

Duality of Link Prediction and Entailment Graph Induction

Mohammad Javad Hosseini^{*§} Shay B. Cohen^{*} Mark Johnson[‡] and Mark Steedman^{*}

^{*}University of Edinburgh [§]The Alan Turing Institute, UK [‡]Macquarie University

javad.hosseini@ed.ac.uk, scohen@inf.ed.ac.uk
mark.johnson@mq.edu.au, steedman@inf.ed.ac.uk

ACL2019

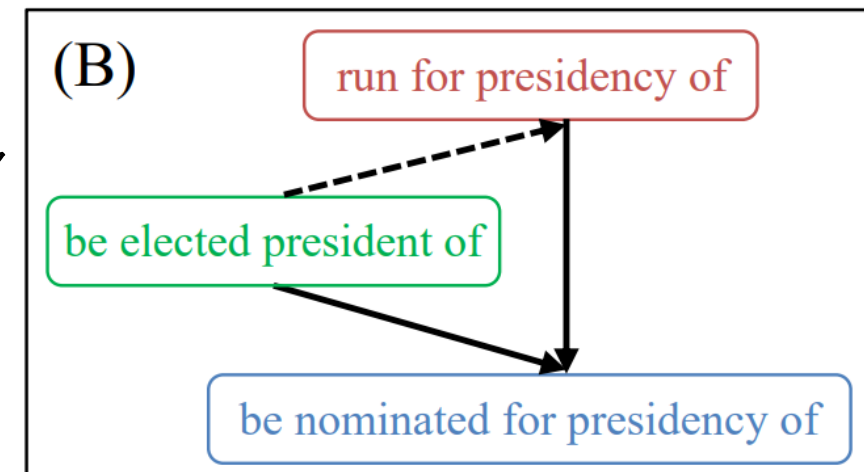
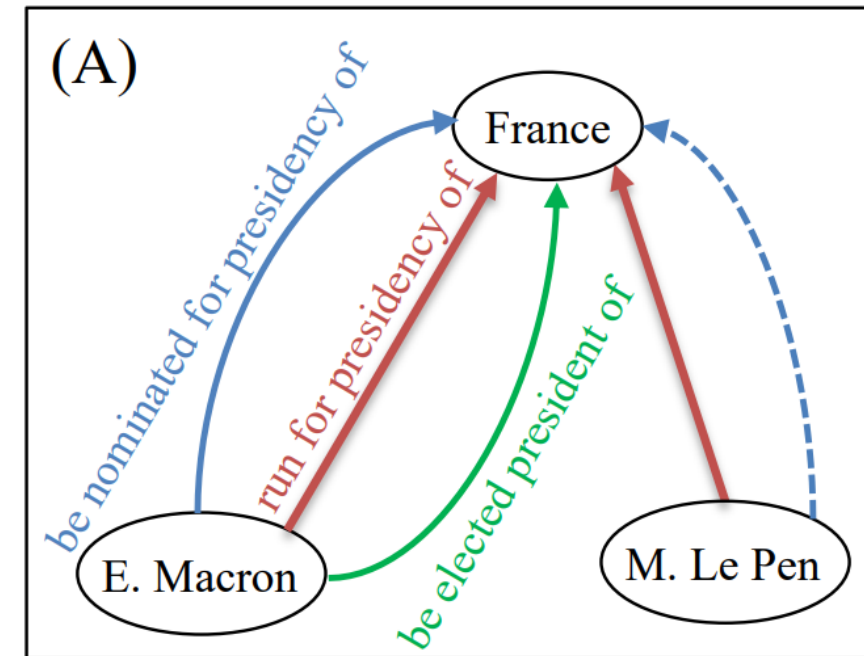
読み手：浅田真生（豊田工業大学）

2019/9/27

図は論文より引用

概要

- Link Prediction (A)
- Entailment Graph Induction (B)
- 2つのタスクは通常異なるタスクとして扱われるが、相関がある
 1. Graph entailmentにLink predictionのスコアを利用し精度向上
 2. 1で得たentailmentスコアを利用し, Link predictionの精度向上



