Learning and Generating from Structured Data

Lei Sha

Institute of Computational Linguistics, Peking University
shalei@pku.edu.cn

May 28, 2018

Table of Contents

Introduction

Structured data can be used in two symmetrical procedures:

- Text $\longrightarrow$ Structured data
- Structured data $\longrightarrow$ Text

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

↓ Learning Structured Data from Raw Text

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

↓ Generating Text from Structured Data

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

Table of Contents

Learning Structured Data from Raw Text

Structured data usually serve for applications like question answering, dialogues, and information retrieval

The tasks of learning structured data includes:

- Relation extraction
- Event extraction

We focused on event extraction in our research.

Event Extraction

Motivation:

- 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*:

Applications of Event Extraction

The Wikipedia infobox of *September 11 attacks*:

September 11 attacks

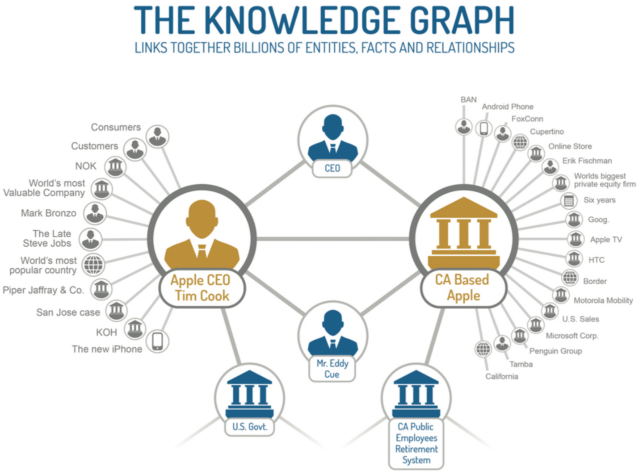
Part of Terrorism in the United States

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.

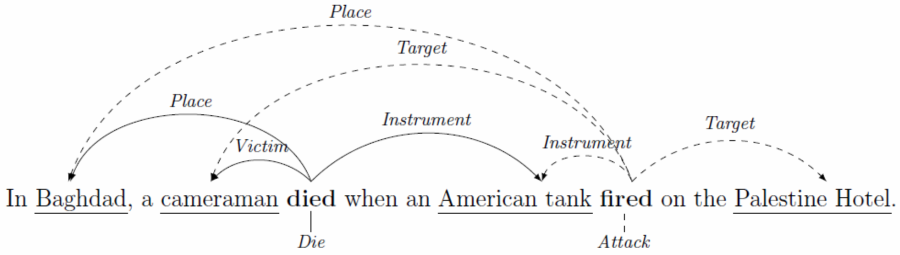
Event Extraction from News Text

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

Event Extraction from News Text

Challenges of event extraction:

- Patterns are accurate (precision $> 96\%$), but cannot cover every case

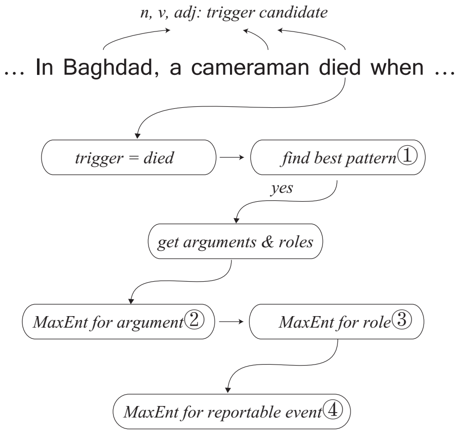
Pattern example

(weapon) tore [through] (building) at (place) $\Rightarrow$ Attack$\{$Roles$\cdots\}$

- Some arguments tend to occur together: Cameraman & American tank (in `Die` event)
- Some arguments tend not to occur together: Baghdad & Palestine Hotel (in `Die` event) (Hint: dependency distance is too long)

