Textual Entailment and Logical Inference

CMSC 473/673
UMBC
December 4th, 2017
Course Announcement 1: Assignment 4

Due Monday December 11th (~1 week)

Any questions?
Course Announcement 2: Final Exam

No mandatory final exam

December 20th, 1pm-3pm: optional second midterm/final

Averaged into first midterm score

No practice questions

Register by Monday 12/11:
https://goo.gl/forms/aXflKkP0BIRxhOS83
Recap from last time...
A Shallow Semantic Representation: Semantic Roles

Predicates (bought, sold, purchase) represent a situation

**Semantic roles** express the abstract role that arguments of a predicate can take in the event

More specific: buyer  agent  proto-agent  More general
FrameNet and PropBank representations
SRL Features

Headword of constituent: Examiner
Headword POS: NNP
Voice of the clause: Active
Subcategorization of pred:
- VP -> VBD NP
- PP
Named Entity type of constituent: ORGANIZATION
First and last words of constituent: The, Examiner
Linear position re: predicate: before

### Frequency | Path | Description
--- | --- | ---
14.2% | VB↑VP↓PP | PP argument/adjunct
11.8% | VB↑VP↑S↓NP | subject
10.1% | VB↑VP↓NP | object
7.9% | VB↑VP↑VP↑S↓NP | subject (embedded VP)
4.1% | VB↑VP↓ADVP | adverbial adjunct
3.0% | NN↑NP↑NP↓PP | prepositional complement of noun
1.7% | VB↑VP↓PRT | adverbial particle
1.6% | VB↑VP↑VP↑VP↑S↓NP | subject (embedded VP)
14.2% | VB↑VP↑VP↑VP↑S↓NP | no matching parse constituent
31.4% | Other |
3-step SRL

1. **Pruning**: use simple heuristics to prune unlikely constituents.

2. **Identification**: a binary classification of each node as an argument to be labeled or a NONE.

3. **Classification**: a 1-of-$N$ classification of all the constituents that were labeled as arguments by the previous stage.

**Pruning & Identification**

Prune the very unlikely constituents first, and then use a classifier to get rid of the rest.

Very few of the nodes in the tree could possibly be arguments of that one predicate.

**Imbalance between positive samples** (constituents that are arguments of predicate)

**Imbalance between negative samples** (constituents that are not arguments of predicate)
Logical Forms of Sentences

\[ \exists e, x, y \, Eating(e) \land Agent(e, x) \land Theme(e, y) \]

Papa ate the caviar
One Way to Represent Selectional Restrictions

\[ \exists e, x, y \, Eating(e) \land Agent(e, x) \land Theme(e, y) \]

\[ \exists e, x, y \, Eating(e) \land Agent(e, x) \land Theme(e, y) \land EdibleThing(y) \]

but do have a large knowledge base of facts about edible things?!

(do we know a hamburger is edible? sort of)
WordNet

Knowledge graph containing *concept* relations

- hypernymy, hyponymy *(is-a)*
- meronymy, holonymy *(part of whole, whole of part)*
- troponymy *(describing manner of an event)*
- entailment *(what else must happen in an event)*
A Simpler Model of Selectional Association
(Brockmann and Lapata, 2003)

Model just the association of predicate \( v \) with a single noun \( n \)

Parse a huge corpus

Count how often a noun \( n \) occurs in relation \( r \) with verb \( v \):

\[
\log \text{count}(n,v,r)
\]

(or the probability)

<table>
<thead>
<tr>
<th>Verb</th>
<th>Plaus./Implaus.</th>
</tr>
</thead>
<tbody>
<tr>
<td>see</td>
<td>friend/method</td>
</tr>
<tr>
<td>read</td>
<td>article/fashion</td>
</tr>
<tr>
<td>find</td>
<td>label/fever</td>
</tr>
<tr>
<td>hear</td>
<td>story/issue</td>
</tr>
<tr>
<td>write</td>
<td>letter/market</td>
</tr>
<tr>
<td>urge</td>
<td>daughter/contrast</td>
</tr>
<tr>
<td>warn</td>
<td>driver/engine</td>
</tr>
<tr>
<td>judge</td>
<td>contest/climate</td>
</tr>
<tr>
<td>teach</td>
<td>language/distance</td>
</tr>
<tr>
<td>show</td>
<td>sample/travel</td>
</tr>
<tr>
<td>expect</td>
<td>visit/mouth</td>
</tr>
<tr>
<td>answer</td>
<td>request/tragedy</td>
</tr>
<tr>
<td>recognize</td>
<td>author/pocket</td>
</tr>
<tr>
<td>repeat</td>
<td>comment/journal</td>
</tr>
<tr>
<td>understand</td>
<td>concept/session</td>
</tr>
<tr>
<td>remember</td>
<td>reply/smoke</td>
</tr>
</tbody>
</table>

