Relation Extraction (2)

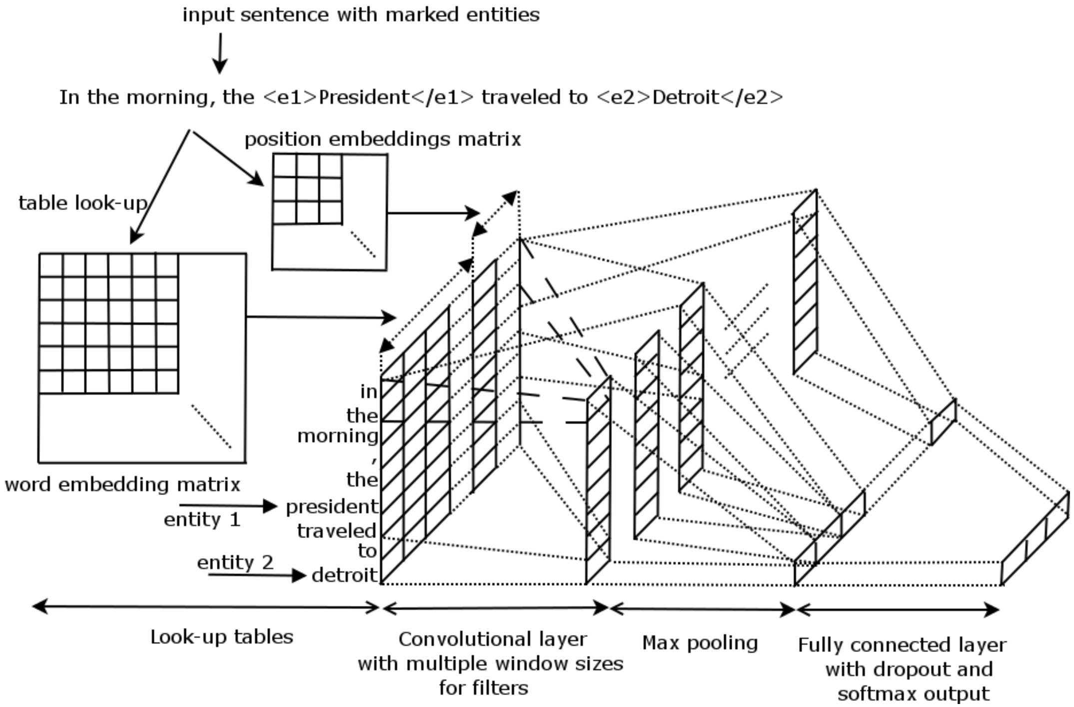
Bonan Min

bonanmin@gmail.com

Some slides are based on class materials from Ralph Grishman, Thien Huu Nguyen, Pedro Domingos, Stanley Kok, Eugene Agichtein, Luis Gravano, Zaiqing Nie

Deep Learning for RE

Avoid feature or kernel design for RE



A Convolutional Neural Network (CNN) for Relation Extraction (Nguyen and Grishman, 2015)

Classifier	Features	F
MaxEnt	POS, WordNet, morphological features, noun compound sys- tem, thesauri, Google n -grams	77.6
SVM	POS, WordNet, prefixes and other morphological fea- tures, dependency parse, Levin classes, PropBank, FrameNet, NomLex-Plus, Google n - grams, paraphrases, TextRunner	82.2
CNN (Zeng et al., 2014)	WordNet	82.7
CNN (Nguyen and Grishman, 2015a)	-	82.8

Performance on SemEval 2010

Position Embeddings

To inform the models about the two entity mentions of interest, we introduce (relative) position embeddings (randomly initialized and updated during training)

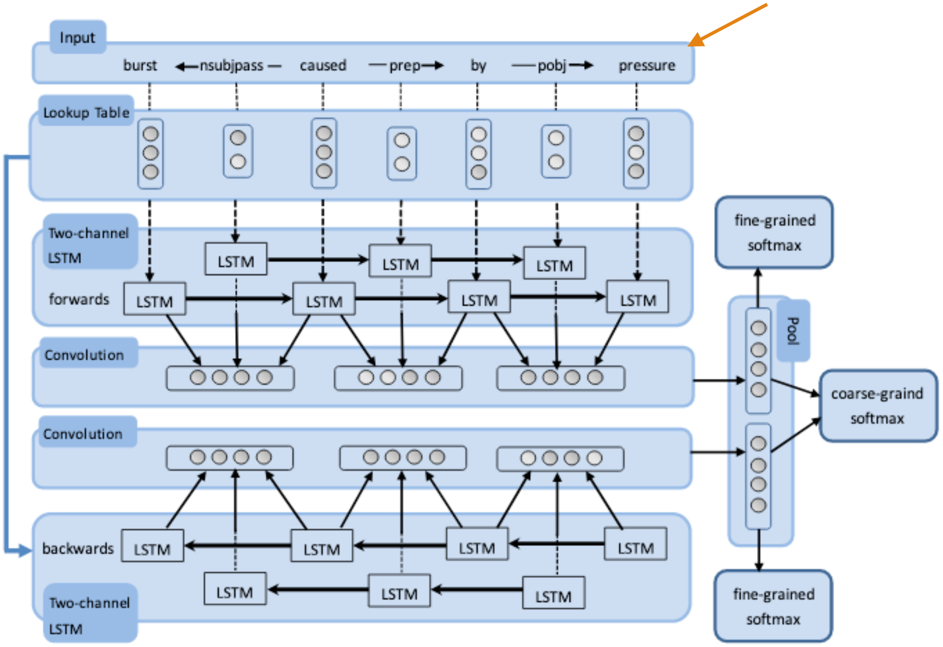
Dist from M1	0	1	2	3	4
Dist from M2	-4	-3	-2	-1	0
	[Bonan Min]	*teaches*	*NLP*	*at*	*[Tufts]*

-4	2	-0.5	1.1	0.3	0.4	-0.5
-3	-1.4	0.4	-0.2	-0.9	0.5	0.9
-2	-1.1	-0.2	-0.5	0.2	-0.8	0
-1	0.7	-0.3	1.5	-0.3	-0.4	0.1
0	-0.8	1.2	1	-0.7	-1	-0.4
1	0	0.3	-0.3	-0.9	0.2	1.4
2	0.8	0.8	-0.4	-1.4	1.2	-0.9
3	1.6	0.4	-1.1	0.7	0.1	1.6
4	1.2	-0.2	1.3	-0.4	0.3	-1.0

Deep Learning for RE

Can also incorporate syntax into deep learning models for RE: to identify important context words (i.e., via the dependency paths) or to guide the computational flows of the neural network models.

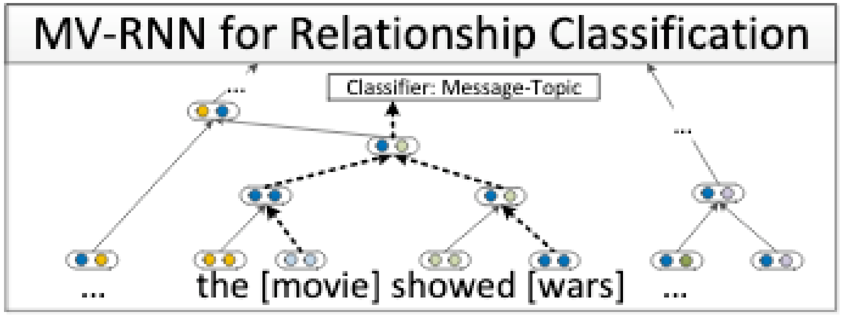
the shortest dependency path between two entity mentions



Cat et al., Bidirectional Recurrent Convolutional Neural Network for Relation Classification (ACL 2016)

Recursive neural networks: building the networks based on the constituent trees

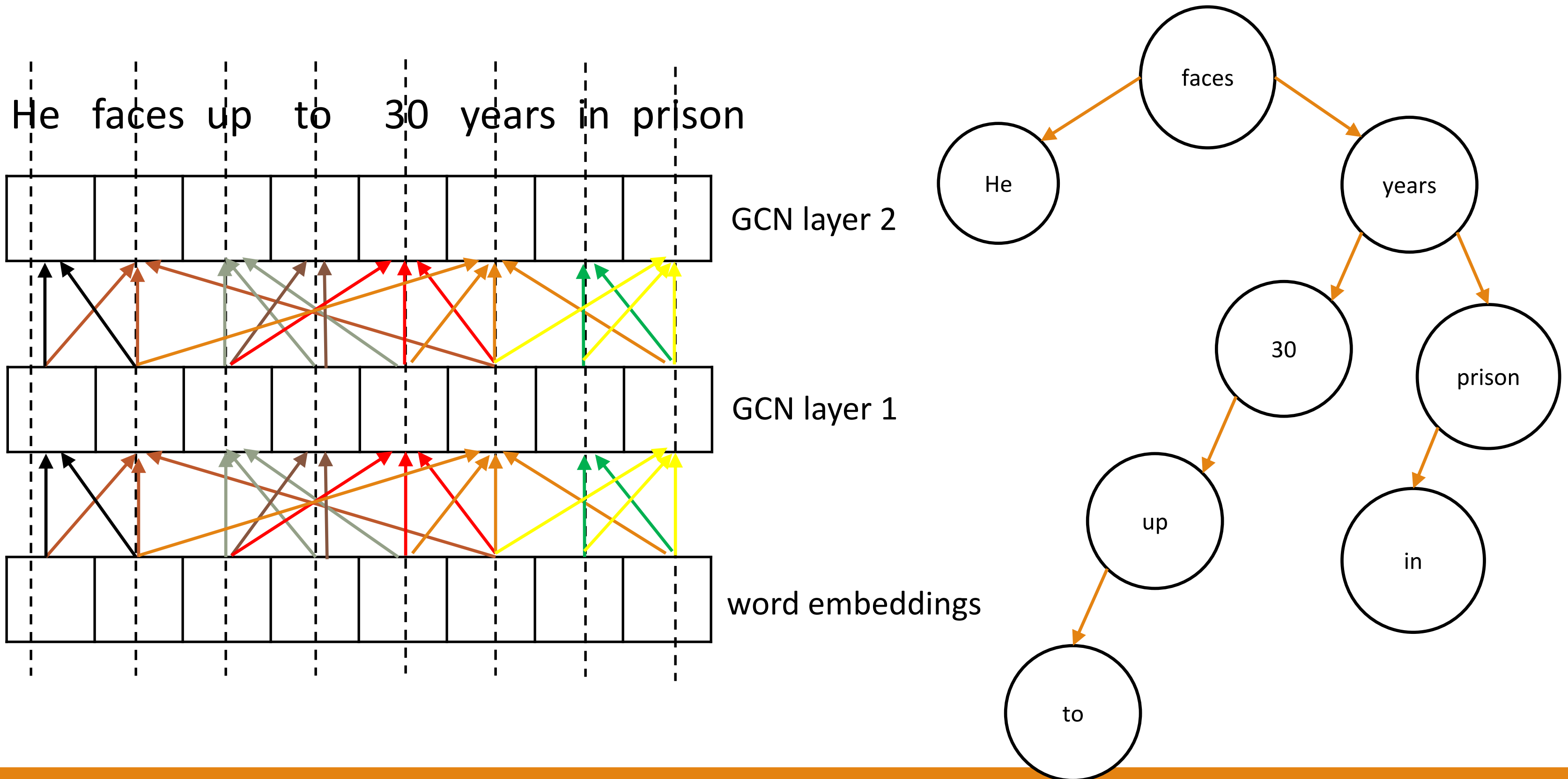
the binarized constituent subtree



Socher et al., Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank (EMNLP 2013)

