Concept hierarchies and Credibility

Jean Mark Gawron
Alex Dodge
Kathleen Burke

August 7, 2013
1 Introduction
   - Group studies
   - Ontologies and concept hierarchies

2 Groups and conceptual frameworks

3 Credibility and importance

4 Tools
Studying groups organized for collective action

Project task

Task 3: Acquire the vocabulary and concept hierarchies for groups organized around ideas or themes or for collective action.

1. What are the defining features of the group? Do they constitute a real world community? An online community?
2. What are their goals?
3. What are their ideas?
4. What are they likely to do?
5. What are the characteristic features of their language and their discourse? What do they talk about? How do they talk about it?
Three groups (kinds of groups)

Website data collected from Google queries, mixture of blogs, forums, and organization websites

1. White militants
2. Vaccine debates [anti vaccine movement, pro vaccine websites, mostly professional or government]
3. Climate change [pro and anti]
Two applications

1. Word clustering
2. Relation extraction with Random Fields
Ontologies

Wikipedia definition

... represents knowledge as a set of concepts within a domain, and the relationships between pairs of concepts. It can be used to model a domain and support reasoning about concepts. ... a "formal specification of a shared conceptualisation"... provides a shared vocabulary, which can be used to model ... the type of objects and/or concepts that exist, and their properties and relations.

consists-of(water, [oxygen, hydrogen])
subtype( car-make, [Volvo, Volkswagen, BMW])
Relation extraction

CONSIST-OF

1. Water consists of hydrogen and oxygen.
2. Water is made up of hydrogen and oxygen.
3. Hydrogen and oxygen combine to make water.
4. Hydrogen bonds with oxygen to make water.
5. Hydrogen combined with oxygen makes water.

Group discourse analogues: Ideology components

What are the salient relations in the data? What are the texts talking about?
The vocabulary-based method

1. Collect data on group documents (websites, blogs, Tweets)
2. Characterize each word usage patterns in a set of documents, using syntactic analysis
3. Clustering words with similar distributions together
4. A hierarchy of words suggesting concepts and conceptual connections
An example

Blood & Honour is a world-wide pan Aryan organization dedicated to the struggle for survival and prosperity of the White Race. Since the Martyr’s death of its founder, Ian Stuart Donaldson, the Movement has experienced a period of turmoil and turbulence. Profiteers and pop stars have -unsuccessfully - tried to abuse B&H for their own selfish puposes. ZOG has geared up the campaign of oppression and the media has constantly smeared the national revolutionary forces. Members and supporters all over the White world have often been misinformed through infiltrators or left on their own due to the problems of the main organization...
White Militant Data II

Since the Martyr’s death of its founder, Ian Stuart Donaldson, the Movement has experienced a period of turmoil and turbulence.
White militant data III
Climate change judgments

**Target/Importance**
Global climate change, conversely, may not have such a substantially negative effect on the world’s population, standards of living, health, and survival.

**Target/Sentiment**
When you start to explore the issue, climate change can be terrifying.

**Information/Source**
Million Climate Jobs Campaign has released a new booklet describing how the transformation of the economy to one that protects people and the environment, can confront two of South Africa's biggest threats: climate change and unemployment.

**Target/credibility**
What is true is that climate change is real and its already happening.
Credibility judgments

1. X is credible
   1. **informational targets**: pronouncements, claims, stories
   2. **sources**

2. Kinds of judgments
   - **Debunking**: Calling into question the credibility or reliability of a story
   - **Legimitizing**: Affirming the credibility of a source
Tasks

Wiebe (2003) [sources, predicates, MPQA corpus]

1 Identify
   - Sources
   - Targets
   - Predicates

2 Infer polarities (positive/negative credibility)
Key sentences: Climate change

Anyone who says they can confidently predict global climate change or effects is either a fool or a fraud.

<table>
<thead>
<tr>
<th>Tags</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>Source of information (thought, speech, writing)</td>
</tr>
<tr>
<td>Predicate</td>
<td>Word or phrase expressing credibility, importance, or sentiment opinion (positive, negative)</td>
</tr>
<tr>
<td>Target</td>
<td>The message or phenomenon about which the opinion is held, by the source, if there is a source, or by the author; simultaneous judgments possible: “lies emanating from the White House”</td>
</tr>
</tbody>
</table>
Example 1

[\textit{source} Anyone who says
[\textit{source} they] can
[\textit{pred} \textit{OPINION}(credibility,positive) confidently]
[\textit{pred} \textit{OPINION}(credibility,positive) predict]
[\textit{target} global climate changes or effects]
]

is either a [\textit{pred} \textit{OPINION}(credibility,negative) fool]
or a [\textit{pred} \textit{OPINION}(credibility, negative) fraud] .
Example II

\[
\begin{align*}
\text{target} & \quad \text{That is how } \text{source} \quad \text{they} \\
\text{pred} & \quad (\text{credibility}, \text{positive}, \text{negative}) \quad \text{make it seem} \\
\text{target} & \quad \text{that all the top scientists are agreed,}
\end{align*}
\]

\[
\begin{align*}
\text{source} & \quad \text{he} \quad \text{pred}(\text{credibility}, \text{positive}) \quad \text{says}
\end{align*}
\]
Some annotated data
Predication by predication

| Line | He | noted | the criticism he took for the comment | , but did not back away from them |
| GF   | source | predicate | target |
| opinion_type | | credibility | |
| polarity | | positive | |
| Line | He | noted the criticism he took for the comment, but did not | back away | from | them |
| GF   | source | | predicate | target |
| opinion_type | | credibility | |
| polarity | | negative | |
Factoring the problem

“He said she said global warming is a myth.”

**Nested brackets:**

\[
[ \text{source He} ] \ [ \text{pred said} ] \ [ \text{target she} ] \ [ \text{pred said} ] \ [ \text{target global warming} ] \text{ is a } [ \text{predicate myth} ].
\]

**Shallow brackets:**

\[
[ \text{source He} ] \ [ \text{pred said} ] [ \text{source she} ] \ [ \text{pred said} ] [ \text{target global warming} ] \text{ is a } [ \text{predicate myth} ].
\]

**Tags:**

He/B-Source said/B-Pred she/B-Source said/B-Pred global/B-Target warming/I-Target is/O a/O myth/B-Pred ./O
Conditional Random Fields

Adapted from C. Sutton, A. McCallum, "An Introduction to Conditional Random Fields", ArXiv, November 2010