Probabilistic Bag-of-Hyperlinks Model for Entity Linking

Octavian Ganea   Marina Ganea   Aurelien Lucchi
Carsten Eickhoff   Thomas Hofmann

Department of Computer Science
ETH Zürich, Switzerland

April 15, 2016

Today's Menu

- Clean and effective solution for entity disambiguation
- State-of-the-art method (on Gerbil platform [Usbeck et al. 2015])
- Expands on the seminal works of
  - [Ferragina et al. 2010]
  - [Hoffart et al. 2011]
  - [Han et al. 2011]
  - [Ratinov et al. 2011]

Entity Linking Stages

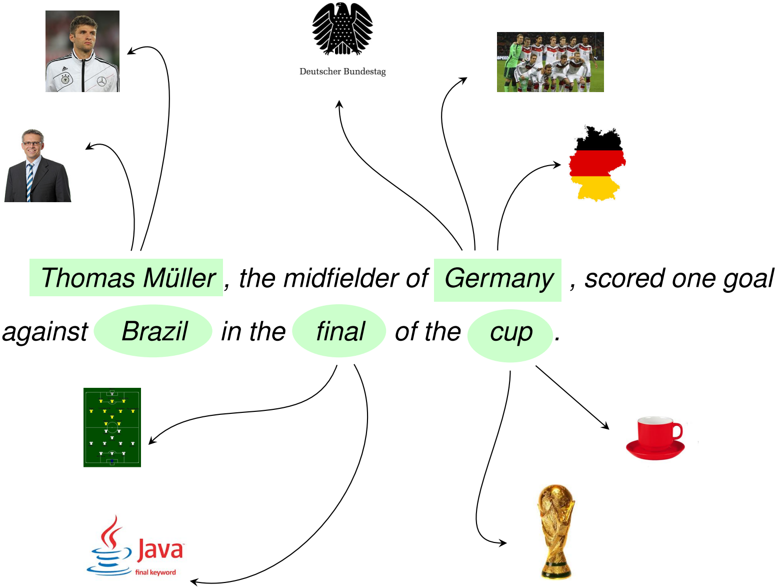
Thomas Müller , the midfielder of Germany , scored one goal against Brazil in the final of the cup.

Entity Linking Stages: Mention Detection (MD)

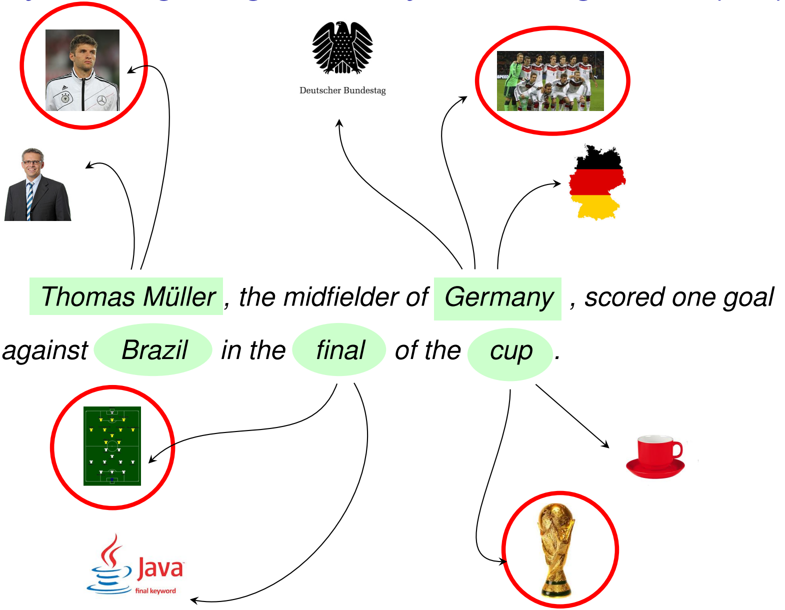
Thomas Müller, the midfielder of Germany, scored one goal against Brazil in the final of the cup.

