Jointly Extracting Event Triggers and Arguments by Dependency-Bridge RNN and Tensor-Based Argument Interaction

Feng Qian*, Lei Sha*, Baobao Chang, Zhifang Sui

Institute of Computational Linguistics, Peking University
{*nickqian, shalei, chbb, szf*}*@pku.edu.cn*

November 29, 2017

Table of Contents

Table of Contents

Introduction

- Event extraction is important for knowledge acquisition from large amounts of news text.
- The result of event extraction can be used to construct knowledge base, which can be applied to question answering, dialogue system, etc.
- Its paradigm is ubiquitous in our daily life:
  - Knowledge Graph
  - Structured summary of search engine
  - Wikipedia infobox

Applications of Event Extraction

The Google search result of *September 11 attacks*:

September 11 attacks

The September 11 attacks were a series of four coordinated terrorist attacks by the Islamic terrorist group al-Qaeda on the United States on the morning of Tuesday, September 11, 2001. Wikipedia

Date: September 11, 2001

Perpetrator: Al-Qaeda

Total number of deaths: 2,997 (2,978 victims + 19 hijackers)

Locations: New York City, Arlington County, Stonycreek Township

Attack types: Aircraft hijacking, Mass murder, Suicide attack

Applications of Event Extraction

The Wikipedia infobox of *September 11 attacks*:

September 11 attacks

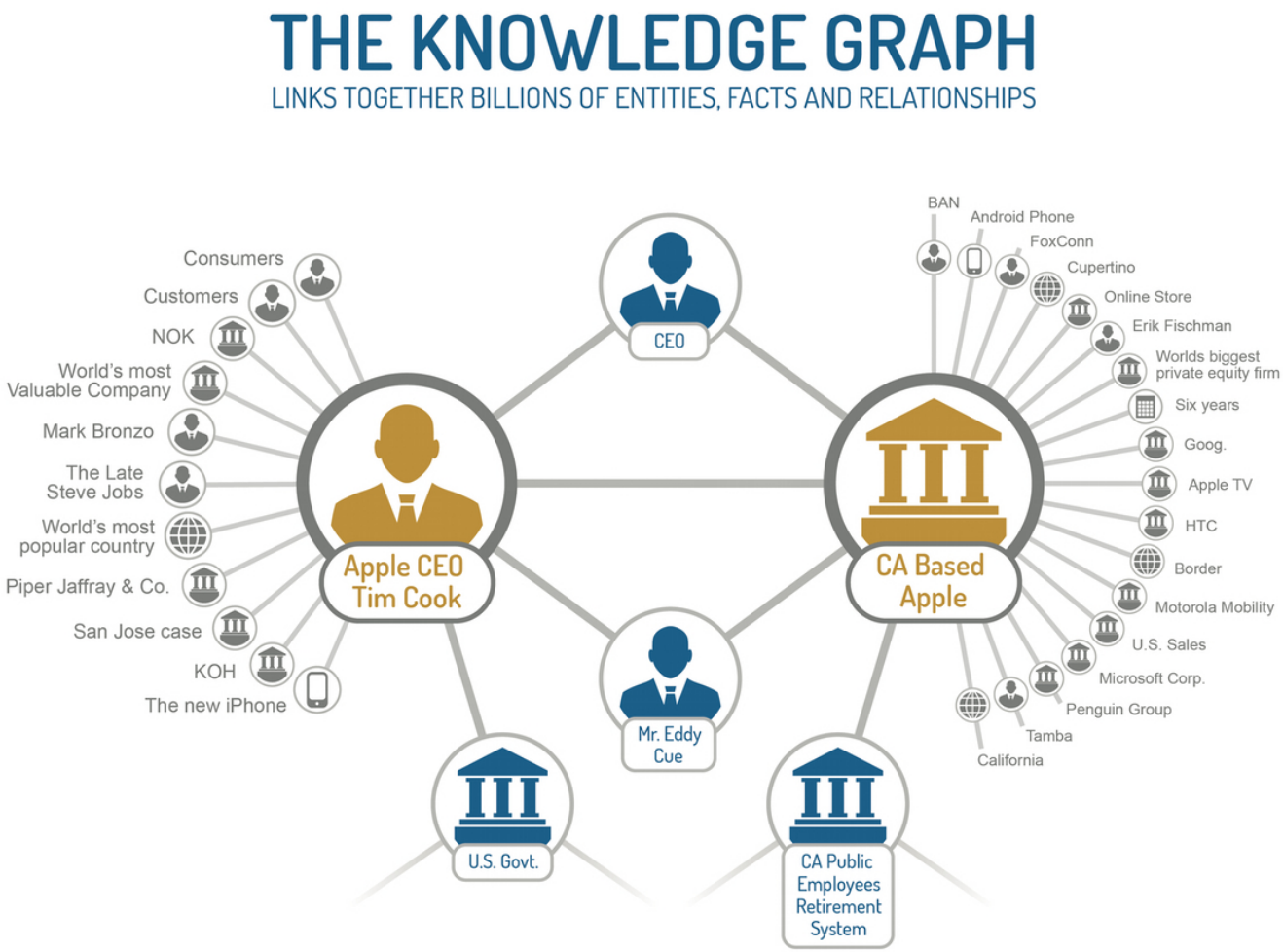
Part of Terrorism in the United States

Top row: The Twin Towers of the World Trade Center burning
2nd row, left to right: Collapsed section of the Pentagon; Flight 175 crashes into 2 WTC;
3rd row, left to right: A firefighter requests assistance at World Trade Center site; Ground Zero; An engine from Flight 93 is recovered
Bottom row: Flight 77's collision with the Pentagon as captured by three consecutive CCTV frames

Location	New York City, New York, U.S.; Arlington County, Virginia, U.S.; Stonycreek Township near Shanksville, Pennsylvania, U.S.
Date	September 11, 2001; 15 years ago 8:46 a.m. – 10:28 a.m. (EDT)
Target	World Trade Center (AA11 and UA 175) The Pentagon (AA77) White House or U.S. Capitol (UA 93; failed)
Attack type	Aircraft hijackings Suicide attacks Mass murder Terrorism
Deaths	2,997 (2,978 victims + 19 hijackers)
Non-fatal injuries	6,000+
Perpetrators	Al-Qaeda[1] (see also responsibility and hijackers)
No. of participants	19

Applications of Event Extraction

- The extracted events can be transferred into triples and store in the knowledge graphs.
- The knowledge graphs can be leveraged by upper applications.



Event Extraction

What's an event?

Event Type:	Business	
Trigger	Release	
Argument	Company	Microsoft
	Product	Surface Pro
	Place	USA

Figure: Microsoft releases surface Pro in USA.

Event Extraction

What's an event?

Event Type:		Attack
Trigger	Crash	
Argument	Attacker	Five hijackers
	Target	World Trade Center's North Tower
	Instrument	American Airlines Flight 11
	Time	September 11th

Figure: On September 11th, five hijackers crashed American Airlines Flight 11 into the World Trade Center's North Tower.

