Cognitive Computation Group
University of Illinois at Urbana-Champaign

A Joint Model for Extended Semantic Role Labeling

Vivek Srikumar and Dan Roth

The punch line

1. Sentences express relations via several surface phenomena
2. These interact with each other
3. We present a model for joint inference *without* re-learning existing state-of-the-art systems

Semantic Role Labeling

The field goal of Brien changed the game in the fourth quarter.

When did the game change?

Semantic Role Labeling

The field goal of Brien changed the game in the fourth quarter.

When did the game change?

Let's ask the Semantic Role Labeler

Punyakanok et al 2008,
Toutanova et al 2008

Semantic Role Labeling

The field goal of Brien **changed** the game in the fourth quarter.

A0: The causer of the transformation

A1: The thing changing

AM-TMP: Temporal

When did the game change?

Let's ask the Semantic Role Labeler

Punyakanok et al 2008,
Toutanova et al 2008

Semantic Role Labeling

The field goal of Brien **changed** the game in the fourth quarter.

A0: The causer of the transformation

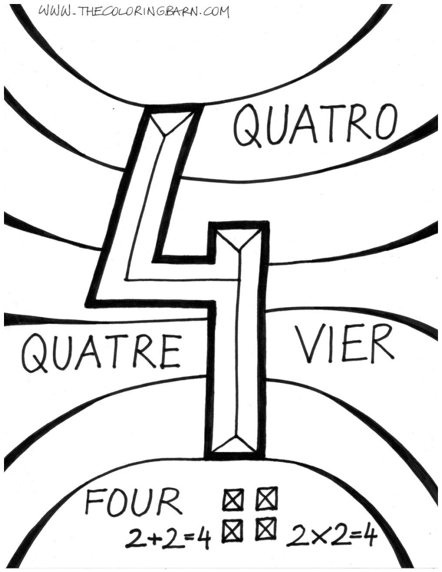
A1: The thing changing

AM-TMP: Temporal

When did the game change?

Let's ask the Semantic Role Labeler

Punyakanok et al 2008,
Toutanova et al 2008



Why extend Semantic Role Labeling?

The field goal of Brien **changed** the game in the fourth quarter.

A0: The causer of the transformation

A1: The thing changing

AM-TMP: Temporal

Who scored the field goal?

Why extend Semantic Role Labeling?

The field goal of Brien **changed** the game in the fourth quarter.

A0: The causer of the transformation

A1: The thing changing

AM-TMP: Temporal

Who scored the field goal?

Why extend Semantic Role Labeling?

The field goal of Brien **changed** the game in the fourth quarter.

A0: The causer of the transformation

A1: The thing changing

AM-TMP: Temporal

Who scored the field goal?

Verb SRL doesn’t help us here!

The field goal and Brien are connected by the preposition *of*

Why extend Semantic Role Labeling?

The field goal of Brien **changed** the game in the fourth quarter.

A0: The causer of the transformation

A1: The thing changing

AM-TMP: Temporal

Who scored the field goal?

Verb SRL doesn't help us here!

The field goal and Brien are connected by the preposition *of*

Sentences express relations via several linguistic phenomena (not just verbs and nominalizations)

Prepositions, compound nouns, commas…

Linguistic phenomena do not operate in vacuum

The field goal of Brien **changed** the game in the fourth quarter.

A0: The causer of the transformation

A1: The thing changing

AM-TMP: Temporal

Linguistic phenomena do not operate in vacuum

The field goal of Brien changed the game in the fourth quarter.

A0: The causer of the transformation

A1: The thing changing

AM-TMP: Temporal

- Prepositions indicate semantic relations

Linguistic phenomena do not operate in vacuum

The field goal **of** Brien **changed** the game **in** the fourth quarter.

A0: The causer of the transformation

A1: The thing changing

AM-TMP: Temporal

Agent of action

Temporal

- Prepositions indicate semantic relations

Linguistic phenomena do not operate in vacuum

The field goal of Brien changed the game in the fourth quarter.

A0: The causer of the transformation

A1: The thing changing

AM-TMP: Temporal

Agent of action

Temporal

- Prepositions indicate semantic relations

Linguistic phenomena do not operate in vacuum

The field goal of Brien changed the game in the fourth quarter.

A0: The causer of the transformation

A1: The thing changing

AM-TMP: Temporal

Agent of action

Temporal

- Prepositions indicate semantic relations

- The different linguistic phenomena interact with each other

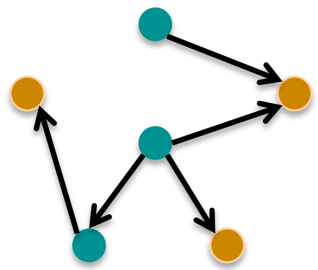
The model: A Constrained Conditional Model

[Chang et al 2007, 2008]

$$arg \max_y \sum_i \mathbf{w}_i \phi_i(x,y) - \sum \rho_i C_i(x,y)$$

Score assigned to structure by learned models

Penalty due to constraint violation. Enforces structure



A set of labeled parts

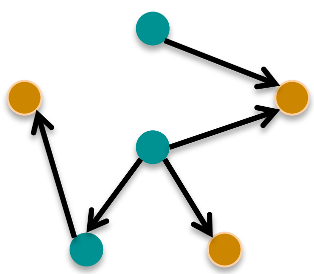
The model: A Constrained Conditional Model

