Cross-document Coreference Resolution over Predicted Mentions

Arie Cattan, Alon Eirew, Gabriel Stanovsky, Mandar Joshi and Ido Dagan

Findings of ACL 2021

Cross-document Coreference Resolution

U.S President nominates new surgeon general: MacArthur “genius grant” fellow Regina Benjamin on the July 13, 2009. Obama emphasize his decision in “her extensive and distinguished career in medicine”.

President Obama will name Dr. Regina Benjamin as U.S. Surgeon General on Monday in a Rose Garden announcement, later this morning.

News that Barack Obama may name Dr. Sanjay Gupta of Emory University and CNN as his Surgeon General has caused a spasm of celebrity reporting.

CNN’s management confirmed that Dr. Gupta had been approached by the Obama team on March 2009. The chief medical correspondent has declined comment.

Cross-document Coreference Resolution

U.S President nominates new surgeon general: MacArthur "genius grant" fellow Regina Benjamin on the July 13, 2009. Obama emphasize his decision in "her extensive and distinguished career in medicine".

President Obama will name Dr. Regina Benjamin as U.S. Surgeon General on Monday in a Rose Garden announcement, later this morning.

News that Barack Obama may name Dr. Sanjay Gupta of Emory University and CNN as his Surgeon General has caused a spasm of celebrity reporting.

CNN's management confirmed that Dr. Gupta had been approached by the Obama team on March 2009. The chief medical correspondent has declined comment.

Related Work

There are two main limitations in previous work:

Related Work

There are two main limitations in previous work:

- ✗ Operate only on gold mentions

Related Work

There are two main limitations in previous work:

- ✗ Operate only on gold mentions
- ✗ Not efficient

Related Work

There are two main limitations in previous work:

- ❌ Operate only on gold mentions
- ❌ Not efficient
  - Barhom et al. (2019) compute pairwise scores at each cluster merge

Related Work

There are two main limitations in previous work:

- ❌ Operate only on gold mentions
- ❌ Not efficient
  - Barhom et al. (2019) compute pairwise scores at each cluster merge
  - Zeng et al. (2020) apply BERT on two sentences for every mention pair (!)

Related Work

There are two main limitations in previous work:

- ❌ Operate only on gold mentions
- ❌ Not efficient
  - Barhom et al. (2019) compute pairwise scores at each cluster merge
  - Zeng et al. (2020) apply BERT on two sentences for every mention pair (!)
  - Caciularu et al. (2021) apply CDLM on two full documents for every mention pair (!)

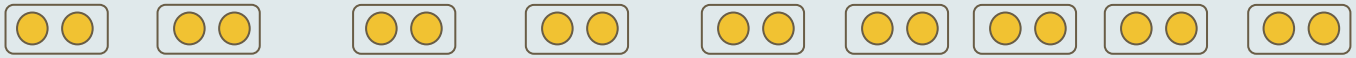
Our model – Extension of E2E-coref

Doc 1: Barack Obama nominates new surgeon general Dr. Sanjay Gupta

Doc 2: President Obama will name Regina Benjamin as surgeon general

…..

