Test Data for Assignment#2 is Out

Download test data from here
http://www.isi.edu/~kozareva/teaching/eng.testing.zip

Output of test data due on 1st of March, 11:59 pm

Send output, source code, train, test arff files and report to kozareva@isi.edu and dirkh@isi.edu

We will be evaluating your homework on Friday 2nd of March!

CS544: Textual Entailment

February 28, 2012

Zornitsa Kozareva
USC/ISI
Marina del Rey, CA
kozareva@isi.edu
www.isi.edu/~kozareva
Textual Entailment - example

Some insects prefer to take a vacation from winter. Monarch butterflies migrate south in large numbers in the fall. Some even travel 1800 miles (2700 km) from Canada to Mexico to escape winter’s chill.

Read the text above and decide if the following statements are true or false based on the reading:

A) Insects, such as mosquitoes, can pass diseases to humans.
B) Some butterflies fly over 2500 kilometres to their winter homes.

Textual Entailment - definition

• (Ido Dagan, 2004) A text \( T \) is said to textually entail a hypothesis \( H \) if the meaning of \( H \) can be inferred from the meaning of \( T \).

\[ T: \text{Cavern Club sessions paid the Beatles £15 evenings and £5 lunchtime.} \]

\[ H: \text{The Beatles perform at Cavern Club at lunchtime.} \]
What is this?

Natural Language and Meaning

Meaning

Variability

Ambiguity

Language

a cup half full of water

a cup half empty of water

bank
Variability of Semantic Expressions

- The Dow Jones Industrial Average closed up 255
- Dow climbs 255
- Dow gains 255 points
- Stock market hits a record high

- Computers do not understand language variability
- Let’s model it as a relation between text expressions

  Textual Entailment: text1 ⇒ text2

Application Needs

- Information Extraction
  - identify relations among Named Entities
    - Yahoo! bought Overture
    - Overture was acquired by Yahoo!
    - Overture is part of Yahoo!
    - Yahoo! purchased Overture
  - extract facts

T: Regan attended a ceremony in Washington to commemorate the leadings in Normandy.

H: Washington is located in Normandy.
Application Needs

• Summarization
  – avoid sentences that infer the same meaning

• Question Answering, Information Retrieval
  – Name “Moby Dick’s” author
    • Herman Melville is the author of Moby Dick
    • Herman Melville wrote Moby Dick

Application Needs

• Machine Translation
  – evaluate how close a machine translation is to human

Watson is an artificial intelligence computer system capable of answering questions posed in natural language, developed at IBM.

Watson is an artificial intelligence computer system can respond to questions posed in natural language, developed at IBM.
Textual Entailment as a Classification Task

- Given a pair of sentences \((T,H)\) decide if:
  - \(T\) implies \(H\) (true)
  - \(T\) does not imply \(H\) (false)
Types of Textual Entailment (TE)

1. T - Euro-Scandinavian media cheer Denmark versus Sweden draw.
   H - Denmark and Sweden tie.

2. T - Jennifer Hawkins is the 21-year-old beauty queen from Australia.
   H - Jennifer Hawkins is Australia’s 21-year-old beauty queen.

3. T - The nomadic Raiders moved to LA in 1982 and won their third Super Bowl a year later.
   H - The nomadic Raiders won the Super Bowl in 1982.

Supervised Learning

- Features that model similarity or mismatch
- Classifier determines relative weights of information sources
- Train on development set of T-H pairs of sentences
**Lexical Information**

- **Bag-of-words model** which uses the words form the lexical constituents
- For each word in $H$, find the “best” word in $T$
- **Normalize** scores across sentence-pairs
- Find a **threshold** to distinguish the good matches from the bad matches
N-gram overlap - Example

<pair id="318" entailment="YES" task="QA">
<T>Mount Olympus towers up from the center of the earth.</T>
<H>Mount Olympus is in the center of the earth. </H>

