COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Computational Frameworks for Semantic Analysis and Wikification

Dan Roth

Department of Computer Science

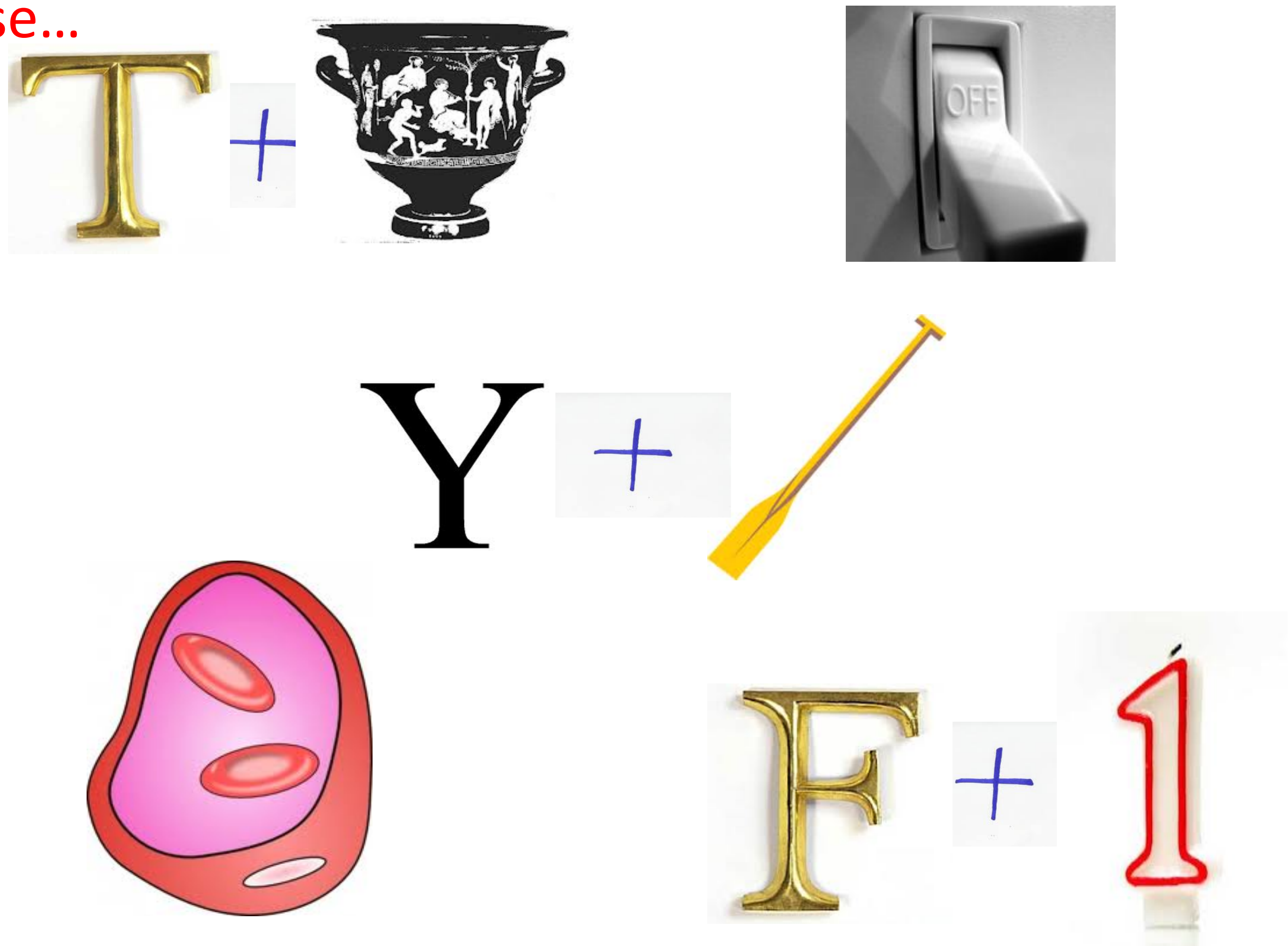
University of Illinois at Urbana-Champaign

With thanks to:

Collaborators: **Ming-Wei Chang, Xiao Cheng, Lev Ratinov, Vivek Srikumar, Many others**

Funding: NSF; DHS; NIH; DARPA; IARPA; ARL
DASH Optimization (Xpress-MP); GUROBI Optimization

Please…



Learning and Inference

- Global decisions in which several local decisions play a role but there are mutual dependencies on their outcome.
  - In current NLP we often think about simpler structured problems: Parsing, Information Extraction, SRL, etc.
  - As we move up the problem hierarchy (Textual Entailment, QA,….) not all component models will be learned simultaneously
  - We need to think about (learned) models for different sub-problems
  - Knowledge relating sub-problems (constraints) becomes more essential and may appear only at evaluation time
- Goal: Incorporate models' information, along with prior knowledge (constraints) in making coherent decisions
  - Decisions that respect the local models as well as domain & context specific knowledge/constraints.

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England. 2. Winnie the Pooh is a title of a book.

3. Christopher Robin is an author. 4. Christopher Robin must be at least 65 now.

This is an Inference Problem

Relational Inference

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party

- What does Socialist Party refer to?
- There is a need to "look up" some information...
  - What and how to look up is determined by understanding local relations
  - These relations need to be coupled with relevant statistical models to support a decision

Inference with General Constraint Structure [Roth&Yih

Recognizing Entities and Relations

Improvement over no inference: 2-5%

other	0.05

other	0.10

othe	0.05

Y = argm

= argm

…

score(R_1 = S-of)¢ [[R_1

Subject to Constraints

An Objective function that incorporates learned models with knowledge (constraints)

A constrained Conditional Model

irrelevant	0.05
spouse_of	**0.45**
born_in	0.50

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

Models could be learned separately; constraints may come up only at decision time.

Constrained Conditional Models

Penalty for violating the constraint.

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

(Soft) constraints component

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

How far y is from a "legal" assignment

Not Today

How to solve?

This is an Integer Linear Program

Solving using ILP packages gives an exact solution.

Cutting Planes, Dual Decomposition & other search techniques are possible

How to train?

Training is learning the objective function

Decouple? Decompose?

How to exploit the structure to minimize supervision?

Outline

- Integer Linear Programming Formulations for Natural Language Processing

- Example 1: Extended Semantic Role Labeling
  - Relaxing the pipeline
  - Dealing with lack of joint annotation: combining structured models

- Example 2: Wikification
  - Knowledge Acquisition by Grounding
  - Relational Inference for Wikification
  - Applications

Examples: CCM Formulations

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

CCMs can be viewed as a general interface to easily combine declarative domain knowledge with data driven statistical models

Formulate NLP Problems as ILP problems (inference may be done otherwise)

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Constrained Conditional Models Allow:

- **Learning a simple model (or multiple; or pipelines)**
- **Make decisions with a more complex model**
- **Accomplished by directly incorporating constraints to bias/re-rank global decisions composed of simpler models' decisions**
- **More sophisticated algorithmic approaches exist to bias the output**
  [CoDL: Cheng et. al 07,12; PR: Ganchev et. al. 10; DecL, UEM: Samdani et. al 12]

Semantic Role Labeling

Archetypical Information Extraction Problem: E.g., Concept Identification and Typing, Event Identification, etc.

