Efficient and Interpretable Neural Models for Entity Tracking

Shubham Toshniwal

TTI Chicago

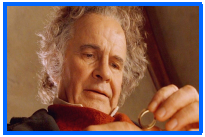
24 Aug 2022

Entities

Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir. Neither hobbit is aware of the Ring's origin, but the wizard Gandalf suspects it is a Ring of Power. Seventeen years later, Gandalf tells Frodo that he has confirmed that the Ring is the one lost by the Dark Lord Sauron long ago and counsels him to take it away from the Shire.

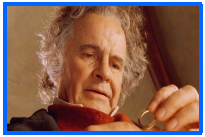
Entities

Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir. Neither hobbit is aware of the Ring's origin, but the wizard Gandalf suspects it is a Ring of Power. Seventeen years later, Gandalf tells Frodo that he has confirmed that the Ring is the one lost by the Dark Lord Sauron long ago and counsels him to take it away from the Shire.




Entities

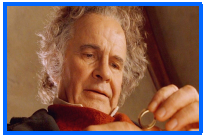
Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir. Neither hobbit is aware of the Ring's origin, but the wizard Gandalf suspects it is a Ring of Power. Seventeen years later, Gandalf tells Frodo that he has confirmed that the Ring is the one lost by the Dark Lord Sauron long ago and counsels him to take it away from the Shire.





Entities

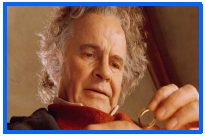
Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir. Neither hobbit is aware of the Ring's origin, but the wizard Gandalf suspects it is a Ring of Power. Seventeen years later, Gandalf tells Frodo that he has confirmed that the Ring is the one lost by the Dark Lord Sauron long ago and counsels him to take it away from the Shire.






Multiplicity in References

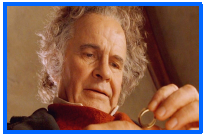
Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing the Ring to [Frodo Baggins], [his cousin and heir]. Neither hobbit is aware of the Ring's origin, but the wizard Gandalf suspects it is a Ring of Power. Seventeen years later, Gandalf tells [Frodo] that he has confirmed that the Ring is the one lost by the Dark Lord Sauron long ago and counsels [him] to take it away from the Shire.






Multiplicity in References

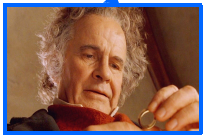
Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing [the Ring] to Frodo Baggins, his cousin and heir. Neither hobbit is aware of [the Ring's] origin, but the wizard Gandalf suspects [it] is a Ring of Power. Seventeen years later, Gandalf tells Frodo that he has confirmed that [the Ring] is the one lost by the Dark Lord Sauron long ago and counsels him to take [it] away from the Shire.






Ambiguity in References

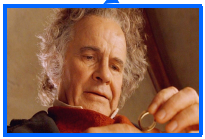
Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing the Ring to Frodo Baggins, [his] cousin and heir. Neither hobbit is aware of the Ring's origin, but the wizard Gandalf suspects it is a Ring of Power. Seventeen years later, Gandalf tells Frodo that [he] has confirmed that the Ring is the one lost by the Dark Lord Sauron long ago and counsels [him] to take it away from the Shire.






Evolving Attributes: Changing Ownership of The Ring

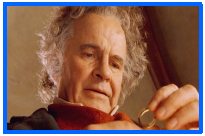
Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir. Neither hobbit is aware of the Ring's origin, but the wizard Gandalf suspects it is a Ring of Power. Seventeen years later, Gandalf tells Frodo that he has confirmed that the Ring is the one lost by the Dark Lord Sauron long ago and counsels him to take it away from the Shire.






Evolving Attributes: Changing Ownership of The Ring

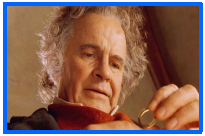
Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir. Neither hobbit is aware of the Ring's origin, but the wizard Gandalf suspects it is a Ring of Power. Seventeen years later, Gandalf tells Frodo that he has confirmed that the Ring is the one lost by the Dark Lord Sauron long ago and counsels him to take it away from the Shire.






Evolving Attributes: Changing Ownership of The Ring

Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir. Neither hobbit is aware of the Ring's origin, but the wizard Gandalf suspects it is a Ring of Power. Seventeen years later, Gandalf tells Frodo that he has confirmed that the Ring is the one lost by the Dark Lord Sauron long ago and counsels him to take it away from the Shire.






Historical Perspective (Kartunnen 1969)

Lauri Karttunen
University of Texas at Austin
Department of Linguistics
Austin, Texas 78712

DISCOURSE REFERENTS

Consider an interpretive device that in some manner keeps track of individuals that have been mentioned in a discourse and what has been said about them. One feature any such device must have is to be able to recognize when a novel individual appears in some sentence. For example, in processing sentence (1), it must recognize that the NP a car refers to some yet unmentioned object, which in the following sentence may be referred to again by any of the alternative ways in (2).

(1) I have a car. (2) (a) It is black. (b) The car is black.

Consider an interpretive device that in some manner keeps track of individuals that have been mentioned in a discourse and what has been said about them. One feature any such device must have is to be able to recognize when a novel individual appears in some sentence.

Entity Tracking Task

Identify and record new entities and their attributes as they are introduced

Identify subsequent references to the entities previously introduced and update their attributes

Useful in downstream NLP applications such as question answering, summarization, story generation.

Entity Tracking Tasks: Coreference Resolution

```
Bilbo celebrates his eleventy-first birthday and leaves the Shire
suddenly, passing the Ring to Frodo Baggins, his cousin and heir.
Neither hobbit is aware of the Ring's origin, but the wizard
Gandalf suspects it is a Ring of Power.  Seventeen years later,
Gandalf tells Frodo that he has confirmed that the Ring is the
one lost by the Dark Lord Sauron long ago and counsels him to
take it away from the Shire.
```

Entity Tracking Tasks: Coreference Resolution

[Bilbo]$_1$ celebrates [his]$_1$ eleventy-first birthday and leaves [the Shire]$_2$ suddenly, passing [the Ring]$_3$ to [Frodo Baggins]$_4$, [[his]$_1$ cousin and heir]$_4$. Neither hobbit is aware of [the Ring's]$_3$ origin, but [the wizard Gandalf]$_5$ suspects [it]$_3$ is a Ring of Power. Seventeen years later, [Gandalf]$_5$ tells [Frodo]$_4$ that [he]$_5$ has confirmed that [the Ring]$_3$ is the one lost by [the Dark Lord Sauron]$_6$ long ago and counsels [him]$_4$ to take [it]$_3$ away from [the Shire]$_2$.

Entity Tracking Tasks: State Tracking

Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir. Neither hobbit is aware of the Ring's origin, but the wizard Gandalf suspects it is a Ring of Power. Seventeen years later, Gandalf tells Frodo that he has confirmed that the Ring is the one lost by the Dark Lord Sauron long ago and counsels him to take it away from the Shire.

Who owns the ring at the end of the story?

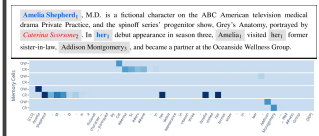
Entity Tracking Tasks: State Tracking

Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir. Neither hobbit is aware of the Ring's origin, but the wizard Gandalf suspects it is a Ring of Power. Seventeen years later, Gandalf tells Frodo that he has confirmed that the Ring is the one lost by the Dark Lord Sauron long ago and counsels him to take it away from the Shire.

Who owned the ring at the start of the story?

Overview

Explicit Entity Tracking Models



Sparsely Supervised Pronoun Resolution

Toshniwal et al
ACL 2020

Long Document Coreference with Bounded Memory

Toshniwal et al
EMNLP 2020

OntoNotes [The Federal Reserve]$_1$ considers interest rates ... On [Tuesday]$_3$, the Federal Reserve Open Market Committee meets for the final time this year to discuss interest rates

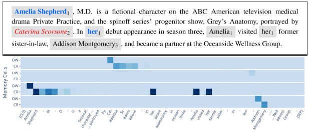
LitBank Down the [Rabbit-Hole]$_1$ [Alice]$_2$ was beginning to get very tired of sitting by [[her]$_2$ sister]$_3$ on [the bank]$_4$, and of having nothing to do: once or twice [she]$_2$ had peeped into ...

Generalization in Coreference Resolution

Toshniwal et al.
CRAC (EMNLP) 2021
Best Short Paper

Overview

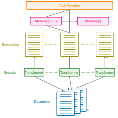
Explicit Entity Tracking Models



Amelia Shepherd$_1$, M.D. is a fictional character on the ABC American television medical drama Private Practice, and the spinoff series' progenitor show, Grey's Anatomy, portrayed by *Caterina Scorsone*$_2$. In **her**$_1$ debut appearance in season three, Amelia$_1$ visited her$_1$ former sister-in-law, Addison Montgomery$_3$, and became a partner at the Oceanside Wellness Group.

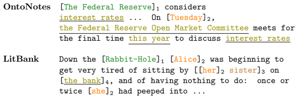
Sparsely Supervised Pronoun Resolution

Toshniwal et al
ACL 2020



Long Document Coreference with Bounded Memory

Toshniwal et al
EMNLP 2020

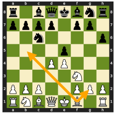


OntoNotes	[The Federal Reserve]$_1$ considers interest rates ... On [Tuesday]$_2$, the Federal Reserve Open Market Committee meets for the final time this year to discuss interest rates
LitBank	Down the [Rabbit-Hole]$_1$ [Alice]$_2$ was beginning to get very tired of sitting by [[her]$_2$ sister]$_3$ on [the bank]$_4$, and of having nothing to do: once or twice [she]$_2$ had peeped into ...

Generalization in Coreference Resolution

Toshniwal et al.
CRAC (EMNLP) 2021
Best Short Paper

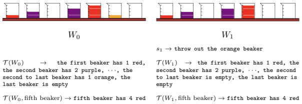
Implicit Entity Tracking via Language Models



e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 f1??

