Stochastic Models for Semantic Parsing, Multi-Faceted Topic Discovery, and Causal Event Inference: Perspectives from Natural Language Processing

ROXANA GIRJU
Linguistics and Computer Science, Beckman Institute, University of Illinois

Semantic Frontiers’ Group: Rania Al-Sabbagh, Brendon Beamer, Chen Li, David Lundgren, Michael Paul (now at JHU), Mehwish Riaz

November 12, 2011
Some Research Projects @ Semantic Frontiers Group

- **P1. Stochastic Models for Semantic Parsing**
  - The problem of knowledge discovery
  - Semantic relation discovery
  - Some text-image applications

- **P2. Perspectives, aspects and sentiment**
  - in scientific literature
  - in Israeli-Palestinian editorials
  - cultural differences from travelers’ experiences

- **P3. Causal Event Inference**
  - Identification of causal relations between events
  - Applications: question answering, textual entailment
Semantic Parsing (1)

- Knowledge Discovery from Text:
  - The process of extracting useful, non-trivial (implicit) knowledge from unstructured data.

- Knowledge Discovery as Semantic Relations:
  - are underlying relations between two concepts expressed by words or phrases
  - Examples:
  - **HYPERNYMY** (IS-A),
  - **MERONYMY** (PART-WHOLE),
  - **CAUSE - EFFECT**, etc.

- Semantic parsing:
  - supports automated reasoning.
The task of semantic relation discovery:
Given a pair of nouns $n_1 - n_2$, determine the pair’s meaning.
Example 1: (Girju et al. 2003, 2006, 2007, 2009)


Semantic Parsing: Basic Approach (1)

- Defined SR list
- Other resources
- Noun-noun pair
- Semantic Parser
- Semantic relation

Diagram showing the process from predefined SR list, other resources, noun-noun pair, to semantic relation via Semantic Parser.
Examples of relations (SemEval 2007):

- **Cause-Effect**: laugh wrinkles
- **Instrument-Agency**: laser printer
- **Product-Producer**: honey bee
- **Origin-Entity**: message from outer-space
- **Theme-Tool**: news conference
- **Part-Whole**: car door
- **Content-Container**: the cookies in the jar
(SemEval 2007):

After the cashier put the `<e1>cash</e1>` in a `<e2>bag</e2>`, the robber saw a bottle of scotch that he wanted behind the counter on the shelf.

Query = “the * in a *”

WordNet(e1) = "cash%1:21:00::"

WordNet(e2) = "bag%1:06:00::"

Content-Container(e1,e2) = “true”
A Stochastic Model for Semantic Parsing

Semantic Scattering$^2$ (SS$^2$)


- The most important component of our full Semantic Parser @SemEval 2007
- Input: $n_1$ (and its sense in context);
  $n_2$ (and its sense in context);
  list of semantic relations;
  WordNet noun hierarchy
- Output: $<n_1, n_2, r>$
Datasets

- 140 training and >70 test examples for each relation;
- Balanced positive and negative examples.
- Definition provided for each relation
- WordNet senses provided for input nouns

Annotation summary:
- high inter-annotator agreement on WordNet senses and semantic relations
- Disagreements discussed and consensus reached (or example thrown out).
Hypothesis:
- Noun – noun pairs with the same/similar meaning tend to encode the same semantic relation.

Approach:
- The semantic class of a noun:
  - specifies its WordNet sense in context and
  - implicitly points to its hypernyms;
A Stochastic Model for Semantic Parsing

Sense 3

bass, basso --
(an adult male singer with the lowest voice)
=> singer, vocalist
    => musician, instrumentalist, player
    => performer, performing artist
    => entertainer
    => person, individual, someone...
    => life form, organism, being...
    => entity, something
=> causal agent, cause, causal agency
    => entity, something

Sense 7

bass --
(the member with the lowest range of a family of musical instruments)
=> musical instrument
    => instrument
    => device
    => instrumentality, instrumentation
    => artifact, artefact
    => object, physical object
    => entity, something
A Stochastic Model for Semantic Parsing

Features (for binary classifiers):

- Semantic class of head noun: $f^h_j$
- Semantic class of modifier noun: $f^m_i$

  - E.g.: hand#1 of a woman#1 [P-W]

