CS544: Textual Entailment

March 24, 2011

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

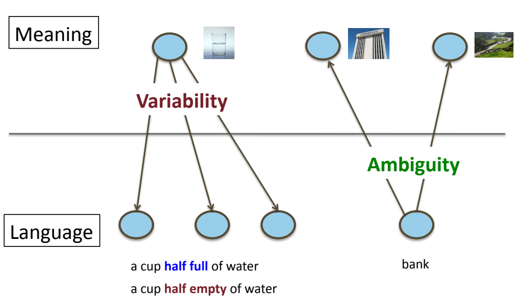
What is this?

glass **half empty** of water

glass **half full** of water

1

Natural Language and Meaning



2

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 variability. One can model it as relations between text expressions:

  Textual Entailment: *text1* ⇒ *text2*

3

Textual Entailment - definition

- A text *T* is said to textually entail a hypothesis *H,* if the meaning of *H* can be most likely inferred from the meaning of *T* (Ido Dagan, 2004)
  - ***T***: The company ***aquired*** four daily newspaper from Sun Enterprises.
  - ***H***: Sun Enterprises ***sold*** four daily newspapers to the company.

True or False ?

4

Application Needs

- Information Extraction
  - identify relations among Named Entities
    - Yahoo! *bough* Overtrue
    - Overtrue *was aquired by* Yahoo!
    - Overtrue *is part of* Yahoo!
    - Yahoo! *purchesed* Overture
  - extract facts

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

H: Washington is located in Normandy.

5

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

6

Application Needs

- Machine Translation
  - evalue 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.

7

Types of Textual Entailment (TE)

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

2. T - Jennifer Hawkins is the 21-year-old beauty queen from Australia. syntactic information
   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.

temporal entities

8

RECIPE FOR SOLVING TEXTUAL ENTAILMENT

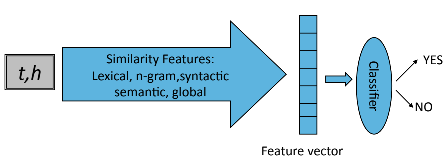
Textual Entailment as Classification Task

- Given a pair of sentences (*T*,*H*) decide if:
  - *T* implies *H* (true)
  - *T* does not imply *H* (false)

Binary classification

- To learn a classifier for TE, we need to:
  - collect annotated examples — Available from TE challenge
  - select a ML algorithm — Any toolkit, for example Weka
  - define a feature space

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

Defining Feature Space

$T_1 \Rightarrow H_1$	
T_1	*"At the end of the year, all solid companies pay dividends."*
H_1	*"At the end of the year, all solid insurance companies pay dividends."*

- Possible features
  - "Distance Features" between *T* and *H*
  - "Entailment triggers"
  - "Pair Feature" representing the content of *T-H*
- Possible representations of the sentences
  - Bag-of-words
  - Syntactic representation
  - Semantic representation

Distance Features

$T \Rightarrow H$	
T	*"At the end of the year, all solid companies pay dividends."*
H	*"At the end of the year, all solid insurance companies pay dividends."*

Possible features:
- Number of words in common (n-grams)
- Longest common subsequence
- Longest common syntactic subtree
- …

Entailment Triggers

- Possible features from (de Marneffe et al.,06):
  - Antonymy features capture the presence/absence of antonymous words in *T* and *H*

    "oil price is **surging**" $\Rightarrow$ "oil prices is **falling down**"

  - Adjunct features capture the dropping/adding of syntactic adjunct when moving from T to H

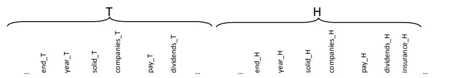
  "companies pay dividends" $\Rightarrow$ "companies pay cash dividends"
  - …

Pair Features

T ⇒ H	
T	*"At the end of the year, all solid companies pay dividends."*
H	*"At the end of the year, all solid insurance companies pay dividends."*

