北京大學
PEKING UNIVERSITY

Modeling Discourse Cohesion for Discourse Parsing via Memory Network

Yanyan Jia, Yuan Ye, Yansong Feng, Yuxuan Lai,
Rui Yan and Dongyan Zhao

Institute of Computer Science and Technology, Peking University

Discourse Dependency Parsing

EDU means **Element Discourse Unit**

EDU_1: President Bush insists

EDU_2: it would be a great tool

EDU_3: for curbing the budget deficit

EDU_4: and slicing the lard out of government programs.

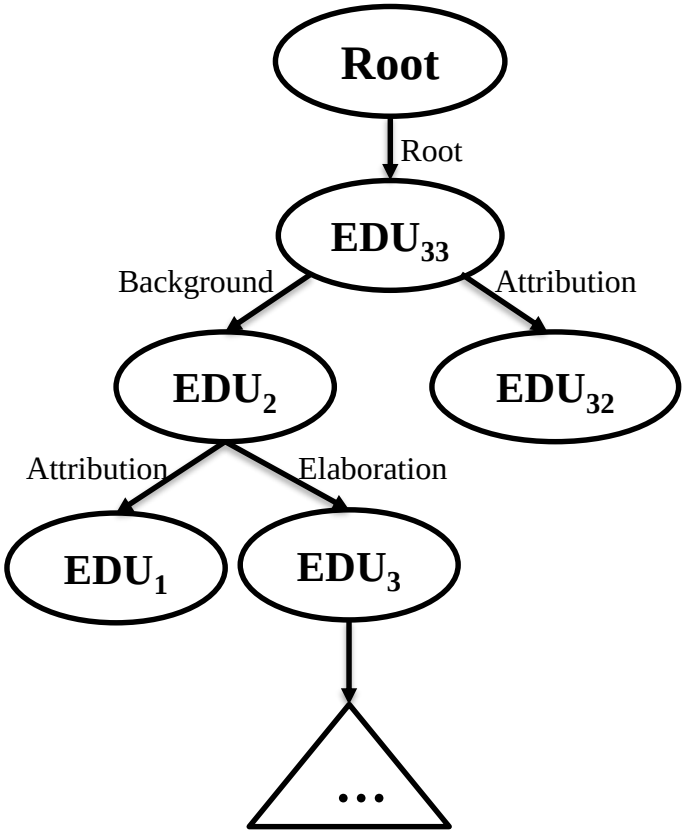
EDU_5: He wants it now .

…

EDU_{32}: Mr. Bush is considering simply declaring

EDU_{33}: that the Constitution gives him the power

…



Discourse Dependency Parsing

EDU_1: President Bush insists

EDU_2: it would be a great tool

EDU_3: for curbing the budget deficit

EDU_4: and slicing the lard out of government programs.

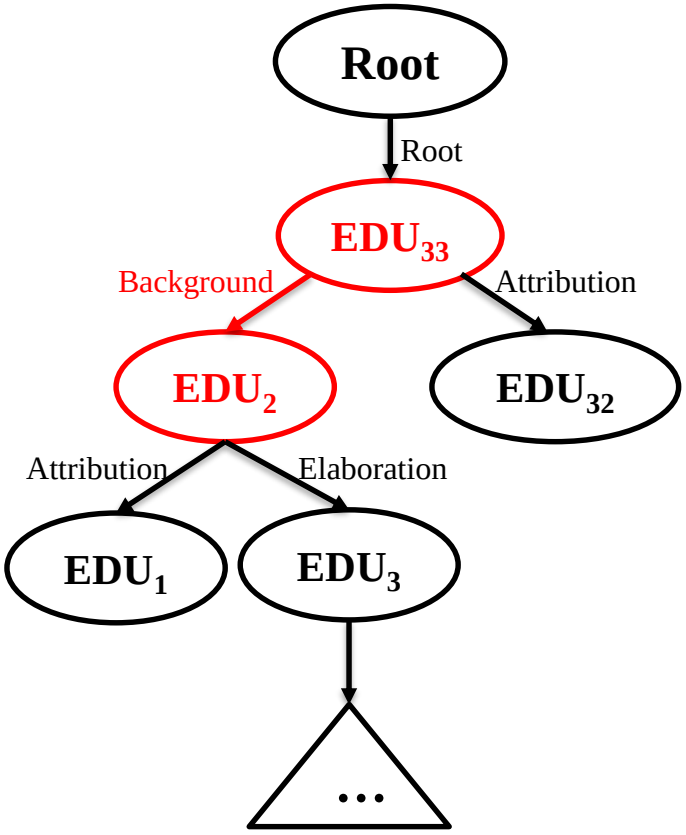
EDU_5: He wants it now .

…

EDU_{32}: Mr. Bush is considering simply declaring

EDU_{33}: that the Constitution gives him the power

…



Motivation

- Identifying **long-span dependencies** between element discourse units
  - Discourse structure
    - Morris and Hirst, 1991 extracts features to characterize discourse structures
  - Discourse cohesion
    - Joty et al., 2013 uses lexical chain features to model discourse cohesion

Motivation

- Identifying **long-span dependencies** between element discourse units
  - Discourse structure
    - Morris and Hirst, 1991 extracts features to characterize discourse structures
  - Discourse cohesion
    - Joty et al., 2013 uses lexical chain feature to model discourse cohesion

Our Work: Use Memory network to implicitly capture discourse cohesion

How Does Memory Network Work?

EDU_1: I feel hungry after wake up,
EDU_2: I rush into the kitchen and make my breakfast.
EDU_3: My breakfast is hamburger.

EDU_4: It is eight o'clock when I leave home.
EDU_5: So late!

EDU_6: I drive into the highway,
EDU_7: but meet a traffic jam.
EDU_8: Oh, I finally arrive at the company.

EDU_9: It is nine o'clock.
EDU_{10}: Thank God, I am not late for work.

EDU_{11}: But the hamburger is cold,
EDU_{12}: order some take-away food is better, maybe.

How Does Memory Network Work?

EDU_1: I feel hungry after wake up,
EDU_2: I rush into the kitchen and make my breakfast.
EDU_3: My breakfast is hamburger.

EDU_4: It is eight o'clock when I leave home.
EDU_5: So late!

EDU_6: I drive into the highway,
EDU_7: but meet a traffic jam.
EDU_8: Oh, I finally arrive at the company.

EDU_9: It is nine o'clock.
EDU_{10}: Thank God, I am not late for work.

EDU_{11}: But the hamburger is cold,
EDU_{12}: order some take-away food is better, maybe.

Food

How Does Memory Network Work?

EDU_1: I feel hungry after wake up,
EDU_2: I rush into the kitchen and make my breakfast.
EDU_3: My breakfast is hamburger.

EDU_4: It is eight o'clock when I leave home.
EDU_5: So late!

EDU_6: I drive into the highway,
EDU_7: but meet a traffic jam.
EDU_8: Oh, I finally arrive at the company.

EDU_9: It is nine o'clock.
EDU_{10}: Thank God, I am not late for work.

EDU_{11}: But the hamburger is cold,
EDU_{12}: order some take-away food is better, maybe.

Time

How Does Memory Network Work?

EDU_1: I feel hungry after wake up,
EDU_2: I rush into the kitchen and make my breakfast.
EDU_3: My breakfast is hamburger.

EDU_4: It is eight o'clock when I leave home.
EDU_5: So late!

EDU_6: I drive into the highway,
EDU_7: but meet a traffic jam.
EDU_8: Oh, I finally arrive at the company.