- Feature pair: $<f^m_i, f^h_j> = f_{ij}$
- Form tuples: $<f_{ij}, r>$

\[
P(r \mid f_{ij}) = \frac{n(r, f_{ij})}{n(f_{ij})}
\]

\[
r = \arg \max P(r \mid f_{ij})
\]
The task: Find the best set of semantic classes (i.e., a boundary $G^*$ – a division in WordNet) that best generalize over the training data and accurately classify unseen Data.
A Stochastic Model for Semantic Parsing

Example for Part-Whole

- **Step 1**: (Create an Initial Boundary (i.e., generalize the training examples):
  - Initial corpus:
    - <n1#sense; n2#sense; target>
    - E.g.: <hand#1; woman#1; YES>
  
    \[<\text{hand}^1, \text{entity}^1; \text{woman}^1, \text{entity}^1; \text{YES}>\]
A Stochastic Model for Semantic Parsing

- Ambiguous examples:
  
  `<n1 hierarchy#sense; n2 hierarchy#sense; Yes/No>`

  `<apartment #1; woman #1; No>`
  `<hand #1; woman #1; Yes>`

  `<entity #1; entity #1; Yes/No>`
Step 2: Specialize ambiguous examples:

<entity#1; entity#1; Yes/No>

specialization

<whole#2; causal_agent#1; No>

<part#7; causal_agent#1; Yes>

woman#1's apartment#1; hand#1 of a woman#1;

entity#1

whole#2

part#7

causal_agent#1

apartment#1

hand#1

woman#1
A Stochastic Model for Semantic Parsing

Specialization example2:

\[<\text{entity}\#1; \text{entity}\#1; \text{Yes}/\text{No}>\]

specialization

\[<\text{part}\#7; \text{organism}\#1; \text{Yes}>\]

\[<\text{part}\#7; \text{organism}\#1; \text{No}>\]

specialization

\[...\]

\[\text{leg}\#2 \text{ of } \text{insect}\#1; \]

\[\text{insect}\#1 \text{'s world}\#7;\]

\[\text{entity}\#1\]

\[\text{part}\#7\]

\[\text{leg}\#2 \quad \text{world}\#7 \quad \text{insect}\#1\]

\[\text{organism}\#1\]
## Experimental results

<table>
<thead>
<tr>
<th>Relation</th>
<th>P</th>
<th>R</th>
<th>F</th>
<th>Acc</th>
<th>Total</th>
<th>Base-F</th>
<th>Base-Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cause-Effect</td>
<td>69.5</td>
<td><strong>100</strong></td>
<td><strong>82</strong></td>
<td><strong>77.5</strong></td>
<td>80</td>
<td>67.8</td>
<td>51.2</td>
</tr>
<tr>
<td>Instrument-Agency</td>
<td><strong>68.2</strong></td>
<td>78.9</td>
<td>73.2</td>
<td><strong>71.8</strong></td>
<td>78</td>
<td>65.5</td>
<td>51.3</td>
</tr>
<tr>
<td>Product-Producer</td>
<td>84.5</td>
<td>79</td>
<td>81.7</td>
<td><strong>76.3</strong></td>
<td>93</td>
<td>80</td>
<td>66.7</td>
</tr>
<tr>
<td>Origin-Entity</td>
<td>86.4</td>
<td>52.8</td>
<td>65.5</td>
<td><strong>75.3</strong></td>
<td>81</td>
<td>61.5</td>
<td>55.6</td>
</tr>
<tr>
<td>Theme-Tool</td>
<td>85.7</td>
<td>41.4</td>
<td><strong>55.8</strong></td>
<td><strong>73.2</strong></td>
<td>71</td>
<td>58</td>
<td>59.2</td>
</tr>
<tr>
<td>Part-Whole</td>
<td>70.8</td>
<td>65.4</td>
<td>68</td>
<td><strong>77.8</strong></td>
<td>72</td>
<td>53.1</td>
<td>63.9</td>
</tr>
<tr>
<td><strong>Content-Container</strong></td>
<td><strong>93.1</strong></td>
<td>71.1</td>
<td>80.6</td>
<td><strong>82.4</strong></td>
<td>74</td>
<td>67.9</td>
<td>51.4</td>
</tr>
<tr>
<td>Average</td>
<td>79.7</td>
<td>71.1</td>
<td>80.6</td>
<td><strong>82.4</strong></td>
<td>74</td>
<td>67.9</td>
<td>51.4</td>
</tr>
</tbody>
</table>
A Stochastic Model for Semantic Parsing