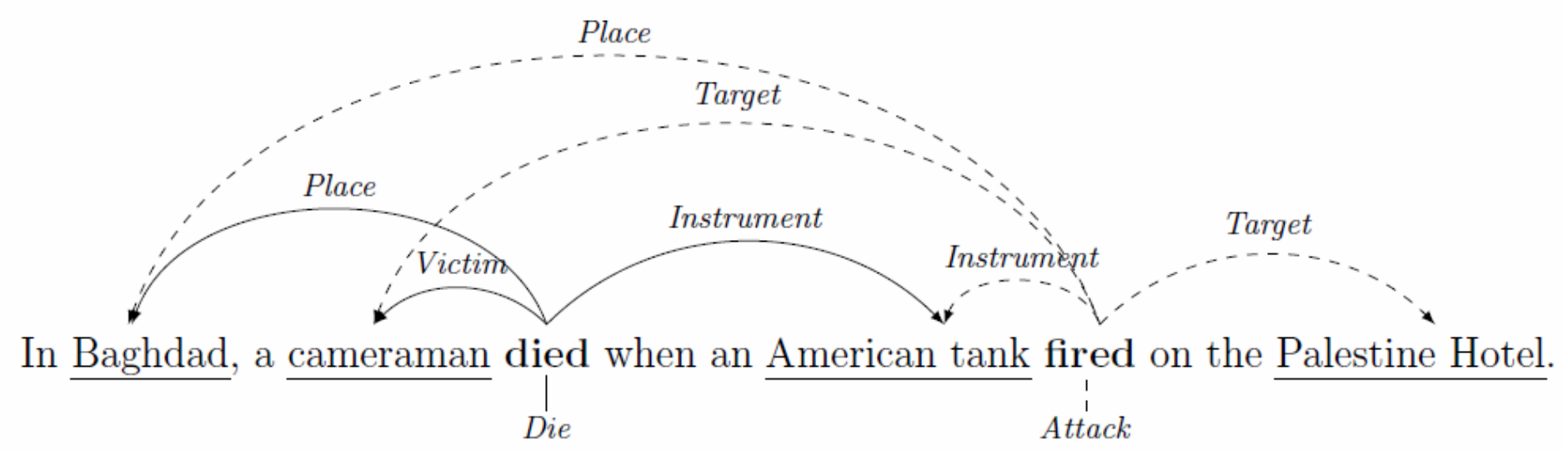
Event Extraction from News Text

What should we do?

- Extract trigger
- Identify arguments
- Classify roles

Event Type: Die

Trigger	Die	
Argument	Victim	cameraman
	Place	Baghdad
	Instrument	American tank



Argument 1	Argument 2	Argument 3
Role:Place	Role:Victim	Role:Instrument

Table of Contents

Motivation

Challenges of event extraction by the previous solutions

- $\surd$ Using syntax information as feature
- $\times$ Using syntax information as architecture
- $\surd$ Capture two kinds of argument-argument relationship (Pos & Neg)
- $\times$ Capture large amount of argument-argument relationship

Table of Contents

Event Extraction from News Text

Motivation 1:

- Dependency relation $\longrightarrow$ Dependency bridge
- According to definition of dependency relation, dependency edges usually contain some information about temporal, consequence, conditional or purpose.

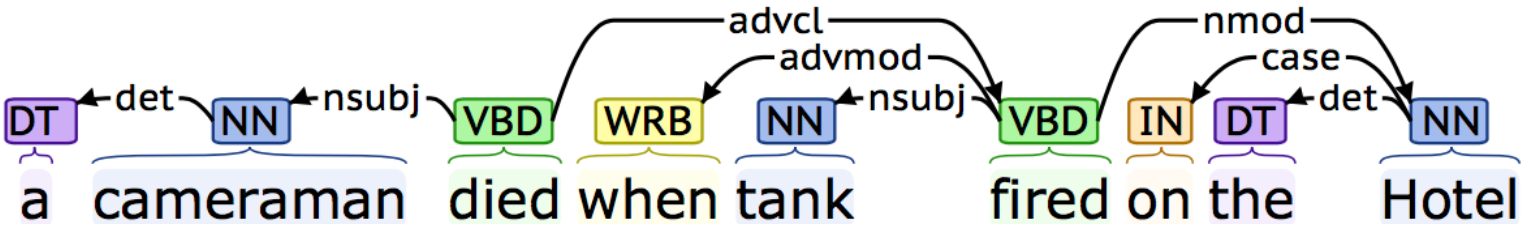


Figure: Example of dependency parse tree.

