

A Model of Coherence Based on Sentence Vector with GRU Neural Network

Team member: Qintai Liu, Xin Guan, Yi Zhou

CONTENTS

- 1 Introduction
- 2 Method
- 3 Result and Analysis
- 4 Conclusion

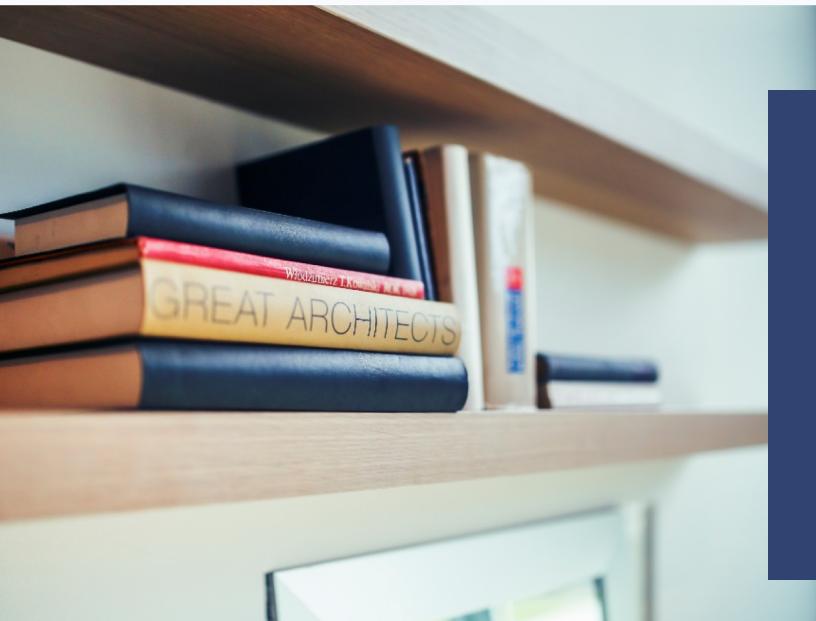


Part 1. Introduction

- Topic
- Problem

Topic

Part 1. INTRODUCTION



Local coherence Detection

Coherence:

- a significant metrics to evaluate text quality and readability
- makes a multi-sentence text meaningful

Problem

Part 1. INTRODUCTION

Part of sample article

BC-Philippines-Quake, 0194 | Strong Quake Rocks Philippines Island.

MANILA-LRB-AP-RRB-A strong earthquake rocked the Philippines island of Mindoro early Tuesday , killing at least two people and causing some damage , authorities said .

The 3:15 am quake had a preliminary magnitude of 6.7 and was centered near Baco on northern Mindoro Island , about 75 miles south of Manila , according to the Philippine Institute of Vulcanology and Seismology .

The U.S. Geological Survey in Menlo Park , Calif. , put the quake 's preliminary magnitude at 7.1 .

.....

Randomly permuted sample

MANILA -LRB- AP -RRB- A strong earthquake rocked the Philippines island of Mindoro early Tuesday , killing at least two people and causing some damage , authorities said .

The 3:15 am quake had a preliminary magnitude of 6.7 and was centered near Baco on northern Mindoro Island , about 75 miles south of Manila , according to the Philippine Institute of Vulcanology and Seismology .

BC-Philippines-Quake ,0194 | Strong Quake Rocks Philippines Island .

The U.S. Geological Survey in Menlo Park , Calif. , put the quake 's preliminary magnitude at 7.1 .

.....

Data: from reports about earthquakes from the Associated Press

Goal: Build a neural network model to measure local coherence



Part 2. Methods

- Baseline Model
- Advanced Model

Baseline Model — Entity Grid + SVM

Part 2. METHODS



BC-Japan-Earthquake | Earthquake Hits Japan Sea Coast. TOKYO -LRB- AP -RRB- A relatively strong earthquake with preliminary magnitude of 6.0 hit the Japan Sea coast Saturday but no damage was reported. The earthquake was centered in the Japan Sea off the coast of Niigata, 256 kilometers -LRB- 160 miles -RRB- northwest of Tokyo. It measured four on the Japanese scale of seven in Niigata. An earthquake of that size can shake houses and topple vases

