Reinforcement Learning for NLP

Advanced Machine Learning for NLP
Jordan Boyd-Graber

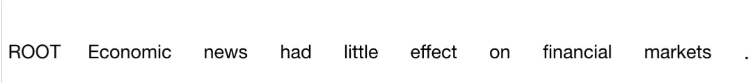
SHIFT-REDUCE PARSERS

Adapted from material by Jimmy Lin and Jason Eisner

Shift-Reduce Parsing

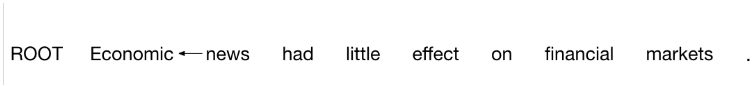
- Alternative to arc-factored models
- Cognitively plausible
- Better at short-range dependencies

Example



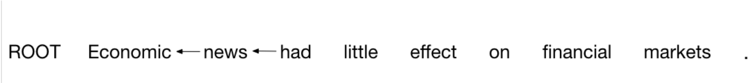
ROOT Economic news had little effect on financial markets .

Example



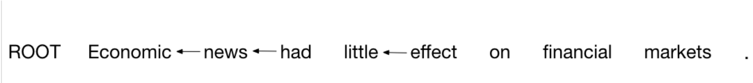
ROOT Economic news had little effect on financial markets .

Example



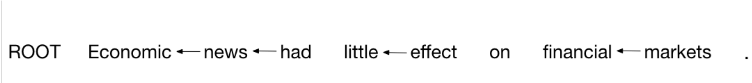
ROOT Economic ← news ← had little effect on financial markets .

Example



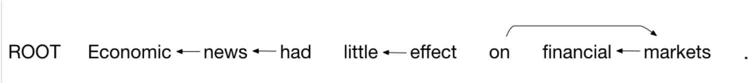
ROOT Economic ← news ← had little ← effect on financial markets .

Example

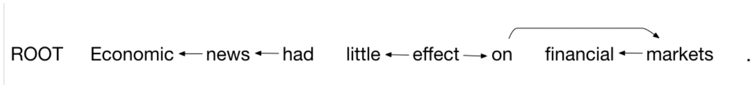


ROOT Economic ← news ← had little ← effect on financial ← markets .

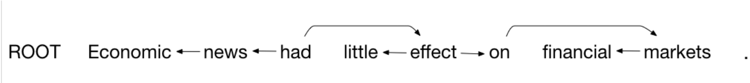
Example



Example

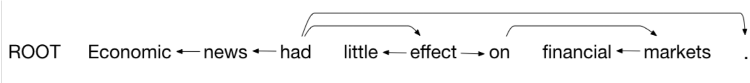


Example

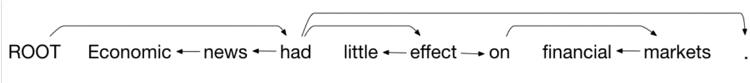


ROOT Economic news had little effect on financial markets .

Example



Example



Components

- Process a sentence word by word from a **buffer**
- You can temporarily place store words on a **stack**
- As you process you can either:

Components

- Process a sentence word by word from a **buffer**
- You can temporarily place store words on a **stack**
- As you process you can either:
  - *Shift*: Move a word from the buffer to the stack

Components

- Process a sentence word by word from a **buffer**
- You can temporarily place store words on a **stack**
- As you process you can either:
    - *Shift*: Move a word from the buffer to the stack
    - *Left*: The top of the stack is the child of the buffer's next word

Components

- Process a sentence word by word from a **buffer**
- You can temporarily place store words on a **stack**
- As you process you can either:
  - *Shift*: Move a word from the buffer to the stack
  - *Left*: The top of the stack is the child of the buffer's next word
  - *Right*: The buffer's next word is the child of the top of the stack

Initial and Final Conditions

- Initially the stack has root, the buffer has the sentence’s words, and there are no edges
- At the end, the buffer must be empty

Action: Left

- Add an edge (w_j, w_i)
- w_i is the top of the stack
- w_j is the first word of the buffer
- Pop the stack

Action: Left

- Add an edge (w_j, w_i)
- w_i is the top of the stack
- w_j is the first word of the buffer
- Pop the stack
- Stack and buffer must be non-empty; w_i cannot be the root

Action: Right

- Add an edge (w_i, w_j)
- w_i is the top of the stack
- w_j is the first word in the buffer
- Pop the stack
- Replace w_j by w_i at the head of buffer

Action: Right

- Add an edge (w_i, w_j)
- w_i is the top of the stack
- w_j is the first word in the buffer
- Pop the stack
- Replace w_j by w_i at the head of buffer
- Stack and buffer must be non-empty

Shift

- Removes w_i from the buffer
- Places it on the stack

Shift

- Removes w_i from the buffer
- Places it on the stack
- Buffer must be non-empty

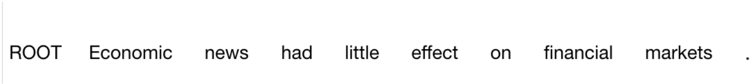
Shift Reduce Example

Stack

[root]

Buffer

[economic, news, had, little, effect, on, financial, markets, .]