Event Extraction from News Text

- We add dependency bridges to conventional LSTM-RNN architecture.
- Bidirectionality:
  - Forward: Set all dependency bridges as forward.
  - Backward: Set all dependency bridges as backward.

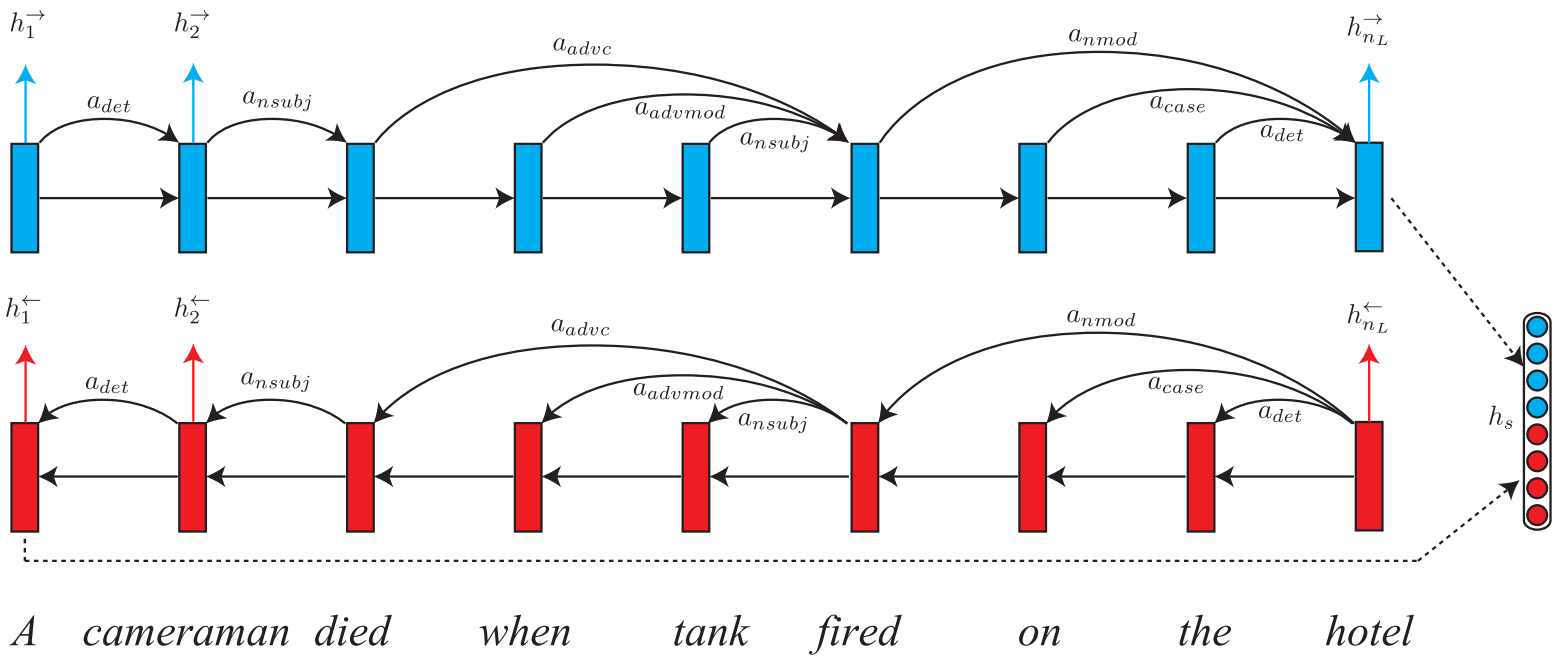


Figure: Dependency bridge on LSTM. Apart from the last LSTM cell, each cell also receives information from former syntactically related cells.