[Chang et al 2007, 2008]

$$\arg\max_{y} \sum_{i} \mathbf{w}_i \phi_i(x, y) - \sum \rho_i C_i(x, y)$$

Score assigned to structure by learned models

Penalty due to constraint violation. Enforces structure



Our structure is a combination of verb SRL and preposition roles

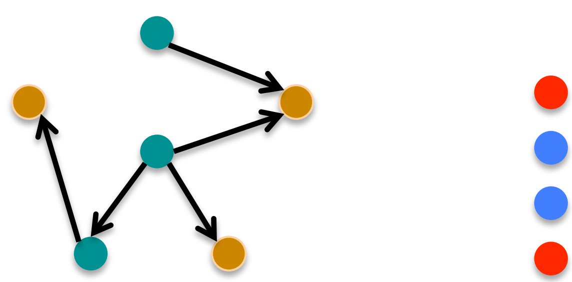
The model: A Constrained Conditional Model

[Chang et al 2007, 2008]

$$arg\max_{y} \sum_{i} \mathbf{w}_i^{SRL}\phi_i^{SRL}(x,y) + \sum_{i} \mathbf{w}_i^{Prep}\phi_i^{Prep}(x,y) - \sum \rho_i C_i(x,y)$$

Score assigned to preposition and SRL structures by *independently learned models*

Penalty due to constraint violation. Enforces structure



Our structure is a combination of verb SRL and preposition roles

The model: A Constrained Conditional Model

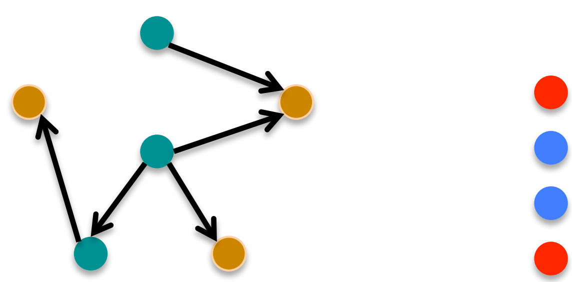
Re-scaling parameters

[Chang et al 2007, 2008]

$$\arg\max_{y} \underbrace{\sum_i \lambda_i \mathbf{w}_i^{SRL} \phi_i^{SRL}(x,y) + \sum_i \mu_i \mathbf{w}_i^{Prep} \phi_i^{Prep}(x,y)} - \sum \rho_i C_i(x,y)$$

Score assigned to preposition and SRL structures by *independently* learned models

Penalty due to constraint violation. Enforces structure



Our structure is a combination of verb SRL and preposition roles

The model: A Constrained Conditional Model

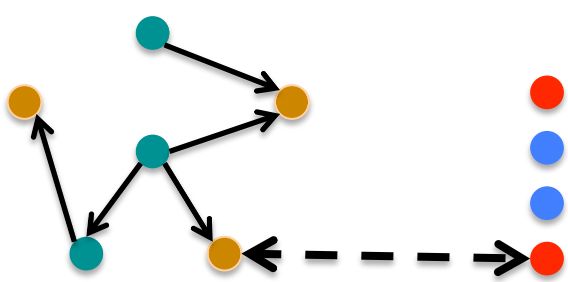
Re-scaling parameters

[Chang et al 2007, 2008]

$$arg\max_{y} \sum_{i} \lambda_i \mathbf{w}_i^{SRL} \phi_i^{SRL}(x,y) + \sum_{i} \mu_i \mathbf{w}_i^{Prep} \phi_i^{Prep}(x,y) - \sum \rho_i C_i(x,y)$$

Score assigned to preposition and SRL structures by *independently* learned models

Penalty due to constraint violation. Enforces structure and coherent joint prediction



Our structure is a combination of verb SRL and preposition roles

Outline

- Tasks and data
- The model
- Experiments and results

Outline

- Tasks and data
- The model
- Experiments and results

Task 1: Verb Semantic Role Labeling

- Identifying the predicate argument structure of verbs in sentences
- Standard Propbank annotation over PennTreebank
- Formed the basis of CoNLL shared tasks (Carreras and Marques, 2004, 2005)

Our baseline

- Modified version of the Illinois Verb SRL system

  [Koomen, Punyakanok, Roth and Yih, 2005], [Punyakanok, Roth and Yih 2008]

- Consists to two learned models:
  - **Argument identifier:** Scores argument candidates for whether they are actually arguments
  - **Argument classifier:** Assigns scores for argument labels for each candidate
    - **Null** is also a label

- Inference:
  - Ensures that the output is a well formed structure
  - Ensures that the argument identifier and the classifier agree

Integer Linear Program inference for verb SRL

$\max$ Total argument classifier score $+$ Total argument identifier score

Subject to

Constraints between output variables

Integer Linear Program inference for verb SRL

$$\max_{\mathbf{v^C}, \mathbf{v^I}} \sum_{i,a} \theta^C_{i,a} v^C_{i,a} +$$ Total argument identifier score

