NYU ML² Machine Learning for Language

Text Generation by Learning from Demonstrations

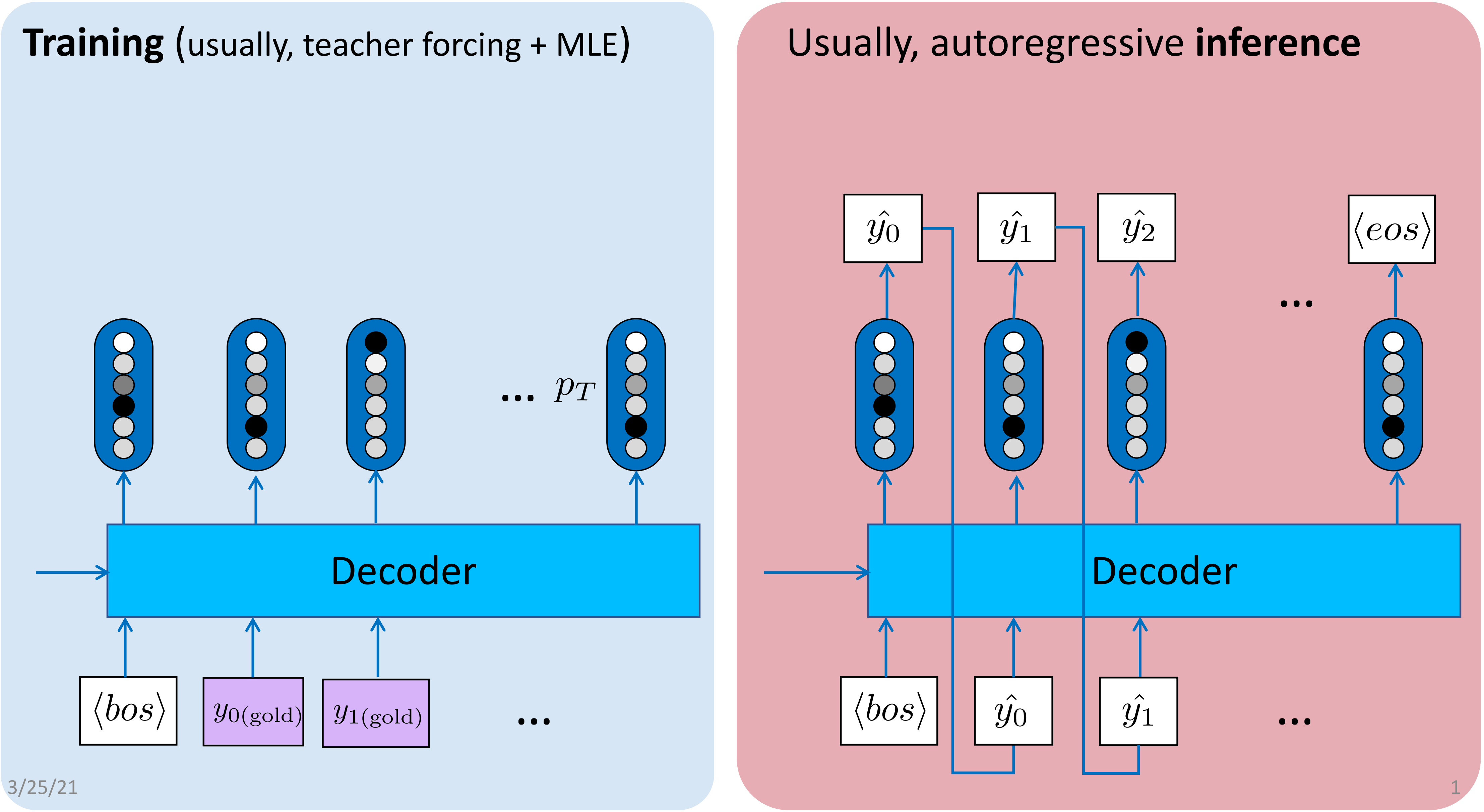
Richard Yuanzhe Pang
yzpang.me

He He
hhexiy.github.io

NEW YORK UNIVERSITY

3/24/21

The most widespread approach for supervised conditional text generation:



Motivation 1: mismatched history (gold vs. model-generated)

Repetition

Beam Search, *b*=32, from GPT-2:
"The study, published in the Proceedings of the National Academy of Sciences of the United States of America (PNAS), was conducted by researchers from the Universidad Nacional Autónoma de México (UNAM) and the Universidad Nacional Autónoma de México (UNAM/Universidad Nacional Autónoma de México/Universidad Nacional Autónoma de México/Universidad Nacional Autónoma de México/Universidad Nacional Autónoma de ..." (Holtzman et al., 2020)

Hallucination

Machine translation hallucination (Wang and Sennrich, 2020):
Source: So höre nicht auf die Ableugner.
Reference: So hearken not to those who deny.
MLE: Do not drive or use machines.

