Introduction

Lexical Chains

Entity-Based Coherence

Lecture 11: Lexical Chains and Entity-based Coherence

Lexical Semantics and Discourse Processing

MPhil in Advanced Computer Science

Simone Teufel

Natural Language and Information Processing (NLIP) Group

UNIVERSITY OF CAMBRIDGE

Simone.Teufel@cl.cam.ac.uk

Slides after Frank Keller, Regina Barzilay and Mirella Lapata

February 25, 2011

Coherence in Text

Coherence:

- is a property of well-written texts;
- makes them easier to read and understand;
- ensures that sentences are meaningfully related;
- and that the reader can work out what expressions mean;
- the text is thematically organized;
- temporally organized;
- rather than a random concatenation of sentences.

Two solutions discussed in this lecture:

- Lexical Chains (Morris and Hearst, 1991)
- Entity-based Coherence (Barzilay and Lapata, 2005)

Reading:


Summary A

Britain said he did not have diplomatic immunity. The Spanish authorities contend that Pinochet may have committed crimes against Spanish citizens in Chile. Baltasar Garzon filed a request on Wednesday. Chile said, President Fidel Castro said Sunday he disagreed with the arrest in London.

Summary B

Former Chilean dictator Augusto Pinochet, was arrested in London on 14 October 1998. Pinochet, 82, was recovering from surgery. The arrest was in response to an extradition warrant served by a Spanish judge. Pinochet was charged with murdering thousands, including many Spaniards. Pinochet is awaiting a hearing, his fate in the balance. American scholars applauded the arrest.
Summary A

Britain said he did not have diplomatic immunity. The Spanish authorities contend that Pinochet may have committed crimes against Spanish citizens in Chile. Baltasar Garzon filed a request on Wednesday. Chile said, President Fidel Castro said Sunday he disagreed with the arrest in London.

Summary B

Former Chilean dictator Augusto Pinochet, was arrested in London on 14 October 1998. Pinochet, 82, was recovering from surgery. The arrest was in response to an extradition warrant served by a Spanish judge. Pinochet was charged with murdering thousands, including many Spaniards. Pinochet is awaiting a hearing, his fate in the balance. American scholars applauded the arrest.

Lexical Chain: Definition

- Sequence of related words in text, spanning short (adjacent sentences) or longer distances (entire text)
- Originally due to Halliday and Hasan (1976) *Cohesion in English*
- First CL definition in Morris and Hearst (1991), *Computational Linguistics*
- Independent of the grammatical structure of the text
- Captures (some of) the cohesive structure of the text
- Can provide context for the resolution of ambiguous terms
- Wordnet can be used in the identification of lexical chains.

Lexical Chains: Applications

- Summarisation (Barzilay, Elhadad 97)
- IR (Stairmand, 96)
- Detection of malapropisms (St Onge 98)
- Topic Segmentation (Hearst 97, Kazman 96)
- Hypertext link generation (Green 97, 99)
- WSD (Okumura and Honda 94)
- Word Prediction (Fazli and Hirst 2002)
Ambiguity Problem of Lexical Chains

- Allowed lexical relations: identity, synonymy, hypo/hypernymy, siblings
- Each noun occurrence must belong to only one lexical chain (in one particular sense)
- If more than one sense fits, then the overall grouping must be optimal (i.e., result in the strongest chains)

Linear-time algorithm for lexical chain computation

- Due to Silber and McCoy (2002), *Computational Linguistics*
- Barzilay and Elhadad algorithm (2007) is exponential, which makes processing chains for long documents impossible
- Idea: create “meta-chains”
- Number of meta-chains: all noun senses in WN + all nouns in document (as they might not be in WN)
- Meta-chains represent all possible chains that can contain that particular sense of a noun encountered in the text

Silber and McCoy’s algorithm

1. Build meta-chains (first pass)
   - For each noun, for each sense, add sense to every meta-chain for which there is a lexical relation with that sense
2. Find the right interpretation (second pass)
   - For each noun, determine the meta-chain that the noun contributes to the most (based on type of relation and distance factors)
   - Distance factors: 1 sentence, 3 sentences, paragraph, default
   - Delete the word from all other meta-chains; update weights
3. Rank chains with relation to each other (third pass)
   - A strong chain is one which is more than 2 standard deviations above the mean of the chain scores in the document

Silber and McCoy: example

*John has a computer. The machine is an IBM.*

Meta-chains after first pass:

<table>
<thead>
<tr>
<th>Index</th>
<th>Synset</th>
<th>Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>person</td>
<td>John – machine/3(sib)</td>
</tr>
<tr>
<td>1</td>
<td>computer/1, data processor...</td>
<td>computer – machine(h) – IBM</td>
</tr>
<tr>
<td>2</td>
<td>computer/2, calculator/1, reckoner/1</td>
<td>computer – machine(h) – IBM</td>
</tr>
<tr>
<td>3</td>
<td>machine/1 (device)</td>
<td>machine – IBM</td>
</tr>
<tr>
<td>4</td>
<td>machine/2 (organisation; war machine)</td>
<td>machine – IBM</td>
</tr>
<tr>
<td>5</td>
<td>machine/4 (car)</td>
<td>machine – IBM</td>
</tr>
<tr>
<td>4</td>
<td>IBM</td>
<td>IBM</td>
</tr>
</tbody>
</table>

Evaluation: compare concepts in strong lexical chains of the full text with concepts in human-written summaries.