Our model – Extension of E2E-coref



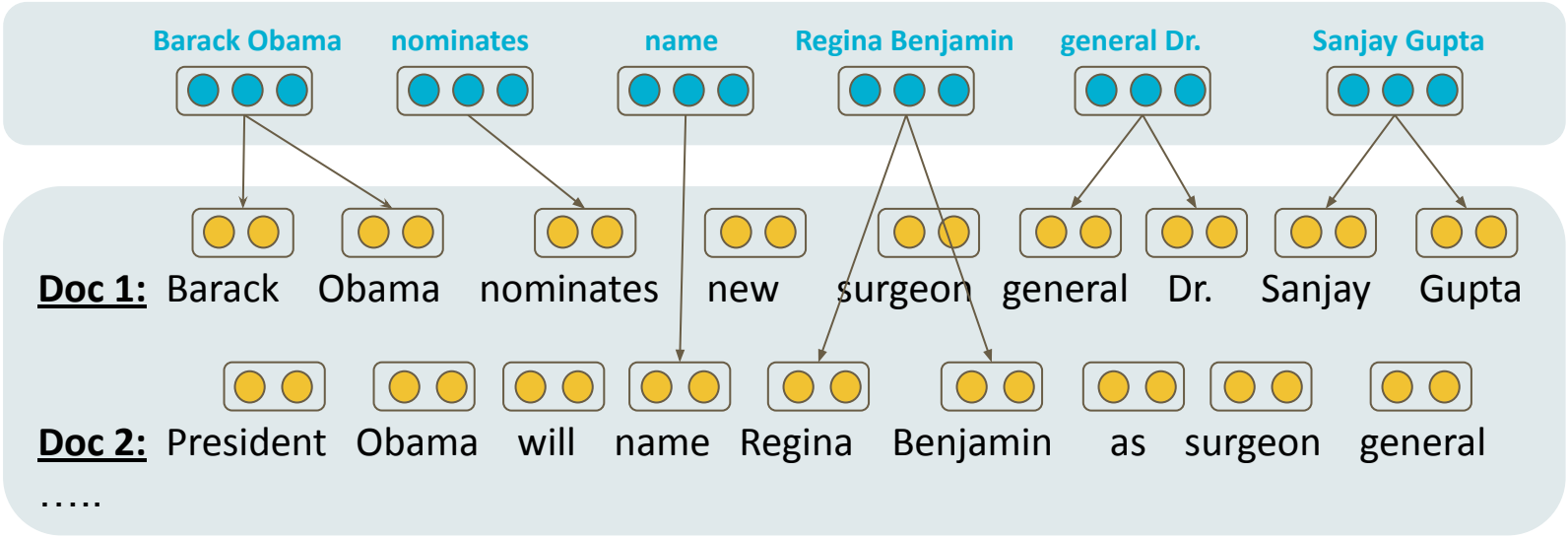
Doc 1: Barack Obama nominates new surgeon general Dr. Sanjay Gupta

Doc 2: President Obama will name Regina Benjamin as surgeon general

…..

Contextualized representations

Our model – Extension of E2E-coref

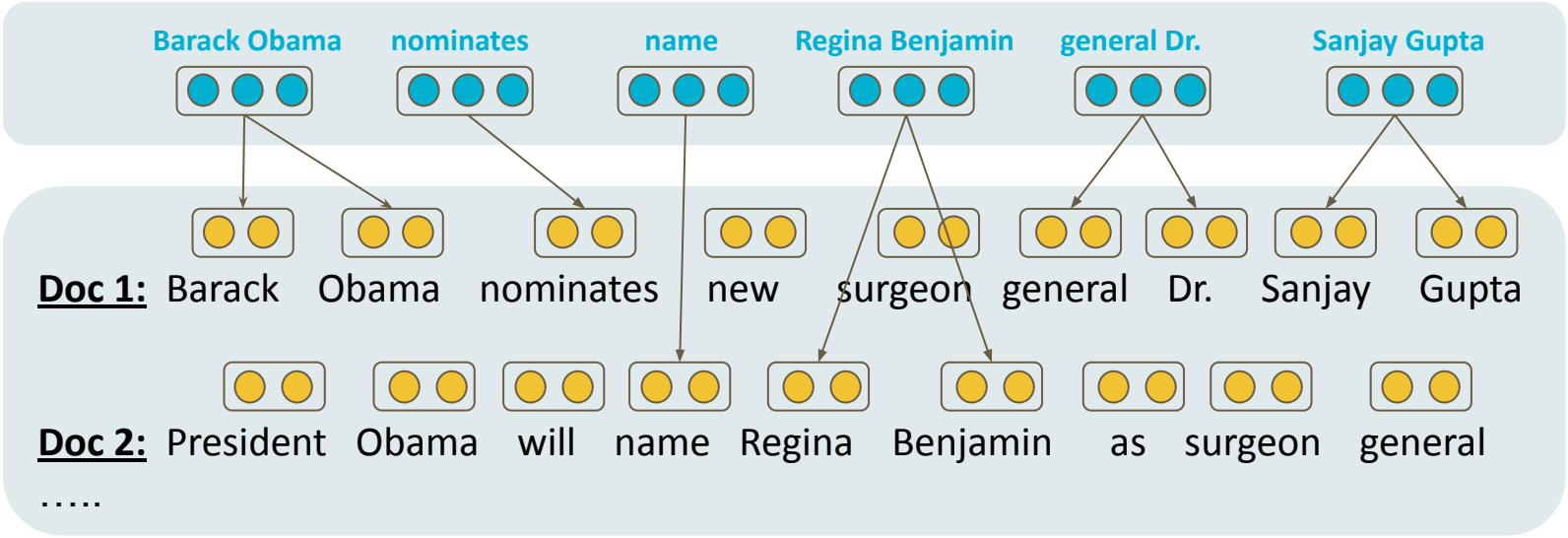


Span representations g(i)

