] Love The Cover* (book, amazon)
- *Defective by design* (music player, amazon)
The Semi-supervised Approach

- Data processing
- Pattern extraction
- Pattern selection
- Data enrichment
- Classification
Data Processing

- For Twitter: actual user, URL, and hashtags tokenized by [USER], [URL], [HASHTAG], respectively
- For Amazon: product, author, company, and book name tokenized by [PRODUCT], [AUTHOR], [COMPANY], [TITLE], respectively
Construct pattern templates

- Words are classified into high-frequency words and content words (less frequent)
- Punctuations considered high-frequency words
- Templates are created using both word classes from knowledge
Pattern Extraction

- Construct pattern templates
  - Words are classified into high-frequency words and content words (less frequent)
  - Punctuations considered high-frequency words
  - Templates are created using both word classes from knowledge

- Instantiation
  - Templates are instantiated from the data, with high-frequency words replaces by actual words
  - Hundreds of patterns collected
  - During matching, partial matching is allowed
Pattern Selection

- Remove patterns that originated from one book/product (Amazon)
- Remove patterns that occur in both classes
Data Enrichment

- Used to get more labeled data
- Motivation: sarcastic sentences usually co-occur
- Use labeled sentences as query to search for more sentences on the web
- Found sentences are assigned similar label to the query
Classification

- Feature vectors are composed of:
  - Match score of each pattern in the sentences
  - Additional punctuation-based features
- k-nearest neighbor classifier is used
Data Annotation

- Sentences are annotated using five labels (1, 2, 3, 4, 5), 1 being not sarcastic, 5 being clearly sarcastic. In classification, 1, 2 are negative labels, 3, 4, 5 are positive labels.
Data Annotation

- Sentences are annotated using five labels (1, 2, 3, 4, 5), 1 being not sarcastic, 5 being clearly sarcastic. In classification, 1, 2 are negative labels, 3, 4, 5 are positive labels.
- The Twitter #sarcasm hashtag is found to be biased and noisy.
Data Annotation

- Sentences are annotated using five labels (1, 2, 3, 4, 5), 1 being not sarcastic, 5 being clearly sarcastic. In classification, 1, 2 are negative labels, 3, 4, 5 are positive labels.
- The Twitter #sarcasm hashtag is found to be biased and noisy.
- Inter-annotator agreement (fair agreement)
  - Twitter: $\kappa = 0.41$
  - Amazon: $\kappa = 0.34$
Sentences are annotated using five labels (1, 2, 3, 4, 5), 1 being not sarcastic, 5 being clearly sarcastic. In classification, 1, 2 are negative labels, 3, 4, 5 are positive labels.

The Twitter #sarcasm hashtag is found to be biased and noisy.

Inter-annotator agreement (fair agreement)

- Twitter: $\kappa = 0.41$
- Amazon: $\kappa = 0.34$

Amazon training data

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed</td>
<td>80</td>
<td>505</td>
</tr>
<tr>
<td>Enriched</td>
<td>471</td>
<td>5020</td>
</tr>
</tbody>
</table>

Evaluation data: for each genre 90 positive + 90 negative sentences.
Experimental Results

- Baseline (star-sentiment, Amazon only): low rating reviews with string positive sentiment
Experimental Results

- Baseline (star-sentiment, Amazon only): low rating reviews with string positive sentiment

<table>
<thead>
<tr>
<th></th>
<th>Prec.</th>
<th>Recall</th>
<th>FalsePos</th>
<th>FalseNeg</th>
<th>F Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Star-sent.</td>
<td>0.5</td>
<td>0.16</td>
<td>0.05</td>
<td>0.44</td>
<td>0.242</td>
</tr>
<tr>
<td>SASI (AM)</td>
<td>0.766</td>
<td>0.813</td>
<td>0.11</td>
<td>0.12</td>
<td>0.788</td>
</tr>
<tr>
<td>SASI (TW)</td>
<td>0.794</td>
<td>0.863</td>
<td>0.094</td>
<td>0.15</td>
<td>0.827</td>
</tr>
</tbody>
</table>

Table 3: Evaluation on the Amazon (AM) and the Twitter (TW) evaluation sets obtained by averaging on 3 human annotations per sentence. TW results were obtained with cross-domain training.
Thanks!
Improving Speech Recognition using Dialogue Acts

- Word recognition (with evidence)
  \[ W_i^* = \arg\max_{W_i} P(W_i|A_i, E) \]

- Dialogue acts can be considered as a factor that the acoustic and language models depend on

- When the dialogue acts are unknown, the models are the mixtures of dialogue act-dependent models
  - Mixture-of-posteriors
    \[ P(W_i|A_i, E) = \sum_{U_i} \frac{P(W_i|U_i)P(A_i|W_i)}{P(A_i|U_i)} P(U_i|E) \]
  - Mixture-of-LMs
    \[ P(W_i|A_i, E) \approx \sum_{U_i} P(W_i|U_i)P(U_i|E) \frac{P(A_i|W_i)}{P(A_i)} \]
# Experimental Results

<table>
<thead>
<tr>
<th>Model</th>
<th>WER (%)</th>
<th>Perplexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>41.2</td>
<td>76.8</td>
</tr>
<tr>
<td>1-best LM</td>
<td>41.0</td>
<td>69.3</td>
</tr>
<tr>
<td>Mixture-of-posteriors</td>
<td>41.0</td>
<td>n/a</td>
</tr>
<tr>
<td>Mixture-of-LMs</td>
<td>40.9</td>
<td>66.9</td>
</tr>
<tr>
<td>Oracle LM</td>
<td>40.3</td>
<td>66.8</td>
</tr>
</tbody>
</table>

**Table 11**
Switchboard word recognition error rates and LM perplexities.