<table>
<thead>
<tr>
<th>Unigram</th>
<th>in T</th>
<th>in H</th>
<th>Common</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mount</td>
<td>1</td>
<td>1</td>
<td>✔</td>
</tr>
<tr>
<td>Olympus</td>
<td>1</td>
<td>1</td>
<td>✔</td>
</tr>
<tr>
<td>towers</td>
<td>1</td>
<td>0</td>
<td>✗</td>
</tr>
<tr>
<td>up</td>
<td>1</td>
<td>0</td>
<td>✗</td>
</tr>
<tr>
<td>from</td>
<td>1</td>
<td>0</td>
<td>✗</td>
</tr>
<tr>
<td>the</td>
<td>2</td>
<td>2</td>
<td>✔</td>
</tr>
<tr>
<td>center</td>
<td>1</td>
<td>1</td>
<td>✔</td>
</tr>
<tr>
<td>of</td>
<td>1</td>
<td>1</td>
<td>✔</td>
</tr>
<tr>
<td>earth</td>
<td>1</td>
<td>1</td>
<td>✔</td>
</tr>
<tr>
<td>is</td>
<td>0</td>
<td>1</td>
<td>✗</td>
</tr>
<tr>
<td>in</td>
<td>0</td>
<td>1</td>
<td>✗</td>
</tr>
</tbody>
</table>

unigrams (7/10)
bigrams (5/9)

Longest Common Subsequence

• Longest common subsequence searches in-sequence matches
• Reflects the sentence level word order and captures the proportion of ordered words found in T and also present in H.

<pair id="413" entailment="NO" task="QA">
<T> A male rabbit is called a buck and a female rabbit is called a doe, just like deer.</T>
<H> A female rabbit is called a buck.</H>
Longest Common Subsequence

• Longest common subsequence searches in-sequence matches
• Reflects the sentence level word order and captures the proportion of ordered words found in $T$ and also present in $H$.

\[\langle \text{pair id="413" entailment="NO" task="QA"} \rangle \]
\[\langle T \rangle \text{ A male rabbit is called a buck and a female rabbit is called a doe, just like deer.} \langle /T \rangle \]
\[\langle H \rangle \text{ A female rabbit is called a buck.} \langle /H \rangle \]

Skip Grams

• Skip-grams are any pair of words in sentence order that allow arbitrary gaps.
• Measure the ratio of overlapping skip-grams between $T$ and $H$ divided by the total number of skip-grams

\[
skip_{overlap} = \frac{\#\text{common\_skip\_grams}(T,H)}{C(m,\#\text{common\_skip\_grams}(T,H))}
\]

$m$ – total number of words in $T$
$\#\text{common\_skip\_grams}(T,H)$ – total number of commons skip grams between $T$ and $H$
$C$ – combinatorial function
Skip Grams

<pair id="419" entailment="YES" task="QA">
  <T> Elizabeth Dowdeswell is the Under Secretary General at the United Nations Offices at Nairobi and Executive Director of the United Nations Environment Programme.</T>
  <H> Elizabeth Dowdeswell is Executive Director of the United Nations Environment Programme.</H>
</pair>

1) generate all possible skip-grams:
   Elizabeth is
   Elizabeth the
   Elizabeth Under
   United Environment
   United Programme
   Nations Programme
   Elizabeth is the
   Elizabeth the Under

2) find common skip-grams:
   Elizabeth is
   Elizabeth Executive
   Elizabeth Director
   United Environment
   United Programme
   Nations Programme
   Elizabeth the United

Comparisons of N-gram, LCS, Skip-gram

S_1: John loves Mary
S_2: John loved Mary
S_3: Mary loves John

• For unigram, LCS S_1, S_2 and S_3 are equally similar
• For Skip-gram S_1 and S_2 are more similar than S_1 and S_3
Levenshtein Distance

• Given strings $T$ and $H$
  – Distance is shortest sequence of edit commands that transform $T$ to $H$, (or equivalently $H$ to $T$).

  – Simple set of operations:
    • copy word from $T$ over to $H$
    • delete a word in $T$
    • insert a word in $H$
    • substitute one word for another

How much is the edit distance of the words?

• Edit (turned, truned) = ?
• Edit (computer, commuter) = ?
• Edit (banana, apple) = ?
• Edit (wombat, worcester) = ?
How much is the edit distance of the words?

- Edit (turned, truned) = 2
  - delete u
  - insert u

- Edit (computer, commuter) = 1
  - replace p with m

- Edit (banana, apple) = 5
  - delete b
  - replace n with p
  - replace a with p
  - replace n with l
  - replace a with e

- Edit (wombat, worcester) = 6

Can you think of a better Edit Distance?