I *left* my pearls to my daughter in my will .

[I]$_{A0}$ *left* [my pearls]$_{A1}$ [to my daughter]$_{A2}$ [in my will]$_{AM-LOC}$.

- ***A0*** Leaver
- ***A1*** Things left
- ***A2*** Benefactor
- ***AM-LOC*** Location

I *left* my pearls to my daughter in my will .

Cognitive Computation Group
University of Illinois at Urbana-Champaign

Algorithmic Approach

can

No duplic
argumen

Learning Based Java: allows a developer to encode constraints in First Order Logic; these are compiled into linear inequalities automatically.

$$\forall y \in \mathcal{Y}_R,\ \sum_{i=0}^{n-1} 1_{\{y_i = y = \text{"R-Ax"}\}} \le \sum_{i=0}^{n-1} 1_{\{y_i = \text{"Ax"}\}}$$

$$\forall j, y \in \mathcal{Y}_C,\ 1_{\{y_j = y = \text{"C-Ax"}\}} \le \sum_{i=0}^{j} 1_{\{y_i = \text{"Ax"}\}}$$

- Identify argument candid
  - Pruning [Xue&Palmer, EM
  - Argument Identifier
    - Binary classification
- Classify argument candid
  - Argument Classifier
    - Multi-class classification
- Inference

argmax $\sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t}\, c_{a=t}$

Subject to:

- One label per argument: $\sum_t y^{a,t} = 1$
- No overlapping or embedding
- Relations between verbs and arguments,….

given

Use the **pipeline architecture's simplicity** while **maintaining uncertainty**: keep probability distributions over decisions & use global inference at decision time.

Constrained Conditional Models [Roth & Yih '04, Chang et. al '12]

Penalty for violating the constraint.

$$\underset{y}{\operatorname{argmax}} \boldsymbol{\lambda} \cdot F(x,y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

(Soft) constraints component

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination (modeled as Boolean variables)

How far y is from a "legal" assignment

How to solve?

This is an Integer Linear Program

Solving using ILP packages gives an exact solution.

Cutting Planes, Dual Decomposition & other search techniques are possible

How to train?

Training is learning the objective function

Decouple? Decompose?

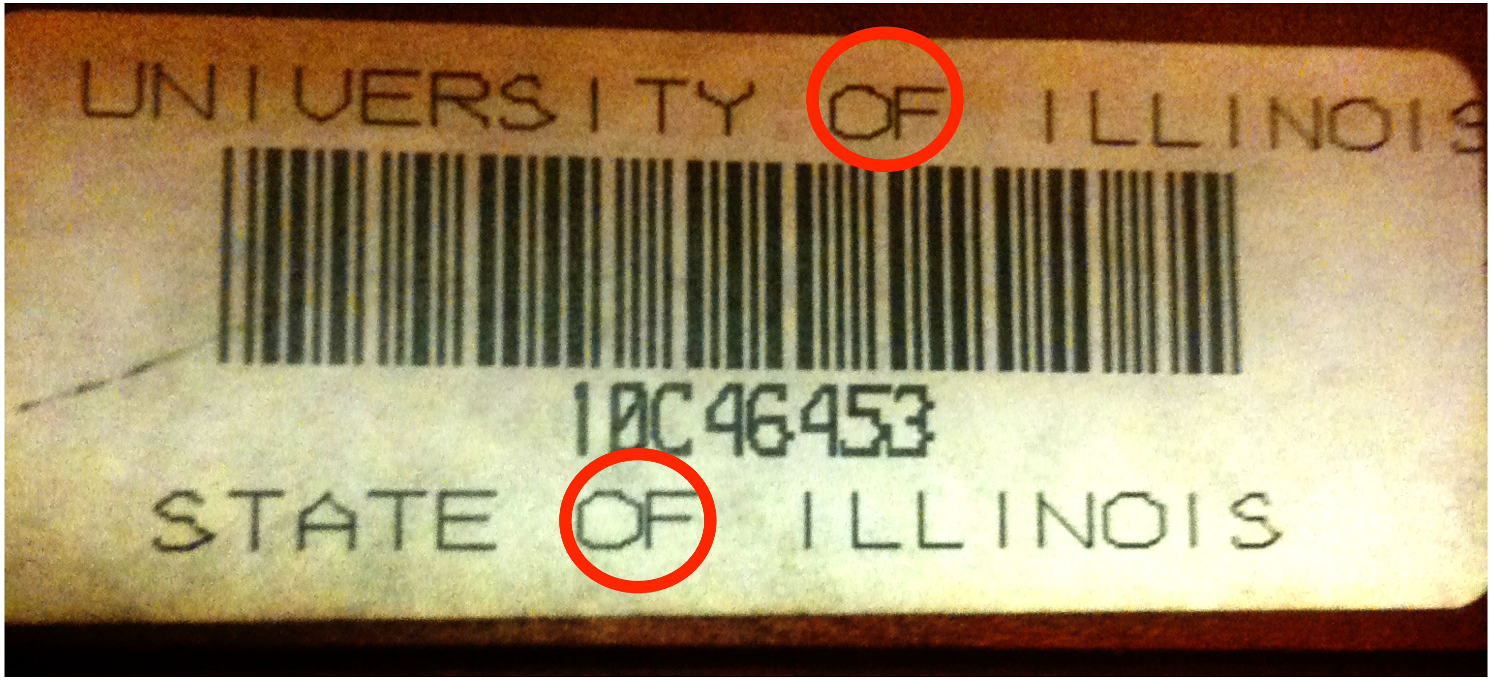
How to exploit the structure to minimize supervision?

Verb SRL is not Sufficient

- *John, a fast-rising politician, slept on the train to Chicago.*
- **Verb Predicate: sleep**
  - Sleeper: John, a fast-rising politician
  - Location: on the train to Chicago
- **Who was John?**
  - Relation: Apposition (comma)
  - John, a fast-rising politician
- **What was John's destination?**
  - Relation: Destination (preposition)
  - train to Chicago