Chess as a Testbed for Language Model State Tracking

Toshniwal et al
AAAI 2022



W_0

$\mathcal{T}(W_0)$ → the first beaker has 1 red, the second beaker has 2 purple, ⋯, the second to last beaker has 1 orange, the last beaker is empty

$\mathcal{T}(W_0, \text{fifth beaker})$ → fifth beaker has 4 red

W_1

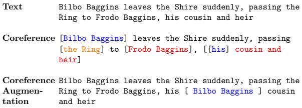
s_1 → throw out the orange beaker

$\mathcal{T}(W_1)$ → the first beaker has 1 red, the second beaker has 2 purple, ⋯, the second to last beaker is empty, the last beaker is empty

$\mathcal{T}(W_1, \text{fifth beaker})$ → fifth beaker has 4 red

Baked-in State Probing

Toshniwal et al
Under Review



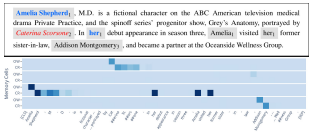
Text	Bilbo Baggins leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir
Coreference	[Bilbo Baggins] leaves the Shire suddenly, passing [the Ring] to [Frodo Baggins], [[his] cousin and heir]
Coreference Augmentation	Bilbo Baggins leaves the Shire suddenly, passing the Ring to Frodo Baggins, his [Bilbo Baggins] cousin and heir

Baking in Coreference Knowledge into Language Models

In Preparation

Overview

Explicit Entity Tracking Models



Sparsely Supervised Pronoun Resolution

Toshniwal et al
ACL 2020

Long Document Coreference with Bounded Memory

Toshniwal et al
EMNLP 2020

OntoNotes [The Federal Reserve]$_1$ considers interest rates ... On [Tuesday]$_2$, the Federal Reserve Open Market Committee meets for the final time this year to discuss interest rates

LitBank Down the [Rabbit-Hole], [Alice]$_2$ was beginning to get very tired of sitting by [[her]$_2$ sister]$_3$ on [the bank]$_4$, and of having nothing to do: once or twice [she]$_2$ had peeped into ...

Generalization in Coreference Resolution

Toshniwal et al.
CRAC (EMNLP) 2021
Best Short Paper

Implicit Entity Tracking via Language Models

Chess as a Testbed for Language Model State Tracking

Toshniwal et al
AAAI 2022

W_0

$T(W_0)$ → the first beaker has 1 red, the second beaker has 2 purple, ···, the second to last beaker has 1 orange, the last beaker is empty

$T(W_0, \text{fifth beaker}) \rightarrow$ fifth beaker has 4 red

W_1

$s_1 \rightarrow$ throw out the orange beaker

$T(W_1)$ → the first beaker has 1 red, the second beaker has 2 purple, ···, the second to last beaker is empty, the last beaker is empty

$T(W_1, \text{fifth beaker}) \rightarrow$ fifth beaker has 4 red

Baked-in State Probing

Toshniwal et al
Under Review

Text Bilbo Baggins leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir

Coreference [Bilbo Baggins] leaves the Shire suddenly, passing [the Ring] to [Frodo Baggins], [[his] cousin and heir]

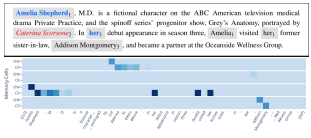
Coreference Augmentation Bilbo Baggins leaves the Shire suddenly, passing the Ring to Frodo Baggins, his [Bilbo Baggins] cousin and heir

Baking in Coreference Knowledge into Language Models

In Preparation

Overview

Explicit Entity Tracking Models



Sparsely Supervised Pronoun Resolution

Toshniwal et al
ACL 2020

Long Document Coreference with Bounded Memory

Toshniwal et al
EMNLP 2020

OntoNotes [The Federal Reserve]$_1$ considers interest rates ... On [Tuesday]$_2$, the Federal Reserve Open Market Committee meets for the final time this year to discuss interest rates

LitBank Down the [Rabbit-Hole]$_1$ [Alice]$_2$ was beginning to get very tired of sitting by [[her]$_2$ sister]$_3$ on [the bank]$_4$, and of having nothing to do: once or twice [she]$_2$ had peeped into ...

Generalization in Coreference Resolution

Toshniwal et al.
CRAC (EMNLP) 2021
Best Short Paper

Implicit Entity Tracking via Language Models

e2e4 e7e5 g1f3 b8c6 d2d4 b7b5 f1f7

Chess as a Testbed for Language Model State Tracking

Toshniwal et al
AAAI 2022

W_0

$\mathcal{T}(W_0)$ → the first beaker has 1 red, the second beaker has 2 purple, ···, the second to last beaker has 1 orange, the last beaker is empty

$\mathcal{T}(W_0, \text{fifth beaker})$ → fifth beaker has 4 red

W_1

$s1$ → throw out the orange beaker

$\mathcal{T}(W_1)$ → the first beaker has 1 red, the second beaker has 2 purple, ···, the second to last beaker is empty, the last beaker is empty

$\mathcal{T}(W_1, \text{fifth beaker})$ → fifth beaker has 4 red

Baked-in State Probing

Toshniwal et al
Under Review

Text Bilbo Baggins leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir

Coreference [Bilbo Baggins] leaves the Shire suddenly, passing [the Ring] to [Frodo Baggins], [[his] cousin and heir]

Coreference Augmentation Bilbo Baggins leaves the Shire suddenly, passing the Ring to Frodo Baggins, his [Bilbo Baggins] cousin and heir

Baking in Coreference Knowledge into Language Models

In Preparation

Roadmap

Roadmap

Explicit Entity Tracking

Coreference Resolution Models for Long Documents

Generalization in Coreference Resolution

Implicit Entity Tracking with Language Models

Chess as a Testbed for Entity Tracking

Baking in Coreference Knowledge into Language Models

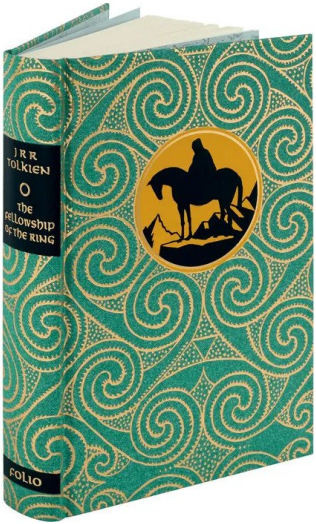
Conclusion

Coreference Resolution

[Bilbo]$_1$ celebrates [his]$_1$ eleventy-first birthday and leaves [the Shire]$_2$ suddenly, passing [the Ring]$_3$ to [Frodo Baggins]$_4$, [[his]$_1$ cousin and heir]$_4$. Neither hobbit is aware of [the Ring's]$_3$ origin, but [the wizard Gandalf]$_5$ suspects [it]$_3$ is a Ring of Power. Seventeen years later, [Gandalf]$_5$ tells [Frodo]$_4$ that [he]$_5$ has confirmed that [the Ring]$_3$ is the one lost by [the Dark Lord Sauron]$_6$ long ago and counsels [him]$_4$ to take [it]$_3$ away from [the Shire]$_2$.

Why Coreference Resolution for Long Documents?

Understanding long narratives requires keeping track of characters introduced

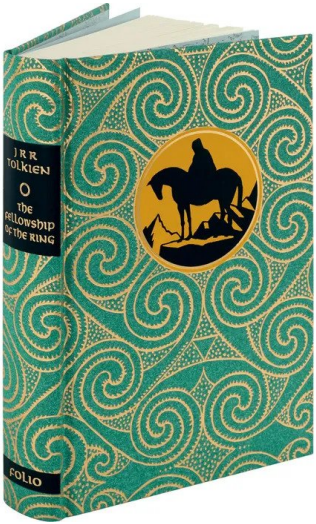


Why Coreference Resolution for Long Documents?

Understanding long narratives requires keeping track of characters introduced

Long documents such as book-length tasks have garnered interest recently

- Computational challenges
- Lack of big annotated datasets



Scaling Issues with Current Models

Given: Input document $\mathcal{D}$ and $\mathcal{T} = |\mathcal{D}|$

Lee et al (2017); Joshi et al (2020); Xu and Choi (2020)

Runtime complexity $\mathcal{O}(\mathcal{T}^2)$ (without heuristics)

Inference can require more than 12GB memory for $\mathcal{T} \leq 4K$ (Xia et al (2020); Kirstain et al (2021))

Wu et al. 2020: Current state-of-the-art; requires $\mathcal{O}(\mathcal{T})$ passes over a document

Scaling Issues with Current Models

Given: Input document $\mathcal{D}$ and $\mathcal{T} = |\mathcal{D}|$

Lee et al (2017); Joshi et al (2020); Xu and Choi (2020)

Runtime complexity $\mathcal{O}(\mathcal{T}^2)$ (without heuristics)

Inference can require more than 12GB memory for $\mathcal{T} \leq 4K$ (Xia et al (2020); Kirstain et al (2021))

Wu et al. 2020: Current state-of-the-art; requires $\mathcal{O}(\mathcal{T})$ passes over a document

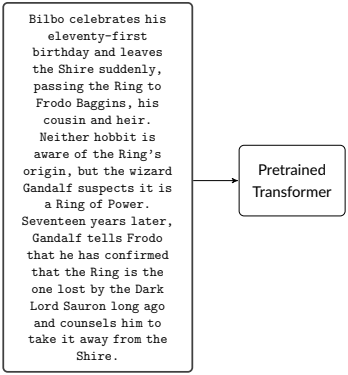
High performance models which can scale to book-length documents

End-to-End Coreference Resolution (Lee et al. 2017; Joshi et al. 2020)

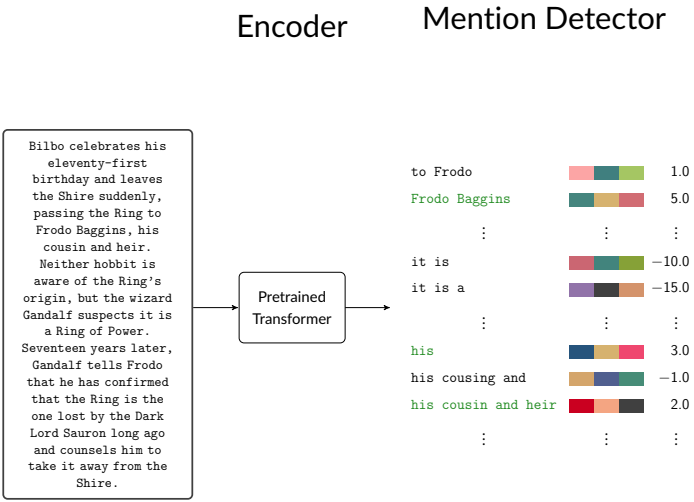
Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir. Neither hobbit is aware of the Ring's origin, but the wizard Gandalf suspects it is a Ring of Power. Seventeen years later, Gandalf tells Frodo that he has confirmed that the Ring is the one lost by the Dark Lord Sauron long ago and counsels him to take it away from the Shire.

