Boosting Entity Linking Performance by Leveraging Unlabeled Documents

Phong Le[1] and **Ivan Titov**[1,2]
[1]University of Edinburgh [2]University of Amsterdam
lephong.xyz@gmail.com ititov@inf.ed.ac.uk

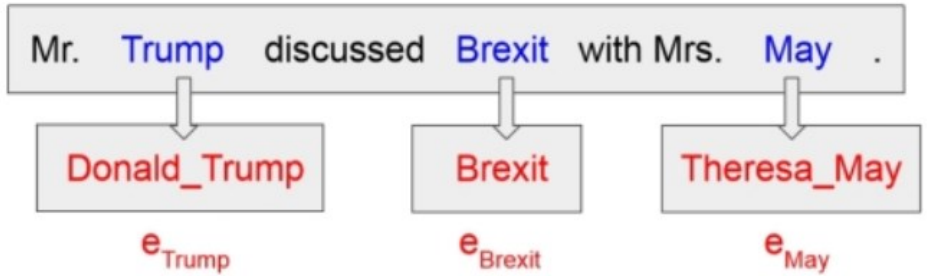
読み手：浅田真生（豊田工業大学）

2019/9/11

図表は論文より引用

背景

- Entity Linkingでは，人手による
  アノテーションを用いた手法が高精度
  - アノテーションのコスト大
  - ドメイン・言語依存大



- Wikipediaのみを用いて学習データをつくることは難しい
  - Wikipediaの説明はrigidな文である
  - 一度リンクされたメンションは再度リンクされない

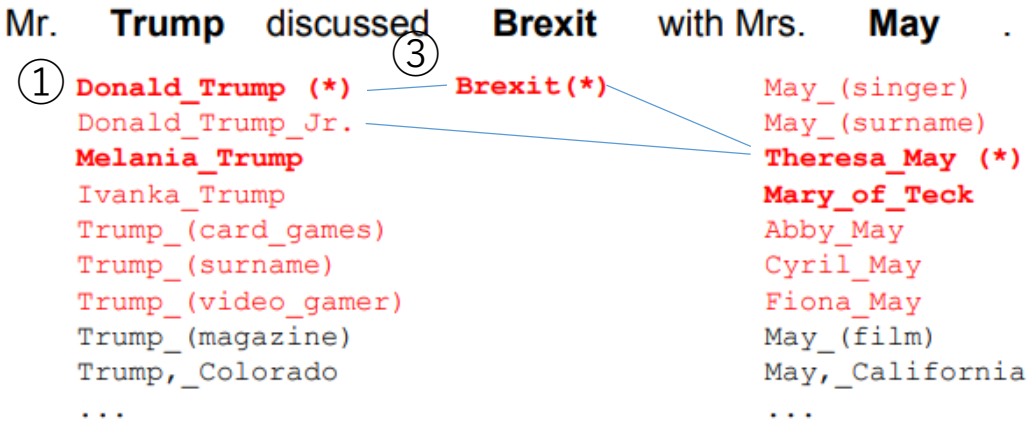
- Wikipediaを用いてUnlabeled-documentにラベル付けをしたい

