Empirical Evaluation of Pretraining Strategies for Supervised Entity Linking

Thibault Fevry, Nicholas FitzGerald, Livio Baldini Soares, Tom Kwiatkowski

Contributions

- New State-of-the-Art on CoNLL and TAC-KBP entity disambiguation with a 4-layer transformer and entity embeddings
- 2. Strong result on end-to-end entity linking
- 3. Ablation studies demonstrate importance of negative selection, input perturbations, and context selection

Entity Disambiguation

"Michael Jordan scored 29 points against Phoenix last Thursday..."



Background

- Traditional entity disambiguation approaches have relied on knowledge bases, complicated modelling and task-specific features
- Recent approaches show using large-scale pretrained language models like
 BERT leads to high performance on tasks related to entity disambiguation
- However, these works focus on different goals:
 - Learning reusable entity representations [Ling et al. 2020]
 - Zero-shot entity linking [Wu et al. 2019]
 - End-to-end entity linking [Broscheit 2019]

Pretraining Data

- Wikipedia (Apr 14, 2019 version)
 - Chunk in to 1000 unicode characters
 - 17.5 million contexts
 - 17 million entity mentions
 - 5.7 million unique entities

Basketball (sport)

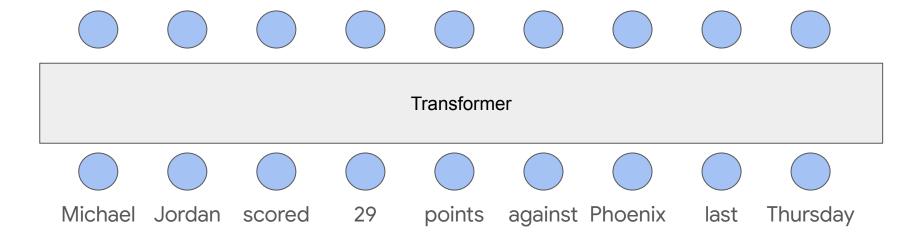
Michael Jordan

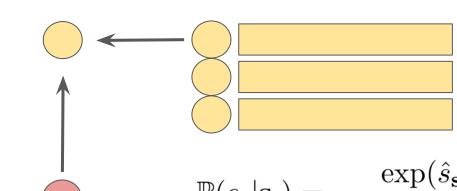
From Wikipedia, the free encyclopedia

This article is about the American basketball player. For other people with the same name, see Michael Jordan (disambig

Michael Jeffrey Jordan (born February 17, 1963), also known by his initials **MJ**,^[7] is an American businessman and former professional basketball player who is currently the principal owner of the Charlotte Hornets of the National Basketball Association (NBA). He played 15 seasons in the NBA, winning six championships with the Chicago Bulls. The official NBA website states that "by acclamation, Michael Jordan is the greatest basketball player of all time" and he was considered instrumental in popularizing the NBA throughout the world during the 1980s and 1990s. [9]

 $egin{aligned} \mathbf{H}_i &= \mathtt{TransformerBlock}(\mathbf{H}_{i-1}) \ &= \mathtt{MLP}(\mathtt{MultiHeadAttention}(\mathbf{H}_{i-1},\mathbf{H}_{i-1},\mathbf{H}_{i-1})) \end{aligned}$





Michael Jordan (Basketball Player)

Michael Jordan (Computer Scientist)

Phoenix Suns

$$\mathbb{P}(e_c|\mathbf{s}_i) = \frac{\exp(\hat{s}_{\mathbf{s}_i} \cdot e_c)}{\sum_{\mathbf{c}' \in \mathcal{E}} \exp(\hat{s}_{\mathbf{s}_i} \cdot e_{c'})}$$















Transformer



















Michael Jordan scored

29

points

against Phoenix

last

Thursday

Entity Candidate Selection

Basketball (sport)

Michael Jordan

From Wikipedia, the free encyclopedia

This article is about the American basketball player. For other people with the same name, see Michael Jordan (disambig

Michael Jeffrey Jordan (born February 17, 1963), also known by his initials MJ,^[7] is an American businessman and former profession it basketball player who is currently the principal owner of the Charlotte Hornets of the National Basketball Association (NBA). He played 15 seasons in the NBA, winning six championships with the Chicago Bulls. The official NBA website states that "by acclamation, Michael Jordan is the greatest basketball player of all time" and he was considered instrumental in popularizing the NBA throughout the world during the 1980s and 1990s. [9]

