Neural Entity Linking: A Survey of Models Based on Deep Learning

Alexander Panchenko

Skolkovo Institute of Science and Technology (Skoltech)

a.panchenko@skoltech.ru

Homepage: https://faculty.skoltech.ru/people/alexanderpanchenko
NLP group at Skoltech: https://sites.skoltech.ru/nlp

March 2, 2021

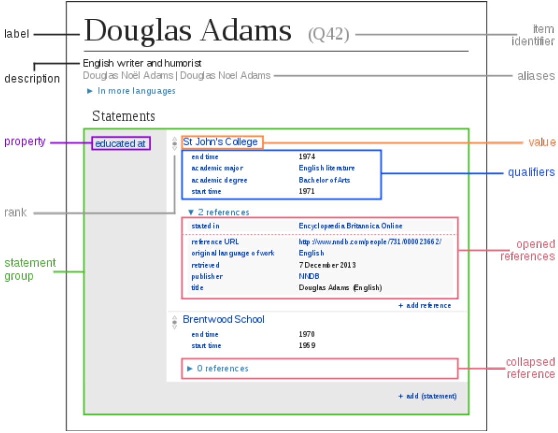
Overview

Acknowledgement

- This presentation is a considerably extended version of presentation by Özge Sevgili (University of Hamburg):
- 
- The materials are based on the following joint (submitted) work with Özge and other co-authors:
- Özge Sevgili, Artem Shelmanov, Mikhail Arkhipov, Alexander Panchenko, Chris Biemann (2021): Neural Entity Linking: A Survey of Models based on Deep Learning. CoRR abs/2006.00575

Motivation

- Knowledge Bases (KBs) like DBpedia, WikiData, and Freebase contain rich information about entities and their typed relationships.



Knowledge Base (KB)

- $KB = (E, R)$ – knowledge base is a multi-label graph

Knowledge Base (KB)

- $KB = (E, R)$ – knowledge base is a multi-label graph
- E – a set of entities (nodes)

Knowledge Base (KB)

- $KB = (E, R)$ – knowledge base is a multi-label graph
- E – a set of entities (nodes)
- $R \subset E \times T \times E$ – a set of directed typed relations between entities (edges)

Knowledge Base (KB)

- $KB = (E, R)$ – knowledge base is a multi-label graph
- E – a set of entities (nodes)
- $R \subset E \times T \times E$ – a set of directed typed relations between entities (edges)
- T – set of all relation types (sometimes just called relations)

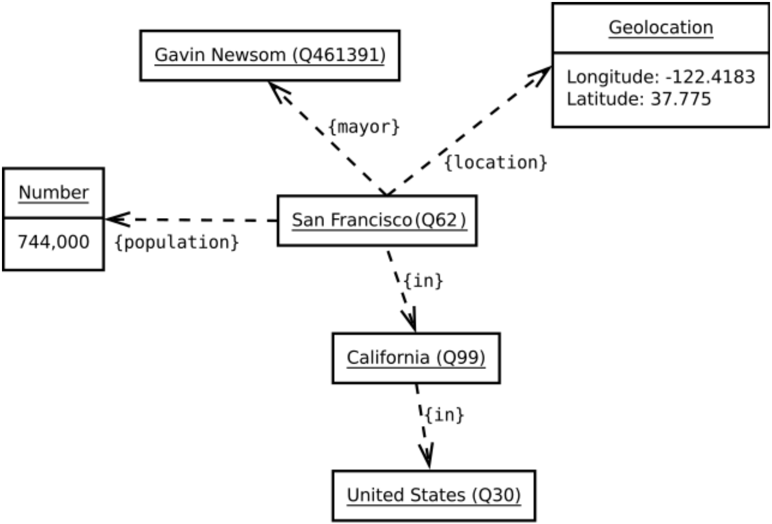
Knowledge Base (KB)

- $KB = (E, R)$ – knowledge base is a multi-label graph
- E – a set of entities (nodes)
- $R \subset E \times T \times E$ – a set of directed typed relations between entities (edges)
- T – set of all relation types (sometimes just called relations)
- $(s, p, o) = (e_i, t_j, e_k) \subset R$ - an spo triple (subject, predicate, object)

Knowledge Base (KB)

- $KB = (E, R)$ – knowledge base is a multi-label graph
- E – a set of entities (nodes)
- $R \subset E \times T \times E$ – a set of directed typed relations between entities (edges)
- T – set of all relation types (sometimes just called relations)
- $(s, p, o) = (e_i, t_j, e_k) \subset R$ - an spo triple (subject, predicate, object)
- Graph-tensor duality: Alternatively, a KB can be represented as a set of $|T|$ adjacency matrices each of dimensionality $|E| \times |E|$. They can be stacked into a 3-dimensional tensor of dimensionality $|E| \times |T| \times |E|$, where an spo triple is a point $(e_i, t_j, e_k) \in \mathbb{R}^3$.

A sample sub-graph from the WikiData KB



Information from KB is useful for semantic processing algorithms

- A search engine that is able to retrieve mentions in the news during the last month of all retired NBA players with a net income of more than 1 billion USD.

Information from KB is useful for semantic processing algorithms

- A search engine that is able to retrieve mentions in the news during the last month of all retired NBA players with a net income of more than 1 billion USD.
- The list of players together with their income and retirement information may be available in a KB.

Information from KB is useful for semantic processing algorithms

- A search engine that is able to retrieve mentions in the news during the last month of all retired NBA players with a net income of more than 1 billion USD.
- The list of players together with their income and retirement information may be available in a KB.
- Equipped with this information, it appears to be straightforward to look up mentions of such retired basketball players in the newswire.

Information from KB is useful for semantic processing algorithms

- A search engine that is able to retrieve mentions in the news during the last month of all retired NBA players with a net income of more than 1 billion USD.
- The list of players together with their income and retirement information may be available in a KB.
- Equipped with this information, it appears to be straightforward to look up mentions of such retired basketball players in the newswire.
- However, the main obstacle for such a direct counting algorithm is the lexical ambiguity of entities.

Information from KB is useful for semantic processing algorithms

- A search engine that is able to retrieve mentions in the news during the last month of all retired NBA players with a net income of more than 1 billion USD.
- The list of players together with their income and retirement information may be available in a KB.
- Equipped with this information, it appears to be straightforward to look up mentions of such retired basketball players in the newswire.
- However, the main obstacle for such a direct counting algorithm is the lexical ambiguity of entities.
- Only retrieve all mentions of “Michael Jordan (basketball player)” and exclude mentions of other persons with the same name such as “Michael Jordan (mathematician)”.

Entity Linking (EL) to the rescue: a technology for disentangling ambiguous entity mentions in text

- There will be more than one entity for the same mention string – “Michael Jordan (basketball player)” vs “Micheal Jordan (mathematician)”.

Entity Linking (EL) to the rescue: a technology for disentangling ambiguous entity mentions in text

- There will be more than one entity for the same mention string – “Michael Jordan (basketball player)” vs “Micheal Jordan (mathematician)”.
- The mapping between a mention in a context and KB entry is required to retrieve the correct information.

