DSC291: Advanced Statistical Natural Language Processing

Text Generation

Zhiting Hu

Lecture 17, May 25, 2022

UC San Diego

HALICIOĞLU DATA SCIENCE INSTITUTE

Outline

- Controllable text generation (cont'd)
- GANs for text

- 3 Paper presentations (15 x 3 mins)

Two Central Goals

Controlled generation in unsupervised settings

- Generating human-like, grammatical, and readable text
  - Exposure bias, criteria mismatch: reinforcement learning (next lecture)

- Generating text that contains desired information inferred from inputs

#supervision data

  - Machine translation
    - Source sentence --> target sentence w/ the same meaning

10s of millions

  - Data description
    - Table --> data report describing the table

10s of 1000s

  - Attribute control
    - Sentiment: positive --> ``I like this restaurant''

10s of 1000s

    - Modify sentiment from positive to negative

0

  - Conversation control
    - Control conversation strategy and topic

0

Unsupervised Controlled Generation of Text

- Sentence-level control
    - Text attribute transfer (style transfer)
    - Text content manipulation
- Conversation-level control
    - Target-guided open-domain conversation

Recap: Text Attribute Transfer

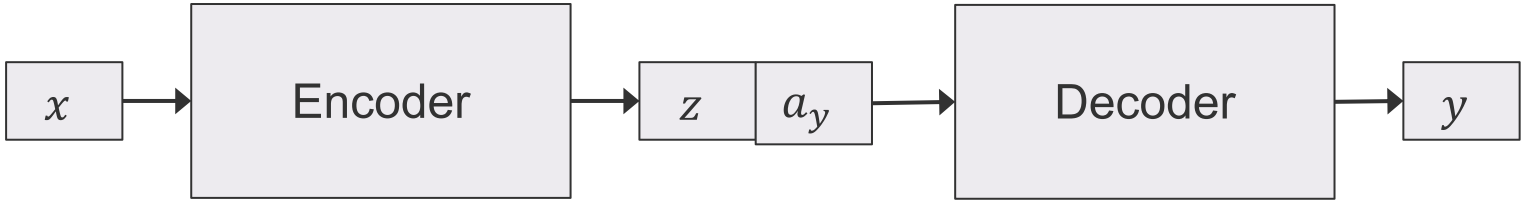
- Modify a given sentence to
  - Have desired attribute values
  - While keeping all other aspects unchanged
- Attribute: sentiment, tense, voice, gender, …

- E.g., transfer sentiment from negative to positive:
  - ``It was super dry and had a weird taste to the entire slice .”
  - ``It was super fresh and had a delicious taste to the entire slice .”

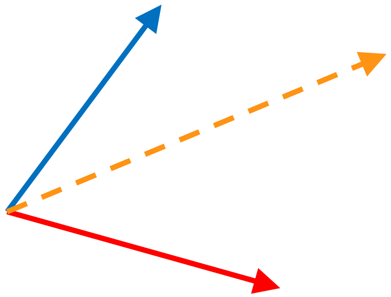
- Applications:
  - Personalized article writing, emotional conversation systems, …

[Hu et al., 17] Toward Controlled Generation of Text

Recap: Text Attribute Transfer: Solution



- Task: $(\boldsymbol{x}, \boldsymbol{a}_y) \rightarrow \boldsymbol{y}$
  - $\boldsymbol{y}$ has the desired attribute $\boldsymbol{a}_y$
  - $\boldsymbol{y}$ keeps all attribute-independent properties of $\boldsymbol{x}$
- Model $p_\theta(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{a}_y)$
- Key intuition for learning:
  - Decompose the task into competitive sub-objectives
  - Use direct supervision for each of the sub-objectives
- Auto-encoding loss: $(\boldsymbol{x}, \boldsymbol{a}_x) \rightarrow \boldsymbol{x}$
- Classification loss: $\hat{y} \sim p_\theta(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{a}_y),\ f(\hat{\boldsymbol{y}}) \rightarrow \boldsymbol{a}_y$
  - where f is a pre-trained attribute classifier
- The above two losses are competitive; minimize jointly to avoid collapse



Unsupervised Controlled Generation of Text

- Sentence-level control
  - Text attribute transfer (style transfer)
  - Text content manipulation

- Conversation-level control
  - Target-guided open-domain conversation

Key idea:
- Decompose the task into competitive sub-objectives
- Use direct supervision for each of the sub-objectives

Unsupervised Controlled Generation of Text

- Sentence-level control
    - Text attribute transfer (style transfer)
    - **Text content manipulation**

- Conversation-level control
    - Target-guided open-domain conversation

Key idea:
- Decompose the task into competitive sub-objectives
- Use direct supervision for each of the sub-objectives

Text Content Manipulation

- Generate a sentence to describe content in a given data record
- We want to control the **writing style**: use the writing style of a reference sentence

	Name	Food	Area	Price	Near
Data Record	Loch Fyne	Italian	Riverside	£20-25	Strada

[Lin et al., 20] Data-to-Text Generation with Style Imitation

Text Content Manipulation

- Generate a sentence to describe content in a given data record

Data Record

Name	Food	Area	Price	Near
Loch Fyne	Italian	Riverside	£20-25	Strada

Exemplar 1	Zizzi is a pub providing fine French dining but with an expensive price, located near Cocum in the city center.
Generation 1	Loch Fyne provides fine Italian dining with a £20-25 price, located near Strada at the riverside.
Exemplar 2	Located near the Blue Spice, there is a highly-rated place, the Mill, as a choice that frugally priced.
Generation 2	Located near Strada by the river, there is a place with Italian foods, Loch Fyne, as a choice that priced £20-25.
Exemplar 3	With a family-friendly atmosphere and a 5-star rating, Aromi is a pub in the city center.
Generation 3	With Italian foods and a moderate price range, Loch Fyne is near Strada at the riverside.