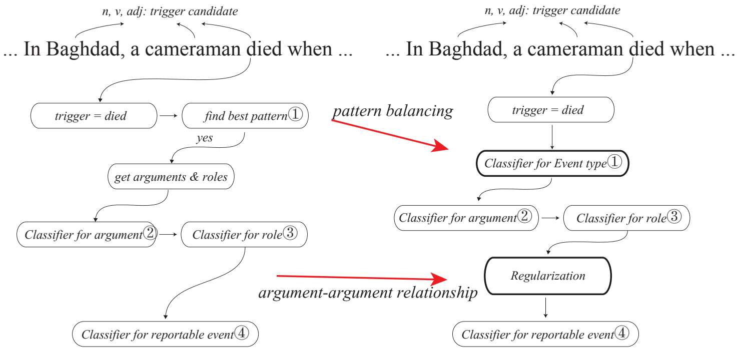
Event Extraction from News Text

Conventional pattern-based and classifier-based event extraction method



Event Extraction from News Text

Sha et al.: RBPB: Regularization-Based Pattern Balancing Method for Event Extraction. (ACL 2016)



- We transfer pattern information into embedded features
- We use regularization method to capture argument-argument relationships

Event Extraction from News Text

How to use pattern better?

- Turn pattern into event type probability distribution
- pattern + other features → trigger identification & classification

Event type	Trigger	Pattern
Attack	shoot	pattern 1
Injure	shoot	pattern 2
Die	shoot	pattern 3
Injure	shoot	pattern 4
Injure	shoot	pattern 5
Die	shoot	pattern 6
Die	shoot	pattern 7
Attack	shoot	pattern 8
Attack	shoot	pattern 9
Attack	shoot	pattern 10

$\Rightarrow$

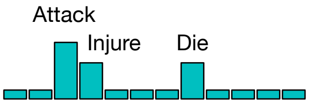


Table: Candidate pattern result for trigger "shoot"

Event Extraction from News Text

Argument-argument relationship regularization

- Arguments tend to occur together: Positive relationship
- Arguments tend not to occur together: Negative relationship

We need to capture these two relationships:

- Use a bunch of position features and dependency features to decide the probability to be "Pos": $P(rel = \text{"Pos"} | arg_i, arg_j)$
- The training set of arg-arg relationship classifier is generated from the original dataset
- We obtain the arg-arg relationship matrix M
- $M_{i,j} = P(rel = \text{"Pos"} | arg_i, arg_j)$

Event Extraction from News Text

We use a score function to evaluate the current configuration of argument: x

- $x(i)$ represents the role taken by i-th candidate argument
- $x_{\text{bin}} = Bin(x)$
- $x_{\text{bin}}(i)$ represents whether i-th candidate argument is an argument for the current trigger
- $Score(x) = x_{\text{bin}}^{\top} M x_{\text{bin}} + f_{\text{arg}}(x_{\text{bin}}) + f_{\text{role}}(x)$
    - f_{arg}: function for argument identification
    - f_{role}: function for role classification

We use Beam Search to find the best configuration (largest score)

Event Extraction from News Text

Example of the quadratic item:

- Assume that "Baghdad" and "Palestine hotel" have negative relationship
- In this example, the closer the value is to 1, the more likely it is positive; the closer the value is to 0, the more likely it is negative

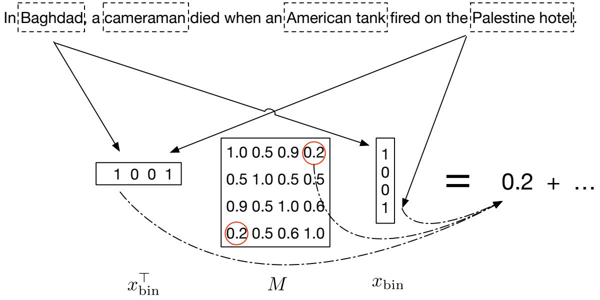


Figure: Example calculation process of negative relation.

