ClusType: Effective Entity Recognition and Typing by Relation Phrase-Based Clustering

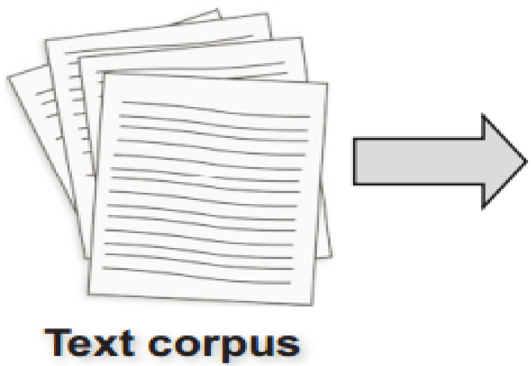
Team Name: DataMind

Pulkit Aggarwal
Usha Amrutha Nookala
Deepak Muralidharan
Anahita Hosseini

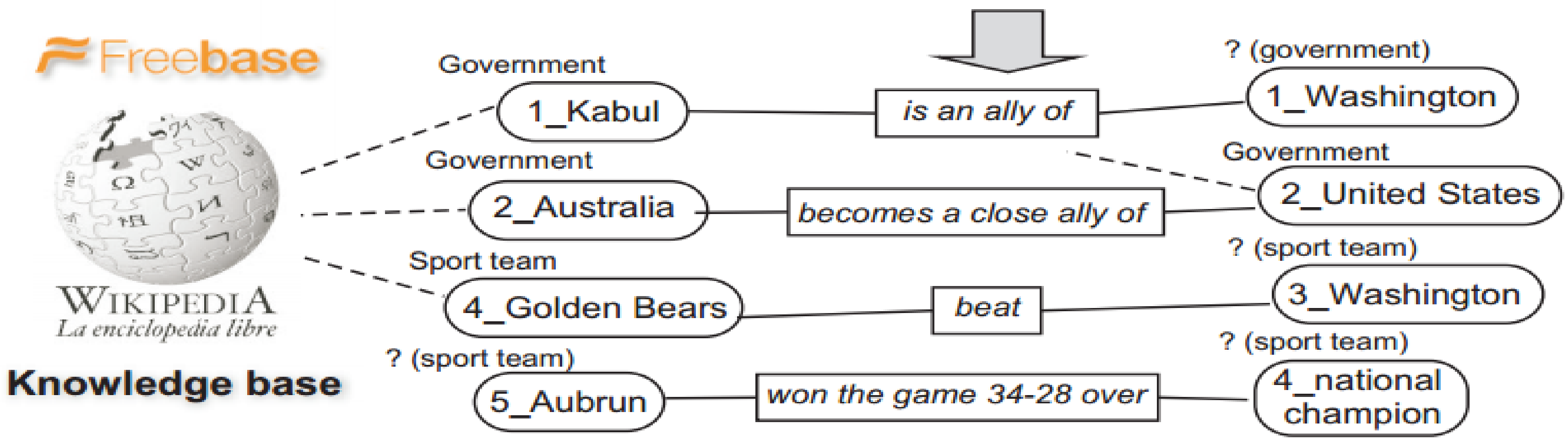
- Overview
- **Goal: Automated typed Entity Recognition**
- **Challenges**
- **Proposed Solution**
- **Heterogeneous Graphs**
- **ClusType Algorithm**
- **Observations and Results**

What is Entity Recognition?

Making sense of large unstructured text corpus



ID	Document Text
1	... has concerns whether **Kabul** *is an ally of* **Washington.**
2	... **Australia** *becomes a close ally of* the **United States**. ...
3	He *has offices in* **Washington, Boston** and **San Francisco**.
4	... The **Cardinal** *will share the title with* **California** if the **Golden Bears** *beat* **Washington** later Saturday. ...
5	... **Auburn** *won the game 34-28 over* the defending **national champions**. ...



What is Entity Recognition?

Entity recognition entails **Identifying** token spans as entity mentions in documents and **labelling** their types

[Obama] arrived this afternoon in [Washington D.C]. [President Obama’s] wife [Michelle] accompanied him.

Entity Types: Person and Location

Linking Entities to Knowledge Base

The criticism consisted primarily of condemnations of mismanagement in response to Hurricane Katrina. Specifically, there was a delayed response to the flooding of New Orleans, Louisiana. New Orleans Mayor Ray Nagin was also criticized for failing to implement his evacuation plan.	Bush was criticized for not returning to Washington, D.C. from his vacation in Texas until after Wednesday afternoon. On the morning of August 28, the president telephoned Mayor Nagin to "plead" for a mandatory evacuation of New Orleans, and Nagin and Gov. Blanco decided to evacuate the city in response to that request

Criticism of government response to the hurricane ...

WIKIPEDIA
The Free Encyclopedia

Link entity mentions to knowledge base entries for in-depth entity information

Linking Entities to Knowledge Base

Drawbacks:

- Human Annotation
- Low Coverage
- Domain Adaptation.
- >50% un-linkable entity mentions in Web Corpus
- > 90% in the paper: Tweets, Yelp reviews.