Evaluation Result: 79% of Strong chains have a corresponding noun in the summary; 80% of noun instances in the summary have a corresponding strong chain in the document.
Coherence as a model of sequences of entity types in text
Assume we know whether two linguistic expression co-refer, i.e., talk about the same entity.
Observations from discourse theory:
- The way entities are introduced and discussed influences coherence (Grosz et al 1995).
- Salience of entities is related to where in the sentence they occur (Sidner, 1992).
- Frequency, syntactic position, pronominalisation are relevant coherence properties.

1 Former Chilean dictator Augusto Pinochet, was arrested in London on 14 October 1998.
2 Pinochet, 82, was recovering from surgery.
3 The arrest was in response to an extradition warrant served by a Spanish judge.
4 Pinochet was charged with murdering thousands, including many Spaniards.
5 He is awaiting a hearing, his fate in the balance.
6 American scholars applauded the arrest.
Lexical Chains

Entity-Based Coherence

Discourse Representation

Entity Transitions

Ranking Model

The Entity Grid

<table>
<thead>
<tr>
<th>Pinocchet</th>
<th>London</th>
<th>October</th>
<th>Surgery</th>
<th>Arrest</th>
<th>Extradition</th>
<th>Warrant</th>
<th>Judge</th>
<th>Thousands</th>
<th>Spaniards</th>
<th>Hearing</th>
<th>Fate</th>
<th>Balance</th>
<th>Scholars</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S</td>
<td>X</td>
<td>X</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>S</td>
<td>–</td>
<td>–</td>
<td>X</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>S</td>
<td>X</td>
<td>X</td>
<td>O</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>O</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>O</td>
<td>O</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>S</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>O</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>S</td>
</tr>
</tbody>
</table>

Columns: entities; lines: sentences

Entity Transitions

Definition

A local entity transition is a sequence \(\{S, O, X, –\}\)\(^n\) that represents entity occurrences and their syntactic roles in \(n\) adjacent sentences.

Feature Vector Notation

Each grid \(x_{ij}\) for document \(d_i\) is represented by a feature vector:

\[
\Phi(x_{ij}) = (p_1(x_{ij}), p_2(x_{ij}), \ldots, p_m(x_{ij}))
\]

\(m\): number of entity transitions (predefined)

\(p_t(x_{ij})\): probability of transition \(t\) in grid \(x_{ij}\)

Linguistic Dimensions

Salience: Are some entities more important than others?
- Discriminate between salient (frequent) entities and the rest.
- Collect statistics separately for each group.

Coreference: Talking about the same entity
- Entities are coreferent if they have (roughly) the same surface form.
- Coreference resolution systems exist (cf. next lecture)

Syntax: Does syntactic knowledge matter?
- Use four categories \(\{S, O, X, –\}\).
- Or just two \(\{X, –\}\).
Learning a Ranking Function

Training Set
Ordered pairs \((x_{ij}, x_{ik})\), where \(x_{ij}\) and \(x_{ik}\) represent the same document \(d_i\), and \(x_{ij}\) is more coherent than \(x_{ik}\) (assume \(j > k\)).

- Source document and permutations of its sentences.
- Original order assumed coherent.
- Given \(k\) documents, with \(n\) permutations, obtain \(k \cdot n\) pairwise rankings for training and testing.
- Two corpora, Earthquakes and Accidents, 100 texts each.

Goal
Find a parameter vector \(\vec{w}\) such that:

\[
\vec{w} \cdot (\Phi(x_{ij}) - \Phi(x_{ik})) > 0 \quad \forall j, i, k \text{ such that } j > k
\]

\(\vec{w}\Phi(x_{ij})\) is a ranking score, such that the violations of pairwise rankings in the training set are minimised.

Support Vector Machines
Constraint optimization problem can be solved using the search technique described in (Joachims 2002).

Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Earthquakes</th>
<th>Accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coreference+Syntax+Salience+</td>
<td>87.2</td>
<td>90.4</td>
</tr>
<tr>
<td>Coreference+Syntax+Salience-</td>
<td>88.3</td>
<td>90.1</td>
</tr>
<tr>
<td>Coreference+Syntax-Salience+</td>
<td>86.6</td>
<td>88.4**</td>
</tr>
<tr>
<td>Coreference+Syntax+Salience-</td>
<td>83.0**</td>
<td>89.9</td>
</tr>
<tr>
<td>Coreference+Syntax-Salience+</td>
<td>86.1</td>
<td>89.2</td>
</tr>
<tr>
<td>Coreference+Syntax+Salience-</td>
<td>82.3**</td>
<td>88.6*</td>
</tr>
<tr>
<td>Coreference+Syntax-Salience+</td>
<td>83.0**</td>
<td>86.5**</td>
</tr>
<tr>
<td>Coreference+Syntax-Salience-</td>
<td>81.4**</td>
<td>86.0**</td>
</tr>
</tbody>
</table>

Evaluation metric: % correct ranks in test set.
**: sig. different from Coreference+Syntax+Salience+
- Omission of coreference causes performance drop.
- Syntax and Salience have more effect on Accidents corpus.
- Linguistically poor model generally worse.

**Summary**

**Strengths:**
- Novel framework for representing and measuring coherence.
- Entity grid and cross-sentential transitions.
- Suited for learning appropriate ranking function.
- Fully automatic and robust, useful for system development.

**Weaknesses:**
- Entity grid doesn’t contain lexical information.
- Doesn’t contain a notion of global coherence.
- Can’t model multi-paragraph text.