Event Extraction from News Text

Another example of the quadratic item:

- Assume that “Baghdad” and “American tank” have positive relationship
- Again, the closer the value is to 1, the more likely it is positive; the closer the value is to 0, the more likely it is negative

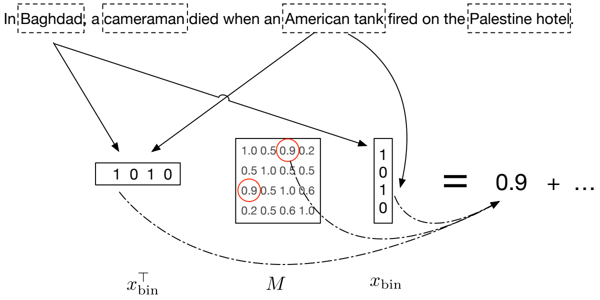


Figure: Example calculation process of positive relation.

Event Extraction from News Text

Argument-argument relationship regularization

- Pos relation vs. Neg relation
- Bilinear in score: $Score(x) = x_{\text{bin}}^{\top} M x_{\text{bin}} + f_{\text{arg}}(x_{\text{bin}}) + f_{\text{role}}(x)$

Event Extraction from News Text

Argument-argument relationship regularization

- Pos relation vs. Neg relation
- Bilinear in score: $Score(x) = x_{\text{bin}}^{\top} M x_{\text{bin}} + f_{\text{arg}}(x_{\text{bin}}) + f_{\text{role}}(x)$

We found that...

Event Extraction from News Text

Argument-argument relationship regularization

- Pos relation vs. Neg relation
- Bilinear in score: $Score(x) = x_{\text{bin}}^{\top} M x_{\text{bin}} + f_{\text{arg}}(x_{\text{bin}}) + f_{\text{role}}(x)$

We found that... Oops! That doesn't work.

Event Extraction from News Text

Argument-argument relationship regularization

- Pos relation vs. Neg relation
- Bilinear in score: $Score(x) = x_{\text{bin}}^{\top} M x_{\text{bin}} + f_{\text{arg}}(x_{\text{bin}}) + f_{\text{role}}(x)$

We found that... Oops! That doesn't work.

Strengthen the two relationships

- Map float numbers of M_{ij} to discrete integers: $[0, 1] \longrightarrow -1, 0, 1$

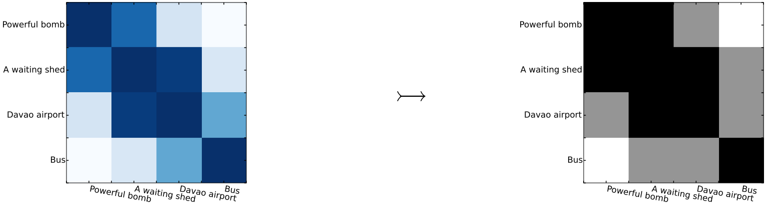


Figure: Visualization of M before and after strengthen.

Event Extraction from News Text

Example of the quadratic item (strengthened):

- Assume that "Baghdad" and "Palestine hotel" have negative relationship
- In this example, 1 represents positive relationship, -1 represents negative relationship, 0 means unclear relationship.

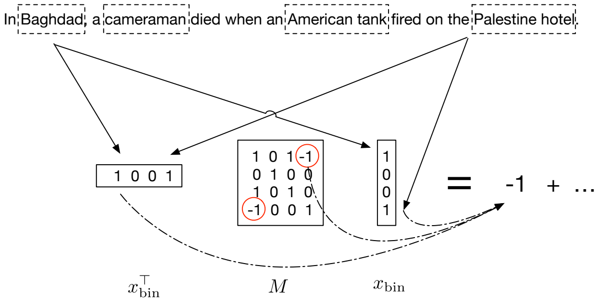


Figure: Example calculation process of negative relation.

Event Extraction from News Text

Another example of the quadratic item (strengthened):

- Assume that “Baghdad” and “American tank” have positive relationship
- Again, 1 represents positive relationship, -1 represents negative relationship, 0 means unclear relationship.