![F-measure bar chart]

- **F-measure**
- **%**
- **Series1**

- Relations:
  - Cause-Effect
  - Prod-Prod
  - Cont-Cont
  - Instr-Agency
  - Part-Whole
  - Origin-Entity
  - Theme-Tool

---

*Note: The chart shows the F-measure percentages for various semantic relations using Series1.*
Application: Text-to-Scene Generation (1)

- **Idea:** Given a snippet of text, generate an image that is a faithful representation of that text

- **Challenges:**
  - Non-visual words (e.g., abstract words: policy, government, feeling)
  - Some knowledge needed for pictures is not explicitly stated in text, but inferred

- **State of the art systems:**
  - Fairy tales
    - WordsEye (Coyle and Sproat 2001): 3D objects (positions, color, texture, etc.); WordNet; (http://www.wordseye.com/)
  - Car simulations (car insurance purposes)
    - CarSim (Dupuy at al. 2001)
    - (Girju et al. 2011)
boat on the lake vs. cabin on the lake

eagle in the nest vs. eagle in the sky

flowers in a vase
Text-to-Scene Generation (3)

Car accident visualizations
P2. Perspectives, Aspects and Sentiment

- in scientific literature
- in Israeli-Palestinian editorials
- cultural differences from travelers’ experiences
Perspectives, Aspects and Sentiment

- TAM (Topic-Aspect Model):
  - Documents can be clustered along a number of dimensions: topics, sentiment/perspective/viewpoint
  - Discovers *topics* and *aspects*
  - Generates token assignment in both dimensions
Probabilistic Topic Models (1/2)

- Each word token associated with hidden “topic” variable
- Probabilistic approach to dimensionality reduction
- Useful for uncovering latent structures in text

Basic formulation:
- \( P(w|d) = P(w|\text{topic}) \ P(\text{topic}|d) \)
“Topics” are latent distributions over words (cluster of words)

There are often other dimensions in which words could be clustered
- Sentiment/perspective/viewpoint

What if we want to model both?
Each document has
- a multinomial topic mixture
- a multinominal aspect mixture

Words may depend on both!
Topic and aspect mixtures are drawn independently of one another
- This differs from hierarchical topic models where one depends on the other
- Can be thought of as two separate clustering dimensions
Each word token also has 2 binary variables:
- the “level” (background or topical)
  - denotes if the word depends on the topic or not
- the “route” (neutral or aspectual)
  - denotes if the word depends on the aspect or not

A word may depend on a topic, an aspect, both, or neither
Topic-Aspect Model (4/8)

**Topic:** SPEECH RECOGNITION

*“Computational” Aspect*

<table>
<thead>
<tr>
<th>Route / Level</th>
<th>Background</th>
<th>Topical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>paper, new, present</td>
<td>speech, recognition</td>
</tr>
<tr>
<td>Aspectual</td>
<td>algorithm, model</td>
<td>markov, hmm, error</td>
</tr>
</tbody>
</table>

**“Linguistic” Aspect**

<table>
<thead>
<tr>
<th>Route / Level</th>
<th>Background</th>
<th>Topical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>paper, new, present</td>
<td>speech, recognition</td>
</tr>
<tr>
<td>Aspectual</td>
<td>language, linguistic</td>
<td>prosody, intonation, tone</td>
</tr>
</tbody>
</table>

- A word may depend on a **topic**, an **aspect**, **both**, or **neither**
### Topic-Aspect Model (5/8)

**Topic:** COMMUNICATION

“Computational” Aspect

<table>
<thead>
<tr>
<th>Route / Level</th>
<th>Background</th>
<th>Topical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>paper, new, present</td>
<td>communication, interaction</td>
</tr>
<tr>
<td>Aspectual</td>
<td>algorithm, model</td>
<td>dialogue, system, user</td>
</tr>
</tbody>
</table>

**“Linguistic” Aspect**

<table>
<thead>
<tr>
<th>Route / Level</th>
<th>Background</th>
<th>Topical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>paper, new, present</td>
<td>speech, recognition</td>
</tr>
<tr>
<td>Aspectual</td>
<td>language, linguistic</td>
<td>conversation, social</td>
</tr>
</tbody>
</table>