Examples of preposition relations

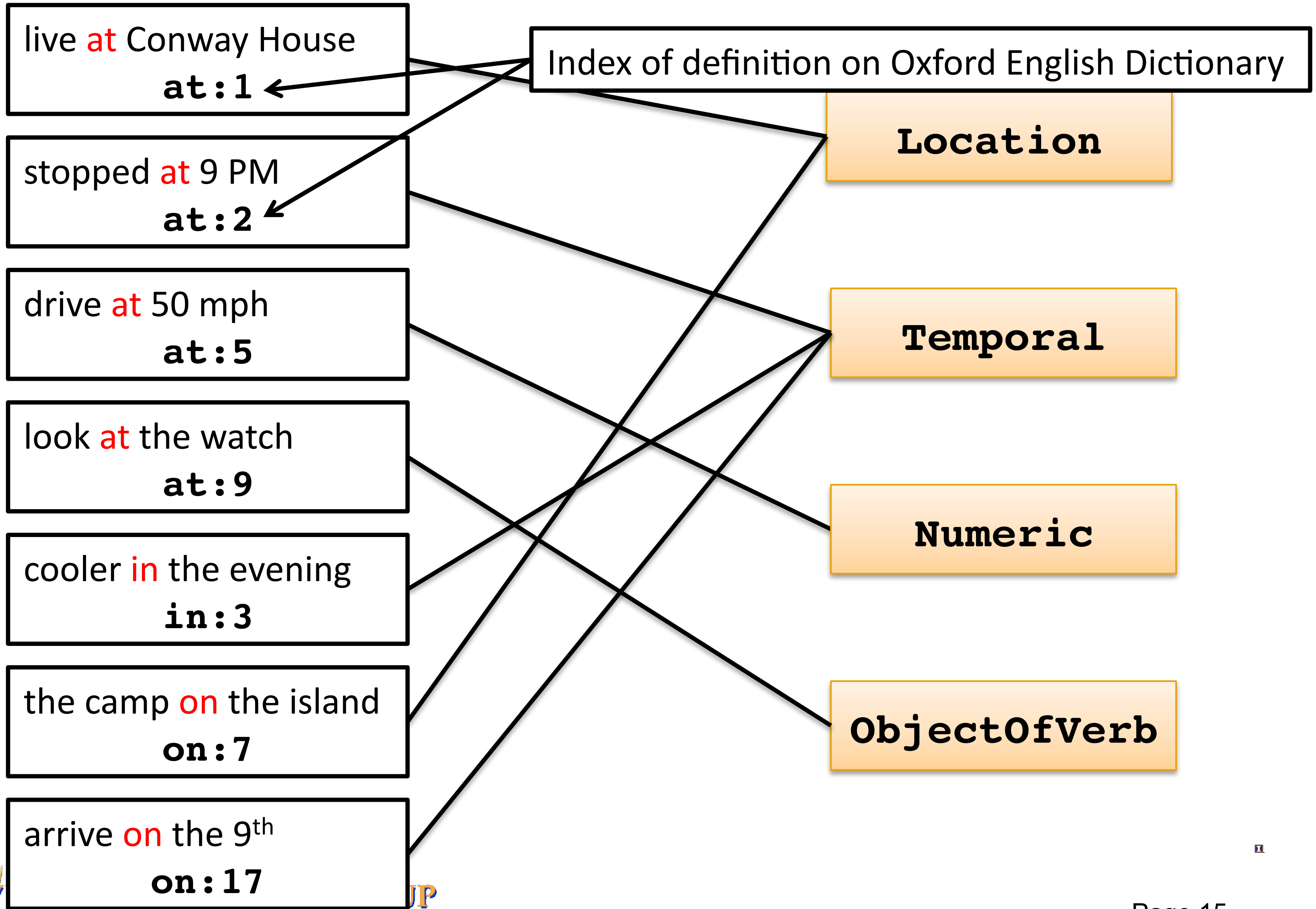
Queen of England



City of Chicago

Predicates expressed by prepositions

Ambiguity & Variability



Preposition relations [Transactions of ACL, '13]

- An inventory of 32 relations expressed by preposition
  - Prepositions are assigned labels that act as predicate in a predicate-argument representation
  - Semantically related senses of prepositions merged
  - Substantial inter-annotator agreement
- A new resource: Word sense disambiguation data, re-labeled
  - SemEval 2007 shared task [Litkowski 2007]
    - **~16K training and 8K test instances; 34 prepositions**
  - Small portion of the Penn Treebank [Dalhmeier, et al 2009]
    - **only 7 prepositions, 22 labels**

Cause

Location

His first patient died of pneumonia. Another, who arrived from NY yesterday suffered from flu. Most others already recovered from flu

cause

Start-state

Computational Questions

1. How do we predict the preposition relations? [EMNLP, ’11]

    - Capturing the interplay with verb SRL?
    - Very small jointly labeled corpus, cannot train a global model!

2. What about the arguments? [Trans. Of ACL, ‘13]

    - Annotation only gives us the predicate
    - How do we train an argument labeler?
    - Exploiting types as latent variables

Coherence of predictions

Predicate arguments from different triggers should be consistent

The bus was heading for Nairobi in Kenya.

Joint constraints
linking the two tasks.

Destination ⇔ A1

Location

Destination

Predicate: head.02
A0 (mover): The bus
A1 (destination): for Nairobi in Kenya

Joint inference (CCMs)

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

Verb arguments

$$\max_{\mathbf{y}} \max_{\mathbf{y}} \sum_t \boxed{\lambda^t} \sum_a y^{a,t} c^{a,t}$$

+ ....

Argument candidates

Re-scaling parameters (one per label)

Preposition relation label

Preposition

Constraints:
Each argument label

Verb SRL constraints

Only one label per preposition

+ Joint constraints between tasks; easy with ILP formulation

Joint Inference – no (or minimal) joint learning

Preposition relations and arguments

1. How do we predict the preposition relations? [EMNLP, '11]
   - **Capturing the interplay with verb SRL?**
   - **Very small jointly labeled corpus, cannot train a global model!**

Enforcing consistency between verb argument labels and preposition relations can help improve both

2. What about the arguments? [Trans. Of ACL, '13]
   - **Annotation only gives us the predicate**
   - **How do we train an argument labeler?**
   - **Exploiting types as latent variables**

Relations depend on argument types

Types are an abstraction that capture common properties of groups of entities.

- Our primary goal is to model preposition relations and their arguments
- But the relation prediction strongly depends also on the semantic type of the arguments.

