Carnegie Mellon University
Language Technologies Institute

11-411/11-611 Natural Language Processing

Discourse Cohesion and Coherence

David R. Mortensen

April 18, 2023

Language Technologies Institute

Please Complete FCEs!

Learning Objectives

- Define DISCOURSE COHESION and its relationship to DISCOURSE SEGMENTATION
- Reproduce an algorithm for discourse segmentation
- Define DISCOURSE COHERENCE and its relationship to DISCOURSE PARSING
- Recognize and justify RHETORICAL STRUCTURE THEORY analyses

Introduction

It bites.

It bites.

- What does this mean?

It bites.

- What does this mean?
- **In an informal dialog**: "it (perhaps a class someone has already mentioned) is very bad"

It bites.

- What does this mean?
- **In an informal dialog**: "it (perhaps a class someone has already mentioned) is very bad"
- **Near a dog or when a particular dog is the topic**: "(it) the dog is prone to biting"

Putting Things in Context

Consider

A: What do you think of of 15-112?

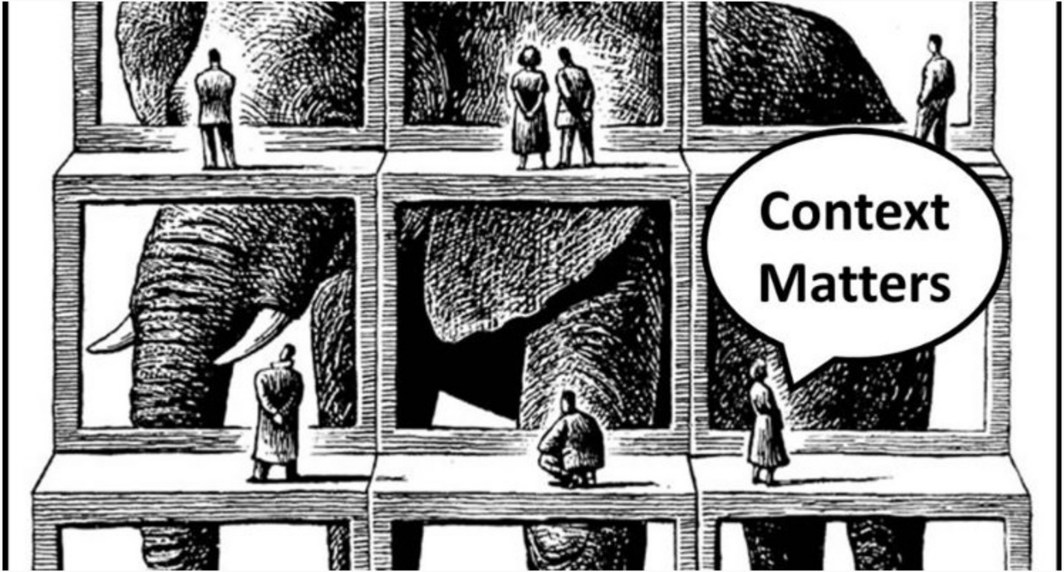
B: It bites.

versus

A: Your dog is so cute. Can I pet it?

B: It bites.

With Language as with Elephants, Context is Everything



Linguistic Context Is Important at All Levels

- Discourse context (where an utterance sits in relation to a document, conversation, speech, etc.)
- Physical context
- Social context
    - Who is speaking
    - Who they are speaking to
    - What they are trying to acheive

What is Discourse?

Discourse is the coherent structure of language above the level of sentences or clauses. A discourse is a coherent structured group of sentences.

What makes a passage coherent?

A practical answer: It has meaningful connections between its utterances.

Discourse Matters

Consider the following sentences:

- However, the incompetence of his managers insures him a steady, six-figure income.
- He only knows Java.
- Worse still, he always optimizes the outermost loop first.
- Eric is a pathetic programmer.

Discourse Matters

There are $4! = 24$ ways to permute these sentences, but only 2 of them are acceptable as discourses:

1. Eric is a pathetic programmer.
2. He only knows Java.
3. Worse still, he always optimizes the outermost loop first.
4. However, the incompetence of his managers insures him a steady, six-figure income.

and

1. Eric is a pathetic programmer.
2. However, the incompetence of his managers insures him a steady, six-figure income.
3. He only knows Java.
4. Worse still, he always optimizes the outermost loop first.