Goal: Automated Typed Entity Recognition

- Minimize human intervention

- Reduce reliability on Knowledge Base (KB)

- Weak Supervision: Relies on manually selecting seed entities.
  - Pattern-based bootstrapping methods & Label Propagation Methods.
  - Assumption: Seeds are frequent and unambiguous

- Distant Supervision : Rely on information in KBs

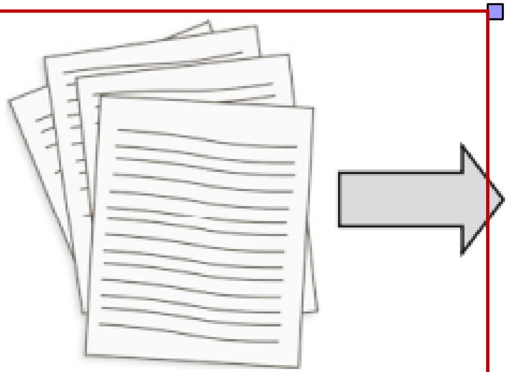
Distant Supervision

Step 1: Extract entities from the text.

Step 2: Map candidates mentions to the KB. (here, Freebase)

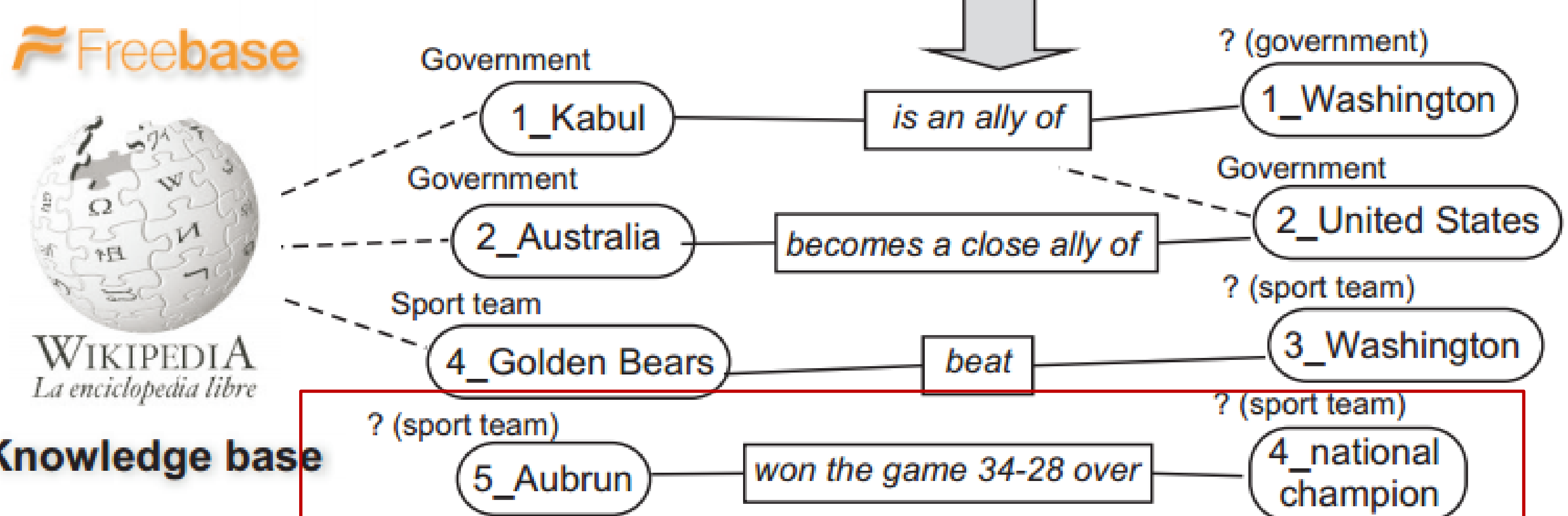
Step 3: Use the type mentions obtained from step 2 to infer the rest of candidate mentions

Distant Supervision



Text corpus

ID	Document Text
1	... has concerns whether **Kabul** *is an ally of* **Washington.**
2	... **Australia** *becomes a close ally of* the **United States**. ...
3	He *has offices in* **Washington, Boston** and **San Francisco**.
4	... The **Cardinal** *will share the title with* **California** if the **Golden Bears** *beat* **Washington** later Saturday. ...
5	... **Auburn** *won the game 34-28 over* the defending **national champions**. ...



Knowledge base

Challenge 1: Domain Restriction

- Existing methods assume entity mentions are already extracted by existing entity detection tools.
- Linguistic features have structured dependency.
- Domain-Specific : tools trained on news articles do not work well with other emerging domains (tweets).

Challenge 2: Name Ambiguity

Entity names are often ambiguous— ***multiple entities may share the same surface name.***

Govt.

- …has concern that Kabul is an ally of Washington.

City

- He has office in Washington, Boston and San Francisco

Team

- While Griffinis not the part of Washington's plan on Sunday's game, …

Challenge 3: Contextual Sparsity

A variety of contextual clues are leveraged to find sources of shared semantics across different entities.