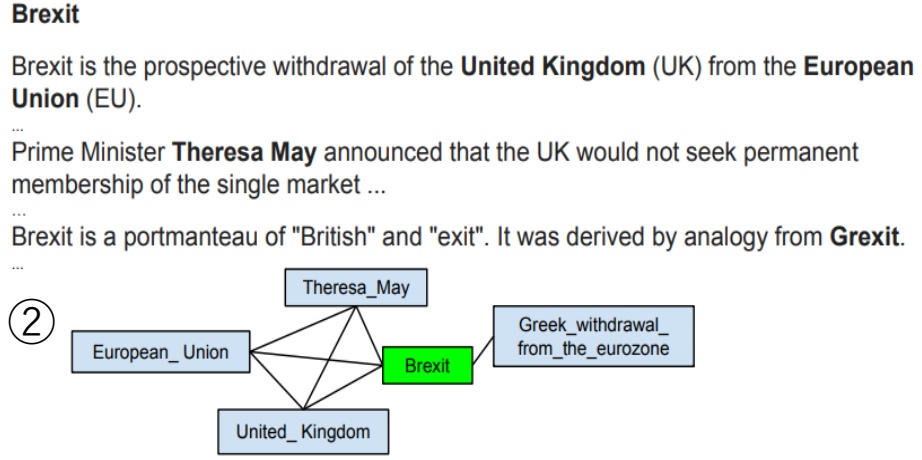
アイディア

① メンションの共起頻度から
エンティティ候補を作成

② Wikipediaのアンカーをもとに
Link Graphを作成

③ Graphのリンクが多いエンティティ
を正例，少ないエンティティを負例
としてラベル付け





提案手法：エンティティ候補作成

- 2つのスコアから，エンティティ候補を作成
  - $p_{wiki}(e|m)$：共起頻度
    - メンションがWikipediaエンティティ中のアンカーテキストとして使用される頻度
  - $q_{wiki}(e|m,c)$：メンション，エンティティembeddingの類似度

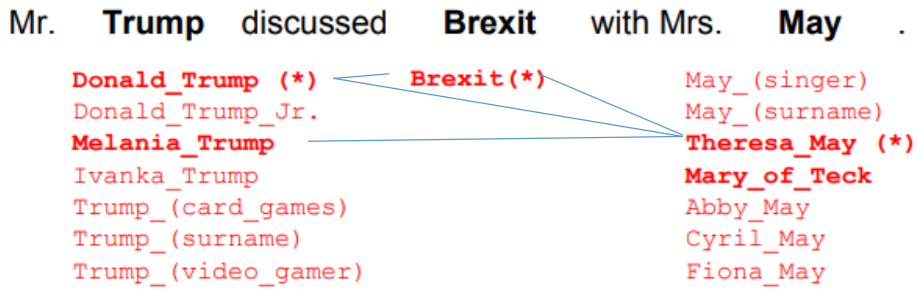
$$q_{wiki}(e|m,c) \propto \exp\{\mathbf{x}_e^T \sum_{w \in (m,c)} \mathbf{x}_w\}$$

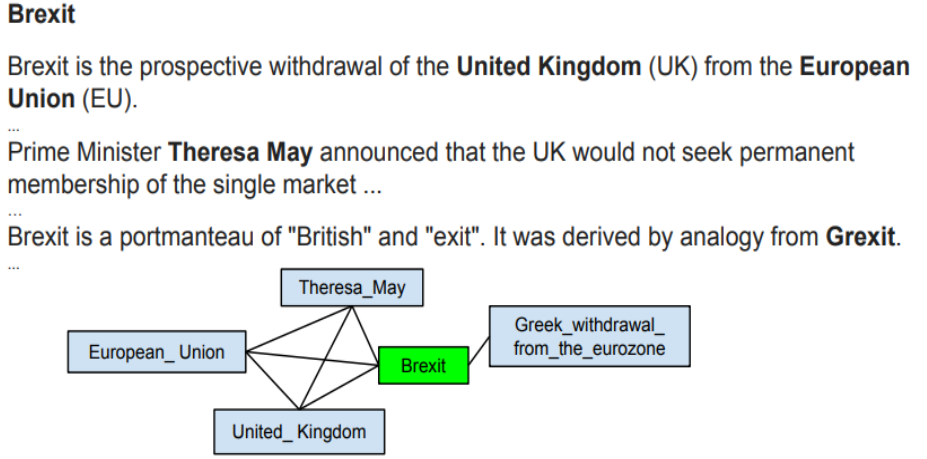
  - p_{wiki}スコアの上位4件とq_{wiki}スコアの上位3件を正例および負例の候補とする