Event Extraction from News Text

Details of dependency bridge

- We add a new gate d_t and change the calculation of hidden state.
- $h_t = o_t \odot \tanh(c_t) + d_t \odot \left(\frac{1}{|S_{in}|} \sum_{(i,p) \in S_{in}} a_p h_i \right)$

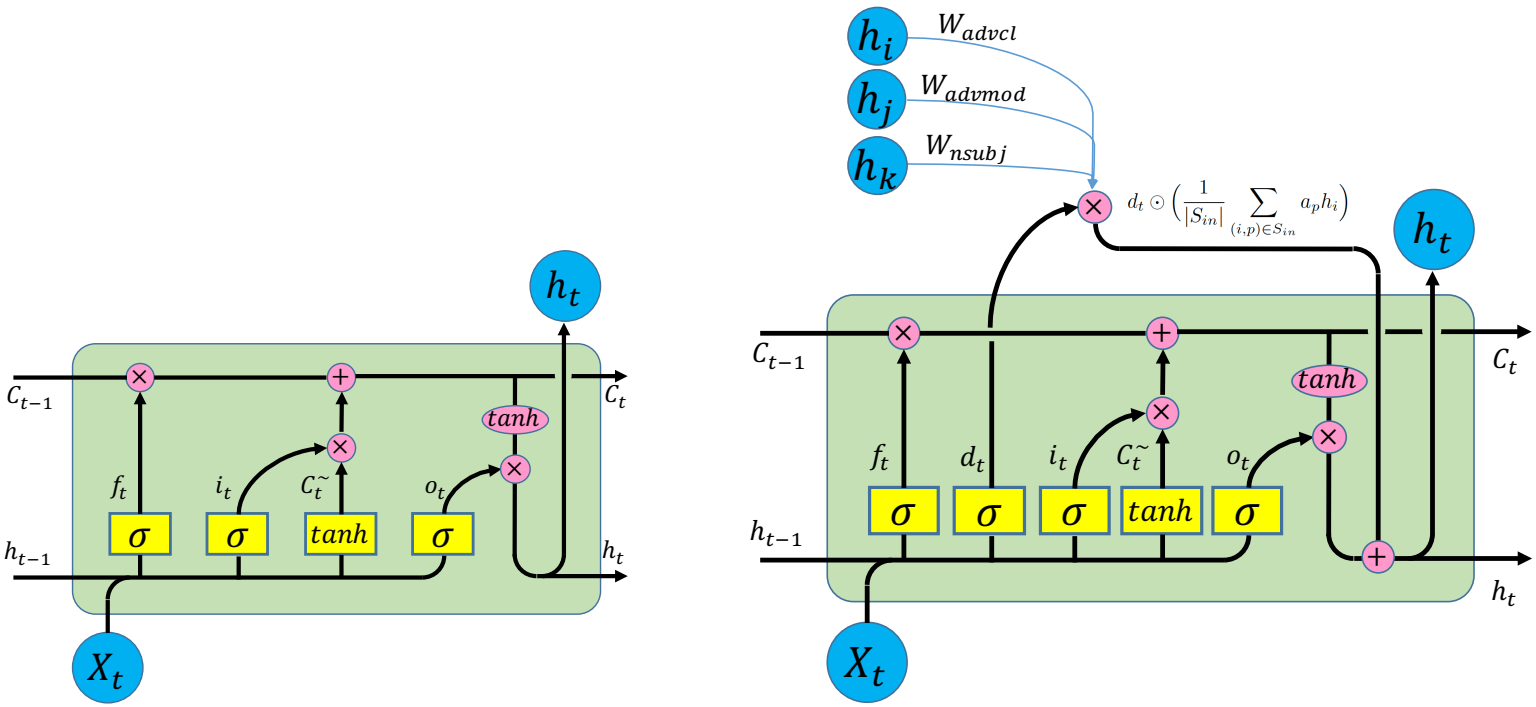


Figure: The calculation detail of dependency bridge.

