COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Constrained Conditional Models

Learning and Inference in Natural Language Understanding

Dan Roth

Department of Computer Science

University of Illinois at Urbana-Champaign

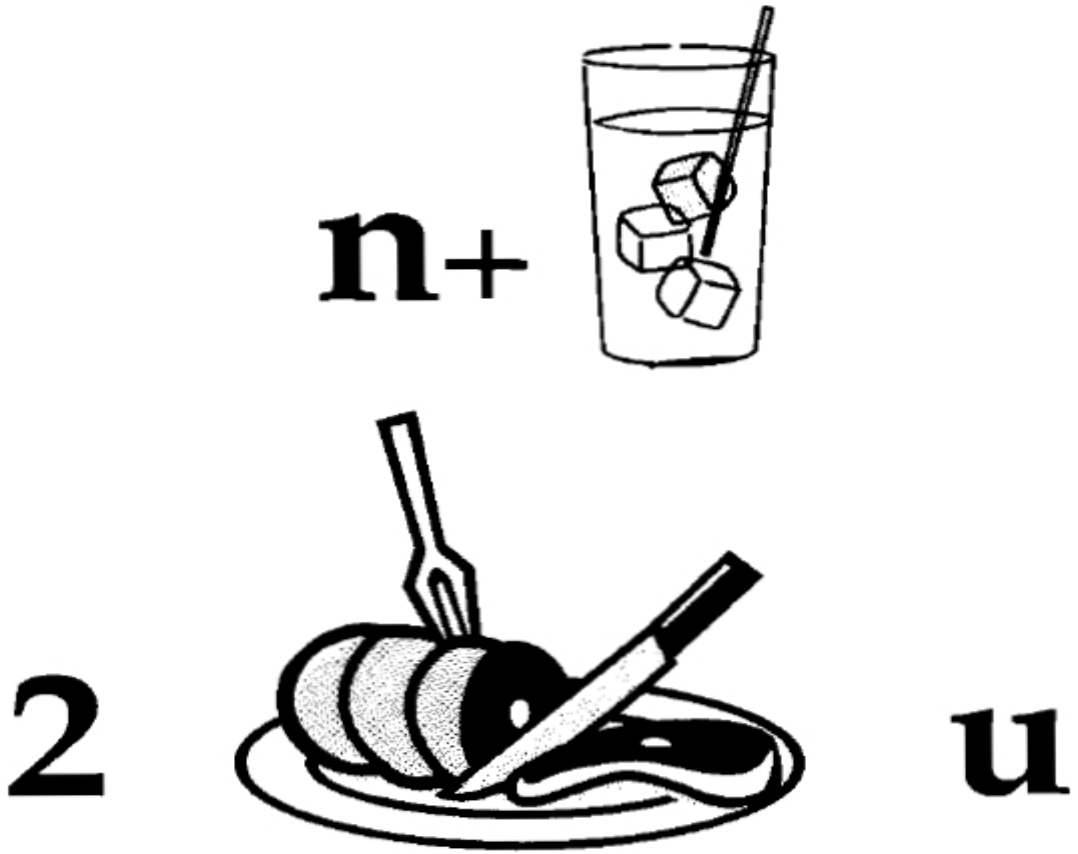
MIAS
MULTIMODAL INFORMATION ACCESS AND SYNTHESIS

With thanks to:

Collaborators: **Ming-Wei Chang, Vasin Punyakanok, Lev Ratinov, Nick Rizzolo, Mark Sammons, Scott Yih, Dav Zimak**

Funding: ARDA, under the AQUAINT program
NSF: ITR IIS-0085836, ITR IIS-0428472, ITR IIS- 0085980, SoD-HCER-0613885
A DOI grant under the Reflex program; DHS
DASH Optimization (Xpress-MP)

Nice to Meet You



Learning and Inference

- Global decisions in which several local decisions play a role but there are mutual dependencies on their outcome.
  - E.g. Structured Output Problems – multiple dependent output variables

- (Learned) models/classifiers for different sub-problems
  - In some cases, not all models are available to be learned simultaneously
  - Key examples in NLP are Textual Entailment and QA
  - In these cases, constraints may appear only at evaluation time

- Incorporate models' information, along with prior knowledge/constraints, in making coherent decisions
  - decisions that respect the learned models as well as domain & context specific knowledge/constraints.

Inference

Comprehension

A process that maintains and updates a collection of propositions about the state of affairs.

(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's dad was a magician.
4. Christopher Robin must be at least 65 now.

This is an Inference Problem

This Talk: Constrained Conditional Models

- A general inference framework that combines
  - Learning conditional models with using declarative expressive constraints
  - Within a constrained optimization framework
    - Formulate a decision process as a constrained optimization problem, or
    - Break up a complex problem into a set of sub-problems and require components' outcomes to be consistent modulo constraints

- Has been shown useful in the context of many NLP problems
  - SRL, Summarization; Co-reference; Information Extraction
  - [Roth&Yih04,07; Punyakanok et.al 05,08; Chang et.al07,08; Clarke&Lapata06,07; Denise&Baldrige07]

- Here: focus on Learning and Inference for Structured NLP Problems

Outline

- Constrained Conditional Models
  - Motivation
  - Examples
- Training Paradigms: Investigate ways for training models and combining constraints
  - Joint Learning and Inference vs. decoupling Learning & Inference
  - Guiding Semi-Supervised Learning with Constraints
    - Features vs. Constraints
    - Hard and Soft Constraints
- Examples
  - Semantic Parsing
  - Information Extraction
  - Pipeline processes

Inference with General Constraint Structure [Rot

Improvement over no inference: 2-5%

other	0.05
per	0.85
loc	0.10

other	0.10
per	0.60
loc	0.30

othe	0.05
[illegible]	0.50
loc	0.45

Dole 's wife, Elizabeth , is a native

E1 $\xrightarrow{R_{12}}$ E2 $\xrightarrow{R_{23}}$

Some Questions:
How to guide the global inference?
Why not learn Jointly?

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.

Task of Interests: Structured Output

- For each instance, assign values to a set of variables
- Output variables depends on each other

- Common tasks in
  - Natural language processing
    - Parsing; Semantic Parsing; Summarization; Transliteration; Co-reference resolution,…
  - Information extraction
    - Entities, Relations,…

- Many pure machine learning approaches exist
  - Hidden Markov Models (HMMs); CRFs
  - Perceptrons…

- However, …

Information Extraction via Hidden Markov Models

Lars Ole Andersen . Program analysis and specialization for the C Programming language. PhD thesis. DIKU , University of Copenhagen, May 1994 .

Prediction result of a trained HMM

[AUTHOR]	Lars Ole Andersen . Program analysis and
[TITLE]	specialization for the
[EDITOR]	C
[BOOKTITLE]	Programming language
[TECH-REPORT]	. PhD thesis .
[INSTITUTION]	DIKU , University of Copenhagen , May
[DATE]	1994 .

Unsatisfactory results !

Strategies for Improving the Results

- (Pure) Machine Learning Approaches
  - Higher Order HMM/CRF?
  - Increasing the window size?
  - Adding a lot of new features
    - Requires a lot of labeled examples
  - What if we only have a few labeled examples?

Increasing the model complexity

Can we keep the learned model simple and still make expressive decisions?

- Any other options?
  - Humans can <u>immediately</u> tell bad outputs
  - The output does not make sense

Information extraction without Prior Knowledge

Lars Ole Andersen . Program analysis and specialization for the C Programming language. PhD thesis. DIKU , University of Copenhagen, May 1994 .