Syntactic Structures for Relation Extraction

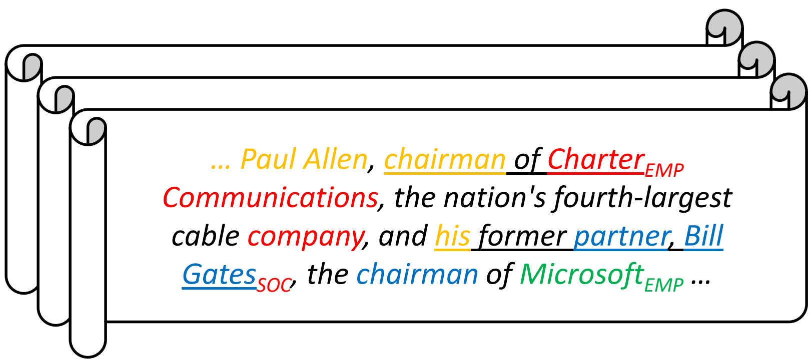
Graph Convolutional Neural Network (GCN) over dependency trees for RE (a recent state-of-the-art approach for RE) (Zhang et al., 2018)



Semi-supervised Methods For RE

Annotation for relation extraction is labor-intensive

- More costly than for names: must annotate entities and relations
- So there is a strong motivation to **minimize training data through semi-supervised methods**



As for names, we discussed a co-training approach:

- Feature set 1: the two entities
- Feature set 2: the contexts between the entities

Semi-Supervised Learning

Bootstrapping relies on **a duality between patterns and pairs/seeds**

- Seed:
  - *[Moby Dick, Herman Melville]*
- Contexts for seed:
  - *… wrote …*
  - *… is the author of …*
- Other pairs appearing in these contexts
  - *[Animal Farm, George Orwell]*
  - *[Don Quixote, Miguel de Cervantes]*
- Additional contexts …

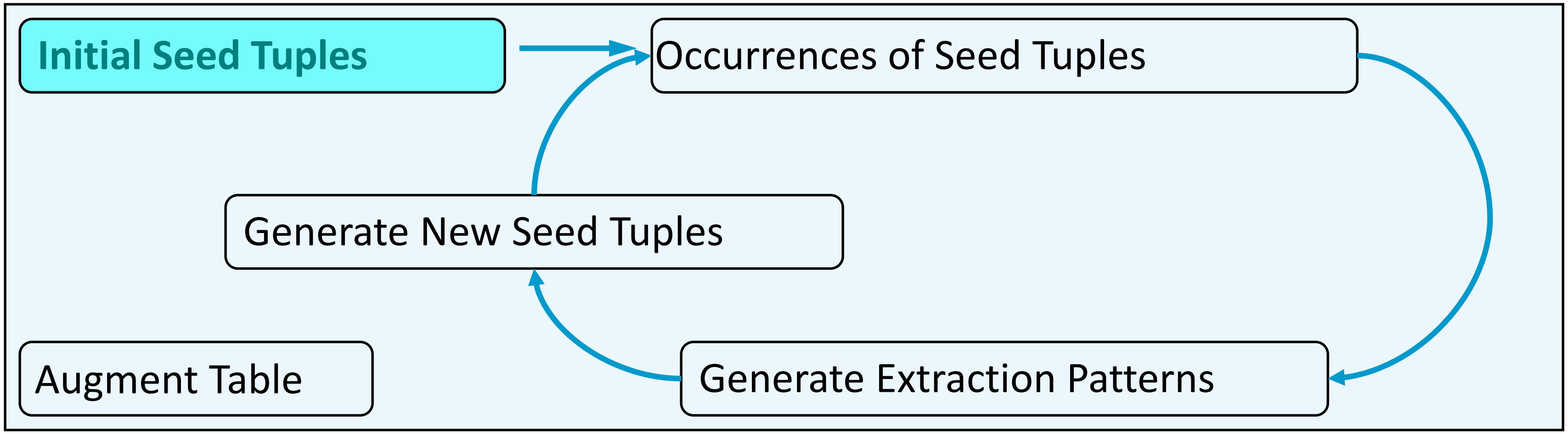
DIPRE (Sergey Brin, 1999): Dual Iterative Pattern Relation Expansion

Snowball (Agichtein and Gravano, 2000)

DIPRE (Brin, 1999)

Initial Seed Tuples:

ORGANIZATION	***LOCATION***
MICROSOFT	REDMOND
IBM	ARMONK
BOEING	SEATTLE
INTEL	SANTA CLARA



DIPRE (Brin, 1999)

Occurrences of seed tuples:

ORGANIZATION	***LOCATION***
MICROSOFT	REDMOND
IBM	ARMONK
BOEING	SEATTLE
INTEL	SANTA CLARA

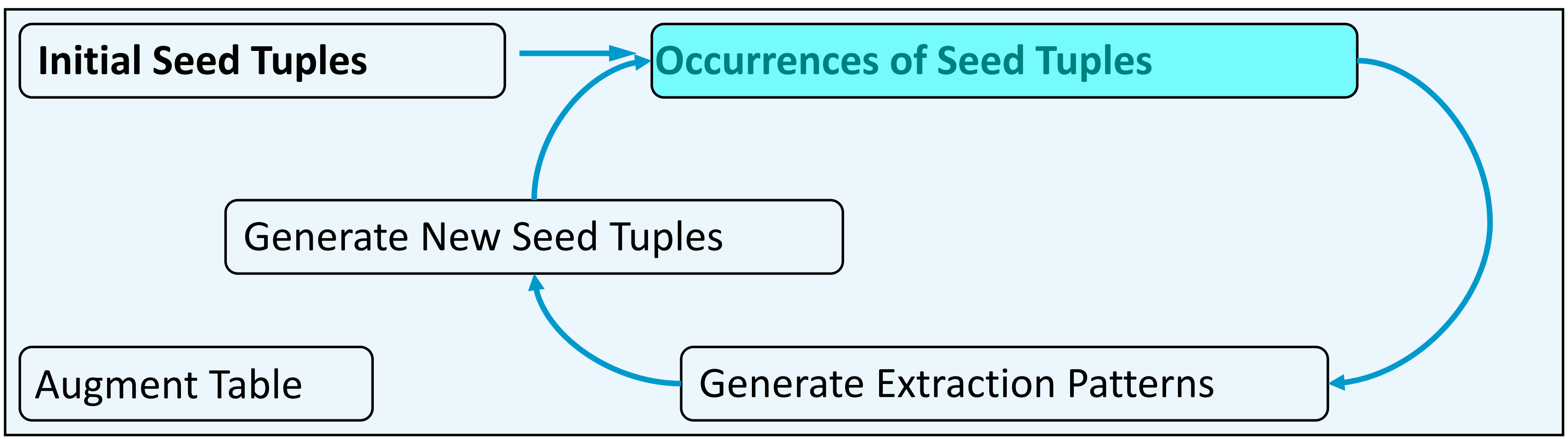
Computer servers at **Microsoft**’s headquarters in **Redmond**…

In mid-afternoon trading, share of **Redmond**-based **Microsoft** fell…

The **Armonk**-based **IBM** introduced a new line…

The combined company will operate from **Boeing**’s headquarters in **Seattle**.

Intel, **Santa Clara**, cut prices of its Pentium processor.



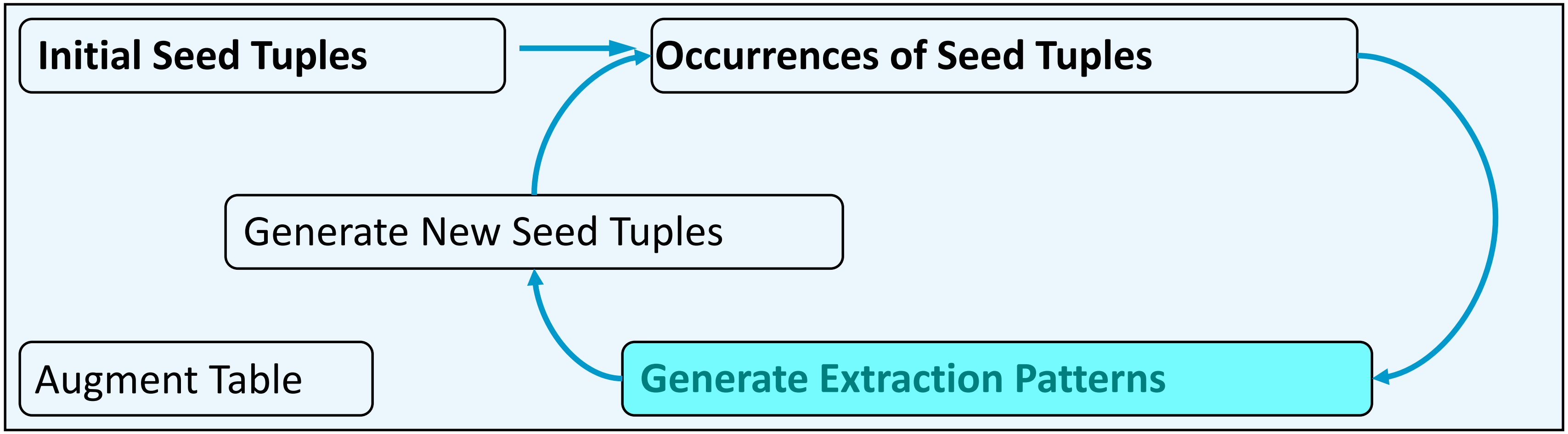
DIPRE (Brin, 1999)

DIPRE Patterns:

<*STRING1*>’s headquarters in <*STRING2*>

<*STRING2*> -based <*STRING1*>

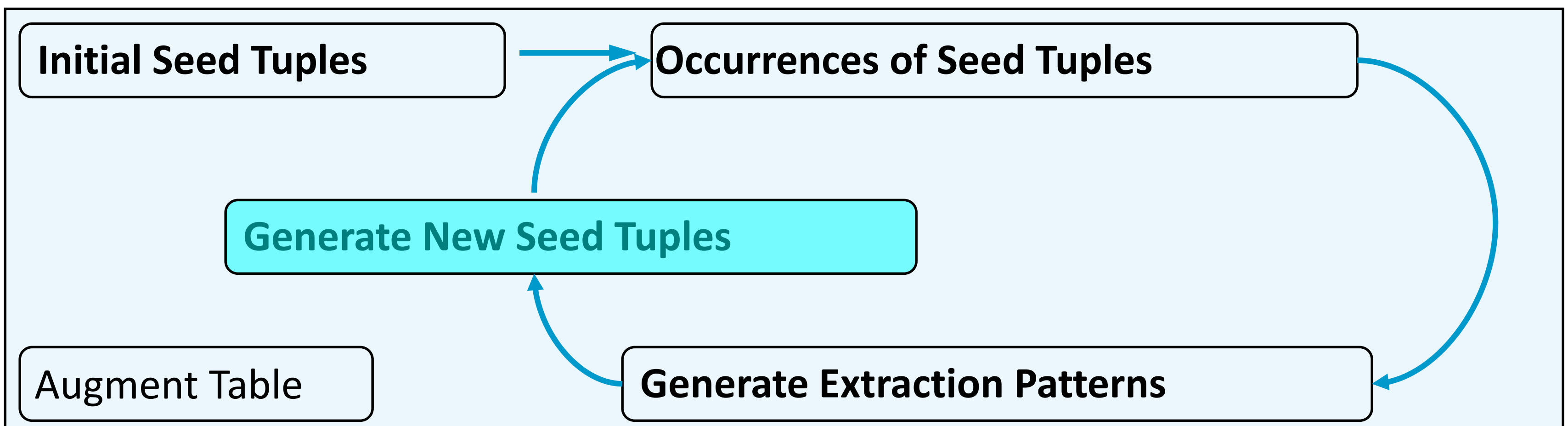
<*STRING1*> , <*STRING2*>



DIPRE (Brin, 1999)

Generate new seed tuples; start new iteration

ORGANIZATION	***LOCATION***
AG EDWARDS	ST LUIS
157TH STREET	MANHATTAN
7TH LEVEL	RICHARDSON
3COM CORP	SANTA CLARA
3DO	REDWOOD CITY
JELLIES	APPLE
MACWEEK	SAN FRANCISCO



Problems with DIRPE

Invalid tuples generated
- Degrade quality of tuples on subsequent iterations
- Must have automatic way to select high quality tuples to use as new seed

Pattern representation
- Patterns need to be reasonably precise
- Patterns must generalize

Snowball (Agichtein and Gravano, 2000) tries to solve these problems
- Pattern representation and generation
- Automatic pattern and tuple evaluation

Patterns Representation

Pattern: <*STRING2*>-based <*STRING1*>

Today's merger with McDonnell Douglas positions ***Seattle*** ***-based*** ***Boeing*** to make major money in space.