Page candidates



Entity Candidate Selection

Michael Jordan

From Wikipedia, the free encyclopedia

This article is about the American basketball player. For other people with the same name, see Michael Jordan (disamble Michael Jeffrey Jordan (born February 17, 1963), also known by his initials MJ,^[7] is an American businessman and former profession (basketball player who is currently the principal owner of the Charlotte Hornets of the National Basketball Association (NBA). He played 15 seasons in the NBA, winning six championships with the Chicago Bulls. The official NBA website states that "by acclamation, Michael Jordan is the greatest basketball player of all time" and he was considered instrumental in popularizing the NBA throughout the world during the 1980s and 1990s. [9]

- Page candidates
- Phrase table candidates
 - [Ringaard et al 2017]

NBA	"Basketball" (Song)
Chicago Bulls	Basketball (video game)
Scottie Pippen	"Basketball" ("The Office" episode)

Entity Candidate Selection

Basketball (sport)

Michael Jordan

From Wikipedia, the free encyclopedia

This article is about the American basketball player. For other people with the same name, see Michael Jordan (disambig

Michael Jeffrey Jordan (born February 17, 1963), also known by his initials MJ, [7] is an American businessman and former profession it basketball player who is currently the principal owner of the Charlotte Hornets of the National Basketball Association (NBA). He played 15 seasons in the NBA, winning six championships with the Chicago Bulls. The official NBA website states that "by acclamation, Michael Jordan is the greatest basketball player of all time" and he was considered instrumental in popularizing the NBA throughout the world during the 1980s and 1990s. [9]

- Page candidates
- Phrase table candidates
- Random candidates

NBA	"Basketball" (Song)	Canada (country)
Chicago Bulls	Basketball (video game)	SpaceX
Scottie Pippen	"Basketball" ("The Office" episode)	Huejotitan

Input Noising

Followed BERT recipe:

- 15% of tokens chosen, of which:
 - 80% replaced with [MASK]
 - 10% replaced with random wordpiece
 - 10% unmodified
- Masked language modelling loss did not improve accuracy

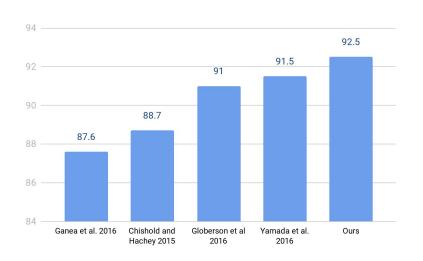


against Phoenix last Thursday

Evaluation Datasets

- AIDA CoNLL YAGO [Hoffart et al 2011]
 - Densely annotated with 34,000 mentions on 1,393 newswire documents
 - Full Wikipedia vocabulary
 - Context: Title and first two sentences
 - Alias tables: [Hoffart et al 2011] and [Pershina et al 2015]
- TAC-KBP [Ji et al 2010]
 - Sparsely annotated, linking to a vocabulary of ~830k entities
 - Context: 256 bytes before and after mention

Results - CoNLL

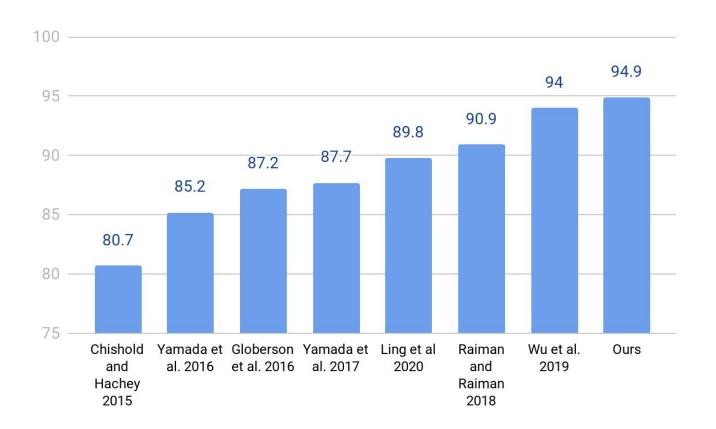


96.7 94.9 94.9 94.3 94 93.1 92.7 91.8 92 -Globerson Yamada et Ling et al Ours Yamada et Raiman and al. 2015 et al. 2016 al. 2017 al. 2017 Raiman 2019

