Position-aware Attention and Supervised Data Improve Slot Filling

Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli,
Christopher D. Manning

Stanford University
September, 2017

Knowledge Base Population

Penner is survived by his brother, John, a copy editor at the Times, and his former wife, Times sportswriter Lisa Dillman.

Knowledge Base Population

Penner is survived by his brother, John, a copy editor at the Times, and his former wife, Times sportswriter Lisa Dillman.

Subject	Relation/Slot	Object
Mike Penner	per:spouse	Lisa Dillman
Lisa Dillman	per:title	Sportswriter
Lisa Dillman	per:employee_of	Los Angeles Times
…	…	…

Knowledge Base Population

Penner is survived by his brother, John, a copy editor at the Times, and his former wife, Times sportswriter Lisa Dillman.

Subject	**Relation/Slot**	**Object**
Mike Penner	per:spouse	?
Lisa Dillman	per:title	?
Lisa Dillman	per:employee_of	?
…	…	…

Key Problems

- Stanford routinely found DL models ineffective at slot filling task.

Key Problems

- Stanford routinely found DL models ineffective at slot filling task.

1. **Existing models insufficiently tailored to relation extraction**
2. **Lack of a large-scale, fully supervised dataset for slot filling**

Overview

Model: A new position-aware attention model

+

Data: A new supervised dataset, TACRED

=

Result: Improved slot filling

Overview

Model: A new position-aware attention model

+

Data: A new supervised dataset, TACRED

=

Result: Improved slot filling

Relation Extraction

Penner is survived by his brother, John, a copy editor at the Times, and his former wife, Times sportswriter **Lisa Dillman**.

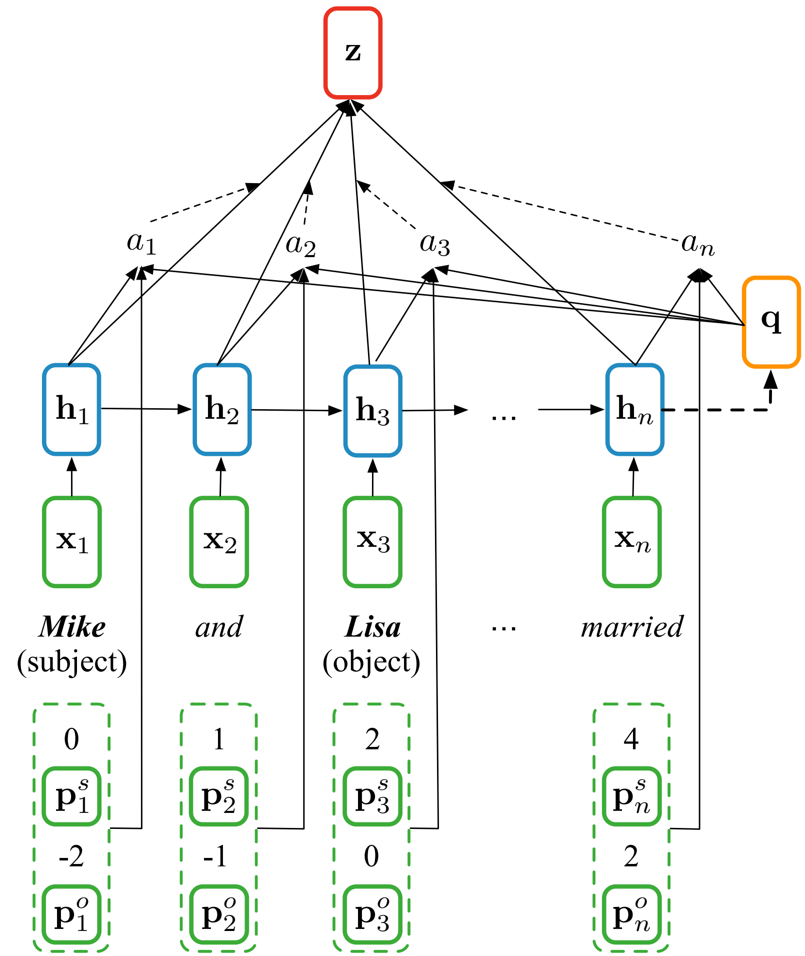
Relation Extraction

> **Penner** is survived by his brother, John, a copy editor at the Times, and his former wife, Times sportswriter **Lisa Dillman**.

Key elements

- Context (relevant + irrelevant)
- Entities (types + positions)