Subject to

Binary vectors of classifier and identifier decisions

Integer Linear Program inference for verb SRL

$$\max_{\mathbf{v^C},\mathbf{v^I}} \sum_{i,a} \boxed{\theta^{C}_{i,a}}\boxed{v^{C}_{i,a}} + \text{Total argument identifier score}$$

Subject to

Binary vectors of classifier and identifier decisions

Score(argument ***i*** has the label ***a***)

Indicator [argument ***i*** takes the label ***a***]

Integer Linear Program inference for verb SRL

$$\max_{\mathbf{v^C},\mathbf{v^I}} \sum_{i,a} \boxed{\theta^C_{i,a}}\boxed{v^C_{i,a}} + \sum_{i} \boxed{\theta^I_i}\boxed{v^I_i}$$

Subject to

Binary vectors of classifier and identifier decisions

Score(argument ***i*** has the label ***a***)

Indicator [argument ***i*** takes the label ***a***]

Score assigned by identifier for argument ***i***

Indicator: candidate ***i*** is an argument

Integer Linear Program inference for verb SRL

$$\max_{\mathbf{v^C}, \mathbf{v^I}} \sum_{i,a} \theta^C_{i,a} v^C_{i,a} + \sum_i \theta^I_i v^I_i$$

Subject to

Constraints between output variables

Integer Linear Program inference for verb SRL

$$\max_{\mathbf{v^C},\mathbf{v^I}} \sum_{i,a} \theta^C_{i,a} v^C_{i,a} + \sum_i \theta^I_i v^I_i$$

Subject to

Each candidate can have exactly one label

No duplicate core arguments

No overlapping or embedding arguments

Some argument labels are illegal, depending on the predicate

R-arg and C-arg constraints

Agreement between identifier and classifier decisions

Task 2: Preposition Role Labeling

- Determining the role of the preposition phrase in relation to its attachment point
  - **Prepositions are highly polysemous**

Task 2: Preposition Role Labeling

- Determining the role of the preposition phrase in relation to its attachment point
  - **Prepositions are highly polysemous**

live at Conway House

stopped at the station

the camp on the island

came on Sep. 26th

stopped at 9 PM

cooler in evening

priced at 9 pounds

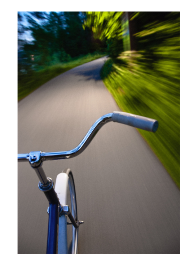
drive at 50 mph

look at the watch

Task 2: Preposition Role Labeling

- Determining the role of the preposition phrase in relation to its attachment point
  - **Prepositions are highly polysemous**

Location	Temporal	Numeric/Level	Object of verb
live at Conway House	came on Sep. 26th	priced at 9 pounds	look at the watch
stopped at the station	stopped at 9 PM	drive at 50 mph	
the camp on the island	cooler in evening		



Preposition Role Labeling

- Related to the preposition sense disambiguation task of SemEval 2007
  - **SemEval 2007** [Litkowski and Hargraves, 2007]
  - **Penn Treebank** (02-04, 23) [Dahlmeier, Ng, Schultz 2009]
- Preposition roles obtained by merging semantically related senses of different prepositions
- Totally 22 preposition labels obtained
  - **Restricted to** at, in, for, of, on, to, with

Predicting preposition roles

- A multiclass classifier based on state-of-the-art sense disambiguation features of [Hovy et al 2009, 2010]

Predicting preposition roles

- A multiclass classifier based on state-of-the-art sense disambiguation features of [Hovy et al 2009, 2010]