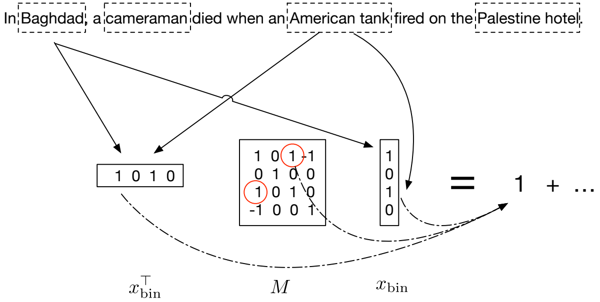


Figure: Example calculation process of positive relation.

Event Extraction from News Text

Can we do better?

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

Event Extraction from News Text

Sha et al.(AAAI 2018) Jointly Extracting Event Triggers and Arguments by Dependency-Bridge RNN and Tensor-Based Argument Interaction

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.

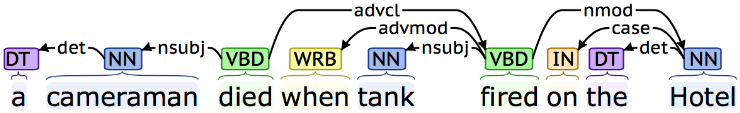


Figure: Example of dependency parse tree.

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.

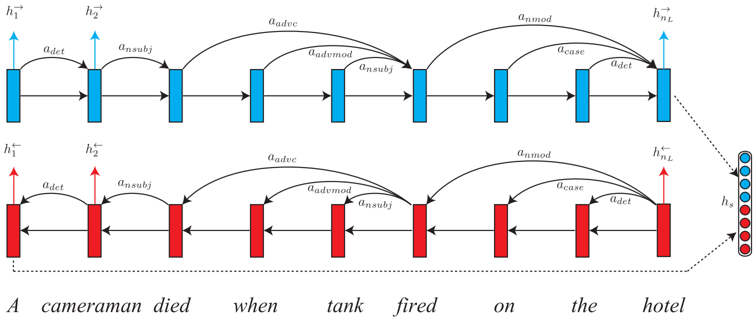


Figure: Dependency bridge on LSTM. Apart from the last LSTM cell, each cell also receives information from former syntactically related cells.

Event Extraction from News Text

Details of dependency bridge

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

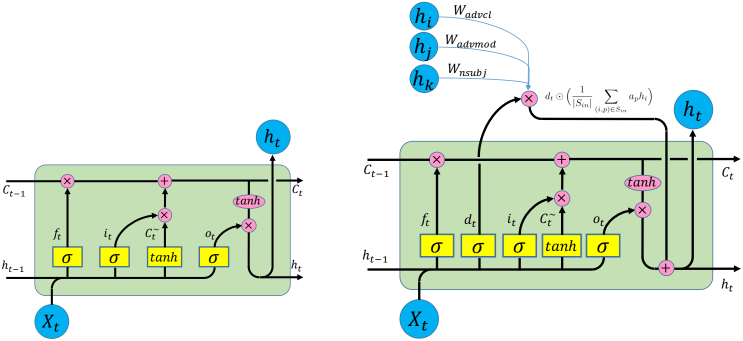


Figure: The calculation detail of dependency bridge.

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

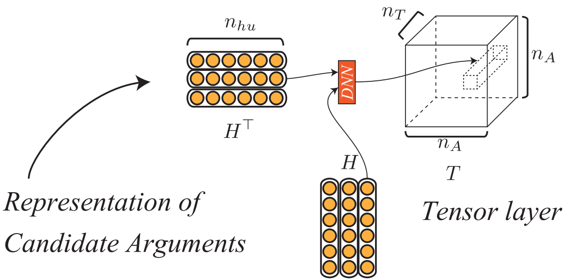
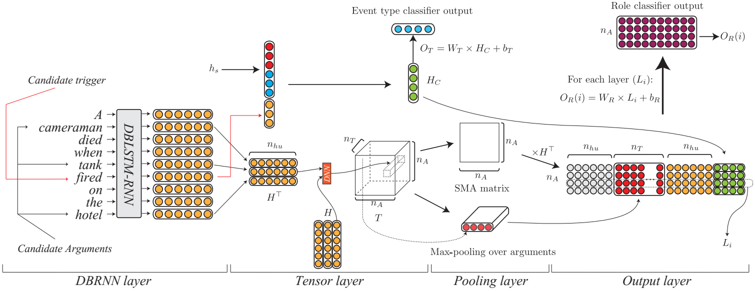


Figure: The calculation detail of tensor layer.

