Global Learning of Textual Entailment Graphs

Supervisors: Eytan Ruppin, Ido dagan, Shimon Edelman
15/8/12
Natural Language Understanding

• Roots in early AI (50’s)

• Information explosion
  • 2008: 1 trillion unique URLs found by Google
Organizing Textual Information

• Acute need to organize textual information
  – Based on **meaning** rather than **form**
    • Etzioni, *Search needs a shake-up*, Nature 476
  – Challenge: inference/language variability

**Auto-Text to Knowledge**
Q&A

Where was Reagan raised?

Reagan was brought up in Dixon.

... IBS is characterized by lower abdominal pain ...
NLU Applications

... Google acquires mobile E-mail utility reMail ...

<table>
<thead>
<tr>
<th>Purchase events:</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM</td>
</tr>
<tr>
<td>Yahoo!</td>
</tr>
<tr>
<td>Google</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Sacramento Kings
The Sacramento Kings are a professional basketball team based in Sacramento, California, United States. They are members of the Western Conference of the National Basketball Association. Wikipedia

Founded: 1945
Arena: Power Balance Pavilion
Head coach: Keith Smart
Division: Pacific Division
Location: Sacramento
Conference: Western Conference

<table>
<thead>
<tr>
<th>Player</th>
<th>Num</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jimmer Fredette</td>
<td>7</td>
<td>Point guard</td>
</tr>
<tr>
<td>DeMarcus Cousins</td>
<td>15</td>
<td>Center</td>
</tr>
<tr>
<td>Tyreke Evans</td>
<td>13</td>
<td>Shooting guard</td>
</tr>
<tr>
<td>Isaiah Thomas</td>
<td>22</td>
<td>Point guard</td>
</tr>
<tr>
<td>Chuck Hayes</td>
<td>42</td>
<td>Power forward</td>
</tr>
</tbody>
</table>

Global Learning of Textual Entailment Graphs
NLU Applications

• Multi-document Summarization

Barack Obama’s AIPAC address yesterday ...

Obama gave a speech last night in the Israeli lobby conference

In his speech at the American Israel Public Affairs Committee yesterday, the president challenged ...
Textual Entailment
Textual Entailment (TE) - Definition

• A directional relation between two text fragments: Text \((t)\) and Hypothesis \((h)\).

\[ t \text{ textually entails } h \ (t \Rightarrow h) \text{ if humans reading } t \text{ will infer that } h \text{ is most likely true} \]

\( t \): Usain Bolt celebrates 100m gold with Swedish women's handball team

\( h \): Usain Bolt won a gold medal

• Recognising Textual Entailment (RTE): given \( t \) and \( h \), recognize whether \( t \) textually entails \( h \)
Appeal of Textual Entailment

• Inferences required by semantic applications can be reduced to recognizing textual entailment:
  • **QA:** (Harabagiu and Hickl, 2006)
    • *Reagan was brought up in Dixon* $\Rightarrow$ *Reagan was raised in X*
  • **IE:** (Romano et al., 2006)
    • *Google acquired e-mail utility reMail* $\Rightarrow$ *X_{company} purchase Y_{company}*
  • **Summarization:** omit sentences entailed by current summary
• Led to seven challenges since 2005 for semantic inference systems
Entailment rules

• TE systems require knowledge
  – encoded as entailment rules:

  \[
  \text{giraffe} \Rightarrow \text{mammal}
  \]

  \[
  X \text{ was brought up in } Y \Rightarrow X \text{ was raised in } Y
  \]

  \[
  X_{\text{city capital of } Y_{\text{country}}} \Rightarrow X_{\text{city be part of } Y_{\text{country}}}
  \]

  \[
  X \left[ \text{VERB}_{\text{trans}} \right] Y \Rightarrow Y \text{ be } \left[ \text{VERB}_{\text{participle}} \right] \text{ by } X
  \]

• Automatic methods for learning large-scale knowledge resources is imperative
Predicative Entailment Rules

• Proposition: NL expression containing one predicate and at least one argument:
  - Andy Murray \textit{jumped up and down}
  - Ryan Lochte \textit{beat} Michael Phelps
  - Yohan Blake \textit{passed the baton to} Ussain Bolt

• Propositional template: proposition where at least one argument is replaced by a variable:
  - $X$ \textit{jumped up and down}
  - $X_{\text{player}}$ \textit{beat} $Y_{\text{player}}$
  - $X$ \textit{passed the baton to} $Y$
Predicative Entailment Rules