Motivation 2: mismatched learning/evaluation objectives

$$\mathbb{E}_{\boldsymbol{y} \sim p_{\text{human}}} \sum_{t=0}^{T} \log p_\theta(y_t \mid \boldsymbol{y}_{0:t-1}, \boldsymbol{x})$$

- High **recall**: p_ϑ must cover all outputs from p_{human}

$$\mathbb{E}_{\boldsymbol{y} \sim p_\theta} \sum_{t=0}^{T} \log p_{\text{human}}(y_t \mid \boldsymbol{y}_{0:t-1}, \boldsymbol{x})$$

- High **precision**: all outputs from p_ϑ must be scored high under p_{human}

Motivation

→ Background: RL formulation of text gen

Offline objective: learning algorithm GOLD

Background: RL in text generation

~~MLE~~ $$\mathbb{E}_{\boldsymbol{y}\sim p_{\text{human}}} \sum_{t=0}^{T} \log p_\theta(y_t \mid \boldsymbol{y}_{0:t-1}, \boldsymbol{x})$$

Eval $$\mathbb{E}_{\boldsymbol{y}\sim p_\theta} \sum_{t=0}^{T} \log p_{\text{human}}(y_t \mid \underline{\boldsymbol{y}_{0:t-1}, \boldsymbol{x})}$$

policy

reward

action

RL $$J(\theta) = \mathbb{E}_{\tau\sim\pi_\theta} \sum_{t=0}^{T} R(a_t, s_t)$$

state

Background: RL in text generation

- Prior approach: directly optimize a sequence-level metric like BLEU, ROUGE, etc., using policy gradient
  - Pros: no exposure bias; may discover high-quality outputs outside the references
  - Cons: degenerate solutions and difficult optimization (gradient estimated by samples from policy has high variance)

$$J(\theta) = \mathbb{E}_{\tau \sim \pi_\theta} \sum_{t=0}^{T} R(a_t, s_t)$$

policy action state reward

Motivation
Background: RL formulation of text gen
→ Offline objective: learning algorithm GOLD

Online vs. offline policy gradient

Online + on-policy policy gradient

- Step 1: sample outputs from the model
- Step 2: get seq-level rewards like BLEU
- Step 3: use policy gradient to optimize

$$\nabla_\theta J(\theta) = \mathbb{E}_{\tau \sim p_\theta} \sum_t \nabla_\theta \log \pi_\theta(a_t \mid s_t) \hat{Q}(s_t, a_t)$$

Online vs. offline policy gradient

No interaction with the environment

Online + on-policy policy gradient

- Step 1: sample outputs from the model
- Step 2: get seq-level rewards like BLEU
- Step 3: use policy gradient to optimize

$$\nabla_\theta J(\theta) = \mathbb{E}_{\tau \sim p_\theta} \sum_t \nabla_\theta \log \pi_\theta(a_t \mid s_t) \hat{Q}(s_t, a_t)$$

Offline + off-policy policy gradient

- Step 1: sample from **demonstrations** (i.e., gold supervised data)
- Step 2: get token-level rewards based on p_{MLE}
- Step 3: use policy gradient to optimize

$$\nabla_\theta J(\theta) = \mathbb{E}_{\tau \sim \pi_b} \sum_t w_t \nabla_\theta \log \pi_\theta(a_t \mid s_t) \hat{Q}(s_t, a_t)$$

$\pi_b = p_{\text{human}}$

use empirical distn

$w_t \approx \pi_\theta(a_{t'} \mid s_{t'})$

model confidence

$\hat{Q}(s_t, a_t) = \sum_{t'=t}^{T} \gamma^{t'-t} r_{t'}$

p_{MLE} based reward (see paper)

Intuition

- Upweight more "confident" examples; focus more on successful data (closer to test-time distribution)
- Intuitively reduce exposure bias

