




Inference Rules and their Application to Recognizing Textual Entailment

Georgiana Dinu & Rui Wang

Saarland University

{dinu,rwang}@coli.uni-sb.de



Inference Rules and their Application to Recognizing Textual Entailment

Georgiana Dinu & *Rui Wang*

Saarland University

{dinu,rwang}@coli.uni-sb.de



Introduction – RTE

- Definition

- Given a text **T** (several sentences) and a hypothesis **H** (one sentence), the goal is to detect whether **H** can be inferred from **T**. (Dagan et al., 2006)

- Example

- **T:** *I am giving a talk in EACL2009, Athens, where there will be a strike outside.*
- **H:** *I am in Greece, where the democracy was born.*

- The RTE Challenges

- RTE-1 (2005, 18 teams), RTE-2 (2006, 23 teams), RTE-3 (2007, 26 teams), RTE-4 (2008, 26 teams)
- The state-of-the-art performance: ~60-70%



The Gap

- Inference rules (in logic forms)
 - ...
- ?
 - ...
- ?
 - ...
- Pairs of natural language sentences



The Gap (cont.)

- Inference rules (in logic forms)
 - ...
- **DIRT**
 - ...
- **Tree Skeleton**
 - ...
- Pairs of natural language sentences



Introduction – DIRT

- Extended Distributional Hypothesis (Lin and Pantel, 2001)

- If two **paths** occur in similar contexts then they have similar meaning.

- An Example

$X \quad \leftarrow(\text{subj})- \quad \mathbf{prevent} \quad -(obj) \rightarrow \quad Y$
 $X \leftarrow(\text{subj})- \text{provide} -(obj) \rightarrow \mathbf{protection} -(mod) \rightarrow \text{against} -(pcomp) \rightarrow Y$

- Statistics
 - Over 12 million rules
 - 1 GB of newstext (parsed with Minipar (Lin, 1998))
 - Estimated accuracy: 50%
- Related Work: Marsi et al. (2007)



Problems

- Missing rules

- *X, founded in Y \rightarrow X, opened in Y*
- *X launch Y \rightarrow X produce Y*
- *X represent Y \rightarrow X work for Y*

- Incorrect rules

- *X win Y \rightarrow X lose Y*
- *X begin Y \rightarrow X end Y*



Extend DIRT

- Extension using WordNet
 - Add rules by allowing every noun or verb in the existing rules to be replaced by a synonym
 - E.g. *X face **threat** of Y ~ X at risk of Y*
→ *X face **menace** of Y ~ X at risk of Y*
- Refinement using WordNet
 - Eliminate rules which contain antonyms
 - E.g. *X **have** confidence in Y ~ X **lack** confidence in Y*



