Modeling Local Coherence: An Entity-Based Approach

Regina Barzilay, Mirella Lapata

Presenter: Richard Littauer
Abstract

This article proposes a novel framework for representing and measuring local coherence. Central to this approach is the entity-grid representation of discourse, which captures patterns of entity distribution in a text. The algorithm introduced in the article automatically abstracts a text into a set of entity transition sequences and records distributional, syntactic, and referential information about discourse entities. We re-conceptualize coherence assessment as a learning task and show that our entity-based representation is well-suited for ranking-based generation and text classification tasks. Using the proposed representation, we achieve good performance on text ordering, summary coherence evaluation, and readability assessment.
Overview

• Introduction
• Previous work
• The Model
• Experiment 1, 2, and 3
• Wrapping it up
Introduction

Local Coherence: text relatedness at the level of sentence-to-sentence transitions.

- Important for global coherence
- Important for automatically produced text
- Difficult to do automatically
Introduction

“The distribution of entities in locally coherent texts exhibits certain regularities.”

The algorithm converts text into a set of entity transition sequences.

Incorporates information that is:

• Distributional
• Syntactic
• Referential
Introduction

Entirely automatic; without manual annotation or a predefined knowledge base.

Tested in three ways:

– Text ordering
– Automatic evaluation of summary coherence
– Readability assessment
Related Work

Entity-Based Approaches to Local Coherence:

– Linguistic Modeling (cf. Centering Theory)
– Computational Modeling

Ranking approaches to NLG

- Manual specifications
Entity-Grids

1 [The Justice Department]s is conducting an [anti-trust trial]o against [Microsoft Corp.]x with [evidence]x that [the company]s is increasingly attempting to crush [competitors]o.
2 [Microsoft]o is accused of trying to forcefully buy into [markets]x where [its own products]s are not competitive enough to unseat [established brands]o.
4 [Microsoft]s claims [its tactics]s are commonplace and good economically.
5 [The government]s may file [a civil suit]o ruling that [conspiracy]s to curb [competition]o through [collusion]x is [a violation of the Sherman Act]o.
6 [Microsoft]s continues to show [increased earnings]o despite [the trial]x.
Entity-Grids

“A local entity transition is a sequence \{S, O, X, -\}n that represents entity occurrences and their syntactic roles in n adjacent sentences.”

Each transition has a probability: frequency of occurrence over total number of transitions of that length.
Example:
Probability of [O-]: $\frac{7}{75} = 0.093$
Exercise!

2 + 7 = __

I cannot answer this question, as it is against my religious principles.

It's worth a shot.
Exercise 1 Answers

<table>
<thead>
<tr>
<th></th>
<th>Curdie</th>
<th>Peter</th>
<th>Father</th>
<th>Mother</th>
<th>Cottage</th>
<th>Mountain</th>
<th>Thing</th>
<th>Me</th>
<th>Thing</th>
<th>terrors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S</td>
<td>X</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>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>S</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>S</td>
<td>-</td>
<td>-</td>
<td>O</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>S</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

What is the probability of [X-]? 0.1 (4/(10*4))
What is the probability of [-xx]? 0.06 (2/(10*3))
Grid Construction

• What linguistic knowledge is useful?
• Considerations of parameters:
  – How important a parameter is
  – Accuracy of automatic computation
  – Size of the resulting feature space
Grid Computation

• Entity Extraction – Coreference +/-
  – Use Ng & Cardie (2002) coreference resolution system

• Grammatical Function – Syntax +/-

• Salience – Salient +/-
  – Frequency
Learning

• Given the feature vector representation, coherence assessment is an ML problem.