1. Y is a symptom of X $\Rightarrow$ X cause Y
2. X cause an increase in Y $\Rightarrow$ X affect Y
3. X’s treatment of Y $\Rightarrow$ X treat Y

• Information is conveyed by propositions
### Contribution 1: Entailment Graphs

<table>
<thead>
<tr>
<th>Statement</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>X affect Y $\Rightarrow$ X treat Y</td>
<td>✗</td>
</tr>
<tr>
<td>X treat Y $\Rightarrow$ X affect Y</td>
<td>✓</td>
</tr>
<tr>
<td>X affect Y $\Rightarrow$ X lower Y</td>
<td>✗</td>
</tr>
<tr>
<td>X lower Y $\Rightarrow$ X affect Y</td>
<td>✓</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>X lower Y $\Rightarrow$ X reduce Y</td>
<td>✓</td>
</tr>
<tr>
<td>X reduce Y $\Rightarrow$ X lower Y</td>
<td>✓</td>
</tr>
</tbody>
</table>

- **Advantage:** employ structural constraints
Contribution 2: Scaling

![Bar Chart]

- **# rules**
- **# nodes**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># rules</td>
<td># rules</td>
<td># rules</td>
</tr>
<tr>
<td></td>
<td># nodes</td>
<td># nodes</td>
<td># nodes</td>
</tr>
</tbody>
</table>

Global Learning of Textual Entailment Graphs
Contribution 3: Applications

• Public release of learned rules:
  1. High precision resource: 100,000 rules
  2. High coverage resource: 10,000,000 rules

• Textual exploration application
BIU Health-care Exploration System

Explore your search results by drilling down and up the medical concepts, their semantic relations, and their mentions in the text.

The following 22 concepts are currently supported: alcohol, asthma, biopsy, brain, cancer, CDC, chemotherapy, chest, cough, diarrhea, FDA, headache, HIV, HPV, lungs, ...
Outline

• Background
• Basic algorithm and ILP formulation
• Scalability 1: Graph decomposition
• Scalability 2: Tree-based approximation
• Text exploration application
Background
Local Learning

**Input:** pair of predicates \((p_1,p_2)\)

**Question:** \(p_1 \rightarrow p_2\) ?

- **Sources of information:**
  1. Lexicographic: WordNet (Szpektor and Dagan, 2009), FrameNet (Ben-aharon et al, 2010)
  2. Pattern-based (Chklovsky and Pantel, 2004)
  3. Distributional similarity (Lin and Pantel, 2001; Szpektor and Dagan, 2008; Bhagat et al, 2007; Yates and Etzioni, 2009; Poon and Domingos 2010; Schoenmackers et al., 2010)
Properties of Information Sources

1. Lexicographic:
   - WordNet relations: hyponym, derivation, entailment
   - Limited coverage: “X cause a reduction in Y”

2. Pattern-based:
   - “he scared and even startled me”
   - Requires very large corpus (web-scale)
   - Distinguishes different semantic relations:
     • “to X and even Y” vs. “Either X or Y”

3. Distributional similarity
## Distributional Similarity

<table>
<thead>
<tr>
<th>X affect Y</th>
<th></th>
<th></th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Y</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>insulin</td>
<td>metabolism</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Zantac</td>
<td>BP</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X treat Y</th>
<th></th>
<th></th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Y</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Zantac</td>
<td>BP</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>diet</td>
<td>diabetes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Vary w.r.t to representation and similarity computation
  - Symmetric vs. directional
- Good coverage
- Discerning exact semantic relation is difficult

**Global Learning of Textual Entailment Graphs**
Input: Set of predicates $P$

Question: Find $E = \{(p_1, p_2) \mid p_1 \Rightarrow p_2\}$

• Applying transitivity constraints:
  – Undirected graphs:
    • Paraphrasing (Yates and Etzioni, 2009)
    • Co-reference resolution (Finkel and Manning, 2008)
  – Directed graph
    • Taxonomy induction (Snow et al., 2006)
    • Ontology induction (Poon and Domingos, 2010)
    • Temporal information extraction (Ling and Weld, 2010)
Global Learning

- Snow et al. (2006) presented an algorithm for taxonomy induction.
- At each step they add the concept that maximizes the likelihood of the taxonomy given the transitivity constraint.
Basic Algorithm and ILP Formulation
Entailment Graphs