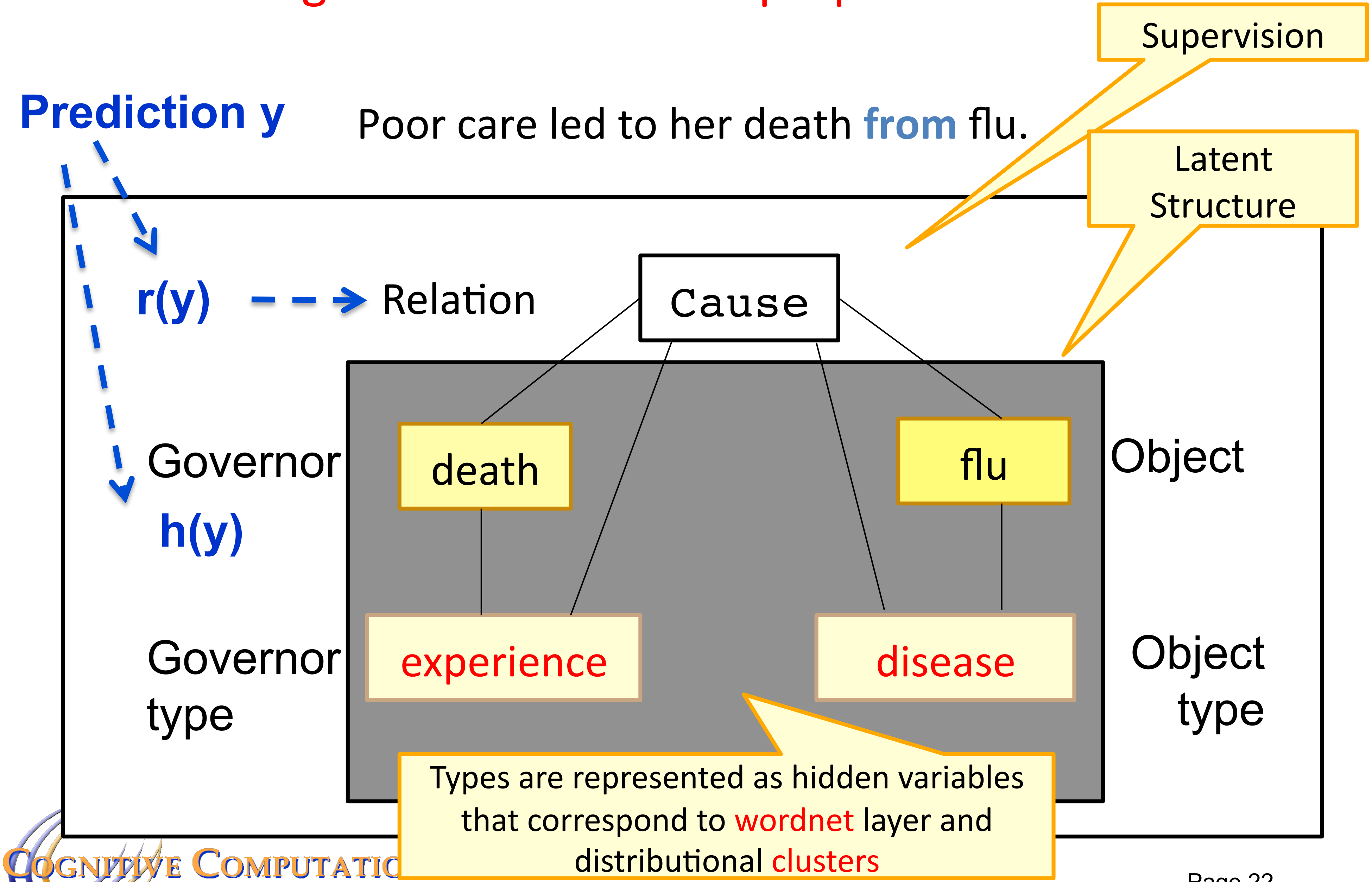
Poor care led to her death **from** pneumonia.

Cause(death, pneumonia)
Cause(death, flu)

Poor care led to her death **from** the flu.

The ability to generalize to unseen words of the same "type" would help argument & relation prediction

Predicate-argument structure of prepositions



COGNITIVE COMPUTATION
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Latent inference

Inference takes into account constrains among parts of the structure y= (r, h), formulated as an ILP

- Standard inference: Find an assignment to the full structure

$$\max_{\mathbf{y}} \mathbf{w}^T \Phi(\mathbf{x}, \mathbf{y})$$

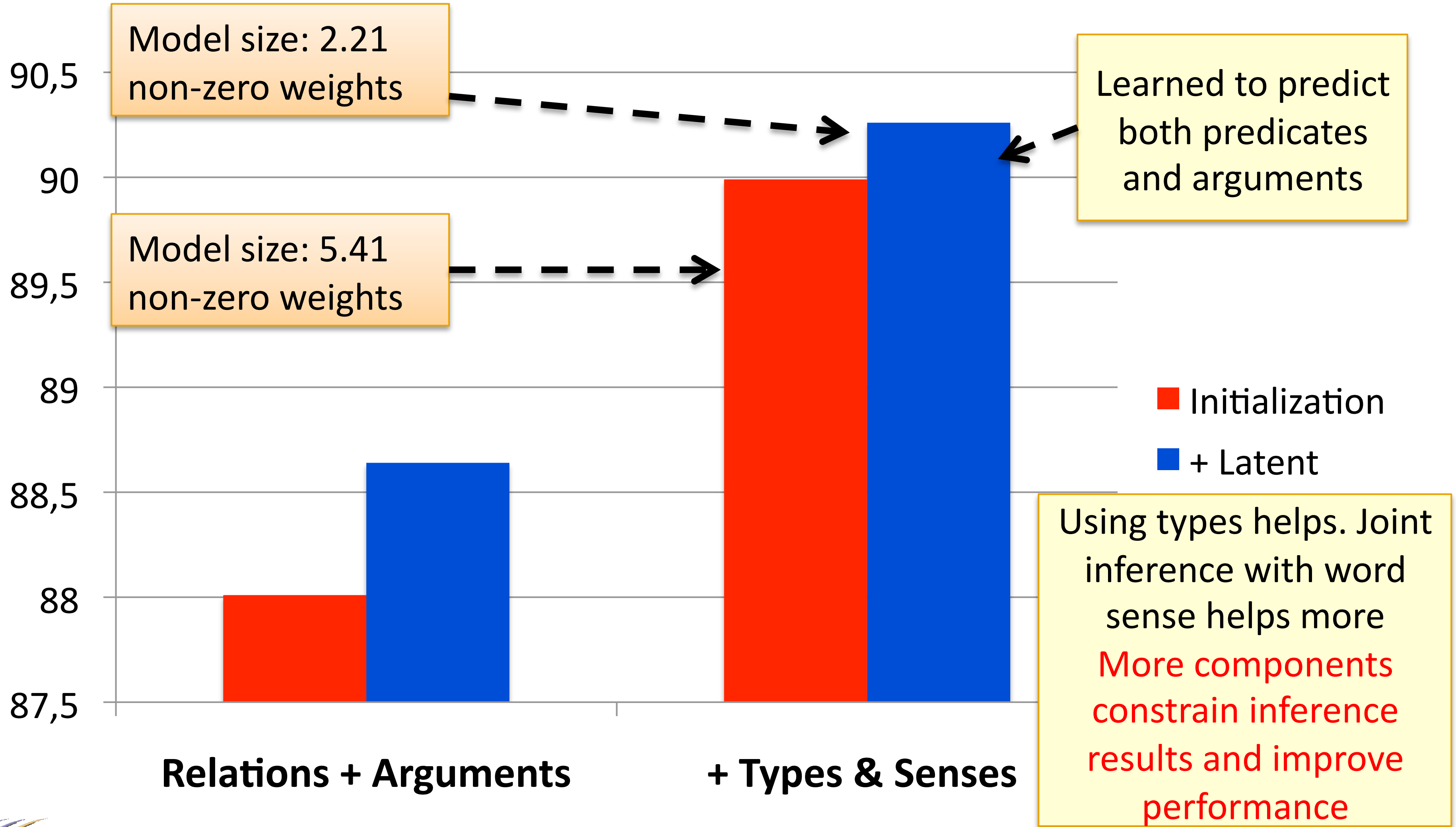
- Latent inference: Given an example annotated with r(y*)

$$\max_{\mathbf{y}} \quad \mathbf{w}^T \Phi(\mathbf{x}, \mathbf{y})$$
$$\text{s.t.} \quad r(\mathbf{y}^*) = r(\mathbf{y})$$

- While satisfying constraints between r(y) and h(y)
- That is: "complete the hidden structure" in the best possible way, to support correct prediction of the supervised variable
  - During training, the loss is defined over the entire structure, where we scale the loss of elements in h(y).