Contextualized representations

Our model – Extension of e2e-coref

$$g_i = [x_{\text{FIRST(i)}}, x_{\text{LAST(i)}}, \hat{x}_i, \phi(i)]$$

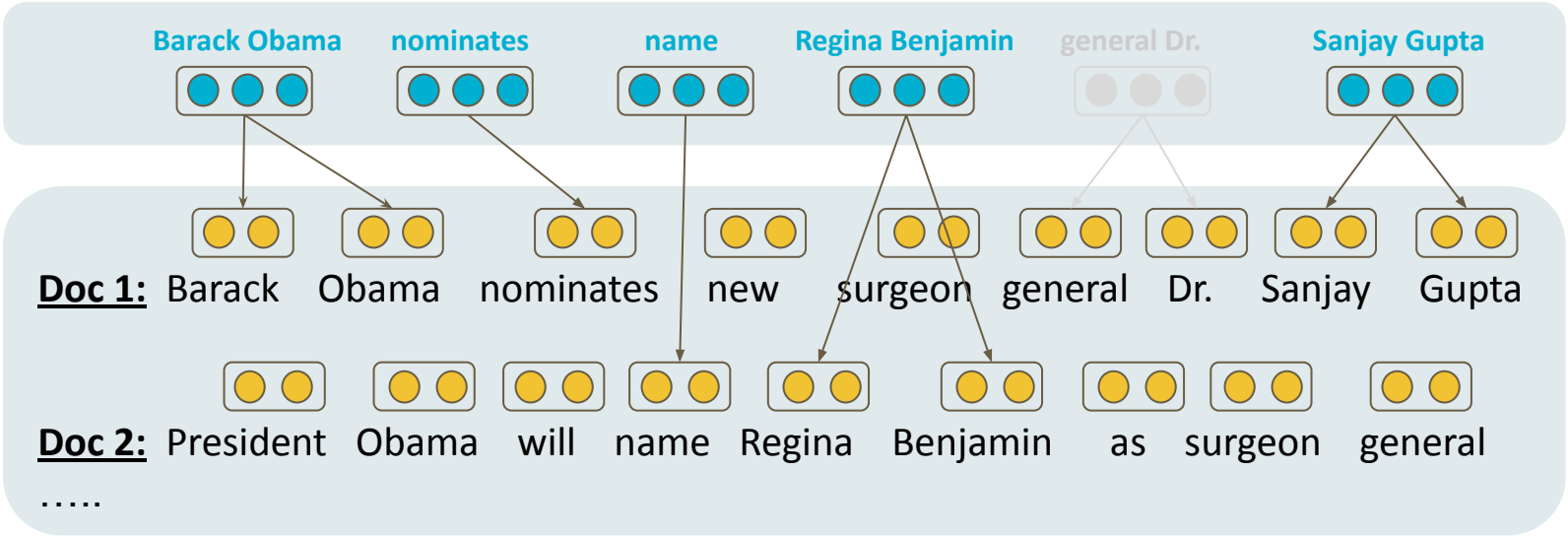


Span representations g(i)