[Lin et al., 20] Data-to-Text Generation with Style Imitation

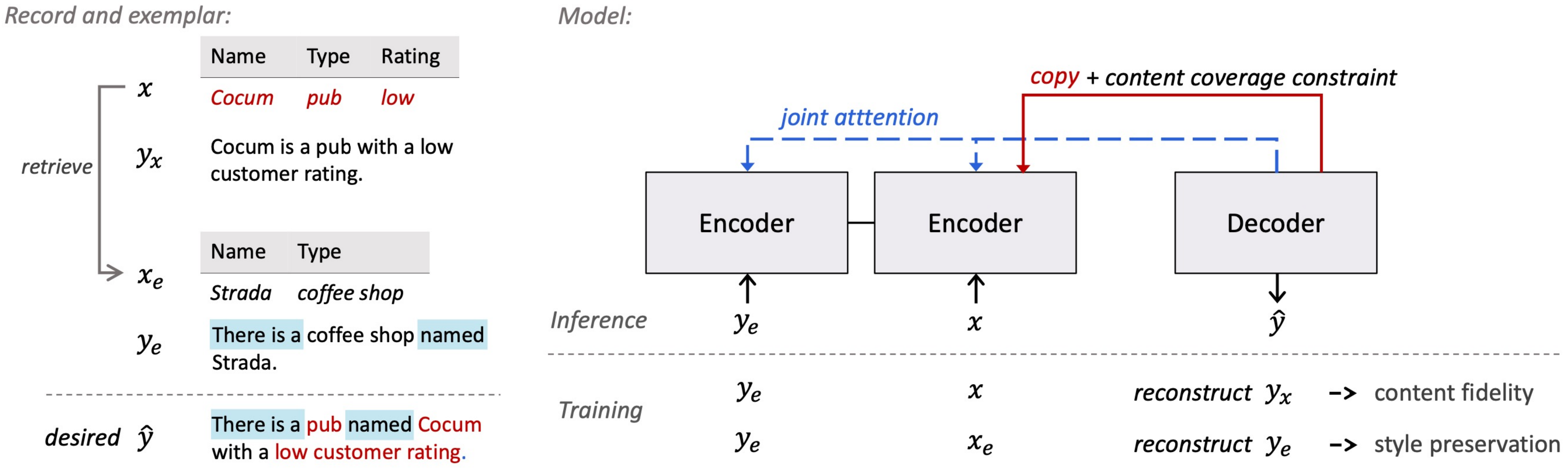
Text Content Manipulation

- Generate a sentence to describe content in a given data record

Content Record	**PLAYER**	**PT**	**RB**	**AS**	**PLAYER**	**PT**
	LeBron_James	32	4	7	Kyrie_Irving	20
Reference Sentence	Jrue_Holiday led the way with 26 points and 6 assists , while Goran_Dragic scored 23 points and pulled down 8 rebounds .					
Output	LeBron_James led the way with 32 points , 7 assists and 4 rebounds , while Kyrie_Irving scored 20 points .					

[Lin et al., 20] Data-to-Text Generation with Style Imitation

Method



[Lin et al., 20] Data-to-Text Generation with Style Imitation

Results

Content Record	Name Cocum	EatType coffee shop	Food Italian	PriceRange £20-25	CustomRating high	FamilyFriendly family friendly
Exemplar 1	Looking for French food near Zizzi? Come try Strada, which has a 3-star customer rating and priced lowly.					
Slot filling	Looking for Italian [...] food near Zizzi? Come try [...] Cocum, which has a high customer rating and priced £20-25.					
AdvST	For Italian [...] place near Zizzi? Come try [...] Cocum, which has a high customer rating with priced £20-25.					
Ours	Looking for an Italian coffee shop? Come try family-friendly Cocum, which has a high customer rating and priced £20-25.					
Exemplar 2	Along the riverside near Cafe Rouge, there is a Japanese food place called The Golden Curry. It has an average customer rating since it is not a family-friendly environment.					
Slot-filling	Along the riverside near Cafe Rouge [...], there is a Italian food [...] place called Cocum. It has an high customer rating since it is not a family-friendly environment.					
AdvST	Along the riverside near the Ranch [...], there is a Italian food [...] place called Cocum. It has [...] high customer rating since it is not a family-friendly environment.					
Ours	Priced £20-25, there is an Italian food coffee shop called Cocum. It has a high customer rating since it is a family-friendly environment.					

Results

		Restaurant Recommendations			NBA Reports		
		Content		Style	Content		Style
	Method	**%Incl.-new**	**%Excl.-old**	**m-BLEU**	**Precision**	**Recall**	**m-BLEU**
Reference	AttnCopy-S2S	78.88±2.08	99.71±0.06	13.95±0.52	81.62±3.25	75.65±7.42	45.5±0.71
	Slot-filling	61.23	66.2	100	56.69	71.34	100
Baselines	MAST	36.28±0.25	37.06±0.16	**91.76±0.28**	23.06±3.90	27.37±3.88	**95.43±2.71**
	AdvST	51.64±4.45	57.06±4.44	76.02±5.27	67.37±0.66	66.79±1.43	64.67±4.81
Ours	Transformer w/o Coverage	60.03±2.16	74.65±2.69	77.81±3.83	62.58±2.88	70.22±3.58	81.75±2.32
	+ Coverage	61.84±1.31	81.14±2.73	80.29±0.35	67.74±0.79	**74.35±1.22**	81.97±2.87
	LSTM w/o Coverage	60.83±1.29	81.45±1.10	78.91±1.05	68.74±3.07	69.35±3.30	79.88±2.44
	+ Coverage	**65.02±4.16**	**82.53±0.70**	82.92±3.18	**69.54±1.16**	73.27±1.18	80.66±1.89