- Are all operations equally likely?
  - copy character from T over to H (cost 0)
  - delete a character in T (cost 1)
  - insert a character in H (cost 1)
  - substitute one character for another (cost 1)

- Improvement, give different weights to different operations
  - replacing a for e is more likely than z for y
Levenshtein Distance - Example

- Distance (William Cohen, William Cohon)

Alignment:

T: W | I | L | L gap | I | A | M _ | C | O | H | E | N

H: W | I | L | L | I | A | M _ | C | O | H | O | N

Edit Op: C C C C I C C C C C C C S C
Cost: 0 0 0 0 1 1 1 1 1 1 1 1 2 2

Problems with Lexical Model

- Lexical overlaps are resource and language independent
- ... but they do not “understand”
  — negation
Negation

- Two texts may be very similar, containing numerous common words, but when one of the texts has a negation, the entailment relation is transformed from true to false, or vice versa
- Resolve the problem capturing negation words like (no, not, never, ...)

```
<pair id="213" entailment="NO" task="IR">
  <T> The death penalty is not a deterrent. </T>
  <H> Capital punishment is a deterrent to crime. </H>
```

Problems with Lexical Model

- Lexical overlaps are resource and language independent
- ... but they do not “understand”
  - negation
  - temporal expressions
  - numeric expressions
Number Matching

• Understand the meaning of numeric expressions
  – (four-thousand) is equivalent to (4000)
  – (4-years-old) has the same meaning as (four-years old)
  – (less than 5), means something (below 5 like 4,3,2,1)

  \[
  \text{<pair id="158" entailment="NO" task="IR">}
  \text{<T> More than 2,000 people lost their lives in the devastating Johnstown Flood. </T>}
  \text{<H> 2,000 people lost their lives. </H>}
  \]

Problems with Lexical Model

• Lexical overlaps are resource and language independent
• ... but they do not “understand”
  – negation
  – temporal expressions
  – numeric expressions
  – named entities
Named Entity Matching

- NE similarity can be captured using rules like acronyms, abbreviated first names, distance etc.
- *String Edit Distance*, given two strings (sequences) return the minimum number of “character edit operations” needed to turn one sequence into the other [like edit distance]

\[
\begin{align*}
\text{Andrew} & \quad \text{Amdrewz} \\
1. \text{substitute } m \text{ with } n \\
2. \text{delete } z \\
\text{distance } = 2
\end{align*}
\]

NE relation Matching

- Match the relations between the NEs

<T> Microsoft Inc. and Google are big competitors just like Toshiba Inc. and Sony. </T>

<H> Microsoft is a competitor of Toshiba. </H>
Problems with Lexical Model

• Common words improve the similarity too much
  – *The king is here*
  – *The salad is cold*

• Ignores syntactic relationships
  – *Mary loves John*  
  – *John loves Mary*  

Solution: perform shallow SOV parsing

Problems with Lexical Model

• Ignores semantic similarities
  – *I own a dog*
  – *I have a pet*  

Solution: supplement word similarity

• Ignores semantic frames/roles
  – *Yahoo bought Flickr*
  – *Flickr was sold to Yahoo*

Solution: analyze verb classes
Word Similarity

• How to capture that
  – buy ⇔ purchase
  – cat ⇔ pet

• Define similarity between words with
  – corpus-based measures (pointwise mutual information)
  – knowledge-based measures relying on WordNet
  – ...

Corpus-based Similarity

Pointwise Mutual Information

- Given two words $w_1$ and $w_2$, their similarity is measured as:

$$PMI(w_1, w_2) = \log_2 \frac{p(w_1, w_2)}{p(w_1) \cdot p(w_2)}$$

where, $p(w_1, w_2)$ is the probability of seeing the two words together

$p(w_i)$ is the probability of seeing word $w_i$ and it is calculated as

$$p(w_i) = \frac{freq(w_i)}{N}$$

Similarity using WordNet Hierarchy

Rank the following terms based on their similarity:

- Sim (wolf, dog)
- Sim (wolf, amphibian)
- Sim (terrier, wolf)
- Sim (dachshund, terrier)
What information/heuristics did you use to rank these?