ID	Sentence	Frequency
1	**The magnitude 9.0 quake caused widespread devastation in [Kesennuma city]**	12
2	**… tsunami that ravaged [northeastern Japan] last Friday**	31
3	**The resulting tsunami devastate [Japan]'s northeast**	244

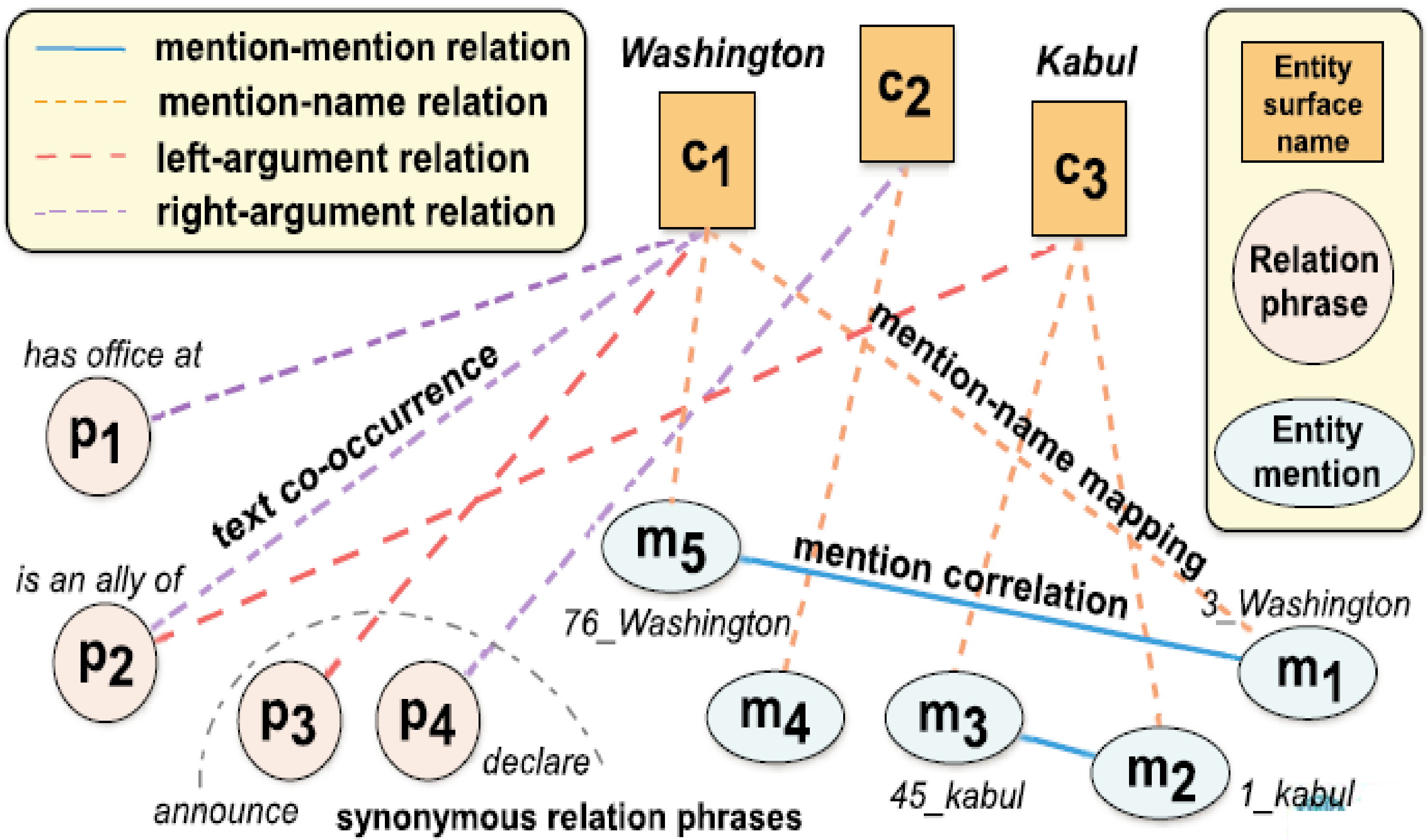
Proposed Solution

- **Avoid Domain Restriction- Phrase Mining Algorithm**
 Extract candidate entity mentions and relation phrases with minimal linguistic/domain assumption

- **Eliminate Name Ambiguity – Mention Correlation**
 model each mention based on its surface name and context, instead of simply merging identical surface names.

- **Overcome Contextual Sparsity- Soft Clustering**
Mine synonymous *relation phrase* co-occurring with entity mentions

A Relation Phrase-Based Entity Recognition Framework



- Generate candidate entity mentions and relation phrases simultaneously.
- Construct heterogeneous graphs to represent information in text corpus.

Candidate Generation – Phrase Mining Algorithm

Generate candidates based on:

- Global Significance Score: Filter low-quality, and insignificant Candidates.
- Generic POS tag patterns: remove phrases which do not comply with syntactic constraints.

Partitions corpus into segments which meet both significance threshold and POS patterns

> Over:RP the weekend the system:EP dropped:RP nearly inches of snow in:RP western Oklahoma:EP and at:RP [Dallas Fort Worth International Airport]:EP sleet and ice caused:RP hundreds of [flight cancellations]:EP and delays. ...... It is forecast:RP to reach:RP [northern Georgia]:EP by:RP [Tuesday afternoon]:EP, Washington:EP and [New York]:EP by:RP [Wednesday afternoon]:EP, meteorologists:EP said:RP.

EP: entity mention candidate; RP: relation phrase

Heterogeneous Graphs

Graphs are constructed based on the following objects:

- Candidate Entity Mentions.
- Entity Surface Names.
- Relation Phrases.

Modeling Type for Entity Mention

- Type of entity mention can be obtained from:

- ✓ Type Distribution (Entity Mention – Surface Names sub graph).
- ✓ Type signature of its surrounding relation phrases.