- **Nodes**: propositional templates
- **Edges**: entailment rules

Properties

- Entailment is **transitive**
- Strong connectivity components correspond to “equivalence”
  - Caveat: ambiguity
Learning Entailment Graph Edges

**Input:** Corpus $C$

**Output:** Entailment graph $G = (P,E)$

1. Extract propositional templates $P$ from $C$
2. Train a **local** entailment classifier: given $(p_1,p_2)$, estimate whether $p_1 \rightarrow p_2$
3. **Decoding:** Find the edges of the graph using the local probabilities and a **transitivity** constraint
Local Entailment Classifier – Step 1+2

Corpus

- \text{treat}(\text{Norvasc},\text{BP})
- \text{affect}(\text{Norvasc},\text{BP})
- \text{treat}(\text{insulin},\text{metab.})
- \text{affect}(\text{diet},\text{diabetes})
- \text{raise}(\text{wine},\text{fatigue})
- \text{lower}(\text{wine},\text{BP})

Distant supervision

Classifier

P = 0.7

Global Learning of Textual Entailment Graphs
Graph Objective Function

- Use local classifier probabilities $p_{ij}$ to express the graph probability:

$$\hat{G} = \arg \max_x \sum_{i \neq j} w_{ij} \cdot x_{ij}$$

$$w_{ij} = \log \frac{p_{ij} \cdot \theta}{(1 - p_{ij}) \cdot (1 - \theta)}$$

"density" prior

1 \quad i \rightarrow j
0 \quad else
Global Learning of Edges – Step 3

**Input**: Set of nodes $V$, weighting function $w: V \times V \rightarrow R$

**Output**: Set of directed edges $E$ that maximizes the objective function under a **global transitivity constraint**

• Problem is NP-hard:
  – Reduction from “Transitive Subgraph” (Yannakakis, 1978)

**Input**: Directed graph $G = (V,E)$

**Output**: Maximal set of edges $A \subseteq E$ such that $G' = (V,A)$ is transitive

• Integer Linear Programming Formulation
Integer Linear Program

\[ \hat{G} = \arg \max \sum_{i \neq j} w_{ij} \cdot x_{ij} \]

\[ \forall i, j, k \in V, x_{ij} + x_{jk} - x_{ik} \leq 1 \]

\[ x_{ij} \in \{0,1\} \]

1+1-0 = 2 > 1

- **Variables**: \( x_{ij} \)
- **Objective function**: maximizes \( P(G) \)
- **Global transitivity constraint**: \( -O(|V|^3) \) constraints
Evaluation: Focused Entailment Graphs

- Argument is instantiated by a target concept (nausea)
- Instantiating an argument reduces ambiguity
BIU Health-care Exploration System

Explore your search results by drilling down and up the medical concepts, their semantic relations, and their mentions in the text.

The following 22 concepts are currently supported: alcohol, asthma, biopsy, brain, cancer, CDC, chemotherapy, chest, cough, diarrhea, FDA, headache, HIV, HPV, lungs,

headache

- associate _ with headache | associate headache with _ (287)
- _ experience headache | _ have headache | _ suffer from headache (82)
- headache accompany _ (59)
- _ treat symptom of headache (18)
- _ treat headache (16)
  - _ relieve headache (5)
    - _ reduce headache (1)
    - _ reduce headache (1)
- symptom of _ poisoning include headache (23)
- _ accompany headache (20)
  - headache common in _ (8)
  - _ prevent headache (7)

Drug, Chemical or Other Substance (7)
Test or Procedure (3)
Occupation or Discipline (2)
Behavior or Activities (1)
Disease or Natural Phenomenon or Process (1)
Food (1)
high blood pressure (1)

Mayo Clinic - Tips for Women (and Men) in Search of a Good Night's Sleep

In addition to the usual sources coffee tea soda be aware of caffeine in chocolate in medications used to treat headaches colds sinus_congestion
Experimental Evaluation

• 50 million word tokens healthcare corpus
• Ten medical students prepared gold standard graphs for 23 medical concepts:
  – Smoking, seizure, headache, lungs, diarrhea, chemotherapy, HPV, Salmonella, Asthma, etc.
• Evaluation:
  – $F_1$ on set of learned edges vs. gold standard
Gold Standard Graph - Asthma
Evaluated algorithms