EDU_9: It is nine o'clock.
EDU_{10}: Thank God, I am not late for work.

EDU_{11}: But the hamburger is cold,
EDU_{12}: order some take-away food is better, maybe.

Traffic

How Does Memory Network Work?

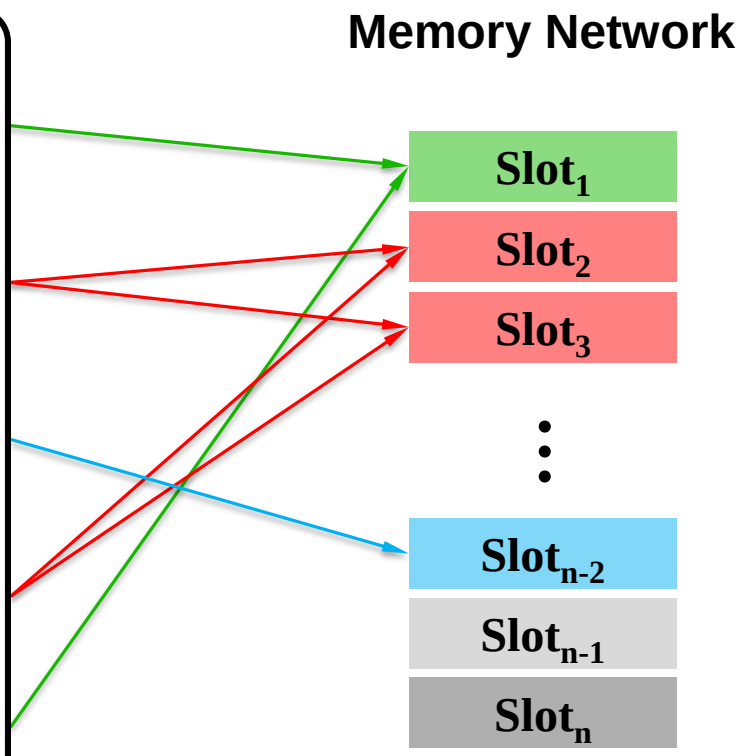
EDU_1: I feel hungry after wake up,
EDU_2: I rush into the kitchen and make my breakfast.
EDU_3: My breakfast is hamburger.

EDU_4: It is eight o'clock when I leave home.
EDU_5: So late!

EDU_6: I drive into the highway,
EDU_7: but meet a traffic jam.
EDU_8: Oh, I finally arrive at the company.

EDU_9: It is nine o'clock.
EDU_{10}: Thank God, I am not late for work.

EDU_{11}: But the hamburger is cold,
EDU_{12}: order some take-away food is better, maybe.



Framework

Transition-based dependency parsing

Arc-eager algorithm (Nivre):

Stack, Buffer, Arcs set

Left-Arc(LA)	$\langle e\|S, e'\|B, Arcs\rangle \rightarrow \langle S, e'\|B, Arcs \cup \{(e', e)\}\rangle$
Right-Arc(RA)	$\langle e\|S, e'\|B, Arcs\rangle \rightarrow \langle e'\|e\|S, B, Arcs \cup \{(e, e')\}\rangle$
Shift	$\langle S, e\|B, Arcs\rangle \rightarrow \langle e\|S, B, Arcs\rangle$
Reduce	$\langle e\|S, B, Arcs\rangle \rightarrow \langle S, B, Arcs\rangle$

Framework

Transition-based dependency parsing

Arc-eager algorithm (Nivre):

Stack, Buffer, Arcs set

Left-Arc(LA)	$\langle e\|S, e'\|B, Arcs\rangle \rightarrow \langle S, e'\|B, Arcs \cup \{(e', e)\}\rangle$
Right-Arc(RA)	$\langle e\|S, e'\|B, Arcs\rangle \rightarrow \langle e'\|e\|S, B, Arcs \cup \{(e, e')\}\rangle$
Shift	$\langle S, e\|B, Arcs\rangle \rightarrow \langle e\|S, B, Arcs\rangle$
Reduce	$\langle e\|S, B, Arcs\rangle \rightarrow \langle S, B, Arcs\rangle$

Framework

Transition-based dependency parsing

Arc-eager algorithm (Nivre):

Stack, Buffer, Arcs set

Left-Arc(LA)	$\langle e\|S, e'\|B, Arcs\rangle \rightarrow \langle S, e'\|B, Arcs \cup \{(e', e)\}\rangle$
Right-Arc(RA)	$\langle e\|S, e'\|B, Arcs\rangle \rightarrow \langle e'\|e\|S, B, Arcs \cup \{(e, e')\}\rangle$
Shift	$\langle S, e\|B, Arcs\rangle \rightarrow \langle e\|S, B, Arcs\rangle$
Reduce	$\langle e\|S, B, Arcs\rangle \rightarrow \langle S, B, Arcs\rangle$

Framework

Transition-based dependency parsing

Arc-eager algorithm (Nivre):

Stack, Buffer, Arcs set

Left-Arc(LA) $\langle e|S, e'|B, Arcs\rangle \rightarrow \langle S, e'|B, Arcs \cup \{(e', e)\}\rangle$

Right-Arc(RA) $\langle e|S, e'|B, Arcs\rangle \rightarrow \langle e'|e|S, B, Arcs \cup \{(e, e')\}\rangle$

- **Shift** $\langle S, e|B, Arcs\rangle \rightarrow \langle e|S, B, Arcs\rangle$

Reduce $\langle e|S, B, Arcs\rangle \rightarrow \langle S, B, Arcs\rangle$

Framework

Transition-based dependency parsing

Arc-eager algorithm (Nivre):

Stack, Buffer, Arcs set

Left-Arc(LA) $\langle e|S, e'|B, Arcs\rangle \rightarrow \langle S, e'|B, Arcs \cup \{(e', e)\}\rangle$

Right-Arc(RA) $\langle e|S, e'|B, Arcs\rangle \rightarrow \langle e'|e|S, B, Arcs \cup \{(e, e')\}\rangle$

Shift $\langle S, e|B, Arcs\rangle \rightarrow \langle e|S, B, Arcs\rangle$

Reduce $\langle e|S, B, Arcs\rangle \rightarrow \langle S, B, Arcs\rangle$

Arc-eager

EDU_1: President Bush insists

EDU_2: it would be a great tool

EDU_3: for curbing the budget deficit

EDU_4: and slicing the lard out of government programs.

EDU_5: He wants it now .

…

EDU_{32}: Mr. Bush is considering simply declaring

EDU_{33}: that the Constitution gives him the power

…

Arc-eager

Transition	**Stack**	**Buffer**
	[]	**[E_1, E_2, E_3, E_4, ⋯]**

EDU_1: President Bush insists

EDU_2: it would be a great tool

EDU_3: for curbing the budget deficit

EDU_4: and slicing the lard out of government programs.

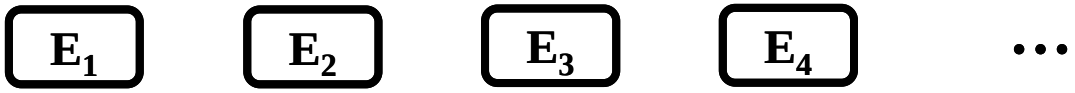
EDU_5: He wants it now .