- The prediction can be represented as an integer linear program

Classifier score for role ***r*** for preposition ***p***

$$\max_{\mathbf{v}} \quad \sum_{p,r} \boxed{\theta_{p,r}}\boxed{v_{p,r}}$$

Indicator that preposition ***p*** is labeled with role ***r***

$$\text{subj. to} \quad \sum_{r} v_{p,r} = 1, \quad \forall p$$

$$v_{p,r} \in \{0,1\}, \quad \forall p, r$$

Exactly one label per preposition

Outline

- Tasks and data
- The model
- Experiments and results

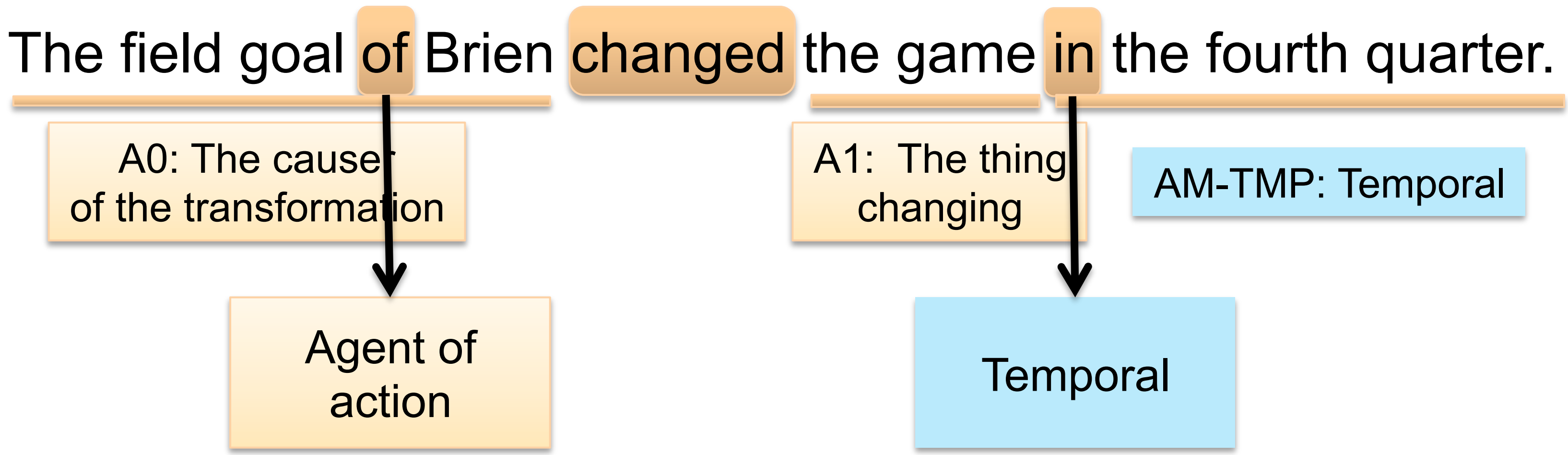
Overview of the model

- Key intuition

The correct interpretation of a sentence is the one that gives a consistent analysis across all the linguistic phenomena expressed in it

- We have two tasks each with their own datasets and learned models

The ideal joint model



1. Should account for dependencies between linguistic phenomena
2. Should be extensible to allow for easy addition of new phenomena
3. Should be able to use existing state of the art models
4. Should minimize use of expensive jointly labeled data

Dependencies between tasks

- Joint constraints linking the two tasks. For example,
  - *If a verb argument that starts with in is labeled* ***AM-TMP****, then the preposition should be labeled* ***Temporal***
  - *If a verb attached preposition is labeled* ***Temporal****, then there should be some verb argument that starts with that preposition and is labeled* ***AM-TMP***
- **Note:** These constraints are universally quantified over verb argument candidates that start with a preposition.

Joint inference

Maximize overall score

Subject to

Phenomena specific constraints

and

Joint constraints

Joint inference

Maximize overall score

For a single phenomenon p, the score is given by:

$$\sum_{Z \in \mathbf{Z}^{\mathbf{p}}} \sum_{y} \theta_{Z,y} \cdot v_{Z,y}$$

Joint inference

Maximize overall score

For a single phenomenon p, the score is given by:

$$\sum_{Z \in \boxed{\mathbf{Z}^{\mathbf{P}}}} \sum_{y} \boxed{\theta_{Z,y}} \cdot \boxed{v_{Z,y}}$$

Indicator that part y is assigned label Z

Score that part y is assigned label Z

All possible labels that can be assigned to parts of the structure for phenomenon p

Joint inference

Maximize overall score

For a several phenomena, the score is given by:

$$\sum_{p} \sum_{Z \in \mathbf{Z}^{\mathbf{P}}} \lambda_Z \sum_{y} \theta^{p}_{Z,y} \cdot \upsilon^{p}_{Z,y}$$

Joint inference

Maximize overall score

For a several phenomena, the score is given by:

$$\boxed{\sum_{p}} \sum_{Z \in \mathbf{Z^P}} \lambda_Z \sum_{y} \theta^{p}_{Z,y} \cdot \upsilon^{p}_{Z,y}$$

Sum over all phenomena

Joint inference

Maximize overall score

For a several phenomena, the score is given by:

$$\boxed{\sum_{p}} \sum_{Z \in \mathbf{Z}^{\mathbf{P}}} \boxed{\lambda_Z} \sum_{y} \theta^{p}_{Z,y} \cdot v^{p}_{Z,y}$$

Sum over all phenomena

Scaling parameter for label Z of phenomenon p

Our individual models (SRL and preposition role) were trained on different datasets

To scale the scoring functions, we associate each label with a scaling parameter

Joint inference

Maximize overall score

Subject to

Phenomena specific constraints

and

Joint constraints

Joint inference

$$\max_{\mathbf{v}} \sum_{p} \sum_{Z \in \mathbf{Z}^{\mathbf{P}}} \lambda_Z \sum_{y} \theta^{p}_{Z,y} \cdot v^{p}_{Z,y}$$

Subject to

Phenomena specific constraints

and

Joint constraints

Learning to scale

- The joint model only requires re-scaling parameters
  - **One parameter per label**
- Trained with Structure Perceptron on a very small jointly labeled corpus

An ideal joint model

1. Should account for dependencies between linguistic phenomena
2. Should be extensible to allow for easy addition of new phenomena
3. Should be able to use existing state of the art models with minimal use of expensive jointly labeled data

