Entailment Graphs for Text exploration

Ido Dagan, Bar Ilan University, Israel
Bernardo Magnini, FBK, Trento, Italy
Outline

1. (Ido)
   Entailment graphs as a basis for
   - Text exploration
   - Knowledge representation

2. (Bernardo)
   Entailment graphs for customer interactions
   - Building and merging entailment graphs
   - Toward integration into application scenarios
Motivation: Text Exploration

Blood pressure - Wikipedia, the free encyclopedia
Blood pressure (BP) is the pressure exerted by circulating blood upon the walls of blood vessels, and is one of the principal vital signs. When used without ...

What Is High Blood Pressure? - NHLBI NH
What Is High Blood Pressure? - NHLBI NH
Blood pressure is the force of blood against the walls of arteries as the heart pumps blood. If this pressure rises and stays high over time, it can ...

Blood pressure is the force of blood against the walls of arteries. Blood pressure is recorded as two numbers—the systolic pressure (as the heart beats) over the ...

Blood Pressure Chart - Normal Blood Pressure Range
Blood Pressure Chart - Normal Blood Pressure Range
Normal blood pressure range chart, with comments about each blood pressure level. High, normal and low blood pressure are defined.

Understanding Blood Pressure Readings
Understanding Blood Pressure Readings
6 Jun 2012 — What are blood pressure readings? Systolic, diastolic? The American Heart Association helps you understand the various levels of blood...
Goal

- Guide users through concrete facts in unfamiliar domains

- Common exploration approaches – via concepts/categories
  - “what are the documents talking about?”

- Our goal – exploration of statements/facts (propositions)
  - “what do the documents actually say?”
  - Full propositions capture the knowledge within texts
Broader Motivation: NL-based Knowledge Representation

- A Knowledge Representation (KR) framework based on natural language constructs
  - Rather than on formal pre-defined terminology
  - When texts are a primary source of knowledge
- Useful for many knowledge-related tasks
  - Information exploration
  - Question answering / fact search
  - Update summaries
  - Modeling user’s knowledge
  - …
Existing KR Approaches

1. Formal representation languages
   - Pre-defined set of target predicates (concepts)
   - Examples:
     - FreeBase / Google Knowledge Graph
     - Semantic-web ontologies
   - Limitations: scalability, supervised extractors

2. Open IE
   - Unsupervised proposition extraction as NL predicate-argument tuples
   - Lack canonization and structure
Goal: NL4KR

- Extend Open-IE to be a viable KR model

- First step: *Statement Entailment Graphs*
  - Nodes: “atomic” NL statements (propositions), comprising of
    - Predicate and arguments
    - Modifiers
  - Edges: entailment relations
Customer Interactions Entailment Graph

- coffee in economy is awful
- no vegetarian food
- journey is too slow
- no clear information
- not happy with the service
- not happy with the catering
- not enough food selection
- food on train is too expensive
- food is bad
- food quality is disappointing
- bad food in premier
- sandwiches are too expensive
- sandwiches are overpriced
- you charge too much for sandwiches
- staff is unfriendly
Entailment-based Text Exploration
Applied to the Health-care Domain

Meni Adler, Jonathan Berant, Ido Dagan
ACL 2012 Demo
Prospects: Extensions and Formal KR

- Extensions of NL Knowledge Graphs
  - Open-IE proposition structure
    - Arity, modifiers, truth status
  - Graph relationships (causality, temporal,...)

- Examine formal-language/NL synergy for KR
  - Integrating text knowledge with expert knowledge
  - NL-graphs vs. emergent schemas for ontologies
  - Cross-leveraging tools/methods
  - *NL as Interlingua for KR-languages/ontologies?*
Outline

1. (Ido)
   Entailment graphs as a basis for
   - Text exploration
   - Knowledge representation

2. (Bernardo)
   Entailment graphs for customer interactions
   - Building and merging entailment graphs
   - Toward integration into application scenarios
Customer interactions

Int-448:
Efficient service. Quick through security and check in. Staff could have been a bit more friendly though and leg room in standard class was quite poor.

Int-202:
Everything ran smoothly and well. Only complaint is lack of leg room with seating with tables. Very cramped when all seats are taken.

Int-275:
The leg room in economy class is not enough I was constantly being kicked by opposite passenger I travel by train lots and this compares badly to other trains

Int-303:
My only gripes, not enough leg room in standard and I think it would be chic to have refreshments served in carriages, either trolley or trays like in theatres.
EXCITEMENT application scenario

Requirements

- Need for customer interaction analytics
  - Compact and informative representation
- Multiple channels: e-mail, speech, social media
- Multiple languages
  - Excitement: English, Italian, German
- Need to manage streams of data

Proposal

- Core technology: entailment graphs based on the EOP platform
Entailment Graph Creation Process

1) EXTRACT RELEVANT STATEMENTS FROM SINGLE INTERACTIONS

- ANNOTATION OF:
  - FRAGMENTS
  - FRAGMENT METADATA
  - FRAGMENT MODIFIERS
    (manual)

2) BUILD AN ENTAILMENT GRAPH FOR EACH STATEMENT

- BUILD-GRAPH PROCEDURE
  (auto)
- FRAGMENT GRAPHS
- XML OUTPUT
- (SUB)FRAGMENT CORRECTION + SUBFRAGMENT METADATA ANNOTATION
  (manual)

3) MERGE ENTAILMENT GRAPHS FROM DIFFERENT INTERACTIONS

- SELECTION OF ENTAILMENT PAIRS
  (auto)
- ENTAILMENT ANNOTATION
  (manual)
- ANNOTATION CONSISTENCY CHECK
  (auto)
- GRAPH MERGING
  (auto)

FINALGRAPH
XML OUTPUT
1) Extracting statements from interactions

*Int-448:*
Efficient service. Quick through security and check in. Staff could have been a bit more friendly though and leg room in standard class was quite poor.