…

EDU_{32}: Mr. Bush is considering simply declaring

EDU_{33}: that the Constitution gives him the power

…



Arc-eager

Transition	**Stack**	**Buffer**
	[]	**[E_1, E_2, E_3, E_4, ⋯]**
Shift	**[E_1]**	**[E_2, E_3, E_4, ⋯]**

EDU_1: President Bush insists

EDU_2: it would be a great tool

EDU_3: for curbing the budget deficit

EDU_4: and slicing the lard out of government programs.

EDU_5: He wants it now .

…

EDU_{32}: Mr. Bush is considering simply declaring

EDU_{33}: that the Constitution gives him the power

…

E_1 **E_2** **E_3** **E_4** …

Arc-eager

Transition	**Stack**	**Buffer**
	[]	**[E_1, E_2, E_3, E_4, ⋯]**
Shift	**[E_1]**	**[E_2, E_3, E_4, ⋯]**
LA(Attribution)	**[]**	**[E_2, E_3, E_4, ⋯]**

EDU_1: President Bush insists

EDU_2: it would be a great tool

EDU_3: for curbing the budget deficit

EDU_4: and slicing the lard out of government programs.

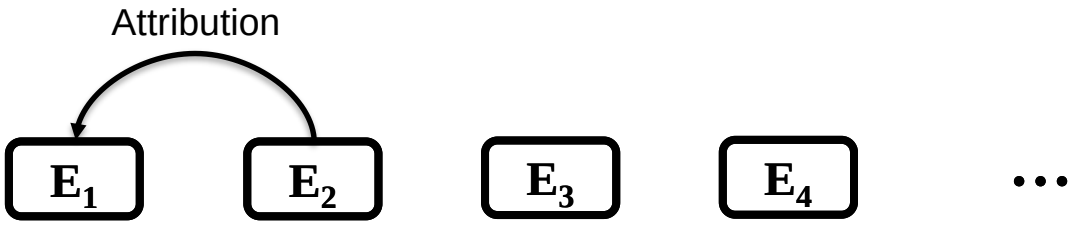
EDU_5: He wants it now .

…

EDU_{32}: Mr. Bush is considering simply declaring

EDU_{33}: that the Constitution gives him the power

…



Arc-eager

Transition	**Stack**	**Buffer**
	[]	**[E_1, E_2, E_3, E_4, ⋯]**
Shift	**[E_1]**	**[E_2, E_3, E_4, ⋯]**
LA(Attribution)	**[]**	**[E_2, E_3, E_4, ⋯]**
SH	**[E_2]**	**[E_3, E_4, ⋯]**

EDU_1: President Bush insists

EDU_2: it would be a great tool

EDU_3: for curbing the budget deficit

EDU_4: and slicing the lard out of government programs.

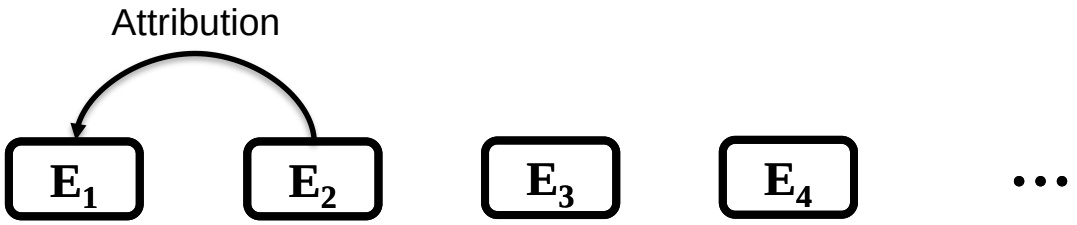
EDU_5: He wants it now .

…

EDU_{32}: Mr. Bush is considering simply declaring

EDU_{33}: that the Constitution gives him the power

…



Arc-eager

Transition	**Stack**	**Buffer**
	[]	**[E_1, E_2, E_3, E_4, ⋯]**
Shift	**[E_1]**	**[E_2, E_3, E_4, ⋯]**
LA(Attribution)	**[]**	**[E_2, E_3, E_4, ⋯]**
SH	**[E_2]**	**[E_3, E_4, ⋯]**
RA(Elaboration)	**[E_2, E_3]**	**[E_4, ⋯]**

EDU_1: President Bush insists

EDU_2: it would be a great tool

EDU_3: for curbing the budget deficit

EDU_4: and slicing the lard out of government programs.

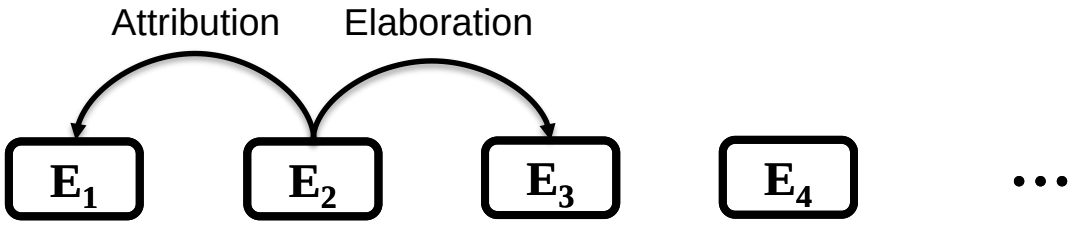
EDU_5: He wants it now .

…

EDU_{32}: Mr. Bush is considering simply declaring

EDU_{33}: that the Constitution gives him the power

…



Arc-eager

Transition	**Stack**	**Buffer**
	[]	**[E_1, E_2, E_3, E_4, ⋯]**
Shift	**[E_1]**	**[E_2, E_3, E_4, ⋯]**
LA(Attribution)	**[]**	**[E_2, E_3, E_4, ⋯]**
SH	**[E_2]**	**[E_3, E_4, ⋯]**
RA(Elaboration)	**[E_2, E_3]**	**[E_4, ⋯]**
RA(Joint)	**[E_2, E_3, E_4]**	**[⋯]**

EDU_1: President Bush insists

EDU_2: it would be a great tool

EDU_3: for curbing the budget deficit

EDU_4: and slicing the lard out of government programs.

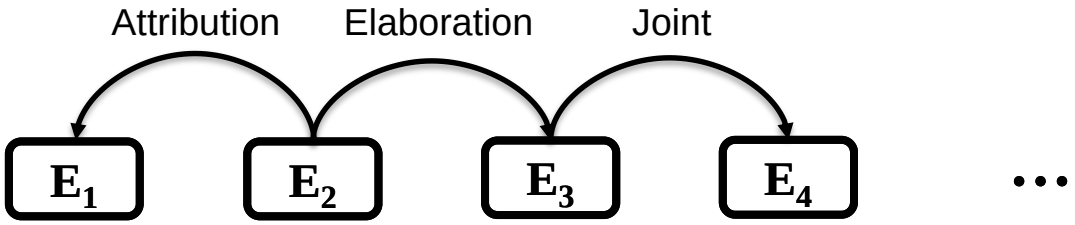
EDU_5: He wants it now .

…

EDU_{32}: Mr. Bush is considering simply declaring

EDU_{33}: that the Constitution gives him the power

…



Arc-eager

Transition	**Stack**	**Buffer**
	[]	**[E_1, E_2, E_3, E_4, ⋯]**
Shift	**[E_1]**	**[E_2, E_3, E_4, ⋯]**
LA(Attribution)	**[]**	**[E_2, E_3, E_4, ⋯]**
SH	**[E_2]**	**[E_3, E_4, ⋯]**
RA(Elaboration)	**[E_2, E_3]**	**[E_4, ⋯]**
RA(Joint)	**[E_2, E_3, E_4]**	**[⋯]**
⋮	⋮	⋮

EDU_1: President Bush insists

EDU_2: it would be a great tool

EDU_3: for curbing the budget deficit

EDU_4: and slicing the lard out of government programs.