See: Bergsma, Lin, Goebel (2008) for evaluation/comparison
Revisiting the PropBank Theory

1. Fewer roles: generalized semantic roles, defined as prototypes (Dowty 1991)
   PROTO-AGENT
   PROTO-PATIENT

2. More roles: Define roles specific to a group of predicates
### Dowty (1991)’s Properties

<table>
<thead>
<tr>
<th>Property</th>
<th>Proto-Agent</th>
<th>Proto-Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>instigated</td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>volitional</td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>awareness</td>
<td></td>
<td>✔️ ?</td>
</tr>
<tr>
<td>sentient</td>
<td></td>
<td>✔️ ?</td>
</tr>
<tr>
<td>moved</td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>physically existed</td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>existed before</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>existed during</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>existed after</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>changed possession</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>changed state</td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>stationary</td>
<td></td>
<td>✔️</td>
</tr>
</tbody>
</table>
Asking People Simple Questions

Reisinger et al. (2015)

He et al. (2015)
Semantic Expectations

Answers can be given by “ordinary” humans

Correlate with linguistically-complex theories

Who did what to whom at where?

The police officer detained the suspect at the scene of the crime

Agent Predicate Theme Location

Reisinger et al. (2015)

He et al. (2015)
Entailment Outline

Basic Definition

Task 1: Recognizing Textual Entailment (RTE)

Task 2: Examining Causality (COPA)

Task 3: Large crowd-sourced data (SNLI)
Entailment Outline

Basic Definition

Task 1: Recognizing Textual Entailment (RTE)

Task 2: Examining Causality (COPA)

Task 3: Large crowd-sourced data (SNLI)
Entailment:
Underlying a Number of Applications

Question
Who bought Overture? >> Expected answer form
X bought Overture

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Entailment:
Underlying a Number of Applications

Question:
Who bought Overture?

Expected answer form:
X bought Overture

Overture’s acquisition by Yahoo

Yahoo bought Overture

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Entailment:
Underlying a Number of Applications

Question
Who bought Overture?

Expected answer form
X bought Overture

Overture’s acquisition by Yahoo

Yahoo bought Overture

Information extraction: X acquire Y
Information retrieval: Overture was bought for ...
Summarization: identify redundant information
MT evaluation

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Chierchia & McConnell-Ginet (2001): 

A text $t$ entails a hypothesis $h$ if 

$h$ is true in every circumstance in which $t$ is true
Classical Entailment Definition

Chierchia & McConnell-Ginet (2001):  
A text \( t \) entails a hypothesis \( h \) if \( h \) is true in every circumstance in which \( t \) is true

Strict entailment - doesn't account for some uncertainty allowed in applications

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
“Almost certain” Entailments

t: The technological triumph known as GPS ... was incubated in the mind of Ivan Getting.

h: Ivan Getting invented the GPS.

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Applied Textual Entailment

A directional relation between two text fragments

$t$ (text) entails $h$ (hypothesis) ($t \rightarrow h$) if humans reading $t$ will infer that $h$ is most likely true

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Probabilistic Interpretation

t probabilistically entails h if:

\[ P(h \text{ is true} \mid t) > P(h \text{ is true}) \]

t increases the likelihood of h being true

Positive PMI – t provides information on h’s truth

the value is the entailment confidence

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Entailment Outline

Basic Definition

Task 1: Recognizing Textual Entailment (RTE)

Task 2: Examining Causality (COPA)

Task 3: Large crowd-sourced data (SNLI)
Generic Dataset by Application Use

PASCAL Recognizing Textual Entailment (RTE) Challenges

7 application settings in RTE-1, 4 in RTE-2/3
  QA, IE, “Semantic” IR, Comparable documents / multi-doc summarization, MT evaluation, Reading comprehension, Paraphrase acquisition