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



Weights of each dependency relation

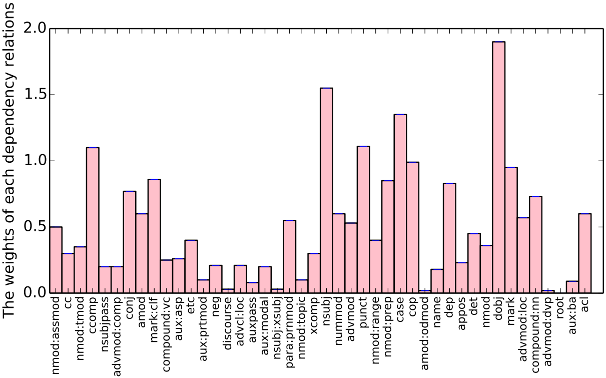


Figure: The visualization of trained weights of each dependency relations.

Event Extraction from News Text

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.

Table of Contents

Table-to-Text Brief Summary Generation

A table can be a list of RBF tuples:

John E Blaha	birthDate	1942,08,26
John E Blaha	birthPlace	San Antonio
John E Blaha	occupation	Fighter pilot
San Antonio	located in	USA

Table-to-Text Brief Summary Generation

A table can be also a list of attributes (like Wiki infobox):

Table:

ID	Field	Value
1	Name	*Arthur Ignatius Conan Doyle*
2	Born	*22 May 1859 Edinburgh, Scotland*
3	Died	*7 July 1930 (aged 71) Crowborough, England*
4	Occupation	*author writer physician*
5	Nationality	*British*
6	Alma mater	*University of Edinburgh Medical School*
7	Genre	*Detective fiction fantasy*
8	Notable work	*Stories of Sherlock Homes*

Figure: An example of Wikipedia infobox.

Table-to-Text Brief Summary Generation

Generate brief summary from structured data is useful

- In the last step of QA system, Table-to-text is used to generate answer.

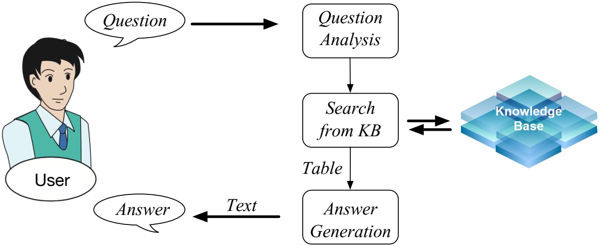


Figure: Table-to-text in question answering system.

Table-to-Text Brief Summary Generation

Table-to-text can also be used to generate response in dialogue system

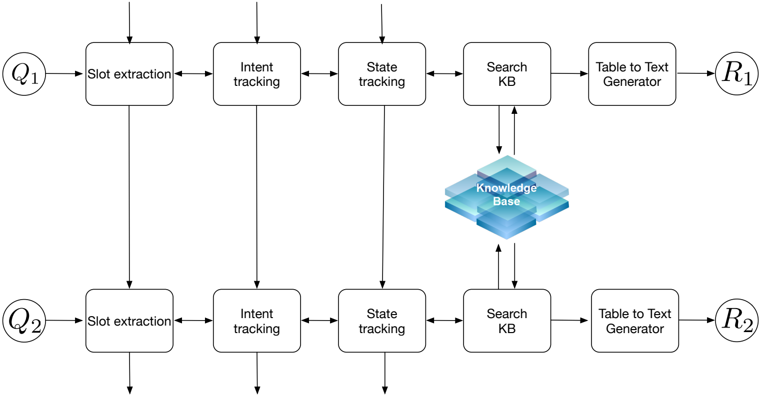


Figure: Table-to-text in dialogue system.

