A Neural Local Coherence Model

Dat Tien Nguyen[1] Shafiq Joty[2]
[1]University of Amsterdam
[2]Qatar Computing Research Institute

Entity-based Coherence

- Text is about entities: objects in the world
- A text addresses a common topic often covering multiple subtopics
- Earlier sentences provide context for later ones
- **Centering theory:** Coherence created by *repeated entity mentions*
- **Coherence modeling:** Model entity transition across sentences to distinguish a coherent from incoherent texts

Entity Grid and Its Extensions

Barzilay and Lapata (2008)

- Model grammatical role transmission of nouns (heads of NPs) across sentences
- Represent documents as distributions defined over **entity transition** (vectors of 4^k transitions probabilities $\{S, O, X, -\}^k$)
- Assessment of text coherence as a ranking problem in an SVM preference ranking framework

Table: Entity grid representation for a WSJ article.

	UNIT	PRODUCTS	RESEARCH	COMPANY	PARTS	CONTROLS	INDUSTRY	ELECTRONICS	TERM	CONCERN	AEROSPACE	EMPLOYEES	SERVICES	LOS ANGELES	EATON
s_0	O	–	X	X	–	–	–	–	–	–	–	X	–	–	X
s_1	–	–	–	–	–	–	–	–	S	–	–	–	–	–	–
s_2	–	O	–	–	–	–	X	–	–	–	–	O	O	X	–
s_3	–	–	–	–	X	X	–	X	–	O	X	–	–	–	S

Elsner and Charniak (2011)

- Include non-head nouns
- Extend grid to distinguish between entities by incorporating entity-specific features: named entity, noun class, modifiers, etc

Limitations of entity grid models

- Cannot capture long entity transitions
- Limit to learn task-specific features

Our Neural Coherence Model

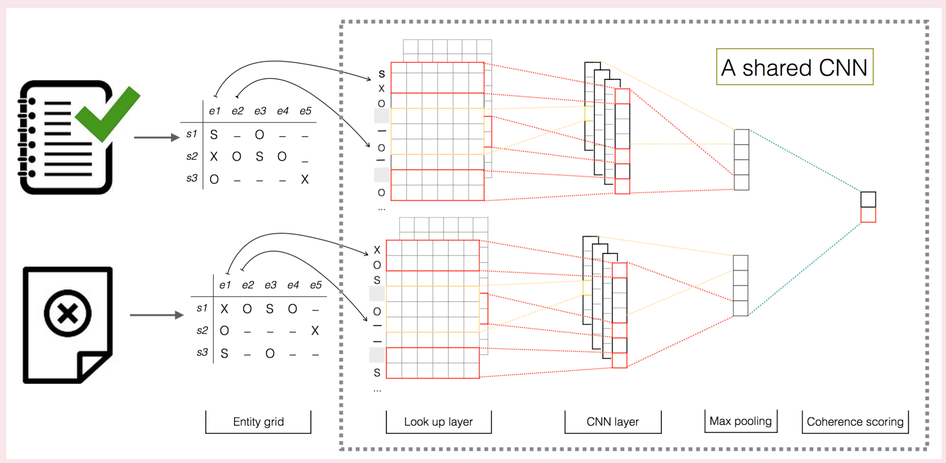


Figure: A Convolutional Neural Network (CNN) architecture for modeling local coherence.

Grid CNN:

- Use a **convolutional** approach
- Transform each grammatical role in grid into distributed representation
- Model sufficiently long entity transitions in location invariant way
- Train in end-to-end fashion on a target task (e.g., discrimination)
- Learn task-specific high level features

Extended Grid CNN:

- Incorporate entity-specific features
- Attach feature values with grammatical roles
  - if an entity e_j of type $PERSON$ appears as a subject (S) in a sentence s_i,
  - the grid entry $G_{i,j}$ can be encoded as $PERSON\_S$.