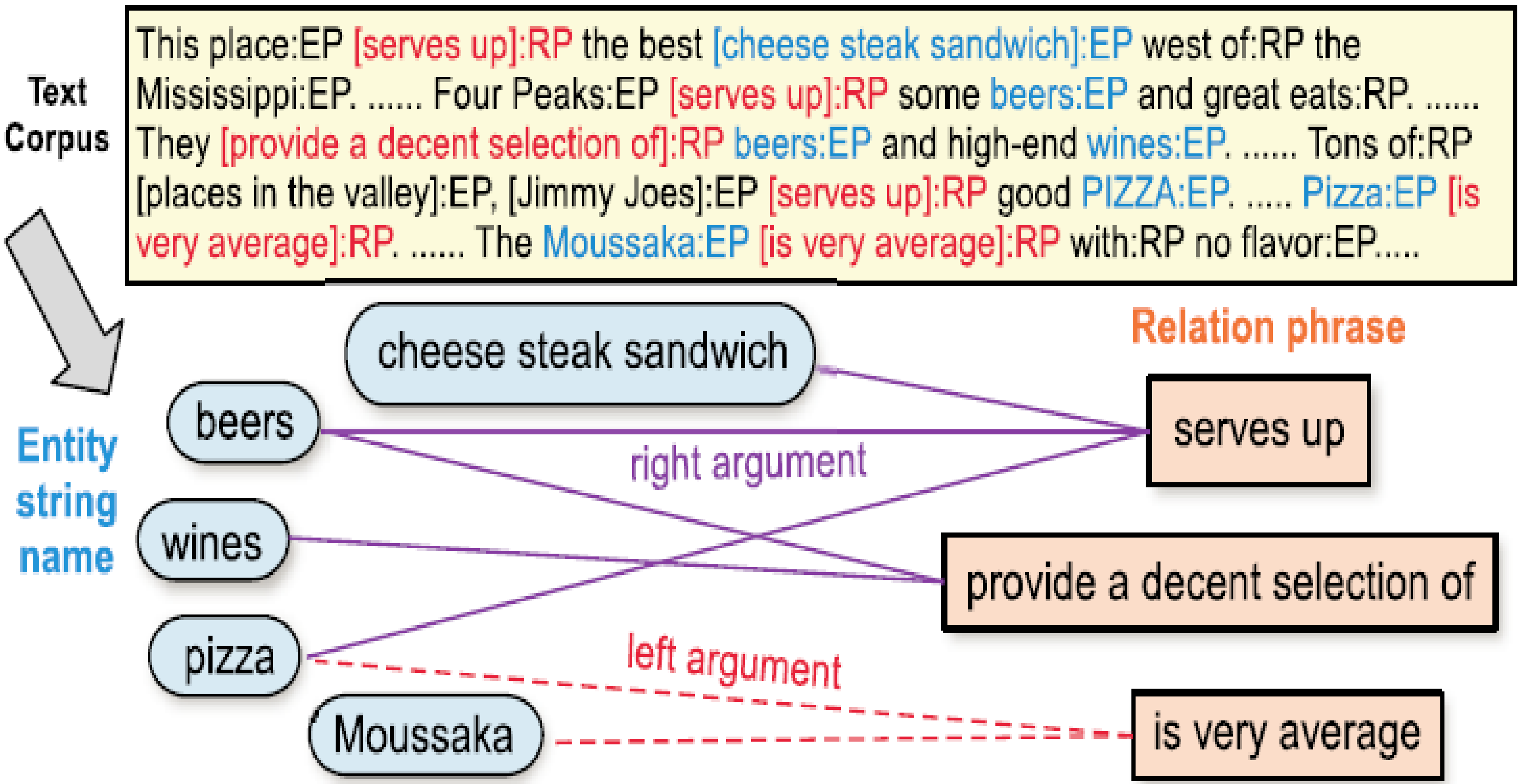
Type Signature of RP: [Type left arg.], [Type Right arg.]

Kabul is an ally of Washington

E.g. “ is an ally” :

[Kabul: Government], [Washington: Government]