Contrast this, for example, with the following:

1. Eric is a pathetic programmer.
2. Worse still, he always optimizes the outermost loop first.
3. However, the incompetence of his managers insures him a steady, six-figure income.
4. He only knows Java.

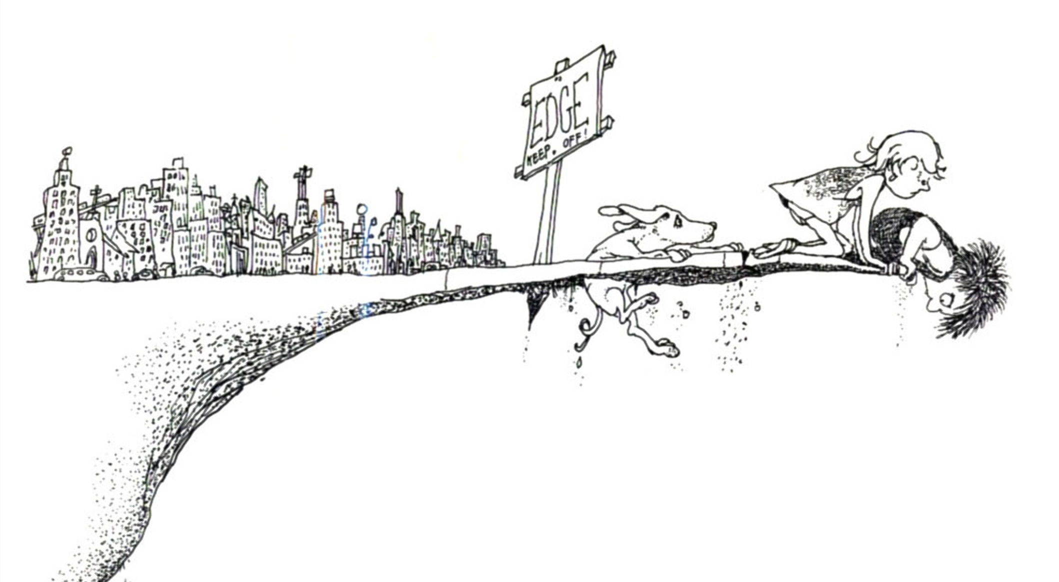
Or this:

1. However, the incompetence of his managers insures him a steady, six-figure income.
2. Worse still, he always optimizes the outermost loop first.
3. Eric is a pathetic programmer.
4. He only knows Java.

There Are Many Applications of Computational Discourse

- Automatic essay grading
- Automatic summarization
- Other forms of text generation
- Argument analysis
- Meeting understanding
- Dialogue systems
- …

The Place of Discourse in NLP



Cohesion and Coherence are Different, but Related, Concepts

cohesion The degree to which two passages of speech/text are "held together" by formal devices like shared words, coreference chains, and DISCOURSE MARKERS that indicate continuity or lack of continuity. **Form**.

coherence The degree to which two passages of text have a meaningful relationship (e.g., passage A explains passage B). **Meaning**.

Cohesion and Discourse Segmentation

Discourse Segmentation Is Dividing a Document into Topical Units

The discourse segmentation task:

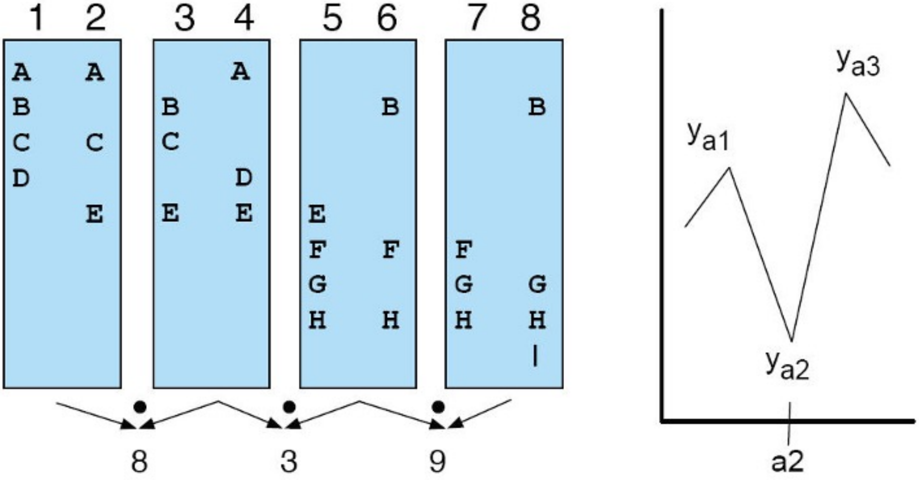
Input unsegmented text

Preprocessing text is divided into sentence-like units

Task Type sequence labeling—each unit is labeled as either the beginning of a new “paragraph” (etc.,) or the continuation of the existing paragraph