Pairwise end-to-end training:

- Input: ordered pairs (d_i, d_j), where document d_i is more coherent than d_j
- Use pairwise ranking approach to learn θ by minimizing the objective:

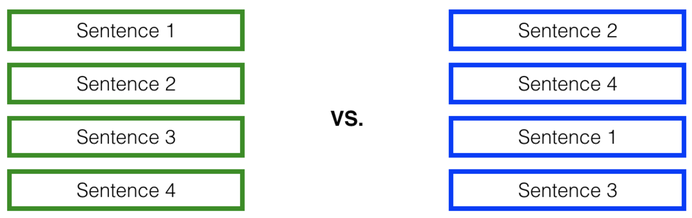
$$\Im(\theta) = max\{0, 1 - \phi(G_i|\theta) + \phi(G_j|\theta)\} \quad (1)$$

where G_i and G_j are entity grids of d_i and d_j, respectively
θ: set of CNN parameters

Experiment 1: Sentence Ordering

Discrimination

- A document is compared to its permutations of sentences



Insertion

- Remove and re-insert one sentence at a time into every position
- Examine permutations closer to the original ordering

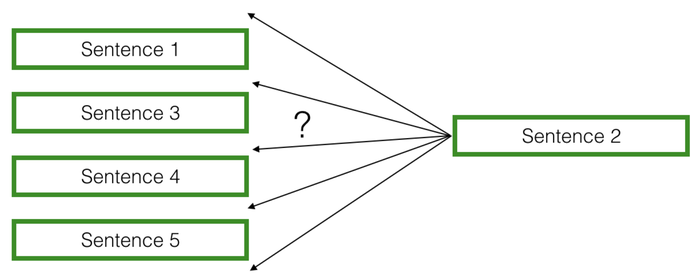


Table: WSJ dataset

	Sections	Doc.	Pairs	Avg. Sen.
TRAIN	00-13	1,378	26,422	21.5
TEST	14-24	1,053	20,411	22.3

Table: Coherence evaluation results on **Discr**imination and **Ins**ertion tasks. † indicates a neural model is significantly superior to its non-neural counterpart with p-value < 0.01.

	Discr.		**Ins.**
	Acc	F_1	
Random	50.00	50.00	12.60
Graph-based (G&S)	64.23	65.01	11.93
Dist. sentence (L&H)	77.54	77.54	19.32
Grid-all nouns (E&C)	81.58	81.60	22.13
Extended Grid (E&C)	84.95	84.95	23.28
Grid-CNN	85.57†	85.57†	23.12
Extended Grid-CNN	**88.69**†	**88.69**†	**25.95**†

Experiment 2: Summary Coherence Rating

- Compare rankings of summaries produce by model against rankings of summaries by human judges
- Dataset: Document Understanding Conference (DUC 2003)
  - Pair summaries from human judgments and automatics summarization systems
  - TRAIN: **144** pairs and TEST: **80** pairs

Table: Evaluation results on the **Summary Coherence Rating** task.

	Acc	F_1
Random	50.0	50.0
Graph-based (G&S)	80.0	81.5
Grid (B&L)	83.8	-
Grid-CNN	85.0	85.0
Extended Grid-CNN	86.3	86.3
Pre-trained Grid-CNN	86.3	86.3
Pre-trained Ext. Grid-CNN	**87.5**	**87.5**

Conclusion

Our contribution

- Local text coherence model based on a convolutional neural network that operates over entity transitions of arbitrary length
- Pairwise ranking approach to train the model end-to-end
- Yield the best performance reported so far
- Code and data: `https://github.com/datienguyen/cnn_coherence/`

Future work:

- Include rhetorical relations
- Extend the model to other forms of discourse, especially, asynchronous conversations (e.g., forum, email, etc)

QCRI معهد قطر لبحوث الحوسبة Qatar Computing Research Institute
جامعة حمد بن خليفة HAMAD BIN KHALIFA UNIVERSITY
UNIVERSITY OF AMSTERDAM

t.d.nguyen@uva.nl