..., a producer of ***apple-based jelly***, ...

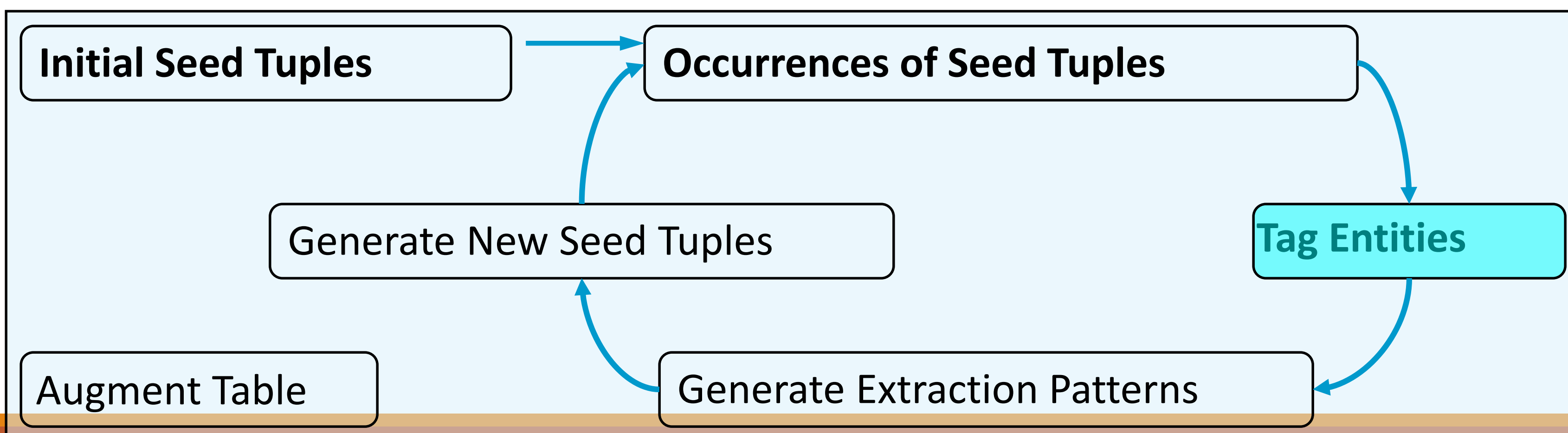
<*jelly*, *apple*>

Incorrect!

Snowball: Pattern Representation

Tag Entities

Computer servers at Microsoft's headquarters in Redmond…
In mid-afternoon trading, share of Redmond-based Microsoft fell…
The Armonk-based IBM introduced a new line…
The combined company will operate from Boeing's headquarters in Seattle.
Intel, Santa Clara, cut prices of its Pentium processor.



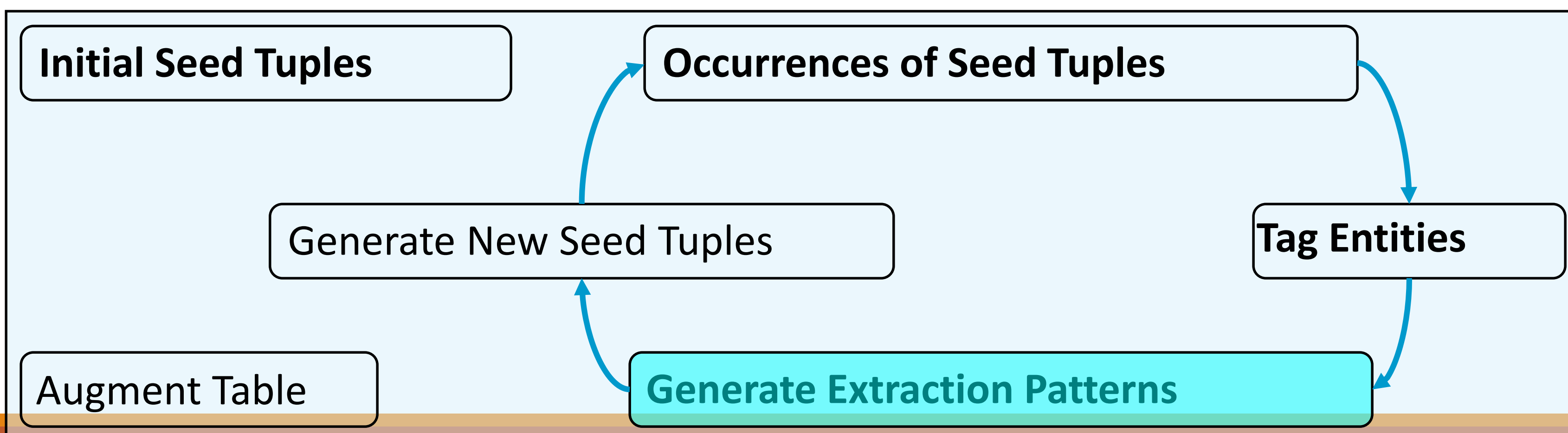
Snowball: Pattern Representation

<*ORGANIZATION*>'s headquarters in <*LOCATION*>

<*LOCATION*> -based <*ORGANIZATION*>

<*ORGANIZATION*> , <*LOCATION*>

PROBLEM: Patterns too specific: have to match text exactly.

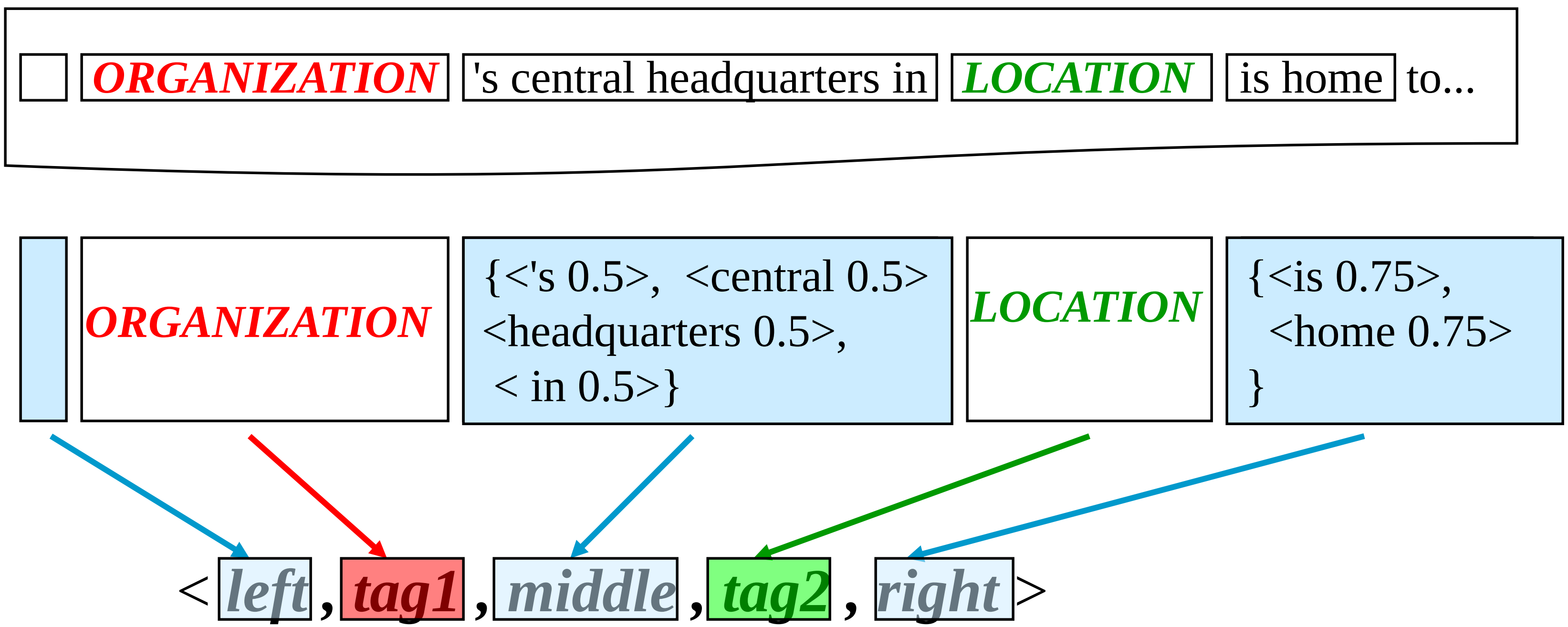


Snowball: Pattern Representation

A *Snowball* pattern vector is a 5-tuple

<***left, tag1, middle, tag2, right***>,

- ***tag1***, ***tag2*** are named-entity tags
- ***left***, ***middle***, and ***right*** are vectors of weighed terms.