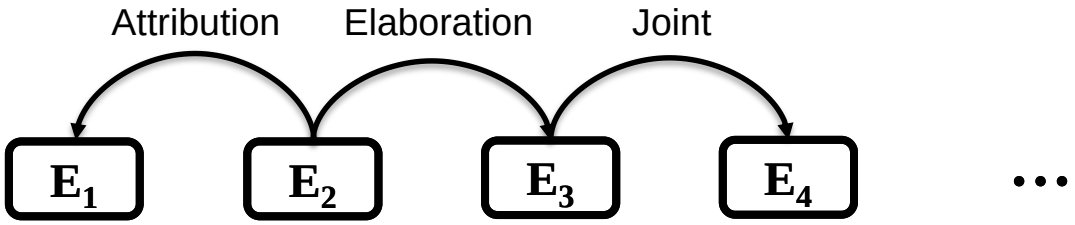
EDU_5: He wants it now .

…

EDU_{32}: Mr. Bush is considering simply declaring

EDU_{33}: that the Constitution gives him the power

…



Arc-eager

Transition	Stack	Buffer
	[]	[E_1, E_2, E_3, E_4, ⋯]
Shift	[E_1]	[E_2, E_3, E_4, ⋯]
LA(Attribution)	[]	[E_2, E_3, E_4, ⋯]
SH	[E_2]	[E_3, E_4, ⋯]
RA(Elaboration)	[E_2, E_3]	[E_4, ⋯]
RA(Joint)	[E_2, E_3, E_4]	[⋯]
⋮	⋮	⋮

EDU_1: President Bush insists

EDU_2: it would be a great tool

EDU_3: for curbing the budget deficit

EDU_4: and slicing the lard out of government programs.

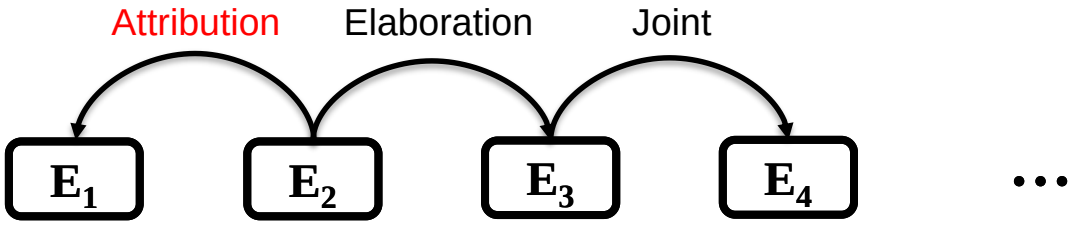
EDU_5: He wants it now .

…

EDU_{32}: Mr. Bush is considering simply declaring

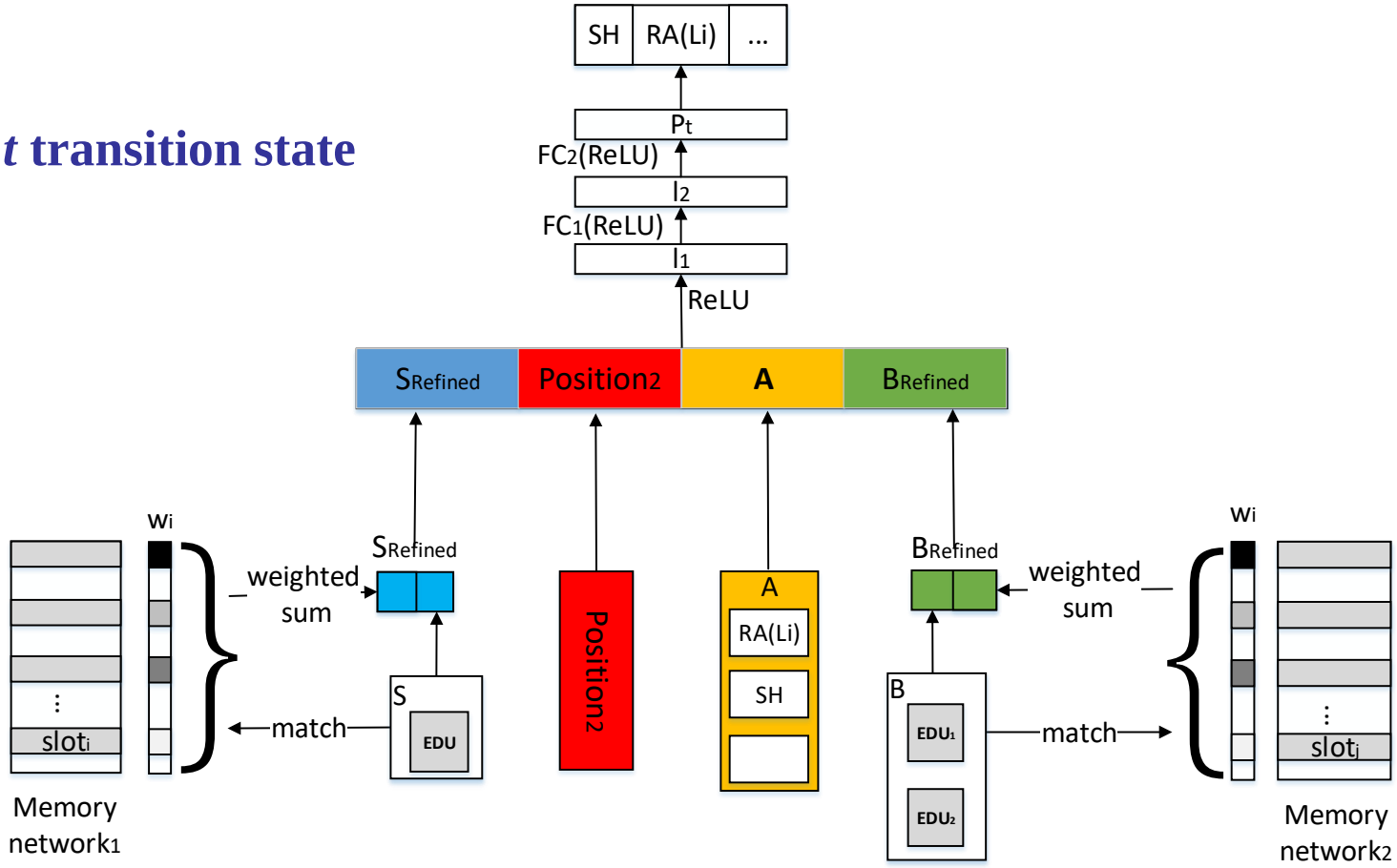
EDU_{33}: that the Constitution gives him the power