Next action: 1. Shift

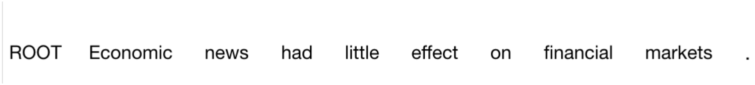
Shift Reduce Example

Stack

[root , economic]

Buffer

[news, had, little, effect, on, financial, markets, .]



ROOT Economic news had little effect on financial markets .

Next action: 2. Left

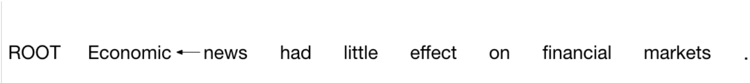
Shift Reduce Example

Stack

[root]

Buffer

[news, had, little, effect, on, financial, markets, .]



Next action: 3. Shift

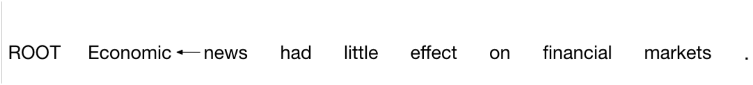
Shift Reduce Example

Stack

[root , news]

Buffer

[had, little, effect, on, financial, markets, .]



Next action: 4. Left

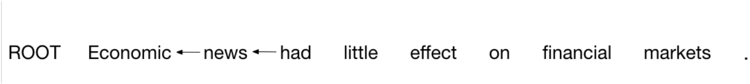
Shift Reduce Example

Stack

[root]

Buffer

[had, little, effect, on, financial, markets, .]



Next action: 5. Shift

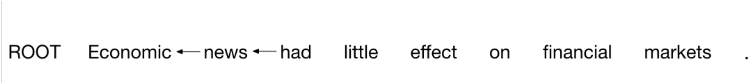
Shift Reduce Example

Stack

[root , had]

Buffer

[little, effect, on, financial, markets, .]



Next action: 6. Shift

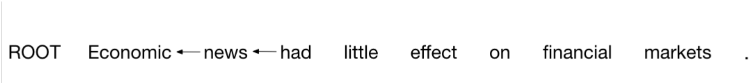
Shift Reduce Example

Stack

[root , had , little]

Buffer

[effect, on, financial, markets, .]



ROOT Economic ← news ← had little effect on financial markets .

Next action: 7. Left

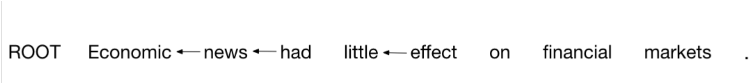
Shift Reduce Example

Stack

[root , had]

Buffer

[effect, on, financial, markets, .]



ROOT Economic ← news ← had little ← effect on financial markets .

Next action: 8. Shift

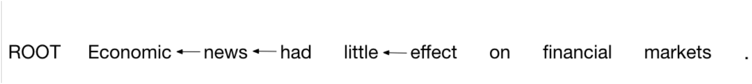
Shift Reduce Example

Stack

[root , had , effect]

Buffer

[on, financial, markets, .]



ROOT Economic ← news ← had little ← effect on financial markets .

Next action: 9. Shift

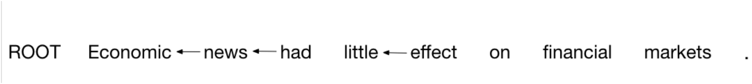
Shift Reduce Example

Stack

[root , had , effect , on]

Buffer

[financial, markets, .]



ROOT Economic ← news ← had little ← effect on financial markets .

Next action: 10. Shift

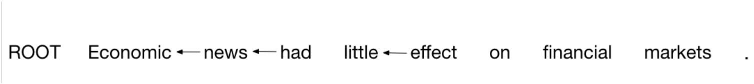
Shift Reduce Example

Stack

[root , had , effect , on , financial]

Buffer

[markets, .]



ROOT Economic ← news ← had little ← effect on financial markets .