Rule Application

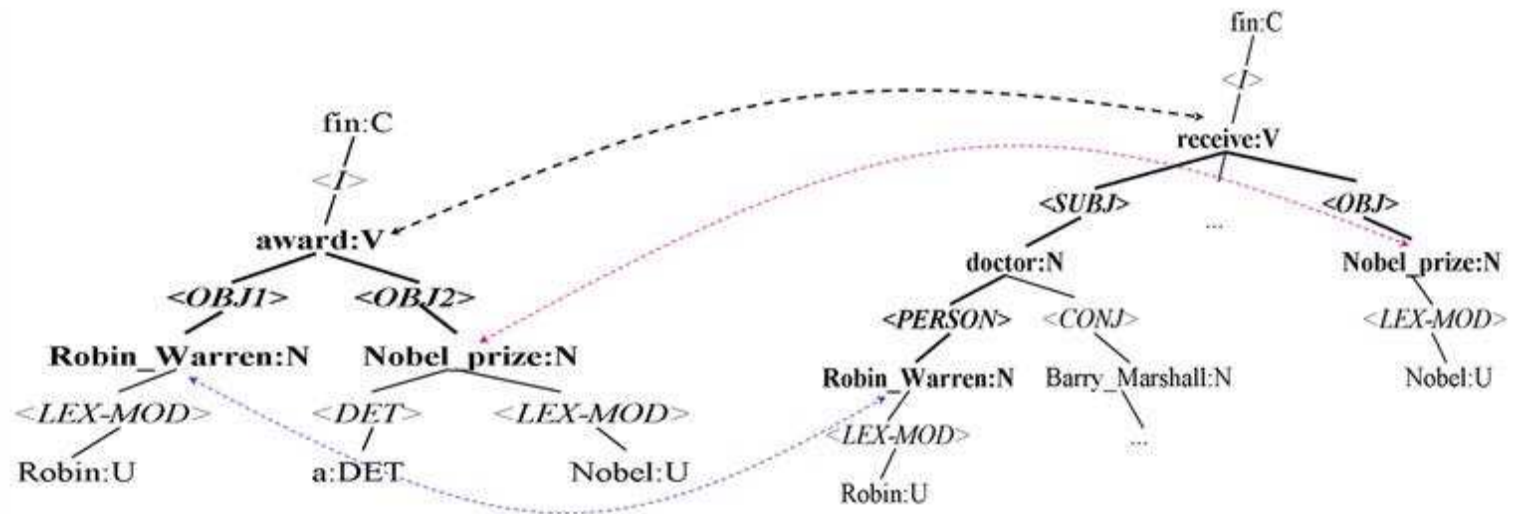
- Motivation

- To find out a proper application base for the inference rules, we use the Tree Skeleton (TS) structure proposed by (Wang and Neumann, 2007)

- An Example

- **T:** *For their discovery of ulcer-causing bacteria, Australian doctors Robin Warren and Barry Marshall have **received** the 2005 Nobel Prize in Physiology or Medicine.*
- **H:** *Robin Warren was **awarded** a Nobel Prize.*

Tree Skeleton



Application Rule: *If a tree skeleton and an inference rule are matched in a pair, then the entailment holds.*



Experiments

- Paraphrase Collections
 - Original DIRT
 - Extended DIRT (Dirt + WordNet)
- Systems
 - BoW: Bag of words overlap system
 - Main: Tree skeleton + Extended DIRT, using BoW back-up on full data sets
- Data sets
 - RTE-2: 800 (dev) + 800 (test) pairs
 - RTE-3: 800 (dev) + 800 (test) pairs



Results

RTE Set	Original DIRT	Extended DIRT
RTE-2	49 pairs / 69.38	130 pairs / 65.38
RTE-3	42 pairs / 69.04	93 pairs / 72.05

RTE Set	BoW	Main
RTE-2 (85 pairs)	51.76%	60.00%
RTE-3 (64 pairs)	54.68%	62.50%

RTE Set	BoW	Main
RTE-2 (800 pairs)	56.87%	57.75%
RTE-3 (800 pairs)	61.12%	61.75%



Error Analysis

Sources of Errors	% pairs
Incorrect Rule Application	32%
Incorrect Inference Rules	16%
Other Errors	52%

- Examples

- Someone founded "*the Institute of Mathematics (Istituto di Matematica) at the University of Milan*" **does not entail** they founded "*The University of Milan*";
- X generate $Y \sim X$ earn Y , in the **restricted** context of money or income; X issue $Y \sim X$ hit Y , it is **difficult** to find a context in which this holds;
- "*More than 1,000 members of the Russian and foreign media*" **does not entail** "*more than 1,000 members from Russia*"



Conclusion → Future Work

- Applying inference rules on recognizing textual entailment
 - Extension and Refinement of DIRT
 - High precision vs. low coverage
- ToDo List
 - The limitations of Distributional Hypothesis
 - From positive rules to negative rules
 - To be integrated into a larger framework



Acknowledgements

- The authors thank Dekang Lin and Patrick Pantel for providing the DIRT collection
- The authors thank Grzegorz Chrupała, Alexander Koller, Manfred Pinkal and Stefan Thater for very useful discussions
- The authors are funded by the IRTG and PIRE PhD scholarship programs

Thank You!

Inference Rules and their Application to Recognizing Textual Entailment

Saarland University
Campus, D-66123 Saarbrücken
(dinu, rwang)@coli.uni-sb.de

Georgiana Dinu & Rui Wang

Discovery of Inference Rules from Text (DIRT)

Extended distributional hypothesis (Lin and Pantel, 2001)
If two paths occur in similar contexts then they have similar meaning.

$$\begin{matrix} X & \xrightarrow{\text{prevent}} & Y \\ X & \xrightarrow{\text{protection}} & \text{against } Y \end{matrix}$$

Example
X put emphasis on Y → X pay attention to Y → X place emphasis on Y
X attach importance to Y → X priority of Y

The DIRT collection
Over 12 million rules (Lin and Pantel, 2001)
1 GB of metadata (paired with Moscar (Lin, 1998))
Estimated accuracy: 50% (most confident rules)

Recognizing Textual Entailment (RTE)

Definition:
Given a text T (several sentences) and a hypothesis H (one sentence), the goal is to detect if H can be inferred from T . (Dagan et al., 2006)

Example:
T: The high-speed train, scheduled for a trial run on Tuesday, is able to reach a maximum speed of up to 430 kilometers per hour, or 139 meters per second.
H: The train accelerates to 430 kilometers per hour.