• Local algorithms
  – Distributional similarity
  – WordNet
  – Local classifier

• Global algorithms
  – ILP/Snow et al. (greedy optimization)
Results

<table>
<thead>
<tr>
<th></th>
<th>recall</th>
<th>Precision</th>
<th>F₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILP-global</td>
<td>46.0</td>
<td>50.1</td>
<td>43.8*</td>
</tr>
<tr>
<td>Greedy</td>
<td>45.7</td>
<td>37.1</td>
<td>36.6</td>
</tr>
<tr>
<td>ILP-local</td>
<td>44.5</td>
<td>45.3</td>
<td>38.1</td>
</tr>
<tr>
<td>Local₁</td>
<td>53.5</td>
<td>34.9</td>
<td>37.5</td>
</tr>
<tr>
<td>Local₂</td>
<td>52.5</td>
<td>31.6</td>
<td>37.7</td>
</tr>
<tr>
<td>Local*₁</td>
<td>53.5</td>
<td>38.0</td>
<td>39.8</td>
</tr>
<tr>
<td>Local*₂</td>
<td>52.5</td>
<td>32.1</td>
<td>38.1</td>
</tr>
<tr>
<td>WordNet</td>
<td>10.8</td>
<td>44.1</td>
<td>13.2</td>
</tr>
</tbody>
</table>

- The algorithm significantly outperforms all other baselines.
Results

<table>
<thead>
<tr>
<th></th>
<th>Global=true/Local=false</th>
<th>Global=false/Local=true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold standard = true</td>
<td>48</td>
<td>42</td>
</tr>
<tr>
<td>Gold standard = false</td>
<td>78</td>
<td>494</td>
</tr>
</tbody>
</table>

• Global algorithm avoids false positives.
Global Learning of Textual Entailment Graphs
Graph Decomposition
• **Proposition:** If we can partition the nodes to $I,J$ such that $w_{ij} < 0$ for every $i,j$ then any $(i,j)$ is not an edge in the optimal solution.
Graph Decomposition

• Entailment graphs are sparse
Scalability – Graph Decomposition

- Local solution
- Trans. violations
- Local solution
- Global Solution
- Local Solution
- Conn. components

Background → Basic algorithm and ILP formulation → **Graph decomposition** → Tree-based approximation → Text exploration application
Scalability – Graph Decomposition

1. Input
2. Undirected positive edges
3. Conn. Components
4. Apply ILP
Scalability – Graph Decomposition

5. Global solution
Decomposition Extension

Background → Basic algorithm and ILP formulation → **Graph decomposition** → Tree-based approximation → Text exploration application
Cutting-plane Method

• Applied in parsing (Riedel and Clarke, 2006)
• Number of constraints: $O(|V|^3)$
• A reasonable local classifier does not violate most of the constraints
• Add constraints only if they are violated
Cutting-plane Method

Constraints:
1. (c,a,b)
2. (d,b,a)
3. (d,b,c)

• Empirically converges in 6 iterations and reduces number of constraints from $10^6$ to $10^3$-$10^4$
Final Algorithm

• Given a set of predicates:
  – Decompose into components
  – Apply a cutting-plane method on each component
Evaluation: Typed Entailment Graphs

• Predicate variables are typed:
  \[ X_{\text{company}} \text{ acquire } Y_{\text{company}} \implies Y_{\text{company}} \text{ is sold to } X_{\text{company}} \]

• Rules of wide-applicability but less ambiguous

• Schoenmackers et al. (EMNLP 2010) used a local algorithm to learn 30,000 entailment rules

• We want to use a global algorithm
Typed Entailment Graphs

• A graph is defined for every pair of types
• Nodes are termed typed predicates
• Types alleviate the ambiguity problem
Experiment 1 - Transitivity

- 1 million TextRunner tuples over 10,672 typed predicates and 156 types
- Consist ~2,000 typed entailment graphs
- 10 gold standard graphs of sizes: 7, 14, 22, 30, 38, 53, 59, 62, 86 and 118
- Evaluation:
  - $F_1$ on set of edges vs. gold standard
  - Area Under the Curve (AUC)
Evaluated algorithms

• Local algorithms
  – Distributional similarity (DIRT, BInc, etc.)
  – Distributional similarity + background knowledge
  – ILP with No transitivity constraints
  – Sherlock rule resource

