L90: Overview of Natural Language Processing

Lecture 12: Natural Language Generation

Weiwei Sun

Department of Computer Science and Technology
University of Cambridge

Michaelmas 2020/21

I have a question about whether you've been attempted to look at generation? [...] That is a rich rich area which so few people address [...]

Well, I find generation completely terrifying [...] I am very interested in the problem [...] That's an important question.

ACL lifetime archievement award lecture (`vimeo.com/288152682`)

Mark Steedman
FBA, FRSE

equally important to language understanding

Lecture 12: Natural Language Generation

1. Overview
2. Text summarization
3. Surface realisation
4. Evaluation

Overview

Generation from what?!

comprehension

natural language expressions $\mathcal{R}$

production

$\mathcal{R}$:
- morphological structure
- syntactic structure
- semantic structure
- discourse structure
- application-related structure

Generation from what?!

comprehension

natural language expressions $\mathcal{R}$

production

$\mathcal{R} \left\{ \begin{array}{l} \text{morphological structure} \\ \text{syntactic structure} \\ \text{semantic structure} \\ \text{discourse structure} \\ \text{application-related structure} \end{array} \right.$

[...] you can get away with incomplete semantics when you are doing parsing, but when you're doing generation, you have to specify everything in semantics. And we don't know how to do that. At least we don't know how to do that completely or properly.

Mark Steedman
FBA, FRSE

Generation from what?!

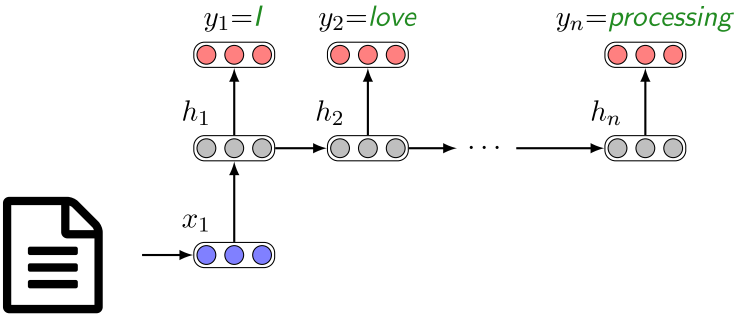
- **logical form**: inverse of (deep) (semantic) parsing.
  aka *surface realisation*
- **formally-defined data**: databases, knowledge bases, etc
- **semantic web ontologies**, etc
- **semi-structured data**: tables, graphs etc
- **numerical data**: weather reports, etc
- **cross-modal input**: image, etc
- **user input** (plus other data sources) in assistive communication.

generating from data often requires domain experts

Components of a classical generation system

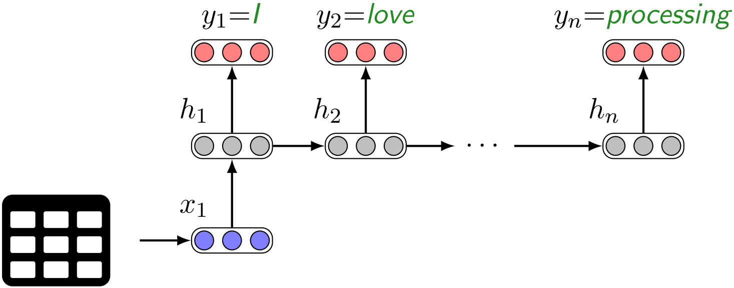
- *Content determination*: deciding what information to convey
- *Discourse structuring*: overall ordering, sub-headings etc
- *Aggregation*: deciding how to split information into sentence-sized chunks
- *Referring expression generation*: deciding when to use pronouns, which modifiers to use etc
- *Lexical choice*: which lexical items convey a given concept (or predicate choice)
- *Realization*: mapping from a meaning representation (or syntax tree) to a string (or speech)
- *Fluency ranking*

A typical framework for neural generation



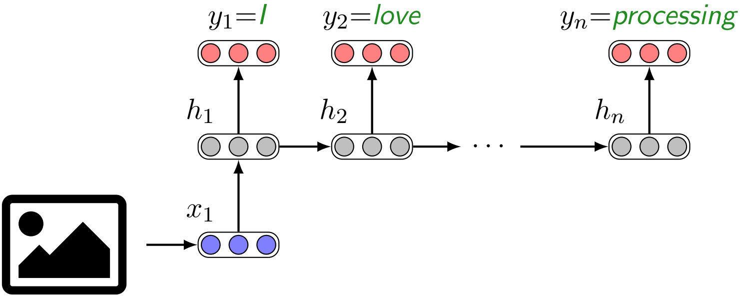
- Many different model designs.
- Need many examples of input and desired output.