…



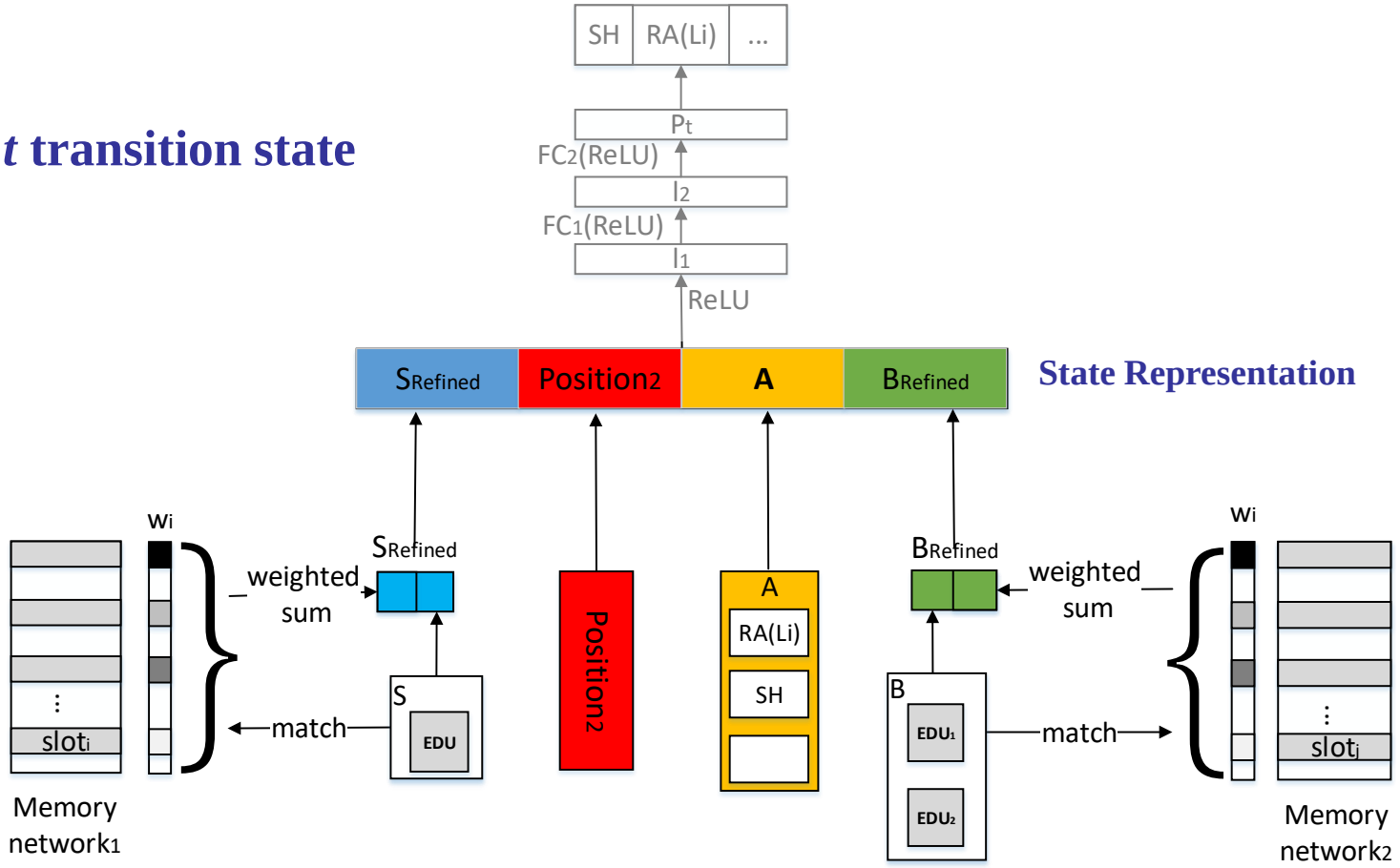
Model Overview

time *t* transition state



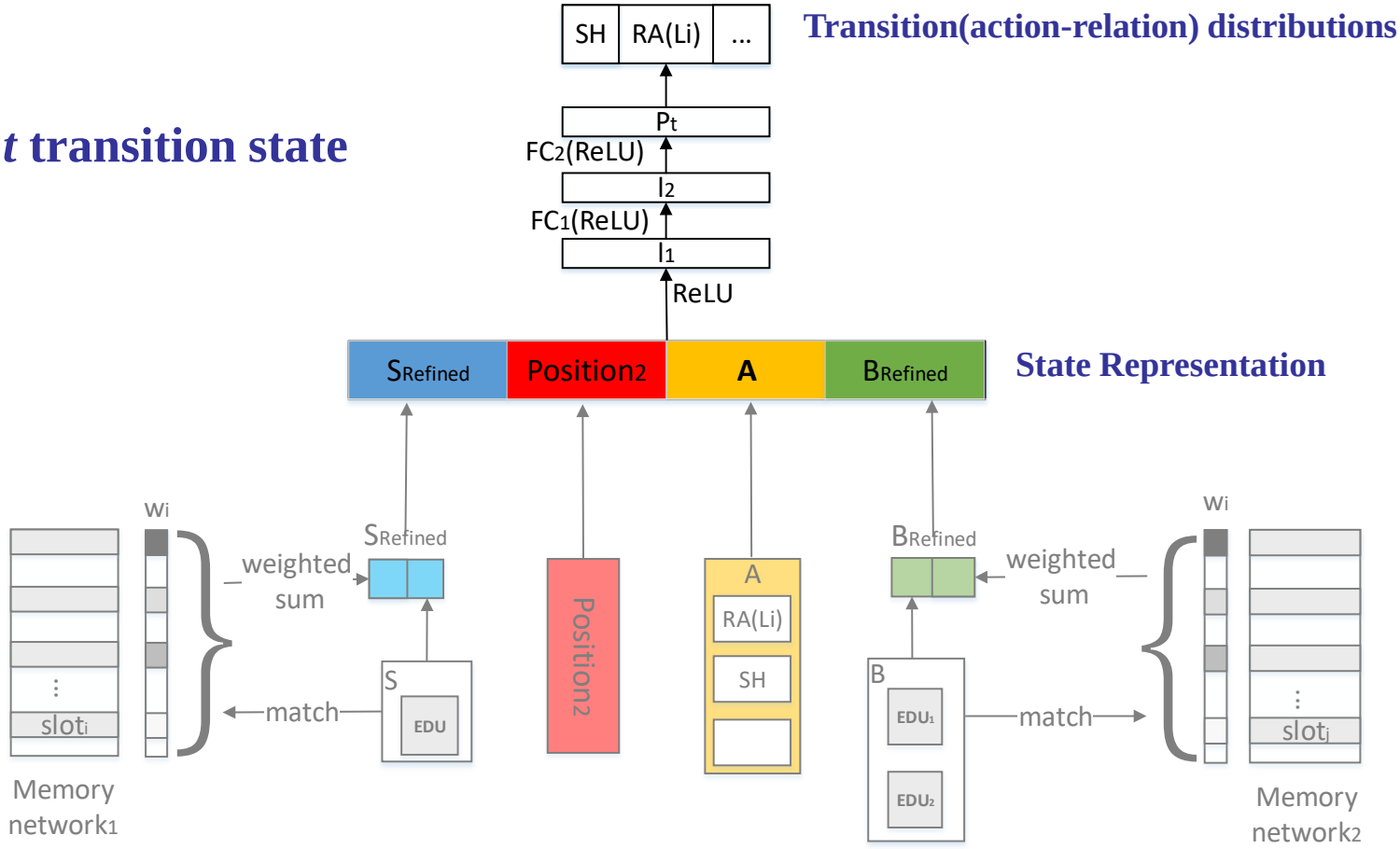
Model Overview

time *t* transition state

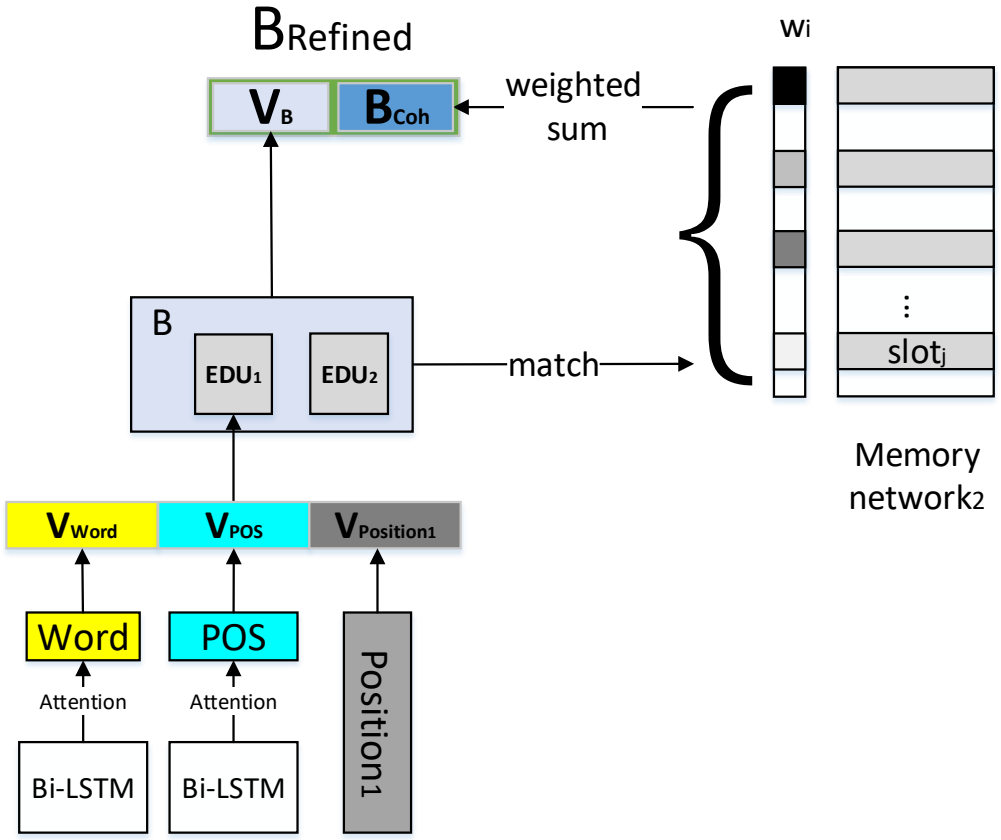


Model Overview

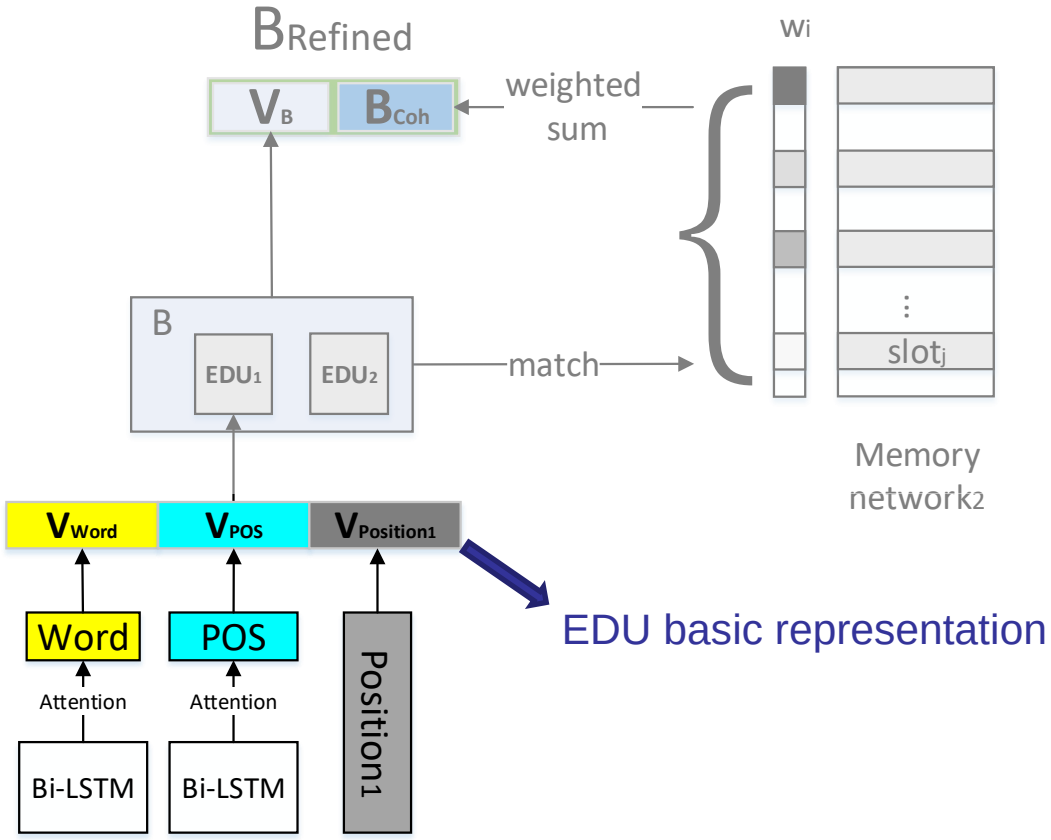
time *t* transition state



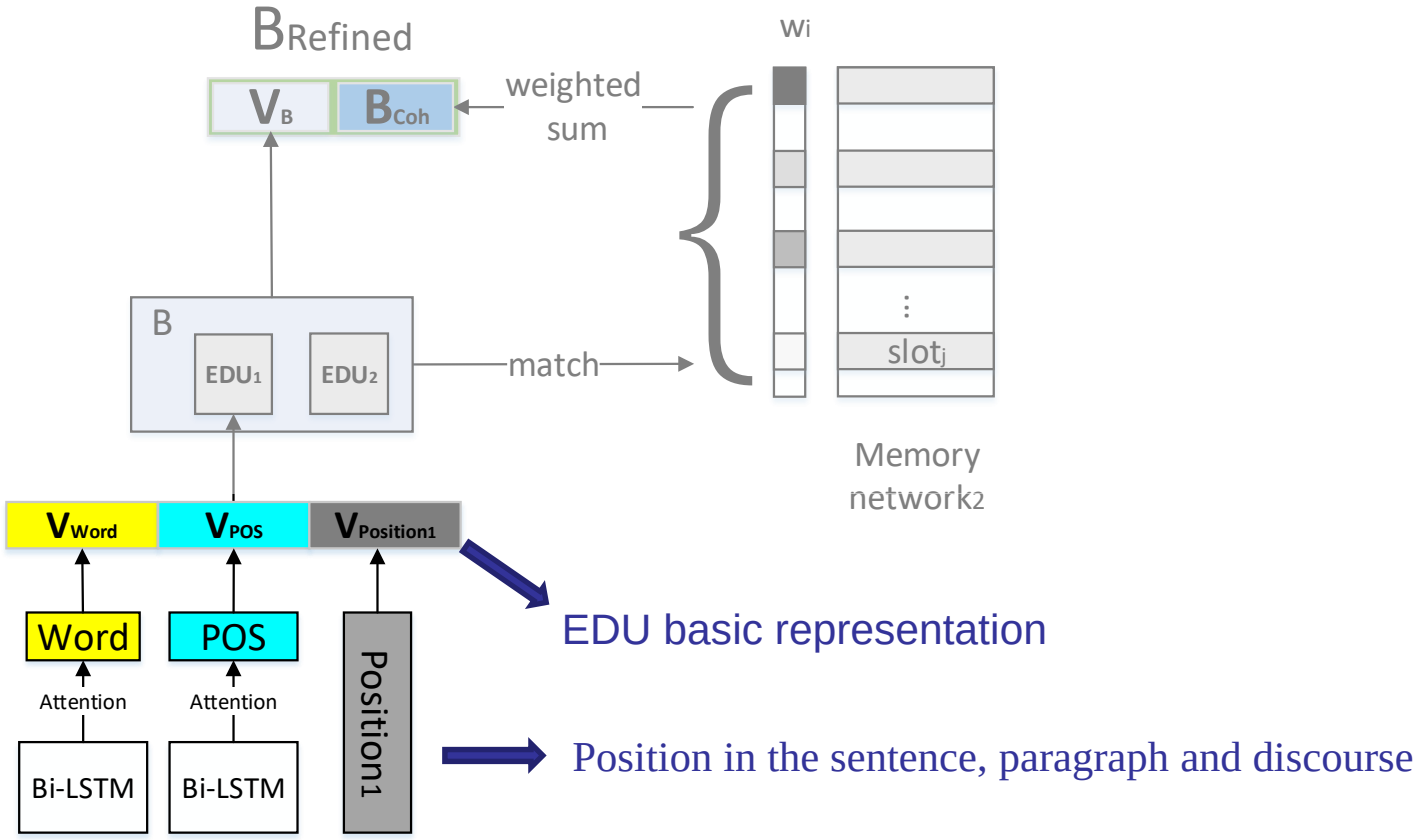
BRefined


BRefined
Wi
