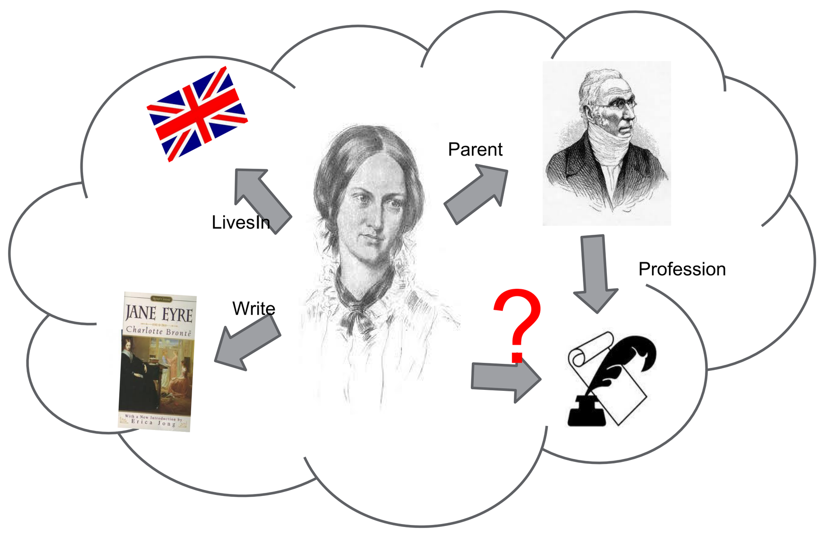
Learning Relational Features with Backward Random Walks

Ni Lao, Google Inc., nlao@google.com
Einat Minkov, University of Haifa, einatm@is.haifa.ac.il
William W. Cohen, Carnegie Mellon University, wcohen@cs.cmu.edu

Knowledge Base Inference

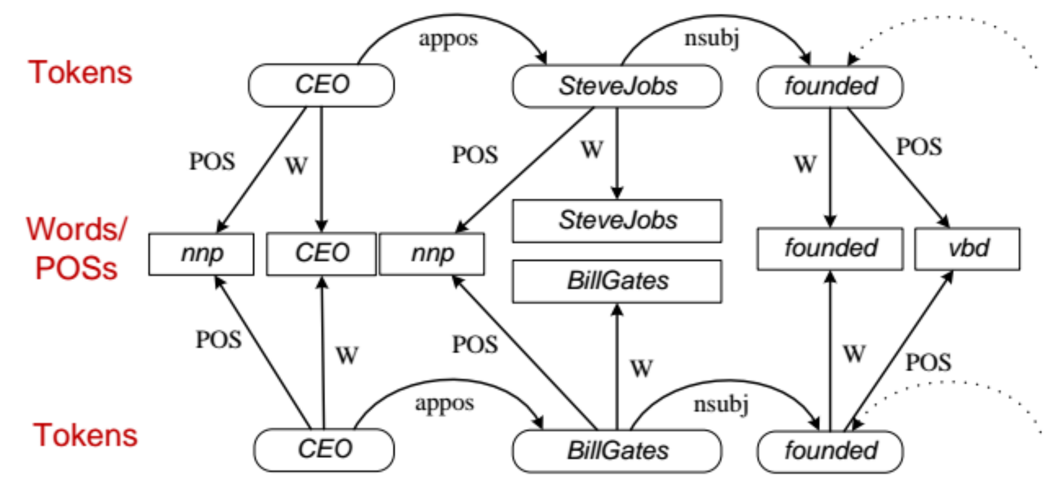


Example Inference rules

$$AthletePlaysForTeam(s, z) \wedge TeamPlaysInLeague(z, t) \rightarrow AthletePlaysForLeague(s, t)$$

Constant path	Interpretation
r=athletePlaysInLeague	
$P(mlb \rightarrow t; \phi)$	Bias toward *MLB*.
$P(boston\_braves \rightarrow t; \langle athletePlaysForTeam^{-1}, athletePlaysInLeague \rangle)$	The leagues played by *Boston Braves* university team members.
r=competesWith	
$P(google \rightarrow t; \phi)$	Bias toward *Google*.
$P(google \rightarrow t; \langle competesWith, competesWith \rangle)$	Companies which compete with Google's competitors.
r=teamPlaysInLeague	
$P(ncaa \rightarrow t; \phi)$	Bias toward *NCAA*.
$P(boise\_state \rightarrow t; \langle teamPlaysInLeague \rangle)$	The leagues played by *Boise State* university teams.