End-to-End Coreference Resolution (Lee et al. 2017; Joshi et al. 2020)

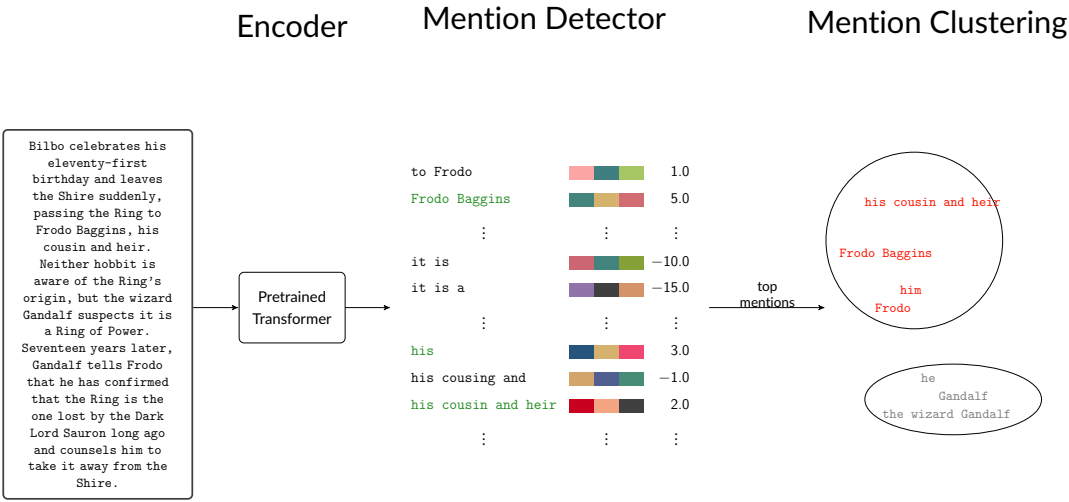
Encoder



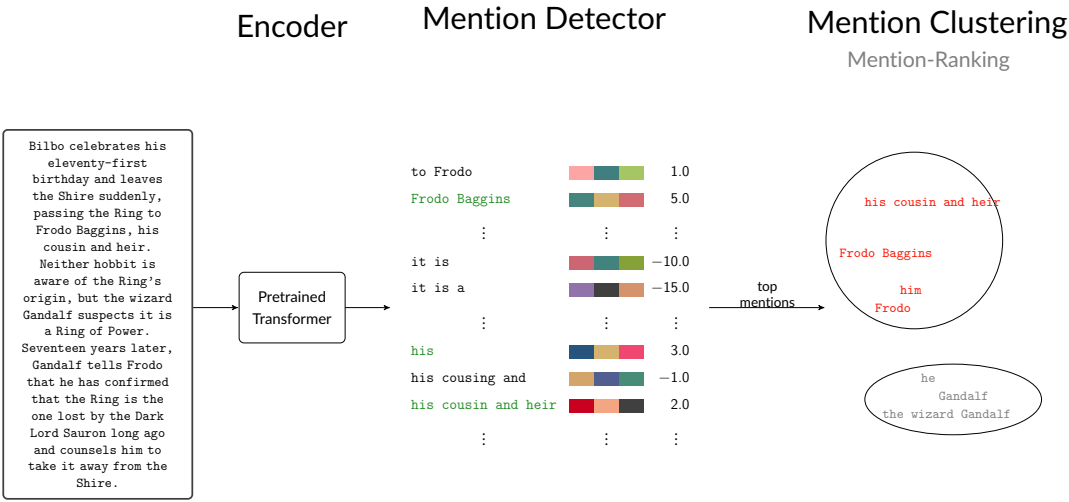
End-to-End Coreference Resolution (Lee et al. 2017; Joshi et al. 2020)



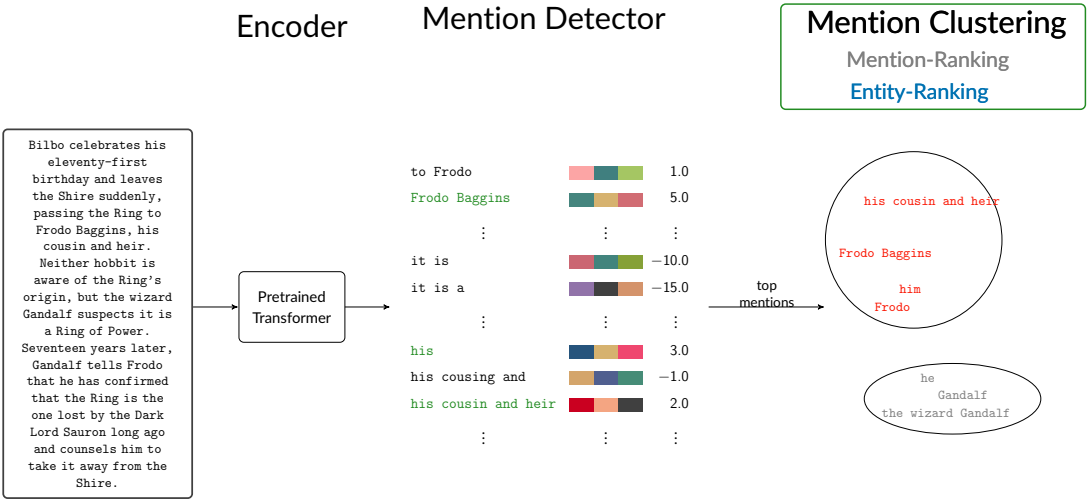
End-to-End Coreference Resolution (Lee et al. 2017; Joshi et al. 2020)



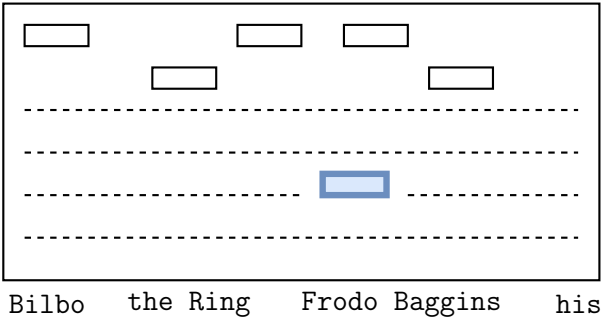
End-to-End Coreference Resolution (Lee et al. 2017; Joshi et al. 2020)



End-to-End Coreference Resolution (Lee et al. 2017; Joshi et al. 2020)

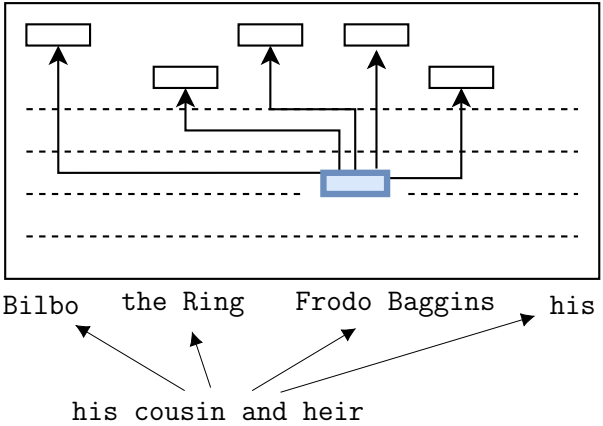


Mention-Ranking Models (Lee et al. 2017, Joshi et al. 2020)

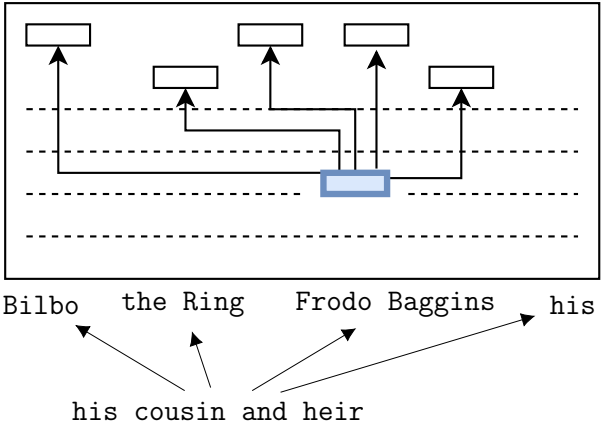


his cousin and heir

Mention-Ranking Models (Lee et al. 2017, Joshi et al. 2020)



Mention-Ranking Models (Lee et al. 2017, Joshi et al. 2020)



Impractical for long documents!
Quadratic runtime!

