

Natural Language Processing for Law and Social Science

11. Local Semantics

Outline

Local Semantics

Reading Comprehension

Textual Entailment

Dialogue Systems

Sequence-to-Sequence Transformers

Group Activity

Local vs global semantics

- ▶ **local semantics** get at linguistic information/relations at the level of sentences or paragraphs.
- ▶ e.g.
 - ▶ producing sentence embeddings
 - ▶ computing sentiment of a sentence
 - ▶ running a parser on a sentence
- ▶ **global semantics** are at the level of documents longer than paragraphs:
- ▶ e.g.:
 - ▶ TF-IDF frequencies of a document
 - ▶ co-reference resolution across paragraphs
 - ▶ information extraction from a whole corpus

Sentiment is a Local Semantic Dimension

- ▶ subjective vs objective statements
 - ▶ positive, negative, neutral
 - ▶ emotions (angry, happy, sad, etc)
 - ▶ stance (whether entity A favors or disfavors entity B)
-
- ▶ these variables could depend on the context of nearby sentences, but not far-off sentences.

Co-Reference Resolution

Finding all expressions that refer to the same entity in a text.

“My sister has a cat. Her name is Roberta.”



[Cat's] name is Roberta \leftrightarrow [Sister's] name is Roberta

Co-Reference Resolution

Finding all expressions that refer to the same entity in a text.

“My sister has a cat. Her name is Roberta.”



[Cat's] name is Roberta ↔ [Sister's] name is Roberta

<https://demo.allennlp.org/coreference-resolution>

The legal pressures facing 0 Michael Cohen are growing in a wide - ranging investigation of 0 his personal business affairs and 0 his work on behalf of 1 0 his former client , President Trump . In addition to 0 his work for 1 Mr. Trump , 0 he pursued 0 his own business interests , including ventures in real estate , personal loans and investments in taxi medallions .

Coherence

“My sister has a cat. Her name is Roberta.”



“My sister has a cat. She attended a conference.”

- ▶ **Coherent sentences** are connected by **structured relations** – e.g., Sentence 2 gives a reason for the action in Sentence 1.

Coherence

“My sister has a cat. Her name is Roberta.”



“My sister has a cat. She attended a conference.”

- ▶ **Coherent sentences** are connected by **structured relations** – e.g., Sentence 2 gives a reason for the action in Sentence 1.
- ▶ Related papers:
 - ▶ Xu et al (2019) use a skip-thought-type model, to predict whether sentences are adjacent, as a measure of coherence.
 - ▶ Nie et al (2019) use “discourse markers” (and, but, because, etc) to build a labeled dataset of discourse relations between sentences, then fine-tune BERT to predict discourse relations.

Outline

Local Semantics

Reading Comprehension

Textual Entailment

Dialogue Systems

Sequence-to-Sequence Transformers

Group Activity

Reading Comprehension \leftrightarrow Local Question Answering

Answering questions about a passage of text to show that the system understands the passage.

Reading Comprehension ↔ Local Question Answering

Answering questions about a passage of text to show that the system understands the passage.

<https://demo.allennlp.org/reading-comprehension>

Passage Context

The institutional framework of Navarre was preserved following the 1512 invasion. Once Ferdinand II of Aragon died in January, the Parliament of Navarre gathered in Pamplona, urging Charles V to attend a coronation ceremony in the town following tradition, but the envoys of the Parliament were met with the Emperor's utter indifference if not contempt. He refused to attend any ceremony and responded with a brief "let's say I am happy and pleases me." Eventually the Parliament met in 1517 without Charles V, represented instead by the **Duke of Najera** pronouncing an array of promises of little certitude, while the acting Parliament kept piling up grievances and demands for damages due to the Emperor, totalling 67—the 2nd Viceroy of Navarre Fadrique de Acuña was deposed in 1515 probably for acceding to send grievances. Contradictions inherent to the documents accounting for the Emperor's non-existent oath pledge in 1516 point to a contemporary manipulation of the records.

Question

Who represented the Charles V at Parliament?

Local Question Answering with BERT

Beyoncé Giselle Knowles-Carter (born September 4, 1981) is an American singer, songwriter, record producer and actress. Born and raised in **Houston, Texas**, she performed in various **singing and dancing** competitions as a child, and rose to fame in the late 1990s as lead singer of R&B girl-group Destiny's Child. Managed by her father, Mathew Knowles, the group became one of the world's best-selling girl groups of all time. Their hiatus saw the release of Beyoncé's debut album, *Dangerously in Love* (**2003**), which established her as a solo artist worldwide, earned five Grammy Awards and featured the Billboard Hot 100 number-one singles "Crazy in Love" and "Baby Boy".