A typical framework for neural generation



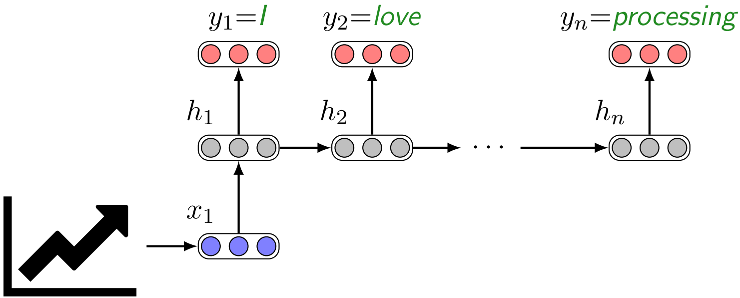
- Many different model designs.
- Need many examples of input and desired output.

A typical framework for neural generation



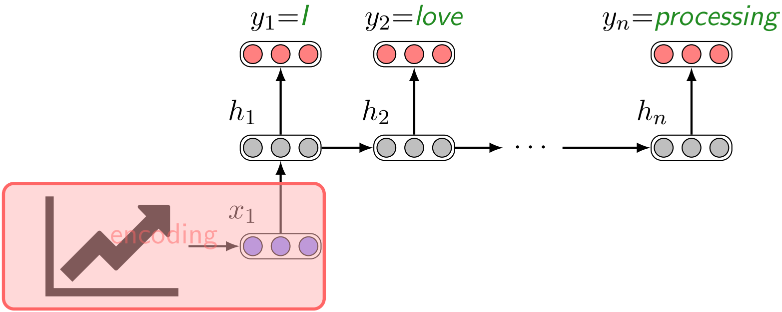
- Many different model designs.
- Need many examples of input and desired output.

A typical framework for neural generation



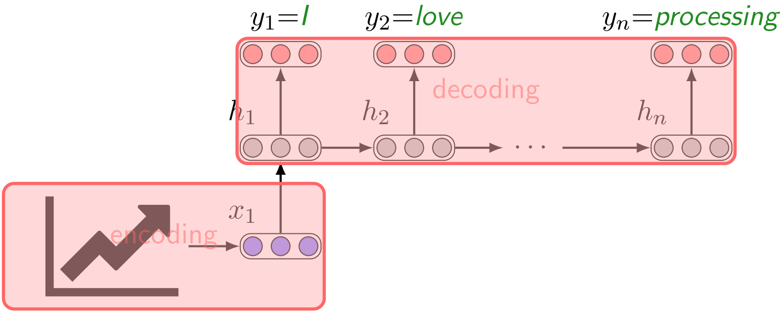
- Many different model designs.
- Need many examples of input and desired output.

A typical framework for neural generation



- Many different model designs.
- Need many examples of input and desired output.

A typical framework for neural generation



- Many different model designs.
- Need many examples of input and desired output.

NLG and me

- I am **NOT** an expert on NLG
- I **MAY** be considered an expert on "neural" NLP methods
- I **sometimes** say controversial things
- I **know enough** about NLG to identify when it is done wrong
- I think neural NLG methods are doing most things wrong

from Y Goldberg's talk

Approaches to generation

- Classical (limited domain): hand-written rules for first five steps, grammar for realization, grammar small enough that no need for fluency ranking (or hand-written rules).
- Templates: most practical systems. Fixed text with slots, fixed rules for content determination.
- Statistical (limited domain): components as above, but use machine learning (supervised or non-supervised).
- Neural (sequence-)to-sequence models.

Text Summarization

Regeneration: transforming text

- Text from partially ordered bag of words: statistical MT.
- Paraphrase
- Summarization (single- or multi-document)
- Wikipedia article construction from text fragments
- Text simplification

Also: mixed generation and regeneration systems, MT.

Overview of summarization

- Pure form of task: reduce the length of a document.
- Most used for search results, question answering etc: different scenarios have different requirements.
- Multidocument summarization: e.g., bringing together information from different news reports.
- Two main system types:

  Extractive: select sentences from a document. Possibly compress selected sentences.

  Abstractive: use partial analysis of the text to build a summary.