Coordinate Term Extraction

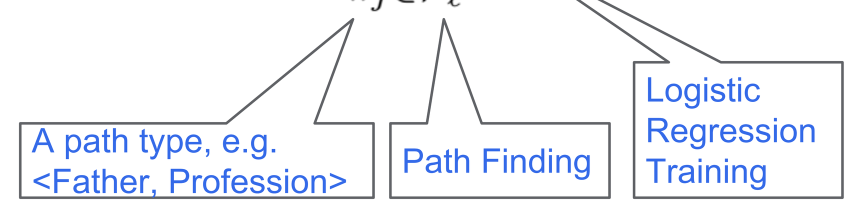


$$P(s \rightarrow t; W^{-1}, conj\_and^{-1}, W, W^{-1}, conj\_and, W)$$
$$P(s \rightarrow t; W^{-1}, nn, W, W^{-1}, appos^{-1}, W)$$
$$P(s \rightarrow t; W^{-1}, appos, W, W^{-1}, appos^{-1}, W)$$

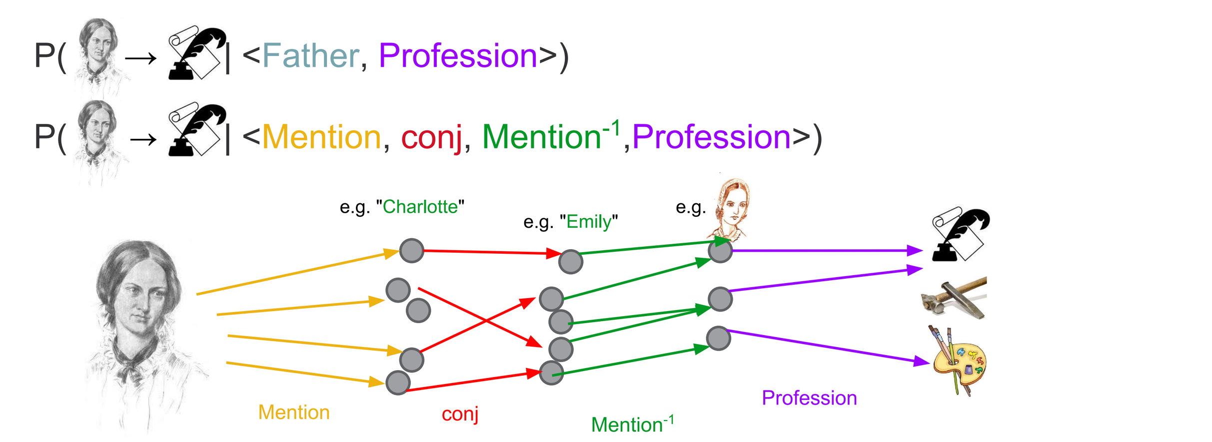
Constant path	Interpretation
$P(said \leftarrow t; W^{-1}, nsubj, W)$ $P(says \leftarrow t; W^{-1}, nsubj, W)$	The subjects of 'said' or 'say' are likely to be a person name.
$P(vbg \leftarrow t; POS^{-1}, nsubj, W)$ $P(nnp \leftarrow t; POS^{-1}, W)$ $P(nn \leftarrow t; POS^{-1}, appos^{-1}, W)$ $P(nn \leftarrow t; POS^{-1}, poss, W)$	Subjects, proper nouns, and nouns with apposition or possessive constructions, are likely to be person names.