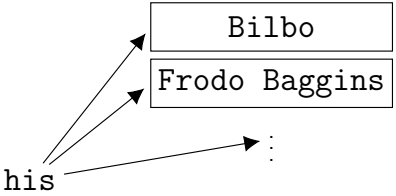
Entity-Ranking Models

Bilbo
Frodo Baggins

⋮

Entities

Entity-Ranking Models



Entities

Use simple average of mention representations

Evaluation Setup

Datasets:

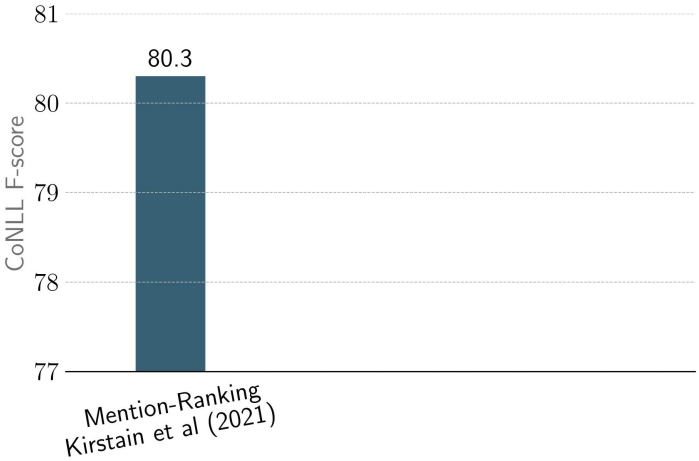
- OntoNotes: Short News Text (460 words)
- LitBank: Long Literary Text (2100 words)

Evaluation Metric: CoNLL F-score

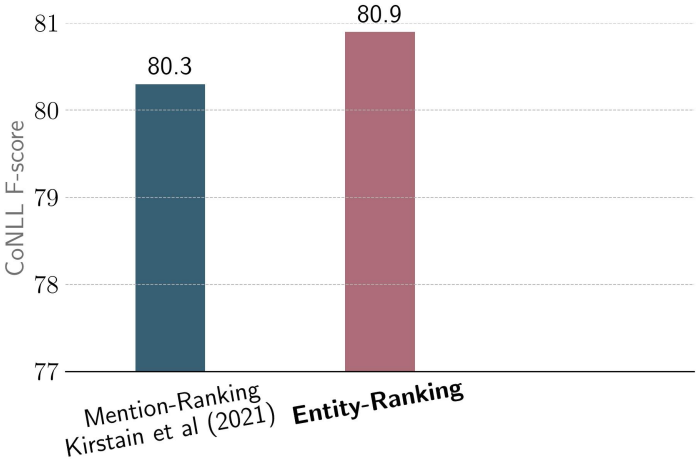
Baselines: Mention-Ranking models

- OntoNotes: Kirstain et al. (2021)
- LitBank: Xu and Choi (2020)

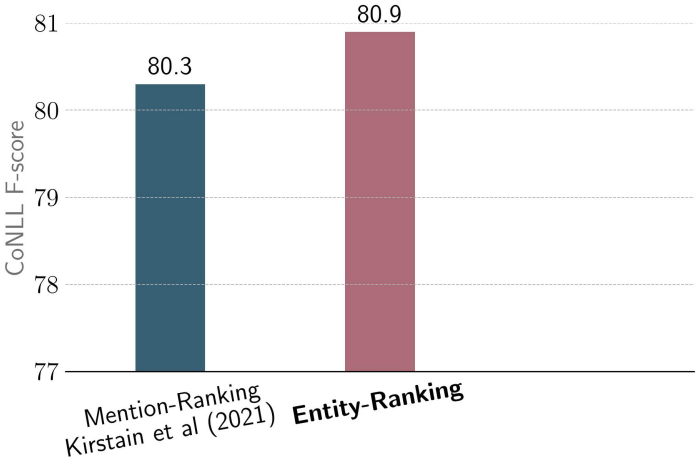
Results for Short News Text



Results for Short News Text

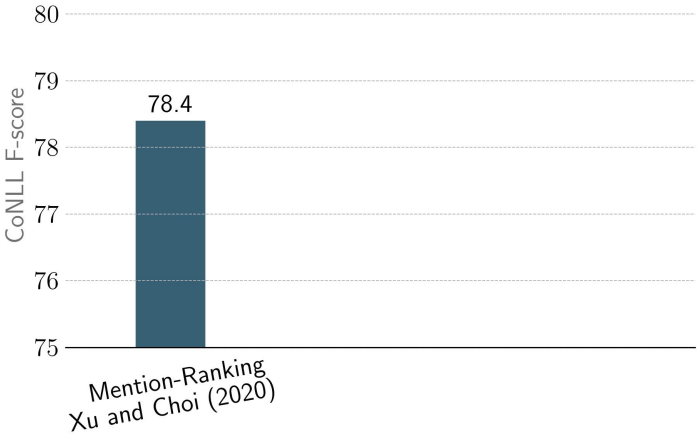


Results for Short News Text

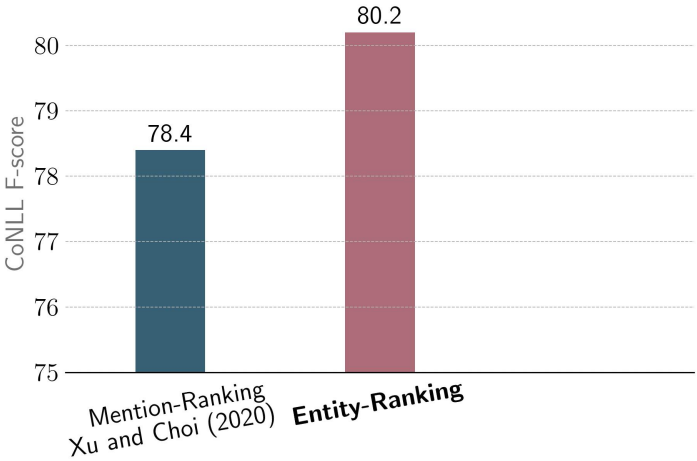


CAN cram all the mentions into one vector :)

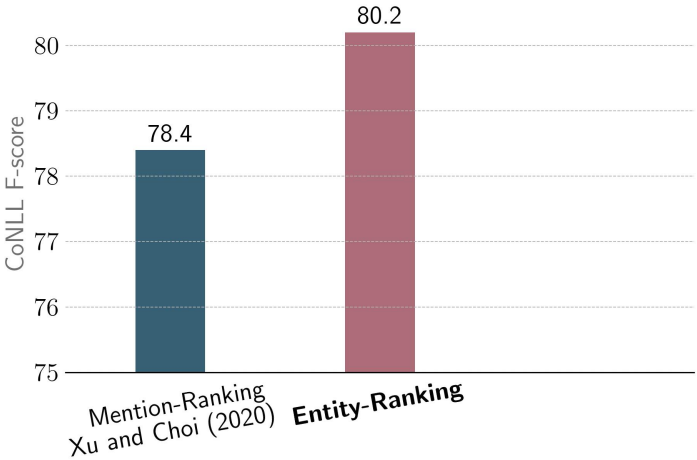
Results for Long Literary Text



Results for Long Literary Text

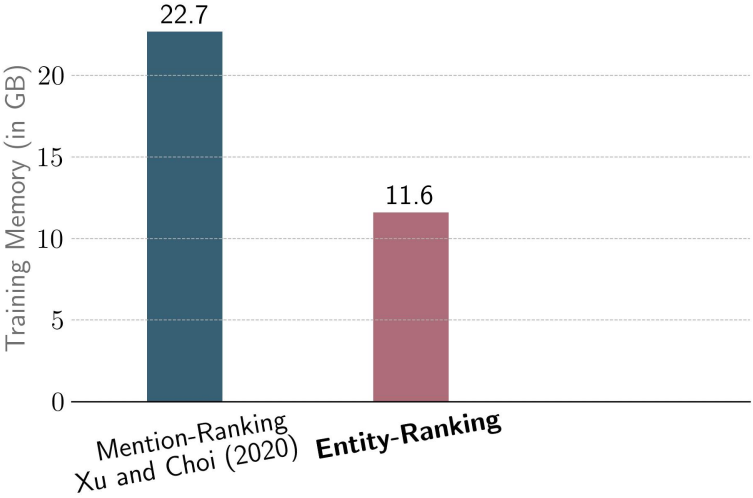


Results for Long Literary Text



State-of-the-art for LitBank!

Memory Comparison for Long Literary Text



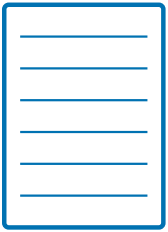
Roughly half the memory!

Can we further reduce memory?

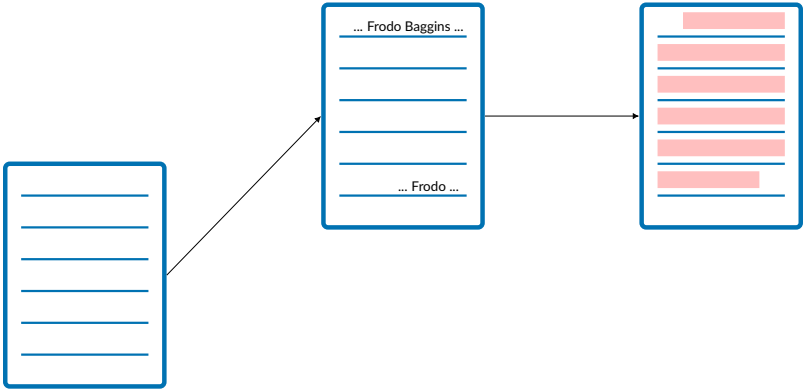
Can we further reduce memory?

Do we need to keep all the entities in the memory?