Prediction result of a trained HMM

[AUTHOR]	Lars Ole Andersen . Program analysis and
[TITLE]	specialization for the
[EDITOR]	C
[BOOKTITLE]	Programming language
[TECH-REPORT]	. PhD thesis .
[INSTITUTION]	DIKU , University of Copenhagen , May
[DATE]	1994 .

Violates lots of natural constraints!

Examples of Constraints

- Each field must be a consecutive list of words and can appear at most once in a citation.
- State transitions must occur on punctuation marks.
- The citation can only start with *AUTHOR* or *EDITOR*.
- The words *pp., pages* correspond to *PAGE*.
- Four digits starting with 20xx and 19xx are *DATE*.
- Quotations can appear only in *TITLE*
- …….

Easy to express pieces of "knowledge"

Non Propositional; May use Quantifiers

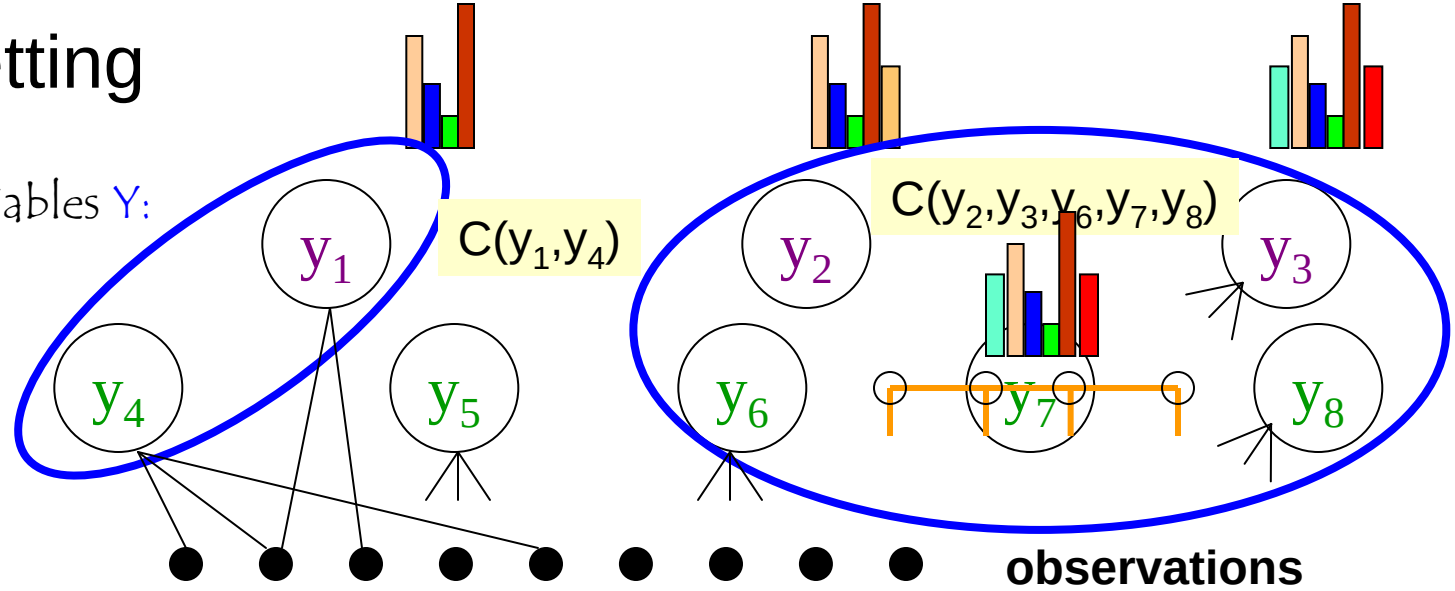
Information Extraction with Constraints

- Adding constraints, we get correct results!
  - Without changing the model

- [AUTHOR] Lars Ole Andersen .
  [TITLE] Program analysis and specialization for the C Programming language .
  [TECH-REPORT] PhD thesis .
  [INSTITUTION] DIKU , University of Copenhagen ,
  [DATE] May, 1994 .

Problem Setting

- Random Variables Y:



- Conditional Distributions P (learned by models/classifiers)
- Constraints C– any Boolean function
  defined on partial assignments (possibly: + weights W)

- Goal: Find the "best" assignment
  - The assignment that achieves the highest global performance.
- This is an Integer Programming Problem

$$Y^* = \text{argmax}_Y \; P \bullet Y \;\; (+ \; W \bullet C) \quad \text{subject to constraints } C$$

Formal Model

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y)$$

Penalty for violating the constraint

Subject to constraints

oft) constraints
mponent

Weight Vector for "local" models

A collection of Classifiers; Log-linear models (HMM, CRF) or a combination

How far away is y from a "legal" assignment

How to solve?

This is an Integer Linear Program

Solving using ILP packages gives an exact solution.

Search techniques are also possible

How to train?

How to decompose global objective function?

Should we incorporate constraints in the learning process?

Example: Semantic Role Labeling

Who did what to whom, when, where, why,…

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

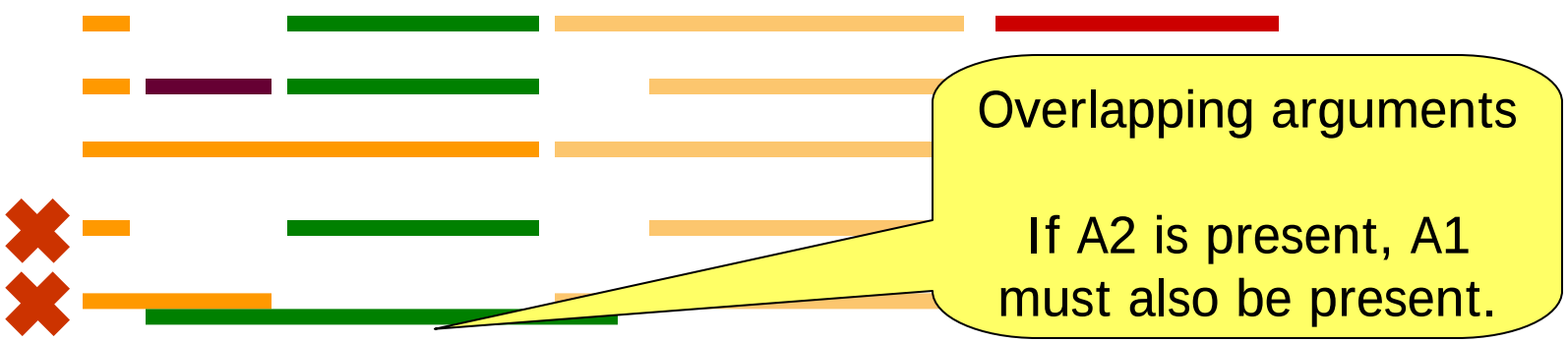
[I]$_{A0}$ *left* [my pearls]$_{A1}$ [to my daughter]$_{A2}$ [in my will]$_{AM\text{-}LOC}$.

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

Special Case (structured output problem): here, all the data is available at one time; in general, classifiers might be learned from different sources, at different times, at different contexts.

Implications on training paradigms

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



Semantic Role Labeling (2/2)

- PropBank [Palmer et. al. 05] provides a large human-annotated corpus of semantic verb-argument relations.
  - It adds a layer of generic semantic labels to Penn Tree Bank II.
  - (Almost) all the labels are on the constituents of the parse trees.

