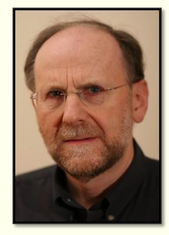
Extending the Entity-based Coherence Model with Multiple Ranks

Vanessa Wei Feng and Graeme Hirst

Department of Computer Science, University of Toronto, Canada



1. Objective

- Extend Barzilay and Lapata (2008)'s **entity-based coherence model** by learning from more fine-grained coherence preferences.
- Assign **multiple ranks** to a set of permutations (not just the original **pairwise rankings**).
- Study the effect of the permutations used in training, and the effect of the coreference component used in entity extraction.
- Evaluate with **sentence ordering** and **summary coherence rating**, compared to B&L's original model.

2. Entity-based Local Coherence

Entity-grid representation for a document *d*

	e_1	e_2	e_3	e_4
s_1	-	X	X	-
s_2	S	O	-	-
s_3	X	X	X	X

The syntactic role of entity e_i in sentence s_j:
S: subject; O: object; X: others; -: not present

Each row represents a sentence.

Each column represents an entity e_i in d.

Sequence of roles gives local transitions for each entity.

- **Entity extraction options**: Coreference resolution or not.
- Represent document as vector $\Phi(d) = (p_1(d), p_2(d), \ldots, p_m(d))$.
  p_t: proportions in text of each possible sequence t.

3. Experimental Setup

In the original model

Sentence ordering task

- Scramble sentences of text to produce random permutations.
- Permutations are considered to be less coherent than their source document.
- Training and testing on the **pairwise** preferences between an original document and its permutations.

Summary coherence rating task

- System-generated and human-composed summaries, rated by human judges for coherence.
- Training and testing on the **pairwise** preferences between summaries generated from the same input cluster.

In our extension

Sentence ordering task

- Assign **multiple ranks** to permutations, indicating the **dissimilarity** between their sentence orders and the original.
- Also train on the pairwise preferences among the permutations.
- Experiment with two sets of permutations:
  PS_{BL} (evenly distributed) and PS_M (favoring swapping near sentences).

Summary coherence rating task

- **Automatically assign** scores to system-generated summaries, by computing the **dissimilarity** between their (rough) sentence orders and the one in the reference summary.

NSERC CRSNG

UNIVERSITY OF TORONTO

Acknowledgement: This work was financially supported by the Natural Sciences and Engineering Research Council of Canada and by the University of Toronto.

4. Multiple Ranks Assignment

Dissimilarity metrics

Reference ordering: $\sigma = (1, 2, \ldots, N)$; test ordering: $\pi = (o_1, o_2, \ldots, o_N)$.

- **Kendall's τ (Lapata, 2006)**: measures the disagreement between π and σ in terms of m, the number of swaps of adjacent sentences to convert π into σ.
- **Average continuity (AC) (Bollegala et al., 2006)**: estimates the quality of σ by the number of correctly arranged continuous sentences, compared to π.
- **Edit distance (ED):** the minimum number of edits (insertions, deletions, and substitutions) needed to convert π into σ.

Rank assignment

Two options for assigning ranks to the permutations:

- **Raw**: rank the permutations by their dissimilarity scores.
- **Stratified**: C (3 to 6) ranks are assigned to the permutations according to their raw dissimilarity scores.

5. Data

Sentence ordering

- **Two datasets**:
  Earthquakes: pronominal realization of entities.
  Accidents: string repetition of entities.
- **Training and testing**: each with 100 texts and up to 20 permutations.

Summary coherence rating

- **Dataset**: MUC 2003 summaries (16 clusters, 5 systems).
- **Training**: 144 pairwise rankings.
- **Testing**:
  Same: 80 pairwise rankings among summaries within the same cluster.
  Full: 1520 pairwise rankings.

6. Results

Sentence ordering

Results: We show the model configurations with the best accuracies.

Perms	Earthquakes Metric	C	F&H	B&L	Accidents Metric	C	F&H	B&L
Condition: full coreference resolution with oracular information								
PS_{BL}	*ED*	3	**86.8**	85.3	*AC*	3	**83.3**	83.2
PS_M	*ED*	*N*	**87.9***	85.3	*ED*	4	**86.3***	81.7
Condition: full coreference resolution without oracular information								
PS_{BL}	*ED*	4	**77.4***	71.7	*AC*	3	**74.5**	73.8
PS_M	τ	3	**55.9**	49.2	*ED*	5	52.3	**53.2**
Condition: no coreference resolution								
PS_{BL}	τ	4	82.8	**83.7**	*AC*	3	**84.2****	80.1
PS_M	*ED*	5	**86.7****	82.6	*AC*	*N*	**86.6****	77.5

Significantly better than B&L: * ($p < .05$), ** ($p < .01$).
C=N: using raw option for rank assignment.

- Multiple ranking is effective in improving accuracies, especially when trained on the more realistic permutations PS_M.
- Different influence on two datasets when trained on PS_{BL}.
- This condition is not a good option when trained on PS_M.
- Coreference resolution is crucial to *Earthquakes*.
- Consistently outperforms B&L's model by a large margin.

Summary coherence rating

Rough sentence orders: via simple sentence alignment.

Entities	Metric	Same	Full
Coreference resolution	*AC*	**82.5**	**72.6***
	ED	**81.3**	**73.0****
	B&L	78.8	70.9
No coreference resolution	*AC*	76.3	72.0
	ED	78.8	71.7
	B&L	**80.8**	**72.3**

Unsupervised score assignment is competitive with B&L's model, which requires human annotations.

References

- Regina Barzilay and Mirella Lapata. 2008. Modeling local coherence: an entity-based approach. *Computational Linguistics*, 34(1):1-34.
- Danushka Bollegala, Naoaki Okazaki, and Mitsuru Ishizuka. 2006. A bottom-up approach to sentence ordering for multi-document summarization. In *Proceedings COLING/ACL 2006*, pages 385-392.
- Mirella Lapata. 2006. Automatic evaluation of information ordering: Kendall's tau. *Computational Linguistics*, 32(4):471-484.