- A word may depend on a **topic**, an **aspect**, **both**, or **neither**
Each token $i$ is associated with: $w_i, z_i, y_i, l_i, x_i$

Generative process for a document $d$:
- Sample a topic $z$ from $P(z|d)$
- Sample an aspect $y$ from $P(y|d)$
- Sample a level $l$ from $P(l|d)$
- Sample a route $x$ from $P(x|l,z)$

Sample a word $w$ from either:
- $P(w|l=0,x=0)$,
- $P(w|z,l=1,x=0)$,
- $P(w|y,l=0,x=1)$,
- $P(w|z,y,l=1,x=1)$
• Distributions have Dirichlet/Beta priors
  ○ Latent Dirichlet Allocation framework

• Number of aspects and topics are user-supplied parameters

• Straightforward inference with Gibbs sampling
Semi-supervised TAM when aspect label is known

Two options:
- Fix $P(y|d) = 1$ for the correct aspect label and 0 otherwise
  - Behaves like ccLDA (Paul and Girju, 2009)
- Define a prior for $P(y|d)$ to bias it toward the true label
Experiments (1/3)

- Datasets and settings:
  - 4,247 abstracts from the ACL Anthology (CL-Only)
    - $Z = 25; Y = 2$;
  - 594 articles from the Bitterlemons corpus (Lin et al., 2006)
    - a collection of editorials on the Israeli/Palestinian conflict (I-P dataset)
    - Both unsupervised and semi-supervised setting ($Z = 12$)
Experiments (2/3)

- Example: **Computational Linguistics (CL-Only)**

<table>
<thead>
<tr>
<th>Background</th>
<th>Topical</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Neutral</strong></td>
<td></td>
<td>Aspect B</td>
</tr>
<tr>
<td>paper</td>
<td>topic</td>
<td>natural</td>
</tr>
<tr>
<td>based</td>
<td>method</td>
<td>language</td>
</tr>
<tr>
<td>approach</td>
<td>corpus</td>
<td>processing</td>
</tr>
<tr>
<td>information</td>
<td>using</td>
<td>structure</td>
</tr>
<tr>
<td>present</td>
<td>data</td>
<td>representation</td>
</tr>
<tr>
<td>language</td>
<td>task</td>
<td>semantic</td>
</tr>
<tr>
<td>new</td>
<td>words</td>
<td>relations</td>
</tr>
<tr>
<td>using</td>
<td>classification</td>
<td>lexical</td>
</tr>
<tr>
<td>model</td>
<td>clustering</td>
<td>relation</td>
</tr>
<tr>
<td>analysis</td>
<td>words</td>
<td>concepts</td>
</tr>
<tr>
<td>different</td>
<td>classification</td>
<td>hierarchy</td>
</tr>
<tr>
<td>problem</td>
<td>distributional</td>
<td>objects</td>
</tr>
<tr>
<td>set</td>
<td>occurrence</td>
<td></td>
</tr>
<tr>
<td>describes</td>
<td>learning</td>
<td></td>
</tr>
<tr>
<td>context</td>
<td>text</td>
<td></td>
</tr>
<tr>
<td>work</td>
<td>evaluation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>methods</td>
<td></td>
</tr>
<tr>
<td></td>
<td>automatic</td>
<td></td>
</tr>
<tr>
<td></td>
<td>features</td>
<td></td>
</tr>
<tr>
<td></td>
<td>experiments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>accuracy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>algorithm</td>
<td></td>
</tr>
</tbody>
</table>

| Topical   | Neutral | | Aspect B |
|-----------|---------| |         |
| **TOPIC 1** | | | |
| similarity | topic   | | ontology|
| patterns  | method  | | conceptual|
| clustering| corpus  | | verbs    |
| words     | using   | | verb     |
| classification | data | | concepts |
| distributional | task | | hierarchy|
| occurrence | learning | | objects  |

| **TOPIC 2** | | | |
| segmentation | topic   | | temporal |
| text        | method  | | expressions |
| segment     | automatic| | tense      |
| segments    | features | | theory     |
| local       | experiments | | aspect     |
| coherence   | accuracy  | | referring  |
| cohesion    | algorithm | | spatial    |
### Experiments (3/3)