Position-aware Attention Model



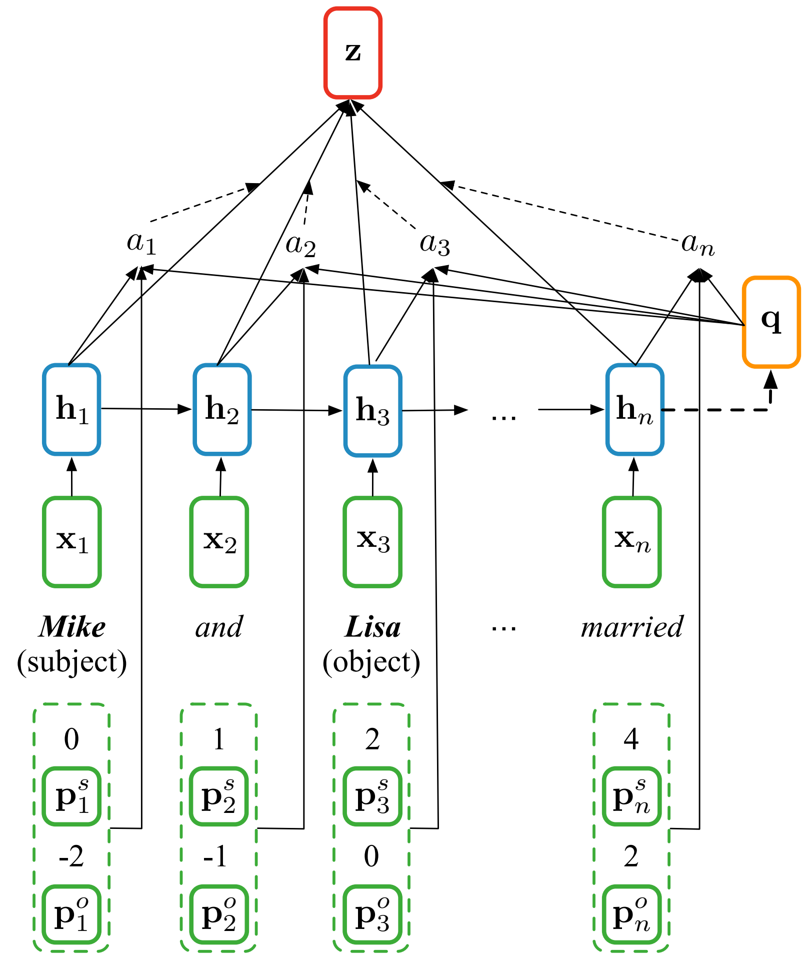
Position-aware Attention Model

Embedding Layers

Word: $\mathbf{x} = [\mathbf{x}_1, ..., \mathbf{x}_n]$

Position: $\mathbf{p}^s = [\mathbf{p}^s_1, ..., \mathbf{p}^s_n]$

$\mathbf{p}^o = [\mathbf{p}^o_1, ..., \mathbf{p}^o_n]$



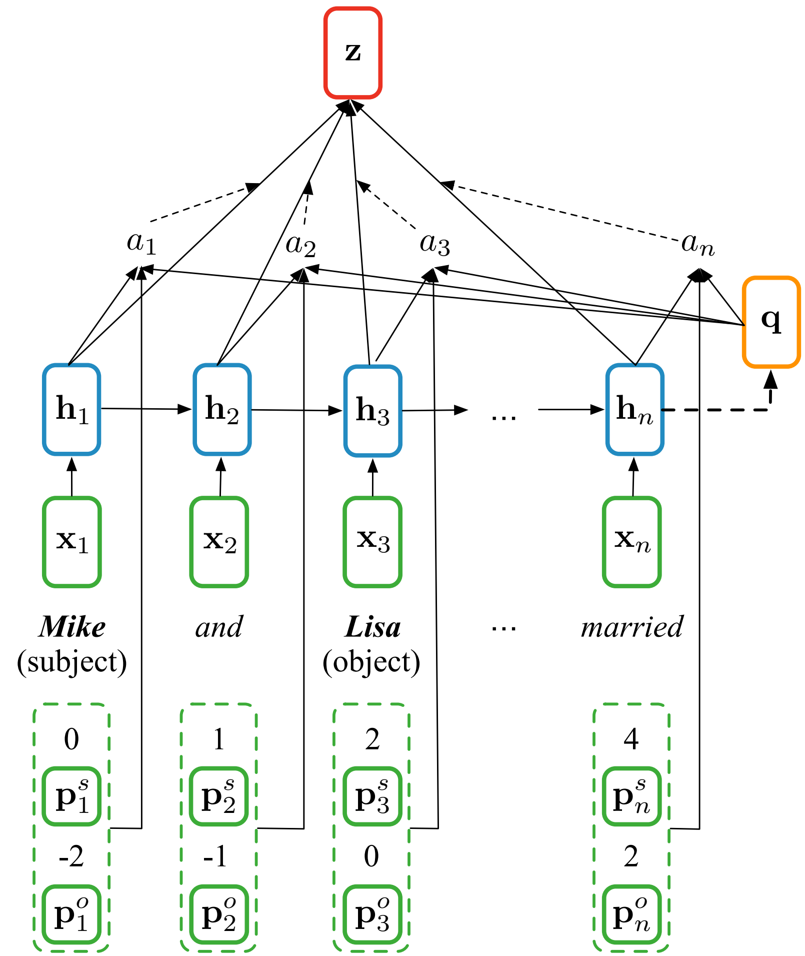
Position-aware Attention Model

Embedding Layers

Word: $\mathbf{x} = [\mathbf{x}_1, ..., \mathbf{x}_n]$

Position: $\mathbf{p}^s = [\mathbf{p}^s_1, ..., \mathbf{p}^s_n]$