- Not our focus
- Assume mentions are given as input in all experiments
- Typically found using a NER system

Entity Linking Stages: Candidate Selection



Entity Linking Stages: Entity Disambiguation (ED)

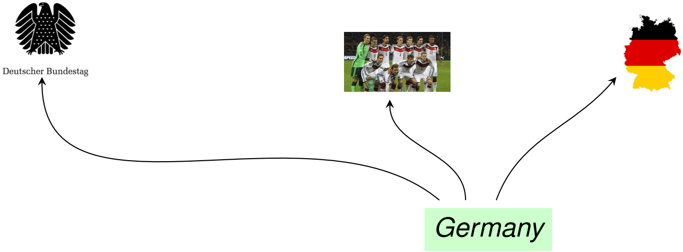


Local Disambiguation

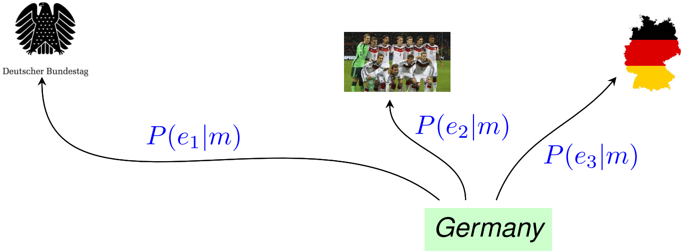
Deutscher Bundestag

Thomas Müller, the midfielder of *Germany* *, scored one goal against Brazil in the final of the cup.*

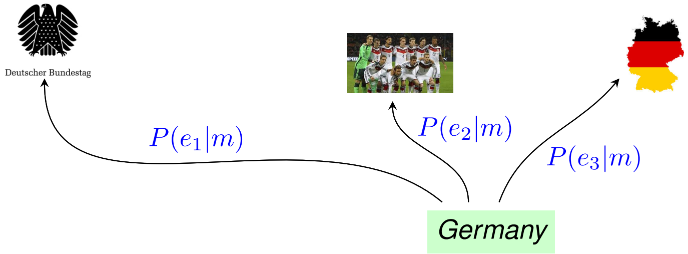
Local Disambiguation: Mention - Entity Compatibility



Local Disambiguation: Mention - Entity Compatibility



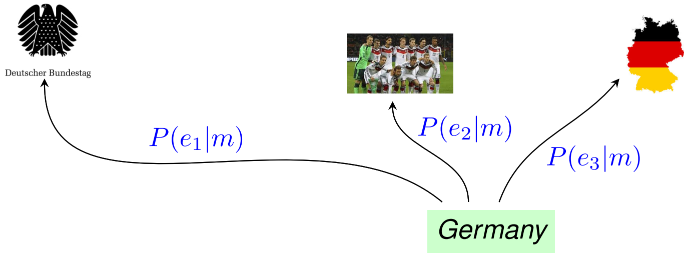
Local Disambiguation: Mention - Entity Compatibility



- Commonness (see e.g. [Milne et al. 2008], [Ferragina et al. 2010])
- Mention-Entity baseline: $e_i^* = \arg\max_{e \in \mathcal{E}} P(e_i|m_i)$
- Estimated from Wikipedia statistics

$$P(e|m) \approx \frac{\text{\# links with } m \text{ that point to } e}{\text{\# links with anchor } m}$$

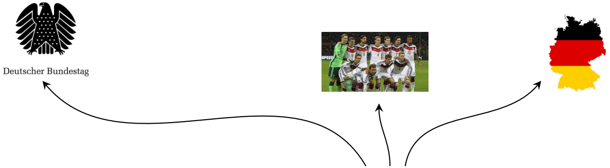
Local Disambiguation: Mention - Entity Compatibility



	CoNLL test A		CoNLL test B	
Baselines	R@MI	R@MA	R@MI	R@MA
Mention-Entity	69.73	69.30	67.98	72.75

Table 1: Accuracy gains of individual *PBoH* components .

Local Disambiguation



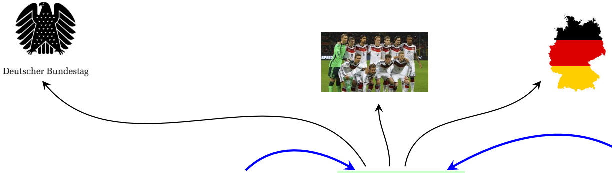
Thomas Müller, the midfielder of Germany, scored one goal against Brazil in the final of the cup.

Local Disambiguation: Surrounding Context

Deutscher Bundestag

Thomas Müller, the midfielder of Germany, scored one goal against Brazil in the final of the cup.