Table of Contents

Event Extraction from News Text

Motivation 2:

- We represent each arg-arg relationship by a vector
- We use a tensor to represent all kinds of arg-arg relationships in a sentence

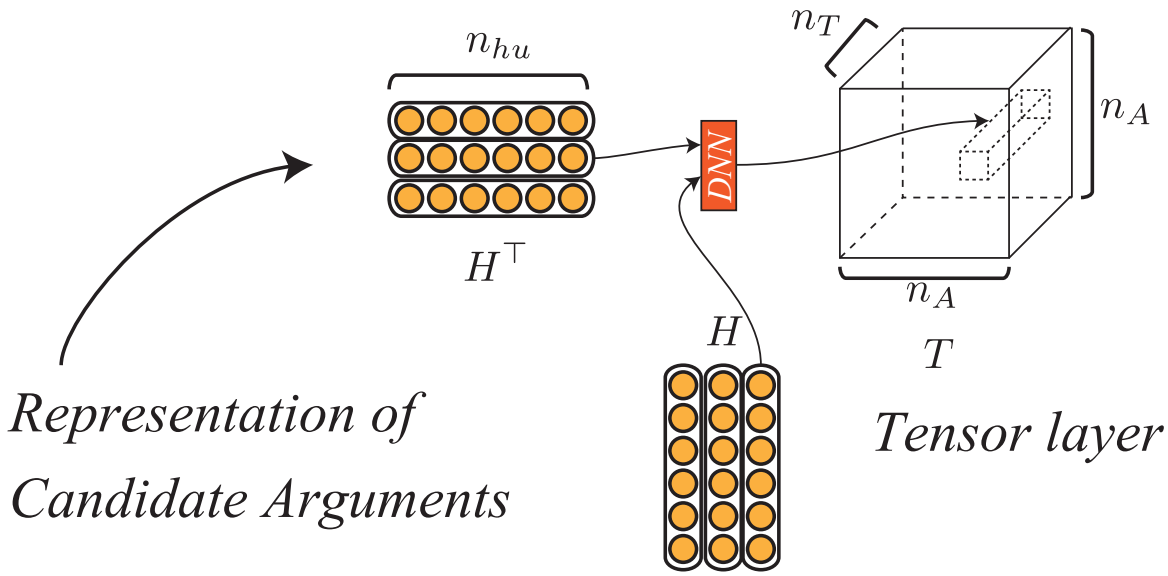


Figure: The calculation detail of tensor layer.

Event Extraction from News Text

The whole architecture ...

- Tensor layer is applied to the hidden layer of the dependency bridge RNN
- Then we apply max-pooling over arguments to find the most important "interactive features" for the arguments