Unsupervised Controlled Generation of Text

- Sentence-level control
  - Text attribute transfer (style transfer)
  - Text content manipulation

- Conversation-level control
  - Target-guided open-domain conversation

Key idea:
- Decompose the task into competitive sub-objectives
- Use direct supervision for each of the sub-objectives

Unsupervised Controlled Generation of Text

- Sentence-level control
  - Text attribute transfer (style transfer)
  - Text content manipulation

- Conversation-level control
  - Target-guided open-domain conversation

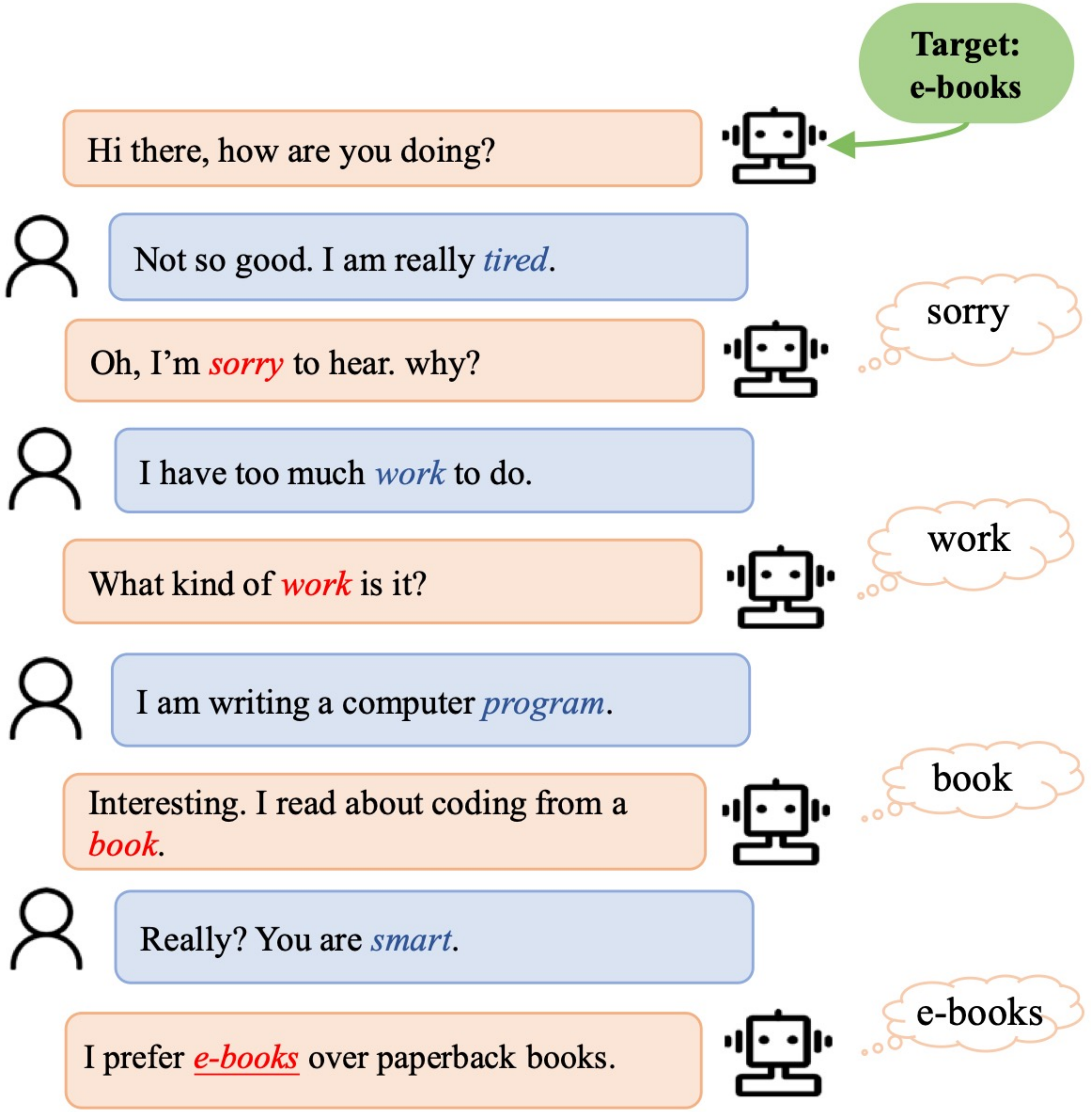
Key idea:
- Decompose the task into competitive sub-objectives
- Use direct supervision for each of the sub-objectives

Target-guided Open-domain Conversation

- Task-oriented dialog:
  - Address a specific task, e.g., booking a flight
  - Close domain
- Open-domain chit-chat:
  - Improve user engagement
  - Random conversation, hard to control
- Target-guided conversation:
  - Open-domain conversation
  - Controlled conversation strategy to reach a *desired topic* in the end of conversation
  - Applications:
    - Bridges task-oriented dialog and open-domain chit-chat
    - Conversational recommender system, education, psychotherapy

