COLING 2022

Pipeline Coreference Resolution Model for Anaphoric Identity in Dialogues

October 17, 2022

Damrin Kim*, Seongsik Park, Mirae Han and Harksoo Kim
Department of Artificial Intelligence, Konkuk University, Seoul, Republic of Korea
{ekafls33, a163912, future26, nlpdrkim}@konkuk.ac.kr

Konkuk University
Natural Language
Processing Lab

CONTENTS

Introduction

Coreference resolution

- Coreference(Anaphoric) resolution
  - The task of extracting mentions from a given document and clustering mentions from the same entity

- Mention detection
  - The task of extracting candidate word ranges that are likely to be mentions within a sentence

- Mention
  - A span of words that are likely to be coreferenced in a sentence

Introduction

Coreference resolution

- Chelsea Football Club is an English professional football club based in Fulham, West London. Founded in 1905, they play their home games at Stamford Bridge. The club competes in the Premier League, the top division of English football. They won their first major honour, the League championship, in 1955. The club won the FA Cup for the first time in 1970, their first European honour, the Cup Winners' Cup, in 1971, and becoming only the third English club to win the Club World Cup in 2022.

Introduction

Problem

- End-to-end Neural Coreference Resolution(Lee et al., EMNLP 2017)
  - Mention detection
    - Extracts all word spans that are likely to be mentions
    - Prunes them with mention scores which are calculated by the model
  - Coreference resolution
    - Pairs pruned mentions to calculate mention pair scores
    - Clusters them into final mention pairs

[End-to-end Neural Coreference Resolution](https://aclanthology.org/D17-1018)(Lee et al., EMNLP 2017)

Introduction

Problem

- Problems of End-to-end Neural Coreference Resolution
  - Uses fixed pruning ratio
    - High pruning rate
      - The increased number of non-correct candidates → Higher computation and complexity
    - Low pruning rate
      - Lower computation and complexity → More likely to remove correct mention

    → A direct impact on the cross-reference resolution model & High cost in time to find the best ratio

  - Long training time and high model complexity($O(n^4)$)

  - Takes long training time and has high model complexity

Konkuk University Natural Language Processing Lab

[End-to-end Neural Coreference Resolution](https://aclanthology.org/D17-1018)(Lee et al., EMNLP 2017)

Introduction

Solution

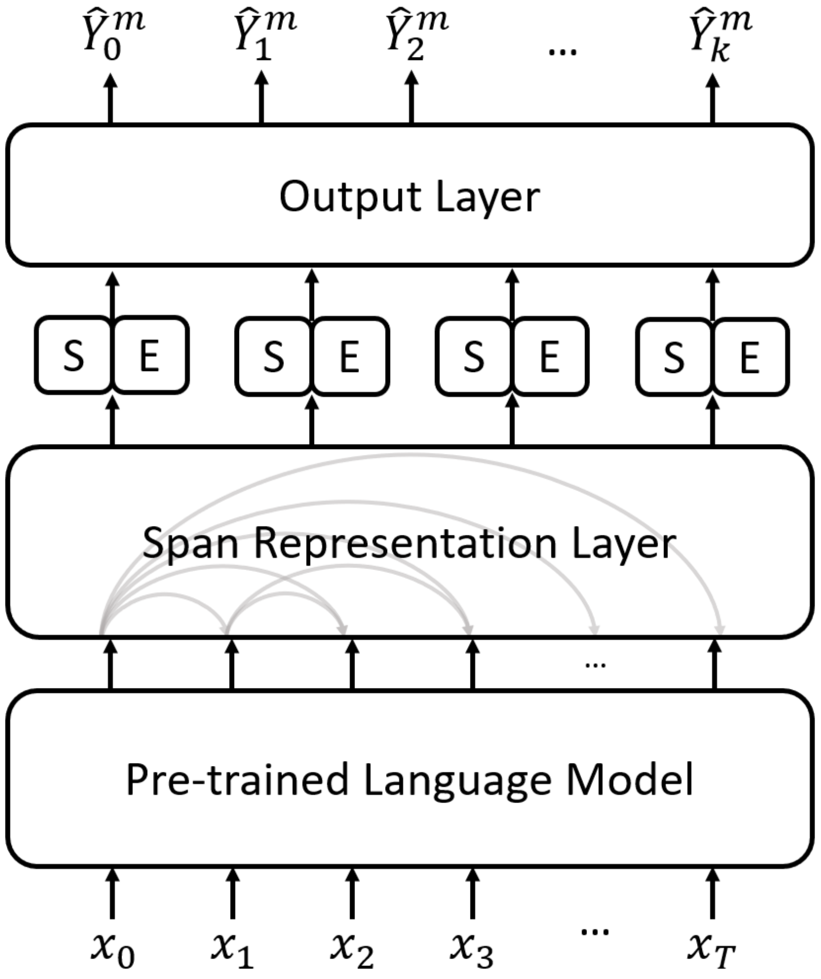
- Our goals
  - Accelerate training time and reduce model complexity
  - Reduce the possibility of missing correct mentions or using incorrectly predicted mentions

- Two-stage coreference resolution pipeline model
  - Step 1: mention detection
    - Learns mentions by calculating mention scores of all candidate word spans
  - Step 2: coreference resolution
    - Calculates mention pair scores by pairing the predicted mentions in the Mention Detection model