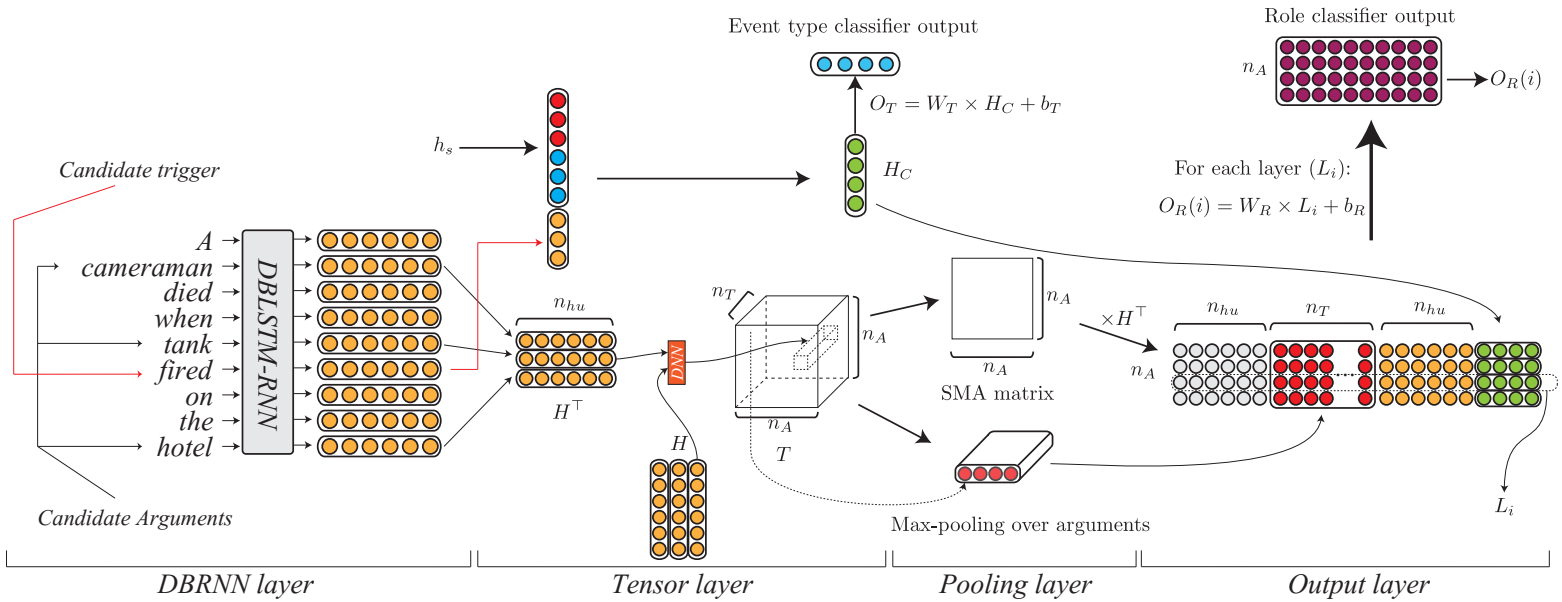


Table of Contents

Weights of each dependency relation

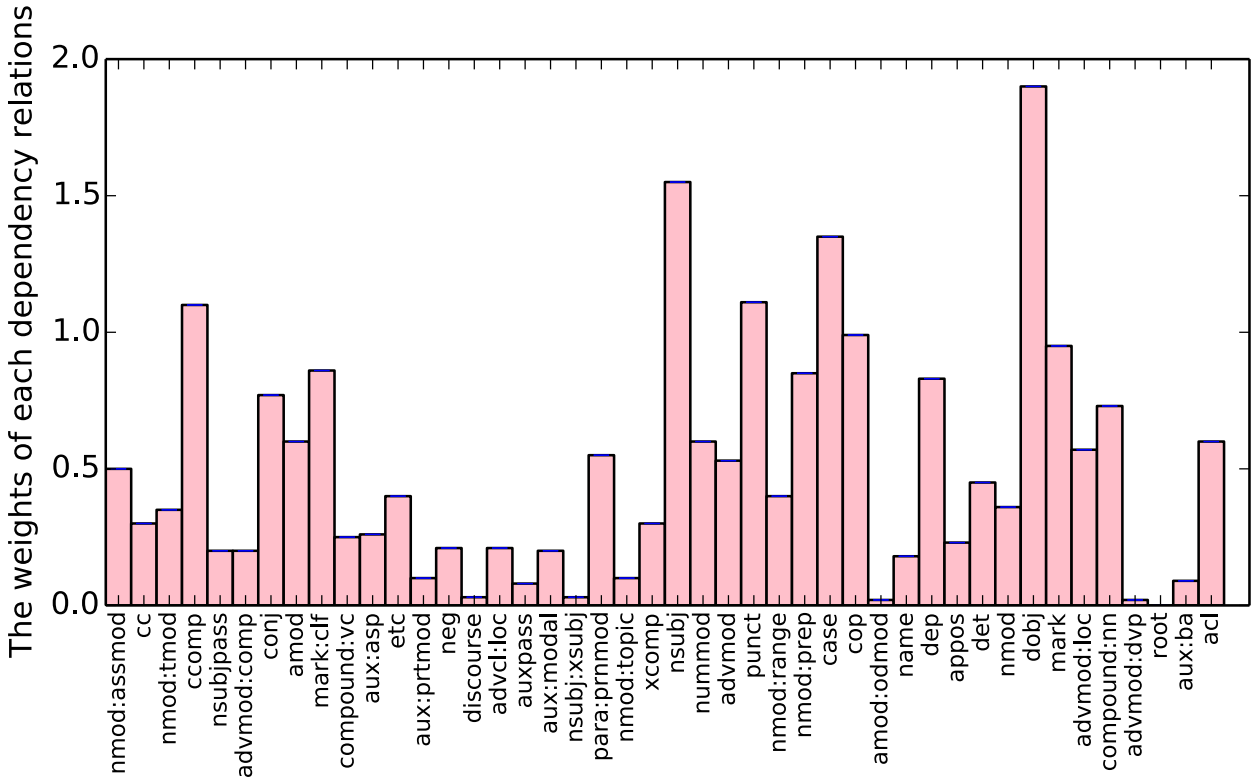


Figure: The visualization of trained weights of each dependency relations.

Overall performance

Method	Trigger Identification +Classification (%)			Argument Identification (%)			Argument Role (%)		
	P	R	F_1	P	R	F_1	P	R	F_1
Cross-Event	68.7	68.9	68.8	50.9	49.7	50.3	45.1	44.1	44.6
Cross-Entity	72.9	64.3	68.3	53.4	52.9	53.1	51.6	45.5	48.3
JointBeam	73.7	62.3	67.5	69.8	47.9	56.8	**64.7**	44.4	52.7
DMCNN	**75.6**	63.6	69.1	68.8	51.9	59.1	62.2	46.9	53.5
RBPB	70.3	67.5	68.9	63.2	59.4	61.2	54.1	53.5	53.8
JRNN	66.0	**73.0**	69.3	61.4	**64.2**	62.8	54.2	**56.7**	55.4
dbRTN	74.1	69.8	**71.9**	**78.3**	54.7	**64.4**	64.2	51.5	**57.2**

Figure: Performances of various approaches on ACE 2005 dataset.

Ablation tests – dependency bridge

- Binary DB: The weight of DB belongs to $0, 1$
- Typed DB: The weight of DB can be any float numbers

Method	Trigger id+cl	Argument id	Argument id+cl
Our model without DB	69.0	62.7	54.6
+ binary DB	71.2	63.9	56.8
+ typed DB (full)	71.9	64.4	57.2

