Modeling Context in Scenario Template Creation

Long Qiu, Min-Yen Kan, Tat-Seng Chua

National University of Singapore

IJCNLP 2008, Hyderabad
Scenario Template: a hierarchy of slots to capture the Salient Aspects of a scenario;

Aviation Disaster 1:
All 143 people aboard a Russian passenger plane are reported to have died after it crashed near the Siberian city of Irkutsk. ...

Aviation Disaster 2:
The American Defence Department has announced that all seven marines on board a military plane that crashed in Pakistan on Wednesday were killed in the incident. ...
Scenario Template: a hierarchy of slots to capture the Salient Aspects of a scenario;

Aviation Disaster 1:
All 143 people aboard a Russian passenger plane are reported to have died after it crashed near the Siberian city of Irkutsk. ...

Aviation Disaster 2:
The American Defence Department has announced that all seven marines on board a military plane that crashed in Pakistan on Wednesday were killed in the incident. ...
Introduction: Scenario Template

- **Scenario Template**: a hierarchy of slots to capture the **Salient Aspects** of a scenario;

**Aviation Disaster 1:**
All 143 people aboard a **Russian passenger plane** are reported to have died after it **crashed** near the **Siberian city of Irkutsk**. ...

**Aviation Disaster 2:**
The American Defence Department has announced that all seven marines on board a **military plane** that **crashed** in **Pakistan** on **Wednesday** were killed in the incident. ...
Introduction: Scenario Template

Scenario Template: a hierarchy of slots to capture the Salient Aspects of a scenario;

Aviation Disaster 1:
All 143 people aboard a Russian passenger plane are reported to have died after it crashed near the Siberian city of Irkutsk. ...

Aviation Disaster 2:
The American Defence Department has announced that all seven marines on board a military plane that crashed in Pakistan on Wednesday were killed in the incident. ...
Scenario Template: a hierarchy of slots to capture the Salient Aspects of a scenario;

Aviation Disaster 1:
All 143 people aboard a Russian passenger plane are reported to have died after it crashed near the Siberian city of Irkutsk. ...

Aviation Disaster 2:
The American Defence Department has announced that all seven marines on board a military plane that crashed in Pakistan on Wednesday were killed in the incident. ...
Introduction: Scenario Template

- **Scenario Template**: a hierarchy of slots to capture the **Salient Aspects** of a scenario;
- **Scenario Template Creation (STC)**:
Introduction: Scenario Template

- **Scenario Template**: a hierarchy of slots to capture the **Salient Aspects** of a scenario;
- **Scenario Template Creation (STC)**:
  - Input: a set of articles on **different** events of the **same** scenario;

  15 Apr., 2002: “Chinese plane crashes in S Korea”
  20 Feb., 2004: “Iranian plane crashes in Emirates”
  10 Jan., 2002: “No survivors in US plane crash”
  3 Jul., 2001: “Siberia air crash kills 143”
  29 Jun., 2004: “UN staff die in S Leone air crash”
**Introduction: Scenario Template**

- **Scenario Template**: a hierarchy of slots to capture the **Salient Aspects** of a scenario;
- **Scenario Template Creation (STC)**:
  - Output:

<table>
<thead>
<tr>
<th>Scenario Name</th>
<th>Aviation Disaster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td></td>
</tr>
<tr>
<td>Airliner</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>Casualty</td>
<td></td>
</tr>
<tr>
<td>Cause</td>
<td></td>
</tr>
</tbody>
</table>
**Scenario Template**: a hierarchy of slots to capture the **Salient Aspects** of a scenario;

**Scenario Template Creation (STC):**
- Output:

<table>
<thead>
<tr>
<th>Scenario Name</th>
<th>Aviation Disaster</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Date</strong></td>
<td>July 03, 2001</td>
</tr>
<tr>
<td><strong>Airliner</strong></td>
<td>Tupolev 154, Vladivostokavia</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td>Siberian city of Irkustk</td>
</tr>
<tr>
<td><strong>Casualty</strong></td>
<td>133 passengers, 10 crew</td>
</tr>
<tr>
<td><strong>Cause</strong></td>
<td>(under investigation)</td>
</tr>
</tbody>
</table>
1. Introduction: Scenario Template