- Core arguments: A0-A5 and AA
  - different semantics for each verb
  - specified in the PropBank Frame files

- 13 types of adjuncts labeled as AM-*arg*
  - where ***arg*** specifies the adjunct type

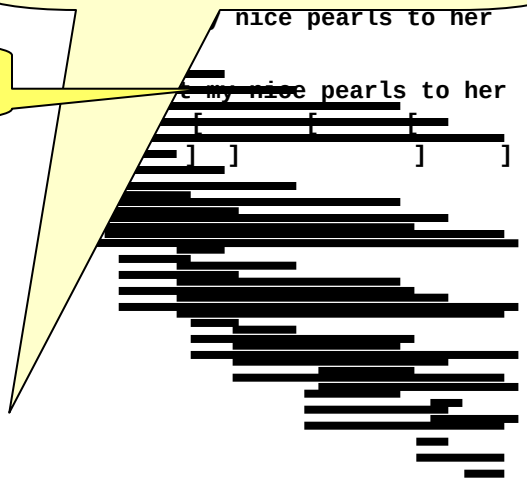
Algorithmic Approach

candidate arguments

- Identify argument candidates
  - Pruning [X &Palmer, EMNLP'04]
  - Argument Identifier
    - Binary classification (SNoW)
- Classify argument candidates
  - Argument Classifier
    - Multi-class classification (SNoW)
- Inference
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output

EASY

Identify Vocabulary



Inference over (old and new) Vocabulary

Inference

- The output of the argument classifier often violates some constraints, especially when the sentence is long.

- Finding the best legitimate output is formalized as an optimization problem and solved via Integer Linear Programming. [Punyakanok et. al 04, Roth & Yih 04;05]

- Input:
  - The probability estimation (by the argument classifier)
  - Structural and linguistic constraints

- Allows incorporating expressive (non-sequential) constraints on the variables (the arguments types).

Integer Linear Programming Inference

- For each argument a_i
  - Set up a Boolean variable: $a_{i,t}$ indicating whether a_i is classified as t

- Goal is to maximize
  - $\sum_i \text{score}(a_i = t)\, a_{i,t}$
  - Subject to the (linear) constraints

The Constrained Conditional Model is completely decomposed during training

- If $\text{score}(a_i = t) = P(a_i = t)$, the objective is to find the assignment that maximizes the expected number of arguments that are correct and satisfies the constraints.

Constraints

Any Boolean rule can be encoded as a linear constraint.

- No duplicate argument classes

  $\Sigma_{a \in POTARG} x_{\{a = A0\}} \leq 1$

- R-ARG

  $\forall\ a2 \in POTARG,\ \Sigma_{a \in POTARG} x_{\{a = A0\}} \geq x_{\{a2 = R\text{-}A0\}}$

If there is an R-ARG phrase, there is an ARG Phrase

- C-ARG

  $\forall\ a2 \in POTARG,$

  $\Sigma_{(a \in POTARG) \wedge (a \text{ is before } a2)} x_{\{a = A0\}} \geq x_{\{a2 = C\text{-}A0\}}$

If there is an C-ARG phrase, there is an A[illegible] before it

Universally quantified

- Many other possible constraints:
  - Unique labels
  - No overlapping or embedding
  - Relations between number of arguments; order constraints
  - If verb is of type A, no argument of type B

LBJ: allows a developer to encode constraints in FOL; these are compiled into linear inequalities automatically.

Joint inference can be used also to combine different SRL Systems.

Semantic Role Labeling

Screen shot from a CCG demo
http://L2R.cs.uiuc.edu/~cogcomp

Semantic Role Labeling Output

Input Text:

A car bomb that exploded outside the U.S. military base in Beniji killed 11 Iraqi citizens.

Result: Complete!

⊞ General Explanation of Argument Labels

A	bomb [A1]	killer [A0]
car		
bomb		
that	bomb (Reference) [R-A1]	
exploded	V: explode	
outside	location [AM-LOC]	
the		
U.S.		
military	temporal [AM-TMP]	
base		
in	location [AM-LOC]	
Beniji		
killed		V: kill
11		corpse [A1]
Iraqi		
citizens		
.		

Semantic parsing reveals several relations in the sentence along with their arguments.

This approach produces a very good semantic parser. F1~90%

Easy and fast: ~7 Sent/Sec (using Xpress-MP)

Top ranked system in CoNLL'05 shared task

Key difference is the Inference

Outline

- Constrained Conditional Models
  - Motivation
  - Examples

- Training Paradigms: Investigate ways for training models and combining constraints
  - Joint Learning and Inference vs. decoupling Learning & Inference
  - Guiding Semi-Supervised Learning with Constraints
    - Features vs. Constraints
    - Hard and Soft Constraints

- Examples
  - Semantic Parsing
  - Information Extraction
  - Pipeline processes

Textual Entailment

Semantic Role Labeling
Punyakanok et. al'05,08

Phrasal verb paraphrasing
[Connor&Roth'07]

Inference for Entailment
Braz et. al'05, 07

Entity matching [Li et. al, AAAI'04, NAACL'04]

Is it true that...?
(Textual Entailment)

Eyeing the huge market potential, currently led by Google, Yahoo took over search company Overture Services Inc. last year

$\Rightarrow$

Yahoo acquired Overture

Overture is a search company

Google is a search company

Google owns Overture

.....

Training Paradigms that Support Global Inference

Incorporating general constraints (Algorithmic Approach)

- Allow both statistical and expressive declarative constraints
- Allow non-sequential constraints (generally difficult)

Coupling vs. Decoupling Training and Inference.

- Incorporating global constraints is important but
- Should it be done only at evaluation time or also at training time?
- How to decompose the objective function and train in parts?
- Issues related to:
  - Modularity, efficiency and performance, availability of training data
  - Problem specific considerations

Training in the presence of Constraints

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

Decompose Model (SRL case)

Decompose Model from constraints

- General Training Paradigm:
  - First Term: Learning from data (could be further decomposed)
  - Second Term: Guiding the model by constraints
  - Can choose if constraints' weights trained, when and how, or taken into account only in evaluation.