Q: "In what city and state did Beyoncé grow up?"

A: "**Houston, Texas**"

Q: "What areas did Beyoncé compete in when she was growing up?"

A: "**singing and dancing**"

Q: "When did Beyoncé release *Dangerously in Love*?"

A: "**2003**"

Figure 23.11 A (Wikipedia) passage from the SQuAD 2.0 dataset (Rajpurkar et al., 2018) with 3 sample questions and the labeled answer spans.

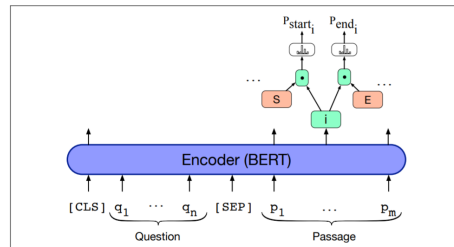


Figure 23.12 An encoder model (using BERT) for span-based question answering from reading-comprehension-based question answering tasks.

- BERT-based models learn to solve these problems by inputting {question + [SEP] + passage} and predicting the token indexes for the start and end of the answer.

Reading Comprehension with GPT-3

Stammbach, Antoniak, and Ash (2022)

Who is the villain in the following text?

Text: In questioning the use of hydraulic fracturing in New York to help produce natural gas, you do not note that the technology has been employed and continuously improved for more than 50 years and that studies by the Environmental Protection Agency and the Ground Water Protection Council have not identified a single instance of groundwater contamination. Wells where fracturing is used are specially constructed to protect drinking water sources. Regulatory oversight is extensive. The fluids mostly water that are forced into a well to create pressure to fracture rock are pushed back out by the oil and gas flowing upward for safe processing. Protecting our water supplies is important, as are reductions in greenhouse gas emissions through use of clean-burning natural gas. Banning hydraulic fracturing would be unwarranted and shortsighted, preventing production of large amounts of natural gas that could directly benefit New York consumers for decades and create thousands of good jobs.

Villain: The villain in this text is the person who is questioning the use of hydraulic fracturing in New York.

Outline

Local Semantics

Reading Comprehension

Textual Entailment

Dialogue Systems

Sequence-to-Sequence Transformers

Group Activity

Textual Entailment \leftrightarrow Natural Language Inference

- ▶ TE is the task of predicting whether, for a pair of sentences, the facts in the first sentence necessarily imply the facts in the second.

Sentence A (Premise)	Sentence B (Hypothesis)	Label
A soccer game with multiple males playing.	Some men are playing a sport.	entailment
An older and younger man smiling.	Two men are smiling and laughing at the cats playing on the floor.	neutral
A man inspects the uniform of a figure in some East Asian country.	The man is sleeping.	contradiction

- ▶ The SNLI (Stanford Natural Language Inference) dataset contains 570k human-written English sentence pairs manually labeled (by Amazon Mechanical Turk Workers) for balanced classification with the labels: entailment, contradiction, neutral.

<https://demo.allennlp.org/textual-entailment>

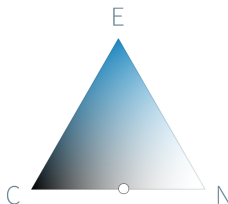
It is somewhat likely that there is no correlation between the premise and hypothesis.

Premise

A handmade djembe was on display at the Smithsonian.

Hypothesis

Visitors could not hear the djembe.



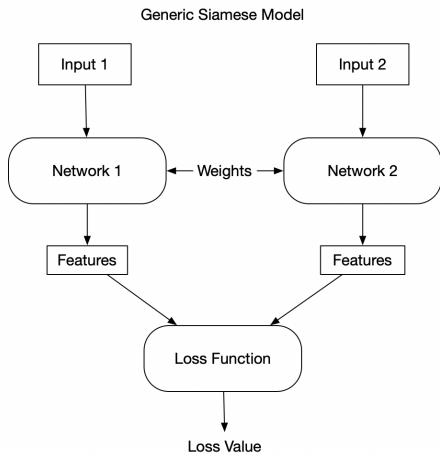
Judgement	Probability
Entailment	0.7%
Contradiction	46.4%
Neutral	52.9%

Sentence-BERT

- ▶ The document embeddings produced by BERT do not perform well for sentence similarity tasks.
- ▶ S-BERT (Reimers and Gurevych 2019):
 - ▶ fine-tune BERT embeddings to classify sentence pairs in textual entailment task.
 - ▶ significantly improves performance of sentence embeddings on standard tasks.