Local Disambiguation: Surrounding Context



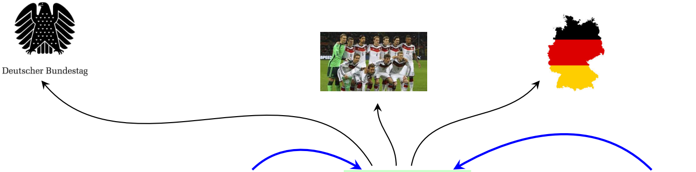
Thomas Müller, the *midfielder* *of* *Germany* *, scored one* *goal* *against Brazil in the final of the cup.*

- Context = text window around mention (say of size 100)
- Similar context models: [Ratinov et al. 2011], [Blanco et al. 2015]
- Bag-of-words context model:

$$P(c|e) = \prod_{w \in c} P(w|e) \qquad P(w|e) \approx \frac{\#\text{times } w \text{ in context of link to } e}{\#\text{words surrounding links to } e}$$

  - Smoothing: absolute discounting with backoff interpolation

Local Disambiguation: Surrounding Context



Thomas Müller, the midfielder of Germany , scored one goal against Brazil in the final of the cup.

- Bayes' rule (w/ conditional indepence assumption: $c \perp\!\!\!\perp m \mid e$)

$$P(e|m,c) \propto P(e|m)P(c|e) = P(e|m) \prod_{w \in c} P(w|e) \quad \text{(Local Context)}$$

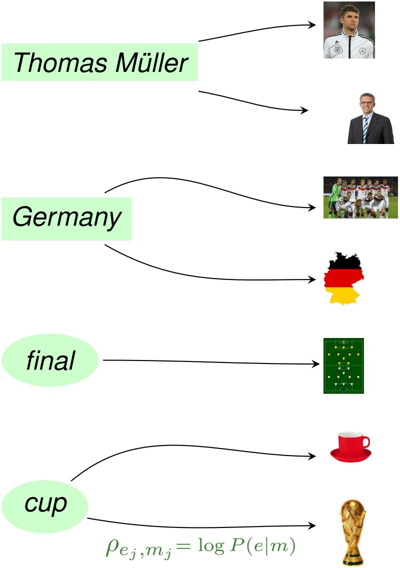
- Incremental accuracy:

	CoNLL test A		CoNLL test B	
Baselines	R@MI	R@MA	R@MI	R@MA
Mention-Entity	69.73	69.30	67.98	72.75
Local Context	82.50	81.56	85.46	84.08

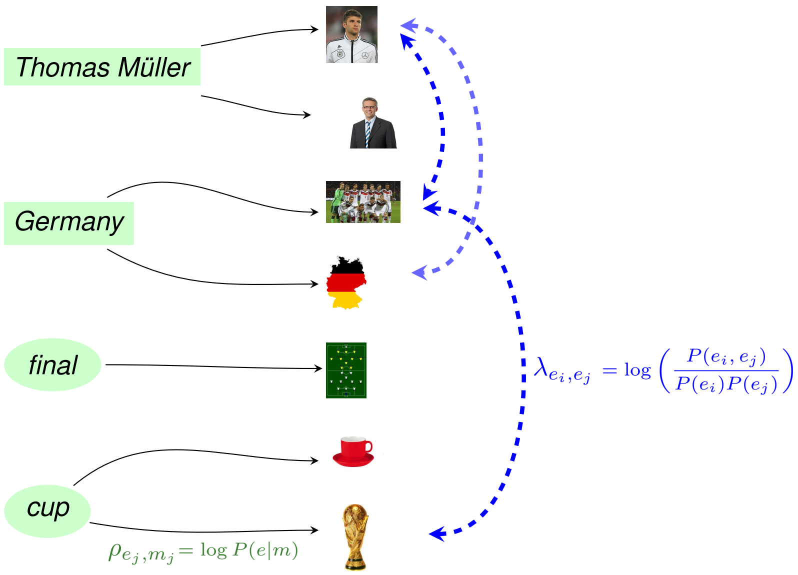
Joint Disambiguation

Thomas Müller, the midfielder of Germany, scored one goal against Brazil in the final of the cup.