• Ranking alternative renderings works as a feature for learning
Experiments

• Exp. 1: Information ordering

• Exp. 2: Summary coherence rating

• Exp. 3: Readability assessment
Exp. 1: Sentence Ordering

- Task: Generate random permutations of a test document and measure how often a permutation is ranked higher than the original.
- Data: 100 source articles, up to 20 random permutations for training, same for testing.
- Original document held to be coherent; corpus includes pairwise rankings that comprise the original document and one permutations.
Exp. 1: 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>
<tr>
<td>HMM-based Content Models</td>
<td>88.0</td>
<td>75.8**</td>
</tr>
<tr>
<td>Latent Semantic Analysis</td>
<td>81.0**</td>
<td>87.3**</td>
</tr>
</tbody>
</table>
Exp. 1: Learning Curve

![Graph showing the learning curve for accuracy as a function of the number of pairwise rankings in training data, with two different datasets: Earthquakes and Accidents.](Image)
Exp. 1: Learning Domains

<table>
<thead>
<tr>
<th></th>
<th>Coreference+Syntax+Salience</th>
<th></th>
<th>HMM-Based Content Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Test</td>
<td>Earthquakes</td>
</tr>
<tr>
<td>Earthquakes</td>
<td></td>
<td>87.3</td>
<td>67.0**</td>
</tr>
<tr>
<td>Accidents</td>
<td></td>
<td>69.7**</td>
<td>90.4</td>
</tr>
<tr>
<td>EarthAccid</td>
<td></td>
<td>86.7</td>
<td>88.5*</td>
</tr>
</tbody>
</table>

** indicates statistically significant results.
Exp. 2: Summary Coherence Rating

• Test: compare model-induced rankings against rankings elicited by human judges
• Data: Training data was 144 student-annotated summaries. Test data was 80 pairwise rankings.
Exp. 2: Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coreference+Syntax+Salience+</td>
<td>80.0</td>
</tr>
<tr>
<td>Coreference+Syntax+Salience−</td>
<td>75.0</td>
</tr>
<tr>
<td>Coreference+Syntax−Salience+</td>
<td>78.8</td>
</tr>
<tr>
<td><strong>Coreference−Syntax+Salience+</strong></td>
<td>83.8</td>
</tr>
<tr>
<td>Coreference+Syntax−Salience−</td>
<td>71.3*</td>
</tr>
<tr>
<td>Coreference−Syntax+Salience−</td>
<td>78.8</td>
</tr>
<tr>
<td>Coreference−Syntax−Salience+</td>
<td>77.5</td>
</tr>
<tr>
<td>Coreference−Syntax−Salience−</td>
<td>73.8*</td>
</tr>
<tr>
<td><strong>Latent Semantic Analysis</strong></td>
<td>52.5**</td>
</tr>
</tbody>
</table>
Exp. 2: Learning Curve

![Learning Curve Diagram]

Accuracy (%) vs. Number of pairwise rankings in training data.
Exp. 3: Readability Assessment

- Task: embed entity-grids into a system that assesses document readability.
- Readability as a classification task: a learner needs to predict if an article is easy or difficulty to read.
- Data: corpus of Encyclopedia articles
Exp. 3: Versions

• 1. Use Schwarm and Ostendorf (2005), who use syntactic, semantic, and combination features.

• 2. Enrich Schwarm and Ostendorf with coherence-based features
Exp. 3: Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schwarm &amp; Ostendorf</td>
<td>78.56</td>
</tr>
<tr>
<td><strong>Schwarm &amp; Ostendorf, Coreference + Syntax + Salience</strong></td>
<td><strong>88.79</strong></td>
</tr>
<tr>
<td>Schwarm &amp; Ostendorf, Coreference − Syntax + Salience</td>
<td>79.49</td>
</tr>
<tr>
<td>Schwarm &amp; Ostendorf, Latent Semantic Analysis</td>
<td>78.56</td>
</tr>
<tr>
<td>Coreference + Syntax + Salience</td>
<td>50.90**</td>
</tr>
<tr>
<td>Coreference − Syntax + Salience</td>
<td>49.55**</td>
</tr>
<tr>
<td>Latent Semantic Analysis</td>
<td>48.58**</td>
</tr>
</tbody>
</table>
Exp. 3: Learning

![Graph showing accuracy vs. number of training documents for different methods.]

- Schwarz & Ostendorf, Coreference+Syntax+Salience
- Schwarz & Ostendorf
Discussion

• Empirical validation of the important of salience and syntactic information for coherence-based models
• Sacrificed linguistic faithfulness by being theory-agnostic
• ?: What about lexico-semantic knowledge?
• ?: Combined with global discourse analysis?
• ?: other languages?
Reference

Thanks!