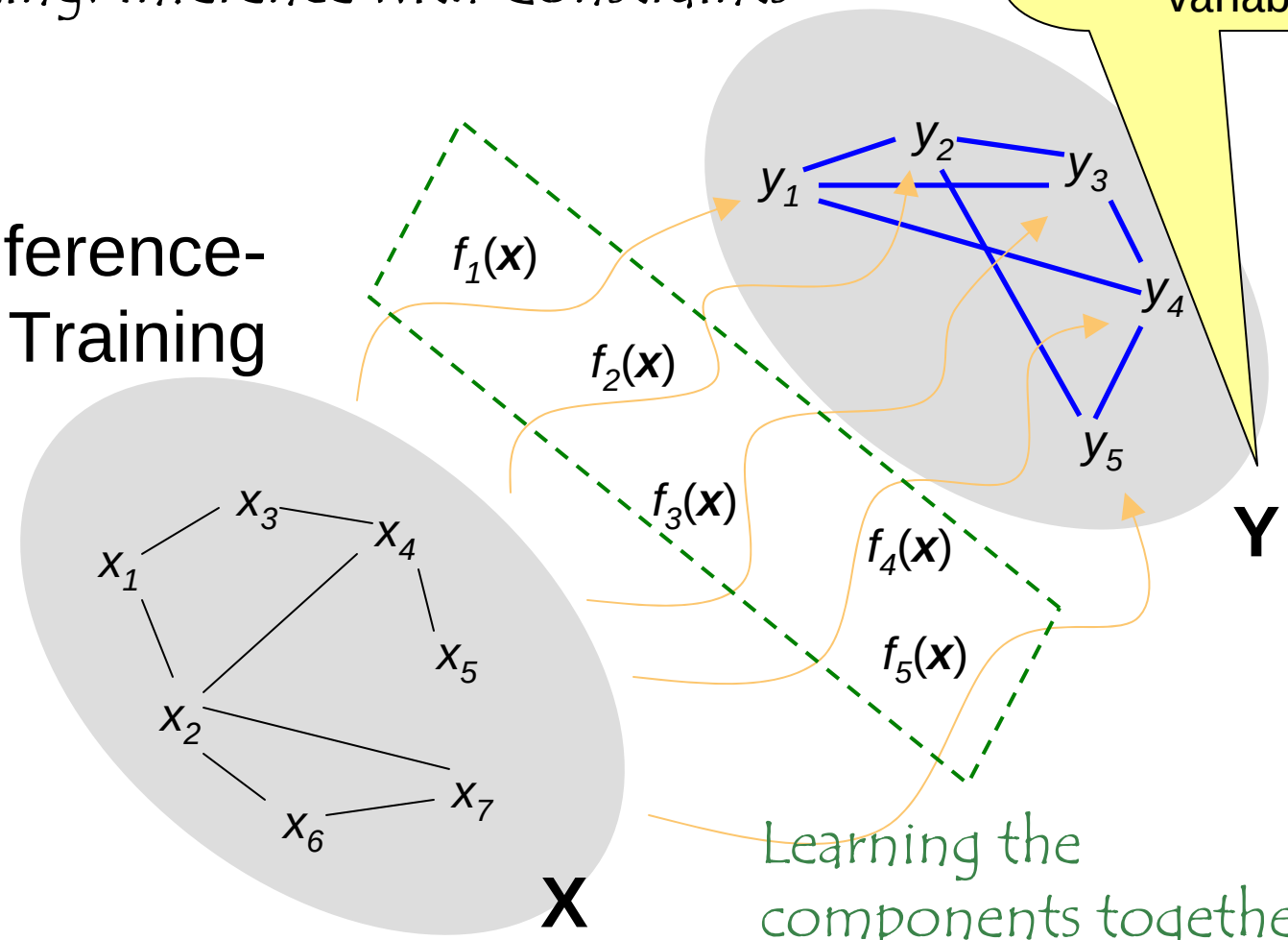
L+I: Learning plus Inference

Training w/o Constraints

Testing: Inference with Constraints

Cartoon: each model can be more complex and may have a view on a set of output variables.

IBT: Inference-based Training



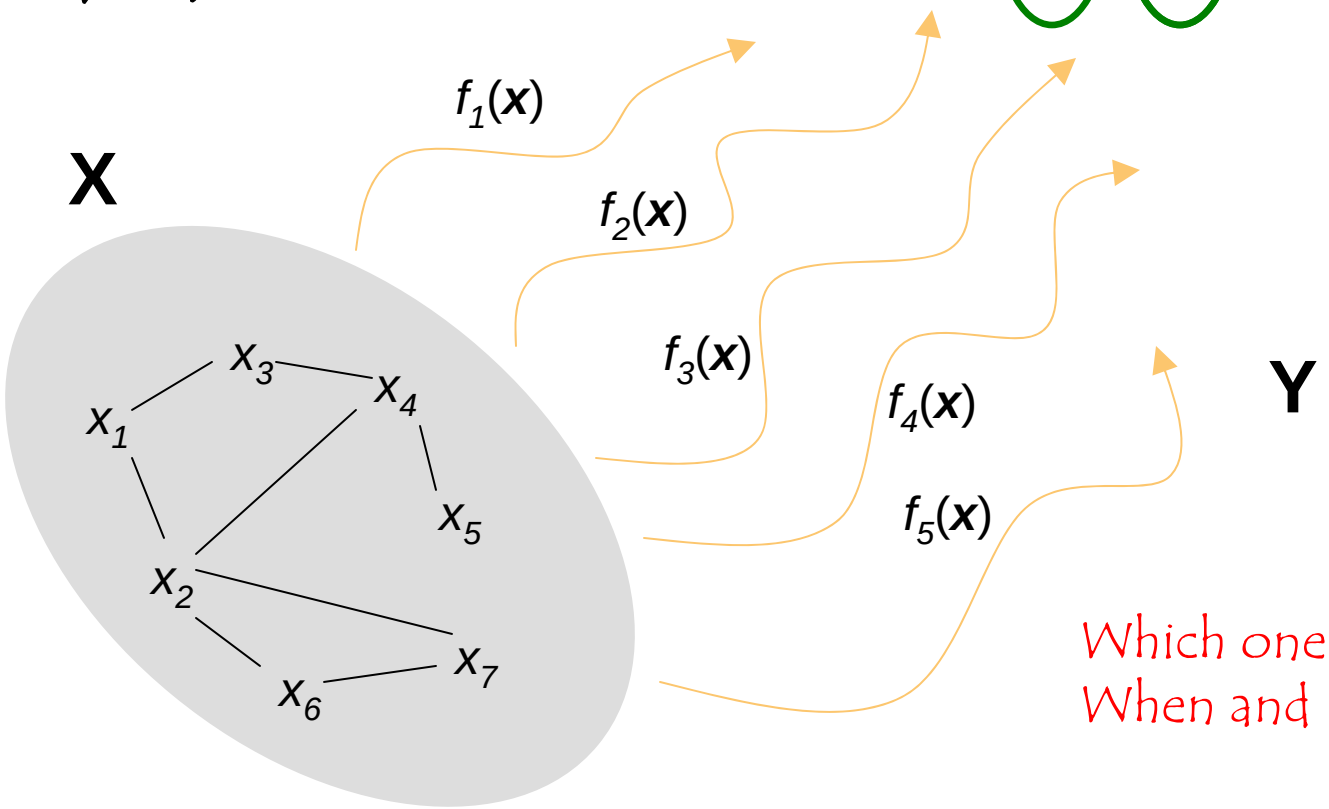
Learning the components together!

Perceptron-based Global Learning

True Global Labeling

Apply Constraints:

Y	-1	1	-1	-1	1
Y'	-1	1	1	1	1



Which one is better?
When and Why?

Claims

- When the local modes are "easy" to learn, L+I outperforms IBT.
  - In many applications, the components are *identifiable* and easy to learn (e.g., argument, open-close, PER).
- Only when the local problems become difficult to solve in isolation, IBT outperforms L+I, but needs a larger number of training examples.
  - When data is scarce, problems are not easy and constraints can be used, along with a "weak" model, to label unlabeled data and improve model.

L+I: cheaper computationally; modular
IBT is better in the limit, and other extreme cases.

- Other training paradigms are possible
- Pipeline-like Sequential Models:
  - Identify a preferred ordering among components
  - Learn k-th model jointly with previously learned models

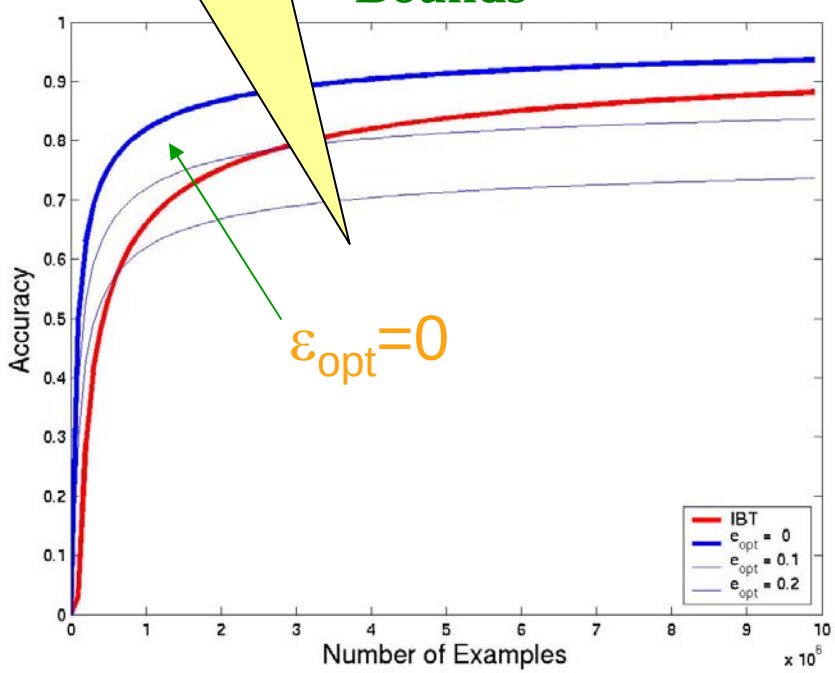
Bound Prediction

L+I vs. IBT: the more identifiable individual problems are, the better overall performance is with L+I