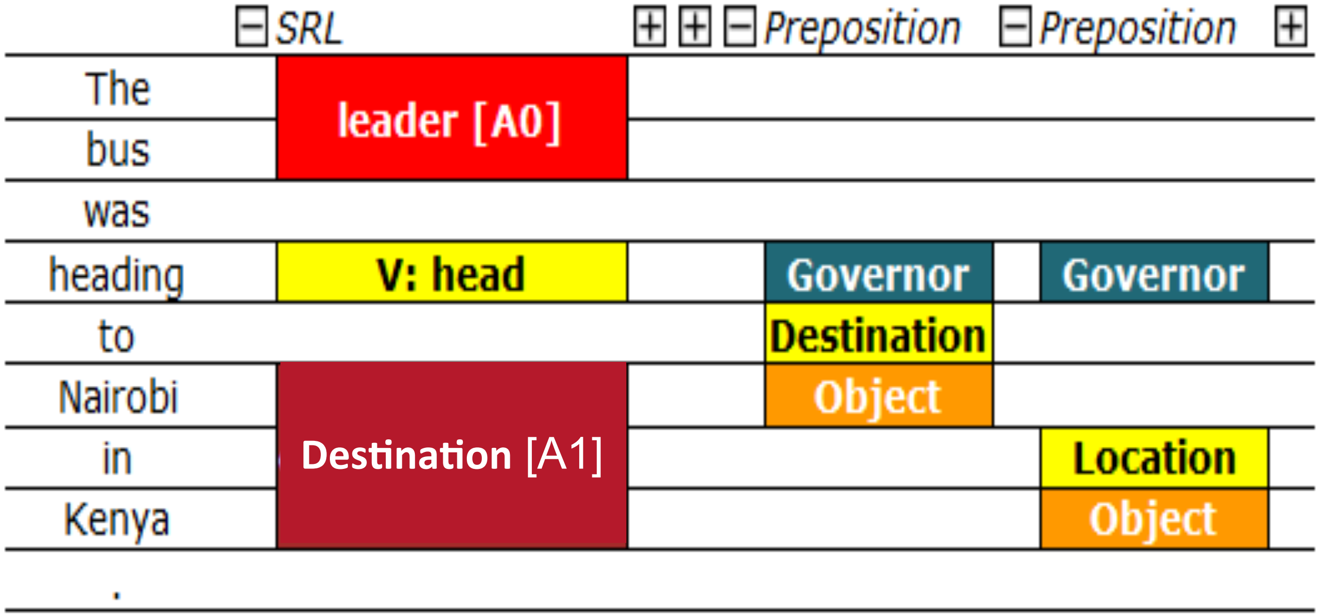
Generalization of Latent Structure SVM [Yu & Joachims '09] &
Indirect Supervision learning [Chang et. al. '10]

Performance on Relation Labeling: The More the Better



Extended SRL [Demo]

Text Annotation: Keep the text – hang multiple semantic annotations on top of it [Roth & Sammons 08]



Joint inference over phenomena specific models to enforce consistency

Models trained with latent structure: senses, types, arguments

- More to do with other relations, discourse phenomena,…

Constrained Conditional Models—ILP Formulations

- **Have been shown useful in the context of many NLP problems**
- [Roth&Yih, 04,07: Entities and Relations; Punyakanok et. al: SRL …]
  - Summarization; Co-reference; Information & Relation Extraction; Event Identifications; Transliteration; Textual Entailment; Knowledge Acquisition; Sentiments; Temporal Reasoning, Dependency Parsing,…
- Some theoretical work on training paradigms [Punyakanok et. al., 05 more; Constraints Driven Learning, PR, Constrained EM…]
- Some work on Inference, mostly approximations, bringing back ideas on Lagrangian relaxation, etc.
- Good summary and description of training paradigms: [Chang, Ratinov & Roth, Machine Learning Journal 2012]
- **Summary of work & a bibliography: http://L2R.cs.uiuc.edu/tutorials.html**

Outline

- Integer Linear Programming Formulations for Natural Language Processing
- Example 1: Extended Semantic Role Labeling
  - Relaxing the pipeline
  - Dealing with lack of joint annotation: combining structured models
- Example 2: Wikification
  - Knowledge Acquisition by Grounding
  - Relational Inference for Wikification
  - Applications

Wikification

Blumenthal (D) is a candidate for the U.S. Senate seat now held by Christopher Dodd (D), and he has held a commanding lead in the race since he entered it. But the Times report has the potential to fundamentally reshape the contest in the Nutmeg State.

Richard Blumenthal
From Wikipedia, the free encyclopedia

Democratic Party (United States)
From Wikipedia, the free encyclopedia

United States Senate
From Wikipedia, the free encyclopedia

[Blumenthal] ([D]) is a candidate for the [U.S. Senate] seat now held by [Christopher Dodd] (D), and he has held a commanding lead in the race since he entered it. But the [Times] report has the potential to fundamentally reshape the contest in [the Nutmeg State].

Chris Dodd
From Wikipedia, the free encyclopedia

The New York Times
From Wikipedia, the free encyclopedia

Connecticut
From Wikipedia, the free encyclopedia

Applications

- Knowledge Acquisition via Grounding
- Coreference Resolution
  - Learning-based multi-sieve co-reference resolution with knowledge (Ratinov et al. 2012)
- Information Extraction
  - Unsupervised relation discovery with sense disambiguation (Yao et al. 2012)
  - Automatic Event Extraction with Structured Preference Modeling (Lu and Roth, 2012)
- Text Classification
  - Gabrilovich and Markovitch, 2007; Chang et al., 2008
- Entity Linking

Challenges

> Blumenthal (D) is a candidate for the U.S. Senate seat now held by Christopher Dodd (D), and he has held a commanding lead in the race since he entered it. But the Times report has the potential to fundamentally reshape the contest in the Nutmeg State.

- Ambiguity

Times → The New York Times
Times → The Times
⋮

- Variability

CT → Connecticut
The Nutmeg State → Connecticut
⋮

- Concepts outside of Wikipedia (NIL)
  - Blumenthal ?
- Scale
  - Millions of labels

Cognitive Computation Group
University of Illinois at Urbana-Champaign

Challenges

Blumenthal (D) is a candidate for the U.S. Senate seat now held by Christopher Dodd (D), and he has held a commanding lead in the race since he entered it. But the Times report has the potential to fundamentally reshape the contest in the Nutmeg State.

- State-of-the-art systems (Ratinov et al. 2011) can achieve the above with local and global statistical features
  - Reaches bottleneck around 70%~ 85% F1 on non-wiki datasets
  - Check out our demo at: http://cogcomp.cs.illinois.edu/demos
  - What is missing?

Relational Inference

- Mubarak, the wife of deposed Egyptian President Hosni Mubarak,…

Relational Inference

Mubarak, **the wife of deposed** Egyptian President Hosni Mubarak, ,…..

- What are we missing with Bag of Words (BOW) models?
  - Who is Mubarak?
- Textual relations provide another dimension of text understanding
- Can be used to constrain interaction between concepts
  - (Mubarak, wife, Hosni Mubarak)
- Has impact in several steps in the Wikification process:
  - From candidate selection to ranking and global decision

Relational Inference for Wikification

- Mubarak, the wife of deposed Egyptian President Hosni Mubarak, …

- Next we will briefly show:
  - How to identify key textual relations for Wikification
  - How to verify relations using external resource
  - A global inference framework to incorporate relational knowledge

- Relational inference yields significant improvements over state-of-the-art systems

Wikification

Mention Detection	Candidate Generation	Candidate Ranking	Determine NILs

Wikification Pipeline 1 - Mention Detection

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party...