Output text segmented into topical units (paragraphs, sections, etc.,)

The TextTiling Algorithm for Discourse Segmentation



(a) (b)

Supervised Discourse Segmentation

- Our instances: place markers between sentences (or paragraphs or clauses)
- Our labels: yes (marker is a discourse boundary) or no (marker is not a discourse boundary)
- What features should we use?
  - Discourse markers or cue words
  - Word overlap before/after boundary
  - Number of coreference chains that cross boundary
  - Others?
- Segmentation can be done fairly well using only on formal criteria (i.e., no direct invocation of meaning).

Discourse Markers (Cue Words) in English

adding
: *also, moreover, furthermore, additionally, besides, in addition*

comparing
: *similarly, likewise, in the same way,*

: *on the whole, in general, broadly speaking, as a rule, in most cases*

For showing cause and effect
: *therefore, thus, consequently, hence, as a result*

contrasting
: *however, although, whereas, despite this fact, on one hand, on the other hand, on the contrary, still, nonetheless, instead, alternatively, in contrast*

For indicating time
: *in the past, not so long ago, recently,*

sequencing
: *firstly, at first, first of all, in the first place, to begin with, in the beginning, once upon a time, secondly, thirdly, subsequently, earlier, meanwhile, later, afterwards*

emphasizing
: *above all, specially, in particular, specifically, as a matter of fact, more importantly*

repeating
: *again and again, over and over, once again, as stated*

For giving examples
: *for example, for instance, such as, namely, in other words*

concluding
: *in conclusion, finally, to sum it up, in the end, lastly, in short, eventually*

Exercise: Divide this Passage into Paragraphs using Features for Cohesion

Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Identify the Lexical Chains

Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Identify the Lexical Chains (Key)

Ping **worked** hard to complete her last *assignment* and did her own **work**. But when she got *homework* scores back, she was very surprised. Because another student had <u>similar</u> **answers**, she was accused of CHEATING. It turns out that the other student was in the <u>same</u> program. He had copied down the answers during a group study ***session***. The ***session*** ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the CHEATER* instead.

Identify the Coreference Chains

Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Identify the Coreference Chains (Key)

Ping worked hard to complete **her** last assignment and did **her** own work. But when **she** got homework scores back, **she** was very surprised. Because *another student* had similar answers, **she** was accused of cheating. It turns out that *the other student* was in the same program. *He* had copied down the answers during ***a group study session***. ***The session*** ran late, and **Ping*** had left her computer in the room when she went to the restroom. Now **Ping*** was distraught, but the professors listened to **her** story. They realized that **she** was telling the truth and punished *the cheater* instead.

Identify the Discourse Markers

Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Identify the Discourse Markers (Key)

Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Final Segmentation

Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. ¶ It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. ¶ Now Ping was distraught, but the professors listened to her story. They realized that she was1 telling the truth and punished the cheater instead.

Coherence and Discourse Parsing

Compare the following examples:

(1) a. John hid Bill's car keys. He was drunk.
    b. John hid Bill's car keys. He likes spinach.

Result Infer that the state or event asserted by S_0 causes or could cause the state or event asserted by S_1.
Lakshmi had an original thought. She was fired.

Explanation Infer that the state or event asserted by S_1 causes or could cause the state or event asserted by S_0.
Lakshmi was fired. She had had an original thought.

Parallel Infer $p(a_1, a_2, \ldots)$ from the assertion of S_0 and $p(b_1, b_2, \ldots)$ from the assertion of S_1, where a_i and b_i are similar, for all i.
Lakshmi was fired; Krishna was promoted.

