Natural Language Understanding, Generation, and Machine Translation

Lecture 26: Natural Language Generation: Data-to-Text Generation

Laura Perez-Beltrachini

(Slides by Mirella Lapata)

School of Informatics
University of Edinburgh
mlap@inf.ed.ac.uk

Outline

1. Introduction
2. Problem Formulation
   - Attention-Based LSTM
   - Decoding Algorithms
3. Automatic Evaluation
   - NLG-specific Metrics
   - Results

What is NLG?

(non-)linguistic input $\Longrightarrow$

$\Longrightarrow$ text

databases
news articles
log files
images

reports
help messages
summaries
captions

Today, we will look at how to generate a document from a database.

An Example: BasketBall

TEAM	WIN	LOSS	PTS	FG_PCT	RB	AST
Nuggets	8	15	85	41	45	24
Wizards	8	13	92	45	35	21

PLAYER	H/V	AST	RB	PTS	FG	CITY
Bradley Beal	H	5	3	26	8	Washington
John Wall	H	5	7	15	5	Washington
Markieff Morris	H	4	3	15	5	Washington
Nikola Jokic	V	1	11	17	6	Denver
Jusuf Nurkic	V	1	7	13	6	Denver
...	...	...	...	...	...	...

PTS: points, FT_PCT: free throw percentage, RB: rebounds, AST: assists, H/V: home or visiting, FG: field goals, CITY: player team city.

These two teams have really struggled this season and the 48 combined turnovers shows just that. **Denver** was the more inept team though, as they set an NBA season-high with 29 turnovers of their own. The shooting wasn't any better, as they shot **41** percent from the field, **67** percent from the free-throw line and **22** percent from three-point range. **Bradley Beal** was the best player in this game, tallying **26** points, **three** rebounds and **five** assists. **John Wall** went just **5**-of-**14** from the field, but still finished with **15** points, **seven** rebounds, **five** assists and **three** steals. **Markieff Morris** also scored **15** points, as all five starters played at least 33 minutes. **Nikola Jokic** was the only effective Nugget on offense, finishing with **17** points and **11** rebounds. **Wilson Chandler** had a game to forget, as he went just **2**-of-**16** from the field, en route to five points.

Another Example: Baseball

TEAM	Inn1	Inn2	Inn3	Inn4	...	R	H	E
Orioles	1	0	0	0	...	2	4	0
Royals	1	0	0	3	...	9	14	1

BATTER	H/V	AB	R	H	RBI	TEAM
C. Mullins	H	4	2	2	1	Orioles
J. Villar	H	4	0	0	0	Orioles
W. Merrifield	V	2	3	2	1	Royals
R. O'Hearn	V	5	1	3	4	Royals
...	...	...	...	...	...	

PITCHER	H/V	W	L	IP	H	R	ER	BB	K
A. Cashner	H	4	13	5.1	9	4	4	3	1
B. Keller	V	7	5	8.0	4	2	2	2	4
...	...	...	...	...	...				

Inn1: innings, R: runs, H: hits, E: errors,
AB: at-bats, RBI: runs-batted-in, H/V: home or visiting, W: wins, L: losses,
IP: innings pitched, ER: earned runs,
BB: walks, K: strike outs.

KANSAS CITY, Mo. – **Brad Keller** kept up his recent pitching surge with another strong outing. **Keller** gave up a home run to the first batter of the game – **Cedric Mullins** – but quickly settled in to pitch eight strong innings in the Kansas City **Royals**' 9–2 win over the Baltimore **Orioles** in a matchup of the teams with the worst records in the majors. **Keller** (7–5) gave up two runs and four hits with two walks and four strikeouts to improve to 3–0 with a 2.16 ERA in his last four starts. **Ryan O'Hearn** homered among his three hits and drove in four runs, **Whit Merrifield** scored three runs, and Hunter Dozier and **Cam Gallagher** also went deep to help the **Royals** win for the fifth time in six games on their current homestand. With the scored tied 1–1 in the fourth, **Andrew Cashner** (4–13) gave up a sacrifice fly to **Merrifield** after loading the bases on two walks and a single. **Dozier** led off the fifth inning with a 423-foot home run to left field to make it 3-1. The **Orioles** pulled within a run in the sixth when **Mullins** led off with a double just beyond the reach of **Dozier** at third, advanced to third on a fly ball and scored on **Trey Mancini**'s sacrifice fly to the wall in right.