2. Related Work

3. A Graph-based Approach

4. Evaluation

5. Conclusion
Related Work

- **Subject-Verb-Object(-PP)**
  - Collier, 1996;
  - Harabagiu and Maiorano, 2002, *etc.*

- **Dependency subtrees**
  - Sudo *et al.*, 2003

- **Syntactic subtrees**
  - Filatova *et al.*, 2006

- **Semantic roles**
  - Harabagiu and Lacatusu, 2005
Related Work

- **Subject-Verb-Object(-PP)**
  - Collier, 1996;
  - Harabagiu and Maiorano, 2002, etc.

- Dependency subtrees
  - Sudo *et al.*, 2003

- Syntactic subtrees
  - Filatova *et al.*, 2006

- Semantic roles
  - Harabagiu and Lacatusu, 2005
Related Work

- **Subject-Verb-Object(-PP)**
  - Collier, 1996;
  - Harabagiu and Maiorano, 2002, *etc.*

- **Dependency subtrees**
  - Sudo *et al.*, 2003

- **Syntactic subtrees**
  - Filatova *et al.*, 2006

- **Semantic roles**
  - Harabagiu and Lacatusu, 2005

... people died ...
... marines were killed ...
(people subj (gov die))
(marine obj (gov kill))
Related Work

- **Subject-Verb-Object(-PP)**
  - Collier, 1996;
  - Harabagiu and Maiorano, 2002, *etc.*
- **Dependency subtrees**
  - Sudo *et al.*, 2003
- **Syntactic subtrees**
  - Filatova *et al.*, 2006
- **Semantic roles**
  - Harabagiu and Lacatusu, 2005

Example: "... people died ..." and "... marines were killed ..."
Related Work

- **Subject-Verb-Object(-PP)**
  - Collier, 1996;
  - Harabagiu and Maiorano, 2002, *etc.*

- **Dependency subtrees**
  - Sudo *et al.*, 2003

- **Syntactic subtrees**
  - Filatova *et al.*, 2006

- **Semantic roles**
  - Harabagiu and Lacatusu, 2005

... people died ...
... marines were killed ...
people<$\text{ARG1}$> die<$\text{PREDICATE}$>
marine<$\text{ARG1}$> kill<$\text{PREDICATE}$>
Related Work

- Subject-Verb-Object(-PP)
  - Collier, 1996;
  - Harabagiu and Maiorano, 2002, etc.
- Dependency subtrees
  - Sudo et al., 2003
- Syntactic subtrees
  - Filatova et al., 2006
- Semantic roles
  - Harabagiu and Lacatusu, 2005
- Semantic Role Labeler + Thesauri
Related Work

- **Subject-Verb-Object(-PP)**
  - Collier, 1996;
  - Harabagiu and Maiorano, 2002, *etc.*

- **Dependency subtrees**
  - Sudo *et al.*, 2003

- **Syntactic subtrees**
  - Filatova *et al.*, 2006

- **Semantic roles**
  - Harabagiu and Lacatusu, 2005

- **Semantic Role Labeler + Thesauri**

- **Not complete without considering Context!**
Motivation

Charley **landed** further south. The hurricane

The storm **hit** on Wednesday and over.
Charley **landed** further south on Florida’s Gulf coast than predicted, ...
The hurricane ... has *weakened* and is *moving* over South Carolina.

At least 21 others are missing after the storm **hit** on Wednesday. ....
But Tokage had *weakened* by the time it passed over Japan’s capital, Tokyo, where it left little damage before *moving* out to sea.
Motivation

Charley **landed** further south on Florida’s Gulf coast than predicted, ...
... The hurricane ... has **weakened** and is **moving** over South Carolina.