$\mathbf{p}^o = [\mathbf{p}^o_1, ..., \mathbf{p}^o_n]$



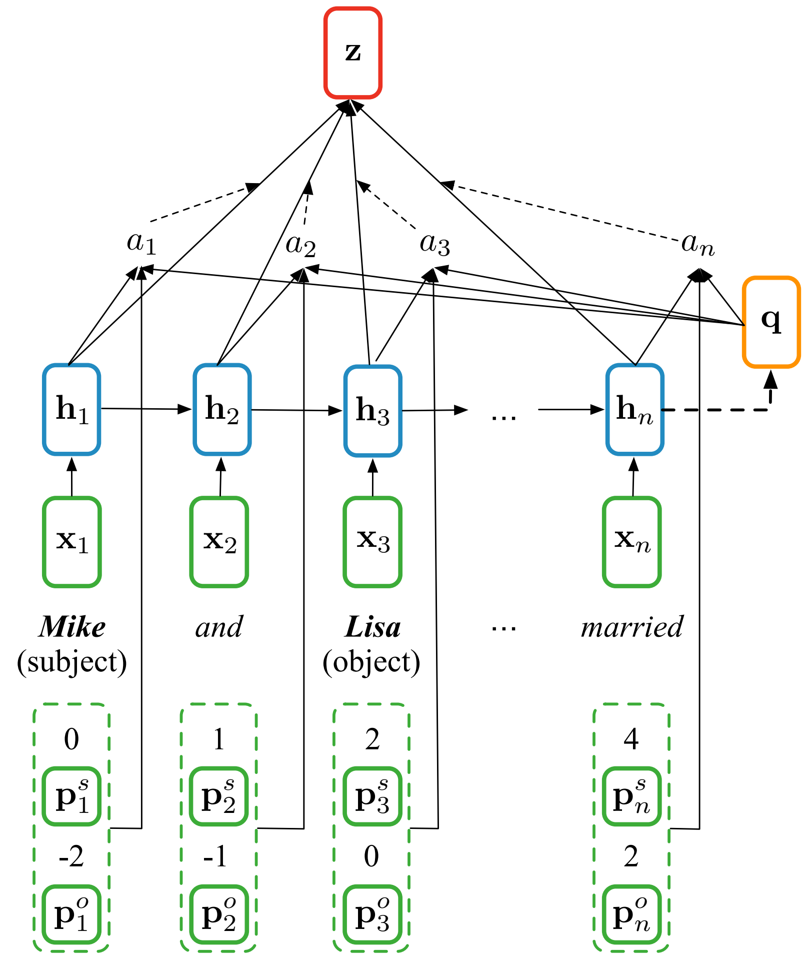
LSTM Layers

$$\{\mathbf{h}_1, ..., \mathbf{h}_n\} = \mathrm{LSTM}(\{\mathbf{x}_1, ..., \mathbf{x}_n\})$$

Position-aware Attention Model

Summary Vector

$$\mathbf{q} = \mathbf{h}_n$$



Position-aware Attention Model

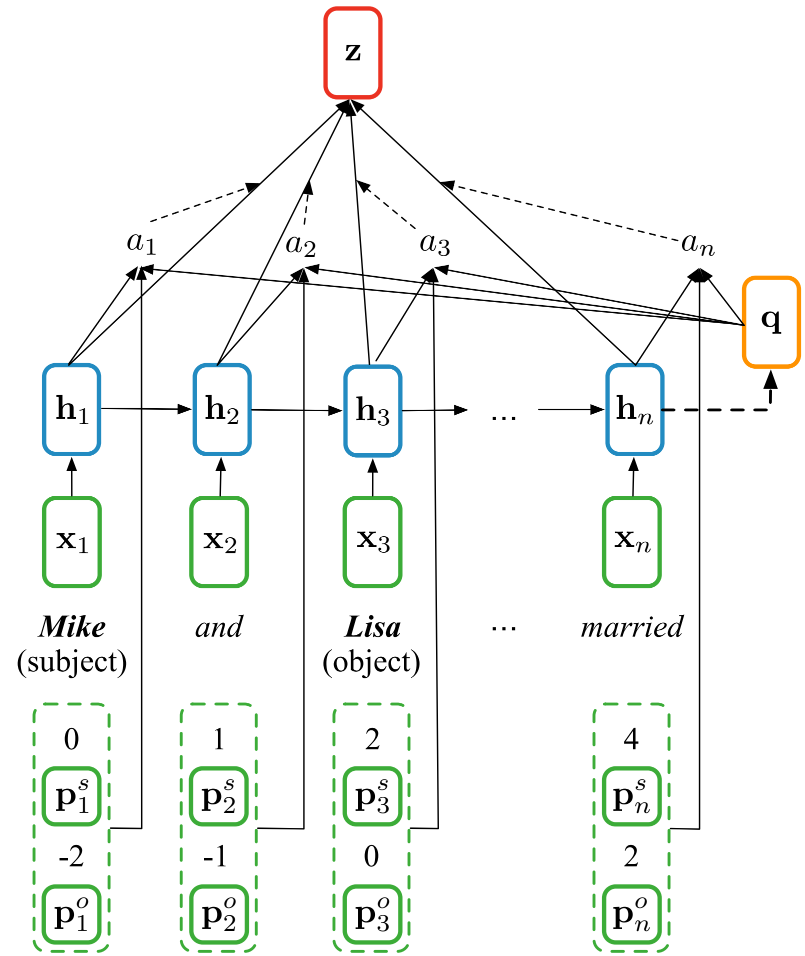
Summary Vector

$$\mathbf{q} = \mathbf{h}_n$$

Attention Layer

$$u_i = \mathbf{v}^\top \tanh(\mathbf{W}_h \mathbf{h}_i + \mathbf{W}_q \mathbf{q} + \mathbf{W}_s \mathbf{p}_i^s + \mathbf{W}_o \mathbf{p}_i^o)$$

