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]$_o$.
3 [The case]$_s$ revolves around [evidence]$_o$ of [Microsoft]$_s$ aggressively pressuring [Netscape]$_o$ into merging [browser software]$_o$.
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]$_x$.

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics	Government	Suit	Earnings	
1	s	o	s	x	o	–	–	–	–	–	–	–	–	–	–	1
2	–	–	o	–	–	x	s	o	–	–	–	–	–	–	–	2
3	–	–	s	o	–	–	–	–	s	o	o	–	–	–	–	3
4	–	–	s	–	–	–	–	–	–	–	–	s	–	–	–	4
5	–	–	–	–	–	–	–	–	–	–	–	–	s	o	–	5
6	–	x	s	–	–	–	–	–	–	–	–	–	–	–	o	6

Entity-Grids

"A <u>local entity transition</u> 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.

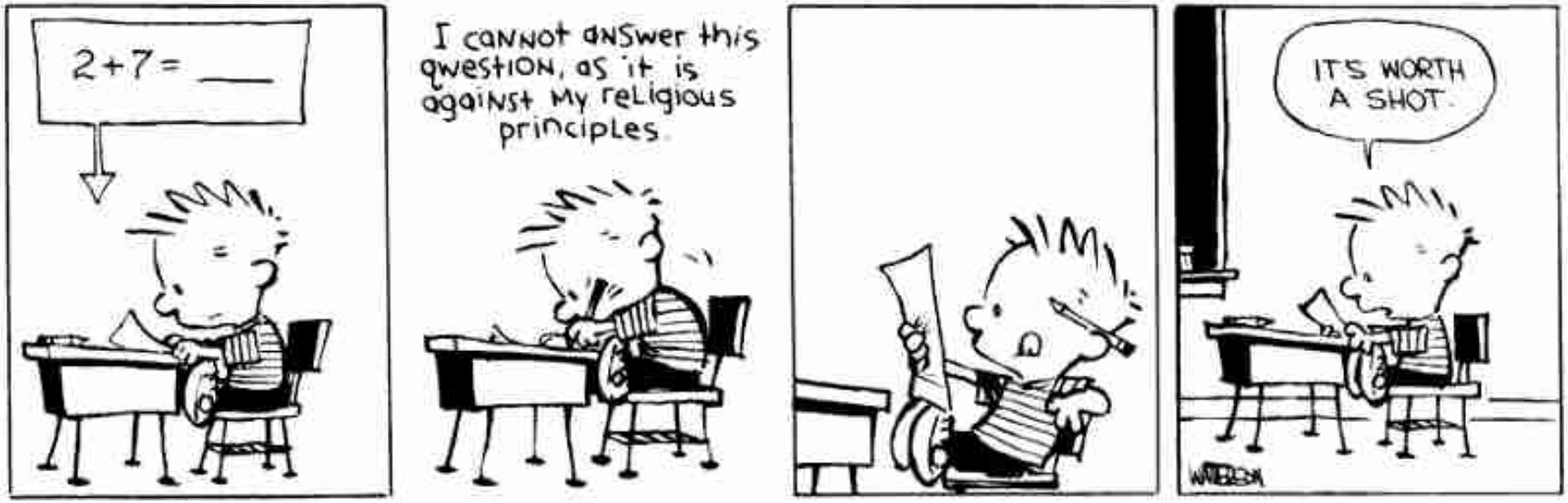
Entity-Grids

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics	Government	Suit	Earnings	
1	s	o	s	x	o	–	–	–	–	–	–	–	–	–	–	1
2	–	–	o	–	–	x	s	o	–	–	–	–	–	–	–	2
3	–	–	s	o	–	–	–	–	s	o	o	–	–	–	–	3
4	–	–	s	–	–	–	–	–	–	–	–	s	–	–	–	4
5	–	–	–	–	–	–	–	–	–	–	–	–	s	o	–	5
6	–	x	s	–	–	–	–	–	–	–	–	–	–	–	o	6

Example:

Probability of [O-]: 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

	Curdie	Peter	Father	Mother	Cottage	Mountain	Thing	Me	Thing	terrors
1	S	X	-	-	-	-	-	-	-	-
2	S	-	X	X	X	X	-	-	-	-
3	S	-	X	-	-	X	-	-	-	-
4	-	-	-	-	-	S	-	-	O	-
5	-	-	-	-	-	S	-	X	-	

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 +/-
  - S,O,X,-. Using Collins (1997) parser.
- 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