weighted sum
VB
BCoh
B
EDU1
EDU2
match
slotj
Memory network2
VWord
VPOS
VPosition1
Word
POS
Position1
Attention
Attention
Bi-LSTM
Bi-LSTM

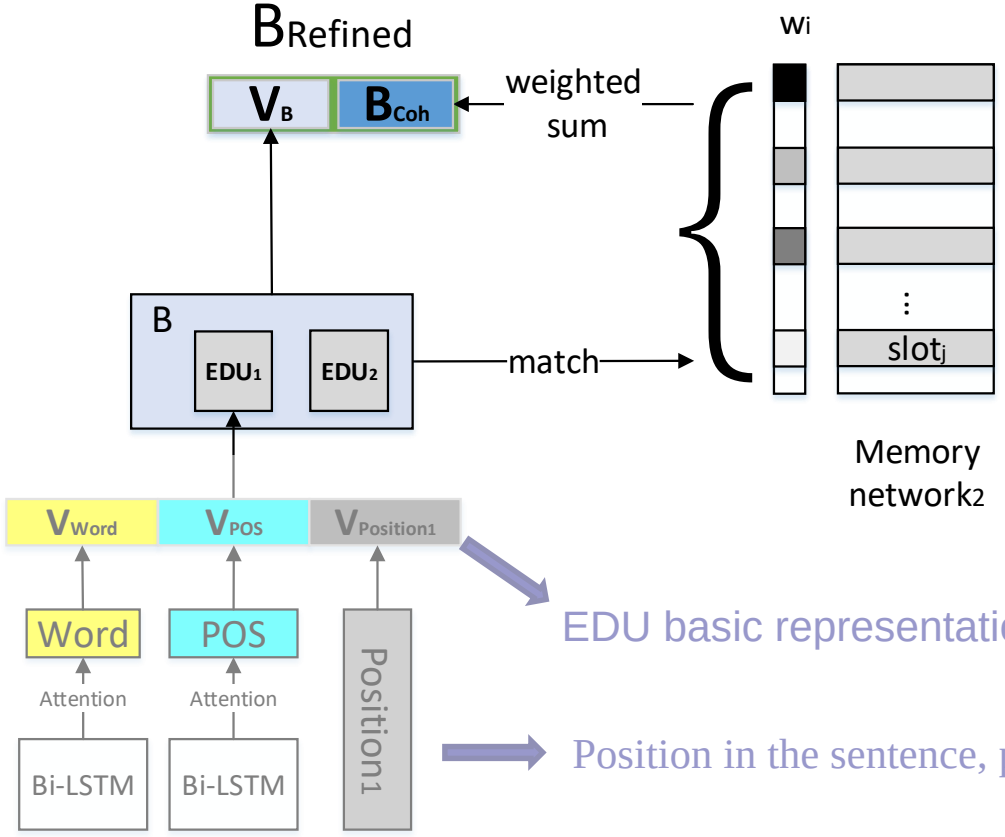
BRefined


BRefined
Wi
weighted sum
VB
BCoh
B
EDU1
EDU2
match
slotj
Memory network2
VWord
VPOS
VPosition1
EDU basic representation
Word
POS
Position1
Attention
Attention
Bi-LSTM
Bi-LSTM

$B_{Refined}$