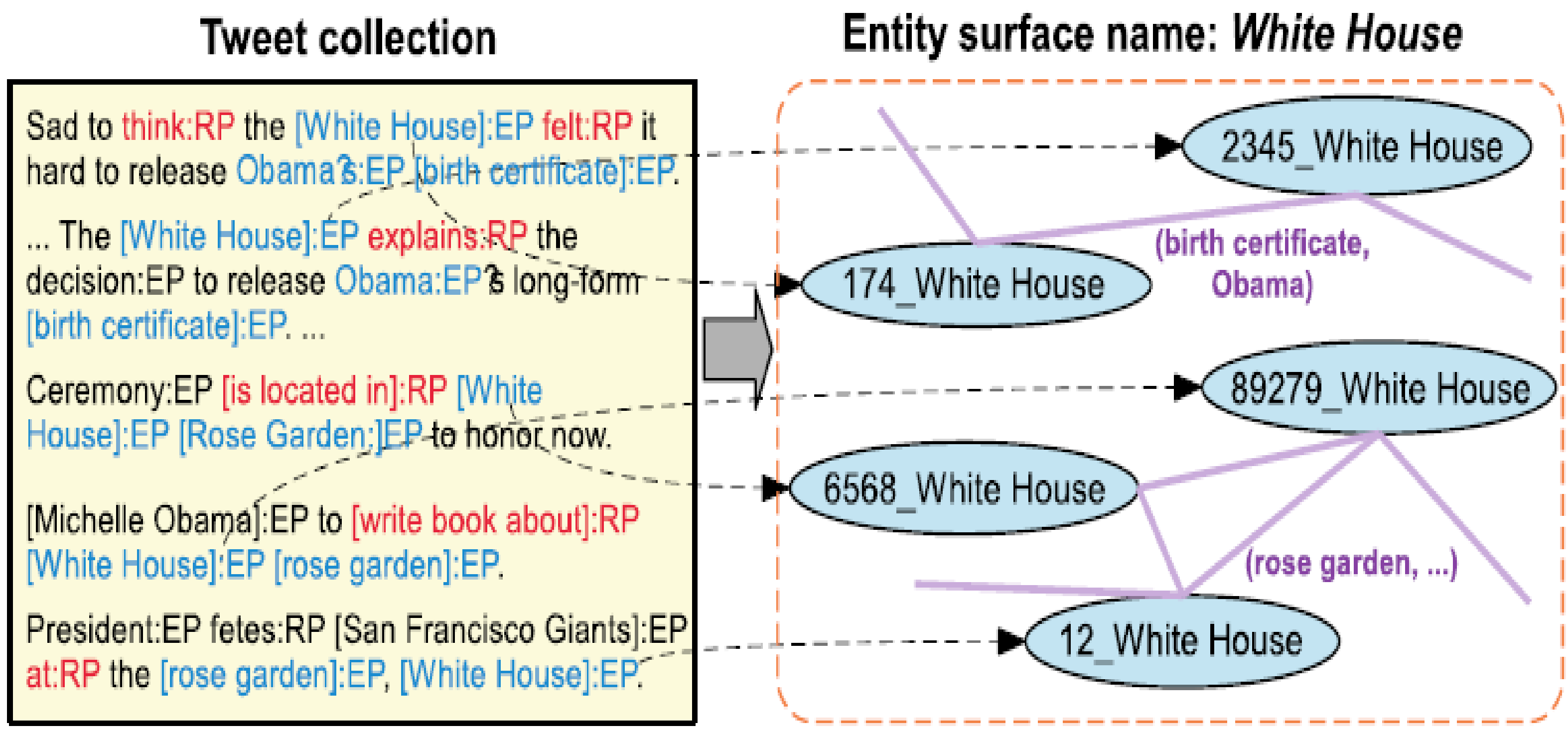
Entity-Name Relation Phrase Sub graph



HYPOTHESIS 1 (ENTITY-RELATION CO-OCCURRENCES) *If surface name c often appears as the left (right) argument of relation phrase p, then c's type indicator tends to be similar to the corresponding type indicator in p's type signature.*

Mention Correlation Sub graph

- Co-occuring mentions can provide good hints to avoid name-ambiguity.



HYPOTHESIS 2 (MENTION CORRELATION). *If there exists a strong correlation (i.e., within sentence, common neighbor mentions) between two candidate mentions that share the same name, then their type indicators tend to be similar.*

How To Cluster Relation Phrases Jointly?

- Many extracted relation phrases have very few occurrences in the corpus.
- 37% of relation phrases have less than 3 unique entity surface names.
- Idea- Infer type signature of infrequent(sparse) relation phrases using type signature of frequent relation phrases.
- What signals are we considering?

1. String similarity

HYPOTHESIS 3 (TYPE SIGNATURE CONSISTENCY). *If two relation phrases have similar cluster memberships, the type indicators of their left and right arguments (type signature) tend to be similar, respectively.*

HYPOTHESIS 4 (RELATION PHRASE SIMILARITY). *Two relation phrases tend to have similar cluster memberships, if (1) their strings are similar; (2) their context words are similar; and (3) the type indicators of their left and right arguments are similar, respectively.*

Type Inference:Joint Optimization Problem

- Two optimization tasks:-

1.Type propagation over both the type indicators of entity names C and the type signatures of relation phrases on the heterogeneous graph G by way of graph-based semi-supervised learning.

2. Multi-view relation phase clustering.

$$\mathcal{O}_{\alpha,\gamma,\mu} = \mathcal{F}(\mathbf{C}, \mathbf{P}_L, \mathbf{P}_R) + \mathcal{L}_\alpha\big(\mathbf{P}_L, \mathbf{P}_R, \{\mathbf{U}^{(v)}, \mathbf{V}^{(v)}\}, \mathbf{U}^*\big) + \Omega_{\gamma,\mu}(\mathbf{Y}, \mathbf{C}, \mathbf{P}_L, \mathbf{P}_R). \quad (2)$$

Mention modeling & mention correlation (H.2)

$$\mathcal{F}(\mathbf{C}, \mathbf{P}_L, \mathbf{P}_R) = \sum_{i=1}^{n}\sum_{j=1}^{l} W_{L,ij}\left\|\frac{\mathbf{C}_i}{\sqrt{D_{L,ii}^{(\mathcal{C})}}} - \frac{\mathbf{P}_{L,j}}{\sqrt{D_{L,jj}^{(\mathcal{P})}}}\right\|_2^2 + \sum_{i=1}^{n}\sum_{j=1}^{l} W_{R,ij}\left\|\frac{\mathbf{C}_i}{\sqrt{D_{R,ii}^{(\mathcal{C})}}} - \frac{\mathbf{P}_{R,j}}{\sqrt{D_{R,jj}^{(\mathcal{P})}}}\right\|_2^2$$

$$\Omega_{\gamma,\mu}(\mathbf{Y}, \mathbf{C}, \mathbf{P}_L, \mathbf{P}_R) = \left\|\mathbf{Y} - f(\Pi_{\mathcal{C}}\mathbf{C}, \Pi_L\mathbf{P}_L, \Pi_R\mathbf{P}_R)\right\|_F^2 + \frac{\gamma}{2}\sum_{c\in\mathcal{C}}\sum_{i,j=1}^{M_c} W_{ij}^{(c)}\left\|\frac{\mathbf{Y}_i}{\sqrt{D_{ii}^{(c)}}} - \frac{\mathbf{Y}_j}{\sqrt{D_{jj}^{(c)}}}\right\|_2^2 + \mu\|\mathbf{Y} - \mathbf{Y}_0\|_F^2$$