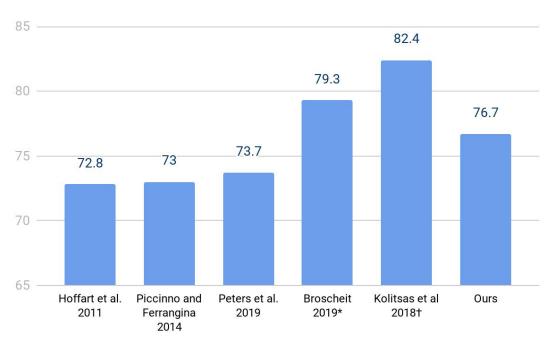
[Hoffart et al] Alias Table

[Pershina et al] Alias Table

Results - TAC-KBP

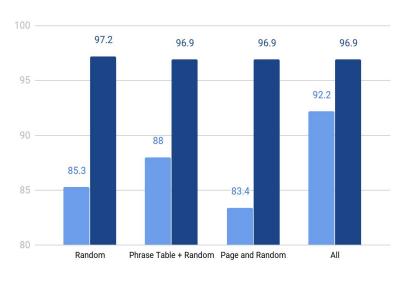


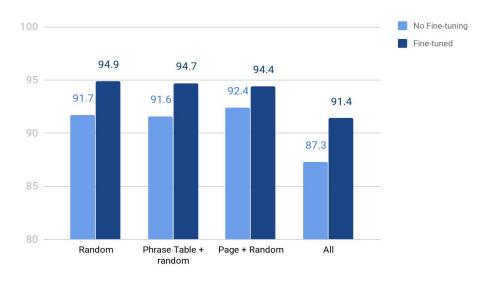
CoNLL End-to-End



^{*} Uses 12-layer pretrained BERT + Uses alias table

Impact of candidate selection

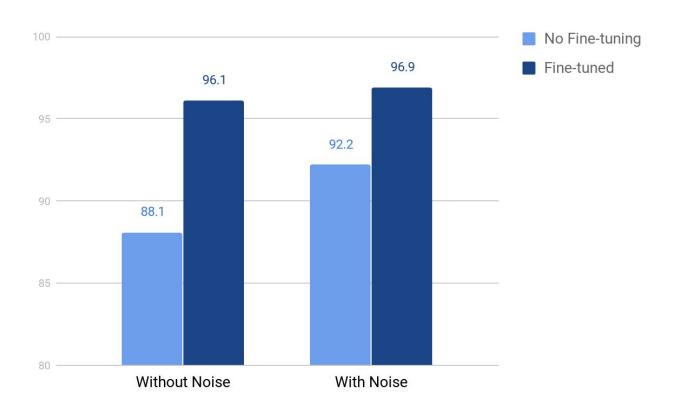




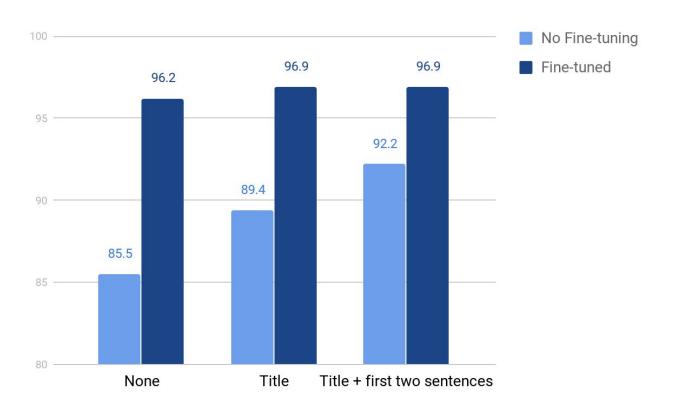
CoNLL

TAC-KBP

Impact of input noising



Impact of context selection



Error analysis

Soccer - Israel beat Bulgaria in European under-21 qualifier.

Mention Text	Prediction	Gold Label
"Israel"	Israel national football team	Israel national football team
"Bulgaria"	Bulgaria national under-21 football team	Bulgaria national football team
"European"	European	European

Error analysis

Scottish labour party narrowly backs referendum.

Mention Text	Prediction	Gold Label
"Labour party"	Scottish labour party	Labour party (UK)

Thank You