Snowball: Pattern Generation

Tagged Occurrences of seed tuples:

Computer servers at **Microsoft**'s central headquarters in **Redmond**…

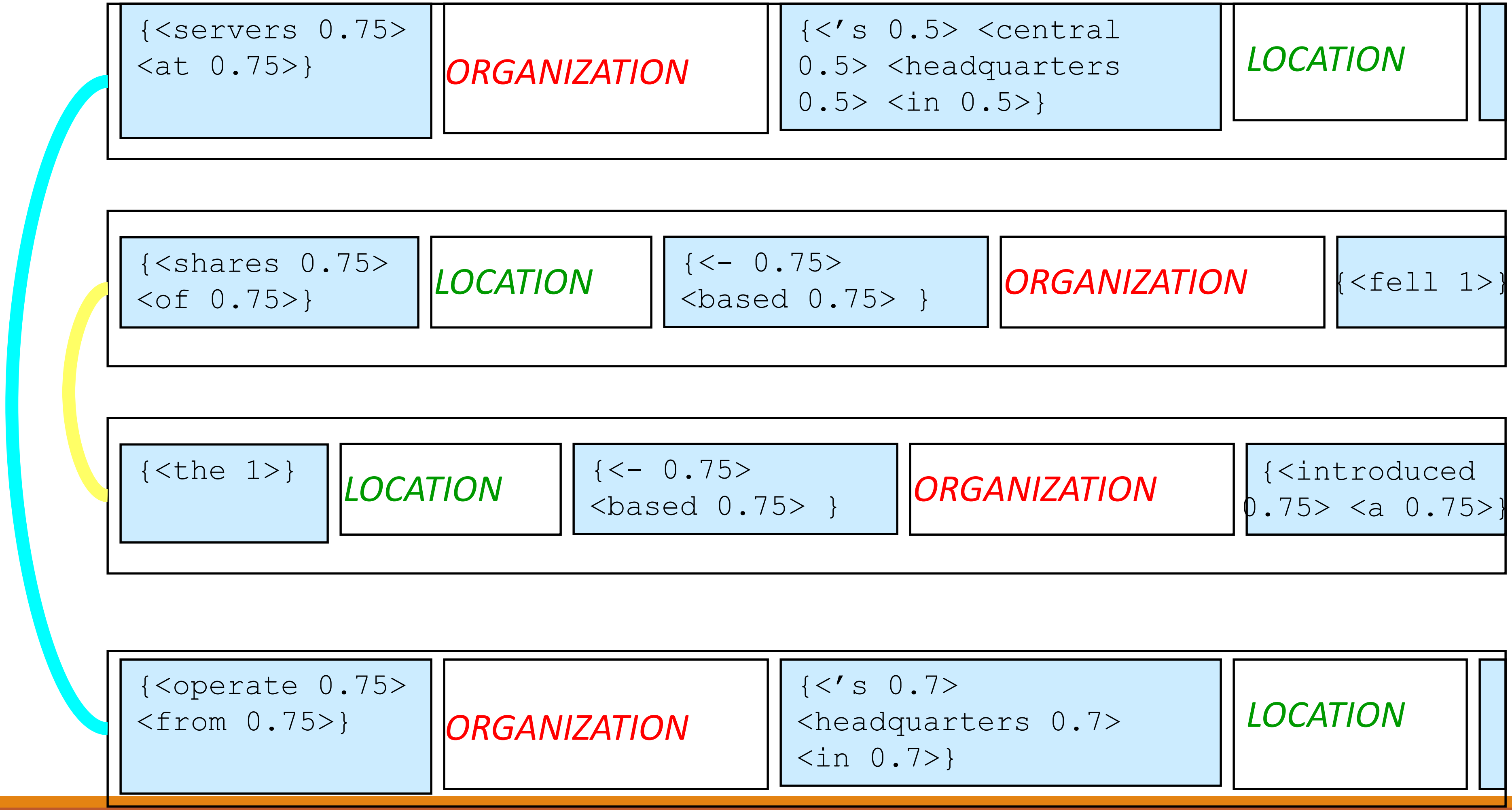
In mid-afternoon trading, share of **Redmond**-based **Microsoft** fell…

The **Armonk** -based **IBM** introduced a new line…

The combined company will operate from **Boeing**'s headquarters in **Seattle**.

Snowball Pattern Generation: Cluster Similar Occurrences

Occurrences of seed tuples converted to *Snowball* representation:



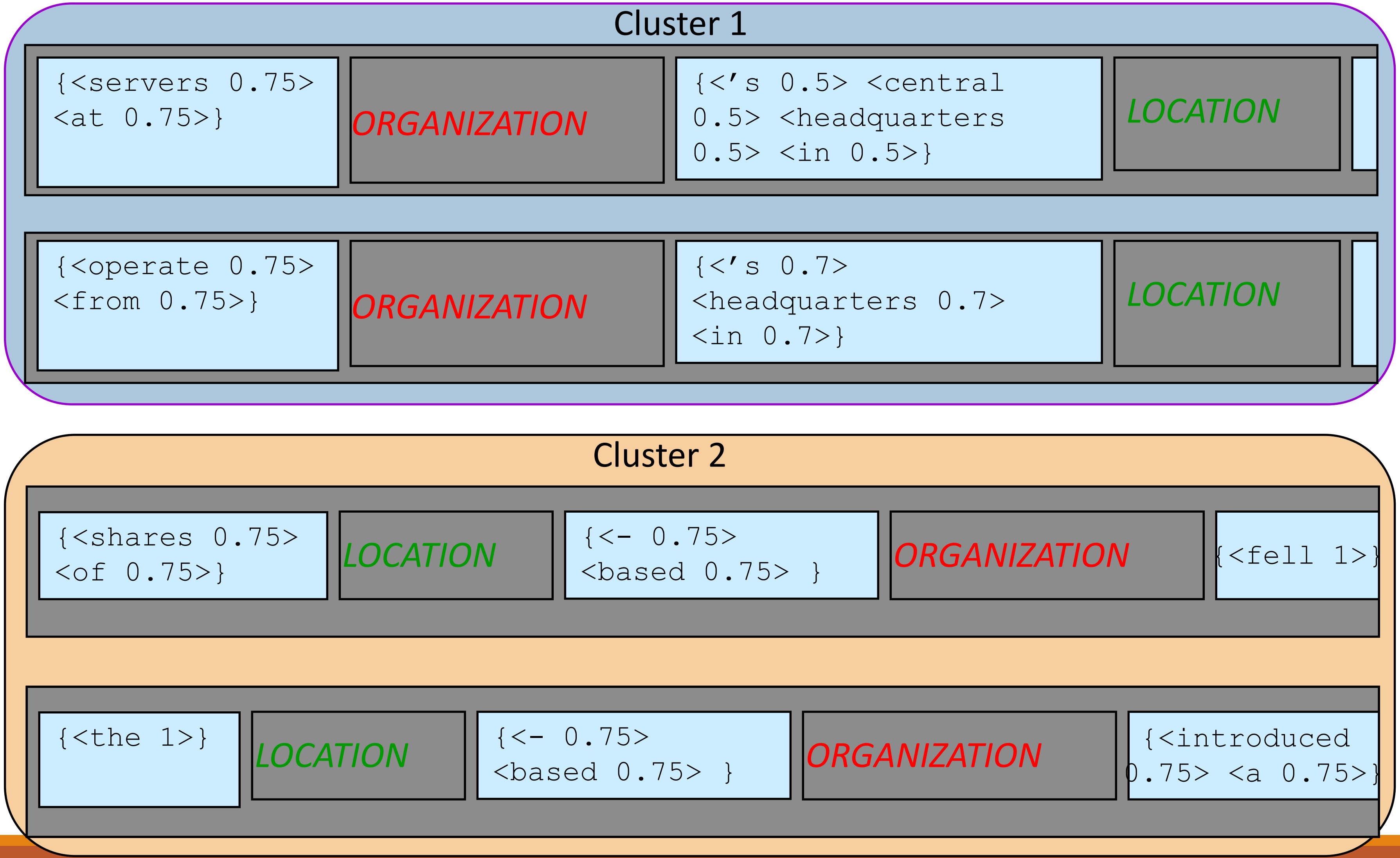
Similarity Metric

$P = \quad < \boxed{Lp}, \boxed{tag1}, \boxed{Mp}, \boxed{tag2}, \boxed{Rp} >$

$S = \quad < \boxed{Ls}, \boxed{tag1}, \boxed{Ms}, \boxed{tag2}, \boxed{Rs} >$

$$\textbf{Match}(P, S) = \begin{cases} \boxed{Lp} \cdot \boxed{Ls} + \boxed{Mp} \cdot \boxed{Ms} + \boxed{Rp} \cdot \boxed{Rs} & \text{if the tags match} \\ 0 & \text{otherwise} \end{cases}$$

Snowball: Pattern Clustering



Snowball: Automatic Pattern Evaluation

Seed tuples

ORGANIZATION	*LOCATION*
MICROSOFT	REDMOND
IBM	ARMONK
BOEING	SEATTLE
INTEL	SANTA CLARA

Pattern “ORGANIZATION, LOCATION” in action:

Boeing, Seattle, said…	**Positive**
Intel, Santa Clara, cut prices…	**Positive**
invest in Microsoft, New York-based analyst Jane Smith said	**Negative**

Pattern Confidence:

Automatically estimate probability of a pattern generating valid tuples:

$$\text{Conf(Pattern)} = \frac{\text{Positive}}{\text{Positive + Negative}}$$

e.g., Conf(Pattern) = 2/3 = 66%

Snowball: Automatic Tuple Evaluation

Brent Barlow, 27, a software analyst and beta-tester at ***Apple Computer*** headquarters in ***Cupertino***, was fired Monday for "thinking a little too different."

Apple's programmers "think different" on a "campus" in Cupertino, Cal.

<*Apple Computer*, *Cupertino*>

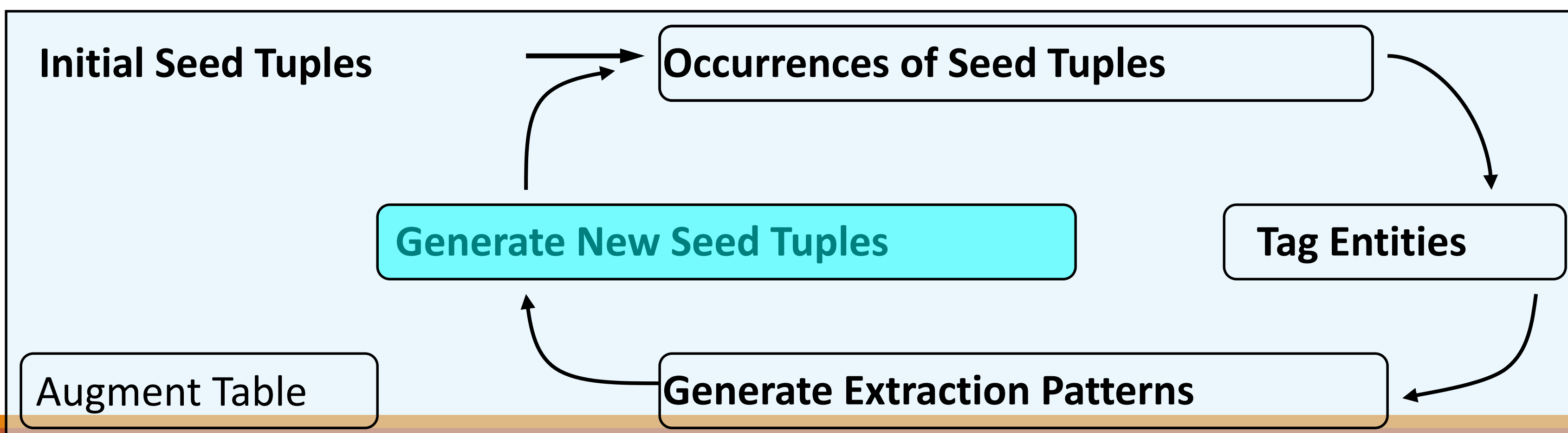
$$\text{Conf(Tuple)} = 1 - \prod (1 - \text{Conf}(P_i))$$

- Estimation of Probability (Correct (Tuple))
- A tuple will have high confidence if generated by multiple high-confidence patterns (P_i).