Target-guided Open-domain Conversation

- Two goals:
  - Starting from any topic, reach a desired topic in the end of conversation
  - Natural conversation: smooth transition



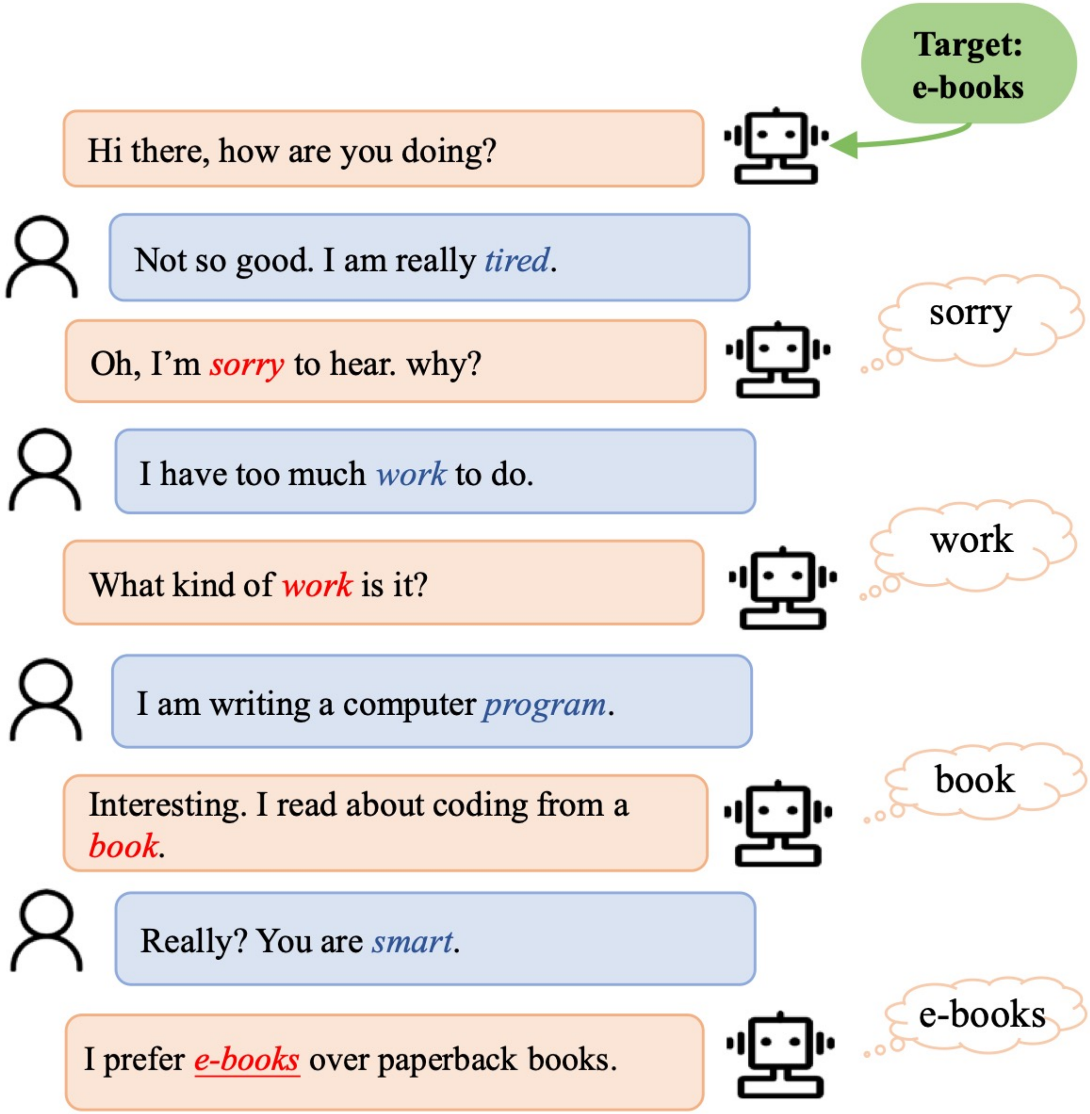
Target-guided Open-domain Conversation

- Two goals:
  - Starting from any topic, reach a desired topic in the end of conversation
  - Natural conversation: smooth transition