Next action: 11. Left

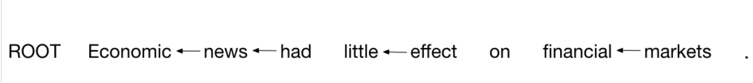
Shift Reduce Example

Stack

[root , had , effect , on]

Buffer

[markets, .]



Next action: 12. Right

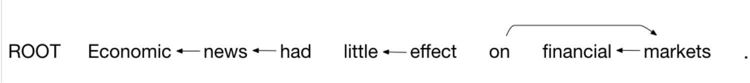
Shift Reduce Example

Stack

[root , had , effect]

Buffer

[on, .]



Next action: 13. Right

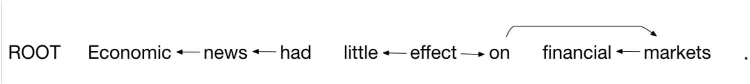
Shift Reduce Example

Stack

[root , had]

Buffer

[effect, .]



Next action: 14. Right

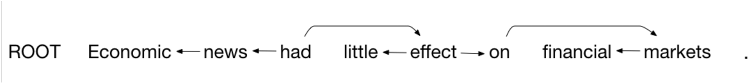
Shift Reduce Example

Stack

[root]

Buffer

[had, .]



Next action: 15. Shift

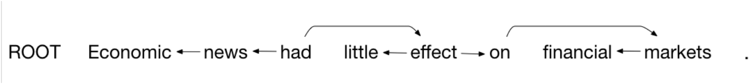
Shift Reduce Example

Stack

[root , had]

Buffer

[.]



Next action: 16. Right

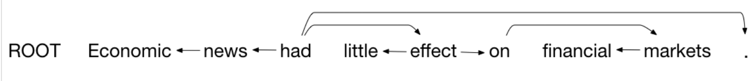
Shift Reduce Example

Stack

[root]

Buffer

[had]



Next action: 17. Right

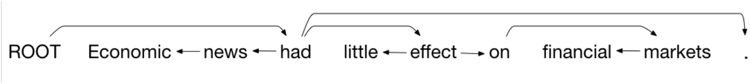
Shift Reduce Example

Stack

[ ]

Buffer

[root]



Next action: 18. Shift

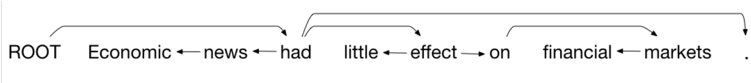
Shift Reduce Example

Stack

[root]

Buffer

[]



Next action:

Transition Sequence Algorithm

- Start with root on stack, buffer with whole sentence
- If there’s nothing on the stack, you must *shift*
- If the top of the stack is the child of the top of the buffer, then make a *left* edge
- If the top of the buffer is is a child of the top of the stack and the top of the buffer has no children that have yet to be added to the tree, then make a *right*

How to apply to data

- Create oracle for all sentences
- Create three-way classifier for each possible actions
- Features
  - The top of the stack
  - Top two words on buffer
  - The parts of speech of the words

Complexity

Complexity

- A word can only enter the stack once
- So complexity is $O(2N)$

Stack

[root]

Buffer

[I, am, the, very, model, of, a, modern, major, general]

Edges

Next action: 1. Shift

Stack

[root , I]

Buffer

[am, the, very, model, of, a, modern, major, general]

Edges

Next action: 2. Left

Stack

[root]

Buffer

[am, the, very, model, of, a, modern, major, general]

Edges

, I ← am

Next action: 3. Shift

Stack

[root , am]

Buffer

[the, very, model, of, a, modern, major, general]

Edges

, I ← am

Next action: 4. Shift

Stack

[root , am , the]

Buffer

[very, model, of, a, modern, major, general]

Edges

, I ← am

Next action: 5. Shift

Stack

[root , am , the , very]

Buffer

[model, of, a, modern, major, general]

Edges

, I ← am

Next action: 6. Left

Stack

[root , am , the]

Buffer

[model, of, a, modern, major, general]

Edges

, I ← am

, very ← model

Next action: 7. Left

Stack

[root , am]

Buffer

[model, of, a, modern, major, general]

Edges

, I ← am

, very ← model

, the ← model

Next action: 8. Shift

Stack

[root , am , model]

Buffer

[of, a, modern, major, general]

Edges

, I ← am

, very ← model

, the ← model

Next action: 9. Shift

Stack

[root , am , model , of]

Buffer

[a, modern, major, general]

Edges

, I ← am

, very ← model

, the ← model

Next action: 10. Shift

Stack

[root , am , model , of , a]