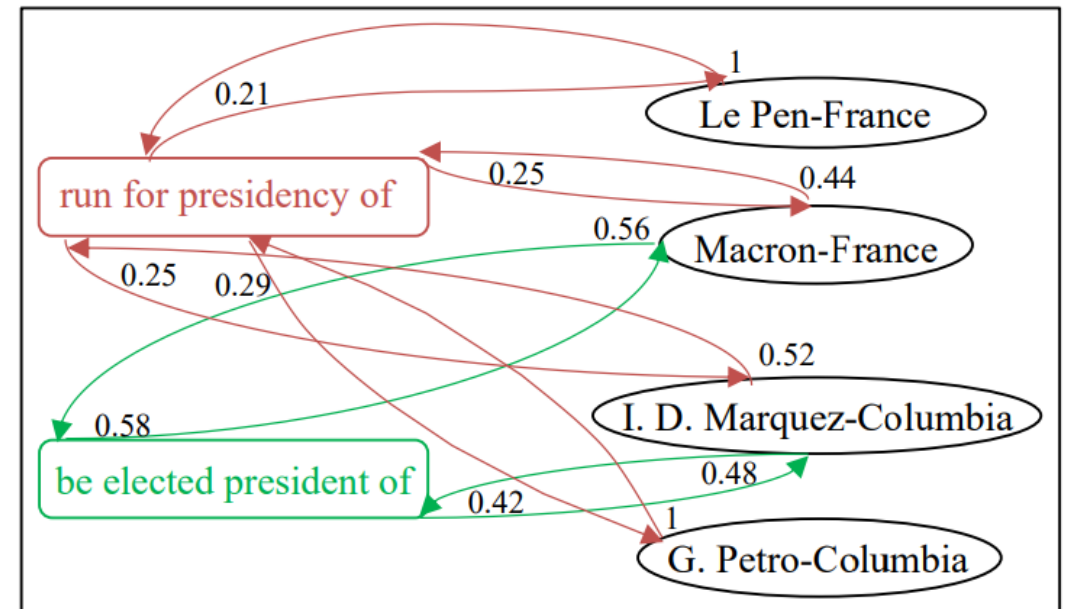
Link predictionを用いたEntailmentスコア

- Knowledge graph上で予めノードとエッジのembeddingを計算
(本手法ではConvE [Dettmers+ 2018] を使用)
- Relationとentityペアで二部
グラフを作成

$$P(\langle e_1, e_2 \rangle | \langle r \rangle) = \frac{P(X_{r,e_1,e_2}=1)}{\sum_{e_1, e_2 \in \mathcal{E}^2} P(X_{r,e_1,e_2}=1)}$$

$$P(\langle r \rangle | \langle e_1, e_2 \rangle) = \frac{P(X_{r,e_1,e_2}=1)}{\sum_{r \in \mathcal{R}} P(X_{r,e_1,e_2}=1)}$$

$$P(\langle q \rangle | \langle r \rangle) = \sum_{e_1, e_2 \in \mathcal{E}^2} P(\langle q \rangle | \langle e_1, e_2 \rangle) P(\langle e_1, e_2 \rangle | \langle r \rangle)$$



Markov Chain

Entailmentスコアを用いたLink prediction

- Distributional Inclusion Hypothesis (DIH) を仮定
 - $r \rightarrow q \Leftrightarrow r$ が使われうるすべてのコンテキストで q も使われる
- これをrelationに適用

$$r \rightarrow q \implies \forall (e_1, e_2) \in \mathcal{E}^2 : X_{r,e_1,e_2} = 1 \rightarrow X_{q,e_1,e_2} = 1$$
$$\implies X_{r,e_1,e_2} \leq X_{q,e_1,e_2}.$$

つまり, $r \rightarrow q$ ならば $P(X_{r,e_1,e_2} = 1) \leq P(X_{q,e_1,e_2} = 1)$ として,
トリプル (q, e_1, e_2) のスコアを以下のように置き換える

$$S_{q,e_1,e_2}^{sent} = \max_{r \in \mathcal{R}: r \rightarrow q} S_{r,e_1,e_2}$$