- sub-NP (Noun Phrase) chunks [Illinois Chunker]
- NER [Illinois NER]
- Regular expressions

Wikification Pipeline 1 - Mention Segmentation

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party...

Wikification Pipeline 2 - Candidate Generation

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party...

k	e^k_3
1	Slobodan_Milošević
2	Milošević_(surname)
3	Boki_Milošević
4	Alexander_Milošević
…	

k	e^k_4
1	Socialist_Party_(France)
2	Socialist_Party_(Portugal)
3	Socialist_Party_of_America
4	Socialist_Party_(Argentina)
…	

Wikification Pipeline 3 - Candidate Ranking

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party...

k	e^k_3	s^k_3
1	Slobodan_Milošević	0.7
2	Milošević_(surname)	0.1
3	Boki_Milošević	0.1
4	Alexander_Milošević	0.05
…		

k	e^k_4	s^k_4
1	Socialist_Party_(France)	0.23
2	Socialist_Party_(Portugal)	0.16
3	Socialist_Party_of_America	0.07
4	Socialist_Party_(Argentina)	0.06
…		

- Local and global statistical features

Wikification Pipeline 4 – Determine NILs

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party...

k	e^k_3	s^k_3
1	Slobodan_Milošević	0.7

k	e^k_4	s^k_4
1	Socialist_Party_(France)	0.23

- Is the top candidate really what the text referred to?
  - If NO, no title is assigned to this mention.

Formulation

- Goal: Promote concepts that are coherent with textual relations
- Formulate as an Integer Linear Program (ILP):

weight to output e_i^k

Whether to output kth candidate of the ith mention

$$\Gamma_D = \arg\max_{\Gamma} \sum_i \sum_k s_i^k e_i^k + \sum_{i,j} \sum_{k,l} w_{ij}^{(k,l)} r_{ij}^{(k,l)}$$

weight of a relation $r_{ij}^{(k,l)}$

Whether a relation exists between e_i^k and e_j^l

$s.t.$	$r_{ij}^{(k,l)} \in \{0,1\}$	Integral constraints
	$e_i^k \in \{0,1\}$	Integral constraints
	$\forall i \sum_k e_i^k = 1$	Unique solution
	$2r_{ij}^{(k,l)} \leq e_i^k + e_j^l$	Relation definition

- If no relation exists, collapses to the non-structured decision

Relation Inference Formulation

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party...

k	e^k_3	s^k_3
1	Slobodan_Milošević	0.7
2	Milošević_(surname)	0.1
3	Boki_Milošević	0.1
4	Alexander_Milošević	0.05
…		

$r^{(1,2)}_{34}$

$r^{(4,3)}_{34}$

k	e^k_4	s^k_4
1	Socialist_Party_(France)	0.23
2	Socialist_Party_(Portugal)	0.16
3	Socialist_Party_of_America	0.07
4	Socialist_Party_(Argentina)	0.06
…		

$$\Gamma_D = \arg\max_{\Gamma} \sum_i \sum_k s_i^k e_i^k + \sum_{i,j} \sum_{k,l} w_{ij}^{(k,l)} r_{ij}^{(k,l)}$$

- e^k_i: whether a concept is chosen
- s^k_i : score of a concept
- $r^{(k,l)}_{ij}$: whether a relation is present
- $w^{(k,l)}_{ij}$: score of a relation

Overall Approach

Wikification

- Candidate Generation
- Candidate Ranking
- Determine NILs

Relation Analysis

- Relation Identification
- Relation Retrieval
- Relational Inference

1. Relation Identification

- ACE style in-document coreference [Chang et. al, EMNLP'13]
  - Extract named entity-only coreference relations with high precision
- Syntactico-Semantic relations [Chan & Roth '10]

Type	Example
Premodifier	Iranian Ministry of Defense
Possessive	NYC's stock exchange
Formulaic	Chicago, Illinois
Preposition	President of the US

  - Easy to extract with high precision
  - Aim for high recall, as false-positives will be verified and discarded
  - These relations covers **~80%** relation instances in ACE2004

1. Relation Identification

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party...

Argument 1	Relation Type	Argument 2
Yugoslav President	apposition	Slobodan Milošević
Slobodan Milošević	coreference	Milošević
Milošević	possessive	Socialist Party

2. Relation Retrieval for Candidate Generation

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party...

- Earlier approach
  - Collect known mappings from Wikipedia page titles, hyperlinks...
  - Limit to top-K candidates based on frequency of links (Ratinov et al. 2011)
- What concepts can "Socialist Party" refer to?

A Lot of Uninformative Mentions

Socialist Party (disambiguation)

From Wikipedia, the free encyclopedia

Socialist Party is the name of many different political parties aro
article.

Socialist Party may also refer to the wide variety of political parti
What follows is an incomplete alphabetical list of such parties:

Names used by several different parties [edi

- Arab Socialist Ba'ath Party (disambiguation)
- Authentic Socialist Party (disambiguation)
- Democratic Socialist Party (disambiguation)
- Independent Socialist Party (disambiguation)
- National Socialist Party (disambiguation)
- New Socialist Party (disambiguation)
- Polish Socialist Party (disambiguation)
- Popular Socialist Party (disambiguation)
- Revolutionary Socialist Party (disambiguation)
- Socialist Action Party (disambiguation)
- Socialist Democratic Party (disambiguation)
- Socialist Equality Party (disambiguation)
- Socialist Labor Party (disambiguation)
- Socialist Labour Party (disambiguation)
- Socialist Left Party (disambiguation)
- Socialist People's Party (disambiguation)
- Socialist Republican Party (disambiguation)
- Socialist Unity (disambiguation)

• • •

2. Relation Retrieval for Candidate Generation

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party...

- What concepts can "Socialist Party" refer to?
- More robust candidate generation
  - Identified relations are verified against a knowledge base (DBPedia)
  - Retrieve relation arguments matching "(Milošević ,?,Socialist Party)" as our new candidates

2. Relation Retrieval for Candidate Generation

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party...