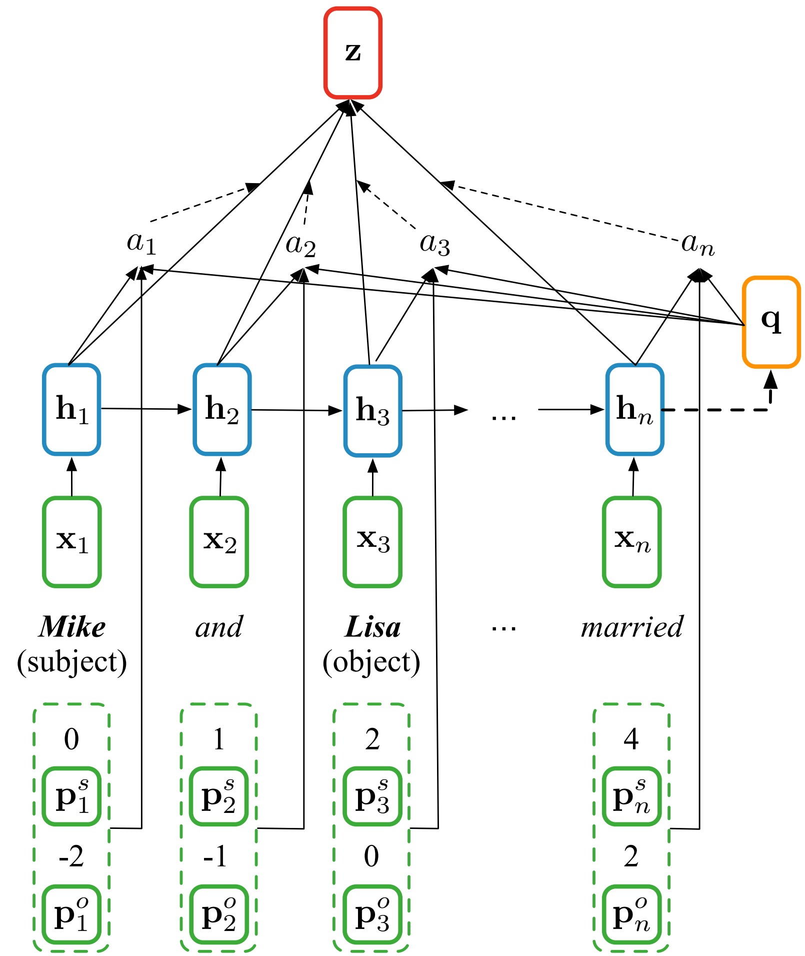
$$a_i = \frac{\exp(u_i)}{\sum_{j=1}^n \exp(u_j)}$$



Position-aware Attention Model

Relation Representation

$$\mathbf{z} = \sum_{i=1}^{n} a_i \mathbf{h}_i$$



Position-aware Attention Model

Relation Representation

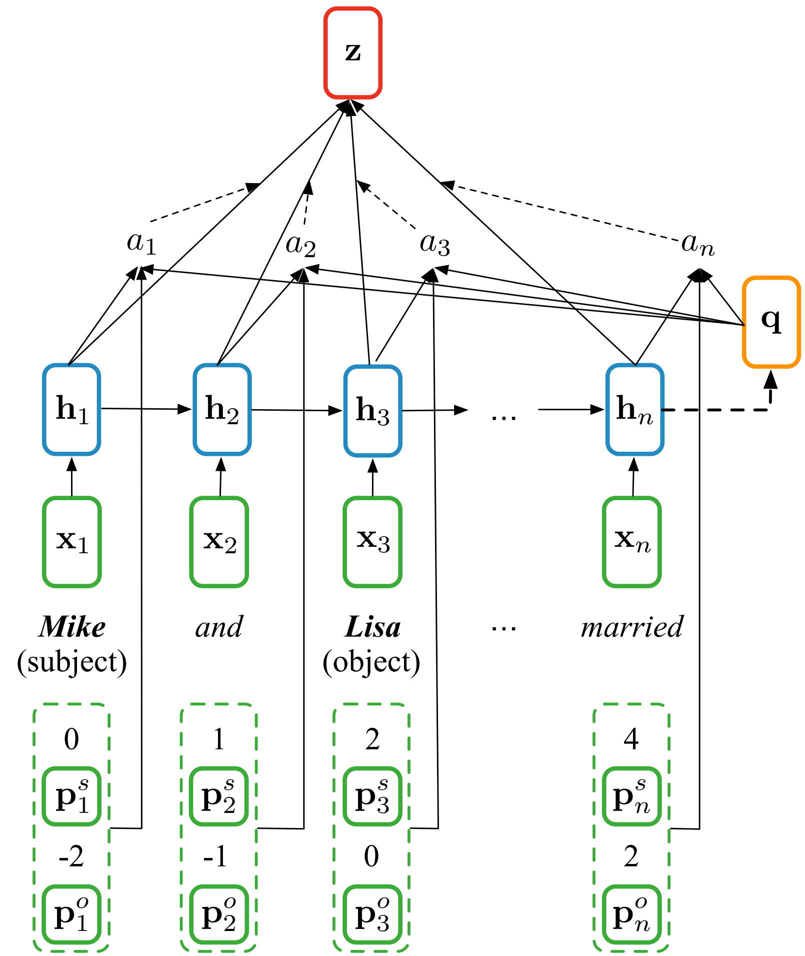
$$\mathbf{z} = \sum_{i=1}^{n} a_i \mathbf{h}_i$$

Softmax Layer

$$\mathbf{y} = \mathrm{softmax}(\mathbf{W}\mathbf{z})$$

$$c = \underset{i}{\mathrm{argmax}}(y_i);$$

$$c \in \{\textit{per:spouse}, \textit{org:founded}, \ldots\}$$



Other Augmentations

- **Word dropout**: balance OOV distribution

Penner is survived by his brother, John

Other Augmentations

- **Word dropout**: balance OOV distribution

Penner is <UNK> by his brother, John

Other Augmentations

- **Entity masking**: focus on relations, not specific entities

Penner is survived by his brother, John

Other Augmentations

- **Entity masking**: focus on relations, not specific entities

SUBJ-PER is survived by his brother, John

Other Augmentations

- **Entity masking**: focus on relations, not specific entities

SUBJ-PER is survived by his brother, OBJ-PER

Other Augmentations

- **Linguistic information**: POS and NER embeddings from Stanford CoreNLP

Penner is survived by …

Other Augmentations

- **Linguistic information**: POS and NER embeddings from Stanford CoreNLP

Penner	is	survived	by	…
NNP	VPZ	VBN	IN	…
PER	O	O	O	…

Other Augmentations

- **Linguistic information**: POS and NER embeddings from Stanford CoreNLP

Penner	is	survived	by	…
NNP	VPZ	VBN	IN	…
PER	O	O	O	…

Other Augmentations

- **Linguistic information**: POS and NER embeddings from Stanford CoreNLP

Penner	is	survived	by	…
NNP	VPZ	VBN	IN	…
PER	O	O	O	…

Try out these tricks!