Table-to-Text Brief Summary Generation

We generate brief summary for wikipedia infobox

Table:

ID	Field	Value
1	Name	*Arthur Ignatius Conan Doyle*
2	Born	*22 May 1859 Edinburgh, Scotland*
3	Died	*7 July 1930 (aged 71) Crowborough, England*
4	Occupation	*author writer physician*
5	Nationality	*British*
6	Alma mater	*University of Edinburgh Medical School*
7	Genre	*Detective fiction fantasy*
8	Notable work	*Stories of Sherlock Homes*

Sir Arthur Ignatius Conan Doyle (22 May 1859 – 7 July 1930) was a British writer best known for his detective fiction featuring the character Sherlock Holmes.

Table-to-Text Brief Summary Generation

Motivation:

- Traditional: language model based generator
  - Use probability of word-by-word: $P(w_t|w_{t-1})$
  - Different from human's generation process
- Human: first plan for order, then write
  - Use probability of field-by-field: $P(f_t|f_{t-1})$
- We propose to add human nature into machine learning models

Table-to-Text Brief Summary Generation

Sha et al. (AAAI 2018) Order-Planning Neural Text Generation From Structured Data (arxiv)

- Content-based attention
  - Use the last output word y_{t-1} to predict the importance of each table content for the next output.
- Link-based attention
  - See which field we are going to generate this time.
- Hybrid attention
  - Combine content-based and link-based attention together.

Table-to-Text Brief Summary Generation

How to build field-by-field probability ($P(f_t|f_{t-1})$)?

- The element in the i-th row and j-th column is the probability of field j occurs after field i

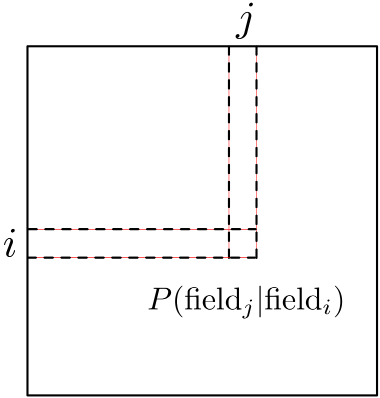


Figure: Field-by-field probability matrix (Link matrix).

Table-to-Text Brief Summary Generation

How to build link sub-matrix?

Field	Content
Name	Arthur
Name	Ignatius
Name	Conan
Name	Doyle
Born	22
Born	May
Born	1859
Occupation	writer
Occupation	physician
Nationality	British

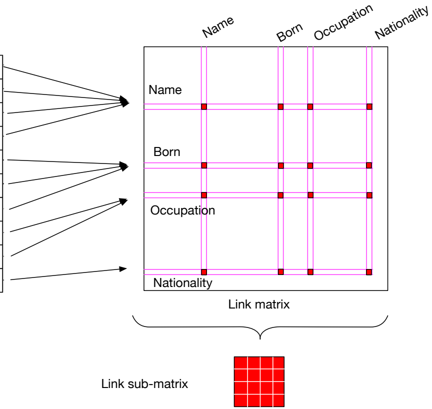


Figure: The process of select link sub-matrix.

Table-to-Text Brief Summary Generation

We calculate the hybrid attention as follows:

- (a) Encoder: Table Representation
- (b) Dispatcher: Planning What to Generate Next

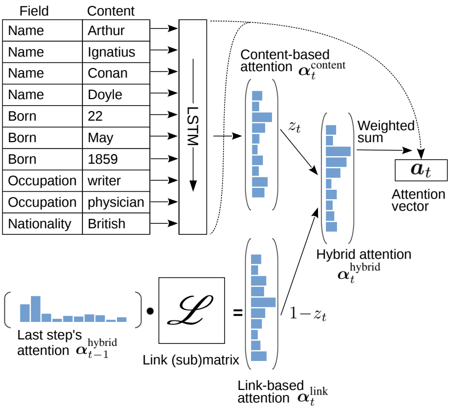


Figure: Illustration of content-based attention and link-based attention.