Contextualized representations

Our model – Extension of e2e-coref

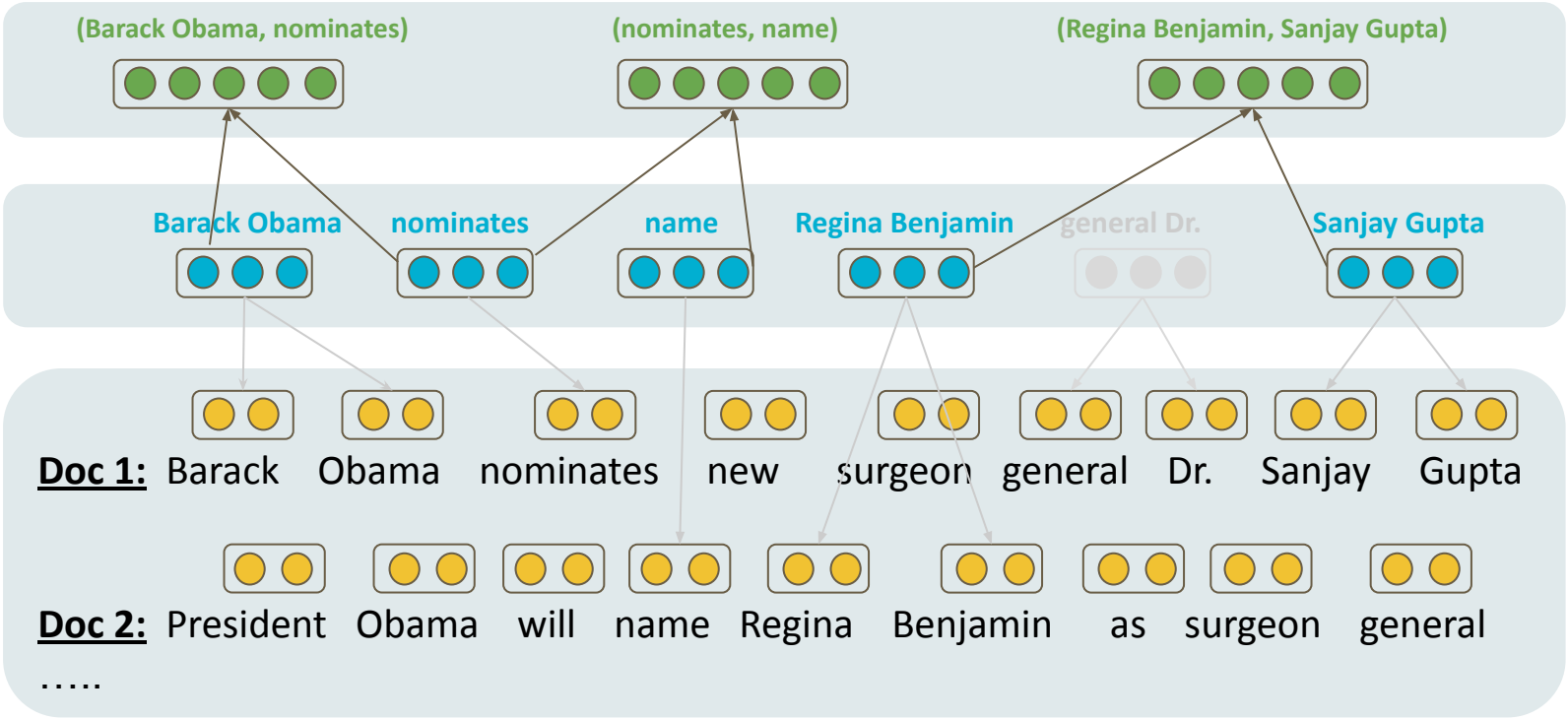
Pruning $s_m(i) = \text{FFNN}_m(g_i)$



Span representations g(i)