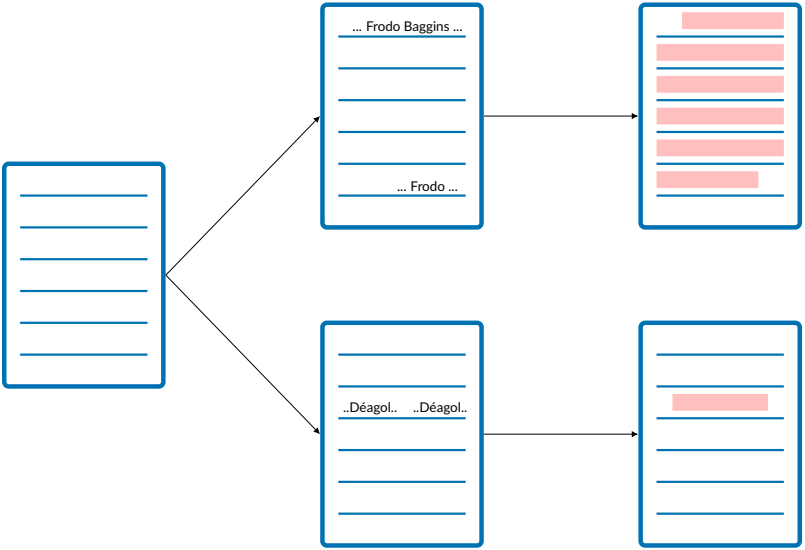
Entity Spread: A Tale of Two Characters



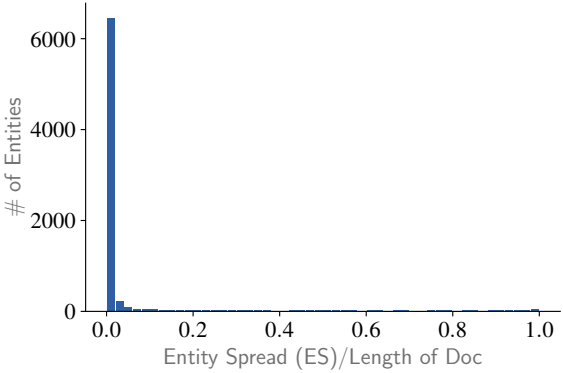
Entity Spread: A Tale of Two Characters



Entity Spread: A Tale of Two Characters



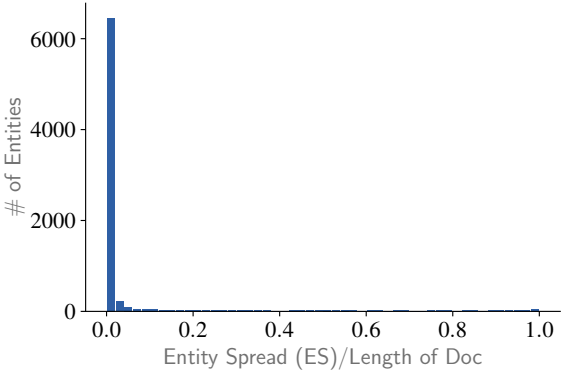
Most Entities Are Transient



LitBank Entity Spread Histogram

Most entities have a small *spread*

Most Entities Are Transient



LitBank Entity Spread Histogram

Most entities have a small *spread*

Not necessary to keep all entities in memory all the time!

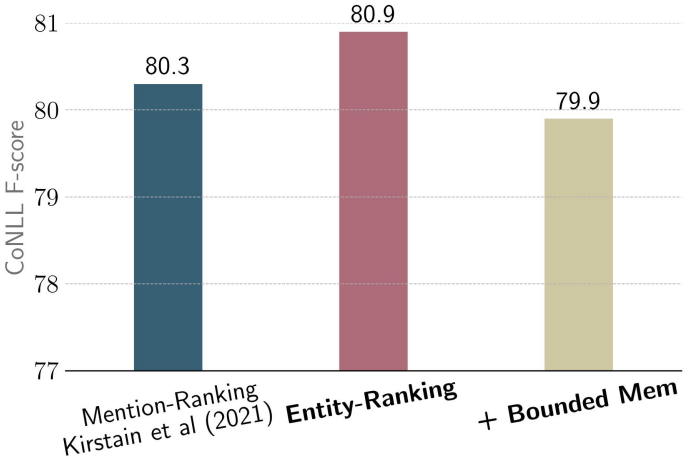
Bounded Memory Model: Ignore and Evict

Track a small, bounded number of entities

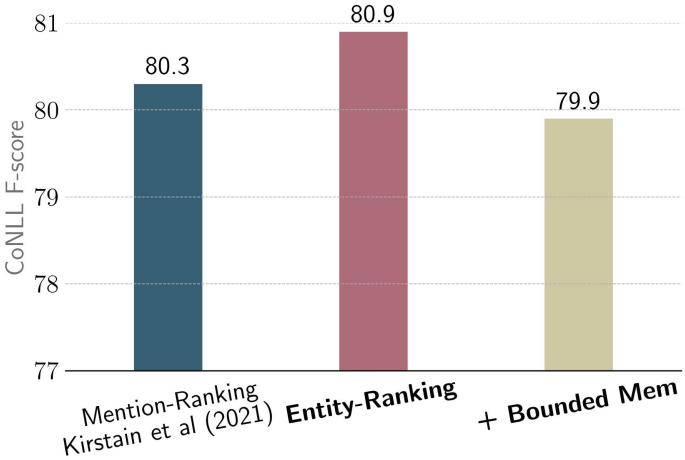
When memory is full, and a mention corresponding to a new entity comes next, then:

- Evict: Remove an entity already being tracked, and start tracking this new entity
- Ignore: Ignore the mention

Results with Bounded Memory for Short News Text

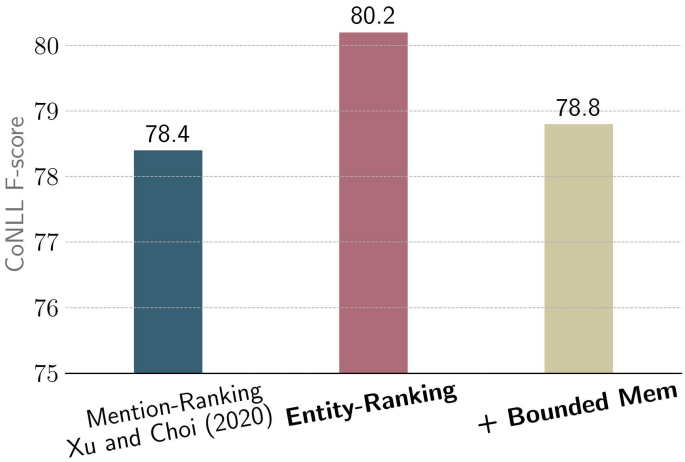


Results with Bounded Memory for Short News Text

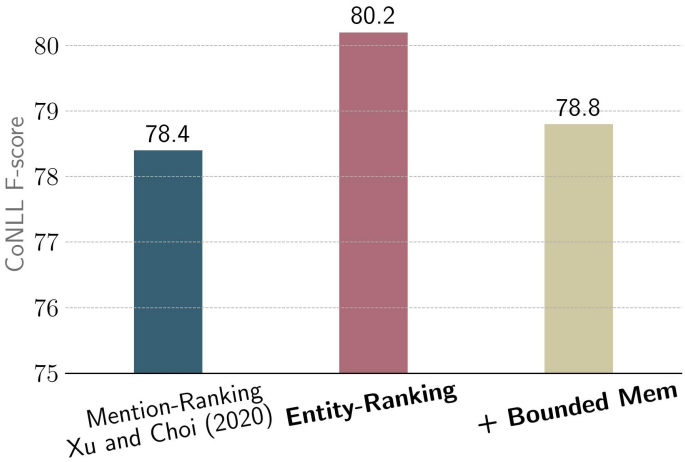


Bounded Memory model keeps **10x** less entities in memory

Results with Bounded Memory for Long Literary Text

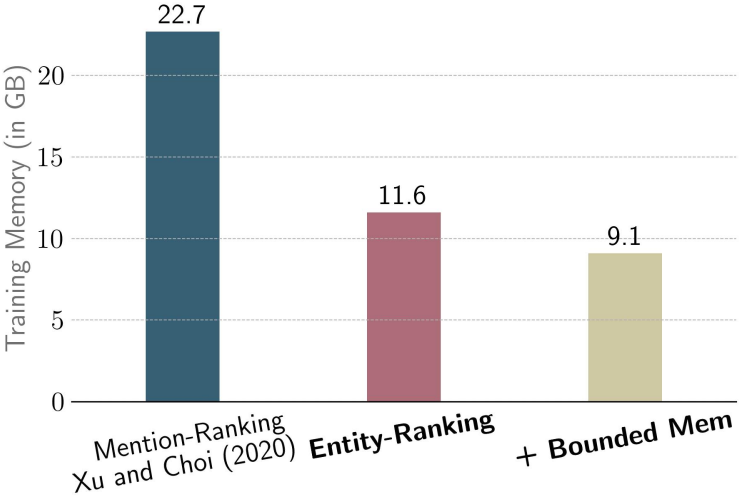


Results with Bounded Memory for Long Literary Text



Slightly bigger drop in performance for longer documents

Memory Comparison for Long Literary Text



25% peak training memory reduction

Takeaways

Proposed approaches to make coreference models more efficient

Proposed models are competitive with prior work and reduce peak training memory

Establish a new state-of-the-art for LitBank

Roadmap

Generalization Capability of Coreference Resolution Models

Prior work has shown coreference models generalize poorly to out-of-domain evaluations

The two big challenges are:

Domain Shift

Annotation Differences

Generalization Capability of Coreference Resolution Models

Prior work has shown coreference models generalize poorly to out-of-domain evaluations

The two big challenges are:

Domain Shift: Joint Training

Annotation Differences: Data Augmentation

Generalization Capability of Coreference Resolution Models

Prior work has shown coreference models generalize poorly to out-of-domain evaluations

The two big challenges are:

Domain Shift: Joint Training

Annotation Differences: Data Augmentation

Proposed models establish state-of-the-art for two more coreference benchmarks: PreCo and WikiCoref

Roadmap

Explicit Entity Tracking

Coreference Resolution Models for Long Documents

Generalization in Coreference Resolution

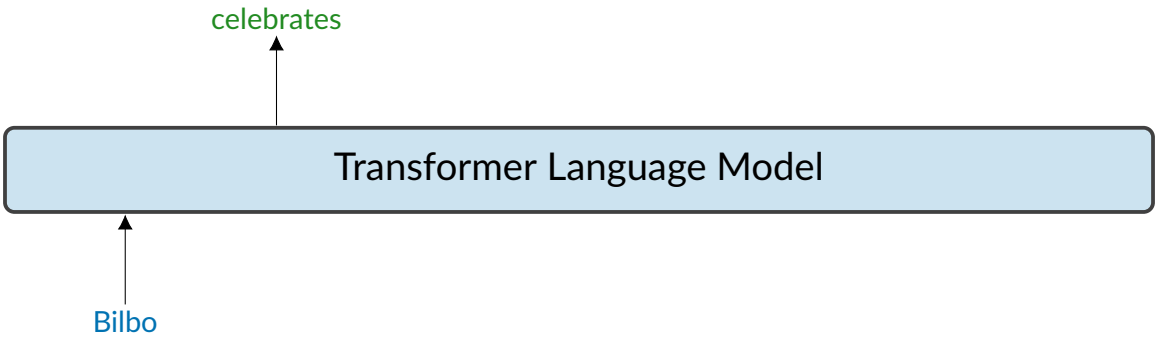
Implicit Entity Tracking with Language Models