Table-to-Text Brief Summary Generation

Then we generate text according to the hybrid attention:

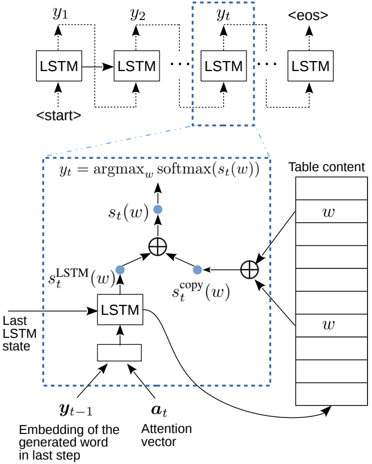


Figure: The decoder in our model, which is incorporated with a copying mechanism.

Table-to-Text Brief Summary Generation

Overall performance of our model:

Group	Model	BLEU	ROUGE	NIST
Previous results	KN	2.21	0.38	0.93
	Template KN	19.80	10.70	5.19
	Table NLM[l]	34.70	25.80	7.98
Our results	Content attention only	41.38	34.65	8.57
	Order planning (full model)	**43.91**	**37.15**	**8.85**

Figure: Comparison of the overall performance between our model and previous methods. [l]Best results in Lebret, Grangier, and Auli (2016).

Table-to-Text Brief Summary Generation

Simple case study:

Name	Emmett John Rice
Birth date	December 21, 1919
Birth place	Florence, South Carolina, United States
Death date	March 10, 2011 (aged 91)
Death place	Camas, Washington, United States
Nationality	American
Occupation	Governor of the Federal Reserve System, Economics Professor
Known for	Expert in the Monetary System of Developing Countries, Father to Susan E. Rice

Reference	emmett john rice (december 21 , 1919 – march 10 , 2011) was a former governor of the federal reserve system , a Cornell university economics professor , expert in the monetary systems of developing countries and the father of the current national security advisor to president barack obama , susan e . rice .
Content-based attention	emmett john rice (december 21 , 1919 – march 10 , 2011) was an economist , author , public official and the former american governor of the federal reserve system , the first african american UNK .
Hybrid attention	emmett john rice (december 21 , 1919 – march 10 , 2011) was an american economist , author , public official and the former governor of the federal reserve system , expert in the monetary systems of developing countries .

Figure: Case study. Left: Wikipedia infobox. Right: A reference and two generated sentences by different attention (both with the copy mechanism).

Table-to-Text Brief Summary Generation

Visualization of attention probabilities in our model.

- x-axis: generated words "...) was an american economist ...";
- y-axis: ⟨field : content word⟩ pairs in the table.

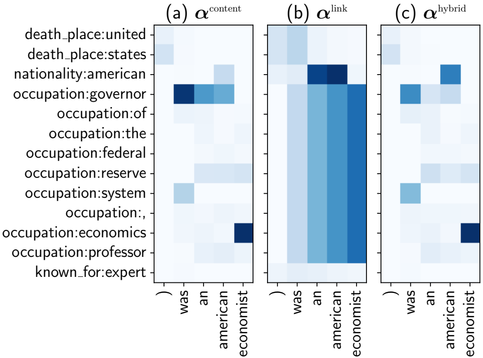


Figure: Subplot (b) exhibits strips because, by definition, link-based attention will yield the same score for all content words with the same field.

Table of Contents

Other works

Research papers:

- Event schema induction (Sha et al, NAACL 2016)
- Chinese SRL (Sha et al, EMNLP 2016)
- Textual Entailment Recognition (Sha et al, EMNLP 2015, Coling 2016)
- Repeated Reading (Sha et al, NLPCC 2017)

Other works

Automatically Design Neural Network Architecture (in Sinovation Ventures)

- Implement a simple version of paper "Neural Architecture Search With Reinforcement Learning"
- Use policy gradient method to increase the probability of sampling child networks with high reward

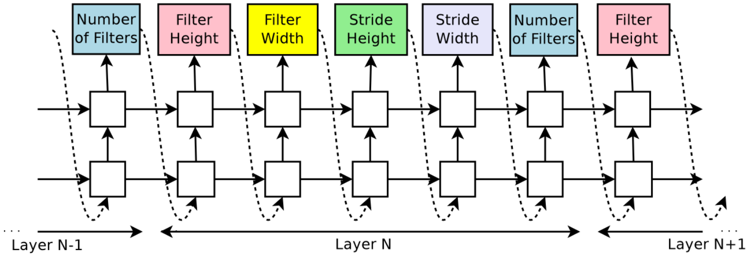


Figure: The process of a parent controller recurrent neural network sampling a child network.