Other Augmentations

- **Linguistic information**: POS and NER embeddings from Stanford CoreNLP

Penner	is	survived	by	...
NNP	VPZ	VBN	IN	...
PER	O	O	O	...

Try out these tricks!

Overview

Model: A new position-aware attention model

+

Data: A new supervised dataset, TACRED

=

Result: Improved slot filling

SemEval 2010: Popular But Suboptimal

- Small in size (10.7k)
- Different and vague relations

A New York Times food **writer** fries the potatoes in a **mixture** of peanut oil and duck fat with bacon added.

SemEval 2010: Popular But Suboptimal

- Small in size (10.7k)
- Different and vague relations

A New York Times food **writer** <u>fries</u> the potatoes <u>in a</u> **mixture** of peanut oil and duck fat with bacon added.	**Instrument-Agency**

The **TAC R**elation **E**xtraction **D**ataset

The **TAC R**elation **E**xtraction **D**ataset

- Crowdsourced
- Tailored for TAC KBP slot filling

- Four explicit goals
  - Large-scale
  - Real-world corpus
  - Negative examples
  - Fully supervised

Goal 1: Large-scale

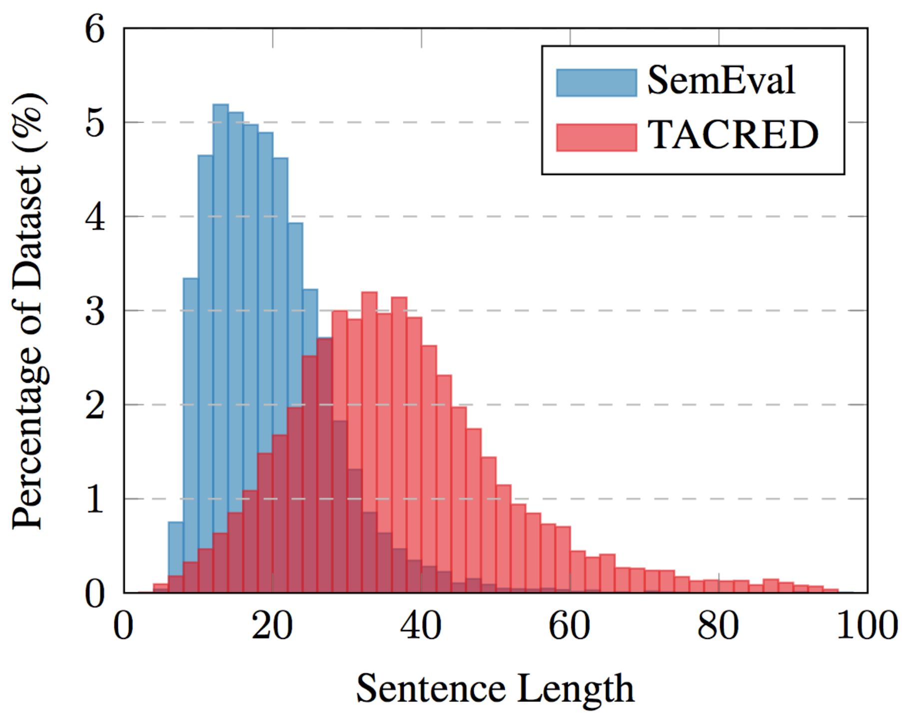
Split	**# examples**
Train	68,124
Dev	22,631
Test	15,509
Total	**106,264**

Goal 1: Large-scale

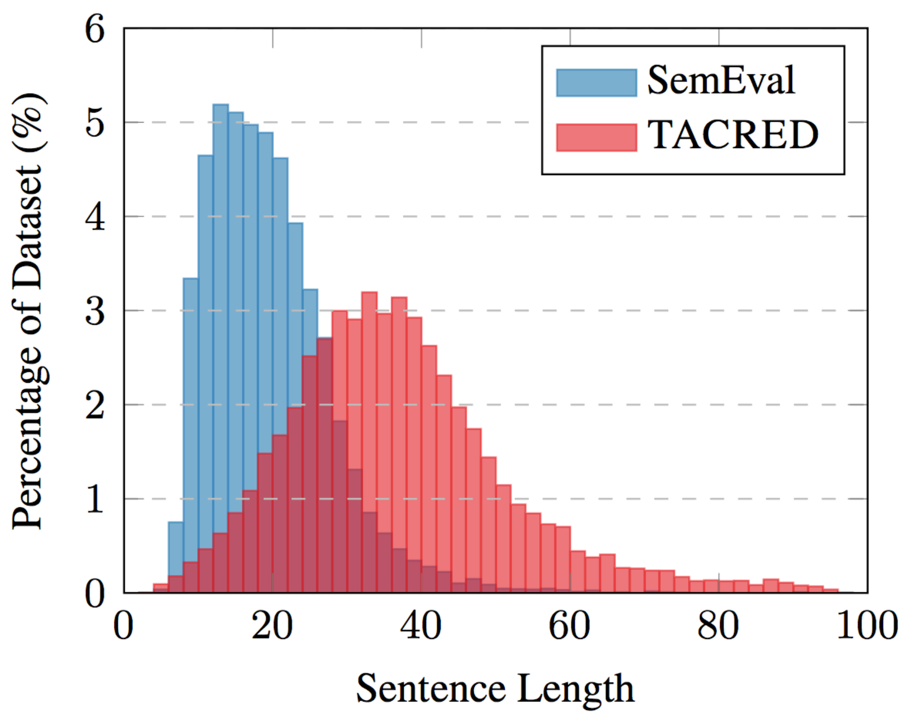
Split	# examples
Train	68,124
Dev	22,631
Test	15,509
Total	**106,264**

An order of magnitude larger!