The RotoWire and MLB Datasets

	RotoWire	MLB
		
Vocab Size	11.3K	38.9K
# Tokens	1.5M	14.3M
# Instances	4.9K	26.3K
Avg Length	337.1	542.05
# Record Types	39	53
Avg Records	628	565

Problem Formulation

- How will we model this generation problem?
  Again, it is a conditional language model, $p(y|s)$
- **Input**: a table of records s
- **Output**: document y containing words $y = y_1 \dots y_{|y|}$

Set of records $\{r_1, \dots, r_S\}$ let $s = \{r_j\}_{j=1}^{|s|}$ denote table of records
Type of record r such that $r.t = \text{PTS}$, $r.e = \text{Bradley Beal}$, $r.m = 26$

PLAYER	H/V	AST	RB	PTS	FG	CITY
Bradley Beal	H	5	3	26	8	Washington
John Wall	H	5	7	15	5	Washington
Markieff Morris	H	4	3	15	5	Washington
Nikola Jokic	V	1	11	17	6	Denver
Jusuf Nurkic	V	1	7	13	6	Denver
...	...	...	...	...	...	...

Record Encoder

Input is a table of unordered records, each represented as features: type of record, entity, value, player is on home- or away-team.

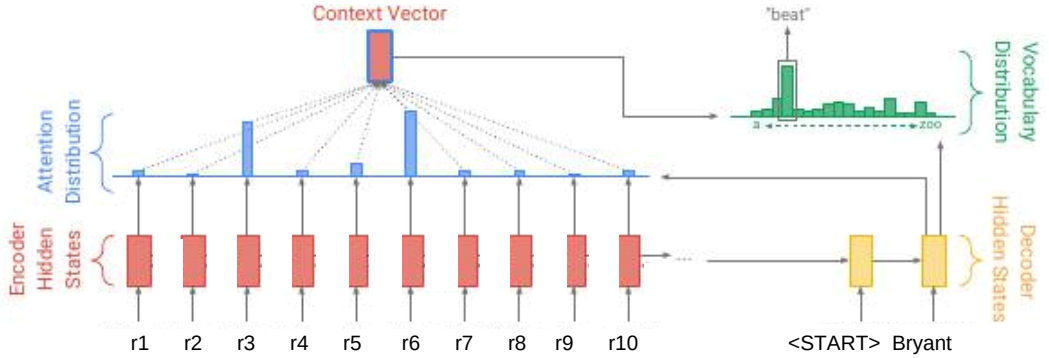
Features are embedded into **vectors** (each feature is a vector).

Multilayer perceptron yields vector representation of each record:

$$\mathbf{r_j} = \text{ReLU}(\mathbf{W}|[r_{j,1}; r_{j,2}; r_{j,3}; r_{j,4}] + \mathbf{b_r}$$

- The MLP is used to perform dimensionality reduction
- ReLU stands for rectified linear activation unit
- It returns the value provided as input or 0 if the input is 0 or less
- It encourages to learn sparse representations

Decoder: Attention-Based LSTM



$s_1, \ldots, s_S$	data record representations
$\mathbf{h_i}$	decoder hidden state
$p(a = j)\|\mathbf{s}, \mathbf{h_i}; \theta)$	attention distribution softmax$(s_i^\intercal h_i)$
$c = \mathbb{E}_a[s_a]$	context vector