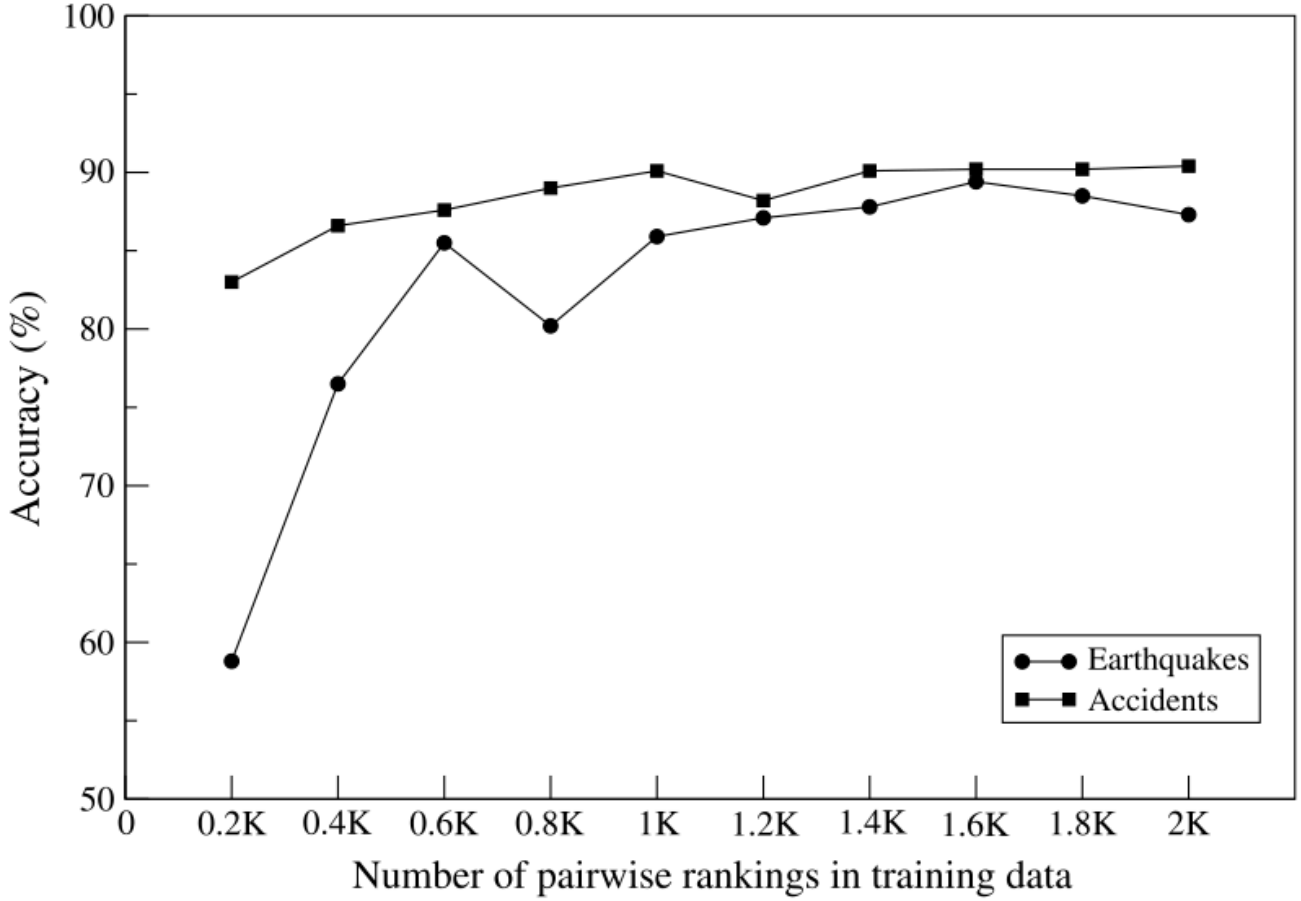
Exp.1: Sentence Ordering

- 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

Model	Earthquakes	Accidents
Coreference+Syntax+Salience+	**87.2**	**90.4**
Coreference+Syntax+Salience−	88.3	90.1
Coreference+Syntax−Salience+	86.6	88.4**
Coreference−Syntax+Salience+	83.0**	89.9
Coreference+Syntax−Salience−	86.1	89.2
Coreference−Syntax+Salience−	82.3**	88.6*
Coreference−Syntax−Salience+	83.0**	86.5**
Coreference−Syntax−Salience−	81.4**	86.0**
HMM-based Content Models	88.0	75.8**
Latent Semantic Analysis	81.0**	87.3**

Exp. 1: Learning Curve



Exp. 1: Learning Domains

Coreference+Syntax+Salience

Train \ Test	Earthquakes	Accidents
Earthquakes	**87.3**	67.0**
Accidents	69.7**	**90.4**
EarthAccid	86.7	88.5*