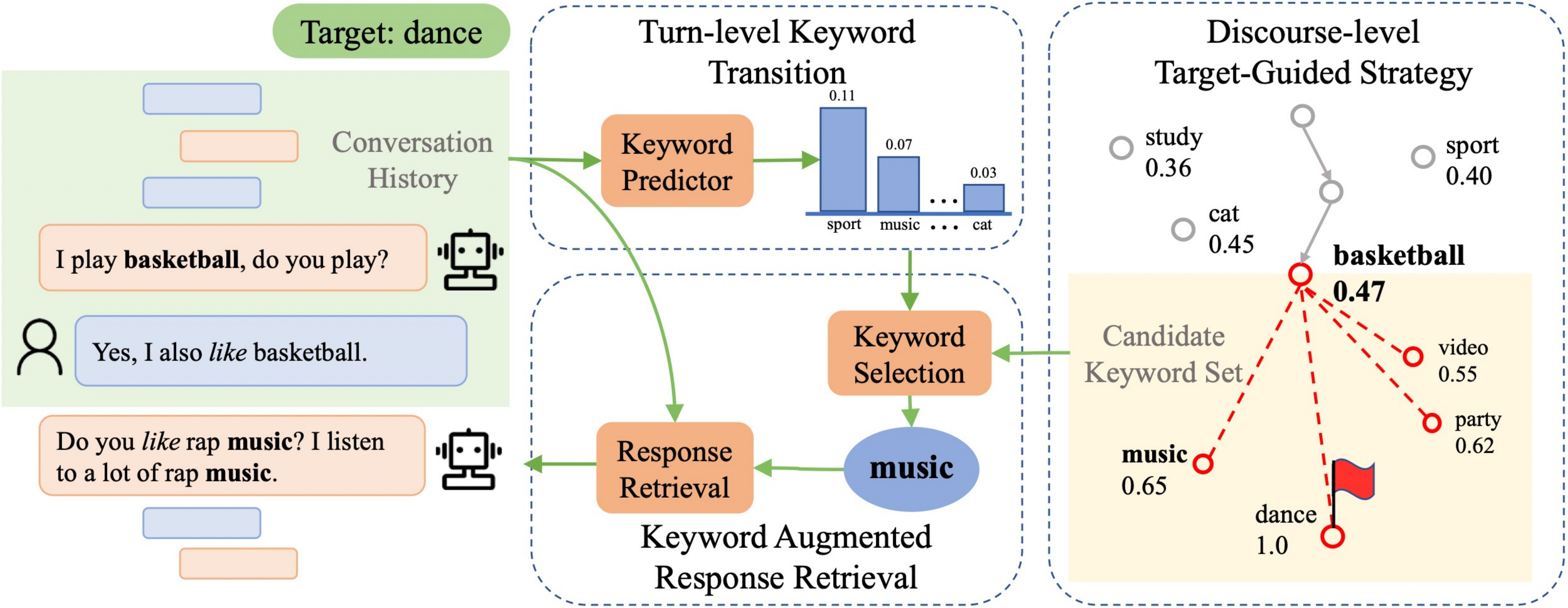
Challenge: No supervised data for the task

Solution: Use competitive sub-objectives and partial supervision

- **Natural conversation**: rich chit-chat data to learn smooth single-turn transition
- **Reaching desired target**: rule-based multi-turn planning

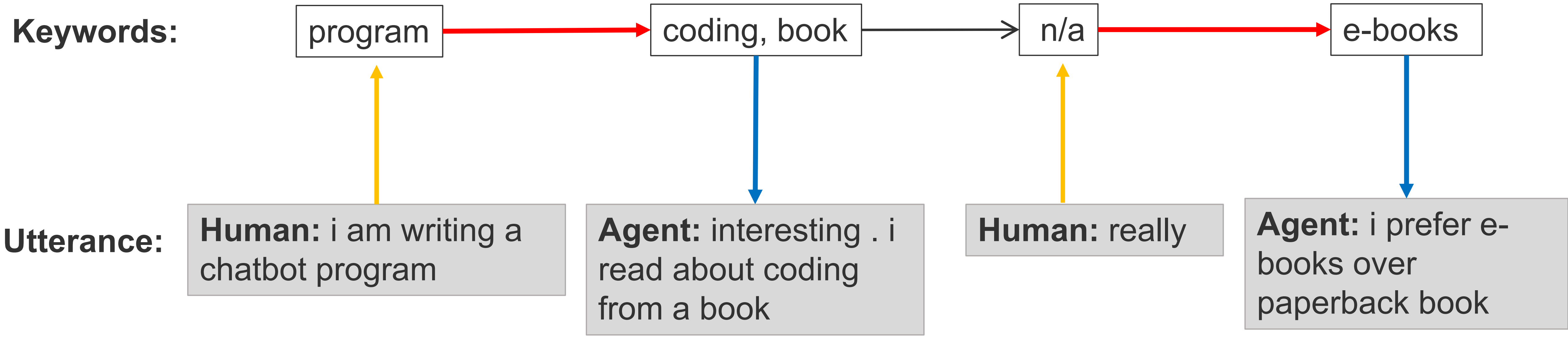


Method



Target: dance
Conversation History
I play **basketball**, do you play?
Yes, I also *like* basketball.
Do you *like* rap **music**? I listen to a lot of rap **music**.
Turn-level Keyword Transition
Keyword Predictor
0.11
0.07
0.03
sport
music
...
cat
Keyword Selection
music
Response Retrieval
Keyword Augmented Response Retrieval
Discourse-level Target-Guided Strategy
study 0.36
sport 0.40
cat 0.45
basketball 0.47
Candidate Keyword Set
video 0.55
party 0.62
music 0.65
dance 1.0