Model

Mention detection

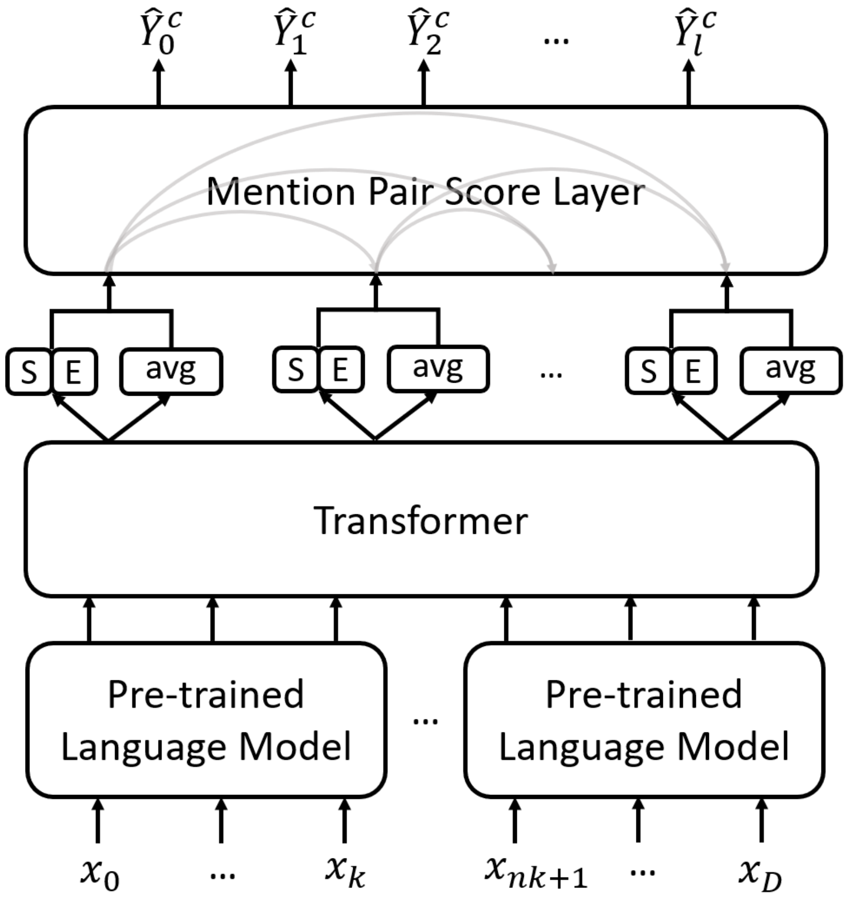
- Pre-trained language model output
  - $X = \{x_1, x_2, \cdots, x_T\}$
- Span representation
  - $g_m(i) = \left[x_{START(i)}, x_{END(i)}\right]$
- Mention score
  - $S_m(i) = W_m \cdot FNN_m\left(g_m(i)\right)$
- Loss function
  - $loss_m = -\sum Y_i^m \log\left(\hat{Y}_i^m\right)$



Model

Coreference resolution

- Pre-trained language model output
  - $X = \{x_1, x_2, \cdots, x_D\}$
- Mention representation
  - $g_c(i) = [x_{START(i)}, x_{END(i)}, avg(x_{START(i)}; x_{END(i)}), \phi(i)]$
- Coreference score
  - $S_c(i,j) = W_c \cdot FNN_c(g_c(i), g_c(j))$
- Loss function
  - $loss_c = -\sum Y_i^c \log(\hat{Y}_i^c)$



Experiments

Dataset

- CODICRAC 2022 Shared-Task datasets
  - Train/Dev : Light, AMI, PSUA, SWBD, ARRAU
  - Test : Light, AMI, PSUA, SWBD

	Light	AMI	PSUA	SWBD	ARRAU
Num of documents	20	7	21	11	202
Num of sentences	909	4,140	813	1,343	4,230
Num of words	11,495	33,741	9,185	14,992	110,440
Num of mentions	3,907	8,918	2,743	4,024	34,454
Num of clusters	1,803	4,391	1,513	2,362	23,238
Avg num of speakers	3	4	2	2	-

Experiments

Mention detection

- Performance of mention detection

	Light	AMI	PSUA	SWBD
Precision	94.76	88.15	90.67	92.60
Recall	89.72	74.01	88.70	78.58
F1-score	92.17	80.46	89.67	85.02

Experiments

Coreference resolution

- Performance of coreference resolution

		Light	AMI	PSUA	SWBD
MUC	P	73.45	36.05	70.04	53.83
	R	83.31	77.67	83.23	83.12
	F1	78.07	49.24	76.07	65.34
B^3	P	76.72	46.22	70.00	58.46
	R	55.14	64.06	69.97	69.08
	F1	64.16	53.70	69.99	63.33
$CEAF_e$	P	63.08	70.76	76.31	70.73
	R	62.07	31.57	51.00	44.07
	F1	62.27	43.66	61.14	54.31
CoNLL F1 score		**68.27**	**48.87**	**69.06**	**60.99**

Experiments

Training time

- Training time per epoch

Model		Training time
end-to-end (Lee et al., 2017)		60 min
pipeline	mention detection	1 min
	coreference resolution	5 min

- Comparison of coreference resolution performances

	Light	AMI	PSUA	SWBD
end-to-end	70.45	35.34	67.52	61.27
ours	68.27 (-2.18)	48.87 (+13.53)	69.06 (+1.54)	60.99 (-0.28)

Conclusion

- Proposal of a two-stage coreference resolution pipeline model
  - A mention detection model using mention spans
  - A coreference resolution model using mention pairs

- Maintains similar performance to the end-to-end model while reducing training time and complexity of the model($O(n^2)$)

- Future work
  - A robust Mention Detection Model for Noun Phrases
  - A coreference resolution model using graphs to reflect information between pairs of mentions

COLING 2022

Q&A

Thank you for listening

ekafls33@konkuk.ac.kr

Konkuk University
Natural Language
Processing Lab