*Int-202:*
Everything ran smoothly and well. Only complaint is lack of leg room with seating with tables. Very cramped when all seats are taken.

*Int-275:*
The leg room in economy class is not enough I was constantly being kicked by opposite passenger I travel by train lots and this compares badly to other trains.

*Int-303:*
My only gripes, not enough leg room in standard and I think it would be chic to have refreshments served in carriages, either trolley or trays like in theatres.
2) Creating entailment graphs for statements

**Int-448:**
Efficient service. Quick through security and check in. Staff could have been a bit more friendly though and leg room in standard class was quite poor.

**F2_S1:** leg room was **quite** poor  
**F2_S2:** leg room **in standard class** was poor  
**F2_S3:** leg room was poor

- A statement is decomposed into its entailed sub-statements  
- Goal: maximize connectivity among interactions
2) Creating entailment graphs for statements

**Int-303**

- Not enough leg room in standard
- Not enough leg room

**Int-448**

- Leg room in standard class was quite poor
- Leg room in standard class was poor
- The leg room is not enough
- leg room was quite poor
- F2_S1
- F2_S2
- F2_S3
- F2
3) Merging entailment graphs

*Int-303*

- Not enough leg room in standard
- F1
- F1_S1
- Not enough leg room

*Int-448*

- Leg room in standard class was quite poor
- F2
- F2_S1
- Leg room was quite poor
- F2_S2
- Leg room in standard class was poor
- F2_S3
- The leg room is not enough
The resulting entailment graph

- **Int-202 F1**: Lack of leg room with seating with tables
- **Int-275 F1**: The leg room in economy class is not enough
- **Int-448 F2**: Leg room in standard class was quite poor

**Int-202 F1_S1**: lack of leg room

**Int-275 F1_S1**: the leg room is not enough

**Int-448 F2_S3**: leg room was poor

**Int-303 F1_S1**: not enough leg room

- **Int-448 F2_S2**: leg room in standard class was poor
- **Int-303 F1**: Not enough leg room in standard
## English Email Dataset

**FINAL DATASET**

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Interactions</th>
<th>Fragments (F + subF)</th>
<th>Nodes (direct)</th>
<th>Edges (direct)</th>
<th>Build phase (relations between different fragment graphs)</th>
<th>Merge phase (relations between different fragment graphs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Edges (direct)</td>
<td>Nodes</td>
</tr>
<tr>
<td>EMAIL0220</td>
<td>4</td>
<td>6</td>
<td>13</td>
<td>27</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>EMAIL0230</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>20</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>EMAIL0240</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>15</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>EMAIL0250</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>EMAIL0310</td>
<td>11</td>
<td>14</td>
<td>37</td>
<td>83</td>
<td>30</td>
<td>24</td>
</tr>
<tr>
<td>EMAIL0320</td>
<td>12</td>
<td>13</td>
<td>24</td>
<td>99</td>
<td>13</td>
<td>21</td>
</tr>
<tr>
<td>EMAIL0330</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>EMAIL0340</td>
<td>14</td>
<td>14</td>
<td>22</td>
<td>88</td>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>EMAIL0350</td>
<td>13</td>
<td>21</td>
<td>33</td>
<td>65</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>EMAIL0360</td>
<td>7</td>
<td>8</td>
<td>19</td>
<td>28</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>EMAIL0370</td>
<td>3</td>
<td>6</td>
<td>11</td>
<td>6</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>EMAIL0380</td>
<td>7</td>
<td>7</td>
<td>14</td>
<td>18</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>EMAIL0390</td>
<td>6</td>
<td>6</td>
<td>13</td>
<td>10</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>EMAIL0410</td>
<td>8</td>
<td>11</td>
<td>15</td>
<td>8</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>28</strong></td>
<td><strong>458</strong></td>
<td><strong>840</strong></td>
<td><strong>3490</strong></td>
<td><strong>369</strong></td>
<td><strong>506</strong></td>
</tr>
</tbody>
</table>
1. Process TextualInput (create CAS, add linguistic and fragment annotation)

2. Create input to fragment graph building (detect base predicates and modifiers)

3. Build fragment graph (based on WP2 build-graph procedure)

4. Update graph (based on WP2 merge-graph procedure: select and process T/H pair candidates; call LAP and store annotation results, detect/resolve inconsistencies, etc.)

5. Collapse graph (select edges, create equivalence classes)

Collection of textual inputs

TextualInput

Annotated CAS

(Updated graph)

EntailmentGraphRaw

EntailmentGraphCollapsed

Final graph

INDUSTRIAL PLATFORMS

OPEN PLATFORM

DECOMPOSITION

COMPOSITION
Ongoing work in the Excitement project

- Extract relevant statements
  - recognize modifiers that preserve entailment
- Merging entailment graphs
  - strategies for controlling the computational complexity (i.e. number of calls to the EOP)
- Optimize global consistency of the graph
  - a new task (pairs in RTE datasets are assumed to be independent)