At least 21 others are missing after the storm **hit** on Wednesday. ....
But Tokage had **weakened** by the time it passed over Japan’s capital, Tokyo, where it left little damage before **moving** out to sea.
Charley landed further south on Florida’s Gulf coast than predicted, ... The hurricane ... has weakened and is moving over South Carolina.

At least 21 others are missing after the storm hit on Wednesday. .... But Tokage had weakened by the time it passed over Japan’s capital, Tokyo, where it left little damage before moving out to sea.
Motivation

Contextual evidence could be helpful:

| Charley **landed** further south on Florida’s Gulf coast than predicted, ...
| The hurricane ... has **weakened** and is **moving** over South Carolina. |
| At least 21 others are missing after the storm **hit** on Wednesday. ....
| But Tokage had **weakened** by the time it passed over Japan’s capital, Tokyo, where it left little damage before **moving** out to sea. |
Context-Sensitive Clustering

w/o Context

B6 \quad B1
B5 \quad B4 \quad B2
B3

A1 \quad Y4
A2
A3
A4
A5
A6

TextSpan_{\text{article\_ID}}

Long Qiu, Min-Yen Kan, Tat-Seng Chua
Context-Sensitive Clustering

w/o Context

A Graph-based Approach

Evaluation

Conclusion

Long Qiu, Min-Yen Kan, Tat-Seng Chua

Modeling Context in Scenario Template Creation
Context-Sensitive Clustering

w/o Context

A1, A2, A3, A4, A5, A6

B1, B2, B3, B4, B5, B6

C1, C2, C3, C4, C5

Long Qiu, Min-Yen Kan, Tat-Seng Chua

Modeling Context in Scenario Template Creation
Context-Sensitive Clustering

w/o Context

B6 B1
B5 B4 B2 B3

C1 C4 C2 C5 C3

A1 Y4
A2 A3
A4
A5
A6

TextSpan_{article ID}
Context-Sensitive Clustering

A Graph-based Approach

Evaluation

Conclusion

Long Qiu, Min-Yen Kan, Tat-Seng Chua

Modeling Context in Scenario Template Creation
Context-Sensitive Clustering

A Graph-based Approach

Evaluation

Conclusion

Long Qiu, Min-Yen Kan, Tat-Seng Chua

Modeling Context in Scenario Template Creation
Context-Sensitive Clustering

**w/o Context**

- B6
- B1
- B5
- B4
- B2
- B3

**CSC**

- C1
- C4
- C2
- C5
- C3

- B6
- B1
- B5
- B4
- B2
- B3

- C1
- C4
- C2
- C5
- C3

- A1
- Y4
- A2
- A3
- A4
- A6
- A5

- A1
- Y4
- A2
- A3
- A4
- A6
- A5

Long Qiu, Min-Yen Kan, Tat-Seng Chua

Modeling Context in Scenario Template Creation
Context-Sensitive Clustering

\[ \text{w/o Context} \]

\[ \text{CSC} \]

\( \text{(the hurricane, weakened)} \)

\( \text{Charley, landed} \)

\( \text{the storm, hit} \)

\( \text{Tokage, weakened} \)

Long Qiu, Min-Yen Kan, Tat-Seng Chua

Modeling Context in Scenario Template Creation
Context-Sensitive Clustering

w/o Context

CSC

(\text{the hurricane, weakened})

(\text{Charley, landed})

(\text{the storm, hit})

\text{TextSpan}_{\text{article ID}}

Long Qiu, Min-Yen Kan, Tat-Seng Chua

Modeling Context in Scenario Template Creation
Context-Sensitive Clustering

w/o Context

\( B_6 \quad B_1 \)
\( B_5 \quad B_4 \)
\( B_3 \quad B_2 \)

CSC

\( C_5 \quad C_3 \) (Tokage, weakened)
\( C_1 \quad C_2 \)
\( B_6 \quad B_5 \quad B_4 \quad B_2 \)

\( B_1 \)

\( \text{(the hurricane, weakened)} \)