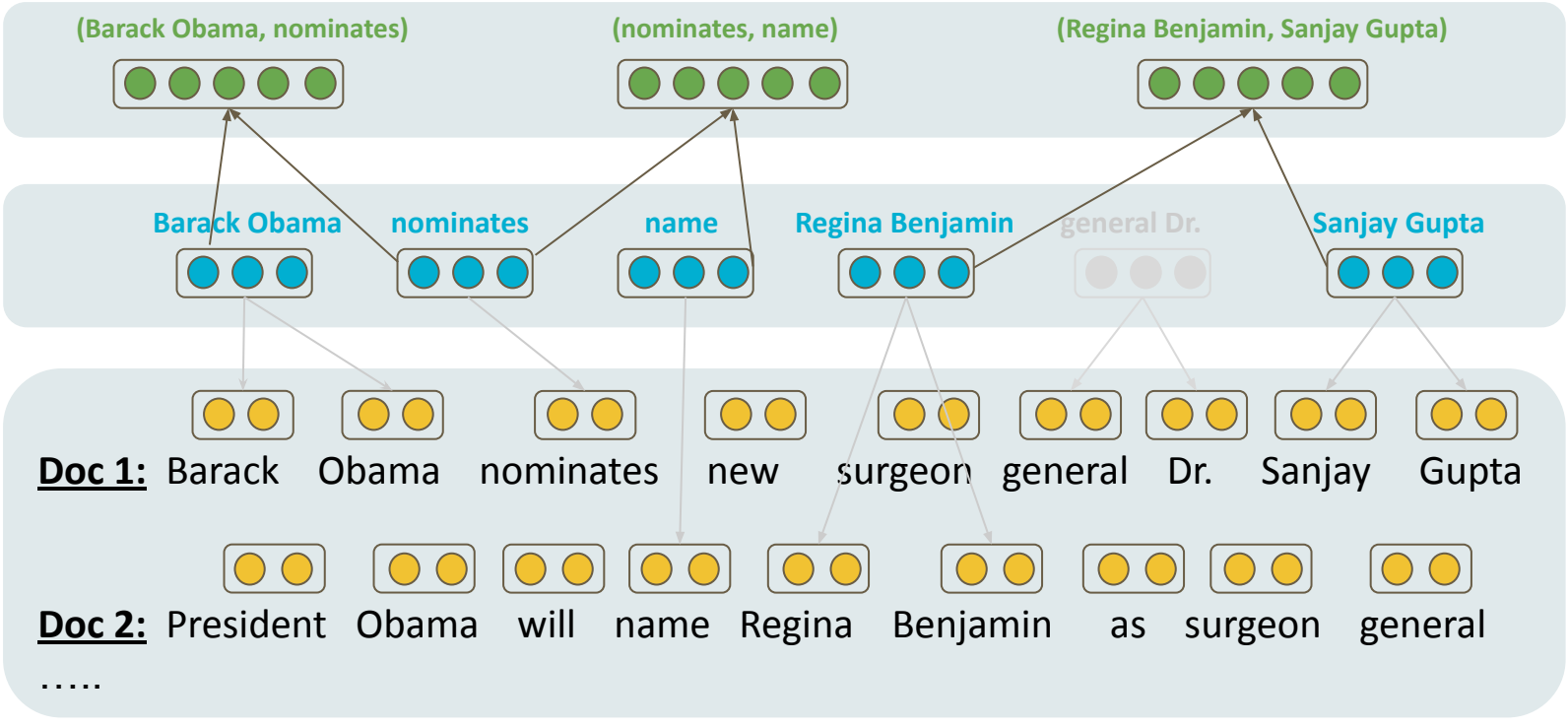
Contextualized representations

Our model – Extension of e2e-coref



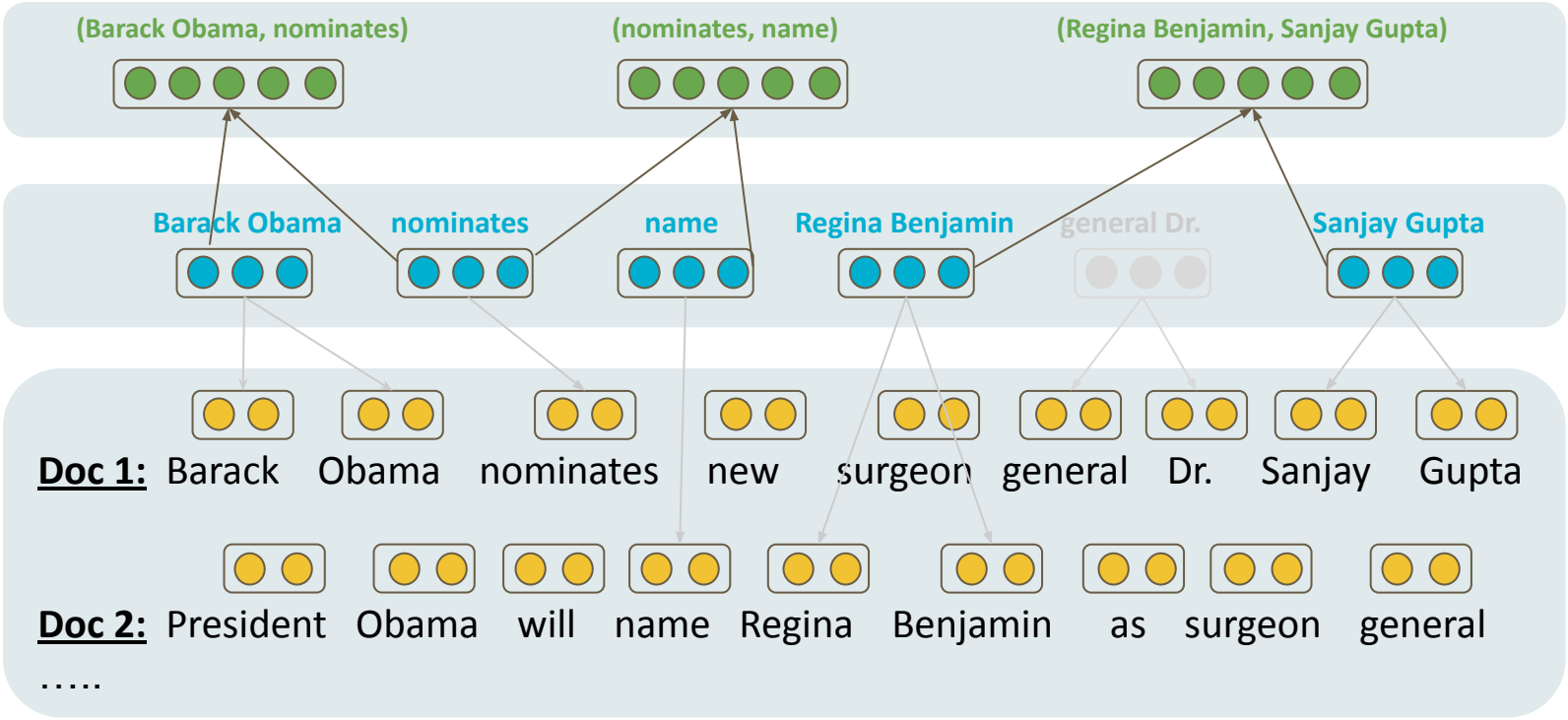
Our model – Extension of e2e-coref

$$s_a(i,j) = \text{FFNN}_a([g_i, g_j, g_i \circ g_j])$$



Our model – Extension of e2e-coref

Overall score: $s(i,j) = s_m(i) + s_m(j) + s_a(i,j)$



Our model – Extension of e2e-coref

- At inference, we use an agglomerative clustering to obtain the clusters over the overall score $s(i, j)$

Our model – Extension of e2e-coref

- At inference, we use an agglomerative clustering to obtain the clusters over the overall score $s(i, j)$

Advantages:

Our model – Extension of e2e-coref

- At inference, we use an agglomerative clustering to obtain the clusters over the overall score $s(i, j)$

Advantages:

✅ **Single pass of RoBERTa for each document**

Our model – Extension of e2e-coref

- At inference, we use an agglomerative clustering to obtain the clusters over the overall score $s(i, j)$

Advantages:

- ✅ **Single pass of RoBERTa for each document**
- ✅ **Pairwise scores are calculated only once**

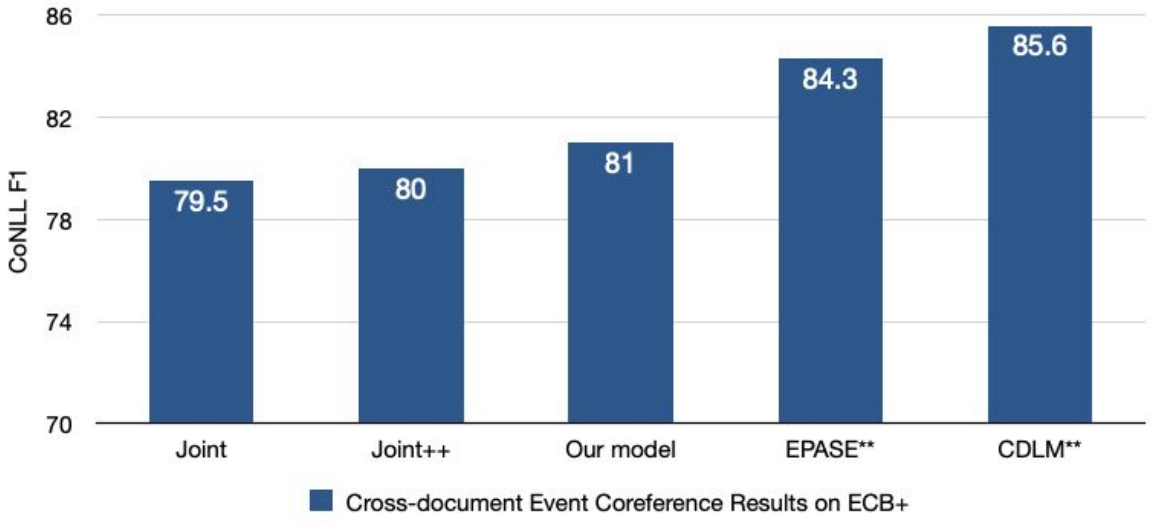
Our model – Extension of e2e-coref

- At inference, we use an agglomerative clustering to obtain the clusters over the overall score $s(i, j)$