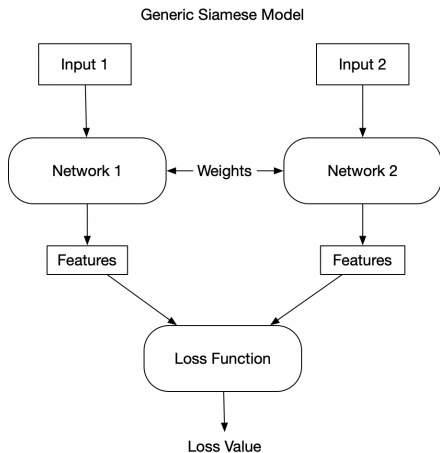
Siamese Networks (more negative sampling)

- ▶ A Siamese Neural Network (SNN) is a class of neural network architectures that contain two or more identical subnetworks (shared architecture and parameters).



Siamese Networks (more negative sampling)

- ▶ A Siamese Neural Network (SNN) is a class of neural network architectures that contain two or more identical subnetworks (shared architecture and parameters).



Standard SNN is similar to StarSpace:

1. start with an “anchor” document
2. take one positive sample (document from the same class) and k negative samples (documents from randomly chosen class)
3. send all of them through the same set of hidden layers to produce embeddings
4. compute **contrastive loss** that rewards high similarity between anchor and positive sample, and rewards low similarity between the anchor and the negative samples

S-BERT Training Objective

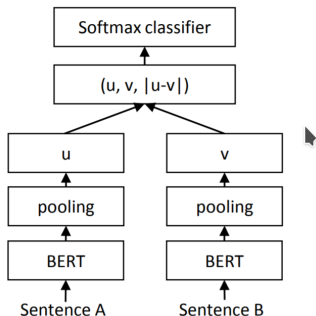


Figure 1: SBERT architecture with classification objective function, e.g., for fine-tuning on SNLI dataset. The two BERT networks have tied weights (siamese network structure).

S-BERT SNN:

1. send sentences through two siamese BERT networks (with shared weights)
2. produce document embeddings
3. use each embedding, and their distance, as features to predict labels in the Stanford Natural Language Inference dataset (entail, contradict, neutral)

- Then at inference time, form embeddings and then compute cosine similarity between sentences.

SentenceTransformers

- ▶ SentenceTransformers (sbert.net) is an amazing python package for embedding texts or short documents.
- ▶ Initially based on S-BERT but expanded to many additional models, including embeddings trained on other tasks besides entailment:
 - ▶ paraphrase identification
 - ▶ semantic textual similarity
 - ▶ duplicate question detection
 - ▶ question-answer retrieval
- ▶ monolingual and multilingual models (for over 100 languages)

Outline

Local Semantics

Reading Comprehension

Textual Entailment

Dialogue Systems

Sequence-to-Sequence Transformers

Group Activity

Chatbots

- ▶ Chatbots have been around since the 1960s, and a system from 1968 already passed the Turing test.
- ▶ Rule-based systems use pattern-action rules (e.g. ELIZA, rule-based therapy chatbot).

function ELIZA GENERATOR(*user sentence*) **returns** *response*

Find the word *w* in *sentence* that has the highest keyword rank

if *w* exists

Choose the highest ranked rule *r* for *w* that matches *sentence*

response \leftarrow Apply the transform in *r* to *sentence*

if *w* = 'my'

future \leftarrow Apply a transformation from the 'memory' rule list to *sentence*

Push *future* onto memory stack

else (no keyword applies)

either

response \leftarrow Apply the transform for the NONE keyword to *sentence*

or

response \leftarrow Pop the top response from the memory stack

return(*response*)

Neural Chatbots

- ▶ Modern chatbots are text-generating transformers that use the history of the conversation as context and then generate the next response
- ▶ Train on:
 - ▶ movie dialogue databases
 - ▶ Twitter conversations
 - ▶ etc
- ▶ Some special features needed to make conversations more creative/interesting, or to allow the chatbot to learn new facts.
 - ▶ e.g. adversarial training / reinforcement learning.

Meena (Google Chatbot, 2019)

Meena is similar to GPT-2:

- ▶ 13 stacked decoder blocks (2.6B parameters)
- ▶ trained on 341GB of text (IT helpdesk chats and movie conversations)

Human: I like how cute cows are

Meena: Ah, interesting. You like how they are cute, right?

Human: Yeah, and they're really smart.

Meena: I heard they go to college

Human: Cows go to college?

Meena: I heard that a cow went to Harvard.

Human: What did the cow study?

Meena: Bovine sciences.

Human: Do horses go to Harvard?

Meena: Horses go to Hayvard.