Snowball: Filtering Seed Tuples

Generate new seed tuples:

ORGANIZATION	LOCATION	CONF
AG EDWARDS	ST LUIS	0.93
AIR CANADA	MONTREAL	0.89
7TH LEVEL	RICHARDSON	0.88
3COM CORP	SANTA CLARA	0.8
3DO	REDWOOD CITY	0.8
3M	MINNEAPOLIS	0.8
MACWORLD	SAN FRANCISCO	0.7
157TH STREET	MANHATTAN	0.52
15TH CENTURY EUROPE	NAPOLEON	0.3
15TH PARTY CONGRESS	CHINA	0.3
MAD	SMITH	0.3



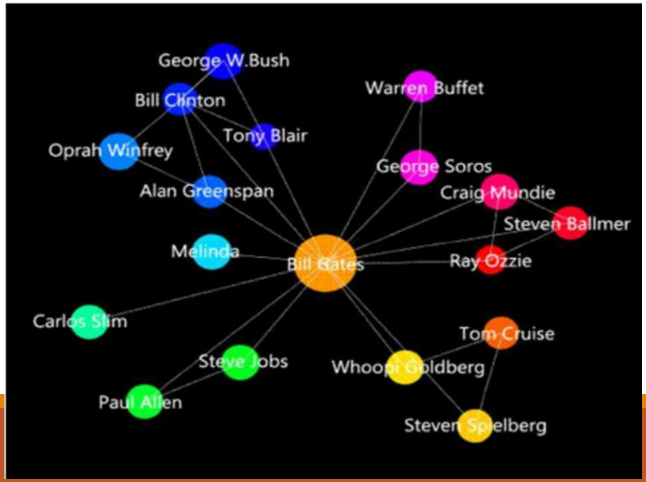
More on Bootstrapping (1): Statistical Snowball

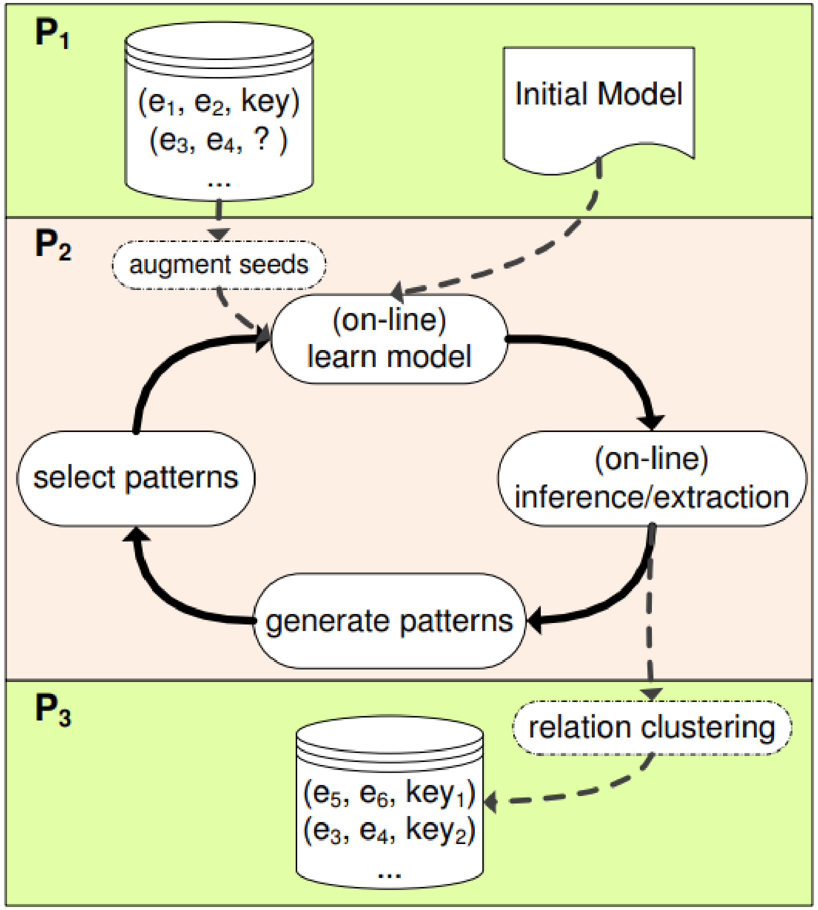
StatSnowball

- Use statistical models (Markov logic networks, MLN) to automatic learn the weight of rules (e.g., patterns)
- Automatically select seeds and patterns

The technology behind EntityCube / Renlinfang(Chinese)

- https://www.microsoft.com/en-us/research/project/entitycube/





StatSnowball, with three parts – P1 (input), P2 (statistical extraction model), and P3 (output)

More on Bootstrapping (2): Coupled Semi-Supervised Learning

Semantic drift

- Ranking / filtering quite effective for functional relations (book → author, company → headquarters)

  - But expansion may occur into other relations generally implied by seed ('semantic drift')
    - Ex: from governor → state governed_to
      person → state born_in

- Precision poor without functional property

More on Bootstrapping (2): Coupled Semi-Supervised Learning

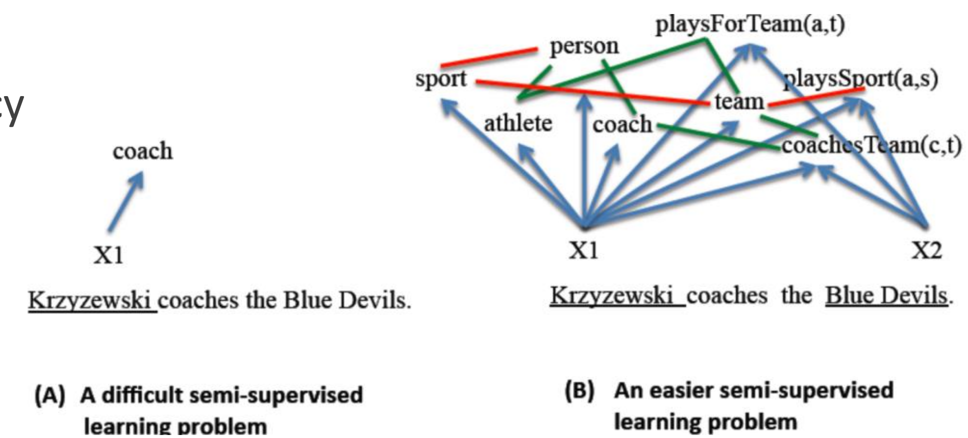
Goal: semi-supervised learning to extract entities (e.g., academic fields, athletes) and relations (e.g., PlaysSport(athlete, sport))

- Per entity/relation: a handful of labeled training examples
- Hundreds of millions of unlabeled web documents

Problem: Semi-supervised training using only a few labeled examples is typically unreliable because the learning task is under-constrained

Hypothesis: Much greater accuracy can be achieved by further constraining the learning task

- By coupling the semi-supervised training of many extractors for different categories and relations



Andrew Carlson, Justin Betteridge, Richard C. Wang, Estevam R. Hruschka Jr., Tom M. Mitchell. Coupled Semi-Supervised Learning for Information Extraction. WSDM'10.

More on Bootstrapping (2): Coupled Semi-Supervised Learning

Couples the training of multiple extraction techniques using a multi view constraint that requires them to agree

- Mutually exclusive, or must-agree

Assumption: Errors made by different extraction techniques should be independent

Algorithm 3: Meta-Bootstrap Learner (MBL)

Input: An ontology $\mathcal{O}$, a set of extractors $\mathcal{E}$
Output: Trusted instances for each predicate

for $i = 1, 2, \ldots, \infty$ **do**
  foreach *predicate* $p \in \mathcal{O}$ **do**
    foreach *extractor* $e \in \mathcal{E}$ **do**
      EXTRACT new candidates for p using e with recently promoted instances;
    end
    FILTER candidates that violate mutual-exclusion or type-checking constraints;
    PROMOTE candidates that were extracted by all extractors;
  end
end

Predicate	Precision (%)					Promoted Instances (#)				
	CPL	UPL	CSEAL	SEAL	MBL	CPL	UPL	CSEAL	SEAL	MBL
AcademicField	70	83	90	97	100	46	903	203	1000	181
Actor	100	33	100	97	100	199	1000	1000	1000	380
Animal	80	50	90	70	97	741	1000	144	974	307
Athlete	87	17	100	87	100	132	930	276	1000	555
AwardTrophyTournament	57	7	53	7	77	86	902	146	1000	79
BoardGame	80	13	70	77	90	10	907	126	1000	31
BodyPart	77	17	97	63	93	176	922	80	1000	61
Building	33	50	30	0	93	597	1000	57	1000	14
Celebrity	100	90	100	100	97	347	1000	72	747	514
CEO	33	30	100	77	100	3	902	322	1000	30
City	97	100	97	87	97	1000	1000	368	1000	603
Clothing	97	20	43	27	97	83	973	167	1000	102

Precision (%) and counts of promoted instances for each category using CPL, UPL, CSEAL, SEAL, and MBL

Andrew Carlson, Justin Betteridge, Richard C. Wang, Estevam R. Hruschka Jr., Tom M. Mitchell. Coupled Semi-Supervised Learning for Information Extraction. WSDM'10.

Distant Supervision

Sometimes a large data base is available involving the type of relation to be extracted

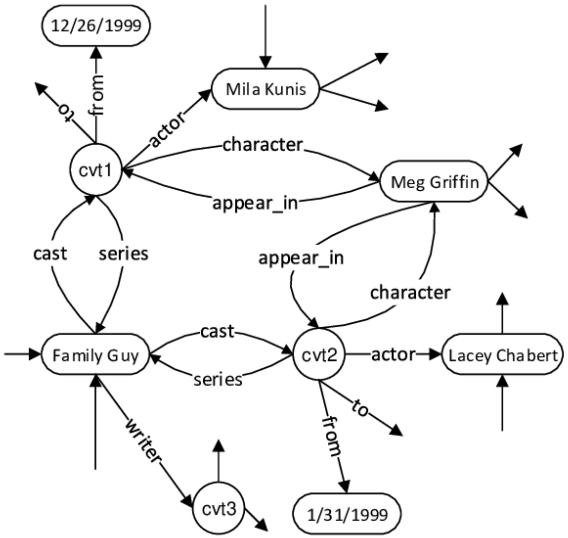
- A number of such public data bases are now available, such as FreeBase and YAGO

Text instances corresponding to some of the data base instances can be found in a large corpus or from the Web

Together these can be used to train a relation classifier

Freebase

yago select knowledge



https://developers.google.com/freebase/

Distant Supervision

Ronaldinho

From Wikipedia, the free encyclopedia

"Ronaldinho Gaucho" redirects here. For the comic strip based on him, see Ronaldinho Gaucho (comic strip). For other uses, see Ronaldinho (disambiguation).

This name uses Portuguese naming customs: the first or maternal family name is Assis *and the second or paternal family name is* Moreira.

Ronaldo de Assis Moreira (born 21 March 1980), commonly known as **Ronaldinho Gaúcho** (Brazilian Portuguese: [ʁonawˈdʒĩɲu ɡaˈuʃu]) or simply **Ronaldinho**,[note 1] is a Brazilian former professional footballer and ambassador for Barcelona.[4] He played mostly as an attacking midfielder, but was also deployed as a forward or a winger. He played the bulk of his career at European clubs Paris Saint-Germain, Barcelona and A.C. Milan as well as playing for the Brazilian national team. Often considered one of the best players of his generation and regarded by many as one of the greatest of all time,[note 2] Ronaldinho won two FIFA World Player of the Year awards and a Ballon d'Or. He was renowned for his technical skills and creativity; due to his agility, pace and dribbling ability, as well as his use of tricks, feints, overhead kicks, no-look passes and accuracy from free-kicks.

Ronaldinho made his career debut for Grêmio, in 1998. At age 20, he moved to Paris Saint-Germain in France before signing for Barcelona in 2003. In his second season with Barcelona, he won his first FIFA World Player of the Year award, as Barcelona won La Liga. The season that followed is considered one of the best in his career as he was instrumental in Barcelona winning the UEFA Champions League, their first in fourteen years, as well as another La Liga title, giving Ronaldinho his first career double. After scoring two spectacular solo goals in *El Clásico*, Ronaldinho became the second Barcelona player, after Diego Maradona in 1983, to receive a standing ovation from Real Madrid fans at the Santiago Bernabéu. Ronaldinho also received his second FIFA World Player of the Year award, as well as the Ballon d'Or.