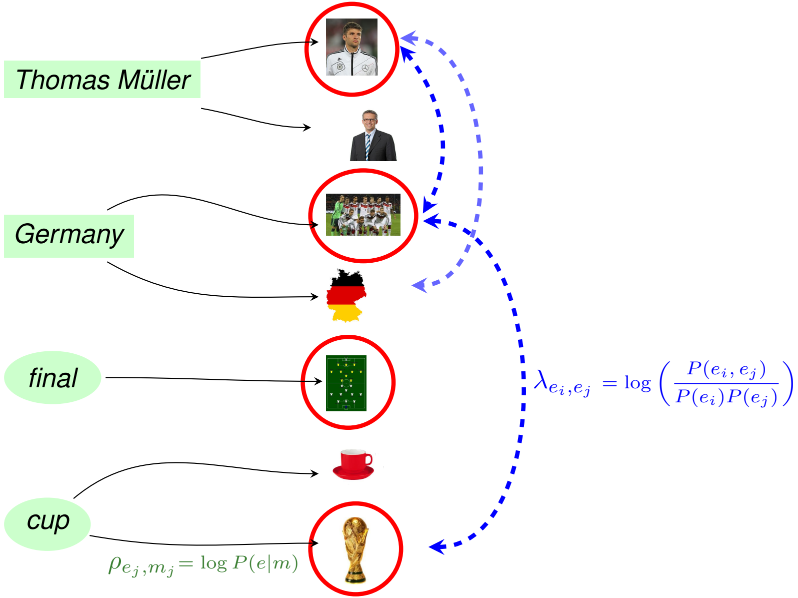
Joint Disambiguation



Joint Disambiguation



Joint Disambiguation



Joint Disambiguation

- Challenge: should not disambiguate single mentions in isolation
- [Han et al. 2011], [Ratinov et al. 2011], [Guo et al. 2014]
- We want to leverage entity - entity co-linking statistics:

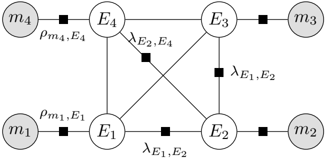
$$P(e, e') \approx \frac{\text{\#articles have links to } e \text{ and } e'}{\text{\#articles}}$$

- Log-linear model:

$$P(\mathbf{e}|\mathbf{m}) = P(e_1 \dots e_n | m_1 \dots m_n) = \frac{1}{Z(\mathbf{m})} \exp \left[\sum_{1 \le i \le n} \rho_{m_i, e_i} + \sum_{1 \le i < j \le n} \lambda_{e_i, e_j} \right]$$

Markov Network and Factor Graph

- Probabilistic model: pairwise Markov Random Field



- Markov assumption: mentions independent given entities.

$$\Rightarrow \quad P(\mathbf{e}|\mathbf{m}) \propto P(\mathbf{e})P(\mathbf{m}|\mathbf{e}) = \underbrace{P(e_1, \ldots, e_n)}_{\text{prior}} \cdot \underbrace{\prod_{i=1}^{n} P(m_i|e_i)}_{\text{likelihood}}$$

- Challenge: estimate $P(\mathbf{e}) = P(e_1, \ldots, e_n)$

Markov Network and Factor Graph

- Plug-in estimators:

$$\left.\begin{aligned} \rho_{m,e} &= \log(P(e|m)), \quad \forall e, m \\ \lambda_{e,e'} &= \log\left(\frac{P(e,e')}{P(e)P(e')}\right), \quad \forall e, e' \end{aligned}\right\} \text{Uncalibrated model}$$

- Related to Bethe free energy (see paper)

- Exact for directed acyclic graphs. Approximation for loopy graphs.

	CoNLL test A		CoNLL test B	
Baselines	R@MI	R@MA	R@MI	R@MA
Mention-Entity	69.73	69.30	67.98	72.75
Local Context	82.50	81.56	85.46	84.08
Uncalibrated	69.77	69.95	75.87	75.12

Calibrated Model

Uncalibrated model:

$$P(\mathbf{e}|\mathbf{m}) = \frac{1}{Z(\mathbf{m})} \exp\left[\underbrace{\sum_{1 \le i \le n} \rho_{m_i, e_i}}_{\substack{\text{likelihood} \\ \text{scales with n}}} + \underbrace{\sum_{1 \le i < j \le n} \lambda_{e_i, e_j}}_{\substack{\text{prior} \\ \text{scales with } \frac{n(n-1)}{2}}} \right]$$