The RTE Challenges:
RTE-1 (2005, 18 teams), RTE-2 (2006, 23 teams), RTE-3 (2007, 26 teams), RTE-4 (2008, 24 teams)
The state-of-the-art performance: ~60-70%

Problems

Missing and incorrect rules:
Use WordNet **synonyms** to:
a) Add rules by allowing in existing rules, every noun or verb to be replaced by a synonym
Existing rule: X face threat of Y → X at risk of Y. Add: X face menace of Y → X at risk of Y
Use WordNet **antonyms** to:
b) Eliminate rules which contain antonyms
Eliminate rule: X have confidence in Y → X lack confidence in Y

What is a correct application of a rule?
Example:
T: Libya's case against Britain and the US concerns the dispute over their demand for extradition of Libyans charged with blowing up a jetliner at over Lockerbie in 1988.
H: One case involved the extradition of Libyan suspects in the Pan Am Lockerbie bombing.
Note: by definition a rule (X pattern Y, X pattern Z) can be applied when identical nouns are instantiated for variables X and Y.

The Tree Skeleton

Motivation:
To find out a proper application of inference rules, we use the Tree Skeleton (TS) structure proposed by (Wang and Neumann, 2007).

Example:
T: For their discovery of ulcer-causing bacteria, Australian doctors **Robin Warren** and **Barry Marshall** have **received** the 2005 Nobel Prize in Physiology or Medicine.
H: **Robin Warren** was **awarded** a Nobel Prize.

Experimental Settings

Experiment: If a tree skeleton and an inference rule are matched in a pair, then decide positive entailment.
Paraphrase Collections:
Original DIRT
Extended DIRT (Dir + WordNet)
Systems:
Bow: Bag of words overlap system
Main: Tree skeleton + Extended DIRT, using Bow backup on full data sets
Data sets:
RTE-2: 800 (dev) + 800 (test) pairs
RTE-3: 800 (dev) + 800 (test) pairs

Results

RTE Set	Original DIRT	Extended DIRT
RTE-2	49 pairs / 69.38	130 pairs / 65.38
RTE-3	42 pairs / 69.04	93 pairs / 72.05

RTE Set	Bow	Main
RTE-2 (85 pairs)	51.76%	60.00%
RTE-3 (64 pairs)	54.68%	62.50%

RTE Set	Bow	Main
RTE-2 (800 pairs)	56.87%	57.75%
RTE-3 (800 pairs)	61.12%	61.75%

Error Analysis

Sources of Errors	% pairs
Incorrect Rule Application	32%
Incorrect Inference Rules	16%
Other Errors	52%

Examples:
Someone founded "the Institute of Mathematics (Institute of Mathematics) at the University of Milan" **does not entail** they founded "The University of Milan".
X generate Y → X earn Y, in the restricted context of money or money, X cause Y → X hit Y, it is difficult to find a context in which this holds.
"More than 1,000 members of the Russian and foreign media" **does not entail** "more than 1,000 members from Russia"

Conclusion

Applying inference rules on recognizing textual entailment

1. Extension and Refinement of DIRT
2. High precision vs. low coverage

Future Work

1. Enhance the tree skeleton structure
2. The limitations of Distributional Hypothesis
3. From positive rules to negative rules

Acknowledgements

We thank Dekang Lin and Patrick Pantel for providing the DIRT collection and to Gregorz Chrupala, Alexander Koller, Manfred Pinkal and Stefan Thater for very useful discussions. Georgiana Dinu and Rui Wang are funded by the IITG and FRB PhD scholarship programs.

References

(Lin and Pantel, 2001) Dekang Lin and Patrick Pantel. 2001. DIRT: Discovery of Inference Rules from Text.
(Lin, 1998) Dekang Lin. 1998. Dependency-based Evaluation of Minipar.
(Dagan et al., 2006) Ido Dagan, Oren Glickman, and Bernardo Magnini. 2006. The Pascal Recognizing Textual Entailment Challenge.
(Wang and Neumann, 2007) Rui Wang and Günter Neumann. 2007. Recognizing Textual Entailment Using a Subsequence Kernel Method.

Thank You!

Inference Rules and their Application to Recognizing Textual Entailment

Saarland University
Campus, D-66123 Saarbrücken
(dinu, rwang)@coli.uni-sb.de

Georgiana Dinu & Rui Wang

Discovery of Inference Rules from Text (DIRT)

Extended distributional hypothesis (Lin and Pantel, 2001)
If two paths occur in similar contexts then they have similar meaning.