k	e^k_3	s^k_3
1	Slobodan_Milošević	0.7
…		

k	e^k_4	s^k_4
1	Socialist_Party_(France)	0.23
…		

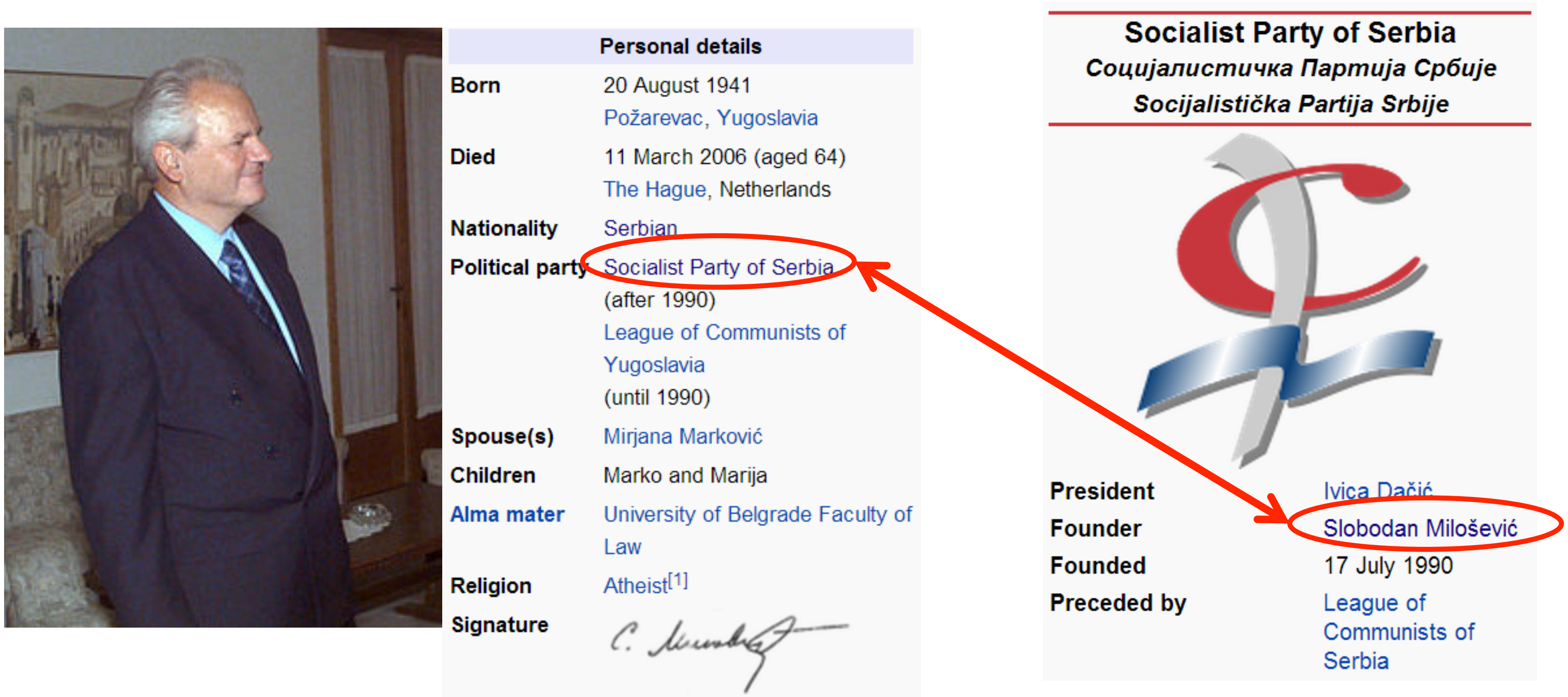
- Query Pruning
  - Only 2 queries per pair necessary due to strong baseline.

q_1=(Socialist Party of France,?, *Milošević*)

q_2=(Slobodan Milošević,?,*Socialist Party*)

2. Relation Retrieval for Candidate Generation

Argument 1	Relation Type	Argument 2
Milošević	possessive	Socialist Party



2. Relation Retrieval for Candidate Generation

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party...

k	e^k_3	s^k_3
1	Slobodan_Milošević	0.7
2	Milošević_(surname)	0.1
3	Boki_Milošević	0.1
4	Alexander_Milošević	0.05
…		

$$r_{34}^{(1,21)} = 1$$

k	e^k_4	s^k_4
1	Socialist_Party_(France)	0.23
2	Socialist_Party_(Portugal)	0.16
3	Socialist_Party_of_America	0.07
4	Socialist_Party_(Argentina)	0.06
…		
21	Socialist_Party_of_Serbia	0.0

3. Relational Inference For Candidate Ranking

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party...

k	e^k_3	s^k_3
1	Slobodan_Milošević	0.7
2	Milošević_(surname)	0.1
3	Boki_Milošević	0.1
4	Alexander_Milošević	0.05
…		

k	e^k_4	s^k_4
1	Socialist_Party_(France)	0.23
2	Socialist_Party_(Portugal)	0.16
3	Socialist_Party_of_America	0.07
4	Socialist_Party_(Argentina)	0.06
…		
21	Socialist_Party_of_Serbia	0.0

$r_{34}^{(1,21)} = 1$

$$\Gamma_D = \arg\max_{\Gamma} \sum_i \sum_k s_i^k e_i^k + \sum_{i,j} \sum_{k,l} w_{ij}^{(k,l)} r_{ij}^{(k,l)}$$

$$w_{34}^{(1,21)} = ?$$

COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

3. Relational Inference For Candidate Ranking - Coreference

...ousted long time Yugoslav President Slobodan Milošević in October. Mr. Milošević's Socialist Party...

k	e^k_3	s^k_3
1	Slobodan_Milošević	0.7
2	Milošević_(surname)	0.1
3	Boki_Milošević	0.1
4	Alexander_Milošević	0.05
…		

$r_{23}^{(1,1)} = 1$

k	e^k_2	s^k_2
1	Slobodan_Milošević	1.0

- Ranking is propagated via other relations to other candidates

4. Relation Inference for Determining Unknown Concepts

Dorothy Byrne, a state coordinator for the Florida Green Party,…

nominal mention

k	e^k_1	s^k_1
1	Dorothy_Byrne_(British _Journalist)	0.6
2	Dorothy_Byrne_(mezzo -soprano)	0.4

k	e^k_2	s^k_2
1	Green_Party_of_Florida	1.0

- How to capture the fact that:
  - "Dorothy Byrne" does not refer to any concept in Wikipedia
- Identify coreferent nominal mention relations
  - Generate better features for NIL classifier