Possible features

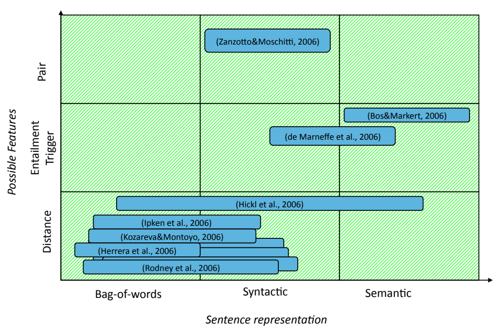
- Bag-of-word spaces of T and H



- Syntactic spaces of T and H

Page 15

ML Methods in the possible feature spaces



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

17

N-gram overlap

- An n-gram is a subsequence of *n* terms from a given text sequence
  - unigram (one word)
  - bigram (two consecutive words)
- Measures the ratio of the n-gram overlaps in the entailing text *T* and hypothesis *H*

$$n-gram-overlap = \frac{m}{n}$$

  - *m* is the number of common n-grams in *T* and *H*
  - *n* is total number of words in *T*
  - *n-gram-overlap* has values between 0 and 1

18

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>

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

Uni-gram	in T	in H	Common
Mount	1	1	✔
Olympus	1	1	✔
towers	1	0	✗
up	1	0	✗
from	1	0	✗
the	2	2	✔
center	1	1	✔
of	1	1	✔
earth	1	1	✔
is	0	1	✗
in	0	1	✗

19

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>

20

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

$$skip\_overlap = \frac{\#common\_skip\_grams(T,H)}{C(m,\#common\_skip\_grams(T,H))}$$

m – total number of words in *T*
#common_skip_grams(*T*,*H*) – total number of commons skip grams between T and H
C – combinatorial function

21

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>

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

22

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

23

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 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)

24

Levenshtein Distance - Example

- Distance (William Cohen, William Cohon)

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

25

Problems with Lexical Model

- Lexical overlaps are resource and language independent
- … but they do not "understand"
  - negation
  - temporal expressions
  - numeric expressions
  - named entities
  - past/present/future tense
  - meanings of words

26

Problems with Lexical Model

- Common words improve the similarity too much
  - *The king is here* vs. *The salad is cold*
- Ignores syntactic relationships
  - *Mary loves John* vs. *John loves Mary*
  - Solution: perform shallow SOV parsing

27

Problems with Lexical Model

- Ignores semantic similarities
  - *I own a dog* vs. *I have a pet*
  - Solution: supplement word similarity
- Ignores semantic frames/roles
  - *Yahoo bought Flickr* vs. *Flickr was sold to Yahoo*
  - Solution: analyze verb classes

28

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>

29

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)

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

30

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]

Andrew
Amdrewz

1. substitute ***m*** with ***n***
2. delete ***z***

distance = 2

31

NE relation Matching

- Match the relations between the NEs

<pair id="355" entailment="NO" task="IE">
<T> Microsoft Inc. and Google are big competitors just like Toshiba Inc. and Sony. </T>
<H> Microsoft is a competitor of Toshiba.</H>

$\text{Microsoft} \xleftrightarrow{\text{competitor}} \text{Google}$

$\text{Toshiba} \xleftrightarrow{\text{competitor}} \text{Sony}$

$\text{Microsoft} \xleftrightarrow{?} \text{Toshiba}$

32

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
  - …

33

Corpus-based Similarity

Pointwise Mutual Information

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

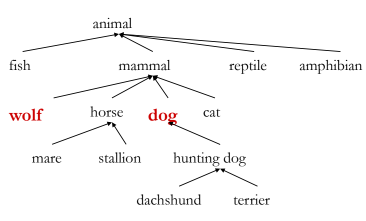
$$PMI(w_1,w_2) = \log_2 \frac{p(w_1,w_2)}{p(w_1) * p(w_2)}$$