Decoder: Attention-Based LSTM with Copying

Copy mechanism: additional binary variable z_t indicating whether target word is copied from the source or generated:

$$p(\hat{y}_t|\hat{y}_{1:t-1}, \mathbf{s}) = \sum_{z \in \{0,1\}} p(\hat{y}_t, z_t = z|\hat{y}_{1:t-1}, \mathbf{s})$$

$$p(\hat{y}_t, z_t|\hat{y}_{1:t-1}, \mathbf{s}) = \begin{array}{ll} p_{copy}(\hat{y}_t|z_t, \hat{y}_{1:t-1}, \mathbf{s})p(z_t|\hat{y}_{1:t-1}, \mathbf{s}) & z_t = 1 \\ p_{gen}(\hat{y}_t|z_t, \hat{y}_{1:t-1}, \mathbf{s})p(z_t|\hat{y}_{1:t-1}, \mathbf{s}) & z_t = 0 \end{array}$$

Words are copied from the value portion of a record.
Switch probability modeled as MLP with binary output:

$$\begin{aligned} p(z_t = 1|\hat{y}_{1:t-1}, \mathbf{s}) &= \sigma(f(\mathbf{s}, \mathbf{h_{t-1}}; \theta)) \\ p(z_t = 0|\hat{y}_{1:t-1}, \mathbf{s}) &= 1 - \sigma(f(\mathbf{s}, \mathbf{h_{t-1}}; \theta)) \end{aligned}$$

Decoding Algorithms

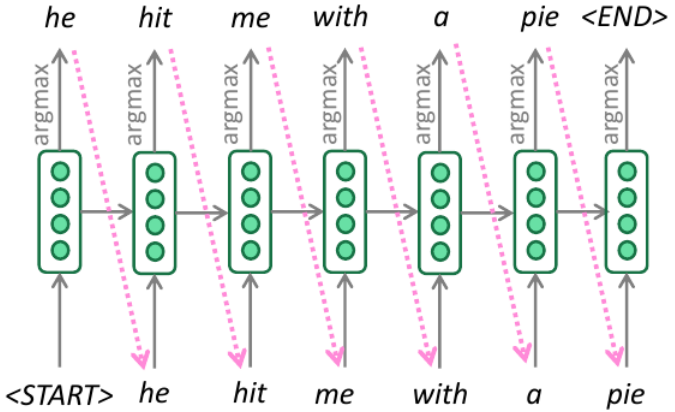
Question: Once you've trained your conditional language model, how do you use it to generate text?

Answer: A decoding algorithm is an algorithm you use to generate text from your language model

Two commonly used decoding algorithms:

- Greedy Decoding
- Beam Search

Greedy Decoding



- On each step, take the most probable word (i.e., argmax)
- Use that as the next word, and feed it as input on the next step
- Keep going until you produce <END> (or reach some max length)

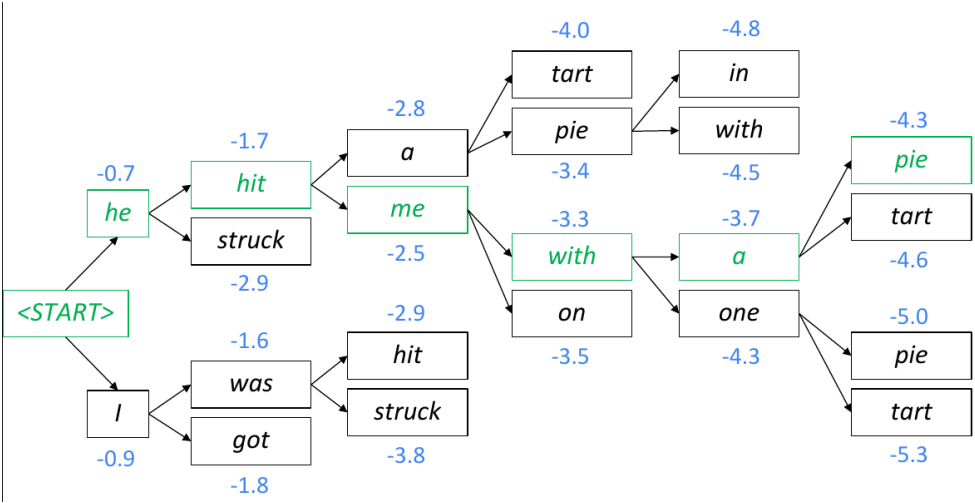
Beam Search Decoding

A search algorithm which aims to find a high-probability sequence (not necessarily the optimal sequence) by tracking multiple possible sequences at once.