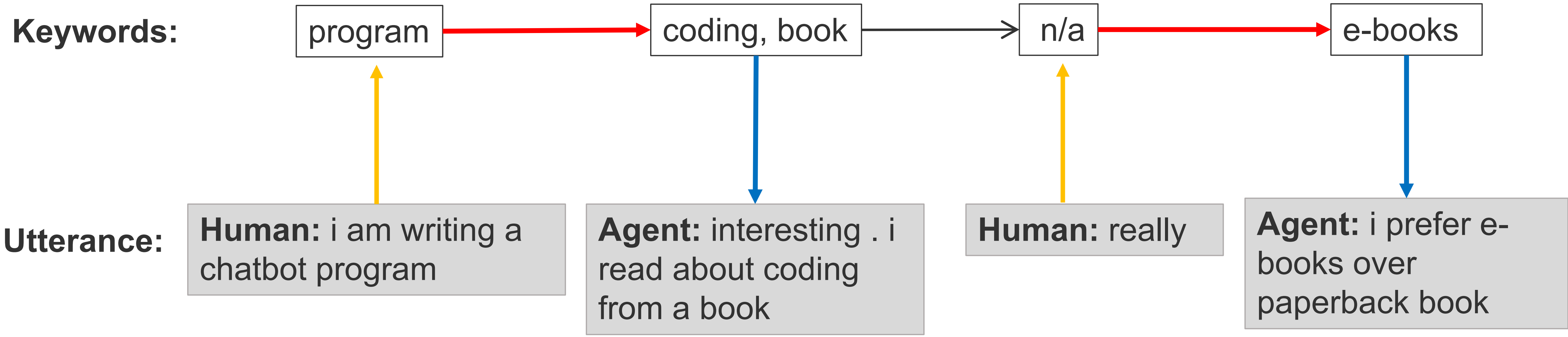
Method



Keywords:
program
coding, book
n/a
e-books
Utterance:
Human: i am writing a chatbot program
Agent: interesting . i read about coding from a book
Human: really
Agent: i prefer e-books over paperback book

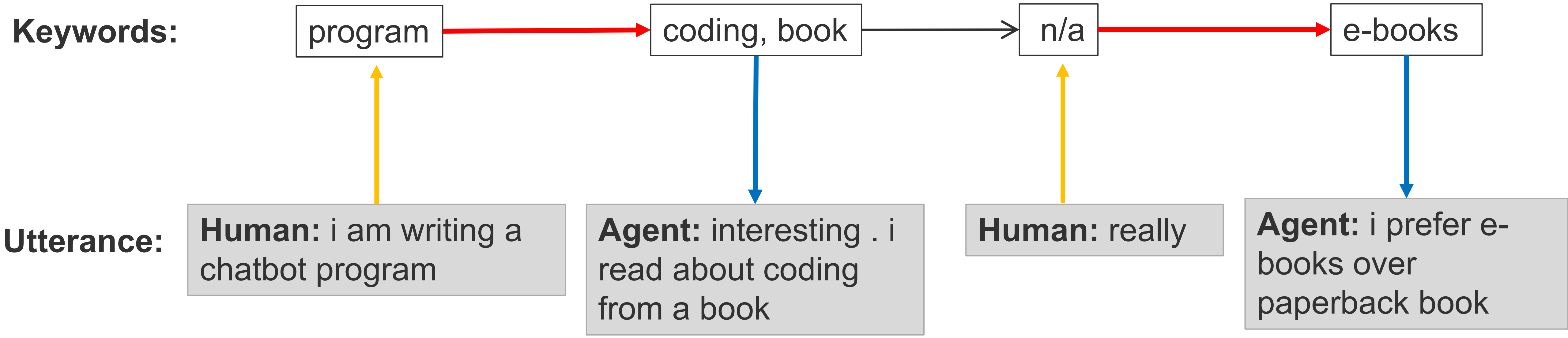
Method

- keyword extraction



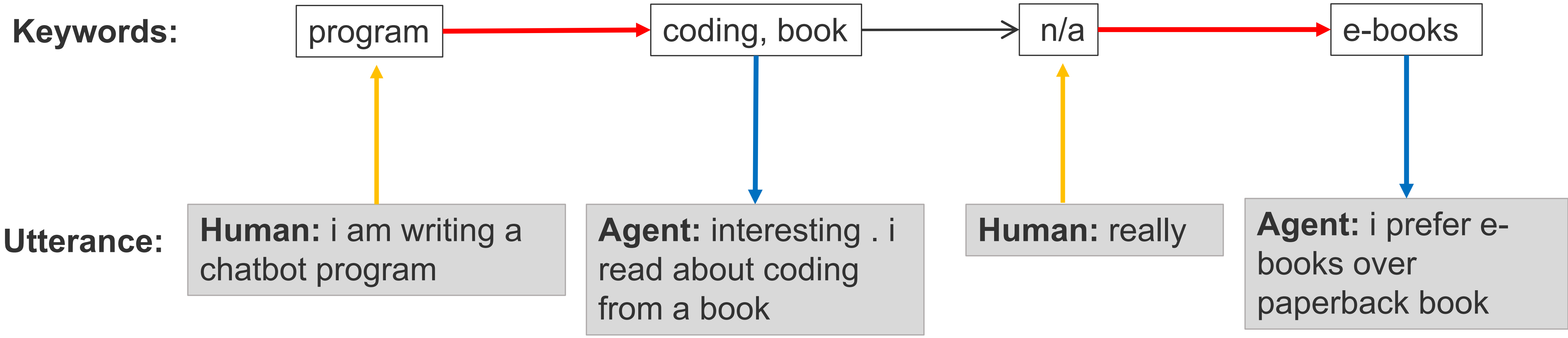
Method

- → keyword extraction
- → keyword conditional response retrieval



Method

- → keyword extraction
- → keyword conditional response retrieval
- → learned kernel-based topic transition + target-guided rule:
  - (1) transit to keywords that are close in the word embedding space – smooth
  - (2) next keywords must get closer to the target keyword – target-guided
  - Tune relative weight of (1) and (2) to control aggressiveness



Example Results

Target: dance

Agent : hi how is it going with you
Human : fine . I'm playing sports
Agent : i play basketball . u play ?
Human : yeah i also like basketball .
Agent : you like rap music ? i listen to a lot of rap music
Human : no i like classic music
Agent : its a taylor swift song
Human : taylor swift is a great singer
Agent : sing to me , what are you going to sing
Human : i know nothing about singing
Agent : oh i think i can dance lol

Example Results – Failure Case

Target: listen

Agent : hi ! how are you ? i am very exciting about my new job !
Human : fine . what is your new job ?
Agent : that sounds exciting .
Human : i like read books
Agent : what kind of books do you read ?
Human : i read comics
Agent : i do not have tv
Human : watch anime is fun
Agent : what position d you play ?
Human : i play a lot of sports
Agent : oh really ? what kind of music . i listen to classical