HMM-Based Content Models

Train \ Test	Earthquakes	Accidents
Earthquakes	**88.0**	31.7**
Accidents	60.3**	**75.8**

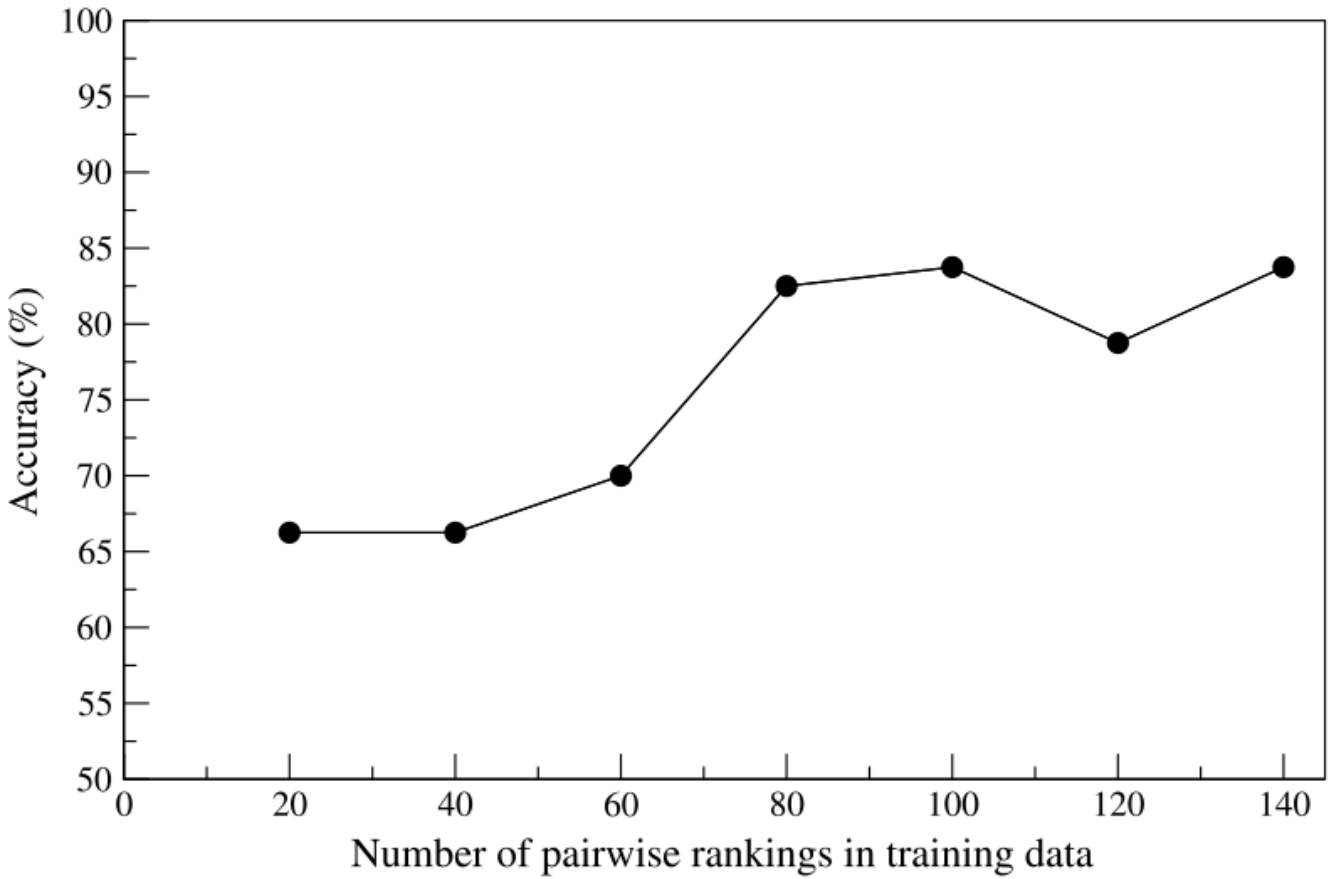
Exp. 2: Summary Coherence Rating

- 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

Model	Accuracy
Coreference+Syntax+Salience+	80.0
Coreference+Syntax+Salience−	75.0
Coreference+Syntax−Salience+	78.8
Coreference−Syntax+Salience+	**83.8**
Coreference+Syntax−Salience−	71.3*
Coreference−Syntax+Salience−	78.8
Coreference−Syntax−Salience+	77.5
Coreference−Syntax−Salience−	73.8*
Latent Semantic Analysis	52.5**