Chess as a Testbed for Entity Tracking

Baking in Coreference Knowledge into Language Models

Conclusion

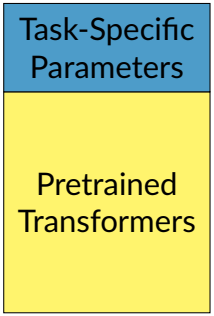
Transformer Language Models

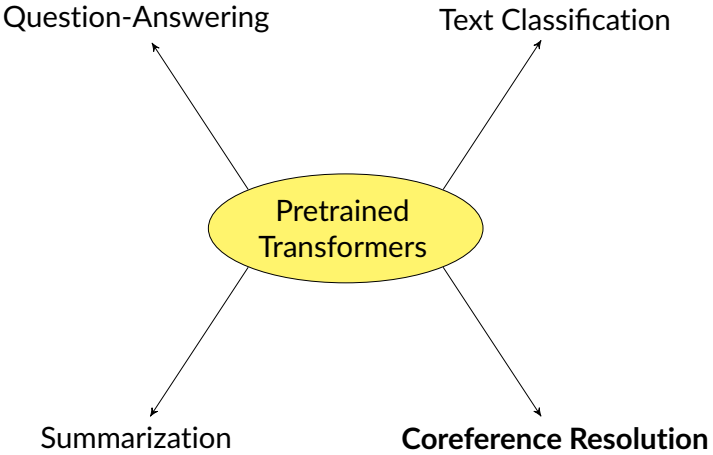


Popular models: GPT-2, GPT-3

Other Training Objectives and Architecture: BERT, BART

Modern NLP Pipeline





Language Models capture Linguistic Knowledge (Tenney et al (2019))

Is the Language Model doing Entity Tracking?

Transformer Language Model

Bilbo celebrates his eleventy-first birthday and leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir. Neither hobbit is aware of the Ring's origin, but the wizard Gandalf suspects it is a Ring of Power. Seventeen years later, Gandalf tells Frodo that he has confirmed that the Ring is the one lost by the Dark Lord Sauron long ago and counsels him to take it away from the Shire.

Is the Language Model doing Entity Tracking?

Probing analysis by Sorodoc et al. 2020 suggests that pretrained transformer LMs lack a global notion of entity

```
Bilbo celebrates his eleventy-first birthday and leaves the Shire
suddenly, passing the Ring to Frodo Baggins, his cousin and heir.
Neither hobbit is aware of the Ring's origin, but the wizard
Gandalf suspects it is a Ring of Power.  Seventeen years later,
Gandalf tells Frodo that he has confirmed that the Ring is the
one lost by the Dark Lord Sauron long ago and counsels him to
take it away from the Shire.
```

Representation of him might be more similar to Sauron than Frodo

Evaluation Challenges: Prompting GPT-3

Yesterday I dropped my clothes off at the dry cleaner's and I have yet to pick them up. Where are my clothes?

Prompt used by Gary Marcus and Ernest Davis to diagnose the entity tracking capability of GPT-3

Evaluation Challenges: Prompting GPT-3

Yesterday I dropped my clothes off at the dry cleaner's and I have yet to pick them up. Where are my clothes? **I have a lot of clothes.**

Prompt used by Gary Marcus and Ernest Davis to diagnose the entity tracking capability of GPT-3

Evaluation Challenges: Prompting GPT-3

Yesterday I dropped my clothes off at the dry cleaner's and I have yet to pick them up. Where are my clothes? **I have a lot of clothes.**

Prompt used by Gary Marcus and Ernest Davis to diagnose the entity tracking capability of GPT-3

Lack of Control over Model's Output

Integrating Entity Tracking into Language Models

Benefits for Entity Tracking:

- Wider application of Entity Tracking
- Easier adoption of Entity Tracking

Integrating Entity Tracking into Language Models

Benefits for Entity Tracking:

- Wider application of Entity Tracking
- Easier adoption of Entity Tracking

Design Goals:

- Preserving the Language Model Architecture
- Interpretability w.r.t. Entity Tracking

Integrating Entity Tracking into Language Models

Benefits for Entity Tracking:

Wider application of Entity Tracking

Easier adoption of Entity Tracking

Design Goals:

Preserving the Language Model Architecture

Interpretability w.r.t. Entity Tracking

Recipe: Train Language Models on Entity State Augmented Text

Roadmap

Explicit Entity Tracking

Coreference Resolution Models for Long Documents

Generalization in Coreference Resolution

Implicit Entity Tracking with Language Models

Chess as a Testbed for Entity Tracking

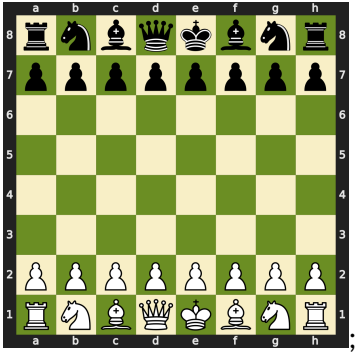
Baking in Coreference Knowledge into Language Models

Conclusion

Entity Tracking in Chess

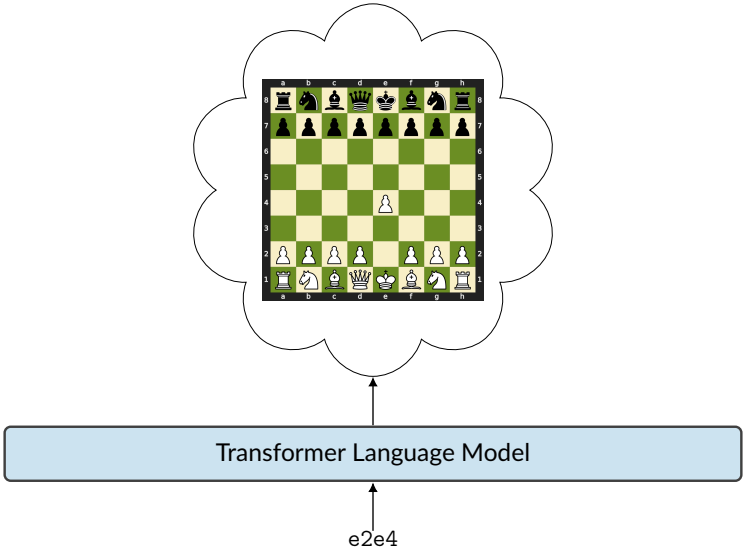
Test out ideas for entity tracking via language models in chess

Why Chess? *Simple, closed domain*



Entities: Chess pieces
Entity State: Piece Location

Learning Chess Blindfolded



Learning Chess Blindfolded

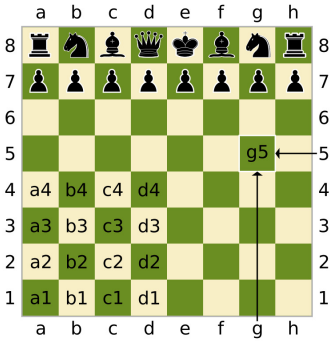
g1f3 d7d5 g2g3 ....

d2d4 d7d5 g1f3 ....

e2e4 e7e5 g1f3 ....

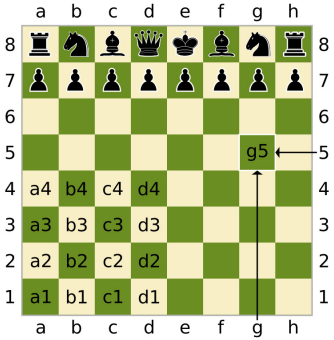
Algebraic Notation

Position Naming



Algebraic Notation

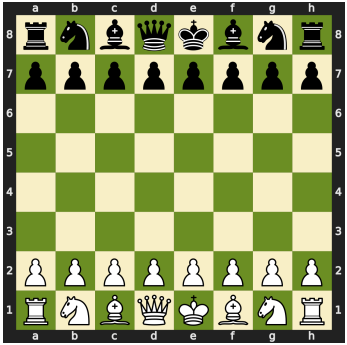
Position Naming



Piece Types

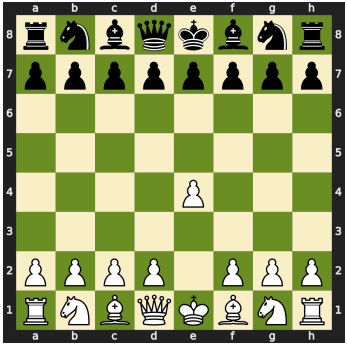
Rook	Knight	Bishop	Queen	King	Pawn
R	N	B	Q	K	P

Chess Notation



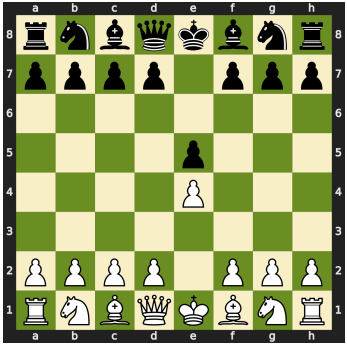
Translation of moves

Chess Notation



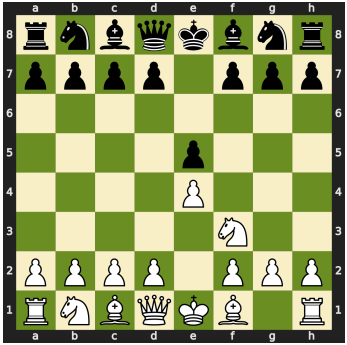
Translation of moves	
e2e4	(Pawn) moved from e2 to e4

Chess Notation



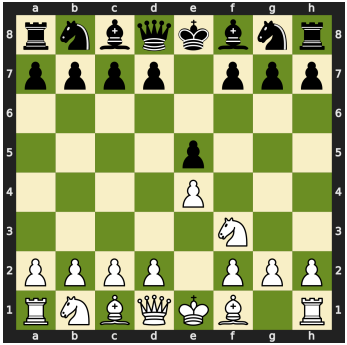
Translation of moves	
e2e4	(Pawn) moved from e2 to e4
e7e5	(Pawn) moved from e7 to e5

Chess Notation



Translation of moves	
`e2e4`	(Pawn) moved from `e2` to `e4`
`e7e5`	(Pawn) moved from `e7` to `e5`
`g1f3`	(Knight) moved from `g1` to `f3`

Chess Notation



Translation of moves	
`e2e4`	(Pawn) moved from `e2` to `e4`
`e7e5`	(Pawn) moved from `e7` to `e5`
`g1f3`	(Knight) moved from `g1` to `f3`
⋮	⋮

Randomly Annotated Piece Type (RAP)

Can a language model benefit from the knowledge of piece types?