• Global algorithm: ILP
Precision-Recall: Global vs. Local

Background → Basic algorithm and ILP formulation → **Graph decomposition** → Tree-based approximation → Text exploration application

Global Learning of Textual Entailment Graphs
Precision-Recall: Global vs. Local

![Graph showing precision vs. recall for Global and Local1-Local4 models.](image)
Background  →  Basic algorithm and ILP formulation  →  **Graph decomposition**  →  Tree-based approximation  →  Text exploration application

**Precision-Recall: Global vs. Local**

- **Global**
- **Sherlock**

Global Learning of Textual Entailment Graphs
## Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Micro-average</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
</tr>
<tr>
<td>ILP</td>
<td>43.4</td>
<td>42.2</td>
</tr>
<tr>
<td>Clsf</td>
<td>30.8</td>
<td>37.5</td>
</tr>
<tr>
<td>Sherlock</td>
<td>20.6</td>
<td>43.3</td>
</tr>
<tr>
<td>SR</td>
<td>38.4</td>
<td>23.2</td>
</tr>
<tr>
<td>DIRT</td>
<td>25.7</td>
<td>31.0</td>
</tr>
<tr>
<td>BINC</td>
<td>31.8</td>
<td>34.1</td>
</tr>
</tbody>
</table>

- R/P/F₁ at point of maximal micro-F₁
- Transitivity improves rule learning over typed predicates
Experiment 2 - Scalability

• Run ILP with and without graph decomposition over ~2,000 graphs
• Compare number of learned rules for various sparseness parameters
Results

<table>
<thead>
<tr>
<th>Sparsity</th>
<th># learned rules</th>
<th>Δ(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.75</td>
<td>6,242/7,466</td>
<td>+20</td>
</tr>
<tr>
<td>-1</td>
<td>16,790/19,396</td>
<td>+16</td>
</tr>
<tr>
<td>-0.6</td>
<td>26,330/29,732</td>
<td>+13</td>
</tr>
</tbody>
</table>

• Scaling techniques add 3,500 new rules to best configuration
• Corresponds to 13% increase in relative recall
Tree-based Approximation
Tree-based Approximation - Outline

1. Forest-reducible graphs (FRG)
2. Tree-Node-Fix algorithm (TNF)
3. Experiments:
   1. Typed entailment graph
   2. Untyped entailment graphs
Directed Forest

• For an edge $i \rightarrow j$ in a DAG we say that $i$ is a child of $j$ and $j$ a parent of $i$

• A directed forest is a DAG where all nodes have no more than one parent
Entailment Graphs are not Forests

- But generally entailment is from "specific" predicates to more "general" ones
Computing SCC Graph

- Contract strongly connected components.
Transitive Reduction

- Delete edges inferred by transitivity
Reduced Graphs

- **Reduced graphs** are an equivalent representation for transitive graphs.
- The reduced graph in this case is a directed forest.
Forest-reducible Graphs (FRG)

- **Hypothesis**: The reduced graph of entailment graphs is usually a directed forest, i.e., entailment graphs are **FRGs**
- Not necessarily true:

```
X person plan to attack Y person
```
```
X person wait for Y person
```
```
X person ambush Y person
```

- But exceptions are rare...
FRG Assumption is reasonable

• Over our Typed entailment graph data set:
  – We manually constructed FRGs >= .95 recall and 1.0 precision on 7 gold standard graphs

• Finding the best FRG is still NP-hard, by a polynomial reduction from X3C.

**Input**: Set $X$ of size $3n$, subsets $S_1, S_2, ..., S_m$ of size 3

**Output**: Subsets $S_{i1}, ..., S_{in}$ covering $X$
FRG Summary

• Entailment graphs are often FRGs
  – Interesting linguistic observation
  – Leads to an efficient method for learning edges
Tree-based Approximation - Outline

• Forest-reducible graphs (FRG)
• Tree-Node-Fix algorithm (TNF)
• Experiments:
  – Typed entailment graph
  – Large untyped entailment graphs + resource
Sequential Approximation

**Input**: Nodes $V$ and weighting function $w$

1. *Initialize graph* (e.g., empty graph)
2. For every node $v$ in $V$:
   1. **Re-attach** $v$ in a locally-optimal way
3. Until convergence
**Sequential Approximation**