Ronaldinho

Ronaldinho in 2019

Personal information	
Full name	Ronaldo de Assis Moreira[1]
Date of birth	21 March 1980 (age 39)[1]
Place of birth	Porto Alegre, Brazil
Height	1.81 m (5 ft 11 in)[1]
Playing position	Attacking midfielder / Forward
Youth career	
1987–1998	Grêmio

Senior career*			
Years	**Team**	**Apps**	**(Gls)**
1998–2001	Grêmio	52	(21)
2001–2003	Paris Saint-Germain	55	(17)
2003–2008	Barcelona	145	(70)
2008–2011	A.C. Milan	76	(20)
2011–2012	Flamengo	33	(15)
2012–2014	Atlético Mineiro	48	(16)
2014–2015	Querétaro	25	(8)
2015	Fluminense	7	(0)

Distant Supervision: Approach

Key idea: **heuristically label a corpus using a knowledge base**

Given:

- Data base for relation R
- Corpus containing information about relation R

Collect $< X, Y >$ pairs from data base relation R

Collect sentences in corpus containing both X and Y

- These are positive training examples

Collect sentences in corpus containing X and some Y' with the same entity type as Y such that $< X, Y' >$ is not in the data base

- These are negative training examples

Use examples to train classifier which operates on pairs of entities

Distant Supervision (Mintz et al., 2009)

Steve Jobs, the co-founder of **Apple** Inc. …	CEO
Most books on **Apple** or **Steve Jobs** have …	CEO
Apple CEO **Steve Jobs** …	CEO
Steve Ballmer joined **Microsoft** on **…**	CEO
Microsoft CEO **Steve Ballmer** is …	CEO
Gary Pard, of NJ's **DeCamp Bus Lines**, said…	NIL
Thamesdown Transport Ltd CEO **Paul Jenkins**	NIL
Steve Jobs thoughts on **Microsoft** …	NIL

CEO (Steve Jobs, Apple)
CEO (Steve Ballmer, Microsoft)
…

Distant Supervision (Mintz et al., 2009)

Steve Jobs, the co-founder of **Apple** Inc. …	E	CEO
Most books on **Apple** or **Steve Jobs** have …	E	CEO
Apple CEO **Steve Jobs** …	E	CEO
Steve Ballmer joined **Microsoft** on **…**	E	CEO
Microsoft CEO **Steve Ballmer** is …	E	CEO
Gary Pard, of NJ's **DeCamp Bus Lines**, said…	E	NIL
Thamesdown Transport Ltd CEO **Paul Jenkins**	E	NIL
Steve Jobs thoughts on **Microsoft** …	E	NIL

CEO (Steve Jobs, Apple)
CEO (Steve Ballmer, Microsoft)
…

Train a statistical Extraction model

Distant Supervision: Limitations

The training data produced through distant supervision may be quite noisy:

- **False positives**: If a pair $< X, Y >$ is involved in multiple relations, $R < X, Y >$ and $R' < X, Y >$ and the data base represents relation R, the text instance may represent relation R', yielding a false positive training instance
  - If many $< X, Y >$ pairs are involved, the classifier may learn the wrong relation
- **False negatives**: If a relation is incomplete in the data base … for example, if *resides_in* $< X, Y >$ contains only a few of the locations where a person has resided … then we will generate many false negatives, possibly leading the classifier to learn no relation at all

	Pairs in Freebase	Sentences
False positive	Y	*Steve Ballmer was invited by **Bill Gates** to join **Microsoft** in 1980.*
	Y	***Bill Gates** showed up at **Microsoft**'s campus.*
False negative	N	***Bonan Min** works at **Tufts.***

Distant Supervision: False Positives

Steve Jobs, the co-founder of **Apple** Inc. … $\rightarrow \boxed{E} \rightarrow$ CEO

Most books on **Apple** or **Steve Jobs** have … $\rightarrow \boxed{E} \rightarrow$ CEO

Apple CEO **Steve Jobs** … $\rightarrow \boxed{E} \rightarrow$ CEO

Steve Ballmer joined **Microsoft** on **…** $\rightarrow \boxed{E} \rightarrow$ CEO

Microsoft CEO **Steve Ballmer** is … $\rightarrow \boxed{E} \rightarrow$ CEO

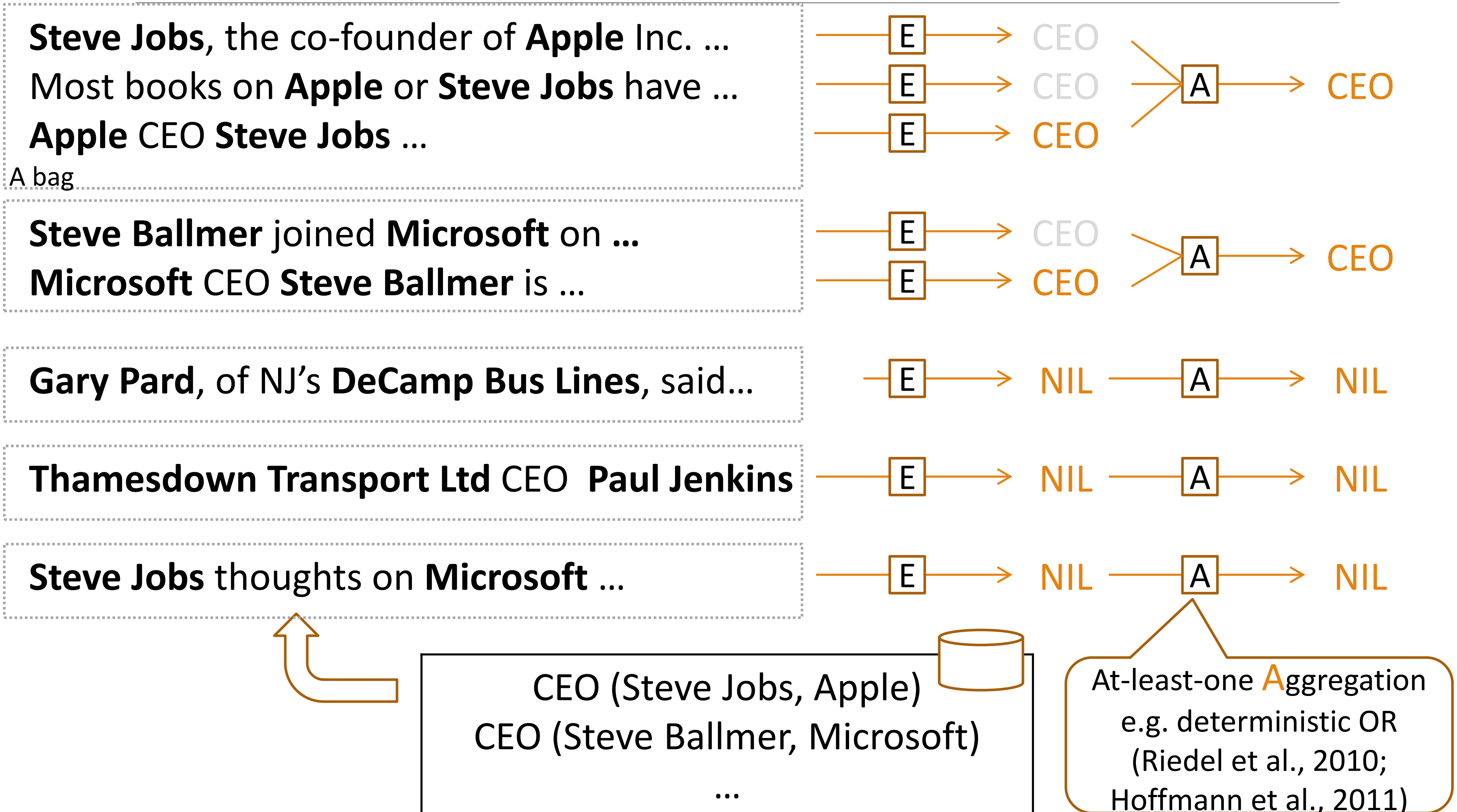
Gary Pard, of NJ's **DeCamp Bus Lines**, said… $\rightarrow \boxed{E} \rightarrow$ NIL

Thamesdown Transport Ltd CEO **Paul Jenkins** $\rightarrow \boxed{E} \rightarrow$ NIL

Steve Jobs thoughts on **Microsoft** … $\rightarrow \boxed{E} \rightarrow$ NIL

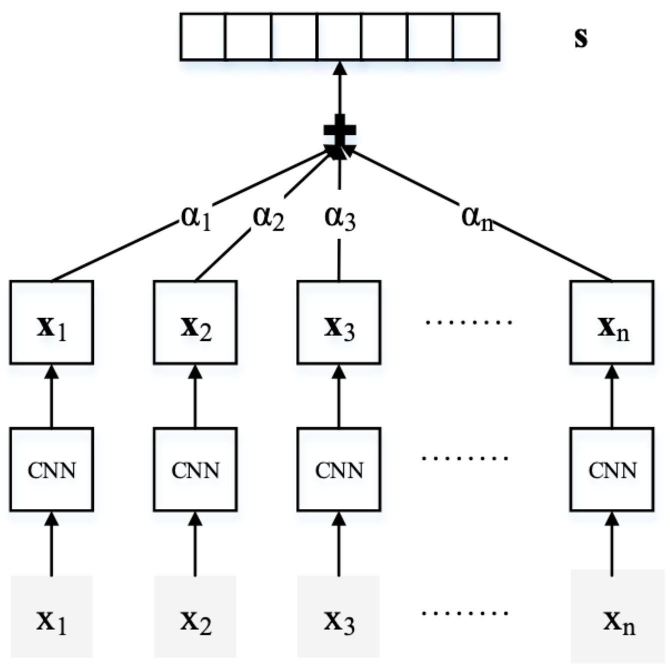
CEO (Steve Jobs, Apple)
CEO (Steve Ballmer, Microsoft)
…

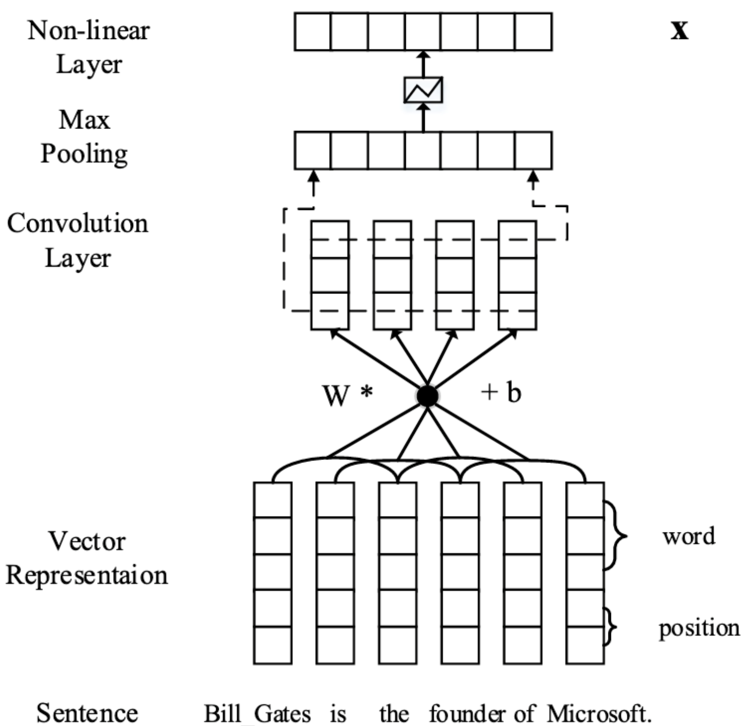
Reduce False Positives: Multi-Instance Learning (MIL)