Extractive

If we consider a discourse relation as a relationship between two phrases, we get a binary branching tree structure for the discourse. In many relationships, such as Explanation, one phrase depends on the other: e.g., the phrase being explained is the main one and the other is subsidiary. In fact we can get rid of the subsidiary phrases and still have a reasonably coherent discourse.

Abstractive summarization with meaning representations

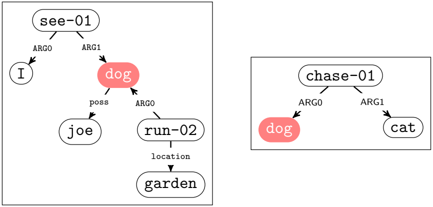
I saw Joe's dog, which was running in the garden.
The dog was chasing a cat.

Abstractive summarization with meaning representations

I saw Joe's dog, which was running in the garden.

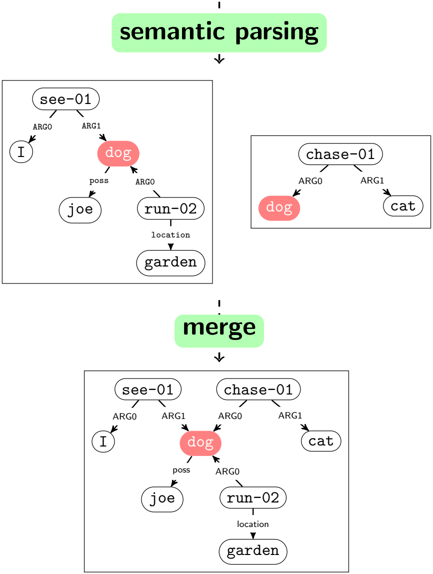
The dog was chasing a cat.

semantic parsing



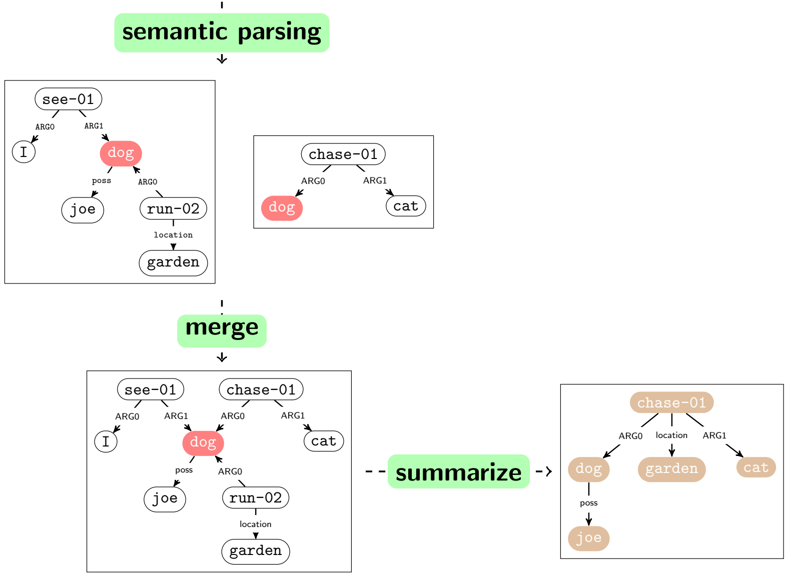
Abstractive summarization with meaning representations

I saw Joe's dog, which was running in the garden.
The dog was chasing a cat.



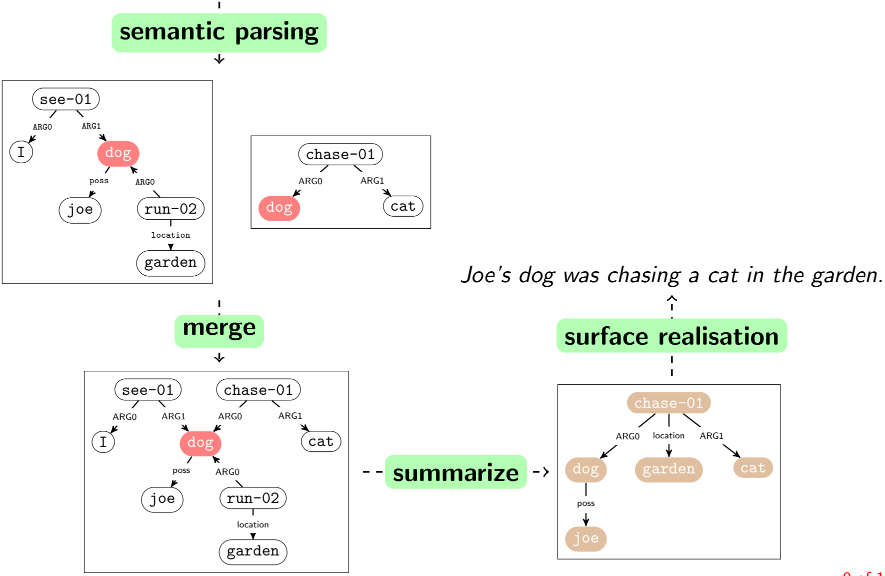
Abstractive summarization with meaning representations