Calibrated Model

Uncalibrated model:

$$P(\mathbf{e}|\mathbf{m}) = \frac{1}{Z(\mathbf{m})} \exp\left[\underbrace{\sum_{1\leq i\leq n} \rho_{m_i,e_i}}_{\substack{\text{likelihood}\\ \text{scales with n}}} + \underbrace{\sum_{1\leq i<j\leq n} \lambda_{e_i,e_j}}_{\substack{\text{prior}\\ \text{scales with } \frac{n(n-1)}{2}}}\right]$$

Calibrated model:

$$P(\mathbf{e}|\mathbf{m}) = \frac{1}{Z(\mathbf{m})} \exp\left[\sum_{1\leq i\leq n} \rho_{m_i,e_i} + \frac{2}{n-1} \cdot \sum_{1\leq i<j\leq n} \lambda_{e_i,e_j}\right]$$

- Based on a combinatorial argument (details in the paper)

Calibrated Model

$$P(\mathbf{e}|\mathbf{m}) = \frac{1}{Z(\mathbf{m})} \exp \left[\sum_{1 \le i \le n} \rho_{m_i,e_i} + \frac{2}{n-1} \cdot \sum_{1 \le i < j \le n} \lambda_{e_i,e_j} \right]$$

where

$$\rho_{m,e} = \log(P(e|m)), \quad \forall e, m$$

$$\lambda_{e,e'} = \log \left(\frac{P(e,e')}{P(e)P(e')} \right), \quad \forall e, e'$$

	CoNLL test A		CoNLL test B	
Baselines	R@MI	R@MA	R@MI	R@MA
Mention-Entity	69.73	69.30	67.98	72.75
Local Context	82.50	81.56	85.46	84.08
Uncalibrated	69.77	69.95	75.87	75.12
Calibrated	75.09	74.25	74.76	78.28

Final Model

- Introduce parameters ζ and τ to control the importance of the context factors and of the entity-entity interaction factors.
- *PBoH* model (**P**robabilistic **B**ag **o**f **H**yperlinks):

$$\log P(\mathbf{e}|\mathbf{m},\mathbf{c}) \propto \sum_{i=1}^{n} \log P(e_i|m_i) + \qquad \text{mention - entity compatibility}$$

$$\zeta \sum_{i=1}^{n} \sum_{w_j \in c_i} \log P(w_j|e_i) + \qquad \text{context - entity interactions}$$

$$\frac{2\tau}{n-1} \sum_{i<j} \log \left(\frac{P(e_i, e_j)}{P(e_i)\,P(e_j)} \right) \qquad \text{entity - entity coherence}$$

Incremental accuracy

	Datasets			
	CoNLL test A		CoNLL test B	
Baselines	R@MI	R@MA	R@MI	R@MA
Mention-Entity	69.73	69.30	67.98	72.75
Uncalibrated	69.77	69.95	75.87	75.12
Calibrated	75.09	74.25	74.76	78.28
Local Context	82.50	81.56	85.46	84.08
PBoH	**85.53**	**85.09**	**87.51**	**86.39**

Table 1: Accuracy gains of individual *PBoH* components.

Inference & Learning

- MAP inference: $\mathbf{e}^* = \arg\max_{\mathbf{e} \in \mathcal{E}^n} P(\mathbf{e}|\mathbf{m}, \mathbf{c})$

- Exact inference - intractable; resort to approximate inference : loopy belief propagation (similar to [Ferragina et al. 2010])

- Fast empirical convergence (typically $<$ 3 iterations, $\sim$400ms/doc)

- Learning all pairwise parameters $\boldsymbol{\rho}$, $\boldsymbol{\lambda}$ was not successful

- Entity candidate pruning:
  - First, top 64 entities based on $P(e|m)$
  - Then, keep only top 10 based on $P(e|m, c)$