**Example:** Israeli/Palestinian Conflict (I-P)

Unsupervised:

<table>
<thead>
<tr>
<th>Aspect A</th>
<th>Aspect B</th>
</tr>
</thead>
<tbody>
<tr>
<td>palestinian</td>
<td>war</td>
</tr>
<tr>
<td>israeli</td>
<td>public</td>
</tr>
<tr>
<td>military</td>
<td>government</td>
</tr>
<tr>
<td>civilians</td>
<td>media</td>
</tr>
<tr>
<td>attacks</td>
<td>society</td>
</tr>
<tr>
<td></td>
<td>terrorist</td>
</tr>
<tr>
<td></td>
<td>soldiers</td>
</tr>
<tr>
<td></td>
<td>incitement</td>
</tr>
<tr>
<td></td>
<td>violence</td>
</tr>
<tr>
<td></td>
<td>palestinians</td>
</tr>
<tr>
<td></td>
<td>occupation</td>
</tr>
<tr>
<td></td>
<td>resistance</td>
</tr>
<tr>
<td></td>
<td>intifada</td>
</tr>
<tr>
<td></td>
<td>violent</td>
</tr>
<tr>
<td></td>
<td>non</td>
</tr>
<tr>
<td></td>
<td>force</td>
</tr>
</tbody>
</table>

Prior for $P(\text{aspect}|d)$ for true label:

<table>
<thead>
<tr>
<th>Israeli</th>
<th>Palestinian</th>
</tr>
</thead>
<tbody>
<tr>
<td>jewish</td>
<td>palestinians</td>
</tr>
<tr>
<td>arab</td>
<td>return</td>
</tr>
<tr>
<td>israeli</td>
<td>right</td>
</tr>
<tr>
<td>jews</td>
<td>refugees</td>
</tr>
<tr>
<td>population</td>
<td>problem</td>
</tr>
<tr>
<td>jordan</td>
<td>refugee</td>
</tr>
<tr>
<td>west</td>
<td>rights</td>
</tr>
<tr>
<td>south</td>
<td>resolution</td>
</tr>
</tbody>
</table>
Evaluation

- Cluster coherence
  - 5 human annotators
  - Compare against ccLDA and LDA (Z=25)
    - TAM clusters are as coherent as other established models
  - “word intrusion” method (Chang et al., 2009)
Cultural differences in tourists’ forums: Topic of ‘weather’ (tourist perspective) (lonelyplanet.com)

<table>
<thead>
<tr>
<th>UK</th>
<th>India</th>
<th>Singapore</th>
</tr>
</thead>
<tbody>
<tr>
<td>wind</td>
<td>leh</td>
<td>hot</td>
</tr>
<tr>
<td>waterproof</td>
<td>monsoon</td>
<td>humid</td>
</tr>
<tr>
<td>ending</td>
<td>road</td>
<td>heat</td>
</tr>
<tr>
<td>rolling</td>
<td>manali</td>
<td>degree</td>
</tr>
<tr>
<td>walkers</td>
<td>ladkh</td>
<td>equator</td>
</tr>
<tr>
<td>rochdale</td>
<td>trekking</td>
<td>sweat</td>
</tr>
<tr>
<td>layers</td>
<td>trek</td>
<td>bring</td>
</tr>
<tr>
<td>snow</td>
<td>season</td>
<td>rain</td>
</tr>
<tr>
<td>footwear</td>
<td>rains</td>
<td>umbrella</td>
</tr>
<tr>
<td>ankle</td>
<td>monsoons</td>
<td></td>
</tr>
</tbody>
</table>
Other Applications: Social Media

- Cultural differences in tourists’ forums: Topic of ‘food’ (blogcatalog.com)