Entity Linking (EL) to the rescue: a technology for disentangling ambiguous entity mentions in text

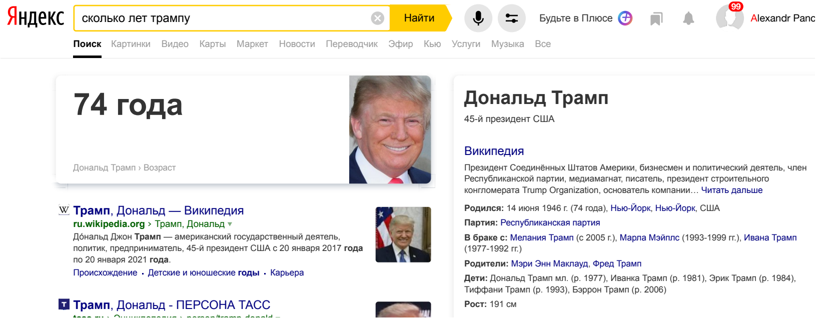
- There will be more than one entity for the same mention string – “Michael Jordan (basketball player)” vs “Micheal Jordan (mathematician)”.
- The mapping between a mention in a context and KB entry is required to retrieve the correct information.
- Entity Linking (EL) is the process of matching a mention, e.g. “Michael Jordan”, in a textual context to a KB entity (e.g. “basketball player” or “mathematician”) fitting the context.

Entity Linking (EL) to the rescue: a technology for disentangling ambiguous entity mentions in text

- There will be more than one entity for the same mention string – “Michael Jordan (basketball player)” vs “Micheal Jordan (mathematician)”.
- The mapping between a mention in a context and KB entry is required to retrieve the correct information.
- Entity Linking (EL) is the process of matching a mention, e.g. “Michael Jordan”, in a textual context to a KB entity (e.g. “basketball player” or “mathematician”) fitting the context.
- This is the key technology enabling various semantic applications.

Another application: KB question answering (KBQA)

- A type of question answering, where an answer is available in a KB.
- Typically, an answer is an entity $e \in E$ or a value (an object of an spo triple which does not belong to E).
- Occasionally an answer may be a relation or a more complex subset of the *KB*.



Another application: KB question answering (KBQA)

Google

which country tahiti belong to

All · Images · News · Maps · Videos · More · Settings · Tools

About 36,900,000 results (0.78 seconds)

French Polynesia

Tahiti itself is the largest of the Society Islands of **French Polynesia**. Oct 5, 2018

tahititourisme.com › en-us › island › where-is-tahiti-and-t...

Where is Tahiti & The Tahitian Islands? | Tahiti Tourisme

About featured snippets • Feedback

People also ask

- What country owns Tahiti?
- Is Tahiti part of USA?
- Is Tahiti owned by France?
- Is Tahiti Australian?

Feedback

en.wikipedia.org › wiki › Tahiti

Tahiti - Wikipedia

Tahiti is the largest island of the Windward group of the Society Islands in **French Polynesia**, ... For example, the languages of Fiji and Polynesia all **belong to** the same Oceanic sub-group, Fijian–Polynesian, ... to retain a considerable hold over **Tahitian** society, thanks to their knowledge of the **country** and its language.

Population: 189,517 (August 2017 census) **Largest settlement:** Papeete (pop. 136,...
Location: Pacific Ocean **Ethnic groups:** Tahitians

Kingdom of Tahiti · Category:Tahiti · Music of Tahiti · Tahitians

en.wikipedia.org › wiki › French_Polynesia

French Polynesia - Wikipedia

French Polynesia is an overseas collectivity of the French Republic and its sole overseas **country**. ... A majority of 54% **belongs to** various Protestant churches, especially the Maohi Protestant Church, which is the largest and accounts for more ...

Country status (nominal title): 27 Februar... **Territorial status:** 27 October 1946
Official languages: French **Recognised regional languages:** Tahiti...

Implementation of the KBQA in the DeepPavlov framework over the WikiData knowledge base

- The following models are used to find the answer:

Implementation of the KBQA in the DeepPavlov framework over the WikiData knowledge base

- The following models are used to find the answer:
  1. BERT model for prediction of query template type. Model performs classification of questions into 8 classes corresponding to 8 query template types.

Implementation of the KBQA in the DeepPavlov framework over the WikiData knowledge base

- The following models are used to find the answer:
  1. BERT model for prediction of query template type. Model performs classification of questions into 8 classes corresponding to 8 query template types.
  2. BERT entity detection model for extraction of entity substrings from the questions.

Implementation of the KBQA in the DeepPavlov framework over the WikiData knowledge base

- The following models are used to find the answer:
  1. BERT model for prediction of query template type. Model performs classification of questions into 8 classes corresponding to 8 query template types.
  2. BERT entity detection model for extraction of entity substrings from the questions.
  3. Substring extracted by the entity detection model is used for entity linking. Entity linking performs matching the substring with one of the Wikidata entities. Matching is based on Levenshtein distance between the substring and an entity title. The result of the matching procedure is a set of candidate entities.

Implementation of the KBQA in the DeepPavlov framework over the WikiData knowledge base

- The following models are used to find the answer:
  1. BERT model for prediction of query template type. Model performs classification of questions into 8 classes corresponding to 8 query template types.
  2. BERT entity detection model for extraction of entity substrings from the questions.
  3. Substring extracted by the entity detection model is used for entity linking. Entity linking performs matching the substring with one of the Wikidata entities. Matching is based on Levenshtein distance between the substring and an entity title. The result of the matching procedure is a set of candidate entities.
  4. BiGRU model for ranking of candidate relations.
  5. BERT model for ranking of candidate relation paths.

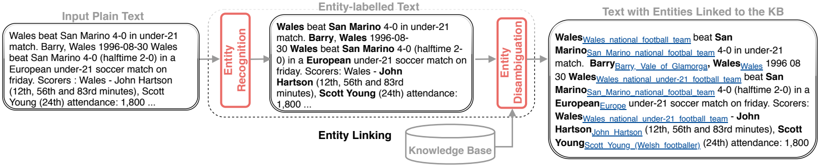
Implementation of the KBQA in the DeepPavlov framework over the WikiData knowledge base

- The following models are used to find the answer:
  1. BERT model for prediction of query template type. Model performs classification of questions into 8 classes corresponding to 8 query template types.
  2. BERT entity detection model for extraction of entity substrings from the questions.
  3. Substring extracted by the entity detection model is used for entity linking. Entity linking performs matching the substring with one of the Wikidata entities. Matching is based on Levenshtein distance between the substring and an entity title. The result of the matching procedure is a set of candidate entities.
  4. BiGRU model for ranking of candidate relations.
  5. BERT model for ranking of candidate relation paths.
  6. Query generator model is used to fill query template with candidate entities and relations.

Problem definition

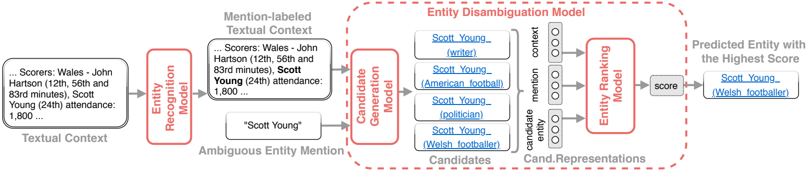
- EL model takes a raw textual input and enriches it with entity mention links in a KB.
- Commonly the task is split into entity recognition (ER) and entity disambiguation (ED) sub-tasks:

$$\mathrm{ER} : C \to M^n, \mathrm{ED} : (M, C)^n \to E^n.$$



General architecture

- Recent neural EL models use a generic architecture that is applicable for the most of the neural models.
- Most of the systems focus on ED by referring it as EL.