Buffer

[modern, major, general]

Edges

, I ← am

, very ← model

, the ← model

Next action: 11. Shift

Stack

[root , am , model , of , a , modern]

Buffer

[major, general]

Edges

, I ← am

, very ← model

, the ← model

Next action: 12. Shift

Stack

[root , am , model , of , a , modern , major]

Buffer

[general]

Edges

, I ← am

, very ← model

, the ← model

Next action: 13. Left

Stack

[root , am , model , of , a , modern]

Buffer

[general]

Edges

, I ← am

, very ← model

, the ← model

, major ← general

Next action: 14. Left

Stack

[root , am , model , of , a]

Buffer

[general]

Edges

, I ← am

, very ← model

, the ← model

, major ← general

, modern ← general

Next action: 15. Left

Stack

[root , am , model , of]

Buffer

[general]

Edges

, I ← am

, very ← model

, the ← model

, major ← general

, modern ← general

, a ← general

Next action: 16. Right

Stack

[root , am , model]

Buffer

[of,]

Edges

, I ← am

, very ← model

, the ← model

, major ← general

, modern ← general

, a ← general

, of → general

Next action: 17. Right

Stack

[root , am]

Buffer

[model,]

Edges

, I ← am

, very ← model

, the ← model

, major ← general

, modern ← general

, a ← general

, of → general

, model → of

Next action: 18. Right

Stack

[root]

Buffer

[am]

Edges

, I ⟵ am

, very ⟵ model

, the ⟵ model

, major ⟵ general

, modern ⟵ general

, a ⟵ general

, of ⟶ general

, model ⟶ of

, am ⟶ model

Next action: 19. Right

Stack

[ ]

Buffer

[root]

Edges

, I ⟵ am

, very ⟵ model

, the ⟵ model

, major ⟵ general

, modern ⟵ general

, a ⟵ general

, of ⟶ general

, model ⟶ of

, am ⟶ model

, root ⟶ am

Next action: 20. Shift

Stack

[root]

Buffer

[]

Edges

, I ← am

, very ← model

, the ← model

, major ← general

, modern ← general

, a ← general

, of → general

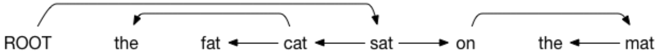
, model → of

, am → model

, root → am

Transition Sequence Algorithm

- Start with root on stack, buffer with whole sentence
- If there’s nothing on the stack, you must *shift*
- If the top of the stack is the child of the top of the buffer, then make a *left* edge
- If the top of the buffer is is a child of the top of the stack and the top of the buffer has no children that have yet to be added to the tree, then make a *right*



Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
S				

Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
s				
s				

Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
s				
s				
l	3	cat	2	fat

Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
s				
s				
l	3	cat	2	fat
l	3	cat	1	the

Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
s				
s				
l	3	cat	2	fat
l	3	cat	1	the
s				

Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
s				
s				
l	3	cat	2	fat
l	3	cat	1	the
s				
l	4	sat	3	cat

Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
s				
s				
l	3	cat	2	fat
l	3	cat	1	the
s				
l	4	sat	3	cat
s				

Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
s				
s				
l	3	cat	2	fat
l	3	cat	1	the
s				
l	4	sat	3	cat
s				
s				

Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
s				
s				
l	3	cat	2	fat
l	3	cat	1	the
s				
l	4	sat	3	cat
s				
s				
s				

Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
s				
s				
l	3	cat	2	fat
l	3	cat	1	the
s				
l	4	sat	3	cat
s				
s				
s				
l	7	mat	6	the

Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
s				
s				
l	3	cat	2	fat
l	3	cat	1	the
s				
l	4	sat	3	cat
s				
s				
s				
l	7	mat	6	the
r	5	on	7	mat

Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
s				
s				
l	3	cat	2	fat
l	3	cat	1	the
s				
l	4	sat	3	cat
s				
s				
s				
l	7	mat	6	the
r	5	on	7	mat
r	4	sat	5	on

Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
s				
s				
l	3	cat	2	fat
l	3	cat	1	the
s				
l	4	sat	3	cat
s				
s				
s				
l	7	mat	6	the
r	5	on	7	mat
r	4	sat	5	on
r	0	None	4	sat

Parse to Transition Sequence

Action	Head Index	Head Word	Dep Index	Dep Word
s				
s				
l	3	cat	2	fat
l	3	cat	1	the
s				
l	4	sat	3	cat
s				
s				
s				
l	7	mat	6	the
r	5	on	7	mat
r	4	sat	5	on
r	0	None	4	sat
s				