Human: That's a pretty good joke, I feel like you led me into it

Meena: You were trying to steer it elsewhere, I can see it

► **GPT-2 and T5 fine-tuned to produce Justice Ruth Bader Ginsburg's oral argument speech**

Mr. Chief Justice, and may it please the Court: The courts below made a simple error. They failed to understand where Employment Division versus Smith controls and where it doesn't. Smith doesn't control when the government uses a system of individualized exemptions or ...



Mr. Frederick, when you say partnership, do you mean a religious organization, or a secular organization?

I mean that the partnership between the City of Philadelphia and Catholic Social Services serves a public purpose. However, our contention is that this partnership does not infringe on rights provided by the Free Exercise Clause.

USER



So, you're saying that the only purpose of the exemption is to allow the religious organization to discriminate.

No, the individualized exemptions under Smith have many applications. Our contention is more narrow. When a reasonable accommodation can be made, as in this case, there is a burden on the government to accommodate a religious organization's free exercise if that accommodation is reasonable.

USER



Mr. Frederick, when you say reasonable, what about an accommodation that would allow a secular organization to discriminate, but not a religious organization?

Ethical issues in dialog systems

- ▶ Machine learning systems replicate biases that occurred in the training data.
- ▶ Could be a problem for chatbots that learn dynamically. e.g. Microsoft's Tay chatbot (Neff and Nagy 2016):
 - ▶ Went live on Twitter in 2016, taken offline 16 hours later
 - ▶ In that time it had started posting racial slurs, conspiracy theories, and personal attacks
 - ▶ Learned from user interactions

Outline

Local Semantics

Reading Comprehension

Textual Entailment

Dialogue Systems

Sequence-to-Sequence Transformers

Group Activity

Seq2Seq Transformers

- ▶ Last time we talked about autoregressive and autoencoding transformers.
- ▶ There is a third category, sequence-to-sequence, which rather than predicting the next word or reproducing the original sequence, has an objective of producing a new sequence.

Seq2Seq Transformers

- ▶ Last time we talked about autoregressive and autoencoding transformers.
- ▶ There is a third category, sequence-to-sequence, which rather than predicting the next word or reproducing the original sequence, has an objective of producing a new sequence.
- ▶ Useful, for example, for machine translation.

EasyNMT

- State-of-the-art machine translation with 3 lines of code
- Translation for 150+ languages
- Sentence & document translation
- Automatic language detection
- 4 pre-trained translation models:
 - opus-mt from Helsinki-NLP
 - mBART50 from Facebook AI Research
 - m2m_100 from Facebook AI Research (418M & 1.2B model)

```
#Install via: pip install -U easynmt
from easynmt import EasyNMT
model = EasyNMT('opus-mt')

#Translate a single sentence to German
print(model.translate('This is a sentence we want to translate to German', target_lang='de'))

#Translate several sentences to German
sentences = ['You can define a list with sentences.',
             'All sentences are translated to your target language.',
             'Note, you could also mix the languages of the sentences.']
print(model.translate(sentences, target_lang='de'))
```

BART, PEGASUS, and T5

These models use the machine-translation (encoder-decoder) transformer architecture, but both the input and output sequence are English.

BART, PEGASUS, and T5

These models use the machine-translation (encoder-decoder) transformer architecture, but both the input and output sequence are English.

- ▶ BART adds noise to input sequence (e.g. BERT-masking, deleting, permuting), then tries to produce output sequence that is the original sequence.

BART, PEGASUS, and T5

These models use the machine-translation (encoder-decoder) transformer architecture, but both the input and output sequence are English.

- ▶ BART adds noise to input sequence (e.g. BERT-masking, deleting, permuting), then tries to produce output sequence that is the original sequence.
- ▶ PEGASUS introduces Gap Sentence Generation:
 - ▶ input multiple sentences, mask out one sentence, and try to reconstruct the missing sentence.
 - ▶ works well as a summarization pre-training objective.

BART, PEGASUS, and T5

These models use the machine-translation (encoder-decoder) transformer architecture, but both the input and output sequence are English.

- ▶ BART adds noise to input sequence (e.g. BERT-masking, deleting, permuting), then tries to produce output sequence that is the original sequence.
- ▶ PEGASUS introduces Gap Sentence Generation:
 - ▶ input multiple sentences, mask out one sentence, and try to reconstruct the missing sentence.
 - ▶ works well as a summarization pre-training objective.
- ▶ T5 is a large model (11B parameters) pre-trained to solve a set of language tasks (SuperGLUE), where the input includes instructions (e.g. “Summarize this sentence: ...”).

Outline

Local Semantics

Reading Comprehension

Textual Entailment

Dialogue Systems

Sequence-to-Sequence Transformers

Group Activity

Group Activity