Path Ranking Algorithm

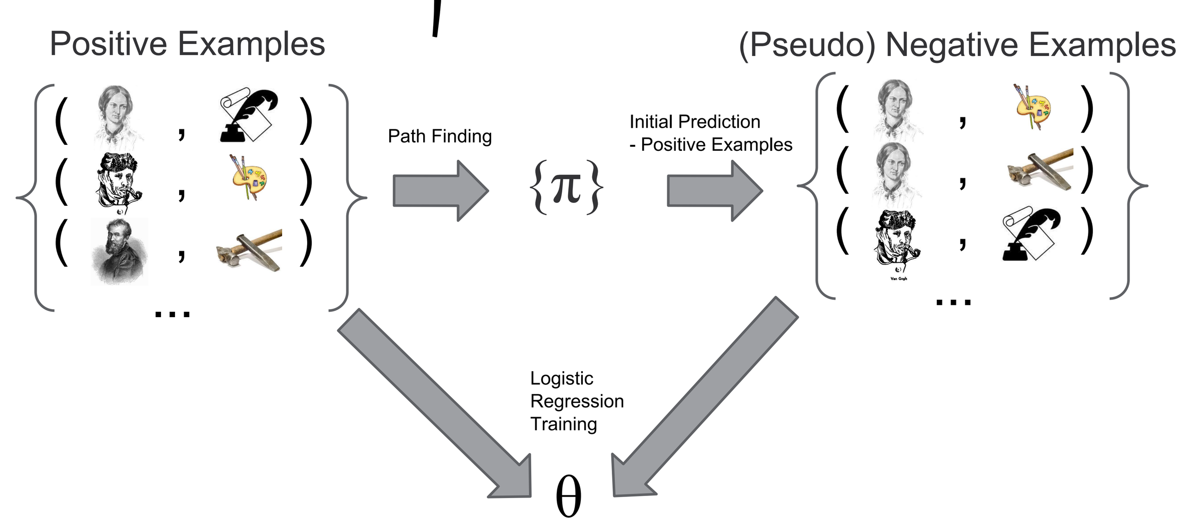
$$score(s, t) = \sum_{\pi_j \in \mathcal{P}_\ell} \theta_j P(s \rightarrow t; \pi_j),$$