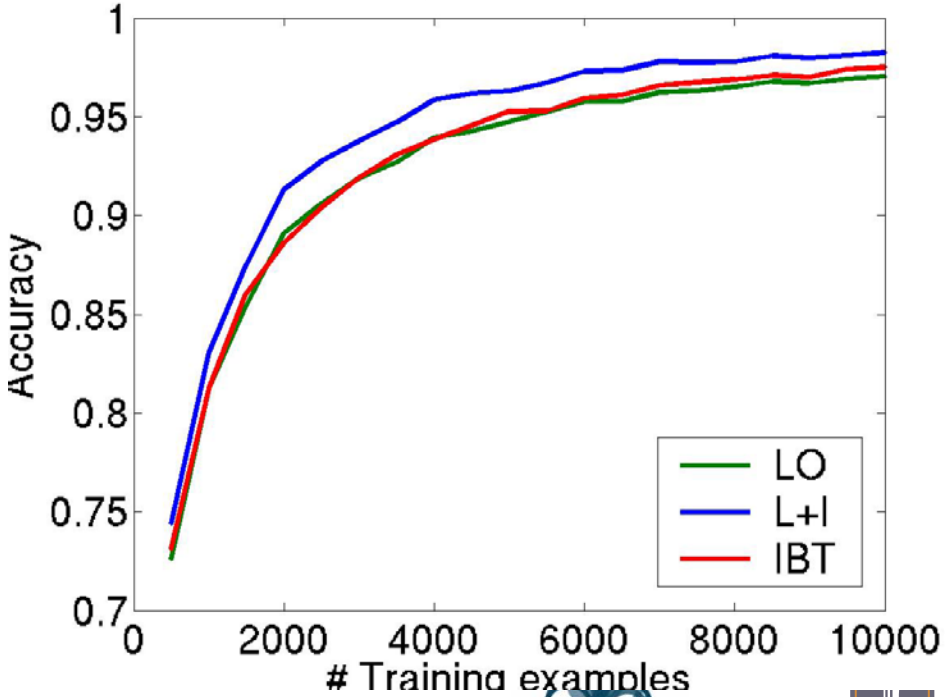
- Local $\varepsilon \le \varepsilon_{opt} + \left((d \log m + \log 1/\delta) / m \right)^{1/2}$
- Global $\varepsilon \le 0 + \left((cd \log m + c^2 d + \log 1/\delta) / m \right)^{1/2}$

Indication for hardness of problem

Bounds

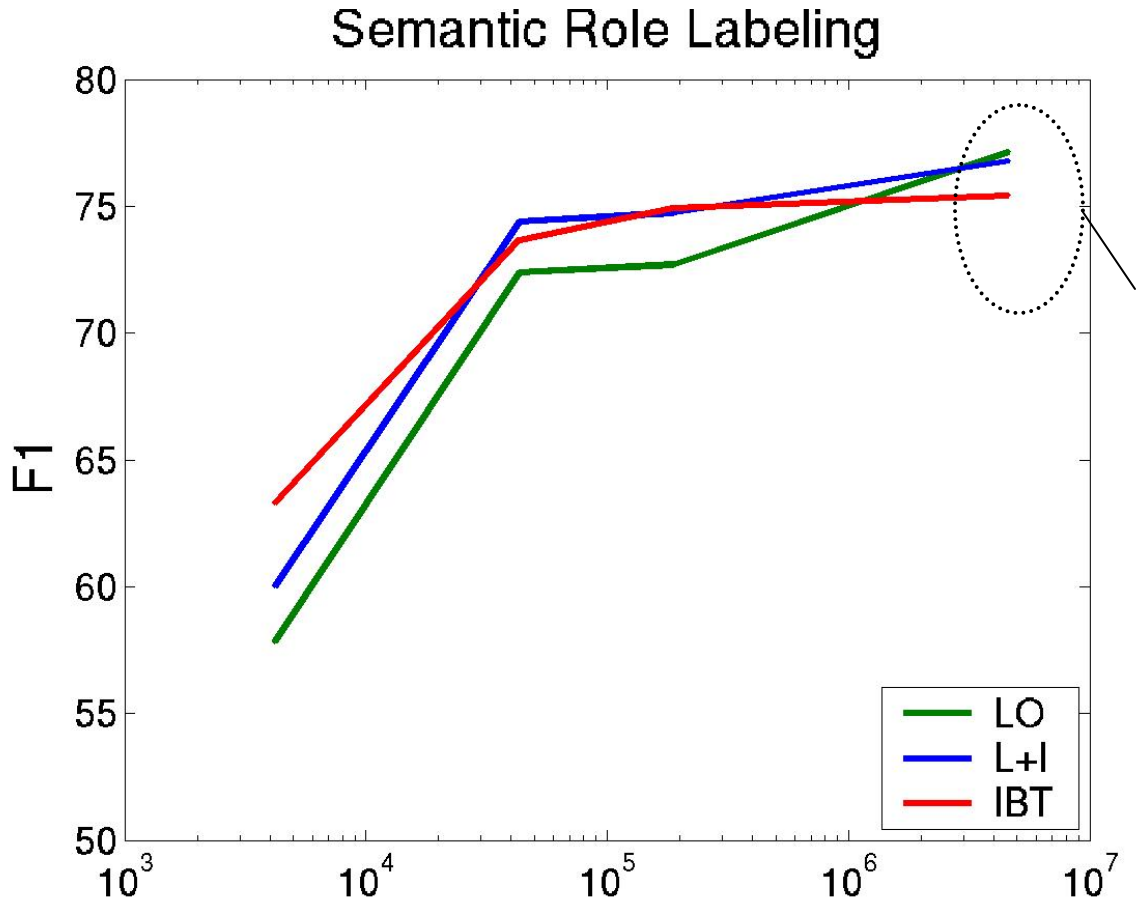


Simulated Data



Relative Merits: SRL

In some cases problems are hard due to lack of training data.
Semi-supervised learning



L+I is better.

When the problem is artificially made harder, the tradeoff is clearer.

hard

Difficulty of the learning problem
(# features)

easy

Outline

- Constrained Conditional Models
  - Motivation
  - Examples

- Training Paradigms: Investigate ways for training models and combining constraints
  - Joint Learning and Inference vs. decoupling Learning & Inference
  - Guiding Semi-Supervised Learning with Constraints
    - Features vs. Constraints
    - Hard and Soft Constraints

- Examples
  - Semantic Parsing
  - Information Extraction
  - Pipeline processes

Information extraction without Prior Knowledge

Lars Ole Andersen . Program analysis and specialization for the C Programming language. PhD thesis. DIKU , University of Copenhagen, May 1994 .

Prediction result of a trained HMM

[AUTHOR]	Lars Ole Andersen . Program analysis and
[TITLE]	specialization for the
[EDITOR]	C
[BOOKTITLE]	Programming language
[TECH-REPORT]	. PhD thesis .
[INSTITUTION]	DIKU , University of Copenhagen , May
[DATE]	1994 .

Violates lots of natural constraints!