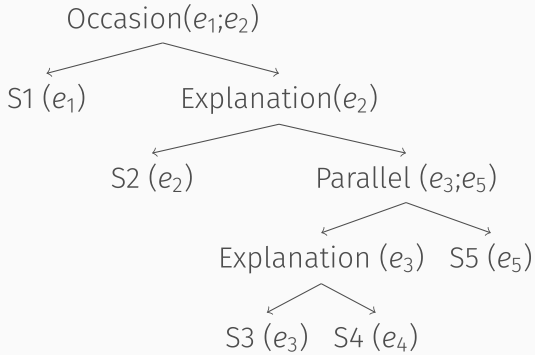
Elaboration Infer the same proposition P from the assertions of S_0 and S_1.
Lakshmi was fired. Her employment was terminated abruptly and without warning.

Occasion A change of state can be inferred from the assertion of S_0, whose final state can be inferred from S_1, or a change of state can be inferred from the assertion of S_1, whose initial state can be inferred from S.
Xudong opened a text editor. He starting pounding out a script similar to those he had written a thousand times before.

A Discourse for Analysis

John went to the bank to deposit his paycheck. (S1)
He then took a train to Bill's car dealership. (S2)
He needed to buy a car. (S3)
The company he works for now isn't near any public transportation. (S4)
He also wanted to talk to Bill about their softball league. (S5)

Discourse Parsing



John went to the bank to deposit his paycheck. (S1)
He then took a train to Bill's car dealership. (S2)
He needed to buy a car. (S3)
The company he works for now isn't near any public transportation. (S4)
He also wanted to talk to Bill about their softball league. (S5)

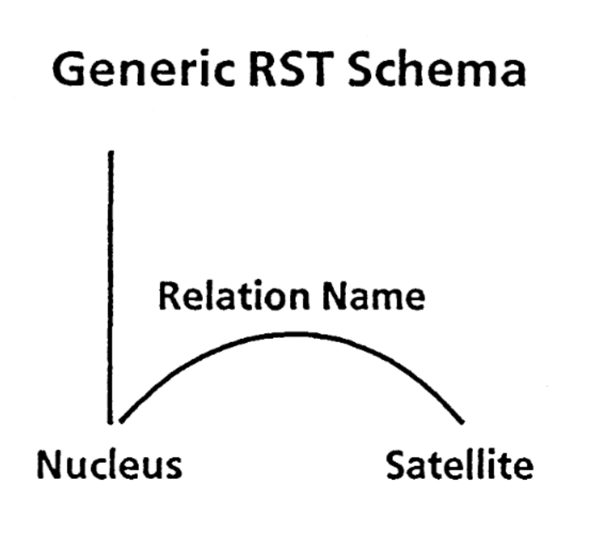
Rhetorical Structure Theory (RST)

Rhetorical Structure Theory is More Explicit Theory of Coherence

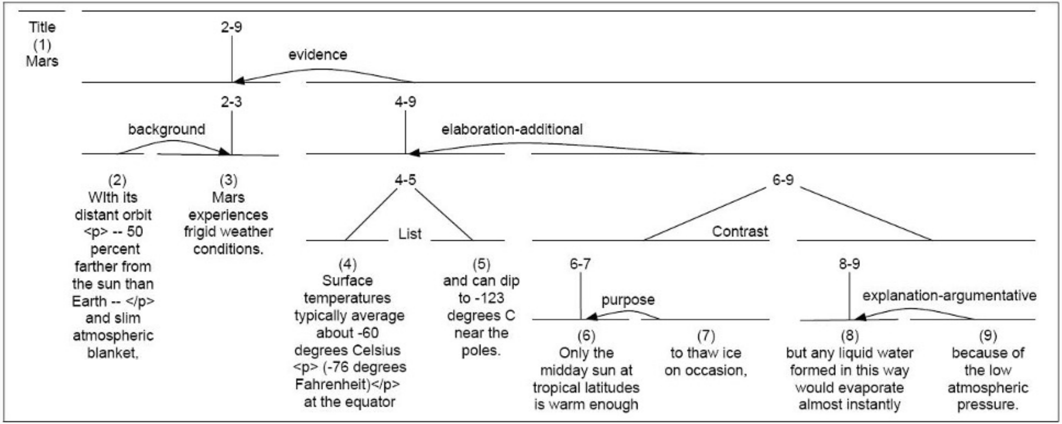
A more explicit theory of coherence

- Vague analogy: dependency grammar
- The passages more central to the discourse are nuclei
- The passages that depend on them are satellites
- The labeled arcs are relations
- Definitions of relations make reference to a READER and WRITER—the producer and the imagined consumer of the discourse