$X \xrightarrow{r_1} \text{prevent} \xrightarrow{r_2} Y$
 $X \xrightarrow{r_3} \text{protection} \xrightarrow{r_4} Y$

Example
 X put emphasis on Y \rightarrow X pay attention to Y \rightarrow X place emphasis on Y
 X attach importance to Y \rightarrow X priority of Y

The DIRT collection
 Over 12 million rules (Lin and Pantel, 2001)
 1 GB of metadata (parsed with Hoggan Lin, 1998)
 Estimated accuracy: 50% (most confident rules)

Recognizing Textual Entailment (RTE)

Definition:
Given a text T (several sentences) and a hypothesis H (one sentence), the goal is to detect if H can be inferred from T . (Dagan et al., 2006)

Example:
 T : The high-speed train, scheduled for a trial run on Tuesday, is able to reach a maximum speed of up to 430 kilometers per hour, or 139 meters per second.
 H : The train accelerates to 430 kilometers per hour.

The RTE Challenges:
 RTE-1 (2005, 18 teams), RTE-2 (2006, 23 teams), RTE-3 (2007, 26 teams), RTE-4 (2008, 26 teams)
 The state-of-the-art performance: ~60-70%

Problems

Missing and incorrect rules:
Use WordNet synonyms to:
a) Add rules by allowing in existing rules, every noun or verb to be replaced by a synonym.
Existing rule: X face threat of Y \rightarrow X at risk of Y . **Add:** X face menace of Y \rightarrow X at risk of Y

Use WordNet antonyms to:
b) Eliminate rules which contain antonyms.
Eliminate rule: X have confidence in Y \rightarrow X lack confidence in Y

What is a correct application of a rule?
Example:
 T : Libya's case against Britain and the US concerns the dispute over their demand for extradition of Libyans charged with blowing up a Pan Am jet over Lockerbie in 1988.
 H : One case involved the extradition of Libyan suspects in the Pan Am Lockerbie bombing.
Note: In definition a rule (X pattern Y, X pattern Z Y) can be applied when identical nouns are instantiated for variables X and Y.

The Tree Skeleton

Motivation:
To find out a proper application of inference rules, we use the Tree Skeleton (TS) structure proposed by (Wang and Neumann, 2007).

Example:
 T : For their discovery of ulcer-causing bacteria, Australian doctors Robin Warren and Barry Marshall have received the 2005 Nobel Prize in Physiology or Medicine.
 H : Robin Warren was awarded a Nobel Prize.

Experimental Settings

Experiment: If a tree skeleton and an inference rule are matched in a pair, then decide positive entailment.
Paraphrase Collections:
Original DIRT
Extended DIRT (DirT + WordNet)
Systems:
Bow: Bag of words overlap system
Main: Tree skeleton + Extended DIRT, using Bow backup on full data sets
Data sets:
 RTE-2: 800 (dev) + 800 (test) pairs
 RTE-3: 800 (dev) + 800 (test) pairs

Results

RTE Set	Original DIRT	Extended DIRT
RTE-2	49 pairs / 69.38	130 pairs / 65.38
RTE-3	42 pairs / 69.04	93 pairs / 72.05

RTE Set	Bow	Main
RTE-2 (85 pairs)	51.76%	60.00%
RTE-3 (64 pairs)	54.68%	62.50%

RTE Set	Bow	Main
RTE-2 (800 pairs)	56.87%	57.75%
RTE-3 (800 pairs)	61.12%	61.75%

Error Analysis

Sources of Errors	% pairs
Incorrect Rule Application	32%
Incorrect Inference Rules	16%
Other Errors	52%

Examples:
 Someone founded "the Institute of Mathematics (Instituto de Matematica) at the University of Milan" does not entail they founded "The University of Milan".
 X generate Y \rightarrow X earn Y , in the restricted context of money or income, X cause Y \rightarrow X hit Y , it is difficult to find a context in which this holds:
 "More than 1,000 members of the Russian and foreign media" does not entail "more than 1,000 members from Russia"

Conclusion

Applying inference rules on recognizing textual entailment

- Extension and Refinement of DIRT
- High precision vs. low coverage

Future Work

- Enhance the tree skeleton structure
- The limitations of Distributional Hypothesis
- From positive rules to negative rules

Acknowledgements

We thank Dekang Lin and Patrick Pantel for providing the DIRT collection and to Gregorz Chrupala, Alexander Koller, Manfred Pinkal and Stefan Thater for very useful discussions. Georgiana Dinu and Rui Wang are funded by the IITG and FRB PhD scholarship programs.

References

(Lin and Pantel, 2001) Dekang Lin and Patrick Pantel. 2001. DIRT: Discovery of Inference Rules from Text.
 (Lin, 1998) Dekang Lin. 1998. Dependency-based Evaluation of Minipar.
 (Dagan et al., 2006) Ido Dagan, Oren Glickman, and Bernardo Magnini. 2006. The Pascal Recognizing Textual Entailment Challenge.
 (Wang and Neumann, 2007) Rui Wang and Günter Neumann. 2007. Recognizing Textual Entailment Using a Subsequence Kernel Method.