An ideal joint model

1. Should account for dependencies between linguistic phenomena

   Joint constraints

2. Should be extensible to allow for easy addition of new phenomena

   As easy as adding new terms to the inference

3. Should be able to use existing state of the art models with minimal use of expensive jointly labeled data

   Only rescales label scores

Outline

- Tasks and data
- The model
- Experiments and results

Constraints: Background knowledge

- A combination of hand written and automatically extracted constraints
- Hand written constraints
  - *Each verb attached preposition that is labeled as **Temporal** should correspond to the start of some **AM-TMP***
  - *For verb attached prepositions, some roles should correspond to an non-**null** argument label*
- Extracted constraints for arguments that start with a preposition using Treebank data
  - Set of allowed verb argument labels for each role and vice versa
    - Included only those constraints that had a support of ten

Settings and data

- Trained on Penn Treebank data
  - **Verb SRL: All sections**
  - **Preposition role: Sections 2-4**
- Test performance on Section 23
- First 500 sentences of Section 02 held out for training rescaling weights
- Inference is done with an ILP solver using a cutting plane solver

Comparisons

- Comparison with baselines and pipeline
  - **Baseline: Existing Verb SRL and preposition role classifiers**
  - **Pipelines:**
    - **Verb → Preposition role**: SRL prediction added to the preposition's feature set if an argument started with the preposition
    - **Preposition →Verb SRL**: Preposition role prediction added to the verb argument candidate's features if the candidate started with the preposition