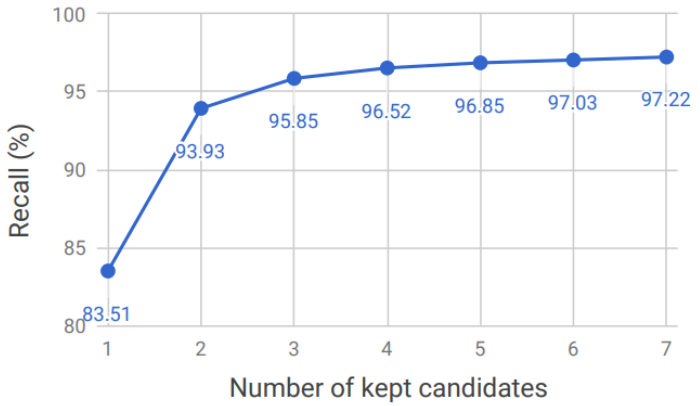
提案手法：Link Graph作成

- 以下の2つのルールでWikipediaグラフを作成
  - D_{wiki} is a Wikipedia article describing e_u, and e_v appears in it, or
  - D_{wiki} contains e_u, e_v and there are less than l entities between them.

- p_{wiki}, q_{wiki}スコア上位7件のうち，リンク数が多い上位2件を正例，その他を負例とする







提案手法：学習

- エンティティ候補ごとにスコアを算出

$$s(e_i|D) = \phi(e_i|c_i, m_i) + \sum_{j \neq i} \alpha_{ij} \max_{e_j \in E_j^+} \xi(e_i, e_j)$$

local score

global score

Ganea and Hofmann (2017) と同様

$$\xi(e_i, e_j) = \mathbf{x}_{e_i}^T \mathbf{R} \mathbf{x}_{e_j}$$

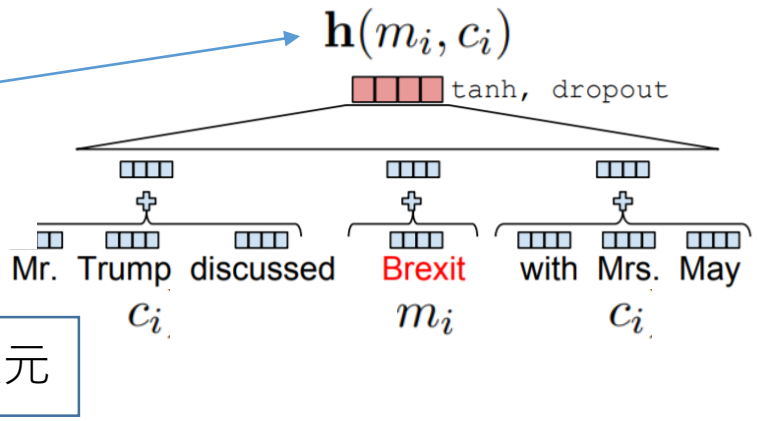
エンティティ embedding

パラメータ

$$\alpha_{ij} \propto \exp\left\{\mathbf{h}(m_i, c_i)^T \mathbf{A} \mathbf{h}(m_j, c_j) / \sqrt{d_c}\right\}$$

パラメータ

hの次元



- 目的関数

$$L(\Theta) = \sum_D \sum_{m_i} \left[\delta + \max_{e_i^- \in E_i^-} \hat{s}(e_i^-|D) - \max_{e_i^+ \in E_i^+} \hat{s}(e_i^+|D)\right]_+$$

E_i^+：正例のエンティティ候補群
E_i^-：負例のエンティティ候補群

実験設定

- Embedding
  - Wikipedia embedding : DeepEd
  - Word embedding : GloVe
- Dataset
  - Train set
    - Randomly selected 30,000 unlabeled documents from RCV1
    - SpaCyによりエンティティ抽出
  - Test sets
    - AIDA CoNLL’testb’, MSNBC, AQUAINT, ACE2004, CWEB, WIKI
    - 評価指標：Wikipediaとリンクしたメンションのみを対象としたmicro-F値