Examples of Constraints

- Each field must be a consecutive list of words and can appear at most once in a citation.
- State transitions must occur on punctuation marks.
- The citation can only start with *AUTHOR* or *EDITOR*.
- The words *pp.*, *pages* correspond to *PAGE*.
- Four digits starting with 20xx and 19xx are *DATE*.
- Quotations can appear only in *TITLE*
- …….

Easy to express pieces of "knowledge"

Non Propositional; May use Quantifiers

Information Extraction with Constraints

- Adding constraints, we get correct results!
  - Without changing the model

- [AUTHOR] Lars Ole Andersen .
  [TITLE] Program analysis and specialization for the C Programming language .
  [TECH-REPORT] PhD thesis .
  [INSTITUTION] DIKU , University of Copenhagen ,
  [DATE] May, 1994 .

Features Versus Constraints

$$f_{\Phi,C}(\mathbf{x},\mathbf{y}) = \sum w_i \phi_i(\mathbf{x},\mathbf{y}) - \sum \rho_i d_{C_i}(\mathbf{x},\mathbf{y}).$$

- $\phi_i : X \times Y \rightarrow R$; $C_i : X \times Y \rightarrow \{0,1\}$; $d: X \times Y \rightarrow R$;
  - In principle, constraints and features can encode the same propeties
  - In practice, they are very different

- Features
  - Local , short distance properties – to allow tractable inference
  - Propositional (grounded):
  - E.g. True if: "the" followed by a Noun occurs in the sentence"
- Constraints
  - Global properties
  - Quantified, first order logic expressions
  - E.g.True if: "all y_is in the sequence y are assigned different values."

Indeed, used differently

Encoding Prior Knowledge

- Consider encoding the knowledge that:
  - Entities of type A and B cannot occur simultaneously in a sentence
- The "Feature" Way
  - Results in higher order HMM, CRF
  - May require designing a model tailored to knowledge/constraints
  - Large number of new features: might require more labeled data
  - Wastes parameters to learn indirectly knowledge we have.

- The Constraints Way
  - Keeps the model simple; add expressive constraints directly
  - A small set of constraints
  - Allows for decision time incorporation of constraints

Outline

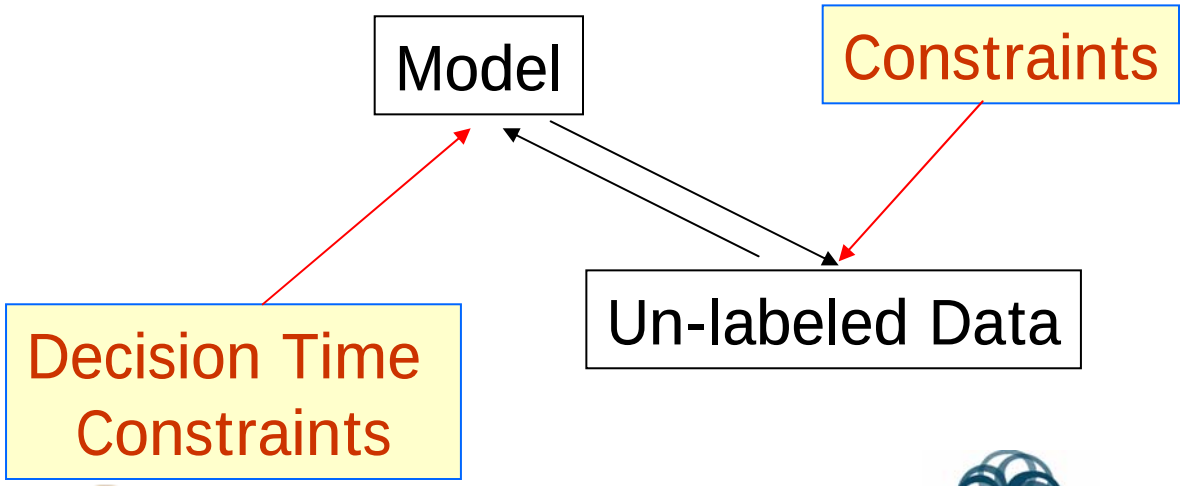
- Constrained Conditional Models
  - Motivation
  - Examples

- Training Paradigms: Investigate ways for training models and combining constraints
  - Joint Learning and Inference vs. decoupling Learning & Inference
  - Guiding Semi-Supervised Learning with Constraints
    - Features vs. Constraints
    - Hard and Soft Constraints

- Examples
  - Semantic Parsing
  - Information Extraction
  - Pipeline processes

Guiding Semi-Supervised Learning with Constraints

- In traditional Semi-Supervised learning the model can drift away from the correct one.
- Constraints can be used
  - At decision time, to bias the objective function towards favoring constraint satisfaction.
  - At training to improve labeling of un-labled data (and thus improve the model)



Training Strategies

$$f_{\Phi,C}(\mathbf{x},\mathbf{y}) = \sum w_i \phi_i(\mathbf{x},\mathbf{y}) - \sum \rho_i d_{C_i}(\mathbf{x},\mathbf{y}).$$

- Hard Constraints or Weighted Constraints
  - Hard constraints: set penalties to infinity
    - No more degrees of violation
  - Weighted Constraints
    - Need to figure out penalties values

Training Algorithms:
L+ CI, L+ wCI

CIBT, wCIBT

- Factored / Jointed Approaches
  - Factored Models (L+I)
    - Learn model weights and constraints' penalties separately
  - Joint Models (IBT)
    - Learn the model weights and constraints' penalties jointly
  - L+I vs IBT: [Punyakanok et. al. 05]

Factored (L+I) Approaches

- Learning model weights
  - HMM

- Constraints Penalties
  - Hard Constraints : infinity
  - Weighted Constraints:
    - $\rho_i = -\log P\{\text{Constraint } C_i \text{ is violated in training data}\}$

Joint Approaches

Algorithm 1 IBT training: CIBT & wCIBT

Require: D is the training dataset, K is the number of constraints, M is the number of iterations

for $i = 1 \ldots K$ **do**
  $if(hardConstraints)$ then $\rho_i = \infty$ else $\rho_i = 0$
end for
for $i = 1 \ldots M$ **do**
  for $(\mathbf{x}, \mathbf{y}^*) \in D$ **do**
    $\hat{\mathbf{y}} = \text{argmax}_{\mathbf{y}} [\sum w_i \phi_i(\mathbf{x}, \mathbf{y}) - \sum \rho_i d_{C_i}(\mathbf{x}, \mathbf{y})]$
    $\mathbf{w} = \mathbf{w} + \Phi(\mathbf{x}, \mathbf{y}^*) - \Phi(\mathbf{x}, \hat{\mathbf{y}})$
    if weightedConstraints **then**
      $\rho = \rho + d_C(\mathbf{x}, \mathbf{y}^*) - d_C(\mathbf{x}, \hat{\mathbf{y}})$
    end if
  end for
end for

Structured Perceptron

Semi-supervised Learning with Constraints

[Chang, Ratinov, Roth, ACL'07]

λ=learn(T)

Supervised learning algorithm parameterized by λ

Inference based augmentation of the training set (feedback) (inference with constraints).

