Features for Modelling Characteristics of Conversations

Gunnar Eriksson  Jussi Karlgren

2012-09-20
Gavagai AB
Stockholm, Sweden
Identifying sexual groomers in conversations

What information to use?
What information to use?

- what content in conversations?
Identifying sexual groomers in conversations

What information to use?

- what content in conversations?
- what behaviour within conversations?
Identifying sexual groomers in conversations

What information to use?

- what content in conversations?
- what behaviour within conversations?
- what types of conversations?
Identifying sexual groomers in conversations

Feature types

Conversation content
- words and short phrases

Conversation behaviour
- turn-taking – the flow of interaction

Conversation type
- number of participants and conversation length
Identifying sexual groomers in conversations

Feature types

Conversation content

words and short phrases
Identifying sexual groomers in conversations

Feature types

**Conversation content**
- words and short phrases

**Conversation behaviour**
- turn-taking – the flow of interaction
Identifying sexual groomers in conversations

Feature types

**Conversation content**
- words and short phrases

**Conversation behaviour**
- turn-taking – the flow of interaction

**Conversation type**
- number of participants and conversation length
Train a maximum entropy classifier with all features from all conversations the author took part in. Some exceptions:

- do not use features with occurrence with only one author.
- remove feature frequency information.
Extracting lexical features

Lexical features

- Normalisation and tokenisation
- Bag of tokens + bigrams for every participant
- Generalize mentions of chat nick-names of other conversation partakers to `OtherName`.
- Generalize mentions of nick-name of the "speaker" to `SelfName`.
Extracting lexical features

Lexical features

- normalisation and tokenisation
- tokens and token bigrams
- bag of tokens + bigrams for every participant

■ generalize mentions of chat nick-names of other conversation partakers to OtherName.

■ generalize mentions of nick-name of the "speaker" to SelfName.
Extracting lexical features

Lexical features

- normalisation and tokenisation
- tokens and token bigrams
- bag of tokens + bigrams for every participant
- generalize mentions of chat nick-names of other conversation partakers to **OtherName**.
- generalize mentions of nick-name of the “speaker” to **SelfName**.
More than one source to the content of a conversation!

**SLEX**  The things you utter in the conversation.
**OLEX**  The things others utter in the conversation.
Extracting lexical features

More than one source to the content of a conversation!

**SLEX** The things you utter in the conversation.

**OLEX** The things others utter in the conversation.

Add the two sets of features to every author profile.

In two ways: Concatenation (SLEX + OLEX) or union (CLEX).
Extracting lexical features

<table>
<thead>
<tr>
<th>Features types used</th>
<th>Precision</th>
<th>Recall</th>
<th>F ($\beta=1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>submitted</td>
<td>0.84</td>
<td>0.89</td>
<td>0.87</td>
</tr>
<tr>
<td>SLEX</td>
<td>0.21</td>
<td>0.84</td>
<td>0.33</td>
</tr>
<tr>
<td>OLEX</td>
<td>0.27</td>
<td>0.80</td>
<td>0.35</td>
</tr>
<tr>
<td>OLEX + SLEX</td>
<td>0.43</td>
<td>0.85</td>
<td>0.57</td>
</tr>
<tr>
<td>CLEX</td>
<td>0.48</td>
<td>0.72</td>
<td>0.58</td>
</tr>
<tr>
<td>CLEX + CONV features</td>
<td>0.56</td>
<td>0.55</td>
<td>0.55</td>
</tr>
</tbody>
</table>
Extracting conversation features

Conversation type, CTYPEx

Monologue
- 0 – 1 participant

Dialogue
- 2 participants

Group
- > 2 participants
Extracting conversation features

Conversation type, CTYPES

Monologue
0 – 1 participant

Dialogue
2 participants

Group
> 2 participants
Extracting conversation features

Conversation length, LTYPE

- fail: 0 – 1
- handshake: 2 – 7
- prelude: 8 – 25
- brief: 26 – 50
- discourse: 51 – 100
- ldiscourse: 101 – 160
- vldiscourse: 161 –
### Conversation length, LTYPE

<table>
<thead>
<tr>
<th>Feature</th>
<th># of utterances</th>
</tr>
</thead>
<tbody>
<tr>
<td>fail</td>
<td>0 – 1</td>
</tr>
<tr>
<td>handshake</td>
<td>2 – 7</td>
</tr>
<tr>
<td>prelude</td>
<td>8 – 25</td>
</tr>
<tr>
<td>brief</td>
<td>26 – 50</td>
</tr>
<tr>
<td>discourse</td>
<td>51 – 100</td>
</tr>
<tr>
<td>ldiscourse</td>
<td>101 – 160</td>
</tr>
<tr>
<td>vldiscourse</td>
<td>161 –</td>
</tr>
</tbody>
</table>
Adding ctype to ltype

Example

\(\text{prelude}_{\text{mono}}, \text{prelude}_{\text{dia}}, \text{or} \ \text{prelude}_{\text{group}}\)
\(\text{brief}_{\text{mono}}, \text{brief}_{\text{dia}}, \text{or} \ \text{brief}_{\text{group}}\)

...
Turn-taking behaviour, TTAKE

The flow of utterances within a conversation is modelled by turn-taking trigrams from each participant’s perspective.
Turn-taking behaviour, TTAKE

The flow of utterances within a conversation is modelled by turn-taking trigrams from each participant’s perspective.

**Example**

<table>
<thead>
<tr>
<th>participants</th>
<th>utterance 1</th>
<th>u2</th>
<th>u3</th>
<th>u4</th>
<th>u5</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Example**

A: Self–Self–Other, Self–Other–Self, Other–Self–Other, ...
B: Other–Other–Self, Other–Self–Other, Self–Other–Self, ...
Extracting conversation features

Adding ctype to ttake
Extracting conversation features

Adding ctype to ttake

Example

A: Self–Self–Other\textsubscript{g}, Self–Other\textsubscript{g}–Self, Other\textsubscript{g}–Self–Other\textsubscript{g}, ...

B: Other\textsubscript{g}–Other\textsubscript{g}–Self, Other\textsubscript{g}–Self–Other\textsubscript{g}, Self–Other\textsubscript{g}–Self, ...

Extracting conversation features

**Combining it all**

<table>
<thead>
<tr>
<th>Features types used</th>
<th>Precision</th>
<th>Recall</th>
<th>$F (\beta=1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-OLEX+CTYPE+LTYPE+TTAKE</td>
<td>0.84</td>
<td>0.89</td>
<td>0.87</td>
</tr>
<tr>
<td>CTYPE+LTYPE+TTAKE</td>
<td>0.68</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>S-OLEX+CTYPE+LTYPE</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>S-OLEX+LTYPE+TTAKE</td>
<td>0.95</td>
<td>0.93</td>
<td>0.94</td>
</tr>
<tr>
<td>S-OLEX+CTYPE+TTAKE</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
</tr>
</tbody>
</table>
Extracting conversation features

Take home

- it takes two to tango – use that!
- and even simplistic situational features helps!
Normalisation and tokenisation

All lower-case

1. let all sequences of black characters delimited by white-space constitute a token.

2. let all initial and final sequences of punctuation characters be a token of its own.

3. add extra tokens from (some) tokens with internal black-space: URL → URL + URL parts.
Appendix

Identifying flagrant utterances

1. use the sexual predator classifier.

2. classify all predator utterances in training material using only SLEX features.

3. rank the utterances by the probability for the utterance “being a sexual predator”.

4. decide a flagrancy threshold.

1. classify all utterances in test material as above.

2. pick all utterances above threshold.