<table>
<thead>
<tr>
<th>Perspective of Locals</th>
<th>Perspective of Tourists</th>
</tr>
</thead>
<tbody>
<tr>
<td>food add chicken recipe cooking take</td>
<td>food eat restaurant tea cheap meal eating</td>
</tr>
<tr>
<td>rice recipes sugar soup</td>
<td>café drink</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>UK</th>
<th>India</th>
<th>Sing.</th>
<th>UK</th>
<th>India</th>
<th>Sing.</th>
</tr>
</thead>
<tbody>
<tr>
<td>food</td>
<td>recipe</td>
<td>coffee</td>
<td>chips</td>
<td>cooking</td>
<td>hawker</td>
</tr>
<tr>
<td>wine</td>
<td>recipes</td>
<td>cup</td>
<td>haggis</td>
<td>spices</td>
<td>satay</td>
</tr>
<tr>
<td>restaurant</td>
<td>powder</td>
<td>oil</td>
<td>fish</td>
<td>sick</td>
<td>stalls</td>
</tr>
<tr>
<td>coffee</td>
<td>indian</td>
<td>comments</td>
<td>respectability</td>
<td>flour</td>
<td>noodles</td>
</tr>
<tr>
<td>cheese</td>
<td>salt</td>
<td>fried</td>
<td>decent</td>
<td>tomato</td>
<td>roti</td>
</tr>
<tr>
<td>soup</td>
<td>tsp</td>
<td>add</td>
<td>veggie</td>
<td>batter</td>
<td>stall</td>
</tr>
<tr>
<td>eat</td>
<td>rice</td>
<td>rice</td>
<td>pudding</td>
<td>ate</td>
<td>seafood</td>
</tr>
<tr>
<td>english</td>
<td>masala</td>
<td>tea</td>
<td>photoblog</td>
<td>cook</td>
<td>malay</td>
</tr>
<tr>
<td>drink</td>
<td>oil</td>
<td>seafood</td>
<td>sausages</td>
<td>olive</td>
<td>sochester</td>
</tr>
</tbody>
</table>
P3. Causal Event Inference

- Identification of causal relations between events
- Applications: question answering, textual entailment
Causal Knowledge

- a pervasive feature of human language and theorising about the world

- important in text comprehension, entailment, automatic question answering and information retrieval (Goldman et al., 1999; Khoo et al., 2001; Girju, 2003)

.. a prerequisite to perform textual reasoning

Despite this, the specification of a satisfactory general analysis of causal relations has long proved difficult.
Qs:
How is causality encoded in language?
How is it perceived and used?

Models:
- unsupervised or semi-supervised
- knowledge poor

- Baselines
- New insights
Hypothesis:

Scenario-specific events, contributing towards same objective in a domain, are likely to be dependent on each other and thus make good candidates for causal relationships.

Focus:

- Event sequences (Events: <[Subject] {verb} [Object]>)
- Scenario-specific events in news articles
  - Hurricane Katrina (447 articles, 189,840 word-tokens, 14,996 word-types)
  - Iraq war (556 articles, 304,481 word-tokens, 20,629 word-types)
### Examples

<table>
<thead>
<tr>
<th>Data set</th>
<th>Hurricane Katrina</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario</td>
<td>Hurricane Katrina disaster and damage.</td>
</tr>
<tr>
<td>Example</td>
<td>Katrina {hit} Florida late last week. Since Friday, Dallas-based Southwest airlines {canceled} more than 250 flights.</td>
</tr>
<tr>
<td>Causal Rel.</td>
<td>“Katrina {hit} Florida” (\rightarrow) “Dallas-based Southwest airlines {canceled} more than 250 flights”</td>
</tr>
<tr>
<td>Type</td>
<td>Inter-sentential causal relationship.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data set</th>
<th>Iraq War</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario</td>
<td>US accusations and the UN inspections.</td>
</tr>
<tr>
<td>Example</td>
<td>Bush {criticized} UN for {being ineffective}</td>
</tr>
<tr>
<td>Causal Rel.</td>
<td>“UN {being ineffective}” (\rightarrow) “Bush {criticized} UN”</td>
</tr>
<tr>
<td>Type</td>
<td>Intra-sentential causal relationship.</td>
</tr>
</tbody>
</table>
Notion of Causality (1)

• Selection criteria:
  - Need for a consistent set of annotation guidelines to capture our perception of causality (as expressed by language)
  - Be based on causal theories
    - provide the annotator with a relatively objective test (without relying on intuitions which will vary significantly from annotator to annotator)
  - The test should also be easy to perform mentally (w/o detailed philosophical knowledge about causality)
    - Subjectivity would certainly arise in cases where the annotator is unaware of how certain things in the world work;
    - But not a problem (many people share more/less the same baggage of commonsense knowledge)
Notion of causality (2)

- annotation test for causality:
  (Answering yes to both would mean the two events are causally related):

  (i) Does event A occur before event B?

