Noisy Or-based model for Relation Extraction using Distant Supervision

Ajay Nagesh[1,2], Gholamreza Haffari[2], Ganesh Ramakrishnan[1]

[1]IIT Bombay, [2]Monash University

DISTANT SUPERVISION

President-of	
Obama	US
P.Mukherjee	India
....	....

CEO-of	
Apple	Tim Cook
Amazon	Jeff Bezos
....	....

Relation

Facts

Freebase

Heuristically **ALIGN** facts in a **knowledge base** to sentences in a large **unlabeled corpus**

Relation Mention

The New York Times

Barack Obama is the 44th and current President of the United States ....

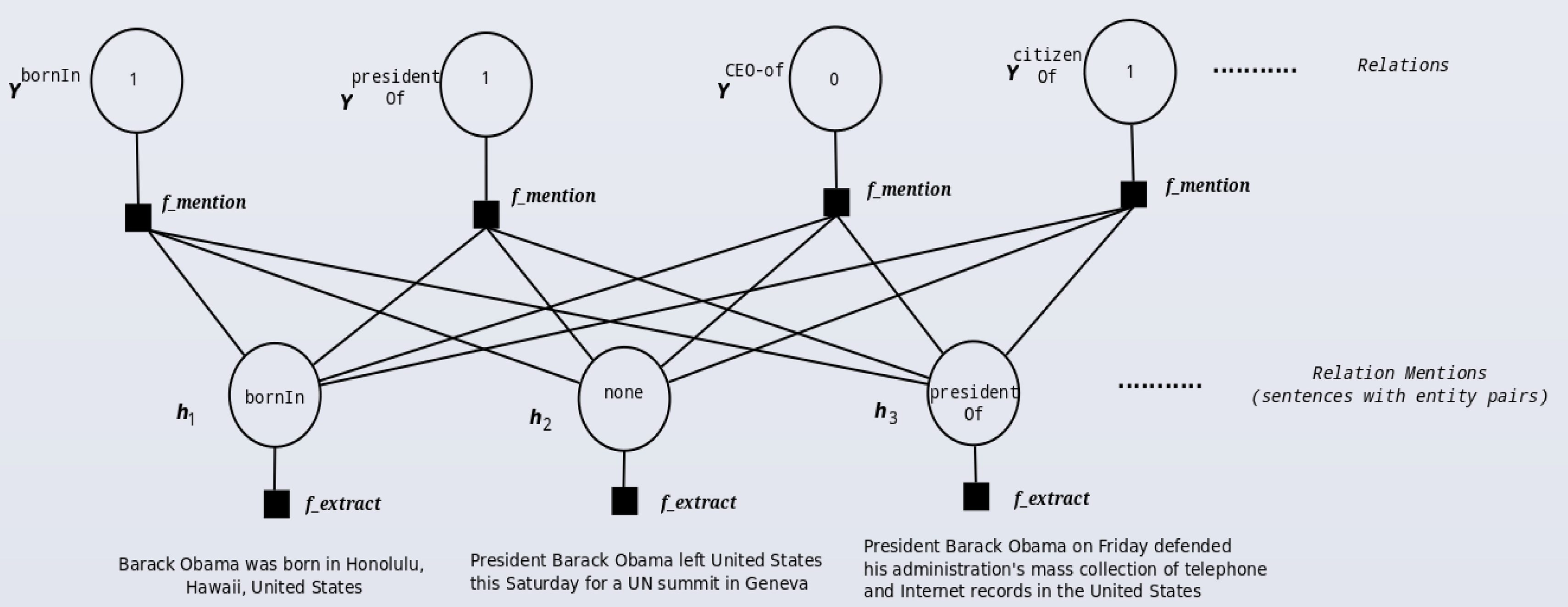
....

Apple's CEO Tim Cook announced the launch of the new brand of iPhones

....

....

GRAPHICAL MODEL



TRAINING ALGORITHM [Hoffmann et. al. 2011]

initialize parameter vector $\Theta \leftarrow 0$
for $t = 1...T$ **do**
  for $i = 1...n$ **do**
    $(\mathbf{y}', \mathbf{z}') \leftarrow \boxed{\arg\max_{\mathbf{y},\mathbf{z}} p(\mathbf{y}, \mathbf{z}|\mathbf{x_i}; \theta)}$
    if $\mathbf{y}' \neq \mathbf{y_i}$ **then**
      $\mathbf{z}^* \leftarrow \boxed{\arg\max_{\mathbf{z}} p(\mathbf{z}|\mathbf{x_i}, \mathbf{y_i}; \theta)}$
      $\Theta \leftarrow \Theta + \phi(\mathbf{x_i}, \mathbf{z}^*) - \phi(\mathbf{x_i}, \mathbf{z}')$
    end if
  end for
end for
Return Θ

Training Objective

$$O(\theta) = \prod_i p(\mathbf{y}_i|\mathbf{x}_i;\theta) = \prod_i \sum_{\mathbf{z}} p(\mathbf{y}_i, \mathbf{z}|\mathbf{x}_i;\theta)$$

most likely sentence labels and inferred facts (ignoring DB facts)

most likely sentence labels given DB facts

CONSTRAINTS ON DISTANT SUPERVISION

Examples of some constraints

1. Each mention of a pair of entities expresses only one relation
2. Each fact is expressed at least once in training corpus (*at-least-one*)
3. Facts present in database might not be present in training corpus (*noisy-or*)
4. Prime Minister has PER as 1st argument and COUNTRY as 2nd argument (*selectional preferences*)
5. Country can have only ONE PM
6. Trustee relationship can be valid for more than 2 tuples
7. A company cannot have HQ in 2 different locations
8. ....