4. Relation Inference for Determining Unknown Concepts

Dorothy Byrne, a state coordinator for the Florida Green Party,…

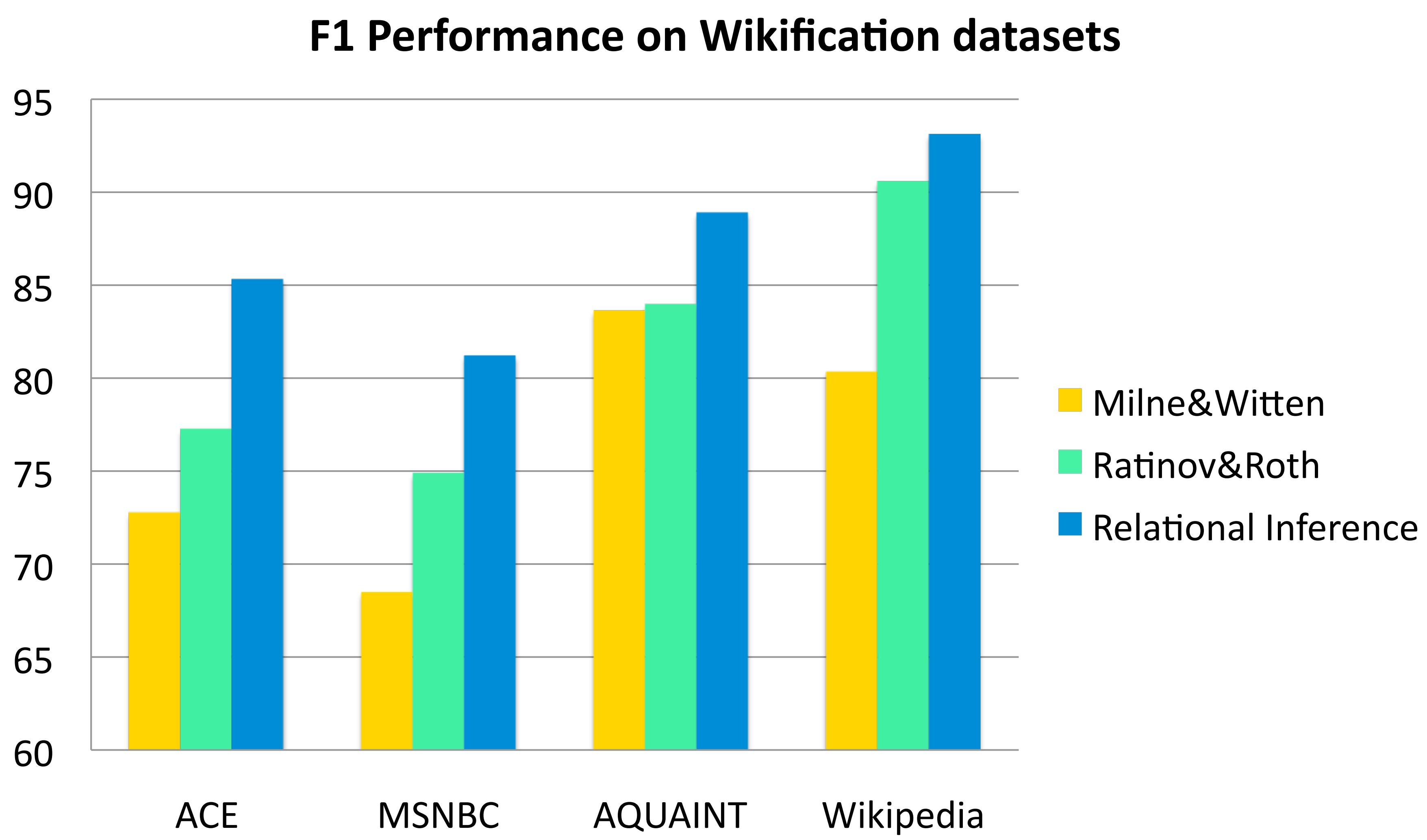
nominal mention

k	e^k_1	s^k_1
0	NIL	1.0
1	Dorothy_Byrne_(British _Journalist)	0.6
2	Dorothy_Byrne_(mezzo -soprano)	0.4

k	e^k_2	s^k_2
1	Green_Party_of_Florida	1.0

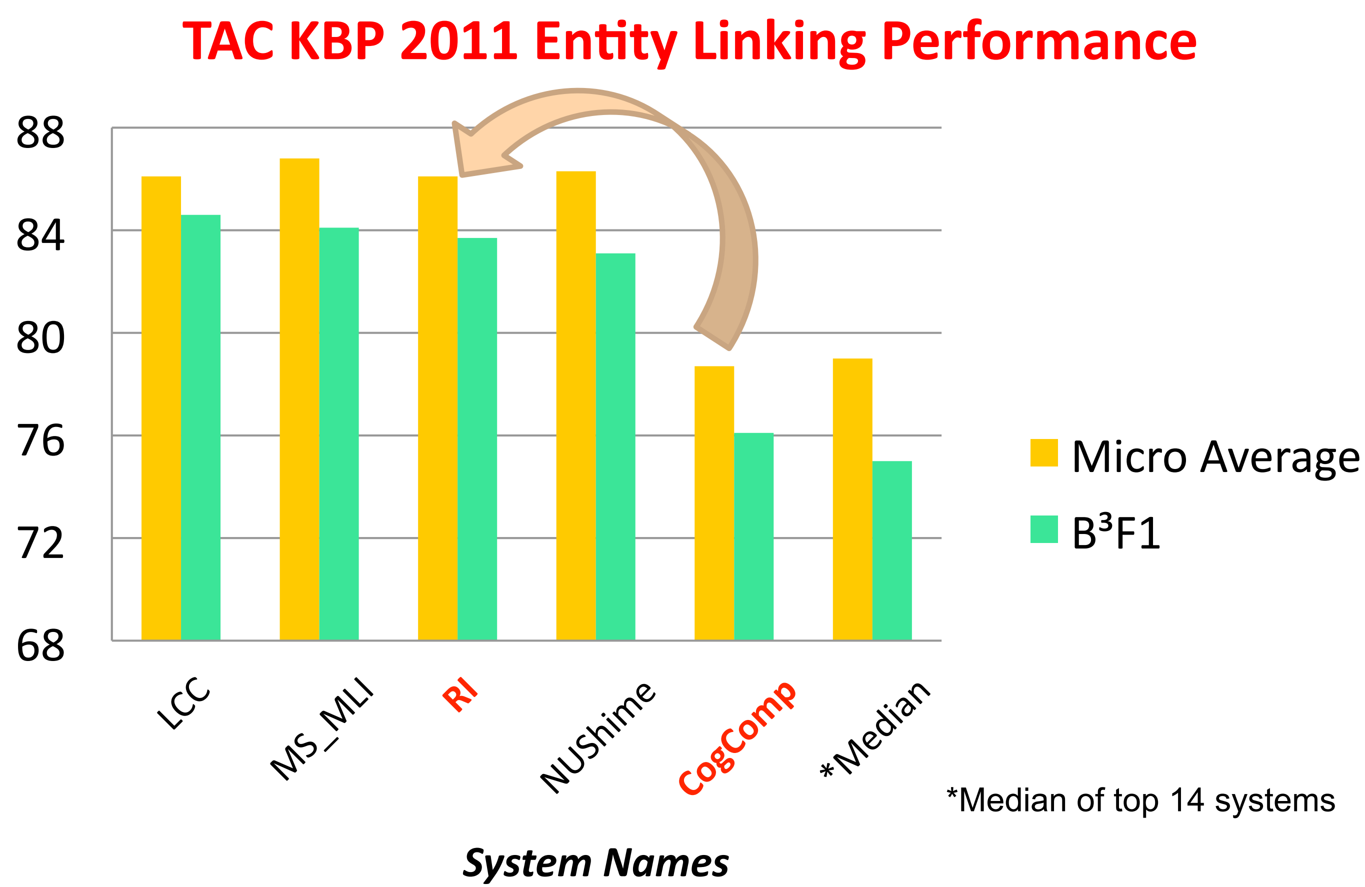
- Create NIL candidate for propagation

Wikification Performance Result



Evaluation – TAC KBP Entity Linking

- Run Relational Inference (**RI**) Wikifier "as-is":
  - No retraining using TAC data



Thank You!

Conclusion

- Presented Constrained Conditional Models
  - A powerful & modular learning and inference paradigm for high level tasks.
- An ILP based computational framework that provides an interface to augment statistically learned linear models with declarative constraints
  - Incorporating knowledge and support decisions in expressive output spaces
  - Flexibility in Training & Inference [E.g., Amortized Inference, ACL'13, EMNLP'12]
- Exemplified the use of CCM in the context of layers of semantic annotations
  - Extended Semantic Role Labeling of Sentences
  - Wikification

Check out our tools, demos, LBJ and CCM tutorial