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

34

Similarity using WordNet Hierarchy



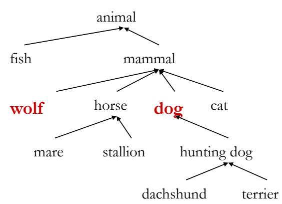
35

Similarity using WordNet

- (Leacock & Chodorow, 1998)

$$sim_{lch} = -\log \frac{length}{2 * D}$$

- *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)

$$sim_{wup} = \frac{2 * depth(LCS)}{depth(concept_1) + depth(concept_2)}$$

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

Semantic Information - Methodology

- Given *T* and *H* sentences
  - determine the POS-tags
  - extract all verbs and nouns
  - measure similarity of terms with WordNet (check WordNet::Similarity package)
  - calculate inter-syntactic similarity

Walk-through example

T: When the defendant and his lawyer walked into the court, some of the victim supporters turned their backs to him.

H: When the defendant walked into the courthouse with his attorney, the crowd turned their backs on him.

Is the meaning of H entailed from the meaning of T?

39

Walk-through example

Text 1	Text 2	maxSim
defendant	defendant	1.0
lawyer	attorney	0.9
walked	walked	1.0
court	courthouse	0.6
victims	courthouse	0.4
supporters	crowd	0.4
turned	turned	1.0
backs	backs	1.0

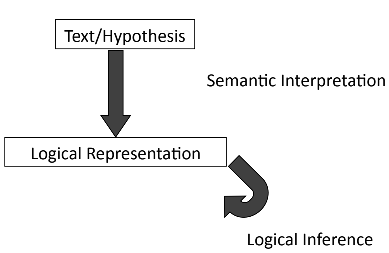
T1: When the **defendant** and his **lawyer walked** into the **court**, some of the **victim supporters turned** their **backs** to him.

T2: When the **defendant walked** into the **courthouse** with his **attorney**, the **crowd turned** their **backs** on him.

- Calculate the semantic similarity score as the sum of all similarities divided by total number of word pairs

40

Deeper Semantics



Logic Forms

- Text "*Peter loves Mary.*"
- Discourse Representation Theory:

x y
peter (x)
mary(y)
love(x,y)

- First Order Logic:
  ∃x∃y(peter (x) & mary(y) & love(x,y))
- Knowledge Base:
  ∀x (peter(x) → man(x))
  ∀x (mary (x) → woman(x))
  ∀x (man(x) → ¬ woman(x))
- Model: D = {d1,d2}
  F(peter)={d1}
  F(mary)={d2}
  F(love)={(d1,d2)}

Problems:
- number of rules
- computation

Results

First Author (Group)	Accuracy	Average Precision
Hickl (LCC)	75.4%	80.8%
Tatu (LCC)	73.8%	71.3%
Zanzotto (Milan & Rome)	63.9%	64.4%
Adams (Dallas)	62.6%	62.8%
Bos (Rome & Leeds)	61.6%	66.9%
11 groups	58.1%-60.5%	
7 groups	52.9%-55.6%	

Average: 60%
Median: 59%

43

Why?

- Most systems report:
  – lack of knowledge (syntactic transformation rules, lexical relations, etc.)
  – lack of training data
- While best performing systems like:
  – Hickl et al. acquired large entailment corpora for training
  – Tatu et al. used large knowledge bases (linguistic and world knowledge)

44

Learning Entailment Rules

Q: What reduces the risk of Heart Attacks?