General architecture: four main components

1. Candidate Generation
2. Mention-Context Encoder
3. Entity Encoder
4. Entity Ranking

Candidate generation

- The goal of this step is given an ambiguous entity mention, such as “Big Blue”, to provide a list of its possible “senses” as specified by entities in a KB:

$$CG : M^n \to (e_1, e_2, ..., e_k)^n$$

Method	10 sample candidate entities for the example mention “Big Blue”
surface form matching based on DBpedia	Santa_Monica_Big_Blue_Bus, Bear_in_the_big_blue_house, The_Big_Blue_Bug, The_Big_Blue_Marble, IBM_Big_Blue_(rugby_union), The_Blue_Mouse_and_the_Big_Faced_Cat, The_Big_Blue_(A-League), The_Big_Blue_Megamix, Millikin_Big_Blue_football, IBM_Big_Blue_(disambiguation)
dictionary lookup based on YAGO-means	Big_Blue_River_(Indiana), Big_Blue_River_(Kansas), Big_Blue_(crane), Big_Red_(drink), **IBM**, IBM_Big_Blue, Millville_Football_&_Athletic_Club, Our_Lady_of_Mount_Carmel_High_School_(Baltimore,_Maryland), The_Big_Blue, Tift_County_High_School
prior probability based on CrossWikis	**IBM**, Big_Blue_River_(Kansas), The_Big_Blue, Utah_State_University, New_York_Giants, Big_Blue_River_(Indiana), Big_Blue_(crane), Big_Blue_(disambiguation), Deep_Blue_(chess_computer), Superman

Table 1

Candidate generation examples. Ten sample candidate entities for the example mention “Big Blue” for each method. The highlighted are “correct” candidates assuming that given mention refers to the IBM corporation and not its sport teams, e.g. IBM_Big_Blue_(rugby_union).

Context-mention encoder

- To capture the information of entity mention from its context, the streamline approach is to construct a dense contextualized vector representation of a mention:

$$\mathrm{mENC} : (C, M)^n \rightarrow (\boldsymbol{y}_{\boldsymbol{m_1}}, \boldsymbol{y}_{\boldsymbol{m_2}}, \ldots, \boldsymbol{y}_{\boldsymbol{m_n}})$$

Context-mention encoder

- To capture the information of entity mention from its context, the streamline approach is to construct a dense contextualized vector representation of a mention:

$$\mathrm{mENC} : (C, M)^n \rightarrow (\boldsymbol{y}_{\boldsymbol{m_1}}, \boldsymbol{y}_{\boldsymbol{m_2}}, ..., \boldsymbol{y}_{\boldsymbol{m_n}})$$

- Early techniques depend on CNN architecture, however in recent models, two approaches prevail: recurrent networks and self-attention.

Context-mention encoder

- To capture the information of entity mention from its context, the streamline approach is to construct a dense contextualized vector representation of a mention:

$$\mathrm{mENC} : (C, M)^n \rightarrow (\boldsymbol{y}_{\boldsymbol{m}_1}, \boldsymbol{y}_{\boldsymbol{m}_2}, ..., \boldsymbol{y}_{\boldsymbol{m}_n})$$

- Early techniques depend on CNN architecture, however in recent models, two approaches prevail: recurrent networks and self-attention.
- A recurrent network with LSTM cells are ubiquitous to encode left and right context of a mention.

Context-mention encoder

- To capture the information of entity mention from its context, the streamline approach is to construct a dense contextualized vector representation of a mention:

$$\mathrm{mENC} : (C, M)^n \rightarrow (\boldsymbol{y}_{\boldsymbol{m_1}}, \boldsymbol{y}_{\boldsymbol{m_2}}, \ldots, \boldsymbol{y}_{\boldsymbol{m_n}})$$

- Early techniques depend on CNN architecture, however in recent models, two approaches prevail: recurrent networks and self-attention.
- A recurrent network with LSTM cells are ubiquitous to encode left and right context of a mention.
- A self-attention based models rely on the outputs from pre-trained BERT layers for context and mention encoding.

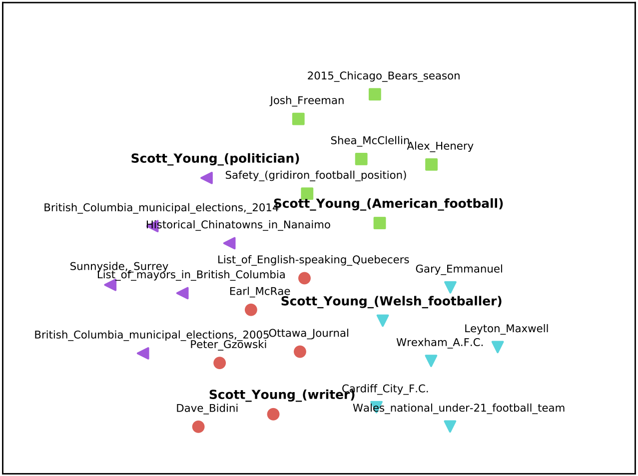
Entity encoder

- Good representations $\boldsymbol{y_e}$ of entity candidates that capture various semantic information are essential for making EL systems robust:

$$\text{eENC} : E^k \rightarrow (\boldsymbol{y_{e_1}}, \boldsymbol{y_{e_2}}, ..., \boldsymbol{y_{e_k}})$$

- Entities are encoded into low-dimensional vectors in such a way that spatial proximity between them in a vector space correlates with their semantic relatedness

Visualization of entity embeddings for “Scott Young”



Entity encoder

- Commonly, entities are represented with their dense vectors to use unstructural (e.g. description pages) or structural entity information (e.g. incoming links).

Entity encoder

- Commonly, entities are represented with their dense vectors to use unstructural (e.g. description pages) or structural entity information (e.g. incoming links).
- Some techniques depend on statistics features like word-entity co-occurrences from labeled/anchor data to train encoder.

Entity encoder

- Commonly, entities are represented with their dense vectors to use unstructural (e.g. description pages) or structural entity information (e.g. incoming links).
- Some techniques depend on statistics features like word-entity co-occurrences from labeled/anchor data to train encoder.
- There are some other models, which directly replace the anchor text with an entity descriptor and train the word representation model like word2vec.