Goal 2: Real-world TAC KBP corpus



Goal 2: Real-world TAC KBP corpus



Longer and more complex context!

Goal 3: Negative examples annotated

Label	Ratio (%)
Positive	20.5
Negative	79.5

Goal 3: Negative examples annotated

Label	Ratio (%)
Positive	20.5
Negative	79.5

Beat false positives in slot filling!

Goal 4: Fully supervised

Pandit worked at the brokerage Morgan Stanley for about 11 years until 2005, when he and some 106 Morgan Stanley colleagues quit and later founded the hedge fund **Old Lane Partners**.	**org:founded_by**

Goal 4: Fully supervised

Pandit worked at the brokerage Morgan Stanley for about 11 years until 2005, when he and some 106 Morgan Stanley colleagues quit and later founded the hedge fund **Old Lane Partners**.	**org:founded_by**

41 TAC KBP slot types + no_relation!

Overview

Model: A new position-aware attention model

+

Data: A new supervised dataset, TACRED

=

Result: Improved slot filling

Experiments

Experiments

- Task 1: Relation Extraction on TACRED

  Q: How well can we do on relation extraction?

  VS: Baseline traditional and neural models.

- Task 2: End-to-end TAC KBP Slot Filling Task

  Q: Does it improve slot filling?

  VS: SOTA slot filling system.

Models Compared Against

Non-Neural

- Stanford's TAC KBP 2015 winning system
  - **Patterns**
  - **Logistic regression (LR)**

Models Compared Against

Non-Neural

- Stanford's TAC KBP 2015 winning system
  - **Patterns**
  - **Logistic regression (LR)**

Neural

- **CNN** with positional encodings (Nguyen and Grishman, 2015)
- **Dependency-based RNN** (Xu et al., 2015)
- **LSTM**: 2-layer Stacked-LSTM

Relation Extraction Results

	Model	**P**	**R**	**F1**
Traditional	Patterns	**86.9**	23.2	36.6
	LR	73.5	49.9	59.4
	LR + Patterns	72.9	**51.8**	**60.5**

Relation Extraction Results

	Model	P	R	F1
Traditional	Patterns	**86.9**	23.2	36.6
	LR	73.5	49.9	59.4
	LR + Patterns	72.9	**51.8**	**60.5**

- Patterns: high precision
- LR: high recall

Relation Extraction Results

	Model	**P**	**R**	**F1**
Traditional	LR + Patterns	72.9	51.8	60.5
Neural	CNN	**75.6**	47.5	58.3
	CNN-PE	70.3	54.2	61.2
	SDP-LSTM	66.3	52.7	58.7
	LSTM	65.7	**59.9**	**62.7**

Relation Extraction Results

	Model	P	R	F1
Traditional	LR + Patterns	72.9	51.8	60.5
Neural	CNN	**75.6**	47.5	58.3
	CNN-PE	70.3	54.2	61.2
	SDP-LSTM	66.3	52.7	58.7
	LSTM	65.7	**59.9**	**62.7**

- CNN higher precision; LSTM higher recall
- CNN-PE and LSTM outperform traditional

Relation Extraction Results

	Model	**P**	**R**	**F1**
Traditional	LR + Patterns	**72.9**	51.8	60.5
Neural	LSTM	65.7	59.9	62.7
	Our model	65.7	**64.5**	**65.1**
	Ensemble (5)	70.1	**64.6**	**67.2**

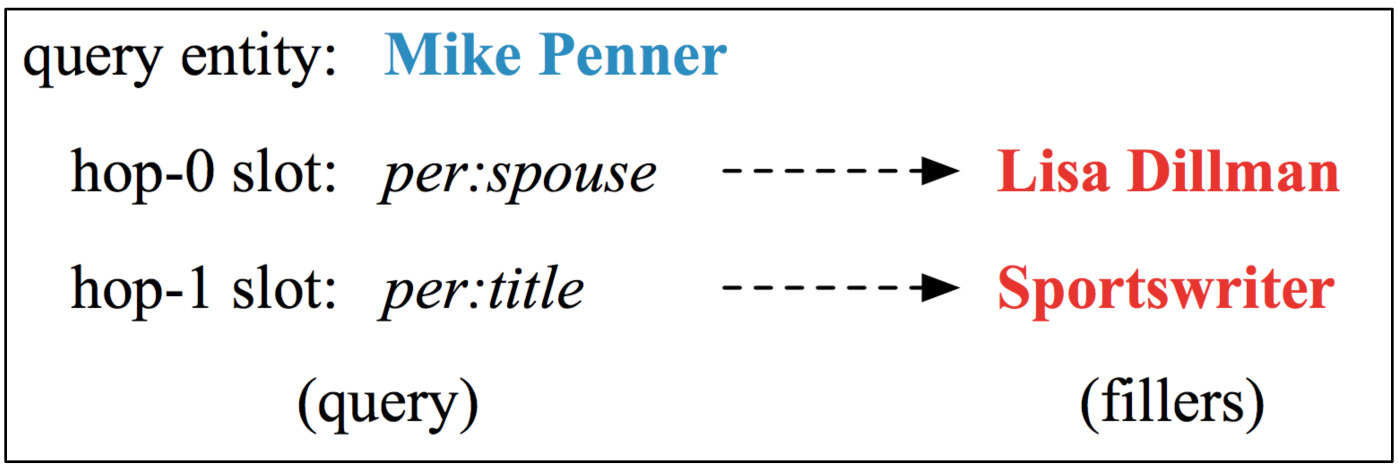
Relation Extraction Results

	Model	**P**	**R**	**F1**
Traditional	LR + Patterns	**72.9**	51.8	60.5
Neural	LSTM	65.7	59.9	62.7
	Our model	65.7	**64.5**	**65.1**
	Ensemble (5)	70.1	**64.6**	**67.2**