Modeling constraints

- Constraints can be modeled effectively by posing the inference problem as an integer linear programming (ILP) problem [Roth & Yih, '04]
- ILP facilitates easy incorporation of constraints

OUR CONTRIBUTIONS

- We reformulate inference procedures during training as ILP problems.
- Introduce soft-constraint in the ILP objective to model noisy-or in training.
- Empirically, our algorithms perform better than Hoffmann et. al. (2011) under certain settings on two benchmark datasets.

ILP INFERENCE FORMULATIONS

Notation

- z_{ji}: mention j taking relation label i
- y_i: relation label being i
- s_{ji}: score of z_j=i (sentential features)
- m: no. of mentions
- R: no. of relation labels

Deterministic-OR

$$\max_{Z,Y} \left\{ \sum_{j=1}^{m} \sum_{i\in\{R,nil\}} [z_{ji} s_{ji}] \right\}$$

s.t
1. $\sum_{i\in\{R,nil\}} z_{ji} = 1 \quad \forall j$
2. $z_{ji} \le y_i \quad \forall j, \forall i$
3. $y_i \le \sum_{j=1}^{m} z_{ji} \quad \forall i$

where $z_{ji} \in \{0,1\}, \quad y_i \in \{0,1\}$

Constraint: at-least-one (Deterministic-OR)

Constraint: each mention has only one label

Noisy-OR

$$\max_{Z,Y,\epsilon} \left\{ \left(\sum_{j=1}^{m} \sum_{i\in\{R,nil\}} [z_{ji} s_{ji}] \right) - \left(\sum_{i\in R} \epsilon_i \right) \right\}$$

s.t
1. $\sum_{i\in\{R,nil\}} z_{ji} = 1 \quad \forall j$
2. $z_{ji} \le y_i \quad \forall j, \forall i$
3. $y_i \le \sum_{j=1}^{m} z_{ji} + \epsilon_i \quad \forall i$

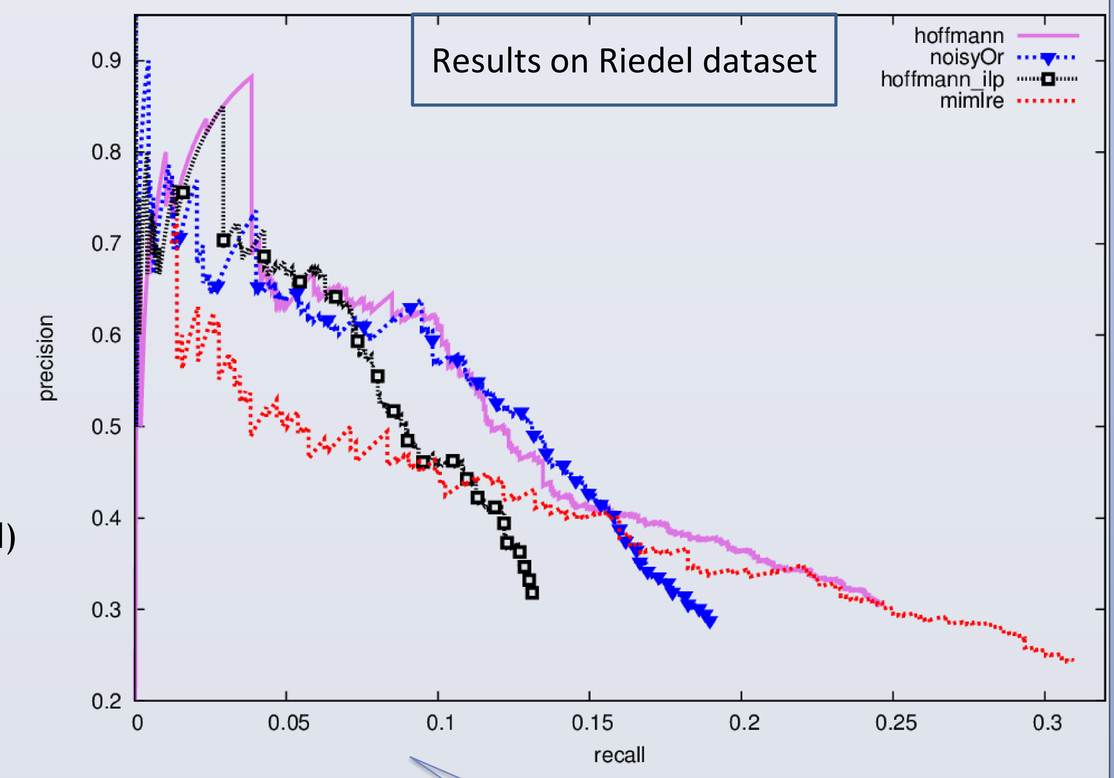
where $z_{ji} \in \{0,1\}, \quad y_i \in \{0,1\}, \quad \epsilon_i \in \{0,1\}$