Reduce False Positives: Multi-Instance Learning with Neural Networks

$$\mathbf{s} = \sum_i \alpha_i \mathbf{x}_i \qquad \alpha_i = \frac{\exp(e_i)}{\sum_k \exp(e_k)} \qquad e_i = \mathbf{x}_i \mathbf{A} \mathbf{r} \qquad \mathbf{o} = \mathbf{M}\mathbf{s} + \mathbf{d} \qquad p(r|S,\theta) = \frac{\exp(o_r)}{\sum_{k=1}^{n_r} \exp(o_k)}$$





Lin et al., Neural Relation Extraction with Selective Attention over Instances (ACL 2016)

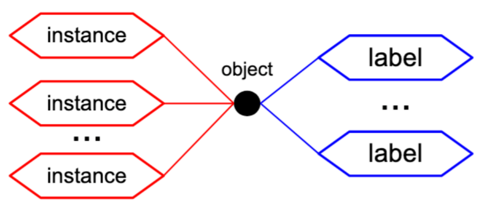
Reduce False Positives: Multi-label Multi-Instance Learning (MIML)

To reduce noise in distant supervision:

- Group instances (sentences) corresponding to the same entity pair $< X, Y >$ in the knowledge base into the a group (a bag of instances)
- Each bag can be assigned to multiple relations to capture the possible relations between X and Y in the knowledge base.
- People might just do multi-instance learning (i.e., a single label for a bag)

$$\text{DB} = \begin{pmatrix} \textit{BornIn}(\text{Barack Obama, United States}) \\ \textit{EmployedBy}(\text{Barack Obama, United States}) \end{pmatrix}$$

Sentence	Latent Label
Barack Obama is the 44th and current President of the United States.	*EmployedBy*
Obama was born in the United States just as he has always said.	*BornIn*
United States President Barack Obama meets with Chinese Vice President Xi Jinping today.	*EmployedBy*
Obama ran for the United States Senate in 2004.	–



Surdeanu et al., Multi-instance Multi-label Learning for Relation Extraction (EMNLP 2012)

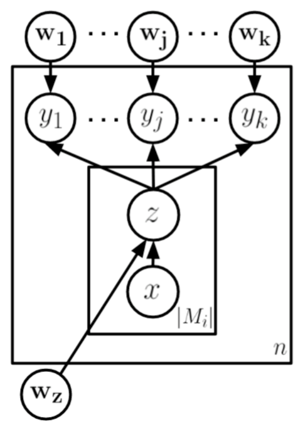
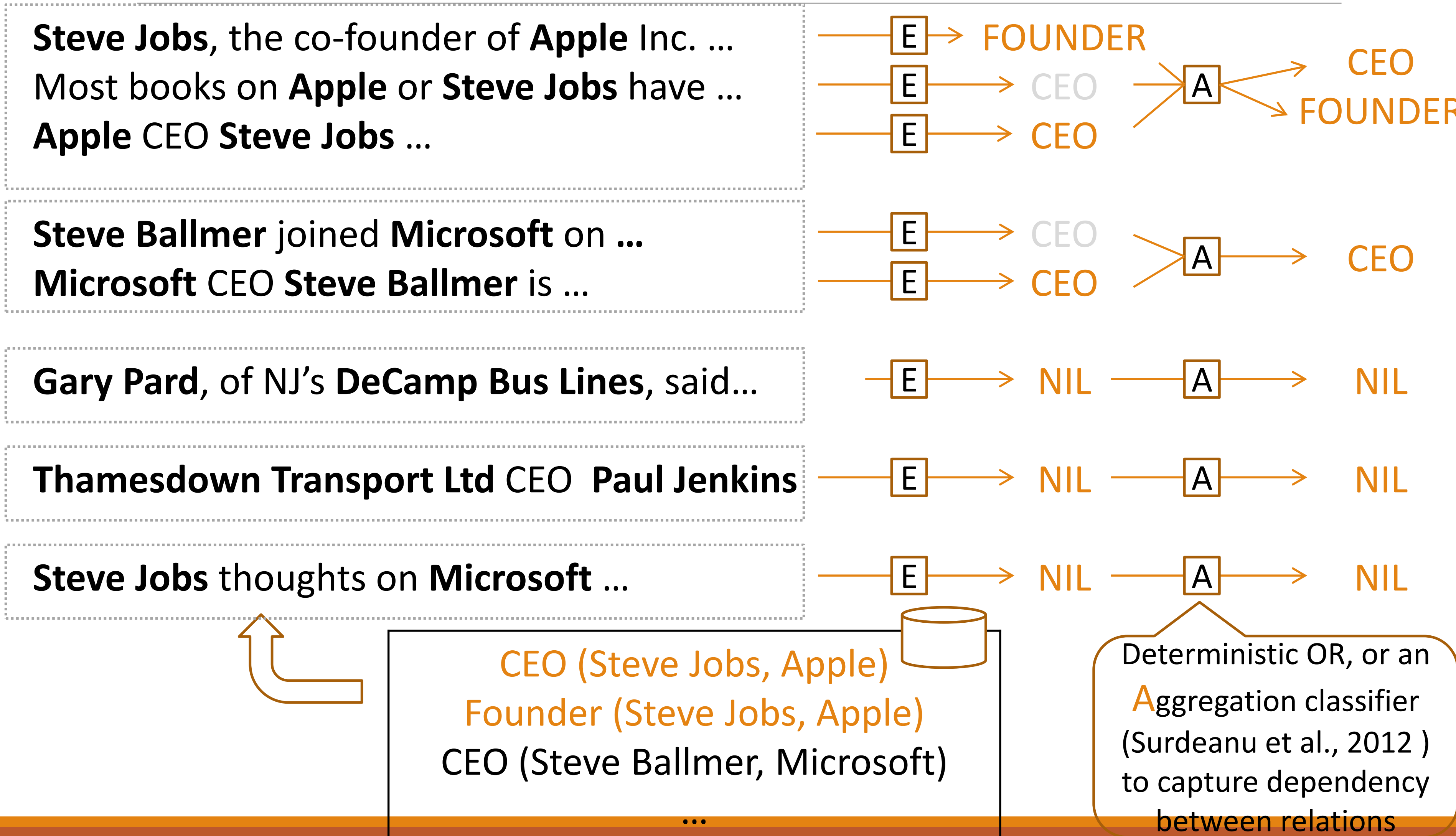


Figure 3: MIML model plate diagram. We unrolled the y plate to emphasize that it is a collection of binary classifiers (one per relation label), whereas the z classifier is multi-class. Each z and y_j classifier has an additional prior parameter, which is omitted here for clarity.

Reduce False Positives: Multi-label Multi-Instance Learning (MIML)



Reduce False Negatives: Semi-supervised MIML (*Positive* and *Unlabeled* Bags)

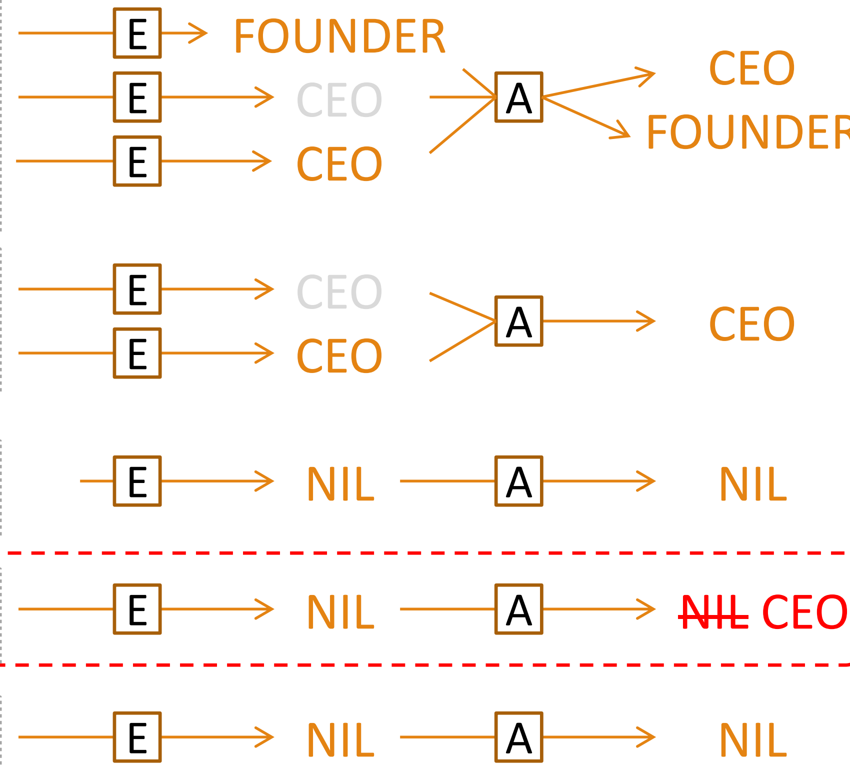
Steve Jobs, the co-founder of **Apple** Inc. …
Most books on **Apple** or **Steve Jobs** have …
Apple CEO **Steve Jobs** …

Steve Ballmer joined **Microsoft** on **…**
Microsoft CEO **Steve Ballmer** is …

Gary Pard, of NJ's **DeCamp Bus Lines**, said…

Thamesdown Transport Ltd CEO **Paul Jenkins**

Steve Jobs thoughts on **Microsoft** …



CEO (Steve Jobs, Apple)
Founder (Steve Jobs, Apple)
CEO (Steve Ballmer, Microsoft)
…

- ***Problem : incorrectly assigned "negative" labels***
- ***Cause:*** *KB is never complete*

Min et al., Distant Supervision for Relation Extraction with an Incomplete Knowledge Base. (NAACL-HLT 2013)

The Incomplete KB problem

Incompleteness (*attr. r*) = percentage of *PER*s has no value for *r*

Freebase relation types	Incompleteness
/people/person/education	0.792
/people/person/employment_history	0.923
/people/person/nationality*	0.785
/people/person/parents*	0.988
/people/person/place_of_birth*	0.938
/people/person/places_lived*	0.966

The incompleteness of Freebase (* are must-have attributes for a person). (Some early observation also appears in Sun et al., 2011)