- Path length is important (but not the only thing)
- Words that share the same ancestor are related
- Words lower down in the hierarchy are finer grained and therefore closer

Knowledge-based Similarity

Similarity using WordNet

• (Leacock & Chodorow, 1998)

\[ sim_{ich} = - \log \frac{\text{length}}{2 * D} \]

- *length* is the length of the shortest path between two concepts using node counting
- *D* is the maximum depth of the taxonomy
Similarity using WordNet

- (Wu & Palmer, 1994)

\[
\text{sim}_{wup} = \frac{2 \times \text{depth}(\text{closest} - \text{common} - \text{ancestor})}{\text{depth}(	ext{concept}_1) + \text{depth}(	ext{concept}_2)}
\]

- (Lesk, 1986)
  - Finds the overlap between the dictionary entries of two words

Semantic Information - Example

<T> He died of blood loss. </T>
<H> He died bleeding. </H>

Step 1: determine POS-tags;
Semantic Information - Example

<T> He_PP died_VVD of_IN blood_NN loss_NN.</T>

<H> He_PP died_VVD bleeding_VVG. </H>

Step2 : extract verbs\&nouns;

<table>
<thead>
<tr>
<th>died_VVD</th>
<th>died_VVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>died_VVD</td>
<td>bleeding_VVG</td>
</tr>
<tr>
<td>blood_NN</td>
<td>died_VVD</td>
</tr>
<tr>
<td>blood_NN</td>
<td>bleeding_VVG</td>
</tr>
<tr>
<td>loss_NN</td>
<td>died_VVD</td>
</tr>
<tr>
<td>loss_NN</td>
<td>bleeding_VVG</td>
</tr>
</tbody>
</table>

Semantic Information - Example

Path measure:

<table>
<thead>
<tr>
<th>die#v#1</th>
<th>die#v#1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>die#v#11</td>
<td>bleed#v#1</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Lin measure:

<table>
<thead>
<tr>
<th>die#v#1</th>
<th>die#v#1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>die#v#1</td>
<td>bleed#v#1</td>
<td>0</td>
</tr>
</tbody>
</table>

Vector measure:

<table>
<thead>
<tr>
<th>blood#n#2</th>
<th>died#v#5</th>
<th>0.0259</th>
</tr>
</thead>
<tbody>
<tr>
<td>blood#n#1</td>
<td>bleeding#v#5</td>
<td>0.0202</td>
</tr>
<tr>
<td>loss#n#5</td>
<td>died#v#9</td>
<td>0.2654</td>
</tr>
<tr>
<td>loss#n#7</td>
<td>bleeding#v#1</td>
<td>0.1632</td>
</tr>
</tbody>
</table>

Step3 : WN Similarity;
Deeper Semantics

Text/Hypothesis

Semantic Interpretation

Logical Representation

Logical Inference

Logic Forms

- Text “Peter loves Mary.”
- Discourse Representation Theory:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>peter (x)</td>
<td>mary(y)</td>
</tr>
</tbody>
</table>

- First Order Logic:
  \[ \exists x \exists y \text{(peter (x) & mary(y) & love(x,y))} \]

- Knowledge Base:
  \[ \forall x \text{(peter(x) \rightarrow man(x))} \]
  \[ \forall x \text{(mary (x) \rightarrow woman(x))} \]
  \[ \forall x \text{(man(x) \rightarrow \neg woman(x))} \]

- Model: \[ D = \{d1,d2\} \]
  \[ F(\text{peter}) = \{d1\} \]
  \[ F(\text{mary}) = \{d2\} \]
  \[ F(\text{love}) = \{(d1,d2)\} \]

- Problems:
  - number of rules
  - computation
# Results

<table>
<thead>
<tr>
<th>First Author (Group)</th>
<th>Accuracy</th>
<th>Average Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hickl (LCC)</td>
<td>75.4%</td>
<td>80.8%</td>
</tr>
<tr>
<td>Tatu (LCC)</td>
<td>73.8%</td>
<td>71.3%</td>
</tr>
<tr>
<td>Zanzotto (Milan &amp; Rome)</td>
<td>63.9%</td>
<td>64.4%</td>
</tr>
<tr>
<td>Adams (Dallas)</td>
<td>62.6%</td>
<td>62.8%</td>
</tr>
<tr>
<td>Bos (Rome &amp; Leeds)</td>
<td>61.6%</td>
<td>66.9%</td>
</tr>
<tr>
<td>11 groups</td>
<td>58.1%-60.5%</td>
<td>Average: 60%</td>
</tr>
<tr>
<td>7 groups</td>
<td>52.9%-55.6%</td>
<td>Median: 59%</td>
</tr>
</tbody>
</table>