GOLD: generation by offline+off-policy learning from demonstrations

	Input	Output
NQG (using SQuAD)	The College of the University of Chicago grants Bachelor of Arts and Bachelor of Science degrees in 50 academic majors and **28** minors	How many academic minors does the university grant in total ?
CNN/DailyMail (**extractive** summarization)	Actor Vince Vaughn and pop star Lady Gaga mustered up the courage to dive into freezing cold water for the Special Olympics charity the day after cities across the country broke cold records for last month. The Chicago-native Wedding Crashers star was the celebrity guest of honor at his hometown's annual Polar Plunge, in which brave swimmers raise money for athletes with special needs by plunging into Lake Michigan. Vaughn, wearing a Blackhawks hockey jersey and jeans, led the way into a patch of frigid, slushy 33 degree water near Lincoln Park that had been cleared of snow, which has accumulated in the city and much of the US. Scroll down for videos . Icy dip: Actor Vince Vaughn was the Special Olympic Polar Plunge celebrity guest on Sunday, when the water was right around the freezing point . Hometown hero: Vaughn, a Chicago native, wore a jersey from the city's Blackhawks hockey team and jeans to take the icy dip and warmed up with a towel after the trying ordeal . Taking the plunge: Lady Gaga joined her new fiance, Chicago Fire actor Taylor Kinney, at the event with his co-stars on the show . Poker face: Celebrities such as Vince Vaughn and Lady Gaga found it difficult not to grimace in the freezing patch of Lake Michigan . The star of upcoming movie Unfinished Business first went in up to his knees before lowering himself into the water backwards. Gaga attended the event with shirtless fiance Taylor Kinney, who was wearing a baseball cap and shorts, and his costars from the show Chicago Fire. The singer rode piggyback-style with Kinney into the water before they fully immersed themselves into the lake. The couple then played with each other in the frozen surf and retreated to the warmth of the shore. 'Taylor gave me his hat I thought my wig was gonna freeze into and become one with the lake,' said the songstress, who will take different sort of plunge with her beau after becoming engaged last month.. Lake Michigan was estimated to be right at freezing when the event began at 10.30am Central Time. Staying warm with love: Lady Gaga attended the charity event with her fiance after getting engaged to the television actor last month . Before and after: Gaga and Kinney looked comfortable while posing for photos (left) before their plunge left them in varied states of bedraggled . California love: A shirtless Kinney wore a baseball cap reminiscent of the warm-weather West Coast when he carried his future wife into the waters of Lake Michigan . Splashing in the snow: Gaga grits her teeth as she finds the energy to play in the lake despite a winter of record-breaking cold weather . Paparazzi: Gaga, whose real name is Stefani Germanotta, lost the glamour she often displays on stage and the red carpet when emerging from Lake Michigan . Organizers say that 5,000 people were expected to attend the plunge, and that they had already made more than $1million during the morning, according to ABC7. The plunge came at the end of a February that had seen cold records shattered across much of the US, which has seen snow in almost every state this winter. Chicago, which has seen weeks of freezing temperatures in the past month, tied a 140-year-old record for its coldest February with an average temperature of 14.6. Temperatures at O'Hare Airport hit minus 10 degrees Saturday morning, according to the Chicago Tribune. Mission accomplished: Special Olympics Chicago's annual Polar Plunge event regularly raises more than $1million for special needs athletes . The Fame: The event draws thousands of participants and spectators to the water each year along with celebrities such as Gaga. Above, the singer poses with a fan on the beach . Fire and ice: The Chicago Fire Department had to clear snow and …	Chicago-native Vaughn was celebrity guest at his hometown's annual Polar Plunge for Special Olympics . Newly-engaged Lady Gaga attended with Chicago Fire fiance Taylor Kinney and his television co-stars . City tied 140-year record for coldest February with 14.6 degree average and more winter weather expected across US . Winter Storm Sparta brings more snow and ice to East Coast after causing tragic weather-related deaths in Midwest . Boston can possibly eclipse its record for most snow in one winter if it receives 5.7 more inches .
XSum (**abstractive** summarization)	The country's consumer watchdog has taken Apple to court for false advertising because the tablet computer does not work on Australia's 4G network. Apple's lawyers said they were willing to publish a clarification. However the company does not accept that it misled customers. The Australian Competition and Consumer Commission (ACCC) said on Tuesday: "Apple's recent promotion of the new 'iPad with wi-fi + 4G' is misleading because it represents to Australian consumers that the product can, with a sim card, connect to a 4G mobile data network in Australia, when this is not the case." The watchdog then lodged a complaint at the Federal Court in Melbourne. At a preliminary hearing, Apple lawyer Paul Anastassiou said Apple had never claimed the device would work fully on the current 4G network operated by Telstra. Apple says the new iPad works on what is globally accepted to be a 4G network. The matter will go to a full trial on 2 May. The Apple iPad's third version went on sale earlier this month, with Australia the first country where it was available. Shoppers lined up by the hundreds at Apple stores on opening day and the company said it had been its strongest iPad launch to date. The ACCC said it was seeking an injunction on sales as well as a financial penalty against Apple, corrective advertising and refunds to consumers. On its website, Apple does state that 4G LTE is only supported on selected networks in the US and Canada.	US technology firm Apple has offered to refund Australian customers who felt misled about the 4G capabilities of the new iPad.
IWLST14 De-En (machine translation)	Im amerikanischen mittelwesten luden bauern getreide auf kähne und sandten es den fluss hoch auf den markt nach chicago .	In the American midwest, farmers used to load grain onto barges and send it upriver to the chicago market.

Our hypotheses to **GOLD**

1. **GOLD improves generation quality**
   - Automatic results

	NQG (BART) (BLEU)	CNN/DM (BART) (R-2)	XSum (BART) (R-2)	IWSLT14 De-En (Transformer) (BLEU)
MLE	20.68	21.28	22.08	34.64
GOLD-*p* (product as *Q*)	21.42	22.01	22.26	35.33
GOLD-*s* (sum as *Q*)	21.98	22.09	22.58	35.45

   - Human evals: pairwise comparison of GOLD-*s* vs. MLE generations

2. **GOLD improves precision at the cost of recall**
   - High precision: larger BLEU/ROUGE, better human evals
   - Low recall: very large perplexity w.r.t. gold standards

High perplexity != low quality

3. **GOLD alleviates exposure bias**

Takeaways

1. MLE encourages high recall
2. GOLD (generation by off-policy and offline learning from demonstration) is easy to implement and optimize
   - Essentially weighted MLE
3. GOLD encourages high-precision generation
   - Instead of distribution matching