Entity encoder

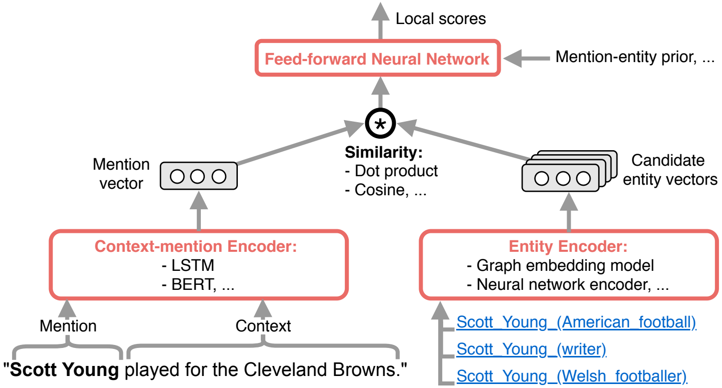
- Commonly, entities are represented with their dense vectors to use unstructural (e.g. description pages) or structural entity information (e.g. incoming links).
- Some techniques depend on statistics features like word-entity co-occurrences from labeled/anchor data to train encoder.
- There are some other models, which directly replace the anchor text with an entity descriptor and train the word representation model like word2vec.
- There are few recent studies, which perform entity encoding without entity annotated text data, using distant supervision or using only structural information.

Features of entity embeddings

	Annotated Text	Entity-Entity Links	Entity-Mention Links	Entity Descriptions	Entity Titles	Entity Types	Joint Space of Entities and Words
Huang et al. (2015) [45]		✘	✘	✘		✘	
Sun et al. (2015) [102]	✘				✘	✘	✘[1,6]
Fang et al. (2016) [25]	✘	✘	✘	✘			✘
Yamada et al. (2016) [116]	✘	✘					✘
Zwicklbauer et al. (2016) [125]	✘[2]			✘			
Tsai and Roth (2016) [104]	✘				✘		✘
Ganea and Hofmann (2017) [32]	✘						✘
Cao et al. (2017) [11]	✘	✘	✘				✘
Moreno et al. (2017) [69]	✘						✘
Gupta et al. (2017) [38]	✘			✘		✘	✘[4,6]
Sil et al. (2018) [98]				✘			✘
Upadhyay et al. (2018) [106]	✘		✘			✘	✘
Newman-Griffis et al. (2018) [75]					✘	✘	✘
Radhakrishnan et al. (2018) [87]	✘						✘
Rijhwani et al. (2019) [90]	✘	✘			✘		✘
Logeswaran et al. (2019) [62]				✘			✘[3,6]
Gillick et al. (2019) [34]	✘			✘	✘	✘	✘[6]
Le and Titov (2019) [55]						✘	✘[6]
Sevgili et al. (2019) [92]		✘		✘			
Shahbazi et al. (2019) [94]	✘						✘
Shi et al. (2020) [97]	✘	✘				✘	✘
Zhou et al. (2020) [124]	✘	✘	✘		✘		✘
Wu et al. (2019) [114]				✘	✘		✘[5,6]
Yamada et al. (2020) [117]	✘						✘[6]

Entity ranking

- Given a list of entity candidates from a KB and a context with a mention to rank these entities:
  RNK : $((e_1, e_2, ..., e_k), C, M)^n \rightarrow \mathbb{R}^{n \times k}$



Entity ranking: unsupervised models

- Most of the state-of-the-art studies compute similarity between representations of a mention and an entity using dot product $s(m, e_i) = \boldsymbol{y}_m \cdot \boldsymbol{y}_{e_i}$; or cosine similarity $s(m, e_i) = \cos(\boldsymbol{y}_m, \boldsymbol{y}_{e_i}) = \frac{\boldsymbol{y}_m \cdot \boldsymbol{y}_{e_i}}{\|\boldsymbol{y}_m\| \cdot \|\boldsymbol{y}_{e_i}\|}$..

Entity ranking: unsupervised models

- Most of the state-of-the-art studies compute similarity between representations of a mention and an entity using dot product $s\,(m, e_i) = \boldsymbol{y}_m \cdot \boldsymbol{y}_{e_i}$; or cosine similarity $s\,(m, e_i) = \cos(\boldsymbol{y}_m, \boldsymbol{y}_{e_i}) = \frac{\boldsymbol{y}_m \cdot \boldsymbol{y}_{e_i}}{\|\boldsymbol{y}_m\| \cdot \|\boldsymbol{y}_{e_i}\|}$..
- The final decision is inferred via probability distribution, which is usually approximated by a softmax function over the candidates. $P(e_i|m) = \frac{\exp(s(m,e_i))}{\sum_{i=1}^{k} \exp(s(m,e_i))}$

Entity ranking: supervised models

- There are several approaches to frame a training objective in the literature on EL. Consider we have k candidates for the target mention m, one of which is a true entity e_*.
- In some works, the models are trained with the standard negative log likelihood objective like in classification tasks [Logeswaran et al., 2019, Wu et al., 2019]. However, instead of classes, negative candidates are used:

$$\mathcal{L}(m) = -s(m, e_*) + \sum_{i=1}^{k} s(m, e_i)$$

Entity ranking: supervised models

- There are several approaches to frame a training objective in the literature on EL. Consider we have k candidates for the target mention m, one of which is a true entity e_*.
- In some works, the models are trained with the standard negative log likelihood objective like in classification tasks [Logeswaran et al., 2019, Wu et al., 2019]. However, instead of classes, negative candidates are used:

$$\mathcal{L}(m) = -s(m, e_*) + \sum_{i=1}^{k} s(m, e_i)$$

- Instead of the the negative log likelihood, some works use variants of a ranking loss.

NIL prediction

- The referent entities of some mentions can be absent in the KBs, e.g. there is no Wikipedia entry about Scott Young as a cricket player of the Stenhousemuir cricket club.

NIL prediction

- The referent entities of some mentions can be absent in the KBs, e.g. there is no Wikipedia entry about Scott Young as a cricket player of the Stenhousemuir cricket club.
- Therefore, an EL system should be able to predict the absence of a reference if a mention appears in specific contexts, which is known as NIL prediction task.

$$\mathrm{NIL} : (C, M)^n \rightarrow \{0, 1\}^n$$

NIL prediction

- The referent entities of some mentions can be absent in the KBs, e.g. there is no Wikipedia entry about Scott Young as a cricket player of the Stenhousemuir cricket club.
- Therefore, an EL system should be able to predict the absence of a reference if a mention appears in specific contexts, which is known as NIL prediction task.

$$\mathrm{NIL} : (C, M)^n \rightarrow \{0, 1\}^n$$

- This is similar to the “reject option”.

Modifications: Joint ER+ED Architectures

- The main difference of joint models is the necessity to produce also mention candidates.

$$\text{EL} : C \rightarrow (M, E)^n.$$

Modifications: Joint ER+ED Architectures

- The main difference of joint models is the necessity to produce also mention candidates.

$$\mathrm{EL} : C \rightarrow (M, E)^n.$$

- Mostly the models treat every span (with a certain width) as a mention candidate and check whether it has possible entity candidate.