Core Idea

- On each step of decoder, keep track of k most probable partial sequences (hypotheses)
- k is the beam size
- After you reach some stopping criterion, choose sequence with highest probability (factoring in some adjustment for length)

Beam Search Decoding



Sampling-based Decoding

Pure Sampling

- On each step t, randomly sample next word from probability distribution P_t.
- Like greedy decoding, but sample instead of argmax.

Top-n Sampling

- On each step t, randomly sample next word from P_t, restricted to just top-n most probable words
- $n = 1$ is greedy search, $n = V$ is pure sampling
- Increase n to get more diverse/risky output
- Decrease n to get more generic/safe output

Are these more efficient than beam search?

Automatic Evaluation for NLG

How can we automatically evaluate system output?

- We can always use BLEU.
- BLEU will mostly evaluate fluency.
- Is the output faithful to source input?
- Does the output cover important content?
- Is the discourse coherent?

Employ a fairly accurate IE system on the gold and automatic summaries and compare whether the identified relations align or diverge.

Automatic Evaluation for NLG

How can we automatically evaluate system output?

- **Relation Generation:** does the output cover important content?
  precision and number of unique relations r extracted from $\hat{y}_{1:T}$ that also appear in s.
- **Content Selection:** is the output faithful to source input?
  precision and count of unique relations r extracted from $\hat{y}_{1:T}$ that are also extracted from $y_{1:T}$.
- **Content Ordering:** is the discourse coherent?
  normalized Damerau-Levenshtein Distance (Brill and Moore, 2000) between the sequences of records extracted from $y_{1:T}$ and that extracted from $\hat{y}_{1:T}$

Template Baseline (RotoWire)

The *team*1 (wins1 – losses1) defeated the team2 (wins2 – losses2) pts1 – pts2 .

×6 player scored pts points (fgm – fga FG, tpm – tpa 3PT, ftm – fta FT) to go with reb rebounds.

The team1 next game will be at home against the Dallas Mavericks, while the team2 will travel to play the Bulls.

Results

RotoWire	Rel. Generation #	Rel. Generation P%	Con. Selection P%	Con. Selection R%	Con. Ordering DLD%	BLEU
TEMPL	**54.23**	**99.94**	26.99	**58.16**	14.92	8.46
ED+CC	23.72	74.80	29.49	36.18	15.42	14.19

MLB	Rel. Generation #	Rel. Generation P%	Con. Selection P%	Con. Selection R%	Con. Ordering DLD%	BLEU
TEMPL	**59.93**	**97.96**	22.82	**68.46**	10.64	3.81
ED+CC	18.69	92.19	**62.01**	50.12	25.44	9.69

- Template good at Relation Generation and Content Selection
- Content selection does not correlate with BLEU
- Results shown with beam size $k = 5$

Discussion

- We can generate summaries of input tables using a sequence-to-sequence model!
- Same modeling framework for MT, summarization, and data-to-text generation.
- But documents are not sequences; many remaining challenges.

Data & Code: `https://github.com/ratishsp/data2text-plan-py`
`http://lstm.seas.harvard.edu/docgen/`

Challenges: Discourse and Reference in Generation

The Atlanta Hawks defeated **the Miami Heat**, 103–95, at **Philips Arena** on Wednesday. **Atlanta** was in desperate need of a win and **they** were able to take care of a shorthanded **Miami** team here.
Defense was key for **the Hawks**, as they held **the Heat** to 42 percent shooting and forced them . . .

Challenges: Content Selection

PLAYER	AST	RB	PTS	FG	FGA	CITY
Tyler Johnson	5	2	27	8	16	Miami
Dwight Howard	11	17	23	9	11	Atlanta
Paul Millsap	2	9	21	8	12	Atlanta
Goran Dragic	4	2	21	8	17	Miami
Wayne Ellington	2	3	19	7	15	Miami
Dennis Schroder	7	4	17	8	15	Atlanta
Rodney McGruder	5	5	11	3	8	Miami

Tyler Johnson led all Miami scorers with 27 points.

Dwight Howard recorded a triple-double on 9 of 11 shooting.