Hypothesis: Aspirin reduces the risk of Heart Attacks	*Text:* Aspirin prevents Heart Attacks

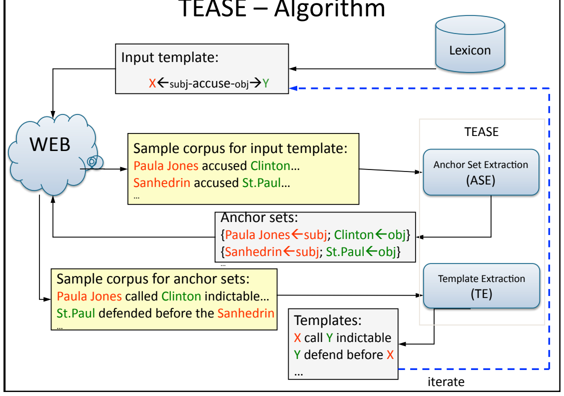
Entailment Rule:

$\underbrace{X \text{ } \textit{prevent} \text{ } Y}_{\text{template}} \Rightarrow \underbrace{X \text{ reduce risk of } Y}_{\text{template}}$

➔ *Need a large knowledge base of entailment rules*

45

TEASE – Algorithm



Sample of Extracted Anchor-Sets for *X prevent Y*

X='sunscreens', **Y**='sunburn'	**X**='gene therapy', **Y**='blindness'
X='sunscreens', **Y**='skin cancer'	**X**='cooperation', **Y**='terrorism'
X='vitamin e', **Y**='heart disease'	**X**='safety valve', **Y**='leakage'
X='aspirin', **Y**='heart attack'	**X**='safe sex', **Y**='cervical cancer'
X='vaccine candidate', **Y**='infection'	**X**='safety belts', **Y**='fatalities'
X='universal precautions', **Y**='HIV'	**X**='security fencing', **Y**='intruders'
X='safety device', **Y**='fatal injuries'	**X**='soy protein', **Y**='bone loss'
X='hepa filtration', **Y**='contaminants'	**X**='MWI', **Y**='pollution'
X='low cloud cover', **Y**= 'measurements'	**X**='vitamin C', **Y**='colds'

47

Sample of Extracted Templates for *X* prevent *Y*

X reduce **Y**	**X** reduce **Y** risk
X protect against **Y**	**X** decrease the risk of **Y**
X eliminate **Y**	relationship between **X** and **Y**
X stop **Y**	**X** guard against **Y**
X avoid **Y**	**X** be cure for **Y**
X for prevention of **Y**	**X** treat **Y**
X provide protection against **Y**	**X** in war on **Y**
X combat **Y**	**X** in the struggle against **Y**
X ward **Y**	**X** a day keeps **Y** away
X lower risk of **Y**	**X** eliminate the possibility of **Y**
X be barrier against **Y**	**X** cut risk **Y**
X fight **Y**	**X** inhibit **Y**

48

Accuracy Extracted Information

- Choose randomly 48 verbs
- Pull all extracted templates (1392 in total)
- Ask humans for pattern correctness/incorrectness

Average Yield per verb	**29** correct templates per verb
Average Precision per verb	**45.30%**

49

Note: not perfect, but this additional knowledge helps the systems

Additional Information

- **Textual Entailment Community**:
  - The *RTE Resource Pool* can now be accessed from: http://aclweb.org/aclwiki/index.php?title=Textual_Entailment_Resource_Pool
  - The *Textual Entailment Subzone* can now be accessed from: http://aclweb.org/aclwiki/index.php?title=Textual_Entailment_Portal
- **Textual Entailment Resource Pool**
  - Textual Entailment Resource Pool
- **PASCAL Challenges**
  - **RTE-1**
  - **RTE-2**
  - **RTE-3**
- Recognizing Textual Entailment (RTE) has been proposed recently as a generic task that captures major semantic inference needs across many natural language processing applications.
- **TAC 2008 challenge**

Textual Entailment Workshops

- ACL 2005 Workshop on Empirical Modeling of Semantic Equivalence and Entailment, 2005
- First PASCAL Recognising Textual Entailment Challenge (RTE-1), 2005
- Second PASCAL Recognising Textual Entailment Challenge (RTE-2), 2006
- Third PASCAL Recognising Textual Entailment Challenge (RTE-3), 2007
- Answer Validation Exercise at CLEF 2006 (AVE 2006)
- Answer Validation Exercise at CLEF 2007 (AVE 2007)