Randomly Annotated Piece Type (RAP)

Can a language model benefit from the knowledge of piece types?

Randomly introduce piece types in text sequences during training

Vanilla Training	`e2e4 e7e5 g1f3 b8c6 d2d4 h7h6`

Randomly Annotated Piece Type (RAP)

Can a language model benefit from the knowledge of piece types?

Randomly introduce piece types in text sequences during training

Vanilla Training	`e2e4 e7e5 g1f3 b8c6 d2d4 h7h6`
+ RAP (p=15)	`e2e4 e7e5 Ng1f3 b8c6 d2d4 h7h6`

Piece Types	 Knight N	 Pawn P

Randomly Annotated Piece Type (RAP)

Can a language model benefit from the knowledge of piece types?

Randomly introduce piece types in text sequences during training

Vanilla Training	`e2e4 e7e5 g1f3 b8c6 d2d4 h7h6`
+ RAP (p=15)	`e2e4 e7e5 Ng1f3 b8c6 d2d4 h7h6`
+ RAP (p=50)	`Pe2e4 e7e5 Ng1f3 b8c6 d2d4 Ph7h6`
+ RAP (p=100)	`Pe2e4 Pe7e5 Ng1f3 Nb8c6 Pd2d4 Ph7h6`

Randomly Annotated Piece Type (RAP)

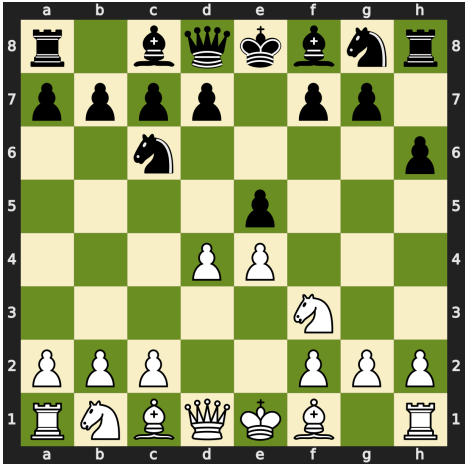
Can a language model benefit from the knowledge of piece types?

Randomly introduce piece types in text sequences during training

Vanilla Training	`e2e4 e7e5 g1f3 b8c6 d2d4 h7h6`
+ RAP (p=15)	`e2e4 e7e5 Ng1f3 b8c6 d2d4 h7h6`
+ RAP (p=50)	`Pe2e4 e7e5 Ng1f3 b8c6 d2d4 Ph7h6`
+ RAP (p=100)	`Pe2e4 Pe7e5 Ng1f3 Nb8c6 Pd2d4 Ph7h6`
Inference	`e2e4 e7e5 g1f3 b8c6 d2d4 h7h6`

Entity Tracking Task: Ending Square

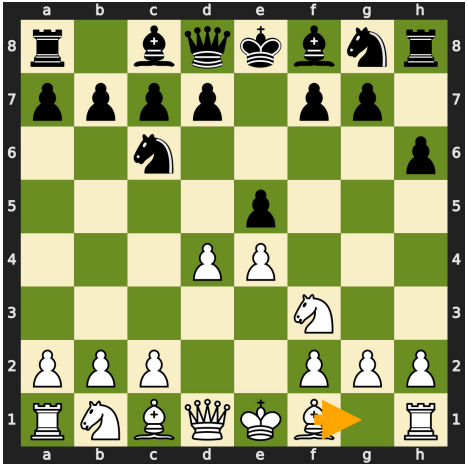
Chess Notation allows for probing for entity state via prompting!



`e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 f1??`

Entity Tracking Task: Ending Square

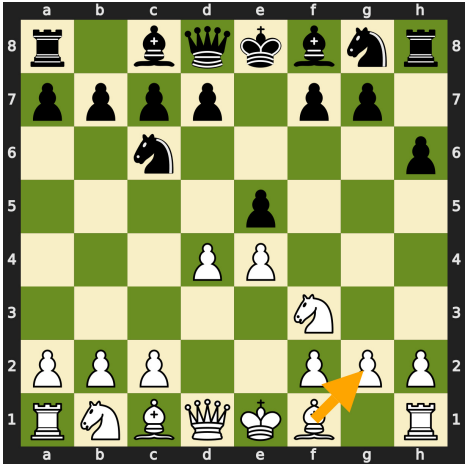
Chess Notation allows for probing for entity state via prompting!



e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 f1g1

Entity Tracking Task: Ending Square

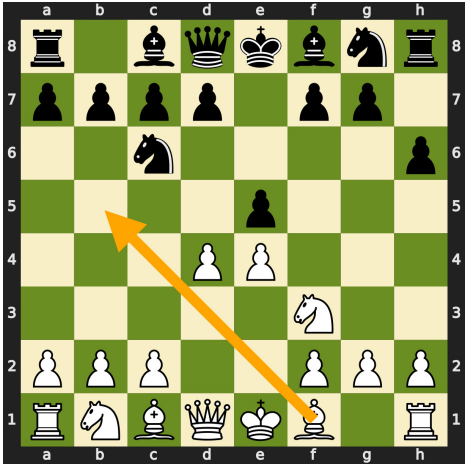
Chess Notation allows for probing for entity state via prompting!



`e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 f1g2`

Entity Tracking Task: Ending Square

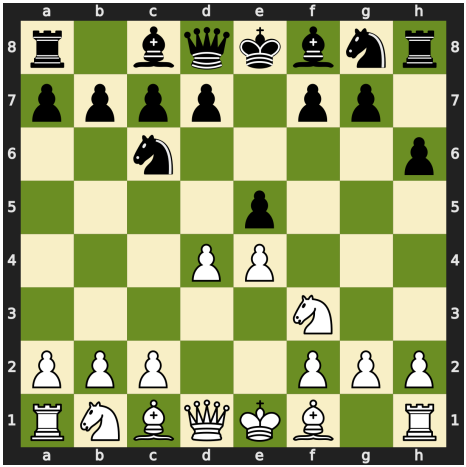
Chess Notation allows for probing for entity state via prompting!



e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 f1b5

Entity Tracking Task: Starting Square

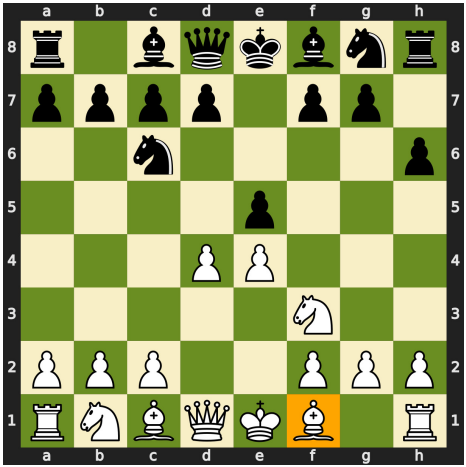
Training with RAP also allows for directly probing for piece location



e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 B??

Entity Tracking Task: Starting Square

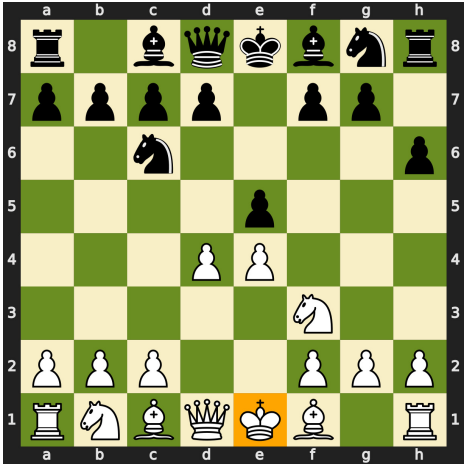
Training with RAP also allows for directly probing for piece location