For N iterations do

$T=\phi$

For each x in unlabeled dataset

$\{y_1,\ldots,y_K\} \leftarrow$ InferenceWithConstraints(x,C, λ)

$T=T \cup \{(x, y_i)\}_{i=1\ldots k}$

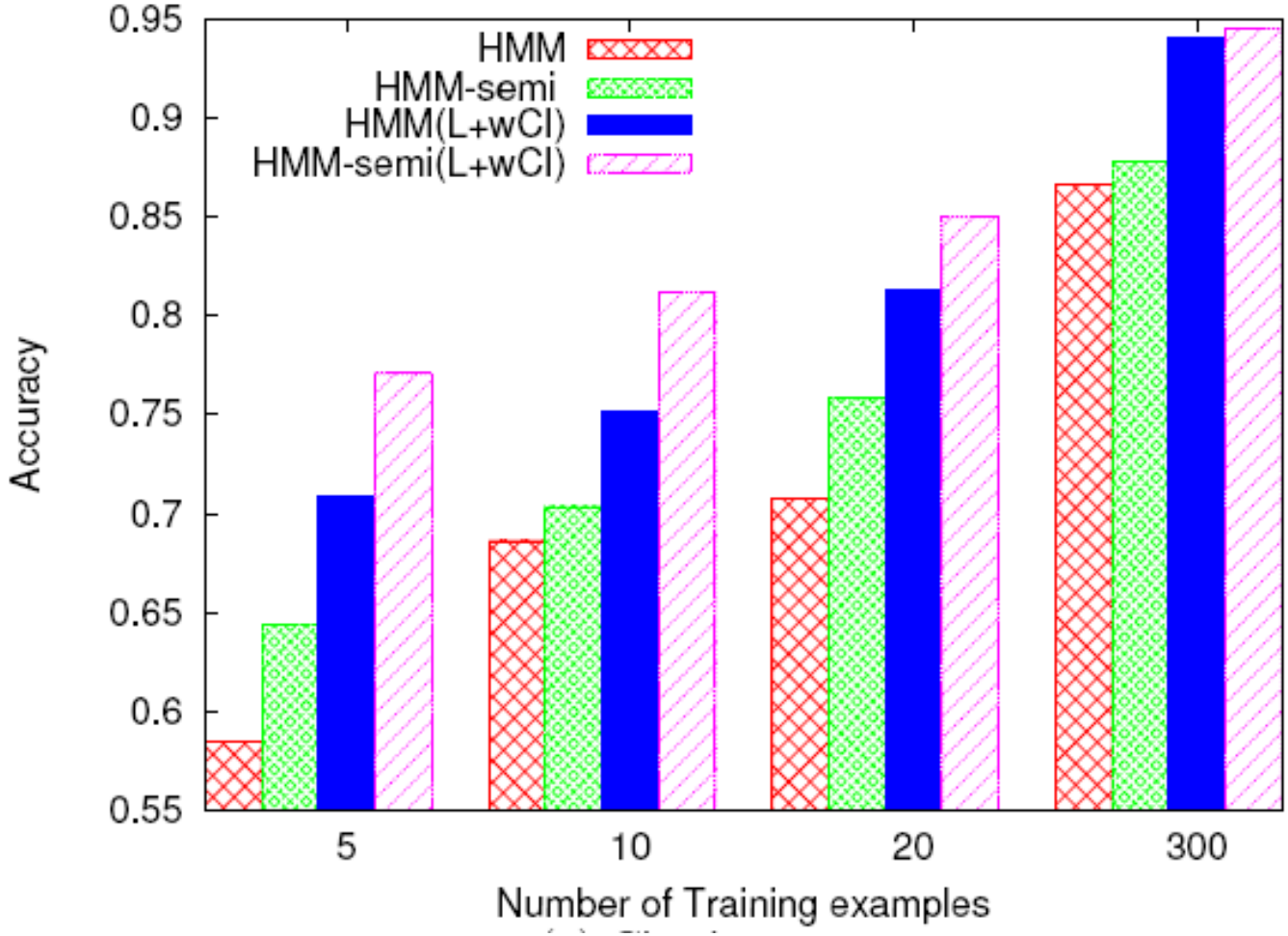
$\lambda= \gamma\ \lambda+(1-\gamma\ )$learn(T)

Learn from new training data. Weigh supervised and unsupervised model.

Outline

- Constrained Conditional Model
  - Feature v.s Constraints
  - Inference
  - Training
  - Semi-supervised Learning
- Results
- Discussion

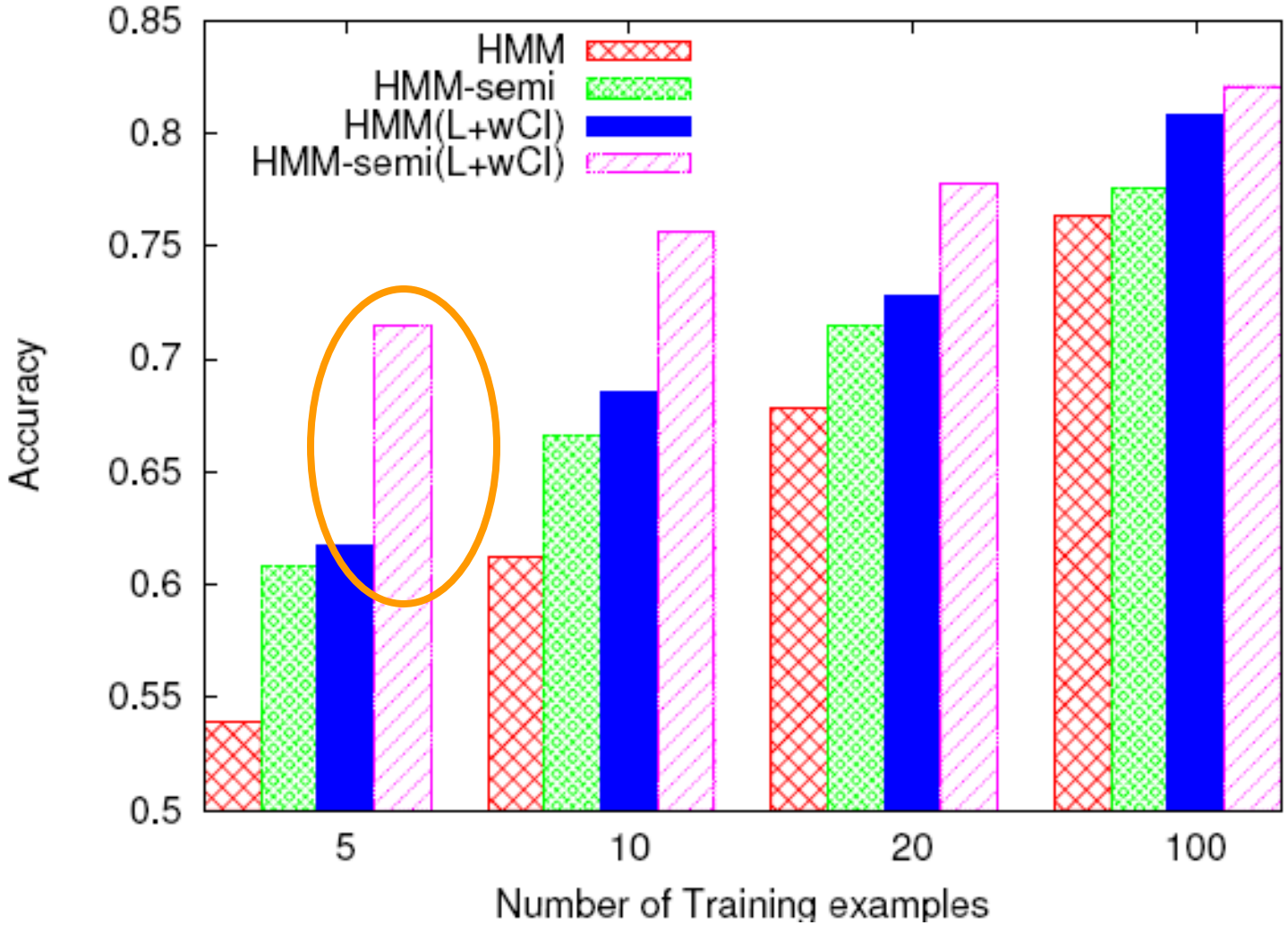
Results on Factored Model -- Citations



In all cases: semi = 1000 unlabeled examples.