Type propagation between entity surface names and relation phrases (H.1)

$$\mathcal{L}_\alpha\big(\mathbf{P}_L, \mathbf{P}_R, \{\mathbf{U}^{(v)}, \mathbf{V}^{(v)}\}, \mathbf{U}^*\big) = \sum_{v=0}^{d}\beta^{(v)}\big(\|\mathbf{F}^{(v)} - \mathbf{U}^{(v)}\mathbf{V}^{(v)T}\|_F^2 + \alpha\|\mathbf{U}^{(v)}\mathbf{Q}^{(v)} - \mathbf{U}^*\|_F^2\big). \quad (3)$$

Multi-view relation phrases clustering (H.3 & 4)

ClusType - Input Parameters

The ClusType algorithm:

Update type indicators and type signatures

$$\mathbf{Y}^{(c)} = \left[(1+\gamma+\mu)\mathbf{I}_c - \gamma \mathbf{S}_{\mathcal{M}}^{(c)}\right]^{-1}\left(\Theta^{(c)} + \mu \mathbf{Y}_0^{(c)}\right), \ \forall c \in \mathcal{C}, \quad (7)$$

$$\mathbf{C} = \frac{1}{2}\left[\mathbf{S}_L\mathbf{P}_L + \mathbf{S}_R\mathbf{P}_R + \Pi_{\mathcal{C}}^T(\mathbf{Y} - \Pi_L\mathbf{P}_L - \Pi_R\mathbf{P}_R)\right]; \quad (8)$$

$$\mathbf{P}_L = \mathbf{X}_0^{-1}\left[\mathbf{S}_L^T\mathbf{C} + \Pi_L^T(\mathbf{Y} - \Pi_{\mathcal{C}}\mathbf{C} - \Pi_R\mathbf{P}_R) + \beta^{(0)}\mathbf{U}^{(0)}\mathbf{V}^{(0)T}\right];$$

$$\mathbf{P}_R = \mathbf{X}_1^{-1}\left[\mathbf{S}_R^T\mathbf{C} + \Pi_R^T(\mathbf{Y} - \Pi_{\mathcal{C}}\mathbf{C} - \Pi_L\mathbf{P}_L) + \beta^{(1)}\mathbf{U}^{(1)}\mathbf{V}^{(1)T}\right];$$

For each view, performs single-view NMF until converges

$$V_{jk}^{(v)} = V_{jk}^{(v)} \frac{[\mathbf{F}^{(v)T}\mathbf{U}^{(v)}]_{jk} + \alpha \sum_{i=1}^{l} U_{ik}^{*} U_{ik}^{(v)}}{\mathbf{\Delta}_{jk}^{(v)} + \alpha\left(\sum_{i=1}^{l} U_{ik}^{(v)2}\right)\left(\sum_{i=1}^{T} V_{ik}^{(v)}\right)}, \quad (9)$$

$$U_{ik}^{(v)} = U_{ik}^{(v)} \frac{[\mathbf{F}^{(v)+}\mathbf{V}^{(v)} + \alpha \mathbf{U}^{*}]_{ik}}{[\mathbf{U}^{(v)}\mathbf{V}^{(v)T}\mathbf{V}^{(v)} + \mathbf{F}^{(v)-}\mathbf{V}^{(v)} + \alpha \mathbf{U}^{(v)}]_{ik}}. \quad (10)$$

Update consensus matrix and relative weights of different views

$$\mathbf{U}^{*} = \frac{\sum_{v=0}^{d} \beta^{(v)}\mathbf{U}^{(v)}\mathbf{Q}^{(v)}}{\sum_{v=0}^{d} \beta^{(v)}}; \quad \beta^{(v)} = -\log\left(\frac{\delta^{(v)}}{\sum_{i=0}^{d} \delta^{(i)}}\right). \quad (12)$$

Until the objective converges

$$\min_{\substack{\mathbf{Y}, \mathbf{C}, \mathbf{P}, \mathbf{P}_R, \mathbf{U}^* \\ \{\mathbf{U}^{(v)}, \mathbf{V}^{(v)}, \beta^{(v)}\}}} \mathcal{O}_{\alpha,\gamma,\mu,\lambda_L,\lambda_\Omega}$$

$$\text{s.t.} \quad \mathbf{Y} \in \{0,1\}^{M \times T}, \quad \mathbf{Y}\mathbf{1} = \mathbf{1};$$

$$\mathbf{U}^* \geq 0, \ \mathbf{U}^{(v)} \geq 0, \ \mathbf{V}^{(v)} \geq 0;$$

$$\sum_{v=0}^{d} \exp(-\beta^{(v)}) = 1, \ \forall 0 \leq v \leq d.$$

Y : Type indicator matrix for mentions
C : Type indicator matrix for surface names
P_L,P_R : Type signature matrix for relation phrases
U : Cluster membership matrix for relation phrases
V : Type indicator matrix for relation phrase clusters
U* : Consensus matrix
β : weightage for information among different views