**Input**: Nodes $V$ and weighting function $w$
1. **Initialize graph** (e.g., empty graph)
2. For every node $v$ in $V$:
   1. **Re-attach** $v$ in a locally-optimal way
3. Until convergence

**Candidate edges:**
- $(1, 2)$
- $(2, 1)$
- $(2, 3)$
- $(3, 2)$
- $(2, 4)$
- $(4, 2)$
- $(2, 5)$
- $(5, 2)$
Sequential Approximation

**Input**: Nodes $V$ and weighting function $w$

1. Initialize graph (e.g., empty graph)
2. For every node $v$ in $V$:
   1. **Re-attach** $v$ in a locally-optimal way
3. Until convergence

**Candidate edges**:
- $(1, 2)$
- $(2, 1)$
- $(2, 3)$
- $(3, 2)$
- $(2, 4)$
- $(4, 2)$
- $(2, 5)$
- $(5, 2)$
Sequential Approximation

**Input:** Nodes $V$ and weighting function $w$

1. Initialize graph (e.g., empty graph)
2. For every node $v$ in $V$:
   1. Re-attach $v$ in a locally-optimal way
3. Until convergence

**Candidate edges:**
- $(1, 4)$
- $(4, 1)$
- $(2, 4)$
- $(4, 2)$
- $(3, 4)$
- $(4, 3)$
- $(4, 5)$
- $(5, 4)$
Sequential Approximation

**Input:** Nodes $V$ and weighting function $w$

1. Initialize graph (e.g., empty graph)
2. For every node $v$ in $V$:
   1. Re-attach $v$ in a locally-optimal way
3. Until convergence
Tree-Node-Fix: Re-attach $v$

**Input:** Graph $G$, node $v$ in $V$, function $w$

1. **Delete $v$**
2. **Compute reduced graph $G_{red}$** (directed forest)
3. **Re-attach $v$ in a locally-optimal way.**

Node re-attachment in FRGs takes linear time!
Objective Function

\[ \arg \max \sum_{i \neq v} W_{iv} \cdot X_{iv} + \sum_{k \neq v} W_{vk} \cdot X_{vk} \]

Incoming edges
Outgoing edges
Scores for Components

\[ S_{v-in}(c) = \sum_{i \in c} w_{iv} + \sum_{d \in \text{child}(c)} S_{v-in}(d) \]

Global Learning of Textual Entailment Graphs
Definitions

\[ S_{v\text{-in}}(c) = \sum_{i \in c} w_{iv} + \sum_{d \in \text{child}(c)} S_{v\text{-in}}(d) \]
Definitions

\[ S_{v-out}(c) = \sum_{i \in c} w_{vi} + S_{v-out}(parent(c)) \]

Global Learning of Textual Entailment Graphs
Definitions

\[ S_{v-out}(c) = \sum_{i \in c} w_{vi} + S_{v-out}(\text{parent}(c)) \]

\[ S_{v-out}(c_5) \]
TNF – case 1

- Insertion into one of the components

\[ S_{v - in}(c) + S_{v - out}(c) \]
TNF – case 2

- Choose a **single** parent
- Only direct children of $p$ can become children of $v$
- Once a parent is chosen, choose children **independently**

$$S_{v-out}(p) + \sum_{c} \max(0, S_{v-in}(c))$$

Global Learning of Textual Entailment Graphs
**TNF – case 3**

- Choose no parent (forest root)
- Choose children independently from roots.

\[ \sum_{r \in \text{roots}} \max(0, S_v - \text{in}(r)) \]
TNF - Summary

• Computing $S_{v-in}(c)$ and $S_{v-out}(c)$ is linear (DP)
• Choosing best case is linear
• Re-attachment is linear
• TNF is quadratic
• **Extension**: Tree-Node-Component-Fix (TNCF)
  – Allow re-attachment of reduced graph components.
## Complexity Results

<table>
<thead>
<tr>
<th></th>
<th>Transitive Graph</th>
<th>FRG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire graph</td>
<td>NP-hard</td>
<td>NP-hard</td>
</tr>
<tr>
<td>Node re-attachment</td>
<td>?</td>
<td>Linear</td>
</tr>
</tbody>
</table>

Background → Algorithm scheme and ILP formulation → Graph decomposition → **Tree-based approximation** → Future work
Outline

• Forest-reducible graphs (FRG)
• Tree-Node-Fix algorithm (TNF)
• Experiments:
  – Typed entailment graph
  – Large untyped entailment graphs + resource
Experiment 1: Typed Graphs