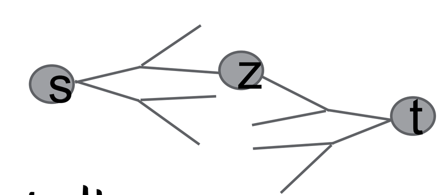
Path-Constrained Random Walks



Distant Supervision



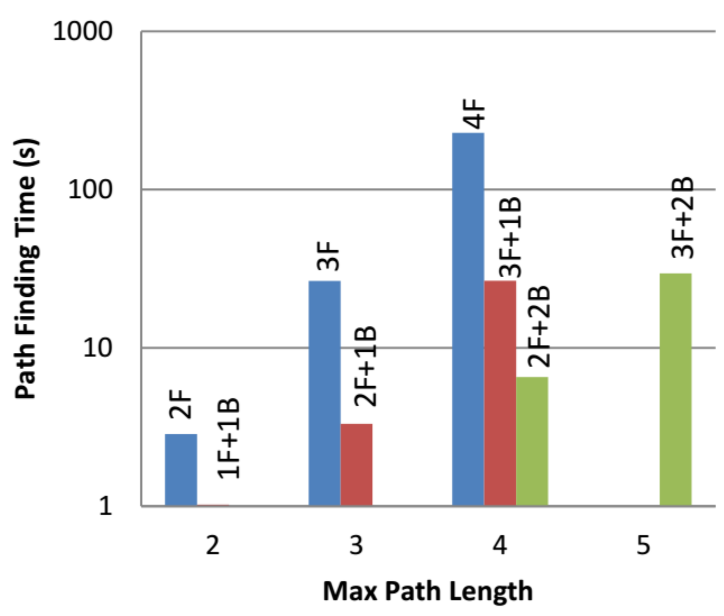
Combine Forward & Backward Random Walks

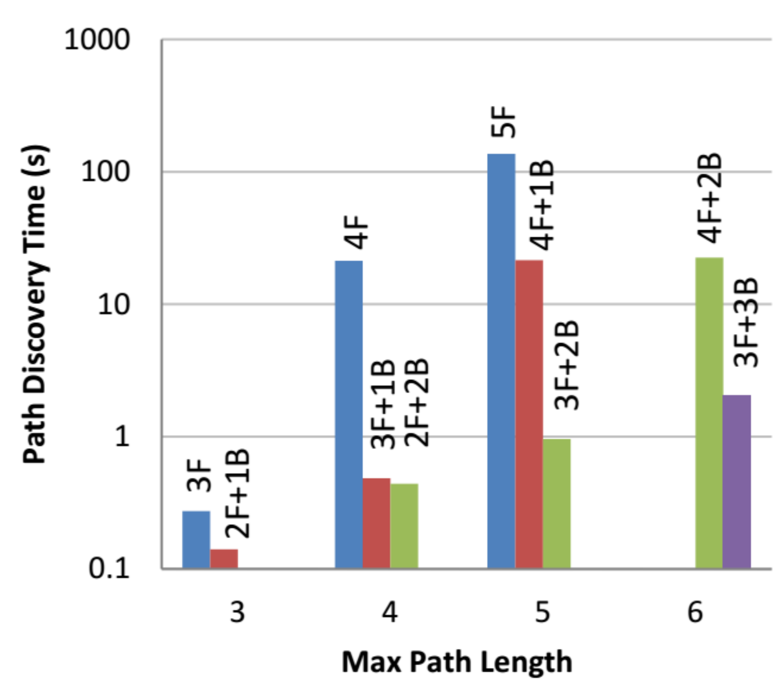


$$P(s \rightarrow t; \pi) = \sum_z P(s \rightarrow z; \pi') P(z \rightarrow t; r)$$

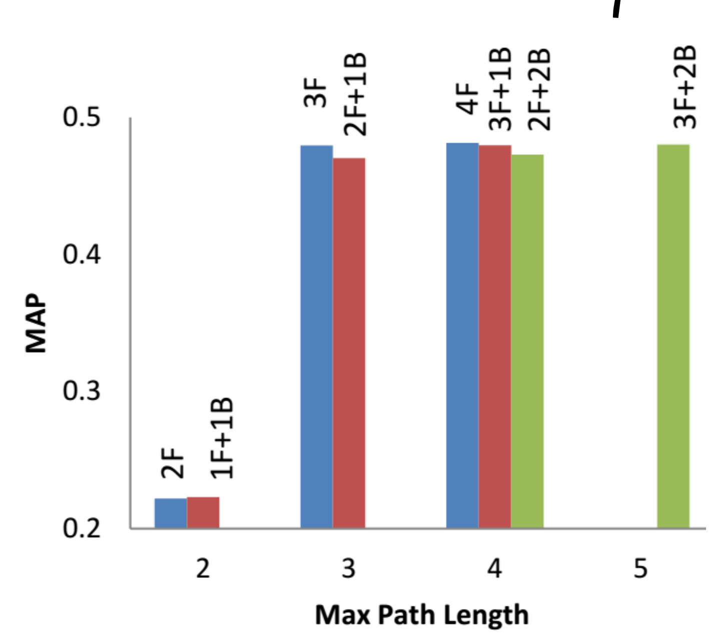
$$P(t \leftarrow s; \pi) = \sum_z P(t \leftarrow z; \pi'^{-1}) P(z \leftarrow s; r^{-1})$$