\( (\text{Charley, landed}) \)
\( A_1 \quad Y_4 \quad A_2 \quad A_3 \) (the storm, hit)
\( A_4 \quad A_6 \quad A_5 \)

\( A_3 \) (the storm, hit)
\( A_4 \quad A_6 \quad A_5 \)

\( C_1 \quad C_4 \quad C_2 \)
\( C_5 \quad C_3 \)

\( Y_4 \quad A_2 \quad A_3 \)

Long Qiu, Min-Yen Kan, Tat-Seng Chua

Modeling Context in Scenario Template Creation
Context-Sensitive Clustering

w/o Context

CSC

(Charley, landed) A1 Y4 A2 A3 (the storm, hit) A4 A6

A1 B1 B5 B4 B3

C1 C4 C2 C5 C3 (Tokage, weakened) B3

A2 A3

A4 A6

A5

A1 A2 A3

A4 A6

A5

TextSpan\textsubscript{article ID}
Context-Sensitive Clustering

w/o Context

CSC

A Graph-based Approach

Evaluation

Conclusion

Long Qiu, Min-Yen Kan, Tat-Seng Chua

Modeling Context in Scenario Template Creation
Expectation-Maximization

- **Initialization:**
  To generate seed predicate argument tuple (PAT) clusters, based on semantic similarity;
Expectation-Maximization

- **Initialization:**
  To generate seed predicate argument tuple (PAT) clusters, based on *semantic* similarity;

- **E-Step:**
  To (re)-estimate parameters of contextual relation *distributions* between PAT clusters;
Initialization:
To generate seed predicate argument tuple (PAT) clusters, based on semantic similarity;

E-Step:
To (re)-estimate parameters of contextual relation distributions between PAT clusters;

M-Step:
To re-assign PATs to clusters, based on both semantic similarity and contextual likelihood.
CSC output

Airliner Crash

- jet, plunged
- Boeing, crashed
- plane, crashed
- killing, people
- marines, were killed
- workers, died
- killing, passengers
Evaluation

- **Data set:**
  - on-line news articles;
  - 10 articles per scenario;
  - first 15 sentences per article;
  - 150 sentences (≈ 400 predicate-argument tuples) per scenario.

- **Gold Standard:** manual PAT clusters

- **Measurement:** Purity, Inverse Purity (Hotho *et al.*, 2003)

- **Baseline:** K-Means
# Cluster Quality: (Inverse) Purity (0,1)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>#Gold Std. Clusters</th>
<th>CSC P</th>
<th>CSC IP</th>
<th>K-means P</th>
<th>K-means IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airliner Crash</td>
<td>23</td>
<td>.61</td>
<td>.42</td>
<td>.52</td>
<td>.28</td>
</tr>
<tr>
<td>Earthquake</td>
<td>18</td>
<td>.60</td>
<td>.44</td>
<td>.53</td>
<td>.30</td>
</tr>
<tr>
<td>Election</td>
<td>10</td>
<td>.77</td>
<td>.49</td>
<td>.75</td>
<td>.21</td>
</tr>
<tr>
<td>Fire</td>
<td>14</td>
<td>.65</td>
<td>.44</td>
<td>.64</td>
<td>.26</td>
</tr>
<tr>
<td>Launch Event</td>
<td>12</td>
<td>.77</td>
<td>.37</td>
<td>.73</td>
<td>.22</td>
</tr>
<tr>
<td>Layoff</td>
<td>10</td>
<td>.71</td>
<td>.28</td>
<td>.70</td>
<td>.19</td>
</tr>
<tr>
<td>Legal Case</td>
<td>8</td>
<td>.75</td>
<td>.37</td>
<td>.75</td>
<td>.18</td>
</tr>
<tr>
<td>Nobel</td>
<td>6</td>
<td>.77</td>
<td>.28</td>
<td>.77</td>
<td>.19</td>
</tr>
<tr>
<td>Obituary</td>
<td>7</td>
<td>.85</td>
<td>.46</td>
<td>.81</td>
<td>.28</td>
</tr>
<tr>
<td>Road Accident</td>
<td>20</td>
<td>.61</td>
<td>.49</td>
<td>.56</td>
<td>.40</td>
</tr>
<tr>
<td>Soccer Final</td>
<td>5</td>
<td>.88</td>
<td>.39</td>
<td>.88</td>
<td>.15</td>
</tr>
<tr>
<td>Storm</td>
<td>14</td>
<td>.61</td>
<td>.31</td>
<td>.61</td>
<td>.22</td>
</tr>
<tr>
<td>Tennis</td>
<td>6</td>
<td>.87</td>
<td>.19</td>
<td>.87</td>
<td>.12</td>
</tr>
<tr>
<td>Terrorist Attack</td>
<td>14</td>
<td>.64</td>
<td>.48</td>
<td>.62</td>
<td>.25</td>
</tr>
<tr>
<td>Volcano</td>
<td>16</td>
<td>.68</td>
<td>.38</td>
<td>.66</td>
<td>.17</td>
</tr>
<tr>
<td>Average</td>
<td>12.2</td>
<td>.72</td>
<td>.39</td>
<td>.69</td>
<td>.23</td>
</tr>
</tbody>
</table>