Advantages:

- ✅ **Single pass of RoBERTa for each document**
- ✅ **Pairwise scores are calculated only once**
- ✅ **Fast inference on a single 12GB GPU**

Results – Gold mentions on ECB+



**** Apply BERT for all mention pairs**

Joint: Barhom, S., Shwartz, V., Eirew, A., Bugert, M., Reimers, N., & Dagan, I. (2019). Revisiting Joint Modeling of Cross-document Entity and Event Coreference Resolution. *ACL*.
Joint++: Meged, Y., Caciularu, A., Shwartz, V., & Dagan, I. (2020). Paraphrasing vs Coreferring: Two Sides of the Same Coin. *FINDINGS*.
Epase: Zeng, Y., Jin, X., Guan, S., Guo, J., & Cheng, X. (2020). Event Coreference Resolution with their Paraphrases and Argument-aware Embeddings. COLING..
CDLM: Caciularu, A., Cohan, A., Beltagy, I., Peters, M.E., Cattan, A., & Dagan, I. (2021). Cross-Document Language Modeling. ArXiv, abs/2101.00406.

Results

- Our model set first results over predicted mentions (54.4 F1)
  - **Large room for improvement under realistic conditions**

- Refer to the paper for the results on entity coreference

Analysis and Ablations

1. Cross-document coreference is **harder** than within-document coreference

	Gold		Predicted	
	WD	CD	WD	CD
Event	**86.6**	81.0	**59.6**	54.4
Entity	**81.2**	73.1	**39.7**	35.7
ALL	**83.9**	76.7	**46.3**	43.4

Table 2: Results (CoNLL F1) of our model, on within-document (WD) vs. cross-document (CD), using gold and predicted mentions.

Analysis and Ablations

1. Cross-document coreference is **harder** than within-document coreference

	Gold		Predicted	
	WD	CD	WD	CD
Event	**86.6**	81.0	**59.6**	54.4
Entity	**81.2**	73.1	**39.7**	35.7
ALL	**83.9**	76.7	**46.3**	43.4

Table 2: Results (CoNLL F1) of our model, on within-document (WD) vs. cross-document (CD), using gold and predicted mentions.

2. Document clustering significantly **simplifies** the task, especially for event

	Gold	**Δ**	**Predicted**	**Δ**
Event	76.0	**–5.0**	48.2	**–6.2**
Entity	70.9	–2.2	34.4	–1.3
ALL	74.1	–2.6	41.4	–2.0

Table 3: CoNLL F1 results of our model without document clustering, using gold and predicted mentions. Δ shows the relative drop in performance.

Analysis and Ablations

1. Cross-document coreference is **harder** than within-document coreference

	Gold		Predicted	
	WD	CD	WD	CD
Event	**86.6**	81.0	**59.6**	54.4
Entity	**81.2**	73.1	**39.7**	35.7
ALL	**83.9**	76.7	**46.3**	43.4

Table 2: Results (CoNLL F1) of our model, on within-document (WD) vs. cross-document (CD), using gold and predicted mentions.

2. Document clustering significantly **simplifies** the task, especially for event

	Gold	Δ	**Predicted**	Δ
Event	76.0	**–5.0**	48.2	**–6.2**
Entity	70.9	–2.2	34.4	–1.3
ALL	74.1	–2.6	41.4	–2.0

Table 3: CoNLL F1 results of our model without document clustering, using gold and predicted mentions. Δ shows the relative drop in performance.

3. Ablation studies show the **effectiveness** of each component

	F1	Δ
Our model	58.1	
– pre-train of mention scorer	54.9	–3.2
– dynamic pruning	54.1	–4.0
– negative sampling	56.7	–1.4

Table 4: Ablation results (CoNLL F1) of our model on the development set of ECB+ event coreference.

Thanks!

Questions?

Arie Cattan

github.com/ariecattan/coref

arie.cattan.github.io