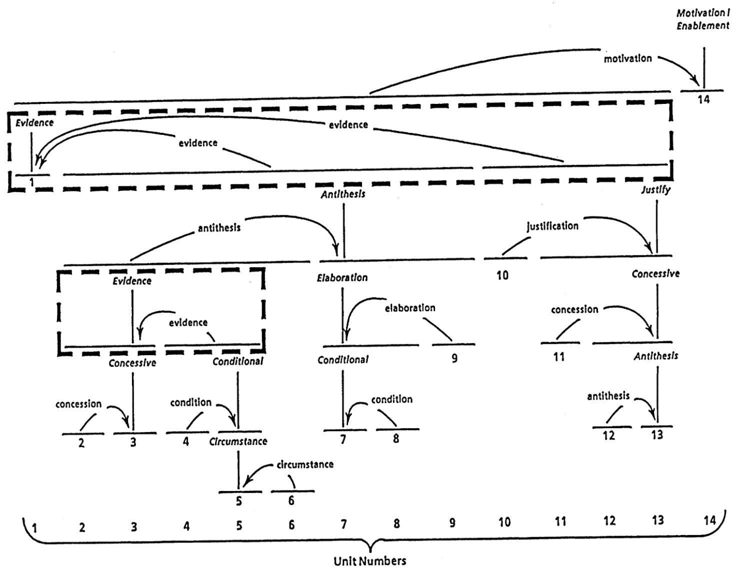
An RST Parse Illustrating Coherence Relations



An RST Parse Illustrating Coherence Relations



An RST Parse Illustrating Coherence Relations



A Formal Definition for an RST Relation Includes Constraints on Nucleus, Satellite, Reader, and Writer

Relation name: Evidence (N = nucleus, S = satellite, R = reader, W = writer)

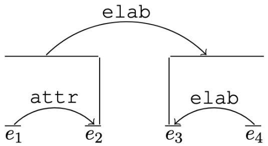
Constr on N: R not believing N enough for W

Constr on S: R believes S, or would

Constr on N+S: R's believing S would increase R's believing N

Effects: R's belief of N is increased

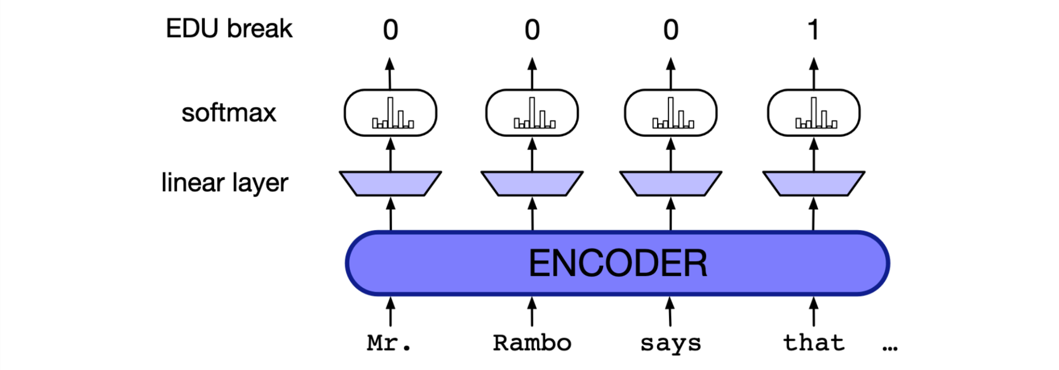
RST Parsing



e_1: American Telephone & Telegraph Co. said it
e_2: will lay off 75 to 85 technicians here , effective Nov. 1.
e_3: The workers install , maintain and repair its private branch exchanges,
e_4: which are large intracompany telephone networks.