Results

Verb SRL

Independent

Prep. Role -> SRL

Joint inference

F1

Preposition Role

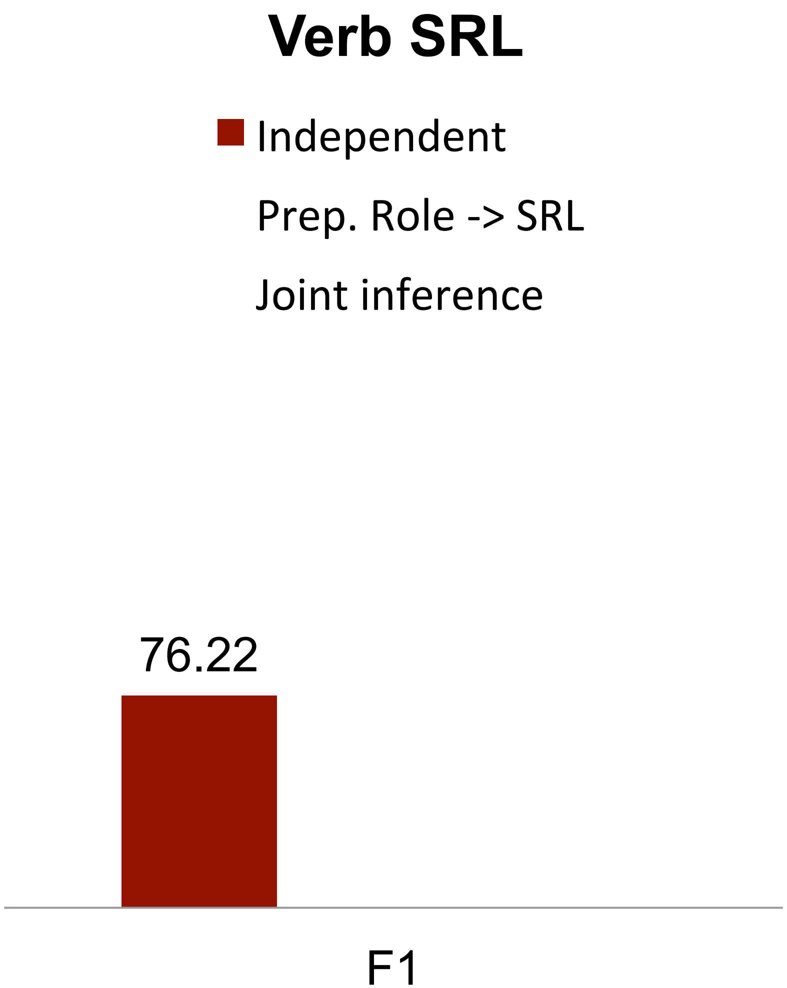
Independent

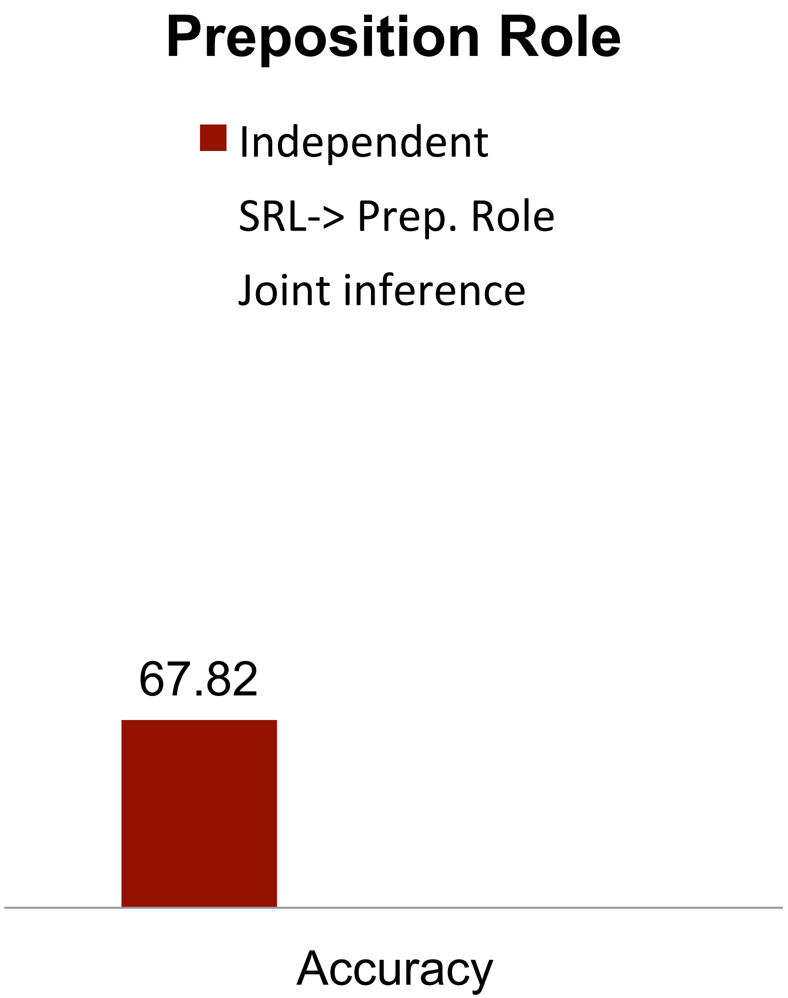
SRL-> Prep. Role

Joint inference

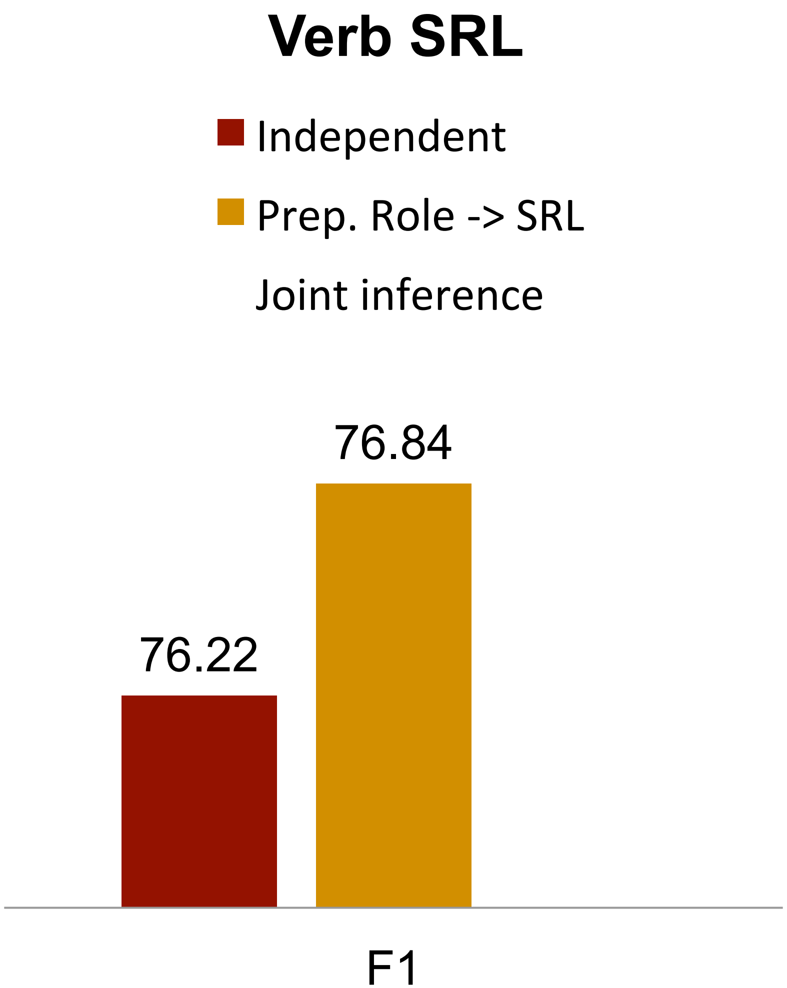
Accuracy

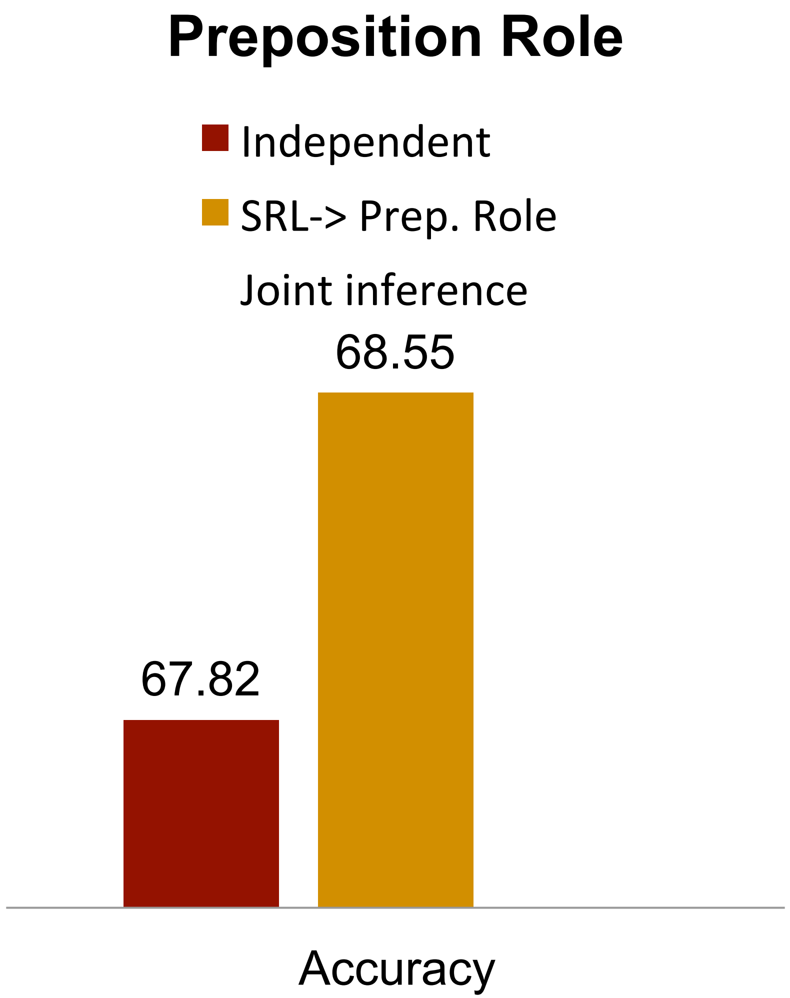
Results


Verb SRL