I saw Joe's dog, which was running in the garden.
The dog was chasing a cat.



Abstractive summarization with meaning representations

I saw Joe's dog, which was running in the garden.
The dog was chasing a cat.



Abstractive summarization: Evaluation

Evaluation on Proxy Report section of AMRBank LCD2017T10.

AMRs	NLG model	ROUGE-1	ROUGE-2	ROUGE-L
gold	amr2seq + LM	40.4	20.3	31.4
	amr2seq	38.9	12.9	27.0
	amr2bow (Liu et al.)	39.6	6.2	22.1
RIGA	amr2seq + LM	42.3	21.2	33.6
	amr2seq	37.8	10.7	26.9
–	OpenNMT	36.1	19.2	31.1

Hardy and Vlachos, 2018

Surface Realisation

Modeling Syntactico-Semantic Composition

The Principle of Compositionality

The meaning of an expression is a function of **the meanings of its parts** *and of* **the way they are syntactically combined**.

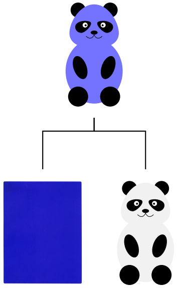
B. Partee

Modeling Syntactico-Semantic Composition

The Principle of Compositionality

The meaning of an expression is a function of **the meanings of its parts** *and of* **the way they are syntactically combined**.

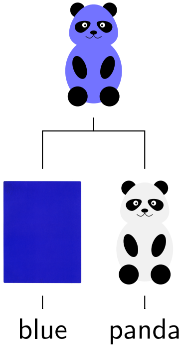
B. Partee



Modeling Syntactico-Semantic Composition

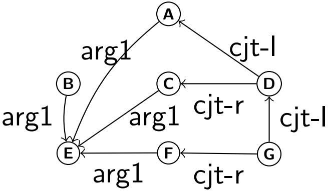
The Principle of Compositionality

> *The meaning of an expression is a function of* **the meanings of its parts** *and of* **the way they are syntactically combined**.
>
> *B. Partee*



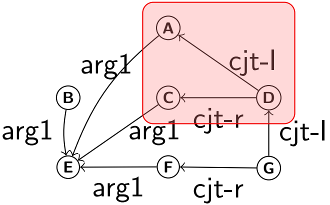
Parse a meaning representation

A dynamic programming algorithm (Chiang et al., 2013)



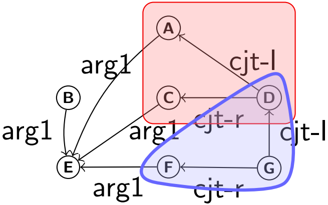
Parse a meaning representation

A dynamic programming algorithm (Chiang et al., 2013)



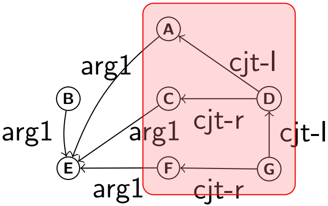
Parse a meaning representation

A dynamic programming algorithm (Chiang et al., 2013)



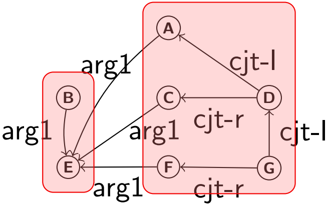
Parse a meaning representation

A dynamic programming algorithm (Chiang et al., 2013)



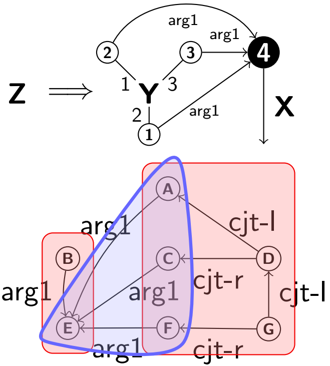
Parse a meaning representation

A dynamic programming algorithm (Chiang et al., 2013)