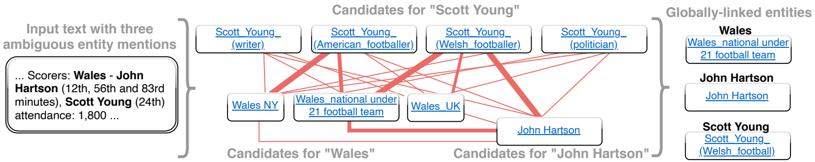
Modifications: Joint ER+ED Architectures

- The main difference of joint models is the necessity to produce also mention candidates.

$$\mathrm{EL} : C \rightarrow (M, E)^n.$$

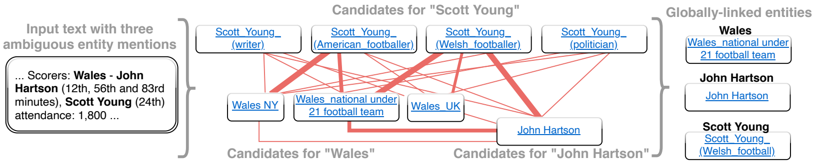
- Mostly the models treat every span (with a certain width) as a mention candidate and check whether it has possible entity candidate.
- Therefore, the decision during the entity disambiguation phase affects entity recognition. However, the interaction between these steps can be beneficial.

Modifications: Global Context Architectures



- Global approaches to ED take into account semantic consistency across multiple entities in a context.

Modifications: Global Context Architectures



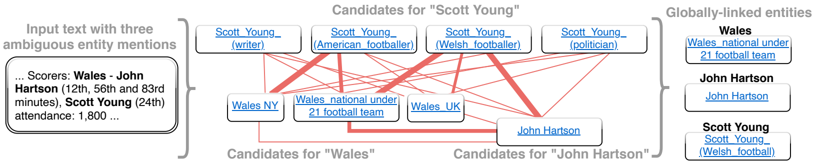
- Global approaches to ED take into account semantic consistency across multiple entities in a context.
- Compare:

$$\text{LED} : (M, C) \rightarrow E$$

and

$$\text{GED} : ((m_1, m_2, ..., m_q), C) \rightarrow E^q$$

Modifications: Global Context Architectures



- Global approaches to ED take into account semantic consistency across multiple entities in a context.
- Compare:

$$\text{LED} : (M, C) \rightarrow E$$

and

$$\text{GED} : ((m_1, m_2, ..., m_q), C) \rightarrow E^q$$

- All entity mentions are disambiguated interdependently: a disambiguation decision for one entity is affected by decisions made for other entities in the context.

Modifications: Global Context Architectures

- Although the extra information of the global context improves the disambiguation accuracy, the number of possible entity assignments is combinatorial, which results in a high time complexity of disambiguation.

Modifications: Global Context Architectures

- Although the extra information of the global context improves the disambiguation accuracy, the number of possible entity assignments is combinatorial, which results in a high time complexity of disambiguation.
- Most of the solutions depend on pairwise entity scores.
- Some studies define the problem as a sequential decision task, where the disambiguation of new entities is based on the already disambiguated ones, using reinforcement learning or LSTM

Modifications: Domain-Independent Architectures

- Annotated resources are very limited and exist only for a few domains. Obtaining labeled data in a new domain requires much labor.

Modifications: Domain-Independent Architectures

- Annotated resources are very limited and exist only for a few domains. Obtaining labeled data in a new domain requires much labor.
- Early solutions are based on unsupervised or semi-supervised models, recently zero-shot models are proposed.

Modifications: Domain-Independent Architectures

- Annotated resources are very limited and exist only for a few domains. Obtaining labeled data in a new domain requires much labor.
- Early solutions are based on unsupervised or semi-supervised models, recently zero-shot models are proposed.
- In zero-shot setting, the only entity information available is its description. For training, texts with mention-entity pairs are also available. The key idea here is to train in one domain and test it in another.

Modifications: Domain-Independent Architectures

- Annotated resources are very limited and exist only for a few domains. Obtaining labeled data in a new domain requires much labor.
- Early solutions are based on unsupervised or semi-supervised models, recently zero-shot models are proposed.
- In zero-shot setting, the only entity information available is its description. For training, texts with mention-entity pairs are also available. The key idea here is to train in one domain and test it in another.
- Recent zero-shot solutions are based on BERT architecture.

Modifications: Cross-lingual Architectures

- There is a big gap between resource-rich Wikipedia languages, like English, and low-resource ones.

Modifications: Cross-lingual Architectures

- There is a big gap between resource-rich Wikipedia languages, like English, and low-resource ones.
- The cross-lingual EL methods aim at overcoming the lack of annotation for some languages.

Modifications: Cross-lingual Architectures

- There is a big gap between resource-rich Wikipedia languages, like English, and low-resource ones.
- The cross-lingual EL methods aim at overcoming the lack of annotation for some languages.
- The inter-language links in Wikipedia is one of the most widely used sources of cross-lingual supervision. These links map pages to equivalent pages in another language.

Modifications: Cross-lingual Architectures

- There is a big gap between resource-rich Wikipedia languages, like English, and low-resource ones.
- The cross-lingual EL methods aim at overcoming the lack of annotation for some languages.
- The inter-language links in Wikipedia is one of the most widely used sources of cross-lingual supervision. These links map pages to equivalent pages in another language.
- Existing techniques of cross-lingual entity linking heavily rely on pre-trained multilingual embeddings for entity ranking. Although there are also zero-shot cross-lingual approaches, they are not powerful.

	Encoder Type	Global	Recog-nition	NIL Prediction	Entity Embeddings	Candidate Generation	Zero-shot	Annotated Text Data	Cross-lingual
Sun et al. (2015) [102]	CNN+ Tensor net.				joint architecture	surface match dictionary		✘	
Francis-Landau et al. (2016) [29]	CNN	✘[4]			joint architecture	surface match prior		✘	
Fang et al. (2016) [25]	n/a	✘			pre-trained[2]	prior[1]		✘	
Yamada et al. (2016) [116]	n/a	✘			pre-trained[2]	prior or dictionary		✘	
Zwicklbauer et al. (2016) [125]	n/a	✘		✘	pre-trained[2]	surface match prior nearest neighbors		✘	
Tsai and Roth (2016) [104]	n/a	✘		✘	pre-trained[2]	prior		✘	✘
Nguyen et al. (2016) [77]	CNN	✘		✘	joint architecture	surface match prior		✘	
Cao et al. (2017) [11]	n/a	✘			pre-trained[2]	dictionary		in entity embedding	
Eshel et al. (2017) [24]	GRU+ Atten.				joint architecture	dictionary		✘	
Ganea and Hofmann (2017) [32]	Atten.	✘			pre-trained[2]	prior+ nearest neighbors		✘	
Moreno et al. (2017) [69]	n/a	✘[4]		✘	pre-trained[2]	surface match		✘	
Gupta et al. (2017) [38]	LSTM	✘[4]			joint architecture	prior	✘		
Sorokin and Gurevych (2018) [99]	CNN	✘	✘		pre-trained[2]	surface match		✘	
Shahbazi et al. (2018) [93]	Atten.	✘			pre-trained	prior		✘	
Le and Titov (2018) [54]	Atten.	✘			pre-trained	prior		✘	
Newman-Griffis et al. (2018) [75]	n/a				pre-trained[2]	dictionary			
Radhakrishnan et al. (2018) [87]	n/a	✘			pre-trained[2]	dictionary		✘	
Kolitsas et al. (2018) [51]	LSTM	✘	✘		pre-trained	prior		✘	
Sil et al. (2018) [98]	LSTM+ Tensor net.	✘[4]		✘	joint architecture	prior	✘[5]	✘	✘
Upadhyay et al. (2018) [106]	CNN				joint architecture	prior		✘	✘
Cao et al. (2018) [12]	FFNN	✘[4]			pre-trained[2]	prior		✘	
Raiman and Raiman (2018) [88]	n/a	✘			n/a	prior type classifier		✘	✘
Mueller and Durrett (2018) [71]	GRU+ Atten.+ CNN				joint architecture	dictionary		✘	
Shahbazi et al. (2019) [94]	ELMo				pre-trained[2]	prior or dictionary		✘	
Logeswaran et al. (2019) [62]	BERT				joint architecture	BM25	✘		
Gillick et al. (2019) [34]	FFNN				joint architecture	nearest neighbors	✘	in entity embedding	
Peters et al. (2019) [85][3]	BERT	✘[4]	✘	✘	pre-trained	prior		in entity embedding	
Le and Titov (2019) [55]	LSTM				joint architecture	surface match			
Le and Titov (2019) [56]	Atten.	✘			pre-trained	prior		in entity embedding	
Fang et al. (2019) [26]	LSTM	✘			pre-trained	dictionary		✘	
Martins et al. (2019) [65]	LSTM		✘	✘	pre-trained	dictionary		✘	
Yang et al. (2019) [118]	Atten. or CNN	✘			pre-trained	prior		✘	
Broscheit (2019) [9]	BERT		✘		n/a	n/a		✘	
Onoe and Durrett (2020) [79]	ELMo+ Atten.+ CNN				n/a	prior or dictionary		✘	
Wu et al. (2019) [114]	BERT				joint architecture	nearest neighbors	✘		
Yamada et al. (2020) [117]	BERT	✘			joint architecture	prior		✘	