Independent
Prep. Role -> SRL
Joint inference
76.22
F1

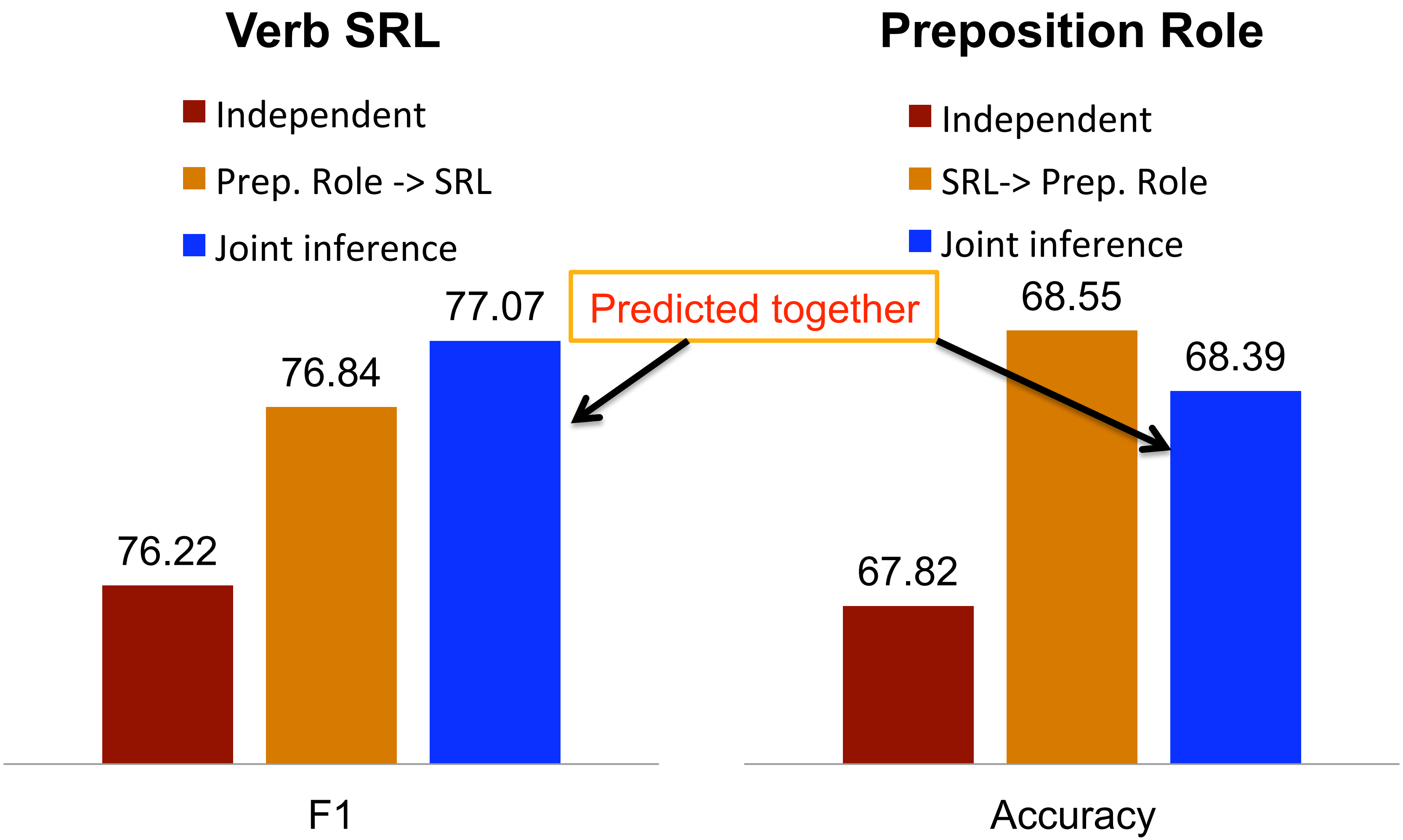

Preposition Role
Independent
SRL-> Prep. Role
Joint inference
67.82
Accuracy

Results


Verb SRL
Independent
Prep. Role -> SRL
Joint inference
76.84
76.22
F1


Preposition Role
Independent
SRL-> Prep. Role
Joint inference
68.55
67.82
Accuracy