Most data created from actual applications output

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
<table>
<thead>
<tr>
<th>TEXT</th>
<th>HYPOTHESIS</th>
<th>TASK</th>
<th>ENTAILMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reagan attended a ceremony in Washington to commemorate the landings in Normandy.</td>
<td>Washington is located in Normandy.</td>
<td>IE</td>
<td>False</td>
</tr>
<tr>
<td>Google files for its long awaited IPO.</td>
<td>Google goes public.</td>
<td>IR</td>
<td>True</td>
</tr>
<tr>
<td>...: a shootout at the Guadalajara airport in May, 1993, that killed Cardinal Juan Jesus Posadas Ocampo and six others.</td>
<td>Cardinal Juan Jesus Posadas Ocampo died in 1993.</td>
<td>QA</td>
<td>True</td>
</tr>
<tr>
<td>The SPD got just 21.5% of the vote in the European Parliament elections, while the conservative opposition parties polled 44.5%.</td>
<td>The SPD is defeated by the opposition parties.</td>
<td>IE</td>
<td>True</td>
</tr>
</tbody>
</table>

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Dominant approach: Supervised Learning

Features model similarity and mismatch

Classifier determines relative weights of information sources

Train on development set and auxiliary $t,h$ corpora

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Common and Successful Approaches (Features)

Measure similarity match between $t$ and $h$
- Lexical overlap (unigram, N-gram, subsequence)
- Lexical substitution (WordNet, statistical)
- Syntactic matching/transformations
- Lexical-syntactic variations ("paraphrases")
- Semantic role labeling and matching
- Global similarity parameters (e.g. negation, modality)

Cross-pair similarity

Detect mismatch (for non-entailment)

Interpretation to logic representation + logic inference

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Common and Successful Approaches (Features)

Measure similarity match between \( t \) and \( h \)
- Lexical overlap (unigram, N-gram, subsequence)
- Lexical substitution (WordNet, statistical)
- Syntactic matching/transformations
- Lexical-syntactic variations (“paraphrases”)
- Semantic role labeling and matching
- Global similarity parameters (e.g. negation, modality)

Cross-pair similarity
- Detect mismatch (for non-entailment)
- Interpretation to logic representation + logic inference

Lexical baselines are hard to beat!

Lack of knowledge (syntactic transformation rules, paraphrases, lexical relations, etc.)

Lack of training data

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Refining the feature space

How do we define the feature space?

\[ T_1 \Rightarrow H_1 \]

<table>
<thead>
<tr>
<th>T_1</th>
<th>“At the end of the year, all solid companies pay dividends.”</th>
</tr>
</thead>
<tbody>
<tr>
<td>H_1</td>
<td>“At the end of the year, all solid insurance companies pay dividends.”</td>
</tr>
</tbody>
</table>

Possible features

“Distance Features” - Features of “some” distance between T and H
“Entailment trigger Features”
“Pair Feature” – The content of the T-H pair is represented

Possible representations of the sentences

Bag-of-words (possibly with n-grams)
Syntactic representation
Semantic representation

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Distance Features

T ⇒ H

<table>
<thead>
<tr>
<th>T</th>
<th>“At the end of the year, all solid companies pay dividends.”</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>“At the end of the year, all solid insurance companies pay dividends.”</td>
</tr>
</tbody>
</table>

Possible features

– Number of words in common
– Longest common subsequence
– Longest common syntactic subtree
– ...

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Entailment Triggers

Possible features from (de Marneffe et al., 2006)

Polarity features
presence/absence of negative polarity contexts (not, no, or few, without)
“Oil price surged” ⇒ “Oil prices didn’t grow”

Antonymy features
presence/absence of antonymous words in T and H
“Oil price is surging” ⇒ “Oil prices is falling down”

Adjunct features
dropping/adding of syntactic adjunct when moving from T to H
“all solid companies pay dividends” ⇒ “all solid companies pay cash dividends”

...
Details of The Entailment Strategy

Preprocessing
- Multiple levels of lexical preprocessing
- Syntactic Parsing
- Shallow semantic parsing
- Annotating semantic phenomena

Representation
- Bag of words, n-grams through tree/graphs based representation
- Logical representations

Knowledge Sources
- Syntactic mapping rules
- Lexical resources
- Semantic Phenomena specific modules
- RTE specific knowledge sources
- Additional Corpora/Web resources

Control Strategy & Decision Making
- Single pass/iterative processing
- Strict vs. Parameter based

Justification
- What can be said about the decision?