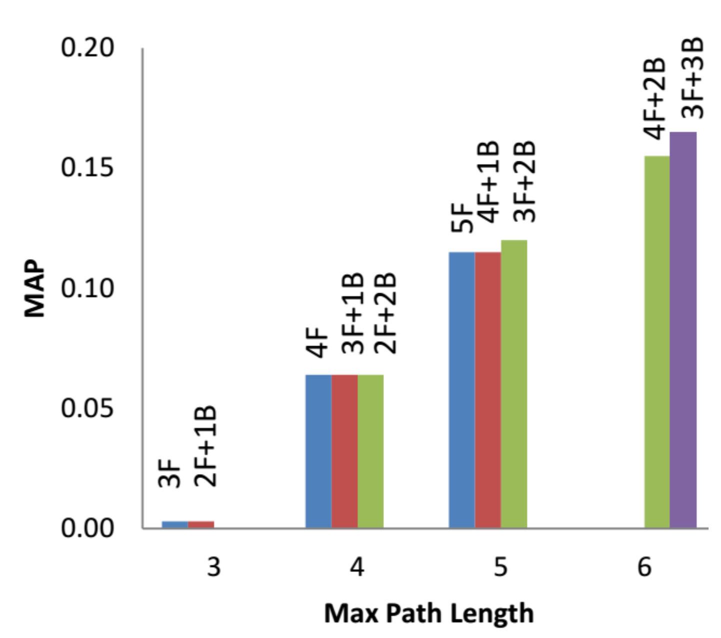
Path Finding Time





Prediction Quality





Main Results

	KB inference		NE extraction	
	Time	MAP	Time	MAP
RWR	25.6	0.429	7,375	0.017
FOIL	18918.1	0.358	366,558	0.167
PRA	10.2	0.477	277	0.107
CoR-PRA-*no-const*	16.7	0.479	449	0.167
CoR-PRA-*const*$_2$	23.3	0.524	556	0.186
CoR-PRA-*const*$_3$	27.1	**0.530**	643	**0.316**

Algorithm

Algorithm 1 Cor-PRA Feature Induction[1]

Input training queries $\{(s_i, G_i)\}, i = 1...n$
for each query (s, G) **do**
  1. Path exploration
  (i). Apply *path-finding* to generate paths $\mathcal{P}_s$ up to length ℓ that originate at s_i.
  (ii). Apply *path-finding* to generate paths $\mathcal{P}_t$ up to length ℓ that originate at every $t_i \in G_i$.
  2. Calculate random walk probabilities:
  for each $\pi_s \in \mathcal{P}_s$: **do**
    compute $P(s \rightarrow x; \pi_s)$ and $P(s \leftarrow x; \pi_s^{-1})$
  end for
  for each $\pi_t \in \mathcal{P}_t$: **do**
    compute $P(G \rightarrow x; \pi_t)$ and $P(G \leftarrow x; \pi_t^{-1})$
  end for
  3. Generate constant paths candidates:
  for each $(x \in N, \pi \in \mathcal{P}_t)$ with $P(G \rightarrow x|\pi_t) > 0$ **do**
    propose path feature $P(c \leftarrow t; \pi_t^{-1})$ setting $c = x$, and update its statistics by $coverage$ += 1.
  end for
  for each $(x \in N, \pi \in \mathcal{P}_t)$ with $P(G \leftarrow x|\pi_t^{-1}) > 0$ **do**
    propose $P(c \rightarrow t; \pi_t)$ setting $c = x$ and update its statistics by $coverage$ += 1
  end for
  4. Generate long (concatenated) path candidates:
  for each $(x \in N, \pi_s \in \mathcal{P}_s, \pi_t \in \mathcal{P}_t)$ with $P(s \rightarrow x|\pi_s) > 0$ and $P(G \leftarrow x|\pi_t^{-1}) > 0$ **do**
    propose long path $P(s \rightarrow t; \pi_s.\pi_t^{-1})$ and update its statistics by $coverage$ += 1, and $precision$ += $P(s \rightarrow x|\pi_s)P(G \leftarrow x|\pi_t^{-1})/n$.
  end for
  for each $(x \in N, \pi_s \in \mathcal{P}_s, \pi_t \in \mathcal{P}_t)$ with $P(s \leftarrow x|\pi_s^{-1}) > 0$ and $P(G \rightarrow x|\pi_t) > 0$ **do**
    propose long path $P(s \leftarrow t; \pi_t.\pi_s^{-1})$ and update its statistics by $coverage$ += 1, and $precision$ += $P(s \leftarrow x|\pi_s^{-1})P(G \rightarrow x|\pi_t)/n$.
  end for
end for