**Table: CSC vs. K-means**

- Higher P values: less noise (in CSC clusters)
- Higher IP values: less divided clusters (for CSC)
CSC vs. K-means: F measure

Figure: CSC vs. K-means
For STC, Semantic Role Labeler helps to normalize syntactical variations;
For STC, Semantic Role Labeler helps to normalize **syntactical variations**;

- String matching is unable to handle **lexical variations**;
  - Thesauri can help with lexical variations;
  - Context can help when thesauri is not accurate.
For STC, Semantic Role Labeler helps to normalize **syntactical variations**;

String matching is unable to handle **lexical variations**;
- Thesauri can help with lexical variations;
- Context can help when thesauri is not accurate.

A graph-based approach proposed: EM-based Context Sensitive Clustering.
- Novel encoding of context and contextual relation;
- Contextual information systematically integrated in clustering.
For STC, Semantic Role Labeler helps to normalize **syntactical variations**;

String matching is unable to handle **lexical variations**;

- Thesauri can help with lexical variations;
- Context can help when thesauri is not accurate.

A graph-based approach proposed: EM-based Context Sensitive Clustering.

- Novel encoding of context and contextual relation;
- Contextual information systematically integrated in clustering.

Evaluation shows:

- Promising recall/precision;
- Satisfactory performance in comparison to baseline system.
For STC, Semantic Role Labeler helps to normalize **syntactical variations**;
String matching is unable to handle **lexical variations**;
  - Thesauri can help with lexical variations;
  - Context can help when thesauri is not accurate.

A graph-based approach proposed: EM-based Context Sensitive Clustering.
  - Novel encoding of context and contextual relation;
  - Contextual information systematically integrated in clustering.

Evaluation shows:
  - Promising recall/precision;
  - Satisfactory performance in comparison to baseline system.

A soon-to-be publicly available STC corpus:
  - 5 articles × 10 events × 15 scenarios
For STC, Semantic Role Labeler helps to normalize **syntactical variations**;

String matching is unable to handle **lexical variations**;
- Thesauri can help with lexical variations;
- Context can help when thesauri is not accurate.

A graph-based approach proposed: EM-based Context Sensitive Clustering.
- Novel encoding of context and contextual relation;
- Contextual information systematically integrated in clustering.

Evaluation shows:
- Promising recall/precision;
- Satisfactory performance in comparison to baseline system.

A soon-to-be publicly available STC corpus:
- 5 articles $\times$ 10 events $\times$ 15 scenarios

Future work:
- Better ways to encode contextual information;
- How to encode more contextual information.
All 143 people aboard a Russian passenger plane are reported to have died.
All seven marines on a military plane that crashed were killed.