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Basic Representations

Meaning Representation

Inference

Logical Forms

Semantic Representation

Syntactic Parse

Local Lexical

Raw Text

Textual Entailment

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Basic Representations (Syntax)

Hyp: The Cassini spacecraft has reached Titan.
Basic Representations (Syntax)

Hyp: The Cassini spacecraft has reached Titan.

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Basic Representations (Syntax)

Syntactic Parse

The bombers had not managed to enter the embassy compounds.

The bombers entered the embassy compounds.

Hyp: The Cassini spacecraft has reached Titan.

Local Lexical

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Enriching Preprocessing

POS tagging
Stemming
Predicate argument representation
  verb predicates and nominalization
Entity Annotation
  Stand alone NERs with a variable number of classes
Co-reference resolution
Dates, times and numeric value normalization
Identification of semantic relations
  complex nominals, genitives, adjectival phrases, and adjectival clauses
Event identification

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Basic Representations (Shallow Semantics)

T: The government purchase of the Roanoke building, a former prison, took place in 1902.

H: The Roanoke building, which was a former prison, was bought by the government in 1902.
Basic Representations (Shallow Semantics)

T: The government purchase of the Roanoke building, a former prison, took place in 1902.

H: The Roanoke building, which was a former prison, was bought by the government in 1902.

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
T: The government purchase of the Roanoke building, a former prison, took place in 1902.

H: The Roanoke building, which was a former prison, was bought by the government in 1902.
Basic Representations (Shallow Semantics)

T: The government *purchase* of the Roanoke building, a former prison, *took* place in 1902.

H: The Roanoke building, which *was* a former prison, *bought* by the government in 1902.

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Characteristics

Multiple paths $\rightarrow$ optimization problem

- Shortest or highest-confidence path through transformations
- Order is important; may need to explore different orderings
- Module dependencies are ‘local’; module B does not need access to module A’s KB/inference, only its output

If outcome is “true”, the (optimal) set of transformations and local comparisons form a proof

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Semantic Role Labeling Could Help with (Some) Semantic Phenomena

Relative clauses
The assailants fired six bullets at the car, which carried Vladimir Skobtsov.
The car carried Vladimir Skobtsov.
Semantic Role Labeling handles this phenomena automatically

Clausal modifiers
But celebrations were muted as many Iranians observed a Shi'ite mourning month.
Many Iranians observed a Shi'ite mourning month.
Semantic Role Labeling handles this phenomena automatically

Passive
We have been approached by the investment banker.
The investment banker approached us.
Semantic Role Labeling handles this phenomena automatically

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Semantic Role Labeling Could Help with (Some) Semantic Phenomena

Relative clauses
- The assailants fired six bullets at the car, which carried Vladimir Skobtsov.
- The car carried Vladimir Skobtsov.
- Semantic Role Labeling handles this phenomena automatically

Appositives
- Frank Robinson, a one-time manager of the Indians, has the distinction for the NL.
- Frank Robinson is a one-time manager of the Indians.

Clausal modifiers
- But celebrations were muted as many Iranians observed a Shi'ite mourning month.
- Many Iranians observed a Shi’ite mourning month.
- Semantic Role Labeling handles this phenomena automatically

Genitive modifier
- Malaysia's crude palm oil output is estimated to have risen.
- The crude palm oil output of Malasia is estimated to have risen.

Conjunctions
- Jake and Jill ran up the hill (Jake ran up the hill)
- Jake and Jill met on the hill (*Jake met on the hill)

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Logical Structure

**Factivity**: Uncovering the context in which a verb phrase is embedded

- The terrorists tried to enter the building.
- The terrorists entered the building.

**Polarity**: negative markers or a negation-denoting verb (e.g. *deny, refuse, fail*)

- The terrorists failed to enter the building.
- The terrorists entered the building.

**Modality/Negation**: Dealing with *modal auxiliary verbs* (can, must, should), that modify verbs’ meanings and with the identification of the scope of negation.

**Superlatives/Comparatives/Monotonicity**: inflecting adjectives or adverbs.

**Quantifiers, determiners and articles**

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Knowledge Acquisition for TE

Explicit Knowledge (Structured Knowledge Bases)

Relations among **words** (or concepts)
- Symmetric: Synonymy, cohyponymy
- Directional: hyponymy, part of, …

Relations among **sentence prototypes**
- Symmetric: Paraphrasing
- Directional: Inference Rules/Rewrite Rules

Implicit Knowledge

Relations among sentences
- Symmetric: paraphrasing examples
- Directional: entailment examples

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Acquisition of Explicit Knowledge

The questions we need to answer

What?
   What we want to learn? Which resources do we need?