- Our model: +2.4 improvement on F1

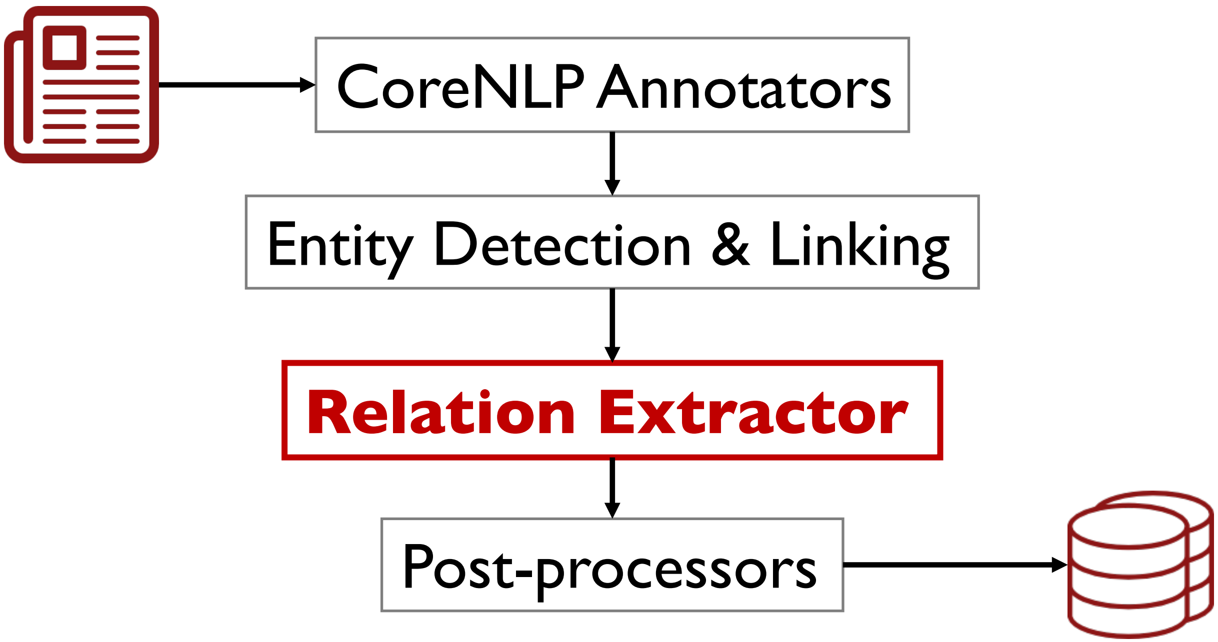
Slot Filling Evaluation

- **Input**: 50k docs + 2-hop queries
- **Output**: slot fillers



Slot Filling Evaluation

- Stanford's 2015 winning system + new extractor



Slot Filling Results

	Hop-0			Hop-all		
Model	**P**	**R**	**F1**	**P**	**R**	**F1**
Patterns	**63.8**	17.7	27.7	**58.9**	13.3	21.8
LR	36.6	21.9	27.4	25.6	16.3	19.9
+ Patterns (2015 winning)	37.5	24.5	29.7	26.6	19.0	**22.2**
LR trained on TACRED	32.7	20.6	25.3	16.8	15.3	16.0
+ Patterns	36.5	**26.5**	**30.7**	20.1	**21.2**	20.6

Slot Filling Results

	Hop-0			Hop-all		
Model	**P**	**R**	**F1**	**P**	**R**	**F1**
Patterns	**63.8**	17.7	27.7	**58.9**	13.3	21.8
LR	36.6	21.9	27.4	25.6	16.3	19.9
+ Patterns (2015 winning)	37.5	24.5	29.7	26.6	19.0	**22.2**
LR trained on TACRED	32.7	20.6	25.3	16.8	15.3	16.0
+ Patterns	36.5	**26.5**	**30.7**	20.1	**21.2**	20.6

- Close results when trained on TACRED only

Slot Filling Results

	Hop-0			Hop-all		
Model	**P**	**R**	**F1**	**P**	**R**	**F1**
LR + Patterns (2015 winning)	37.5	24.5	29.7	26.6	19.0	22.2
LR trained on TACRED	32.7	20.6	25.3	16.8	15.3	16.0
+ Patterns	36.5	26.5	30.7	20.1	21.2	20.6
Our model	39.0	28.9	33.2	28.2	21.5	24.4
+ Patterns	**40.2**	**31.5**	**35.3**	**29.7**	**24.2**	**26.7**

Slot Filling Results

	Hop-0			Hop-all		
Model	**P**	**R**	**F1**	**P**	**R**	**F1**
LR + Patterns (2015 winning)	37.5	24.5	29.7	26.6	19.0	22.2
LR trained on TACRED	32.7	20.6	25.3	16.8	15.3	16.0
+ Patterns	36.5	26.5	30.7	20.1	21.2	20.6
Our model	39.0	28.9	33.2	28.2	21.5	24.4
+ Patterns	**40.2**	**31.5**	**35.3**	**29.7**	**24.2**	**26.7**

- Neural vs LR: **+7.9** hop-0, **+8.4** hop-all!
- Best vs 2015 winning: **+5.6** hop-0, **+4.5** hop-all!