  (ii) Keeping constant as many other states of affairs of the world in the given text context as possible, does modifying event A entail predictably modifying event B?

- Causal relations in a broad sense: contingency discourse relations:
  - purpose, reason, explanation, argument-claim (Sanders et al, 1992; Mann & Thompson, 1988)
Unsupervised model:

1. Identify Topic-Specific Scenarios and their Events
   - Discover topic-specific scenarios
   - Identify scenario-specific events

2. Generate Appropriate Event-Pair Candidates
   - Group events
   - Identify frequent event pairs

3. Learn Causal Relationships
   - Identify causal dependency
   - Assign cause and effect roles
1. Identifying Topic-Specific Scenarios and their Events

- cluster the input sentences according to their probability distributions into topic-specific scenarios.
- Use topic models to cluster semantically related text units

Pachinko Allocation Model (PAM) (Lei & McCallum, 2006)
- Assume documents are represented by Directed Acyclic Graph (DAG)
- Cluster text units
- Also find relationships between clusters
(C1) “War effects-economic progress in Iraq and side effects on the world’s economy”,
(C2) “US accusations and the UN inspection,
(C3) “Pre-war: War strategies and planning”.

The three topic-specific scenarios for the Iraq War collection.
Recovering Sentences and their events from Discovered Scenarios

- **Representation:**
  - Each scenario cluster \( v \) is a vector of words and each word’s weight is its probability of assignment to scenario \( v \) cf. PAM.
  - Each sentence \( s \) is a vector of words. Each word’s weight is its probability of occurrence in a sentence.
  - Assign sentence \( s \) to scenario cluster \( v \) with which it has highest cosine similarity measure (\( N \) is vocabulary size)

\[
\text{Cosine - Sim}(\vec{s}, \vec{v}) = \frac{\vec{s} \cdot \vec{v}}{\sqrt{\sum_{i=1}^{N} s_i^2} \sqrt{\sum_{i=1}^{N} v_i^2}}.
\]
Extracting Events from Recovered Sentences

- use Semantic Role Labeler (Surdeanu et al, 2005)
  - A0 – agent and A1 – theme

Sentences along with their events (shown in italic) assigned to the scenarios identified for the Iraq war collection.

Scenario 1 – “War effects - economic progress in Iraq and side effects on the world’s economy”

**Event in context:** <Financial markets {wobble}> as Iraq war unfolds.

Scenario 2 -- “US accusations and the UN inspections”

**Event in context:** <Pentagon {fears} last-ditch Iraqi chemical attack>.

Scenario 3 -- “Pre-war: War strategies and planning”

**Event in context:** If the <Kurds join the Shiites> in a general offensive against the Sunnis, <the Sunnis will probably lose>.
2. Generating Appropriate Event Pair Candidates

- **Grouping Events**
  - \{“UN council suspects Iraq”, “UN security council suspects Iraq”\}
  - basic clustering approach
    - distance measure dependent on lexical similarity of two events with same verb lemma.

- **Identifying Frequent Event Pairs**
  - FP-Growth (Han et al, 2004) algorithm to collect frequent event pairs (a, b) that appear in at least \( n \) news articles.
FP-Growth

\[ D_i = \{e_1, e_2, \ldots, e_n\} \]

**FP-Growth:**
- output: \((G_i, G_j)\)
- min. support of 3

E.g.: \((G_1\text{.suspect}, G_2\text{.kill})\)
\((G_1\text{.suspect}, G_2\text{.fall})\)
(other pairs were rejected; E.g., \((G_1\text{.suspect}, G_4\text{.go})\)
3. Learning Causal Relationships

- (1) determine if the events of a frequent event pair \((a,b)\) encode a contingency relation
- (2) identify the Cause and the Effect roles

(1) Causal dependency:
- Rank candidate \((a, b)\) based on how strongly dependent the events are.
- Condition: Cause can appear independently with other events, while Effect is expected to have a high likelihood of occurrence in the presence of the causing event (similar to Suppes, 1970)
- The causal events can appear anywhere in or across documents \(\rightarrow\) direct or indirect relationships.