Using what?
   Which are the principles we have?

How?
   How do we organize the “knowledge acquisition” algorithm
Acquisition of Explicit Knowledge: what?

Symmetric

Co-hyponymy
Between words: cat ≈ dog

Synonymy
Between words: buy ≈ acquire
Sentence prototypes (paraphrasing): X bought Y ≈ X acquired Z% of the Y’s shares

Directional semantic relations

Words: cat → animal, buy → own, wheel partof car
Sentence prototypes: X acquired Z% of the Y’s shares → X owns Y

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Verb Entailment Relations

Given the expression

\[ \text{player wins} \]

as a selectional restriction:

\[ \text{win}(x) \rightarrow \text{play}(x) \]

as a selectional preference:

\[ P(\text{play}(x) \mid \text{win}(x)) > P(\text{play}(x)) \]

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Knowledge Acquisition

Direct Algorithms
- Concepts from text via clustering (Lin and Pantel, 2001)
- Inference rules – aka DIRT (Lin and Pantel, 2001)

Indirect Algorithms
- Hearst’s ISA patterns (Hearst, 1992)
- Question Answering patterns (Ravichandran and Hovy, 2002)

Iterative Algorithms
- Entailment rules from Web (Szepktor et al., 2004)
- Espresso (Pantel and Pennacchiotti, 2006)

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Acquisition of Implicit Knowledge

Symmetric

Acme Inc. bought Goofy ltd. ≈ Acme Inc. acquired 11% of the Goofy ltd.’s shares

Directional semantic relations

Entailment between sentences

Acme Inc. acquired 11% of the Goofy ltd.’s shares → Acme Inc. owns Goofy ltd.

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Context Sensitive Paraphrasing

He used a Phillips head to **tighten** the screw.

The bank owner **tightened** security after a spat of local crimes.

The Federal Reserve will aggressively **tighten** monetary policy.

Adapted from Dagan, Roth and Zanzotto (2007; tutorial)
Entailment Outline

Basic Definition

Task 1: Recognizing Textual Entailment (RTE)

Task 2: Examining Causality (COPA)

Task 3: Large crowd-sourced data (SNLI)
Choice of Plausible Alternatives (COPA; Roemmele et al., 2011)

Goal: test *causal* implication, not (likely) entailment

http://ict.usc.edu/~gordon/copa.html
Choice of Plausible Alternatives (COPA; Roemmele et al., 2011)

Goal: test *causal* implication, not (likely) entailment

1000 questions
Premise, prompt, and 2 plausible alternatives
Forced choice, 50% random baseline
Forward and backward causality
Cohen’s Kappa = 0.95 (only 30 disagreements)

http://ict.usc.edu/~gordon/copa.html

Adapted from Roemmele et al. (2011)
Example Items

Forward causal reasoning:
The chef hit the egg on the side of the bowl. What happened as a RESULT?

A. The egg cracked.
B. The egg rotted.

Backward causal reasoning:
The man broke his toe. What was the CAUSE of this?

A. He got a hole in his sock.
B. He dropped a hammer on his foot.
The Role of Background Knowledge

Balloons rise

Event A
The child let go of the string attached to the balloon

ok?

A causes B

Event B
The balloon flew away

Adapted from Roemmele et al. (2011)
The Role of Background Knowledge

The child let go of the string attached to the balloon

Event A

The balloon is filled with air!

Balloons rise

Bridging inference

The balloon flew away

Event B

ok?

Adapted from Roemmele et al. (2011)
Baseline Test Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Test Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMI (window of 5)</td>
<td>58.8</td>
</tr>
<tr>
<td>PMI (window of 25)</td>
<td>58.6</td>
</tr>
<tr>
<td>PMI (window of 50)</td>
<td>55.6</td>
</tr>
</tbody>
</table>

Performance of purely associative statistical NLP techniques?

Statements that are causally related often occur close together in text
   Connected by causal expressions
      (“because”, “as a result”, “so”)

Approach: choose the alternative with a stronger correlation to the premise
   PMI a la Church and Hanks, 1989

Adapted from Roemmele et al. (2011)
Goodwin et al. (2012) Approach

Adapted from Roemmele et al. (2011)
Performance of purely associative statistical NLP techniques?