$$\text{ここで } S_{r,e_1,e_2} = P(X_{r,e_1,e_2} = 1)$$

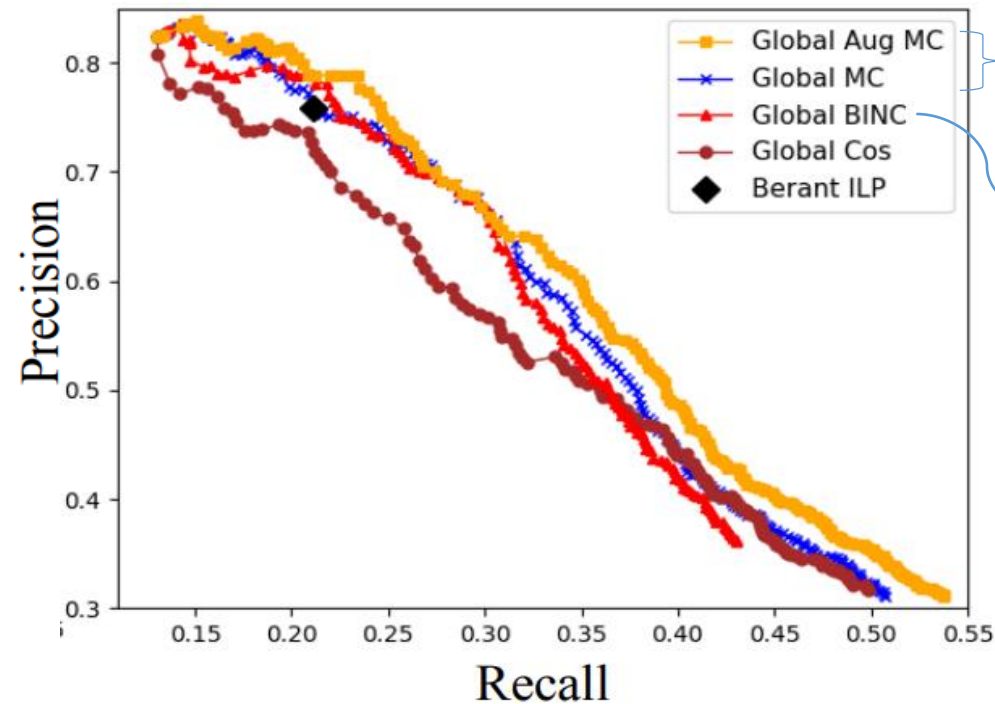
実際は

$$S_{q,e_1,e_2}^{sent} = \max \left(S_{q,e_1,e_2}, \sum_{r \in \mathcal{R}} W'_{r,q} S_{r,e_1,e_2} \right)$$

実験設定

- Knowledge graph
 - NewsSpikeコーパスから, Hosseini+ 2018がトリプルを作成したものを使用 (3.9M relations)
 - 95%をTrain, 4%をDevelopとし, ConvEでembeddingを学習, チューニング
- 評価データ
 - Entailmentタスク
 - Levy/Holt's dataset [Levy and Dagan 2016, Holt 2018]
 - 18,407 examples (3,916 positive, 14,491 negative)
 - Link predictionタスク
 - Knowledge graphの1%

結果：Entailment graph induction



提案手法 (Markov Chain; MC)
AugはKB上にないトリプルを使用

BOWベースでsotaな手法

Link predictionを用いた
提案手法が高い精度

PR曲線のAUC

SBOW

Link Prediction

Weeds	Lin	BInc	Cos	MC	Aug MC
.147	.149	.165	.150	.174	.187

提案手法

結果：Link prediction

	Hits@1	Hits@10	MR	MRR
	ALL entities			
ConvE	20.36	47.93	1999.29	29.58
+ Global MC	20.68	49.13	1012.54	30.19
+ Global Aug MC	20.64	49.16	987.13	30.19
	INFREQUENT entities			
ConvE	19.05	45.59	2124.71	27.94
+ Global MC	19.29	46.60	1154.28	28.41
+ Global Aug MC	19.28	46.66	1118.09	28.43

- ConvEにGraph Entailmentの情報を加えて精度向上

解析

Target Triple	Alternative Triple
John Kerry nominee for secretary of state	John Kerry confirmed as secretary of state
Lady Gaga canceled performance in Hamilton	Lady Gaga canceled show in Hamilton
Dave Toub considers anyone from Jon Gruden	Dave Toub considers everyone from Jon Gruden
Zeke Spruill traded in exchange for Justin Upton	Justin Upton sent in return for Zeke Spruill

- ターゲットのトリプルではスコアが低かった (<0.05) が, entailするrelationのトリプルでのスコアが高かった (>0.95) ため, 正しくリンクできた例

まとめ

- Link predictionとEntailment graph inductionは補完的なタスク
- Entailmentタスクで、link predictionのスコアを利用し、高い精度
- Link predictionタスクで、Entailmentスコアを利用し、精度向上
- 二つのタスクを相互に利用することが有用であると示した