The ClusType Algorithm

Algorithm 1 The **ClusType** algorithm

Input: biadjacency matrices $\{\Pi_\mathcal{C}, \Pi_L, \Pi_R, \mathbf{W}_L, \mathbf{W}_R, \mathbf{W}_\mathcal{M}\}$, clustering features $\{\mathbf{F}_s, \mathbf{F}_c\}$, seed labels $\mathbf{Y}_0$, number of clusters K, parameters $\{\alpha, \gamma, \mu\}$

Initialize $\{\mathbf{Y}, \mathbf{C}, \mathbf{P}_L, \mathbf{P}_R\}$ with $\{\mathbf{Y}_0, \Pi_\mathcal{C}^T\mathbf{Y}_0, \Pi_L^T\mathbf{Y}_0, \Pi_R^T\mathbf{Y}_0\}$, $\{\mathbf{U}^{(v)}, \mathbf{V}^{(v)}, \beta^{(v)}\}$ and $\mathbf{U}^*$ with positive values.
repeat
    Update candidate mention type indicator $\mathbf{Y}$ by Eq. (7)
    Update entity name type indicator $\mathbf{C}$ and relation phrase type signature $\{\mathbf{P}_L, \mathbf{P}_R\}$ by Eq. (8)
    for $v = 0$ to 3 **do**
        repeat
            Update $\mathbf{V}^{(v)}$ with Eq. (9)
            Normalize $\mathbf{U}^{(v)} = \mathbf{U}^{(v)}\mathbf{Q}^{(v)}$, $\mathbf{V}^{(v)} = \mathbf{V}^{(v)}\mathbf{Q}^{(v)-1}$
            Update $\mathbf{U}^{(v)}$ by Eq. (10)
        until Eq. (11) converges
    end for
    Update consensus matrix $\mathbf{U}^*$ and relative feature weights $\{\beta^{(v)}\}$ using Eq. (12)
until the objective $\mathcal{O}$ in Eq. (6) converges
Predict the type of $m_i \in \mathcal{M}_U$ by $\text{type}(m_i) = \text{argmax}\, Y_i$.

- Efficiently solved by alternate minimization based on block coordinate descent algorithm
- Algorithm complexity is linear to # entity mentions, #relation phrases, #cluster, #clustering features and #target types

Comparison of Clustype with other methods

Data sets	NYT			Yelp			Tweet		
Method	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Pattern [9]	0.4576	0.2247	0.3014	0.3790	0.1354	0.1996	0.2107	0.2368	0.2230
FIGER [16]	0.8668	0.8964	0.8814	0.5010	0.1237	0.1983	**0.7354**	0.1951	0.3084
SemTagger [12]	0.8667	0.2658	0.4069	0.3769	0.2440	0.2963	0.4225	0.1632	0.2355
APOLLO [29]	0.9257	0.6972	0.7954	0.3534	0.2366	0.2834	0.1471	0.2635	0.1883
NNPLB [15]	0.7487	0.5538	0.6367	0.4248	0.6397	0.5106	0.3327	0.1951	0.2459
ClusType-NoClus	0.9130	0.8685	0.8902	0.7629	0.7581	0.7605	0.3466	0.4920	0.4067
ClusType-NoWm	0.9244	0.9015	0.9128	0.7812	0.7634	0.7722	0.3539	**0.5434**	0.4286
ClusType-TwoStep	0.9257	0.9033	0.9143	0.8025	0.7629	0.7821	0.3748	0.5230	0.4367
ClusType	**0.9550**	**0.9243**	**0.9394**	**0.8333**	**0.7849**	**0.8084**	0.3956	0.5230	**0.4505**

- F1 score, precision, recall as evaluation metrics.

- ClusType obtained **46.08%** improvement on F1 score and **168%** improvement in recall compared to best baseline FIGER on Tweet datasets.

Comparison of Clustype with other methods

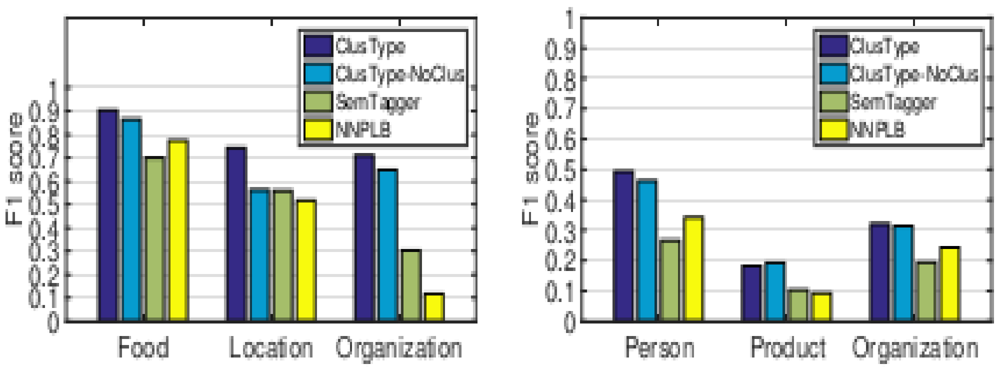
Data sets	NYT			Yelp			Tweet		
Method	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Pattern [9]	0.4576	0.2247	0.3014	0.3790	0.1354	0.1996	0.2107	0.2368	0.2230
FIGER [16]	0.8668	0.8964	0.8814	0.5010	0.1237	0.1983	**0.7354**	0.1951	0.3084
SemTagger [12]	0.8667	0.2658	0.4069	0.3769	0.2440	0.2963	0.4225	0.1632	0.2355
APOLLO [29]	0.9257	0.6972	0.7954	0.3534	0.2366	0.2834	0.1471	0.2635	0.1883
NNPLB [15]	0.7487	0.5538	0.6367	0.4248	0.6397	0.5106	0.3327	0.1951	0.2459
ClusType-NoClus	0.9130	0.8685	0.8902	0.7629	0.7581	0.7605	0.3466	0.4920	0.4067
ClusType-NoWm	0.9244	0.9015	0.9128	0.7812	0.7634	0.7722	0.3539	**0.5434**	0.4286
ClusType-TwoStep	0.9257	0.9033	0.9143	0.8025	0.7629	0.7821	0.3748	0.5230	0.4367
ClusType	**0.9550**	**0.9243**	**0.9394**	**0.8333**	**0.7849**	**0.8084**	0.3956	0.5230	**0.4505**