BRefined
VB
BCoh
weighted sum
Wi
B
EDU1
EDU2
match
slotj
Memory network2
VWord
VPOS
VPosition1
Word
POS
Position1
Attention
Attention
Bi-LSTM
Bi-LSTM
EDU basic representation
Position in the sentence, paragraph and discourse

BRefined



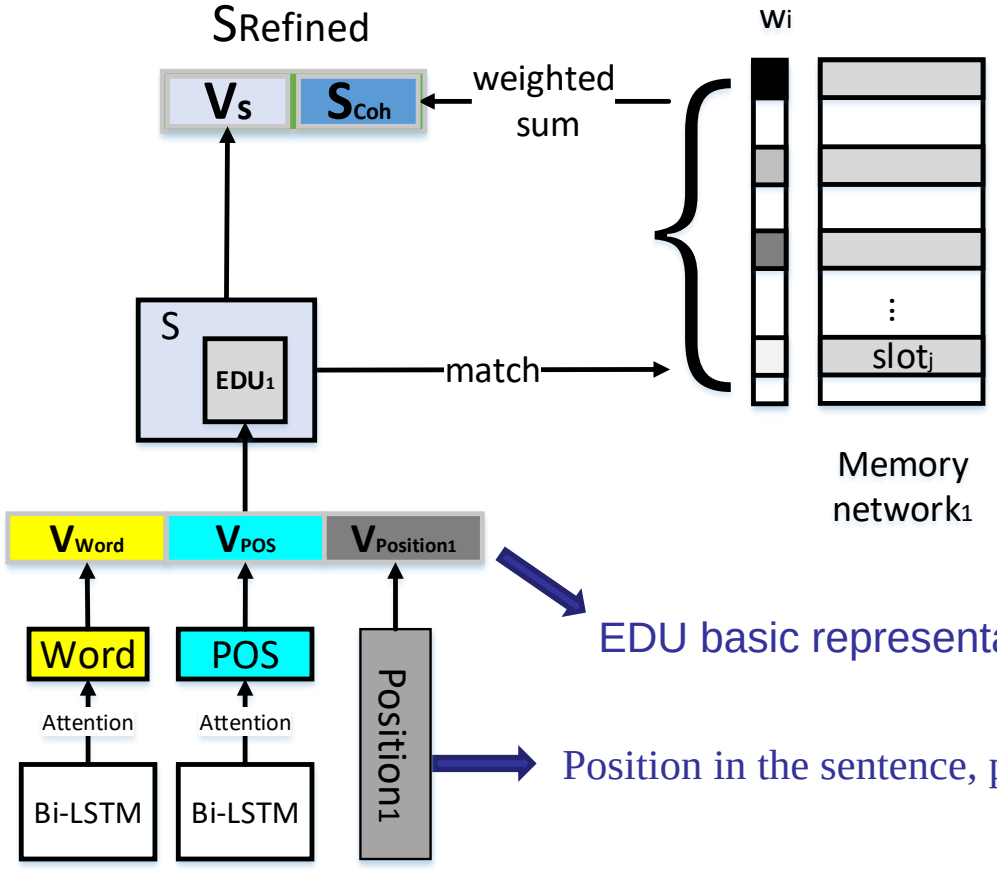
$$w_i = \frac{exp(\lambda Cos(V_B, slot_i))}{\sum_j exp(\lambda Cos(V_B, slot_j))}$$