- ACL 11' experimental setting
  - Medium-size graphs
  - Predicates/nodes are disambiguated
Evaluated Algorithms

- No-trans: no transitivity constraints ($w_{ij}$)
- Exact-graph: ACL 11’ exact method
- Exact-forest: Find optimal FRG with ILP solver
- LP-relax (Martins et al., 09’)
- Tree-Node-Fix (TNF)
- Graph-Node-Fix (GNF)
Run time

![Graph showing run time comparison between different algorithms: Exact-graph, LP-relax, GNF, and TNF. The x-axis represents the data points, while the y-axis shows the run time in seconds. The graph illustrates the performance of each algorithm as the data points increase.](image-url)
Background → Basic algorithm and ILP formulation → Graph decomposition → **Tree-based approximation** → Text exploration application
Experiment 2: Untyped Graphs

• Large graph: 10,000 predicates
• Based on REVERB (fader et al., 11’)
• Predicates are ambiguous
  – Transitivity and FRG assumption
• Features: distributional similarity, lexicographic and string-similarity features for every pair of predicates
Training and Test Set Generation

• Gold standard annotation:
  – Instance-based evaluation with crowdsourcing (Zeichner et al., 2012)
  – Training set: ~3,300 pairs of predicates
  – Test set: ~1,750 pairs of predicates

• Trained classifier provides $w_{ij}$

• We managed to run only No-trans, TNF, and TNCF
Experiment 2 - Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall</th>
<th>Precision</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>No-trans</td>
<td>0.14</td>
<td>0.7</td>
<td>0.24</td>
</tr>
<tr>
<td>TNF</td>
<td>0.15</td>
<td>0.7</td>
<td>0.25</td>
</tr>
<tr>
<td>TNCF</td>
<td>0.16</td>
<td>0.8</td>
<td>0.27</td>
</tr>
</tbody>
</table>

- FRG and Transitivity assumptions limit recall
- TNCF – improved precision for recall ~ 0.15
- *High precision resource of > 100,000 entailment rules*
Resource

• High-precision resource of > 100,000 entailment rules
• In addition, 15 million rules learned over 100,000 predicates
  – Outperforms DIRT (Lin and Pantel, 2001)
• Syntactic and lexical representations
• http://u.cs.biu.ac.il/~nlp/downloads/
BIU Health-care Exploration System

Explore your search results by drilling down and up the medical concepts, their semantic relations, and their mentions in the text.

The following 22 concepts are currently supported: alcohol, asthma, biopsy, brain, cancer, CDC, chemotherapy, chest, cough, diarrhea, FDA, headache, HIV, HPV, lungs,

headache

- associate ___ with headache | associate headache with ___ (287)
- ___ experience headache | ___ have headache | ___ suffer from headache (82)
- headache accompany ___ (59)
- ___ treat symptom of headache (18)
- treat headache (16)
  - ___ relieve headache (5)
    - ___ reduce headache (1)
    - ___ reduce headache (1)
  - symptom of ___ poisoning include headache (23)
  - ___ accompany headache (20)
    - headache common in ___ (8)
  - ___ prevent headache (7)

Drug, Chemical or Other Substance (7)
Test or Procedure (3)
Occupation or Discipline (2)
Behavior or Activities (1)
Disease or Natural Phenomenon or Process (1)
Food (1)
high blood pressure (1)

Mayo Clinic - Tips for Women (and Men) in Search of a Good Night's Sleep

In addition to the usual sources coffee tea soda be aware of caffeine in chocolate in medications used to treat headaches colds sinuses congestion
Local Results
Re-attachment Examples

mean a lot of

generate a lot of

give a lot of

offer plenty of

cause of

cause

offer a lot of

offer a wealth of

provide a lot of

provide plenty of

Global Learning of Textual Entailment Graphs
Re-attachment Examples

- **convert**
- **convinced**
- **encrypt**
- **code**
- **encode**
- **compress**
Re-attachment Examples

deter

discourage

depress

bring down

lower
Conclusions

• NLU Applications require large knowledge resources of predicative entailment rules

• In this dissertation we presented
  – A model that utilizes graph structure to improve rule learning
  – Algorithms that scale for learning entailment graphs
  – A large knowledge resource of rules
  – An application for text exploration

• Future work: ambiguity
Thanks for coming!