e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 Bf1

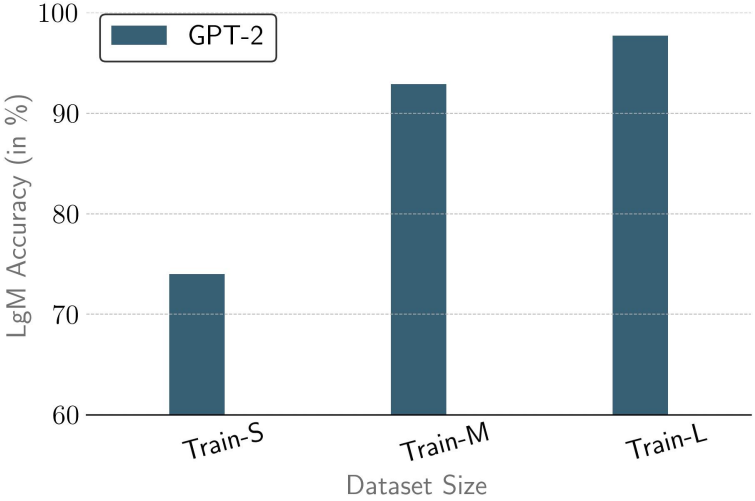
Entity Tracking Task: Starting Square

Training with RAP also allows for directly probing for piece location

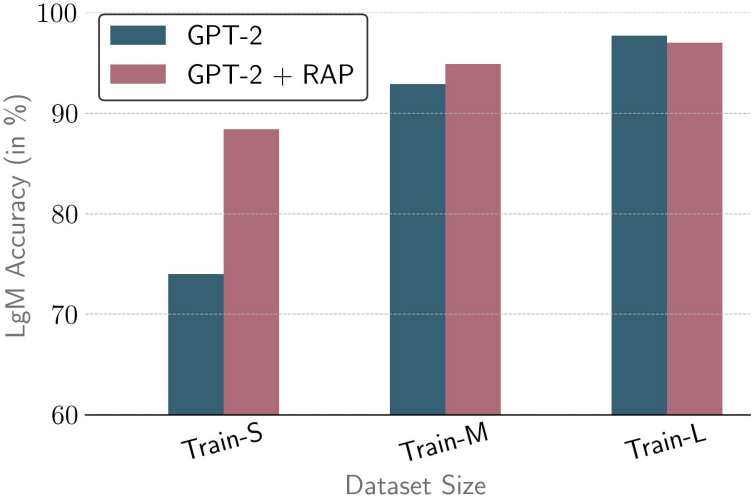


e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 Be1

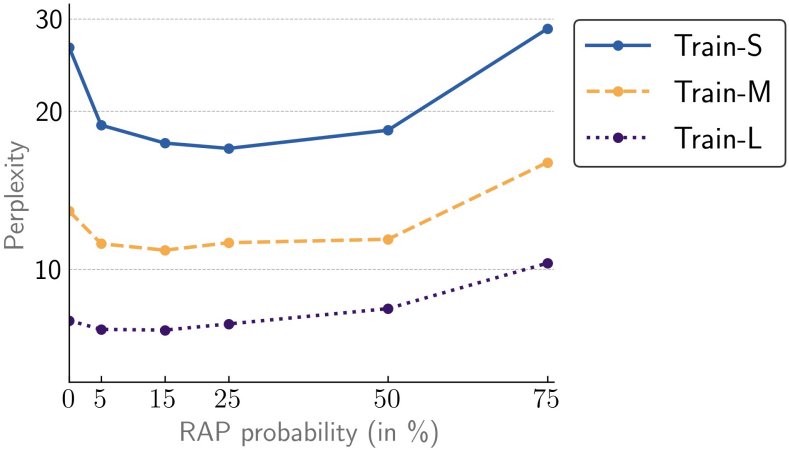
Entity Tracking Results



Entity Tracking Results



Language Modeling Results



Takeaways

Proposed chess as a testbed for entity tracking in language models

Data augmentation using RAP improves both entity tracking and language modeling results for low data settings

Roadmap

Explicit Entity Tracking

Coreference Resolution Models for Long Documents

Generalization in Coreference Resolution

Implicit Entity Tracking with Language Models

Chess as a Testbed for Entity Tracking

Baking in Coreference Knowledge into Language Models

Conclusion

Integrating Entity Tracking Into Natural Language Models

Text	`Bilbo Baggins leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir`

Integrating Entity Tracking Into Natural Language Models

Text	`Bilbo Baggins leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir`
Coreference	`[Bilbo Baggins] leaves the Shire suddenly, passing [the Ring] to [Frodo Baggins], [[his] cousin and heir]`

Integrating Entity Tracking Into Natural Language Models

Text	`Bilbo Baggins leaves the Shire suddenly, passing the Ring to Frodo Baggins, his cousin and heir`
Coreference	`[Bilbo Baggins] leaves the Shire suddenly, passing [the Ring] to [Frodo Baggins], [[his] cousin and heir]`
Coreference Augmentation	`Bilbo Baggins leaves the Shire suddenly, passing the Ring to Frodo Baggins, his [ Bilbo Baggins ] cousin and heir`

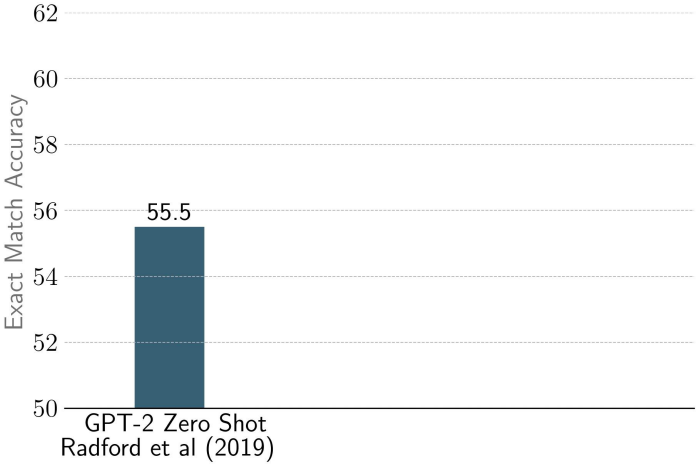
LAMBADA Cloze Task

gemma had never been introduced to any of them before, but she knew of them. she had heard her parents and hawke discuss mr. percival at great length. he was next in line for the earldom of worcester, and one of the royal duke's favorite cousins. no doubt they had designs on him as a match for _____

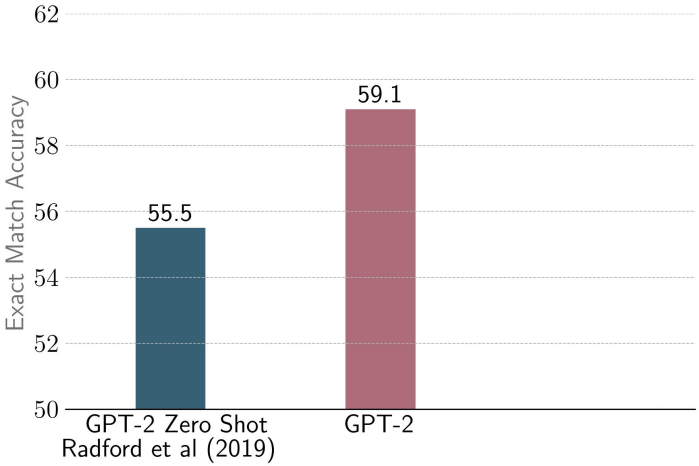
LAMBADA Cloze Task

gemma had never been introduced to any of them before, but she knew of them. she had heard her parents and hawke discuss mr. percival at great length. he was next in line for the earldom of worcester, and one of the royal duke's favorite cousins. no doubt they had designs on him as a match for gemma

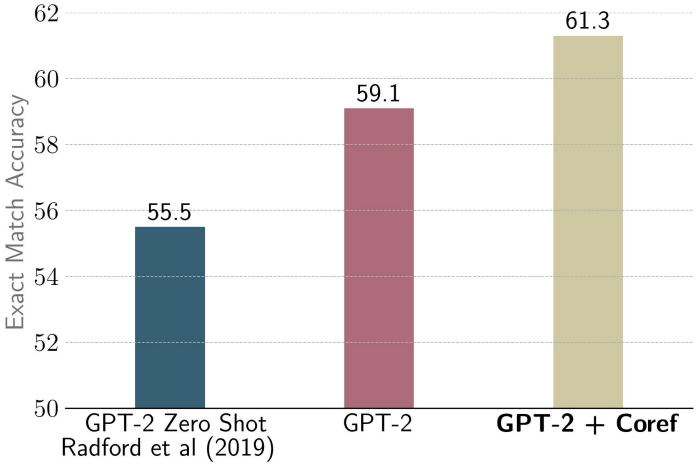
LAMBADA Results



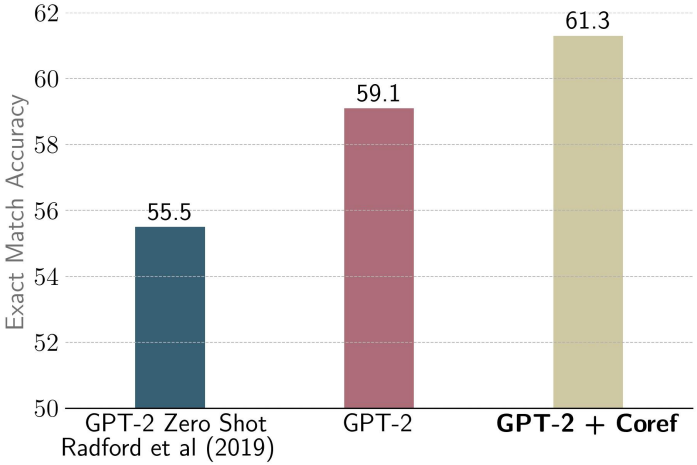
LAMBADA Results



LAMBADA Results



LAMBADA Results



Error analysis shows that most of the gains are due to improvement in entity tracking!

Roadmap

Explicit Entity Tracking

Coreference Resolution Models for Long Documents

Generalization in Coreference Resolution

Implicit Entity Tracking

Chess as a Testbed for Entity Tracking

Baking in Coreference Knowledge into Language Models

Conclusion

Conclusion

We presented our work on two approaches to the entity tracking task:

Explicit Entity Tracking

Proposed efficient coreference models with strong performance on coreference resolution benchmarks

Implicit Entity Tracking

Proposed chess as a testbed for entity tracking

Proposed data augmentation based training recipe to integrate entity tracking into language models

Future Work

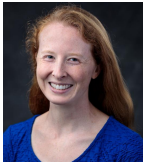
Long context understanding: Dialog logs, Books (NarrativeQA)

Thanks!

Kevin Gimpel

Karen Livescu

Sam Wiseman



Allyson Ettinger

Tara Sainath

Mohit Bansal

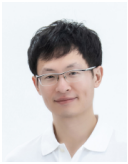
Trang Tran

Mari Ostendorf

Ron Weiss

Patrick Xia

Freda Shi



Hao Tang

Questions?