$$B_{Coh} = \sum_i w_i slot_i$$

EDU basic representation

Position in the sentence, paragraph and discourse

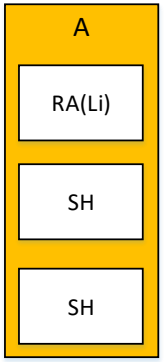
SRefined



$$w_i = \frac{exp(\lambda Cos(V_S, slot_i))}{\sum_j exp(\lambda Cos(V_S, slot_j))}$$

$$S_{Coh} = \sum_i w_i slot_i$$

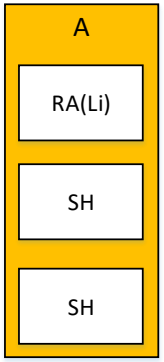
A and Position2



Top three transition information

Concatenate every transition's embedding

A and Position2



Top three transition information

Concatenate every transition's embedding

Position2

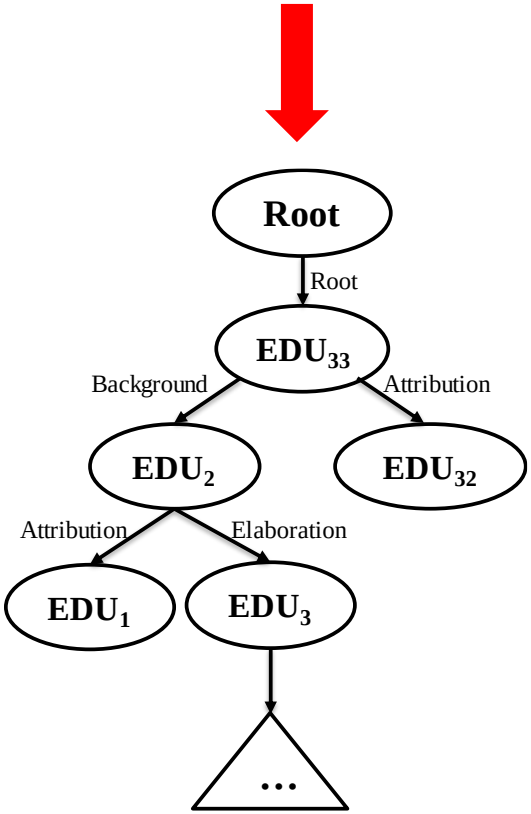
The spatial relationship between the top EDUs of S and B

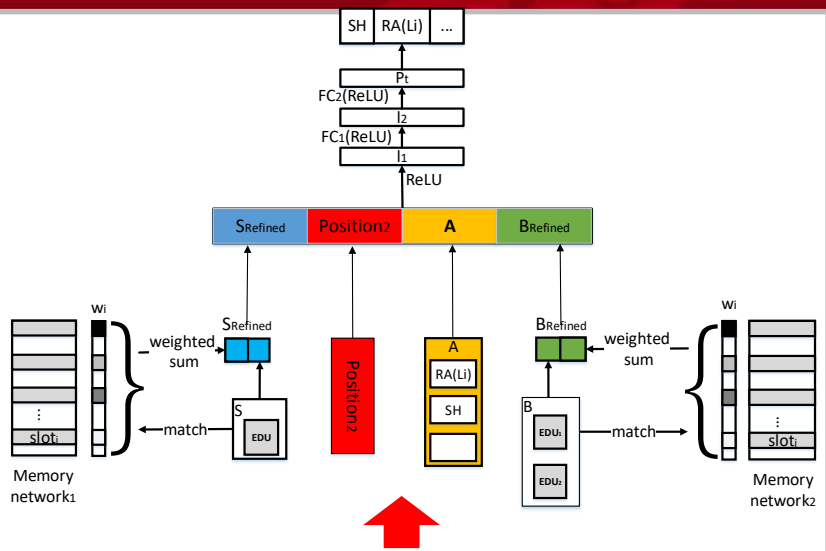
- Same sentence
- Same paragraph
- Distance in paragraph

Overall Process

Transitions Sequence:

Shift, LA-attribution, SH, RA-elaboration , RA-joint, …





EDU_1: President Bush insists

EDU_2: it would be a great tool

EDU_3: for curbing the budget deficit

EDU_4: and slicing the lard out of government programs.

EDU_5: He wants it now .

…

EDU_{32}: Mr. Bush is considering simply declaring

EDU_{33}: that the Constitution gives him the power

…

Experiment

Dataset:

RST Discourse Treebank

- 380 discourses
  - **312 training, 30 validation, 38 testing**
- 111 relation types for fine-grained
- 19 relation types for coarse-grained

Experiment

Dataset:

RST Discourse Treebank

- 380 discourses
  - **312 training, 30 validation, 38 testing**
- 111 relation types for fine-grained
- 19 relation types for coarse-grained

Evaluation metrics:

- UAS, LAS

Experiment(Cont.)

Method	UAS	LAS(Fine)	LAS(Coarse)
Perceptron	0.5422	0.3231	0.3777
Basic(word+POS)	0.5588	0.367	0.3985
Basic(word+POS+position)	**0.5933**	**0.3832**	**0.4305**
Main-full	0.6197	0.3947	0.4445
MST-full	0.7331	0.4309	0.4851

Position features provide useful **structural clues** to our parser

Experiment(Cont.)

Method	UAS	LAS(Fine)	LAS(Coarse)
Perceptron	0.5422	0.3231	0.3777
Basic(word+POS)	0.5588	0.367	0.3985
Basic(word+POS+position)	0.5933	0.3832	0.4305
Main-full	**0.6197**	**0.3947**	**0.4445**
MST-full	0.7331	0.4309	0.4851

Memory Network could **model the discourse cohesion info** such as lexical chains, topical infos so as to provide clues to our parser.

Experiment(Cont.)

Method	UAS	LAS(Fine)	LAS(Coarse)
Perceptron	0.5422	0.3231	0.3777
Basic(word+POS)	0.5588	0.367	0.3985
Basic(word+POS+position)	0.5933	0.3832	0.4305
Main-full	0.6197	0.3947	0.4445
MST-full	**0.7331**	**0.4309**	**0.4851**

MST-full (graph-based) can directly analyze the relationship between **any EDU pairs**

Conclusions & Future work

Conclusions:

We propose to utilize **memory networks** to model **discourse cohesion** automatically.

- Capture the **topic change** or **lexical chains** within a discourse

Conclusions & Future work

Conclusions:

We propose to utilize **memory networks** to model **discourse cohesion** automatically.

- Capture the **topic change** or **lexical chains** within a discourse

Improve the discourse parsing performance

Conclusions & Future work

Conclusions:

We propose to utilize **memory networks** to model **discourse cohesion** automatically.

- Capture the **topic change** or **lexical chains** within a discourse

Improve the discourse parsing performance

Future work:

Apply our method on the **graph-based** parsing system

Optimize memory network structure

Thanks