Unsupervised Controlled Generation of Text

- Sentence-level control
  - Text attribute transfer (style transfer)
  - Text content manipulation
- Conversation-level control
  - Target-guided open-domain conversation

Key idea:
- Decompose the task into competitive sub-objectives
- Use direct supervision for each of the sub-objectives

Key Takeaways: Two Central Goals

- Generating human-like, grammatical, and readable text
  - Exposure bias, criteria mismatch: reinforcement learning (next lecture)
- Generating text that contains desired information inferred from inputs
  - Machine translation
    - Source sentence --> target sentence w/ the same meaning
  - Data description
    - Table --> data report describing the table
  - Attribute control
    - Sentiment: positive --> ``I like this restaurant''
  - Conversation control
    - Control conversation strategy and topic

Generative Adversarial Networks

Generative modeling

- In generative modeling, we'd like to train a network that models a distribution, such as a distribution over images.
- One way to judge the quality of the model is to sample from it.
- This field has seen rapid progress:

2009

2015

2018

Courtesy: Grosse CSC321 Lecture 19

Generative modeling

- In generative modeling, we'd like to train a network that models a distribution, such as a distribution over images.
- One way to judge the quality of the model is to sample from it.
- This field has seen rapid progress:

Figure courtesy: Ian Goodfellow

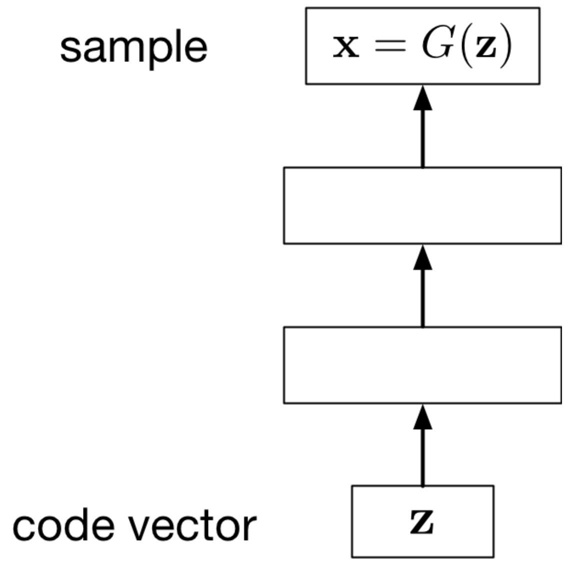
Generative modeling

- Modern approaches to generative modeling:
  - Variational Auto-encoder (Lecture #8)
  - Auto-regressive models (e.g., language model) (Lecture #3)
  - Generative adversarial networks (today)
  - Flow-based models, diffusion models (not covered)

Courtesy: Grosse CSC321 Lecture 19

Implicit Generative Models

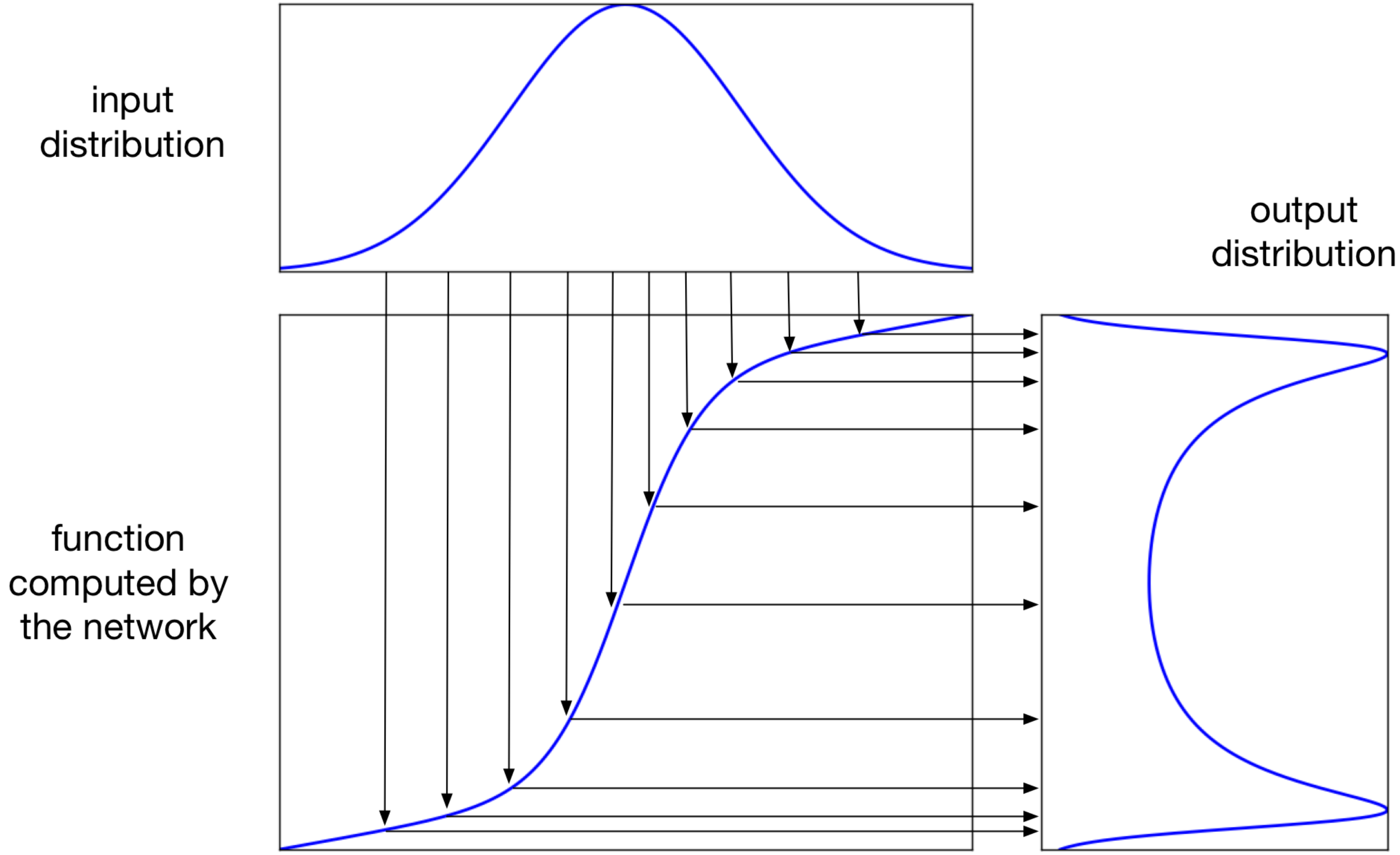
- Implicit generative models implicitly define a probability distribution
- Start by sampling the code vector $\mathbf{z}$ from a fixed, simple distribution (e.g. spherical Gaussian)
- The generator network computes a differentiable function G mapping $\mathbf{z}$ to an $\mathbf{x}$ in data space