結果

Methods	AIDA-B	MSNBC	AQUAINT	ACE2004	CWEB	WIKI	Avg
Wikipedia							
(Milne and Witten, 2008)	-	78	85	81	64.1	**81.7**	77.96
(Ratinov et al., 2011a)	-	75	83	82	56.2	67.2	72.68
(Hoffart et al., 2011)	-	79	56	80	58.6	63	67.32
(Cheng and Roth, 2013)	-	90	90	86	67.5	73.4	81.38
(Chisholm and Hachey, 2015)	84.9	-	-	-	-	-	-
Wiki + unlab							
(Lazic et al., 2015)	86.4	-	-	-	-	-	
Our model	**89.66** ±0.16	**92.2** ±0.2	**90.7** ±0.2	**88.1** ±0.0	**78.2** ±0.2	**81.7** ±0.1	**86.18**
Wiki + Extra supervision							
(Chisholm and Hachey, 2015)	88.7	-	-	-	-	-	-
Fully-supervised (Wiki + AIDA CoNLL train)							
(Guo and Barbosa, 2016)	89.0	92	87	88	77	84.5	85.7
(Globerson et al., 2016)	91.0	-	-	-	-	-	-
(Yamada et al., 2016)	91.5	-	-	-	-	-	-
(Ganea and Hofmann, 2017)	92.22 ±0.14	93.7 ±0.1	88.5 ±0.4	88.5 ±0.3	77.9 ±0.1	77.5 ±0.1	85.22
(Le and Titov, 2018)	93.07 ±0.27	93.9 ±0.2	88.3 ±0.6	89.9 ±0.8	77.5 ±0.1	78.0 ±0.1	85.5

提案手法

- Wikipediaベースの手法より高精度
- Avgでfully-supervisedな手法より高精度

まとめ

- Weakly-supervisedなエンティティリンキング手法の提案
- Wikipediaを用いて自動ラベル付けした生文を用いて学習
- Unlabeled dataを用いないWikipedia-based手法と比較して高い精度 fully-supervised手法を（タスクによっては）上回る精度
- 今後は，人手によるラベル付きデータと組み合わせる手法を検討

感想

- Link Graphを用いたDisambiguationは妥当か
- 言語・ドメイン-specificでないことをウリにしているが Wikipediaが使えない言語・ドメインでは全く使えない

4.4 Analysis and ablations

- constraint-driven learningは有効か

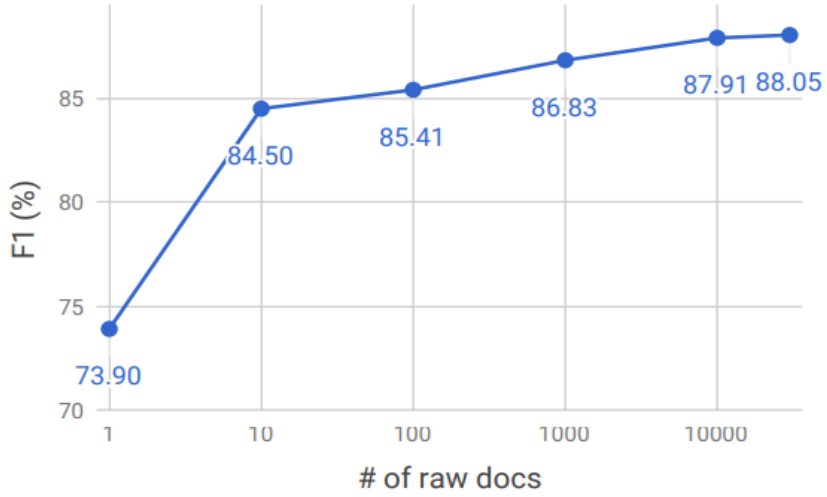
Our model	AIDA-A	AIDA-B	Avg
weakly-supervised	88.05	89.66	86.18
fully-supervised			
on Wikipedia	87.23	87.83	85.84
on AIDA CoNLL	91.34	91.87	84.55

- document-level disambiguationは有効か
  local and global disambiguationは有効か

Model	AIDA-A
Our model	88.05
without local	82.41
without attention	86.82
No disambiguation model (s_c)	86.42

4.4 Analysis and ablations

- unlabeled documentsはどの程度必要か



Unlabeledテキストを増加させると精度は上がる
10,000件以降の精度上昇は緩やか

- リンクさせやすいエンティティは何か

Type	Our model	Fully-supervised learning on AIDA CoNLL
LOC	85.53	89.41
MISC	75.71	83.27
ORG	89.51	92.70
PER	97.20	97.73

AIDA, CoNLLデータセットにおけるNERタグごとに見たlinkingの精度
PERタグの精度が高い