Table: Comparison after adding dependency bridges (DB). The numbers are F_1 scores. We compare with two baselines: no dependency bridges considered and only binary dependency bridges.

Ablation tests – Tensor layer

- dbRNN-SMA: only cast SMA away from the whole model
- dbRNN-MP: means cast the max-pooling feature matrix away
- dbRNN-TL: dbRNN without tensor layer
- dbRNN full model

	Method	1/1	1/*N*	All
Argument Identification	dbRNN-SMA	59.5	67.0	64.1
	dbRNN-MP	59.7	64.8	62.0
	dbRNN-TL	59.6	55.8	58.2
	dbRNN	59.9	69.5	67.7
Argument Role Classification	dbRNN-SMA	54.6	56.5	56.0
	dbRNN-MP	54.7	55.8	55.2
	dbRNN-TL	54.9	52.3	53.1
	dbRNN	54.6	60.9	58.7

Table: Comparison between different models. Here, we report the argument performance since the tensor layer is only applied to argument extraction.

Table of Contents

Conclusion

In this paper:

- We propose to add dependency bridges to sequential architecture
- We propose to add tensor layer for capturing various of argument relationships
- The weights of dependency bridges after training illuminates the importance of each dependency type in event extraction task
- The full model achieves high performance in all the three evaluation metrics, trigger classification, argument identification and role classification

Thank you. Any questions?