Other works

Multi-intent switch dialogue system (in MSRA system group)

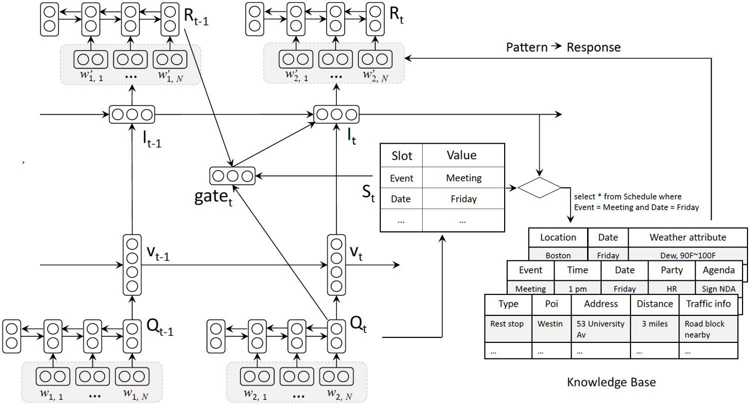


Figure: The architecture of ISwitch.

Publications (lead author)

- Sha et al: Order-Planning Neural Text Generation From Structured Data. In **AAAI 2018**.
- Sha et al: Joint Extracting Event Trigger and Arguments Using dependency bridge Recurrent Tensor Network. In **AAAI 2018**.
- Sha et al: Multi-View Fusion Neural Network for Answer Selection. In **AAAI 2018**.
- Sha et al: Will Repeated Reading Benefit Natural Language Understanding? In **NLPCC 2017**.
- Sha et al: Reading and Thinking: Re-read LSTM Unit for Textual Entailment Recognition. In **Coling 2016**.
- Sha et al: Capturing Argument Connection for Chinese Semantic Role Labeling. In **EMNLP 2016**.
- Sha et al: RBPB: Regularization-Based Pattern Balancing Method for Event Extraction. In **ACL 2016**.
- Sha et al: Joint Learning Templates and Slots for Event Schema Induction. In **NAACL 2016**.
- Sha et al: Recognizing Textual Entailment Using Probabilistic Inference. In **EMNLP 2015**.

Publications (co-author)

- Feng Qian, **Lei Sha**, Baobao Chang, Lu-chen Liu, Ming Zhang: Syntax Aware LSTM Model for Chinese Semantic Role Labeling. In **EMNLP 2017**
- Qiaolin Xia, **Lei Sha**, Baobao Chang, Zhifang Sui: A Progressive Learning Approach to Chinese SRL Using Heterogeneous Data. In **ACL 2017**
- Xiaodong Zhang, **Lei Sha**, Sujian Li, Houfeng Wang: Attentive Interactive Neural Networks for Answer Selection in Community Question Answering. In **AAAI 2017**
- Tingsong Jiang, Tianyu Liu, Tao Ge, **Lei Sha**, Sujian Li, Baobao Chang and Zhifang Sui: Encoding Temporal Information for Time-Aware Link Prediction. In **EMNLP 2016**.
- Tingsong Jiang, Tianyu Liu, Tao Ge, **Lei Sha**, Sujian Li, Baobao Chang and Zhifang Sui: Towards Time-Aware Knowledge Graph Completion. In **Coling 2016**.
- Li Li, Houfeng Wang, **Lei Sha**, Xu Sun, Baobao Chang, Shi Zhao: Multi-label Text Categorization with Joint Learning Predictions-as-Features Method. In **EMNLP 2015**.
- Tingsong Jiang, **Lei Sha** and Zhifang Sui: Event Schema Induction Based on Relational Co-occurrence over Multiple Documents. In **NLPCC 2014**.

Thank you. Any questions?