- Sentence 1: 'bc-japan-earthquake earthquake hits japan sea coast',
- Sentence 2: 'tokyo ap earthquake magnitude japan sea coast saturday damage',
- Sentence 3: 'earthquake japan sea coast niigata kilometers miles tokyo',
- Sentence 4: 'it scale niigata',
- Sentence 5: 'earthquake size houses vases',
- Sentence 6: 'television footage niigata sado island damage',
- Sentence 7: 'earthquake japan jan magnitude scale parts city kobe',
- Sentence 8: 'earthquake people'

	bc	japan	earthquake	hits	sea	coast	tokyo	ap	magnitude	saturday	...	vases
sentence1	0	1	0	1	0	1	0	1	0	0	0	0
sentence2	1	0	0	0	1	1	1	0	0	0	0	0
sentence3	0	0	0	0	1	0	1	0	0	0	0	0
sentence4	0	0	0	0	0	0	0	0	0	0	0	0
sentence5	0	0	0	0	0	1	0	0	0	0	0	0
sentence6	0	0	0	0	1	0	1	0	0	0	1	0
sentence7	0	0	1	0	0	1	0	0	0	0	0	1
sentence8	0	0	0	0	0	1	0	0	0	0	0	0

8 rows × 28 columns



	0->0	0->1	1->0	1->1
0	0.627551	0.147959	0.168367	0.056122

SVM

Advanced Model

Part 2. METHODS



Sentence Model

- Recurrent neural network
- Recursive neural network



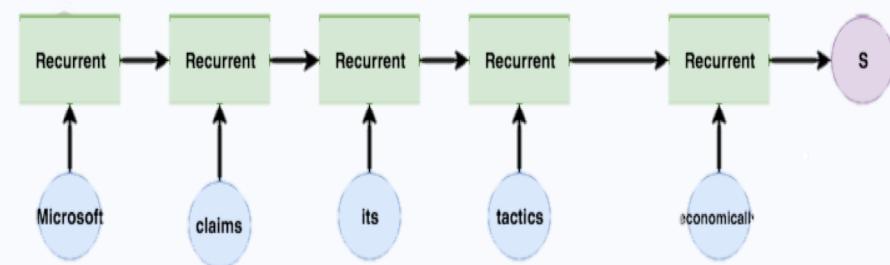
Coherence Model

- GRU

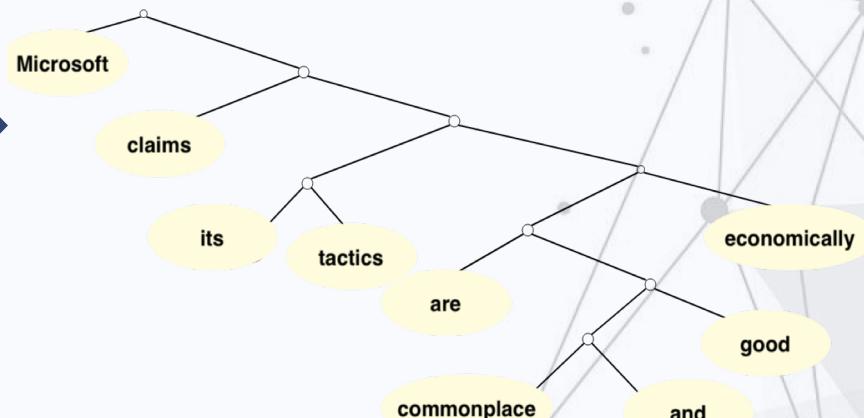
Sentence Model——Tree RNN

Part 2. METHODS

Recurrent neural network



Recursive neural network



e.g. Microsoft claims its tactics are commonplace and good economically

Implementation—Tree RNN

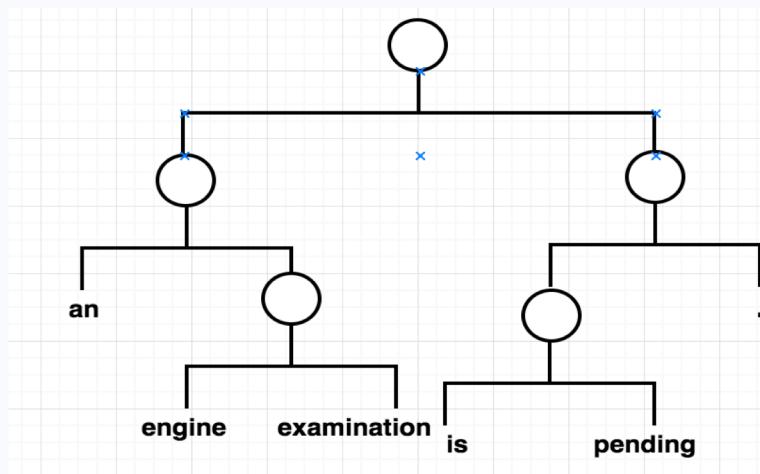
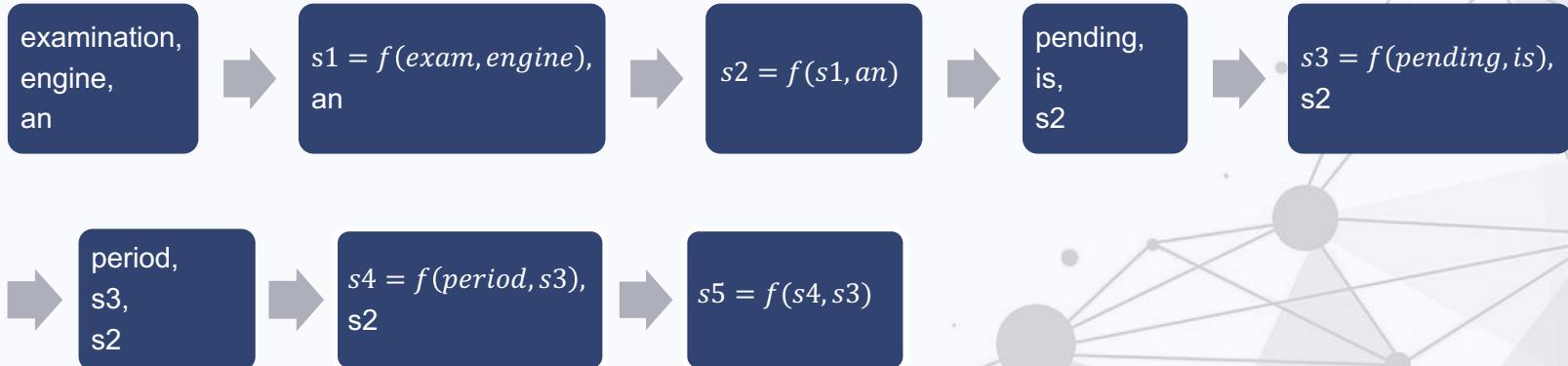
Part 2. METHODS

CoreNLP

Example:

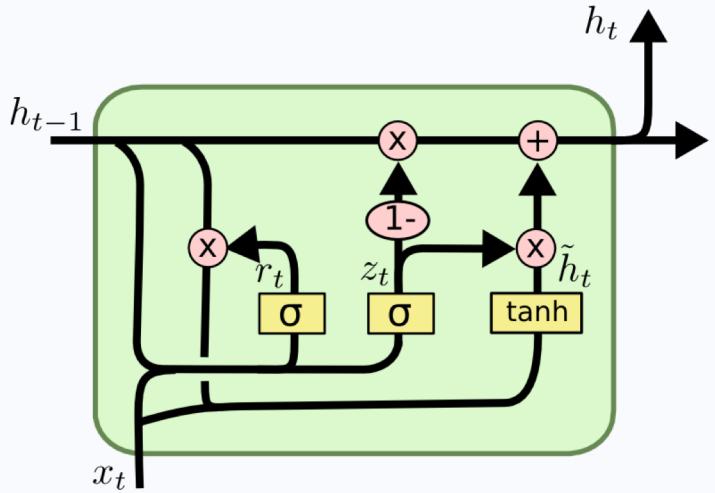
- an engine examination is pending.
- ((an (engine examination)) ((is pending) .))
- an engine examination) is pending) .))