Statements that are causally related often occur close together in text. Connected by causal expressions (“because”, “as a result”, “so”)

Approach: choose the alternative with a stronger correlation to the premise

PMI a la Church and Hanks, 1989

Adapted from Roemmele et al. (2011)
Updated Test Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Test Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMI (window of 5)</td>
<td>58.8</td>
</tr>
<tr>
<td>PMI (window of 25)</td>
<td>58.6</td>
</tr>
<tr>
<td>PMI (window of 50)</td>
<td>55.6</td>
</tr>
<tr>
<td>Goodwin et al.: bigram PMI</td>
<td>61.8</td>
</tr>
<tr>
<td>Goodwin et al.: SVM</td>
<td>63.4</td>
</tr>
</tbody>
</table>

Performance of purely associative statistical NLP techniques?

Statements that are causally related often occur close together in text

Connected by causal expressions (“because”, “as a result”, “so”)

Approach: choose the alternative with a stronger correlation to the premise

PMI a la Church and Hanks, 1989

Adapted from Roemmele et al. (2011)
Entailment Outline

Basic Definition

Task 1: Recognizing Textual Entailment (RTE)

Task 2: Examining Causality (COPA)

Task 3: Large crowd-sourced data (SNLI)
SNLI (Bowman et al., 2015)

Stanford Natural Language Inference corpus

https://nlp.stanford.edu/projects/snli/

570k human-written sentence pairs for entailment, contradiction, and neutral judgments

balanced dataset
Given *just* the caption for a photo:

Write one alternate caption that is definitely a true description of the photo.

Write one alternate caption that might be a true description of the photo.

Write one alternate caption that is definitely a false description of the photo.
## Examples of SNLI Judgments

<table>
<thead>
<tr>
<th>Text</th>
<th>Hypothesis</th>
<th>Judgments</th>
</tr>
</thead>
<tbody>
<tr>
<td>A man inspects the uniform of a figure in some East Asian country.</td>
<td>The man is sleeping.</td>
<td>contradiction</td>
</tr>
<tr>
<td>An older and younger man smiling.</td>
<td>Two men are smiling and laughing at the cats playing on the floor.</td>
<td>neutral</td>
</tr>
<tr>
<td>A black race car starts up in front of a crowd of people.</td>
<td>A man is driving down a lonely road.</td>
<td>contradiction</td>
</tr>
<tr>
<td>A soccer game with multiple males playing.</td>
<td>Some men are playing a sport.</td>
<td>entailment</td>
</tr>
<tr>
<td>A smiling costumed woman is holding an umbrella.</td>
<td>A happy woman in a fairy costume holds an umbrella.</td>
<td>neutral</td>
</tr>
</tbody>
</table>

Bowman et al. (2015)
SNLI (Bowman et al., 2015)

<table>
<thead>
<tr>
<th>Bowman et al. (2015)</th>
<th>SNLI Test Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexicalized</td>
<td>78.2</td>
</tr>
<tr>
<td>Unigrams Only</td>
<td>71.6</td>
</tr>
<tr>
<td>Unlexicalized</td>
<td>50.4</td>
</tr>
</tbody>
</table>

BLEU score between hypothesis and premise

# words in hypothesis - # words in premise

word overlap

unigram and bigrams in the hypothesis

Cross-unigrams: for every pair of words across the premise and hypothesis which share a POS tag, an indicator feature over the two words.

Cross-bigrams: for every pair of bigrams across the premise and hypothesis which share a POS tag on the second word, an indicator feature over the two bigrams.
**SNLI (Bowman et al., 2015)**

<table>
<thead>
<tr>
<th>Bowman et al. (2015)</th>
<th>SNLI Test Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexicalized</td>
<td>78.2</td>
</tr>
<tr>
<td>Unigrams Only</td>
<td>71.6</td>
</tr>
<tr>
<td>Unlexicalized</td>
<td>50.4</td>
</tr>
<tr>
<td>Neural: sum of word vectors</td>
<td>75.3</td>
</tr>
<tr>
<td>Neural: LSTM</td>
<td>77.6</td>
</tr>
</tbody>
</table>

**BLEU score between hypothesis and premise**

# words in hypothesis - # words in premise

**word overlap**

unigram and bigrams in the hypothesis

Cross-unigrams: for every pair of words across the premise and hypothesis which share a POS tag, an indicator feature over the two words.

Cross-bigrams: for every pair of bigrams across the premise and hypothesis which share a POS tag on the second word, an indicator feature over the two bigrams