Step	Stack	Queue	Action	Relation
1	$\varnothing$	e_1, e_2, e_3, e_4	SH	$\varnothing$
2	e_1	e_2, e_3, e_4	SH	$\varnothing$
3	e_1, e_2	e_3, e_4	RD(attr,SN)	$\varnothing$
4	$e_{1:\mathbf{2}}$	e_3, e_4	SH	$\widehat{e_1\mathbf{e_2}}$
5	$e_{1:\mathbf{2}}$, e_3	e_4	SH	$\widehat{e_1\mathbf{e_2}}$
6	$e_{1:\mathbf{2}}$, e_3, e_4	$\varnothing$	RD(elab,NS)	$\widehat{e_1\mathbf{e_2}}$
7	$e_{1:\mathbf{2}}$, $e_{\mathbf{3}:4}$	$\varnothing$	RD(elab,SN)	$\widehat{e_1\mathbf{e_2}}, \widehat{\mathbf{e_3}e_4}$
8	$e_{1:4}$	$\varnothing$	PR	$\widehat{e_1\mathbf{e_2}}, \widehat{\mathbf{e_3}e_4}, \widehat{e_{1:2}\mathbf{e_{3:4}}}$

Extract Elementary Discourse Units (EDUs)



Encoder-decoder Architecture for Shift-Reduce RST Parsing

Encoder represents the input span of words and ELEMENTARY DISCOURSE UNITS using a hierarchical BiLSTM.

- Layer 1: words inside an EDU
- Layer 2: EDU sequence

Input: sentences $w_1, w_2, \ldots, w_m$ represented as embeddings $\mathbf{x}_1^w, \mathbf{x}_2^w, \ldots, \mathbf{x}_m^w$

Word level biLSTM is applied, yielding a sequence of $\mathbf{h}^w$ values:

$$\mathbf{h}_1^w, \mathbf{h}_2^w, \ldots, \mathbf{h}_m^w, = \text{biLSTM}(\mathbf{x}_1^w, \mathbf{x}_2^w, \ldots, \mathbf{x}_m^w 1) \quad (1)$$

Consequently, an EDU with the span $w_s, w_{s+1}, \ldots, w_t$ has the biLSTM output representation $\mathbf{h}_s^w, \mathbf{h}_{s+1}^w, \ldots, \mathbf{h}_t^w$,

Encoder-decoder Architecture for Shift-Reduce RST Parsing

The representation $\mathbf{h}^w_s, \mathbf{h}^w_{s+1}, \ldots, \mathbf{h}^w_t$, is then mean-pooled:

$$\mathbf{x}^e = \frac{1}{t-s+1} \sum_{k=s}^{t} \mathbf{h}^w_k \tag{2}$$

This input is then used by the second layer to compute $\mathbf{h}^e$, the final representation of the sequence of EDU representations:

$$\mathbf{h}^e_1, \mathbf{h}^e_2, \ldots \mathbf{h}^e_n = \mathrm{biLSTM}(\mathbf{x}^e_1, \mathbf{x}^e_2, \ldots, \mathbf{x}^e_n,) \tag{3}$$

This is passed to the decoder.

The decoder is a FFNN $\mathbf{W}$ that computes o based on a concatenation of the top three subtrees on the stack (s_o, s_1, s_2) plus the first EDU on the queue (q_0):

$$\mathbf{o} = \mathbf{W}(\mathbf{h}^t_{s_0}, \mathbf{h}^t_{s_1}, \mathbf{h}^t_{s_2}, \mathbf{h}^e_{q_0},) \tag{4}$$

- $\mathbf{h}^e_{q_0}$ comes directly from the encoder
- $\mathbf{h}^t_{s_0}$, $\mathbf{h}^t_{s_1}$, and $\mathbf{h}^t_{s_2}$ are computed via mean pooling over encoder output for the EDUs in trees $\mathbf{h}^e_{s_0}$, $\mathbf{h}^e_{s_1}$, and $\mathbf{h}^e_{s_2}$

$$\mathbf{h}^t_s = \frac{1}{j - i + 1} \sum_{k=1}^{j} \mathbf{h}^e_k \tag{5}$$

Training a Shift-Reduce RST Parser

1. Map each tree in the training data to an oracle sequence of operations (as with transition-based dependency parsing)
2. Use standard cross entropy loss (with L_2 regularization) to train the system to take actions of the same kind

Given a state S and an oracle action a, compute o (the hypothesized action), then apply softmax to compute probabilities:

$$p_a = \frac{\exp(\mathbf{o}_a)}{\sum_{d' \in A} \exp(\mathbf{o}_{a'})} \tag{6}$$

We can then compute the cross-entropy loss:

$$L_{CE} = -\log(p_a) + \frac{\lambda}{2} ||\Phi||^2 \tag{7}$$

and back-propagate and optimize following the typical pattern.

Please Complete FCEs!

Questions?