Collaboratively-edited KBs are incomplete and can't stay current

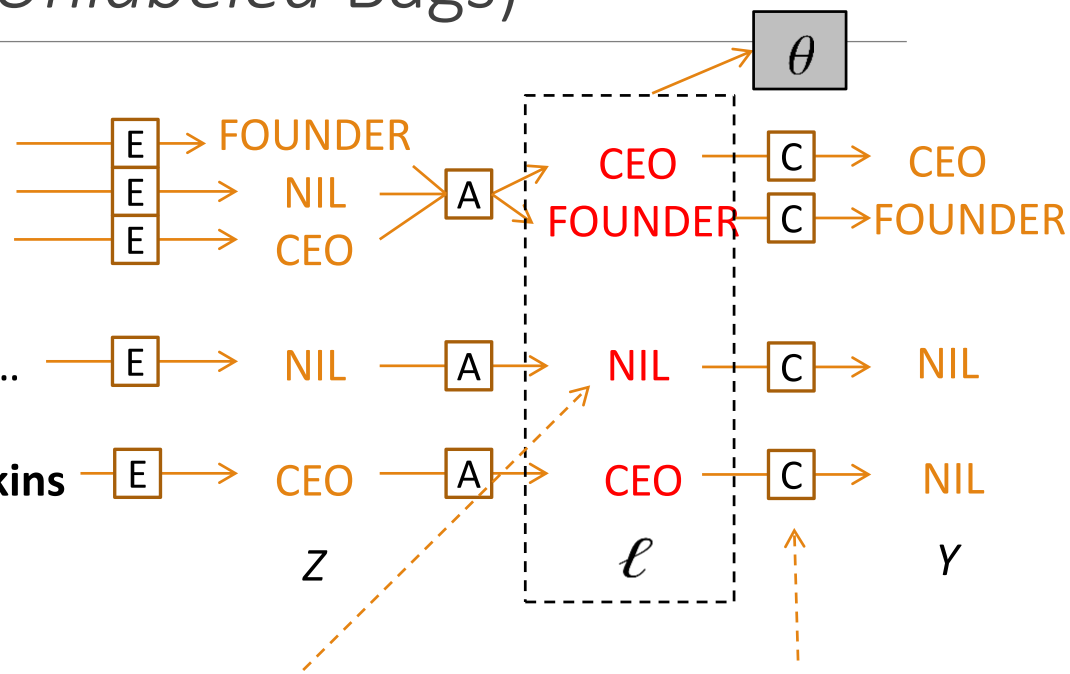
Min et al., Distant Supervision for Relation Extraction with an Incomplete Knowledge Base. (NAACL-HLT 2013)

Reduce False Negatives: Semi-supervised MIML (*Positive* and *Unlabeled* Bags)

Steve Jobs, the co-founder of **Apple** Inc. …
Most books on **Apple** or **Steve Jobs** have …
Apple CEO **Steve Jobs** …

Gary Pard, of NJ's **DeCamp Bus Lines**, said…

Thamesdown Transport Ltd CEO **Paul Jenkins**



X

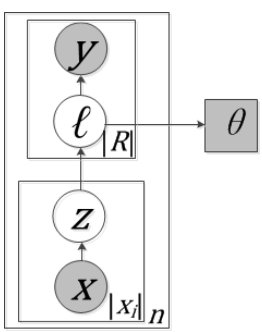


Plate diagram of semi-supervised MIML

Negatives in the unlabeled bags of instances

$$p(y_i^r|\ell_i^r) = \begin{cases} 1/2 & \text{if } y_i^r = P \wedge \ell_i^r = P; \\ 1/2 & \text{if } y_i^r = U \wedge \ell_i^r = P; \\ 1 & \text{if } y_i^r = U \wedge \ell_i^r = N; \\ 0 & \text{otherwise ;} \end{cases}$$

(Compatibility)

Use hard Expectation-Maximization (EM) to approximate $\ell^* = \arg\max_{\ell} p(\ell|\mathbf{y}, \theta, \mathbf{x}; \mathbf{w}_z, \mathbf{w}_\ell)$

NN Relation Extraction: An Open-Source Implementation

An Open-Source Package for Neural Relation Extraction (NRE)
- https://github.com/thunlp/OpenNRE

Largely follow the “distant supervision” paradigm
- Though you can use it for training with supervised dataset such as ACE

Included several algorithms
- CNN, Piecewise CNN
- BERT as classifier
- Attention mechanism at instance or bag (MIL) levels, etc

Supported datasets
- NYT, Wiki80, SemEval
- Few-shot learning dataset

Unsupervised Relation Extraction

Supervised approaches

- Manual annotation of training data; not scalable to Web

Unsupervised relation instance extraction, e.g., TextRunner [Banko et. al., IJCAI'07]

- Extracts noisy & sparse ground facts
- No high-level knowledge that generalizes ground facts

Unsupervised relation extraction

- Unsupervised, domain-independent
- Scales to Web
  - Text → semantic network
  - Abundance of Web text → KB

Semantic Network Extractor

Use open IE (e.g., TextRunner) to extraction triples

- Extracts *(object,relation,object)* triples from webpages in a single pass
- Identify nouns with noun phrase chunker
- Heuristically identify string between two nouns as relation
- Classify each triple as true or false using Naïve Bayes classifier

Semantic Network Extractor

- Input: tuples r(x,y) → Output: simple semantic network
- Clusters objects and relations simultaneously
  - Co-clustering of objects and relations
- Number of clusters need not be specified in advance
- Cluster relations by objects they relate and vice versa

Semantic Network Extractor

Markov Logic

- A logical KB is a set of **hard constraints** on the set of possible worlds
- Let's make them **soft constraints**:
  - When a world violates a formula, it becomes less probable, not impossible
- Give each formula a **weight** (Higher weight $\Rightarrow$ Stronger constraint)

$$P(\text{world}) \propto \exp\left(\sum \text{weights of formulas it satisfies}\right)$$

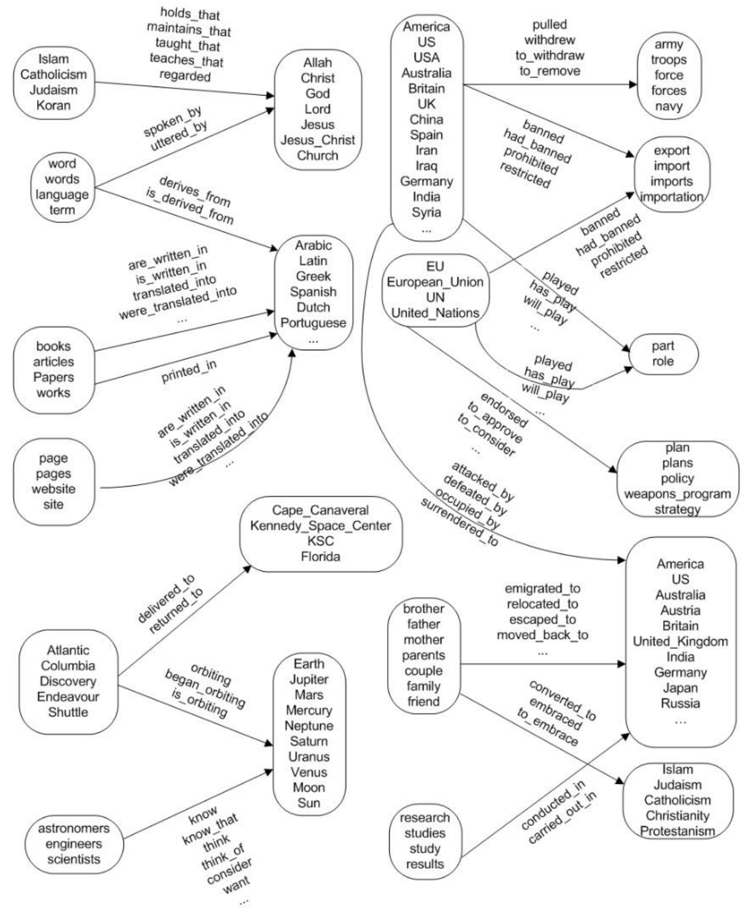
Four simple rules in MLN

- Each symbol belongs to exactly one cluster
- Exponential prior on #cluster combinations
- Most symbols tend to be in different clusters
- Atom prediction rule: Truth value of atom is determined by cluster combination it belongs to

Semantic Network Extractor

Greedy, agglomerative, hard clustering

- Approximation: Hard assignment of symbols to clusters
- Searches over cluster assignments, evaluate each by its log-posterior
- Agglomerative clustering
  - Start with each *r, x, y* symbols in own cluster
  - Merge pairs of clusters in bottom-up manner



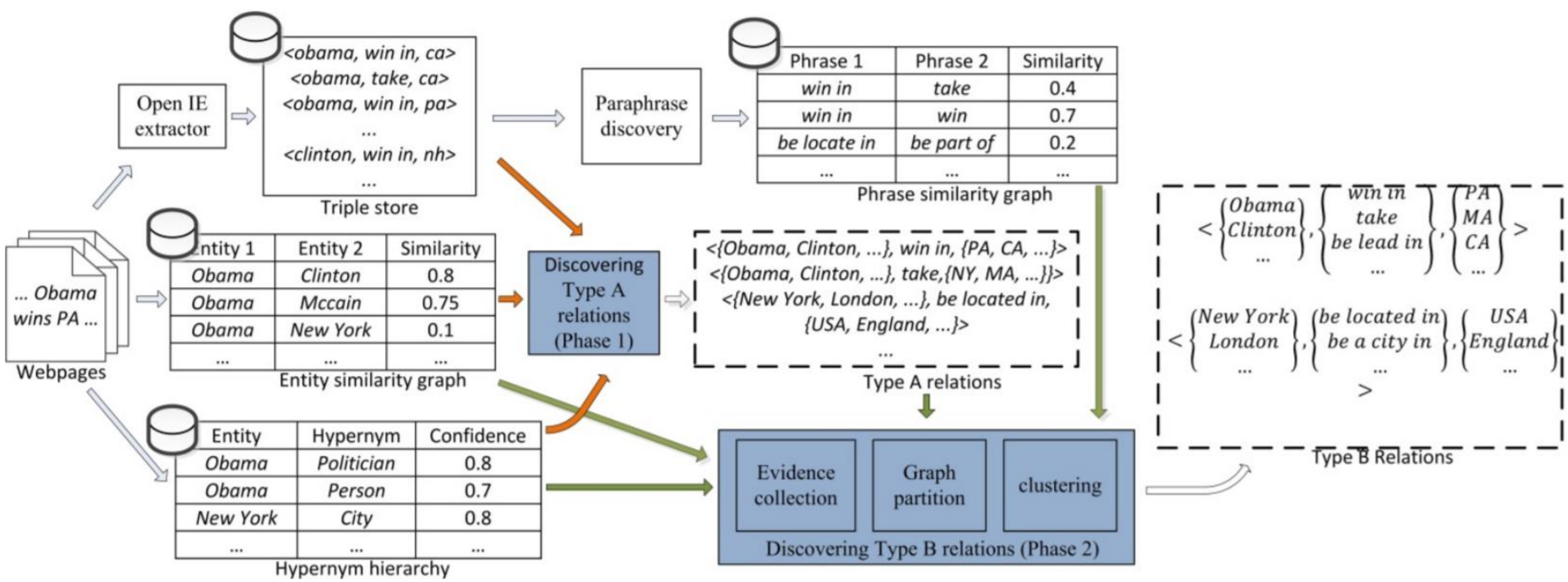
Fragments of a Semantic Network Learned by SNE

Web-Scale Unsupervised Relation Extraction

Applies a hierarchical agglomerative clustering algorithm (similar to SNE)

Key innovations:

- Type A (disambiguating phrases by entity pairs) vs Type B relations
- Ensemble various knowledge sources harvested at web scale:
  - NeedleSeek, http://research.microsoft.com/en-us/projects/needleseek
  - Distributional similarity, phrases that share the same hypernym. What else?
- Various speed optimization for web-scale clustering, e.g., canopy clustering (McCallum et al., 2009)



Min et al., Ensemble Semantics for Large-scale Unsupervised Relation Extraction . EMNLP-CoNLL 2012

A Hierarchy of Relations

Relations gradually merged together when we lowered the threshold in the clustering algorithm

- Different thresholds → relations at different target granularities
- This could provide us the ability to query relations at different granularity.