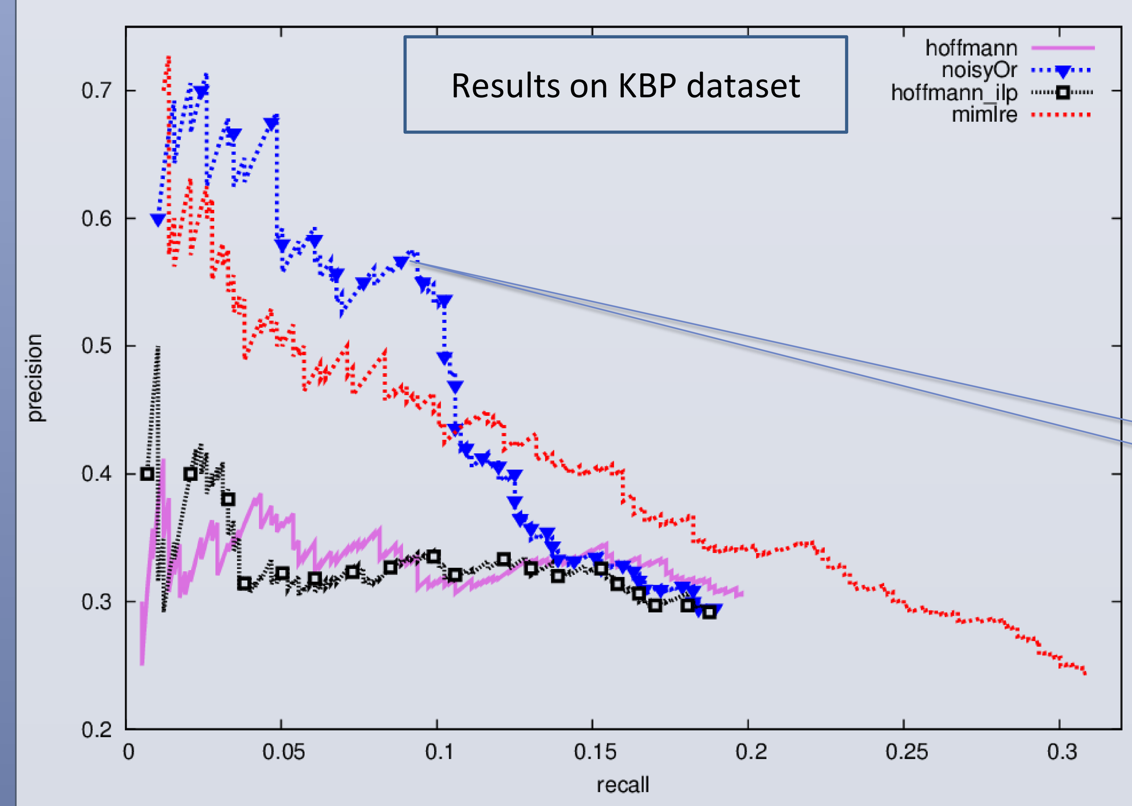
Constraint: Noisy-OR

EXPERIMENTS

Experimental Setup

- *Datasets:*
- KBP(shared task←→Wikipedia Infobox)
- Riedel(NYTimes←→Freebase)
- *Algorithms compared:*
- hoffmann_ilp , noisyOr
- hoffmann (baseline), MIMLRE (EM-based)





noisyOr : Comparable precision wrt mimlre but lower recall

noisyOr : Big jump in precision when compared to hoffmann (upto 0.4 points)

SUMMARY

- Addition of constraints using a ILP formulation
- Relaxation of deterministic-OR by a soft constraint (noisy-OR)
- Experiments on two benchmark datasets
- *Future Work*: Augment with other type of constraints (e.g. : selectional preferences of entity types, global constraints)

REFERENCES

- Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer, and Daniel S. Weld. 2011. Knowledge-based weak supervision for information extraction of overlapping relations. In *ACL-HLT '11*.
- Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *ACL '09*.
- Sampath Srinivas. 2013. A generalization of the noisyor model. *CoRR, abs/1303.1479*.
- Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati, and Christopher D. Manning. 2012. Multi-instance multi-label learning for relation extraction. In *EMNLP-CoNLL '12*.
- Dan Roth and Wen tau Yih. 2004. A linear programming formulation for global inference in natural language tasks. In *CoNLL-2004*.

ajaynagesh@cse.iitb.ac.in