Classical application of entity linking

- Biomedical: Clinical text processing – COVIDASK a system to answer coronavirus related questions. EL is used to link objects, like drugs, symptoms, disease mentions.

Classical application of entity linking

- Biomedical: Clinical text processing – COVIDASK a system to answer coronavirus related questions. EL is used to link objects, like drugs, symptoms, disease mentions.
- Relation extraction: extraction of relations between mentions such as “child-of”, “politician-from”, “born-in”, etc. EL helps to build a resource.

Classical application of entity linking

- Biomedical: Clinical text processing – COVIDASK a system to answer coronavirus related questions. EL is used to link objects, like drugs, symptoms, disease mentions.
- Relation extraction: extraction of relations between mentions such as “child-of”, “politician-from”, “born-in”, etc. EL helps to build a resource.
- Semantic parsing, question answering, information retrieval: EL helps to restrict the search space of a query. “Who first voiced Meg on Family Guy?”, after linking “Meg” and “Family Guy” to entities in a KB, the task becomes to resolve the predicates to the “Family Guy (the TV show)” entry rather than all entries in the KB.

Novel applications: training of neural language models

- Neural EL models have unlocked the new category of application.
- Neural models allow the integration of an entire entity linking system inside a larger neural network such as BERT [Devlin et al., 2019].

$$\mathcal{L}_{\text{JOINT}} = \mathcal{L}_{\text{BERT}} + \mathcal{L}_{\text{EL-related}} .$$

- EL helps in language models to benefit from information stored in KBs by incorporating EL into deep models for transfer learning.

Novel applications: the use-case of KnowBERT

- The original objective of BERT consists of the masked language model (MLM) task and the next sentence prediction (NSP) task:

$$\mathcal{L}_{\text{BERT}} = \mathcal{L}_{\text{NSP}} + \mathcal{L}_{\text{MLM}}.$$

Novel applications: the use-case of KnowBERT

- The original objective of BERT consists of the masked language model (MLM) task and the next sentence prediction (NSP) task:

$$\mathcal{L}_{\text{BERT}} = \mathcal{L}_{\text{NSP}} + \mathcal{L}_{\text{MLM}}.$$

- KnowBERT [Peters et al., 2019] injects one or several entity linkers between top layers of the BERT architecture.

Novel applications: the use-case of KnowBERT

- The original objective of BERT consists of the masked language model (MLM) task and the next sentence prediction (NSP) task:

$$\mathcal{L}_{\mathrm{BERT}} = \mathcal{L}_{\mathrm{NSP}} + \mathcal{L}_{\mathrm{MLM}}.$$

- KnowBERT [Peters et al., 2019] injects one or several entity linkers between top layers of the BERT architecture.
- It optimizes the whole network for three tasks: (1) the masked language model (MLM) task, (2) next sentence prediction (NSP) from the original BERT model, and (3) EL:

$$\mathcal{L}_{\text{KnowBert}} = \mathcal{L}_{\mathrm{NSP}} + \mathcal{L}_{\mathrm{MLM}} + \mathcal{L}_{\text{EL}}.$$

Novel applications: other similar applications

- ERNIE [Zhang et al., 2019] expands the BERT [Devlin et al., 2019] architecture with a knowledgeable encoder (K-Encoder), which fuses contextualized word representations obtained from the underlying self-attention network with entity representations from a pre-trained TransE model [Bordes et al., 2013]:

$$\mathcal{L}_{\text{ERNIE}} = \mathcal{L}_{\mathrm{NSP}} + \mathcal{L}_{\mathrm{MLM}} + \mathcal{L}_{\text{dEA}} .$$

- [Wang et al., 2019] train a disambiguation network using the composition of two losses: regular MLM and a Knowledge Embedding (KE) loss based on the TransE [Bordes et al., 2013] objective for encoding graph structures:

$$\mathcal{L}_{\mathrm{KEPLER}} = \mathcal{L}_{\mathrm{MLM}} + \mathcal{L}_{\mathrm{KE}}.$$

Two main types of evaluation settings

Entity disambiguation evaluation

- **Input**: a text with a set of provided entity mentions.
- **Output**: an entity-linked text.
- The list of candidates can be fixed to ensure a better comparability of the disambiguation models.

End-to-end entity linking evaluation

- **Input**: a raw text
- **Output**: an entity-linked text
- End-to-end evaluation performs mention detection / entity recognition + entity disambiguation)

Common evaluation dataset used to compare entity linking models and perform experiments

Corpus	Text Type	# of Docs	# of Mentions
AIDA-B	News	231	4485
MSNBC	News	20	656
AQUAINT	News	50	727
ACE2004	News	36	257
CWEB	ClueWeb & Wikipedia	320	11154
WW	ClueWeb & Wikipedia	320	6821
TAC KBP 2010	News & Web	1013	1020
TAC KBP 2015 Chinese	News & Forums	166	11066
TAC KBP 2015 Spanish	News & Forums	167	5822

Common evaluation dataset used to compare entity linking models and perform experiments

Corpus	Text Type	# of Docs	# of Mentions
AIDA-B	News	231	4485
MSNBC	News	20	656
AQUAINT	News	50	727
ACE2004	News	36	257
CWEB	ClueWeb & Wikipedia	320	11154
WW	ClueWeb & Wikipedia	320	6821
TAC KBP 2010	News & Web	1013	1020
TAC KBP 2015 Chinese	News & Forums	166	11066
TAC KBP 2015 Spanish	News & Forums	167	5822

- Note that, both evaluation setups can be used with these dataset

Common evaluation dataset used to compare entity linking models and perform experiments