Further Analysis

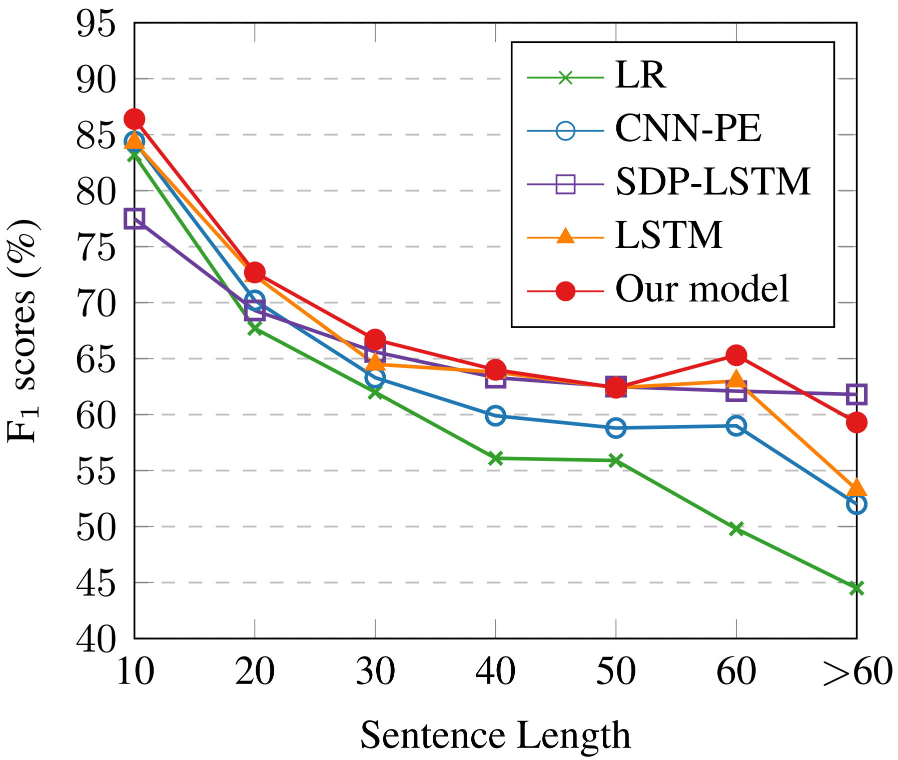
- Model Ablation

Model	Dev F1
Our model (single)	**66.0**
- Position-aware attention	65.2
- All attention	64.1
- Word dropout	65.4
- All Above	62.2

Attention plays an important role!

Further Analysis

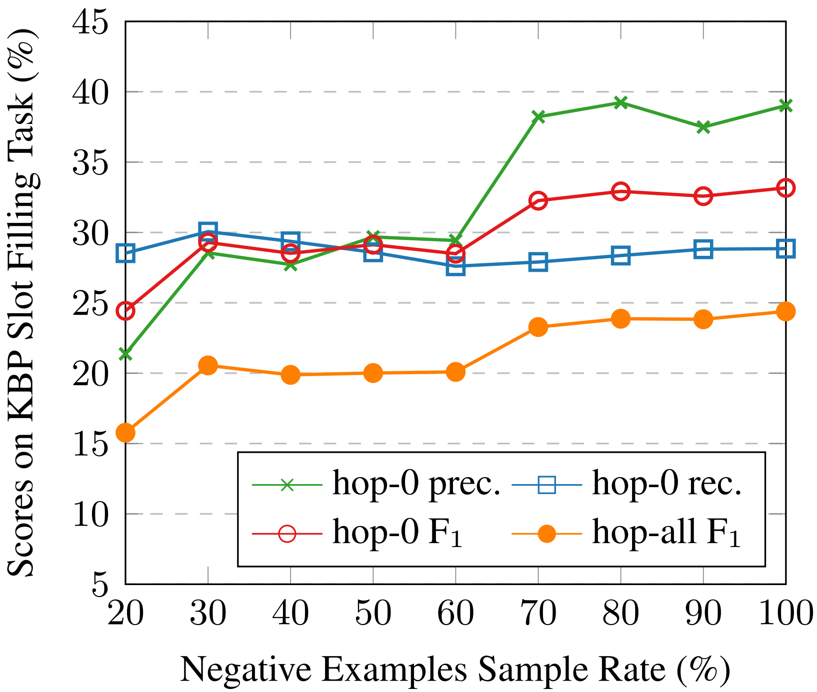
- Performance by sentence length



- LR and CNN drops drastically
- SDP-LSTM least sensitive to lengths
- Our model achieves best performance

Further Analysis

- Impact of negative examples



- Hop-0 precision increases
- Hop-0 recall stays
- Hop-0 and hop-all F1 increases

Further Analysis

- Attention visualization

PER-SUBJ graduated from North Korea 's elite Kim Il Sung University and ORG-OBJ ORG-OBJ .

per:schools_attended

The cause was a heart attack following a case of pneumonia , said PER-SUBJ 's niece , PER-OBJ PER-OBJ .

per:other_family

Independent ORG-SUBJ ORG-SUBJ ORG-SUBJ (ECC) chairman PER-OBJ PER-OBJ refused to name the three , saying they would be identified when the final list of candidates for the august 20 polls is published on Friday .

org:top_members/employees

Summary

Model: A new position-aware attention model

+

Data: A new supervised dataset, TACRED

=

Result: Improved slot filling results

Availability

- Code will be available soon at: https://github.com/yuhaozhang/tacred-relation
- The TACRED dataset will be made publicly available via LDC.
- Let us know how you use TACRED!

Availability

- Code will be available soon at: https://github.com/yuhaozhang/tacred-relation
- The TACRED dataset will be made publicly available via LDC.
- Let us know how you use TACRED!

Thank you! Questions?