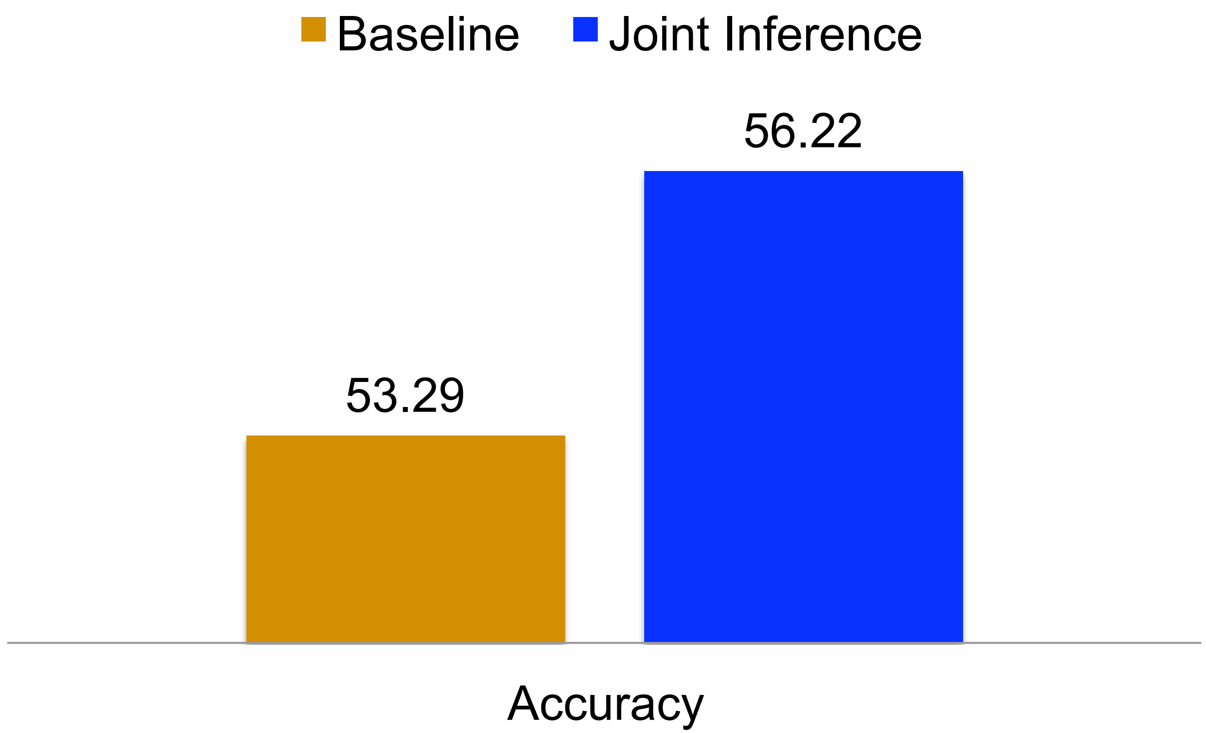
Results



COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Adaptation

- Replaced preposition classifier with one trained on SemEval data



Conclusion

1. Sentences express relations via several linguistic phenomena, which interact with each other

2. A model for joint inference without re-learning existing state-of-the-art systems to improve them

3. Easily extensible to include other tasks

Conclusion

1. Sentences express relations via several linguistic phenomena, which interact with each other
2. A model for joint inference without re-learning existing state-of-the-art systems to improve them
3. Easily extensible to include other tasks

See http://cogcomp.cs.illinois.edu

Questions?

Preposition Roles

Role	Train	Test
ACTIVITY	57	23
ATTRIBUTE	119	51
BENEFICIARY	78	17
CAUSE	255	116
CONCOMITANT	156	74
ENDCONDITION	88	66
EXPERIENCER	88	42
INSTRUMENT	37	19
LOCATION	1141	414
MEDIUMOFCOMMUNICATION	39	30
NUMERIC/LEVEL	301	174
OBJECTOFVERB	365	112
OTHER	65	49
PARTWHOLE	485	133
PARTICIPANT/ACCOMPANIER	122	58
PHYSICALSUPPORT	32	18
POSSESSOR	195	56
PROFESSIONALASPECT	24	10
RECIPIENT	150	70
SPECIES	240	58
TEMPORAL	582	270
TOPIC	148	54

Table 3: Preposition role data statistics for the Penn Treebank preposition dataset.