Corpus	Text Type	# of Docs	# of Mentions
AIDA-B	News	231	4485
MSNBC	News	20	656
AQUAINT	News	50	727
ACE2004	News	36	257
CWEB	ClueWeb & Wikipedia	320	11154
WW	ClueWeb & Wikipedia	320	6821
TAC KBP 2010	News & Web	1013	1020
TAC KBP 2015 Chinese	News & Forums	166	11066
TAC KBP 2015 Spanish	News & Forums	167	5822

- Note that, both evaluation setups can be used with these dataset
- ... and even more, e.g. entity typing (predicting “hypernym of an entity”)

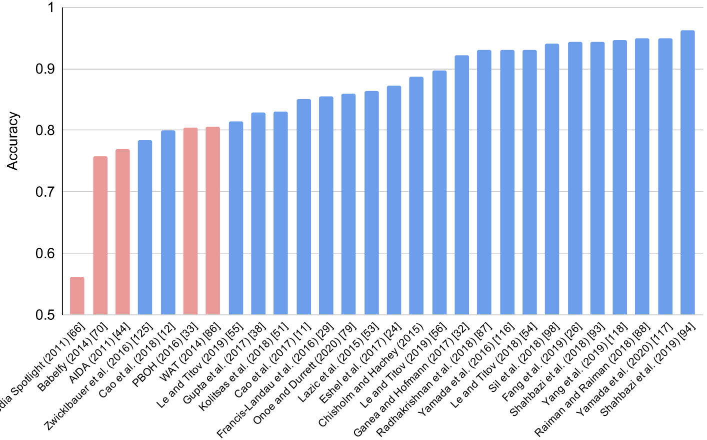
Common evaluation dataset used to compare entity linking models and perform experiments

Corpus	Text Type	# of Docs	# of Mentions
AIDA-B	News	231	4485
MSNBC	News	20	656
AQUAINT	News	50	727
ACE2004	News	36	257
CWEB	ClueWeb & Wikipedia	320	11154
WW	ClueWeb & Wikipedia	320	6821
TAC KBP 2010	News & Web	1013	1020
TAC KBP 2015 Chinese	News & Forums	166	11066
TAC KBP 2015 Spanish	News & Forums	167	5822

- Note that, both evaluation setups can be used with these dataset
- ... and even more, e.g. entity typing (predicting “hypernym of an entity”)
- ... or even the simple entity recognition.

Entity disambiguation: classic vs neural models

- Performance of the best classic entity linking models (red) with the more recent neural models (blue) on the AIDA dataset shows an improvement over 15 points of accuracy.



Entity disambiguation: Sparsity of the evaluation

	AIDA-B	KBP'10	MSNBC	AQUAINT	ACE-2004	CWEB	WW	KBP'15 (es)	KBP'15 (zh)
	Accuracy	Accuracy	Micro F1	Micro F1	Micro F1	Micro F1	Micro F1	Accuracy	Accuracy
Non-Neural Baseline Models									
DBpedia Spotlight (2011) [66]	0.561	-	0.421	0.518	0.539	-	-	-	-
AIDA (2011) [44]	0.770	-	0.746	0.571	0.798	-	-	-	-
Ratinov et al. (2011) [89]	-	-	0.750	0.830	0.820	0.562	0.672	-	-
WAT (2014) [86]	0.805	-	0.788	0.754	0.796	-	-	-	-
Babelfy (2014) [70]	0.758	-	0.762	0.704	0.619	-	-	-	-
Lazic et al. (2015) [53]	0.864	-	-	-	-	-	-	-	-
Chisholm and Hachey (2015) [15]	0.887	-	-	-	-	-	-	-	-
PBOH (2016) [33]	0.804	-	0.861	0.841	0.832	-	-	-	-
Neural Models									
Sun et al. (2015) [102]	-	0.839	-	-	-	-	-	-	-
Tsai and Roth (2016) [104]	-	-	-	-	-	-	-	0.824	0.851
Fang et al. (2016) [25]	-	0.889	0.755	0.852	0.808	-	-	-	-
Yamada et al. (2016) [116]	0.931	0.855	-	-	-	-	-	-	-
Zwicklbauer et al. (2016) [125]	0.784	-	0.911	0.842	0.907	-	-	-	-
Francis-Landau et al. (2016) [29]	0.855	-	-	-	0.899	-	-	-	-
Eshel et al. (2017) [24]	0.873	-	-	-	-	-	-	-	-
Ganea and Hofmann (2017) [32]	0.922	-	0.937	0.885	0.885	0.779	0.775	-	-
Gupta et al. (2017) [38]	0.829	-	-	-	0.907	-	-	-	-
Cao et al. (2017) [11]	0.85	-	-	-	-	-	-	-	-
Sil et al. (2018) [98]	0.940	0.874	-	-	-	-	-	0.823	0.844
Shahbazi et al. (2018) [93]	0.944	0.879	-	-	-	-	-	-	-
Kolitsas et al. (2018) [51]	0.831	-	0.864	0.832	0.855	-	-	-	-
Le and Titov (2018) [54]	0.931	-	0.939	0.884	0.899	0.775	0.780	-	-
Radhakrishnan et al. (2018) [87]	0.930	0.896	-	-	-	-	-	-	-
Cao et al. (2018) [12]	0.800	0.910	-	0.870	0.880	-	0.860	-	-
Raiman and Raiman (2018) [88]	0.949	0.909	-	-	-	-	-	-	-
Upadhyay et al. (2018) [106]	-	-	-	-	-	-	-	0.844	0.860
Gillick et al. (2019) [34]	-	0.870	-	-	-	-	-	-	-
Le and Titov (2019) [55]	0.815	-	-	-	-	-	-	-	-
Le and Titov (2019) [56]	0.897	-	0.922	0.907	0.881	0.782	0.817	-	-
Fang et al. (2019) [26]	0.943	-	0.928	0.875	0.912	0.785	0.828	-	-
Yang et al. (2019) [118]	0.946	-	0.946	0.883	0.901	0.756	0.788	-	-
Shahbazi et al. (2019) [94]	0.962	0.883	-	-	-	-	-	-	-
Onoe and Durrett (2020) [79]	0.859	-	-	-	-	-	-	-	-
Wu et al. (2019) [114]	-	0.940	-	-	-	-	-	-	-

End-to-end evaluation: results of joint ER-ED models on AIDA and MSNBC datasets

	AIDA-B	**MSNBC**
	Micro F1	Micro F1
Non-Neural Baseline Models		
DBpedia Spotlight [Mendes et al., 2011]	0.578	0.406
AIDA [Hoffart et al., 2011]	0.728	0.651
WAT [Piccinno and Ferragina, 2014]	0.730	0.645
Babelfy [Moro et al., 2014]	0.485	0.397
Neural Models		
End-to-end [Kolitsas et al., 2018]	**0.824**	**0.724**
[Martins et al., 2019]	0.819	-
KnowBERT [Peters et al., 2019]	0.744	-

Other types of evaluation

Extrinsic evaluation

- Take an application, e.g. KBQA and measure its performance.

Other types of evaluation

Extrinsic evaluation

- Take an application, e.g. KBQA and measure its performance.
- Compare two entity linkers (A and B) by integration them inside the system in the same way.