Exp. 2: Learning Curve



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

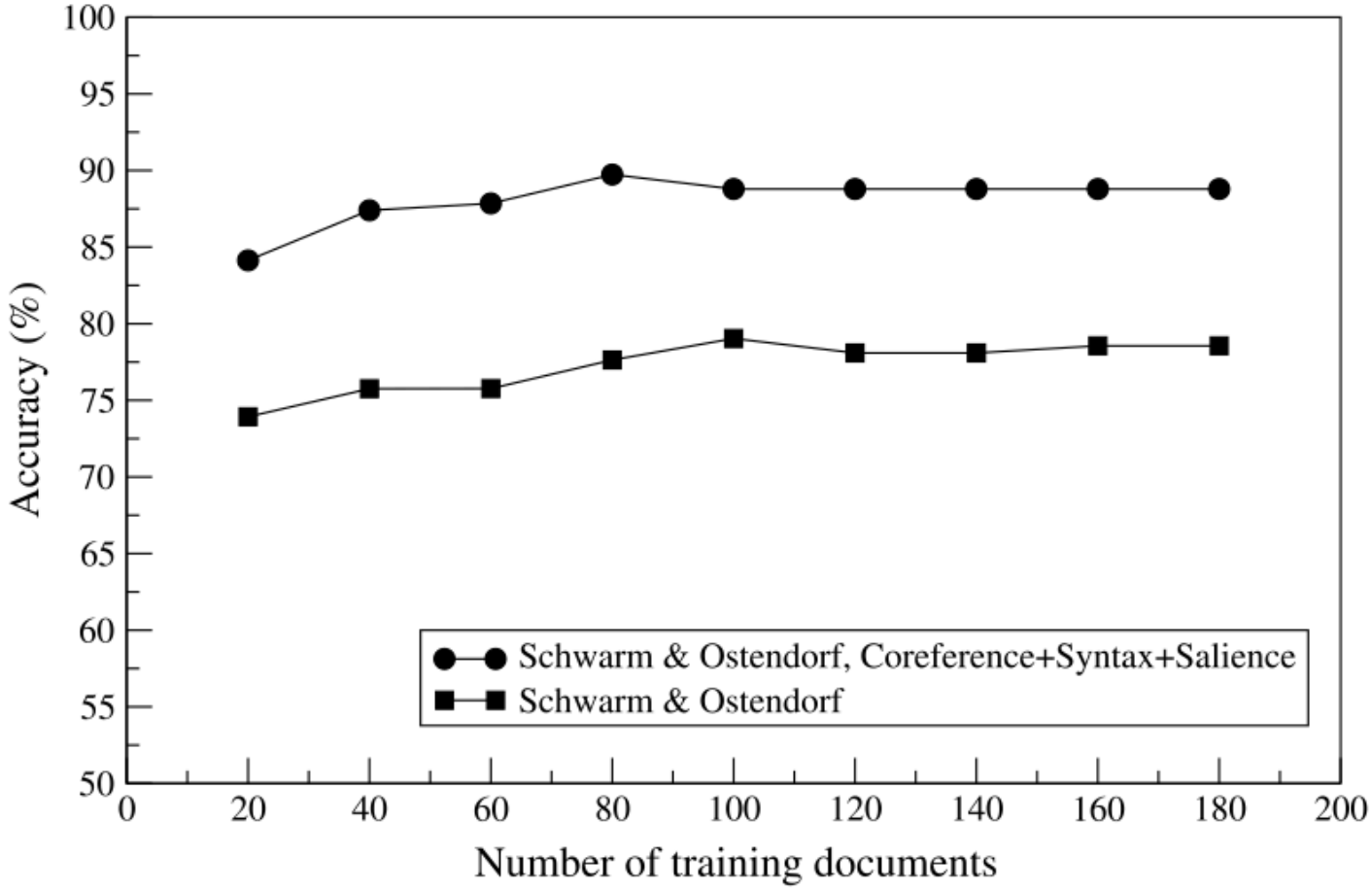
Exp. 3: Versions

- 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

Model	Accuracy
Schwarm & Ostendorf	78.56
Schwarm & Ostendorf, Coreference+Syntax+Salience+	**88.79***
Schwarm & Ostendorf, Coreference−Syntax+Salience+	79.49
Schwarm & Ostendorf, Latent Semantic Analysis	78.56
Coreference+Syntax+Salience+	50.90**
Coreference−Syntax+Salience+	49.55**
Latent Semantic Analysis	48.58**

Exp. 3: Learning



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

- Barzilay and Lapata (2008). Modeling Local Coherence: An Entity-Based Approach. ACL.

Thanks!