In all cases: Significantly better results than existing results [Chang et. al. '07]

Results on Factored Model -- Advertisements



Hard Constraints vs. Weighted Constraints

Constraints are close to perfect

(a)-Citations

Training samples	5	10	20	300
Weighted constraints	77.09	81.25	85.00	94.51
Hard constraints	78.18	81.11	85.16	92.80

(b)-Advertisement

Training samples	5	10	20	300
Weighted constraints	71.46	75.61	77.76	82.06
Hard constraints	69.91	73.46	75.25	79.59

Labeled data might not follow the constraints

Factored vs. Jointed Training

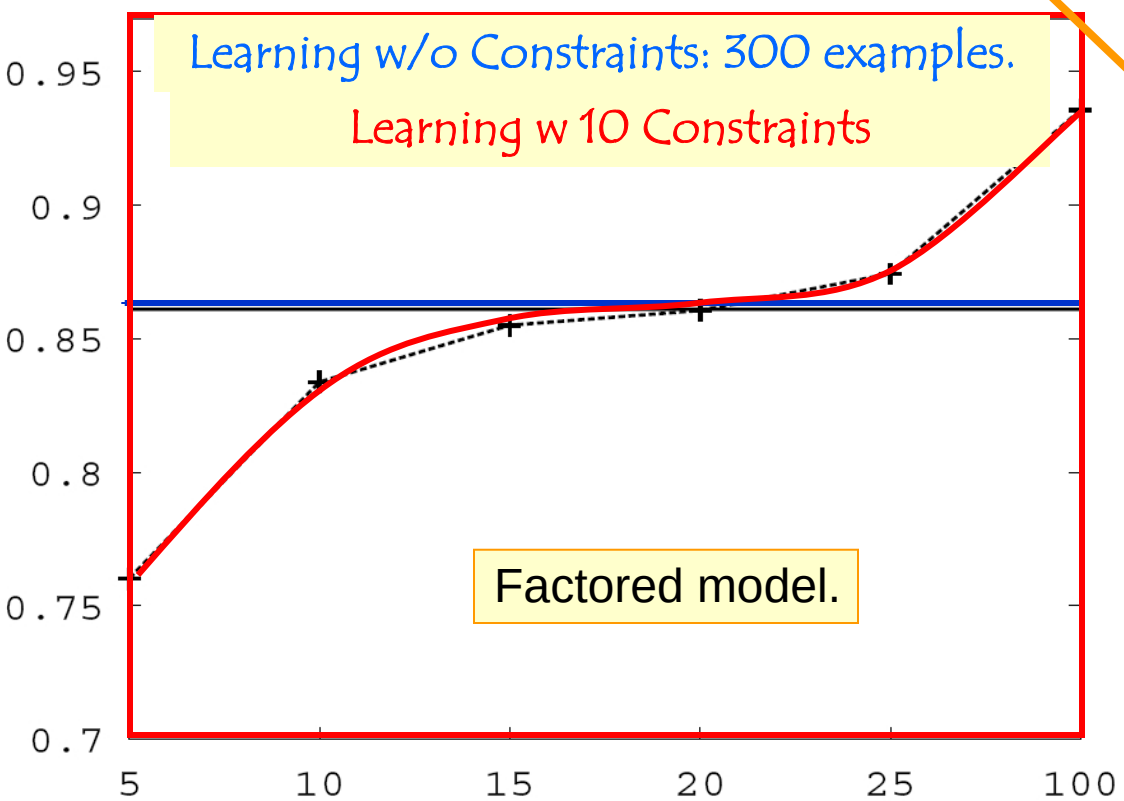
- Using the best models for both settings
  - Factored training: HMM + weighted constraints
  - Jointed training: Perceptron + weighted constraints
  - Same feature set
- Without constraints
  - Factored Model is better, HMM > perceptron
  - Many labelled examples, perceptron > HMM
  - Few labelled examples, HMM > perceptron

Agrees with earlier results in the supervised setting ICML'05, IJCAI'5

	Discriminative		Maximum Likelihood	
Training	L	wCIBT	L	L+wCI
5	50.14	61.65	58.48	70.85
10	59.90	69.64	68.61	75.11
20	68.26	78.46	70.81	81.31
300	89.83	93.89	86.66	94.08

Value of Constraints in Semi-Supervised Learning

Objective function: $f_{\Phi,C}(\mathbf{x},\mathbf{y}) = \sum w_i \phi_i(\mathbf{x},\mathbf{y}) - \sum \rho_i d_{C_i}(\mathbf{x},\mathbf{y}).$

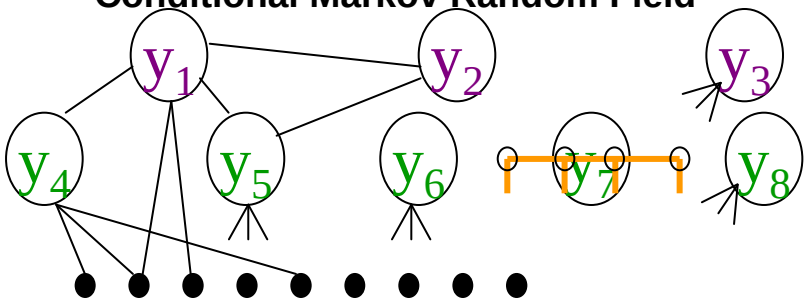


# of available labeled examples

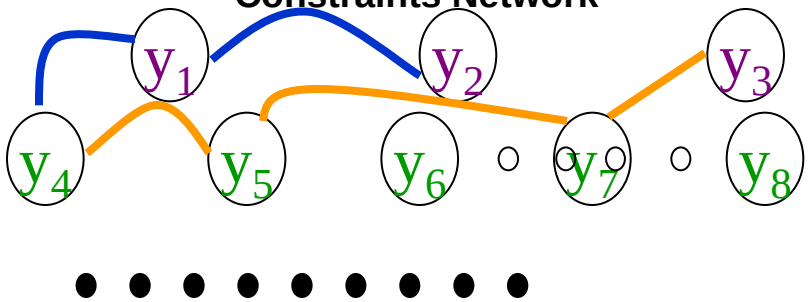
Constraints are used to Bootstrap a semi-supervised learner
Poor model + constraints used to annotate unlabeled data, which in turn is used to keep training the model.

Summary: Constrained Conditional Models

Conditional Markov Random Field



Constraints Network



$$y^* = \text{argmax}_y \sum w_i \, \phi(x; y)$$

- Linear objective functions
- Typically $\phi(x,y)$ will be local functions, or $\phi(x,y) = \phi(x)$

$$- \sum_i \rho_i \, d_C(x,y)$$

- Expressive constraints over output variables
- Soft, weighted constraints
- Specified declaratively as FOL formulae

- Clearly, there is a joint probability distribution that represents this mixed model.
- We would like to:
  - Learn a simple model or several simple models
  - Make decisions with respect to a complex model

Key difference from MLNs which provide a concise definition of a model, but the whole joint one.

Conclusion

- Constrained Conditional Models combine
  - Learning conditional models with using declarative expressive constraints
  - Within a constrained optimization framework
- Use constraints! The framework supports:
  - A clean way of incorporating constraints to bias and improve decisions of supervised learning models
    - Significant success on several NLP and IE tasks (often, with ILP)
  - A clean way to use (declarative) prior knowledge to guide semi-supervised learning

LBJ (Learning Based Java): http://L2R.cs.uiuc.edu/~cogcomp
A modeling language for Constrained Conditional Models. Supports programming along with building learned models, high level specification of constraints and inference with constraints

Questions?

- Thank you