Experiments

F1@MI **F1@MA**	ACE2004	CoNLL-Comp.	CoNLL-Test A	CoNLL-Test B	CoNLL-Train	AQUAINT	DBpediaSpotl.	IITB	KORE50	μposts14-Test	μposts14-Train	MSNBC	N3-Reuters-128	N3-RSS-500
AGDISTIS	65.83 77.63	60.27 56.97	59.06 53.36	58.32 58.03	61.05 57.53	60.10 58.62	36.61 33.25	41.23 43.38	34.16 30.20	42.43 61.08	50.39 62.87	75.42 73.82	67.95 75.52	59.88 70.80
Babelfy	63.20 76.71	78.00 73.81	75.77 71.26	80.36 74.52	78.01 74.22	72.27 73.23	51.05 51.97	57.13 55.36	73.12 69.77	47.20 62.11	50.60 61.02	78.17 75.73	58.61 59.87	69.17 76.00
DBpediaSpotlight	70.38 80.02	58.84 60.59	54.90 54.11	57.69 61.34	60.04 62.23	74.03 73.13	69.27 67.23	65.44 62.81	37.59 32.90	56.43 71.63	56.26 67.99	69.27 69.82	56.44 58.77	57.63 65.03
Dexter	18.72 16.97	48.46 45.29	45.44 42.17	48.59 46.20	49.25 45.85	38.28 38.15	26.70 22.75	28.53 28.48	17.20 12.54	31.27 44.02	35.21 42.07	36.86 39.42	32.74 31.85	31.11 33.55
EntityClassifier.eu	12.74 12.3	46.6 42.86	44.13 42.36	44.02 41.31	47.83 43.36	21.67 19.59	22.59 18.0	18.46 19.54	27.97 25.2	29.12 39.53	32.69 38.41	41.24 40.3	28.4 24.84	21.77 22.2
Kea	80.08 87.57	73.39 73.26	70.9 67.91	72.64 73.31	74.22 74.47	81.84 81.27	73.63 76.60	72.03 70.52	57.95 53.17	63.4 76.54	64.67 74.32	85.49 87.4	63.2 64.45	69.29 75.93
NERD-ML	54.89 72.22	54.62 52.35	52.85 49.6	52.59 51.34	55.55 53.23	49.68 46.06	46.8 45.59	51.08 49.91	29.96 24.75	38.65 57.91	39.83 53.74	64.03 67.28	54.96 62.9	61.22 67.3
TagMe 2	81.93 89.09	72.07 71.19	69.07 66.5	70.62 70.38	73.2 72.45	76.27 75.12	63.31 65.1	57.23 55.8	57.34 54.67	56.81 71.66	59.14 70.45	75.96 77.05	59.32 67.55	78.05 83.2
WAT	80.0 86.49	83.82 83.59	81.82 80.25	84.34 84.12	84.21 84.22	76.82 77.64	65.18 68.24	61.14 59.36	58.99 53.13	59.56 73.89	61.96 72.65	77.72 79.08	64.38 65.81	68.21 76.0
Wikipedia Miner	77.14 86.36	64.72 66.17	61.65 61.67	60.71 63.19	66.48 67.93	75.96 74.63	62.57 61.43	58.59 56.98	41.63 35.0	54.88 69.29	55.93 67.0	64.25 64.68	60.05 66.51	64.54 72.23
PBoH	87.19 90.40	86.72 86.85	86.63 85.48	87.39 86.32	86.59 87.30	86.64 86.14	79.48 80.13	62.47 61.04	61.70 55.83	74.19 84.48	73.08 81.25	89.54 89.62	76.54 83.31	71.24 78.33

Table 2: Micro and macro F1 scores on the Gerbil[1] (v1.1.4) platform. We highlight the best system and the 2^{nd} best system.

[1] [Usbeck et al. 2015]

Conclusions

- Presented a new state of the art entity disambiguation system
- Light-weight probabilistic model based on simple data statistics - scalable to massive amounts of data.
- Plug-in parameter estimators
- Loopy belief propagation inference technique
- Very good generalization performance across many datasets
- Future work:
  - Alleviate data sparseness using low-dimensional entity vector representations
  - Joint MD and ED
- Code soon : `github.com/dalab/pboh-entity-linking`

References

Ferragina, Paolo and Scaiella, Ugo (2010)
Tagme: on-the-fly annotation of short text fragments (by wikipedia entities)

Usbeck, Ricardo et al. (2015)
GERBIL: General Entity Annotator Benchmarking Framework

Blanco, Roi et al. (2015)
Fast and Space-Efficient Entity Linking in Queries

Hoffart, Johannes et al. (2011)
Robust disambiguation of named entities in text

Han, Xianpei et al. (2011)
Collective entity linking in web text: a graph-based method

Guo, Zhaochen and Barbosa, Denilson (2014)
Robust Entity Linking via Random Walks

Milne, David and Witten, Ian H (2008)
Learning to link with wikipedia

Ratinov, Lev et al. (2011)
Local and global algorithms for disambiguation to wikipedia

Thank you!