Other types of evaluation

Extrinsic evaluation

- Take an application, e.g. KBQA and measure its performance.
- Compare two entity linkers (A and B) by integration them inside the system in the same way.
- If the overall performance of the application improved using linker B then the linker B is better than the original linker A.

Evaluation of separate components

- Entity disambiguation evaluation.

Other types of evaluation

Extrinsic evaluation

- Take an application, e.g. KBQA and measure its performance.
- Compare two entity linkers (A and B) by integration them inside the system in the same way.
- If the overall performance of the application improved using linker B then the linker B is better than the original linker A.

Evaluation of separate components

- Entity disambiguation evaluation.
- Given a set of relevant and irrelevant entity pairs, use entity embeddings to perform the relevancy prediction.

Entity relatedness evaluation

- Reported results for entity relatedness evaluation on the dataset of [Ceccarelli et al., 2013].

	nDCG@1	nDCG@5	nDCG@10	MAP
[Milne and Witten, 2008]	0.540	0.520	0.550	0.480
[Huang et al., 2015]	0.810	0.730	0.740	0.680
[Yamada et al., 2016]	0.590	0.560	0.590	0.520
[Ganea and Hofmann, 2017]	0.632	0.609	0.641	0.578
[Cao et al., 2017]	0.613	0.613	0.654	0.582
[El Vaigh et al., 2019]	0.690	0.640	0.580	-
[Shi et al., 2020]	0.680	0.814	0.820	-

Summary

- Neural entity linking models generally perform the task with higher accuracy than classical methods.

Summary

- Neural entity linking models generally perform the task with higher accuracy than classical methods.
- Generic neural entity linking architecture is applicable for most of the neural EL systems and features:
  - candidate generation
  - mention-context encoding
  - entity encoding
  - entity ranking

Summary

- Neural entity linking models generally perform the task with higher accuracy than classical methods.
- Generic neural entity linking architecture is applicable for most of the neural EL systems and features:
  - candidate generation
  - mention-context encoding
  - entity encoding
  - entity ranking
- The four main modifications of general architecture are:
  - joint entity recognition and linking models
  - global entity linking models
  - domain-independent approaches including zero-shot and distant supervision methods
  - cross-lingual techniques

Future Directions

- End-to-end models featuring the candidate generation step.

Future Directions

- End-to-end models featuring the candidate generation step.
- Further development of zero-shot approaches.

Future Directions

- End-to-end models featuring the candidate generation step.
- Further development of zero-shot approaches.
- More use-cases of EL-enriched language models.
- Integration of EL loss in more neural models.

Thank you! Questions?

References I

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. (2013).
Translating embeddings for modeling multi-relational data.
In *Advances in neural information processing systems*, pages 2787–2795, Stateline, Nevada, USA.

Cao, Y., Huang, L., Ji, H., Chen, X., and Li, J. (2017).
Bridge text and knowledge by learning multi-prototype entity mention embedding.
In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1623–1633, Vancouver, Canada. Association for Computational Linguistics.

References II

Ceccarelli, D., Lucchese, C., Orlando, S., Perego, R., and Trani, S. (2013).
Learning relatedness measures for entity linking.
In *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management*, CIKM '13, pages 139–148, New York, NY, USA. ACM.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019).
BERT: Pre-training of deep bidirectional transformers for language understanding.
In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*

References III

(Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

El Vaigh, C. B., Goasdoué, F., Gravier, G., and Sébillot, P. (2019).
Using knowledge base semantics in context-aware entity linking.
In *Proceedings of the ACM Symposium on Document Engineering 2019*, DocEng '19, New York, NY, USA. ACM.

References IV

Ganea, O.-E. and Hofmann, T. (2017).
Deep joint entity disambiguation with local neural attention.
In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2619–2629, Copenhagen, Denmark. Association for Computational Linguistics.

Hoffart, J., Yosef, M. A., Bordino, I., Fürstenau, H., Pinkal, M., Spaniol, M., Taneva, B., Thater, S., and Weikum, G. (2011).
Robust disambiguation of named entities in text.
In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 782–792. Association for Computational Linguistics.

References V

Huang, H., Heck, L., and Ji, H. (2015).
Leveraging deep neural networks and knowledge graphs for entity disambiguation.
arXiv preprint arXiv:1504.07678.

Kolitsas, N., Ganea, O.-E., and Hofmann, T. (2018).
End-to-end neural entity linking.
In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 519–529, Brussels, Belgium. Association for Computational Linguistics.

References VI

Logeswaran, L., Chang, M.-W., Lee, K., Toutanova, K., Devlin, J., and Lee, H. (2019).
Zero-shot entity linking by reading entity descriptions.
In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3449–3460, Florence, Italy. Association for Computational Linguistics.

Martins, P. H., Marinho, Z., and Martins, A. F. T. (2019).
Joint learning of named entity recognition and entity linking.

In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student*

References VII

Research Workshop, pages 190–196, Florence, Italy. Association for Computational Linguistics.

Mendes, P. N., Jakob, M., García-Silva, A., and Bizer, C. (2011).
Dbpedia spotlight: Shedding light on the web of documents.
In *Proceedings of the 7th International Conference on Semantic Systems*, I-Semantics '11, pages 1–8, New York, NY, USA. ACM.

References VIII

Milne, D. and Witten, I. H. (2008).
Learning to link with Wikipedia.
In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM '08, pages 509–518, New York, NY, USA. ACM.

Moro, A., Raganato, A., and Navigli, R. (2014).
Entity linking meets word sense disambiguation: a unified approach.
Transactions of the Association for Computational Linguistics, 2:231–244.

References IX

- Peters, M. E., Neumann, M., Logan, R., Schwartz, R., Joshi, V., Singh, S., and Smith, N. A. (2019).
  Knowledge enhanced contextual word representations.
  In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 43–54, Hong Kong, China. Association for Computational Linguistics.

- Piccinno, F. and Ferragina, P. (2014).
  From tagme to wat: A new entity annotator.
  In *Proceedings of the First International Workshop on Entity Recognition —& Disambiguation*, ERD '14, pages 55 – 62, New York, NY, USA. Association for Computing Machinery.

References X

- Shi, W., Zhang, S., Zhang, Z., Cheng, H., and Yu, J. X. (2020).
  Joint embedding in named entity linking on sentence level.
  arXiv preprint arXiv:2002.04936.

- Wang, X., Gao, T., Zhu, Z., Liu, Z., Li, J., and Tang, J. (2019).
  Kepler: A unified model for knowledge embedding and pre-trained language representation.
  arXiv preprint arXiv:1911.06136.

- Wu, L., Petroni, F., Josifoski, M., Riedel, S., and Zettlemoyer, L. (2019).
  Zero-shot entity linking with dense entity retrieval.
  arXiv preprint arXiv:1911.03814.

References XI

Yamada, I., Shindo, H., Takeda, H., and Takefuji, Y. (2016).
Joint learning of the embedding of words and entities for named entity disambiguation.
In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 250–259, Berlin, Germany. Association for Computational Linguistics.

Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., and Liu, Q. (2019).
ERNIE: Enhanced language representation with informative entities.
In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages

References XII

1441–1451, Florence, Italy. Association for Computational Linguistics.