$$f(a, b) = \tanh(W \cdot [a, b] + bias)$$



Coherence Model — GRU

Part 2. METHODS



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

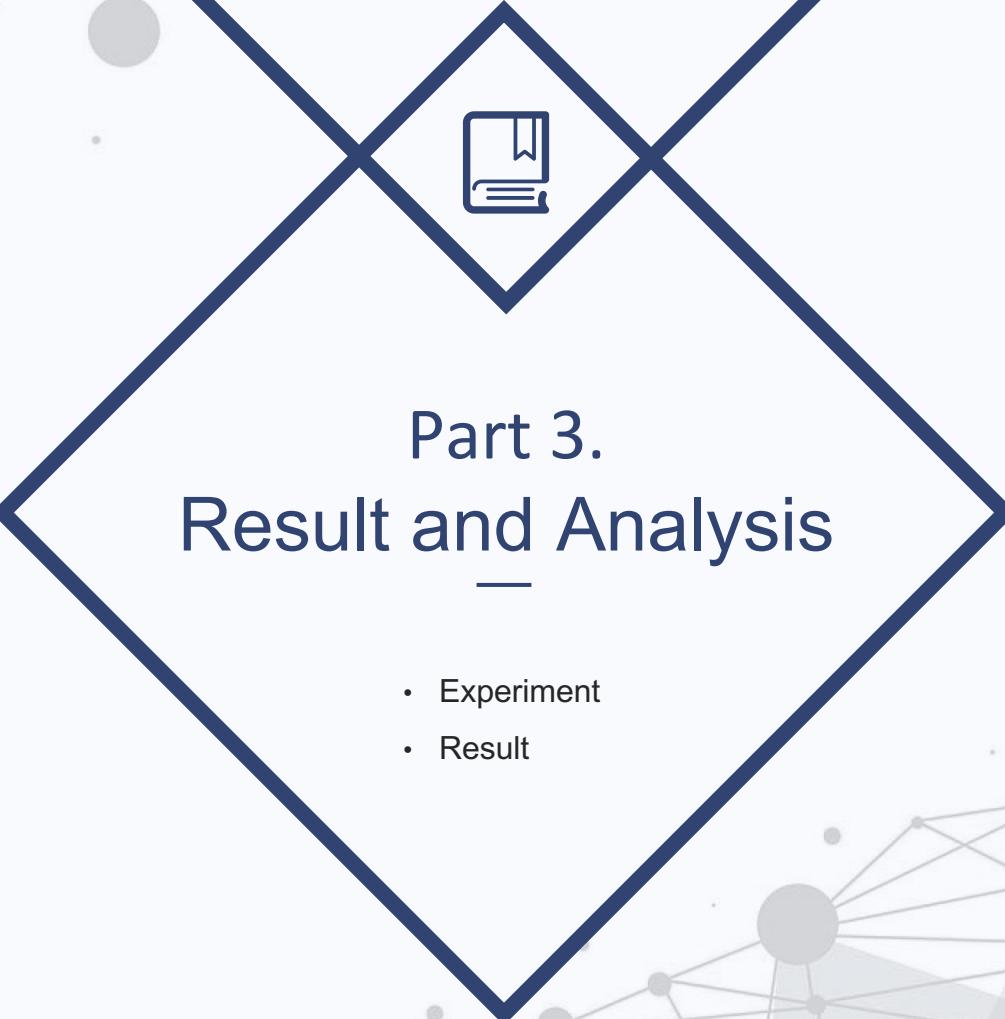
$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Objective function:

$$J(\theta) = \frac{1}{N} \sum_{T \in \text{trainset}} \{-y_T \log[p(y_T = 1)] - (1 - y_T) \log[1 - p(y_T = 1)]\} + Q \sum_{\theta \in \Theta} \theta^2$$

where $\Theta = [W_{\text{Recursive}}, W_z, W_r, U]$, $p(y_T = 1) = \sinh(U^T h_{t_n} + b)$



A blue diamond-shaped frame surrounds the title and a list of bullet points. The background features a light gray network of interconnected circles and dots.

Part 3. Result and Analysis

- Experiment
- Result

Experiments

Part 3. RESULT AND ANALYSIS



Dataset:

two corpora widely employed for coherence prediction
(Barzilay and Lee, 2004; Barzilay and Lapata, 2008; Elsner et al., 2007).

- One contains reports on airplane accidents from the National Transportation Safety Board
- The other contains reports about earthquakes from the Associated Press



Training and Testing:

- 200 articles for training
- 190 articles for testing
- A maximum of 20 random permutations were generated for each test article to create the pairwise data
- total of 1990 test pairs for the accident corpus and 1720 for earthquakes

Result

Part 3. RESULT AND ANALYSIS

Comparison of Different Coherence Models

	Acci	Earthquake	Average
Entity Gird	0.803	0.838	0.820
Recursive NN+GRU	0.898	0.997	0.948
Recursive	0.864	0.976	0.920
Recurrent	0.840	0.951	0.895
HMM	0.822	0.938	0.880
Graph	0.846	0.635	0.740

Analysis:

- Accuracy
- Robustness
- Time Consumption

Code: <https://github.com/qltf8/ds1012/blob/master/core.ipynb>



Part 4. Conclusion

Conclusion

Part 4. CONCLUSION

In this project, we combine Tree RNN and GRU approaches to the sentence-ordering (coherence) task, using compositional sentence representations.

The proposed approach obtains state-of-art performance on the standard coherence evaluation tasks.

Reference

- Regina Barzilay and Lillian Lee. 2004. Catching the drift: Probabilistic content models, with applications to generation and summarization. In *HLT-NAACL*, pages 113-120.
- Regina Barzilay and Mirella Lapata. 2008. Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34.
- Micha Elsner, Joseph L Austerweil, and Eugene Char- niak. 2007. A unified local and global model for dis- course coherence. In *HLT-NAACL*, pages 436–443.
- Jiwei Li and Eduard Hovy. 2014. A model of coherence based on distributed sentence representation. In *Proceedings of Empirical Methods in Natural Language Processing* , volume arXiv:1510.03055.

Thanks

Team member: Qintai Liu, Xin Guan, Yi Zhou

2018-5-2

DS-GA 1012 NLU project