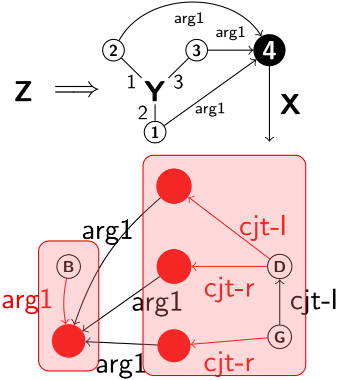
Parse a meaning representation

A dynamic programming algorithm (Chiang et al., 2013)



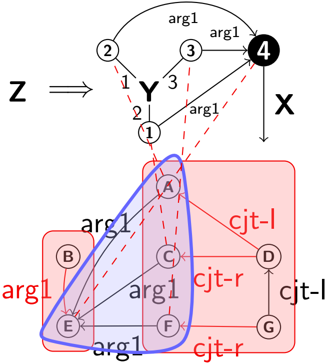
Parse a meaning representation

A dynamic programming algorithm (Chiang et al., 2013)



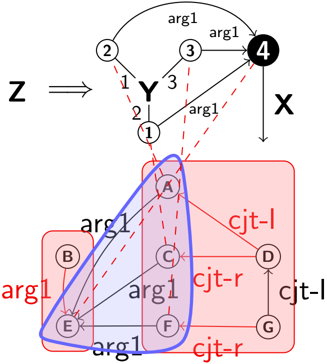
Parse a meaning representation

A dynamic programming algorithm (Chiang et al., 2013)



Parse a meaning representation

A dynamic programming algorithm (Chiang et al., 2013)



Evaluation

Tokenwise evaluation

complete match?

Tokenwise evaluation

complete match?

POS tagging

$$\frac{|\,\{\langle \text{word}, \text{tag}\rangle\}_{\text{system}} \cap \{\langle \text{word}, \text{tag}\rangle\}_{\text{gold}}\,|}{|\,\{\text{word}\}\,|}$$

Phrase structure parsing

$$\text{precision} = \frac{|\,\{\langle \text{left}, \text{right}, \text{category}\rangle\}_{\text{system}} \cap \{\langle \text{left}, \text{right}, \text{category}\rangle\}_{\text{gold}}\,|}{|\,\{\langle \text{left}, \text{right}, \text{category}\rangle\}_{\text{system}}\,|}$$

$$\text{recall} = \frac{|\,\{\langle \text{left}, \text{right}, \text{category}\rangle\}_{\text{system}} \cap \{\langle \text{left}, \text{right}, \text{category}\rangle\}_{\text{gold}}\,|}{|\,\{\langle \text{left}, \text{right}, \text{category}\rangle\}_{\text{gold}}\,|}$$

$$\mathsf{F}_\beta = (1+\beta^2) \times \frac{\text{precision} \times \text{recall}}{\beta^2\text{precision} + \text{recall}}$$

f-score: `en.wikipedia.org/wiki/F-score`

ROUGE

ROUGE-N: Overlap of N-grams between the system and *reference summaries*.

ROUGE-*L*: Longest Common Subsequence.

- A sequence $Z = [z_1, z_2, \ldots, z_k]$ is a subsequence of another sequence $X = [x_1, x_2, \ldots, x_m]$, if there exists a strict increasing sequence $[i_1, i_2, \ldots, i_k]$ of indices of X such that for all $j = 1, 2, \ldots, k$, we have $x_{i_j} = z_j$.
- The longest common subsequence (LCS) of X and Y is a common subsequence with maximum length.

Sentence-level LCS (X: reference):

$$\begin{aligned} \mathsf{R}_{\mathsf{lcs}} &= \frac{\#LCS(X,Y)}{\#X} \\ \mathsf{P}_{\mathsf{lcs}} &= \frac{\#LCS(X,Y)}{\#Y} \end{aligned}$$

Lin (2004): `www.aclweb.org/anthology/W04-1013.pdf`

Readings

- Ann's lecture notes.
  `https://www.cl.cam.ac.uk/teaching/1920/NLP/materials.html`

* Y Goldberg. Neural Language Generation. `https://inlg2018.uvt.nl/wp-content/uploads/2018/11/INLG2018-YoavGoldberg.pdf`