- **FIGER**: effectiveness of our candidate generation and proposed hypotheses on type propagation

- **NNPLB and APOLLO**: ClusType not only utilizes semantic-rich relation phrase as type cues, but only cluster synonymous relation phrases to tackle context sparsity

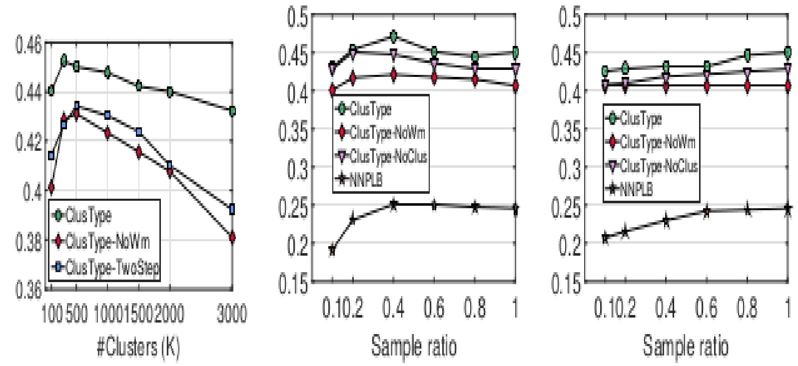
- **variants**: (i) models mention correlation for name disambiguation; (ii) integrates clustering in a mutually enhancing way

Further Comparisons



(a) Yelp (b) Tweet

Figure 6: Performance breakdown by types.



(a) #Clusters (b) Seed set size (c) Corpus size

Figure 7: Performance changes in F1 score with #clusters, #seeds and corpus size on Tweets.

- ❑ It obtains larger gain on types organization and person, which have more entities with ambiguous surface names.

- ❑ This depicts that type propagation on mention-mention subgraph is crucial.

Case study

Table 8: Example relation phrase clusters and their corpus frequency from the NYT dataset.

ID	Relation phrase
1	recruited by (5.1k); employed by (3.4k); want hire by (264)
2	go against (2.4k); struggling so much against (54); run for re-election against (112); campaigned against (1.3k)
3	looking at ways around (105); pitched around (1.9k); echo around (844); present at (5.5k);

- Synonymous as well as sparse relation phrases are clustered together.
- Type information of sparse relation phrases is boosted by using frequent relation phrases.

Table 7: Example output of ClusType and the compared methods on the Yelp dataset.

ClusType	SemTagger	NNPLB
The best **BBQ:Food** I've tasted in **Phoenix:LOC** ! I had the **[pulled pork sandwich]:Food** with **coleslaw:Food** and **[baked beans]:Food** for lunch. ...	The best BBQ I've tasted in **Phoenix:LOC** ! I had the pulled **[pork sandwich]:LOC** with **coleslaw:Food** and **[baked beans]:LOC** for lunch. ...	The best **BBQ:Loc** I've tasted in **Phoenix:LOC** ! I had the pulled pork **sandwich:Food** with coleslaw and baked **beans:Food** for **lunch:Food**. ...
I only go to **ihop:LOC** for **pancakes:Food** because I don't really like anything else on the menu. Ordered **[chocolate chip pancakes]:Food** and a **[hot chocolate]:Food**.	I only go to ihop for pancakes because I don't really like anything else on the menu. Ordered **[chocolate chip pancakes]:LOC** and a **[hot chocolate]:LOC**.	I only go to ihop for pancakes because I don't really like anything else on the menu. Ordered chocolate chip pancakes and a hot chocolate.

ClusType extracts more entity mention candidates (e.g., "BBQ", "ihop") and predicts their types with better accuracy (e.g., "baked beans", "pulled pork sandwich").

References

[1] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Tay-lor. Freebase: a collaboratively created graph database for structuring human knowledge. In SIGMOD, 2008

.

[2] X. L. Dong, T. Strohmann, S. Sun, and W. Zhang. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In SIGKDD, 2014.

[3] A. El-Kishky, Y. Song, C. Wang, C. R. Voss, and J. Han. Scalable topical phrase mining from text corpora. VLDB,2015

.

[4] A. Fader, S. Soderland, and O. Etzioni. Identifying relations for open information extraction. In EMNLP, 2011.

[5] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In ACL, 2005.

[6] L. Galárraga, G. Heitz, K. Murphy, and F. M. Suchanek. Canonicalizing open knowledge bases. In CIKM, 2014.

[7] S. Gupta and C. D. Manning. Improved pattern learning for bootstrapped entity extraction. In CONLL, 2014.

Thank You !!

Q&A