Event-Control Dependency: ECD \((a,b)\):

\[
\max \left( \frac{P(a,b)}{P(b)-P(a,b)+\gamma} \times \frac{P(a,b)}{\max_i P(a,c_i)-P(a,b)+\gamma}, \frac{P(a,b)}{P(a)-P(a,b)+\gamma} \times \frac{P(a,b)}{\max_i P(c_i,b)-P(a,b)+\gamma} \right)
\]

2) Cause and Effect roles
Example

Pair-1 = ("US accused Iraq of developing chemical weapons" [92],
"UN inspected Iraqi scientists" [51]) [50]

Pair-2 = ("UN Security Council held an emergency session" [55],
"Security Council closed emergency session" [6]) [2]
Experiments and Evaluation

- Evaluating Scenarios
  - Three scenarios for each domain
  - Relatedness:
    - Annotation of top-50 words in each scenario (i.e. “YES” if a word is semantically similar to other words in top-50 list, otherwise “NO”)

<table>
<thead>
<tr>
<th>Test</th>
<th>Data</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relatedness</td>
<td>Katrina</td>
<td>66%</td>
<td>57%</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td>Iraq</td>
<td>90%</td>
<td>83%</td>
<td>39.5%</td>
</tr>
<tr>
<td>Annotator-Agreement</td>
<td>Katrina</td>
<td>86%</td>
<td>94%</td>
<td>92%</td>
</tr>
<tr>
<td></td>
<td>Iraq</td>
<td>80%</td>
<td>96%</td>
<td>86%</td>
</tr>
</tbody>
</table>

Evaluation of word relatedness and inter-annotator agreement for all three scenario clusters
Evaluating the Ranked Causal Relationships

Human Annotations

- Annotation guidelines: Manipulation Theory of Causality Test (Beamer & Girju, 2009)
- Observations:
  - the Iraq war instances are much more difficult to annotate.
  - cases where the annotators had to have some advanced domain knowledge in order to annotate the examples.

Human perception task:
- Does distance between events reduce strength of relationship between events?
- HK(C1+C3) and IW(C1+C2): 80 event pairs each at distances 1 to 4 (20 examples)
- The smaller the distance between two events, the more likely it is to be perceived as causal
Causal Dependency Evaluation through interpolated precision-recall curve
- Distance 1 and 2 only
- top 100 ranked causal pair (a,b) examples from top two scenarios of both domains
Experiments and Evaluation (2)

- Assignment of Cause and Effect Roles
  - Distance 1 and 2 only

<table>
<thead>
<tr>
<th>Task</th>
<th>HK:ECD</th>
<th>HK:PMI</th>
<th>IW:ECD</th>
<th>IW:PMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>63.4%</td>
<td>71.0%</td>
<td>72.5%</td>
<td>66.6%</td>
</tr>
<tr>
<td>CA-agreement</td>
<td>98%</td>
<td>93%</td>
<td>90%</td>
<td>85%</td>
</tr>
<tr>
<td>RA-agreement</td>
<td>100%</td>
<td>97%</td>
<td>95%</td>
<td>94%</td>
</tr>
</tbody>
</table>

Table 5: Roles Accuracy (RA) for all test sets. Roles accuracy = # of correct roles predicted/# of causal examples. CA-agreement and RA-agreement show the inter-annotator agreement on the causality annotation and causality roles annotation tasks.
Conclusions

- Unsupervised, knowledge-poor model
  - Baseline

- Improvements:
  - Better similarity measures
  - Better clustering models
  - Better analysis of the extracted causal relations:
    - Role of context vs. statistical tendencies
    - Types of knowledge
Model people interactions:

- **Direct observations** (explicitly stated)
- **Inferences** of what they feel/think
- **Predictions** of impending actions
- **Decisions/suggestions** about subsequent interactions
When perceiving, explaining, judging human behavior, people differentiate between intentional / unintentional actions

- This can set the course of social interactions

Example:

**Behavior perceived as intentional:**
- Critical remark → insult
- Collision in the driveway → provocation
- Charming smile → hint of seduction

**Behavior perceived as unintentional:**
- Critical remark → constructive feedback
- Collision in the driveway → new friendship
- Charming smile → good mood

*Negatively evaluated behavior*  
*Positively evaluated behavior*
Thank You