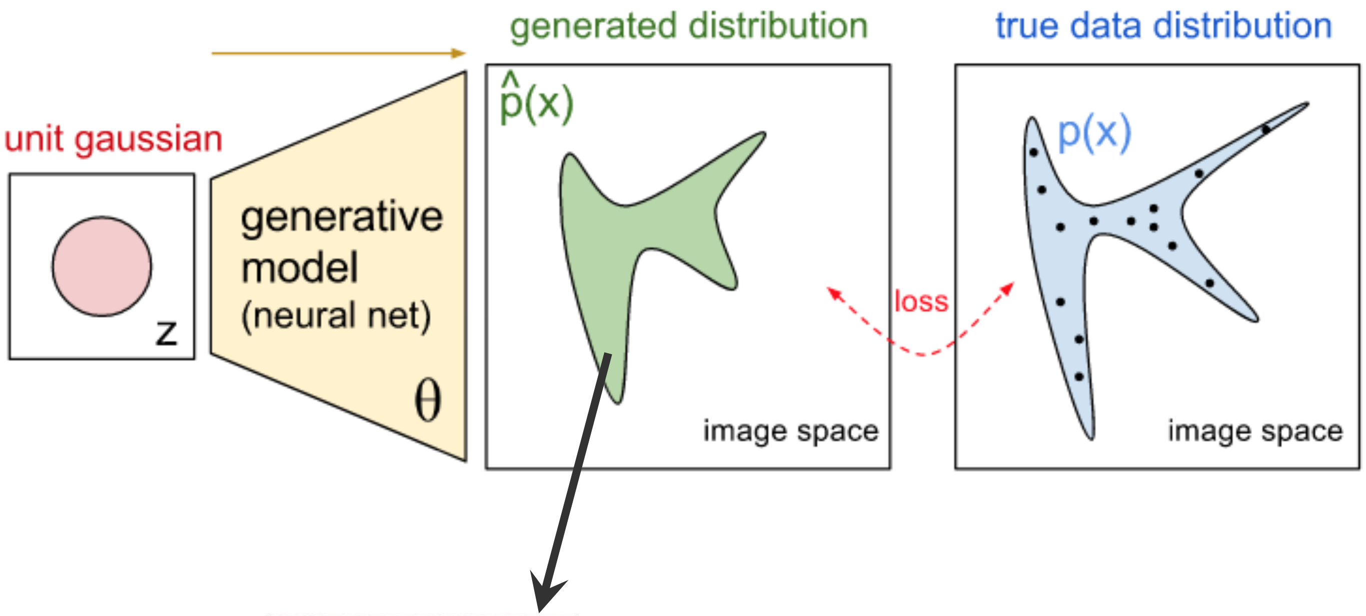
- a stochastic process to simulate data $\boldsymbol{x}$
- Intractable to evaluate likelihood

Implicit Generative Models

A 1-dimensional example:



Implicit Generative Models



https://blog.openai.com/generative-models/

Courtesy: Grosse CSC321 Lecture 19

Implicit Generative Models

- The advantage of implicit generative models: if you have some criterion for evaluating the quality of samples, then you can compute its gradient with respect to the network parameters, and update the network's parameters to make the sample a little better
- The idea behind **Generative Adversarial Networks (GANs)**: train two different networks
  - The generator network tries to produce realistic-looking samples
  - The discriminator network tries to figure out whether an image came from the training set or the generator network
- The generator network tries to fool the discriminator network

Generative Adversarial Nets (GANs)

- Generative model $\boldsymbol{x} = G_\theta(\boldsymbol{z}), \;\; \boldsymbol{z} \sim p(\boldsymbol{z})$
  - Maps noise variable $\boldsymbol{z}$ to data space $\boldsymbol{x}$
  - Defines an implicit distribution over $\boldsymbol{x}$: $p_{g_\theta}(\boldsymbol{x})$
- Discriminator $D_\phi(\boldsymbol{x})$
  - Output the probability that $\boldsymbol{x}$ came from the data rather than the generator

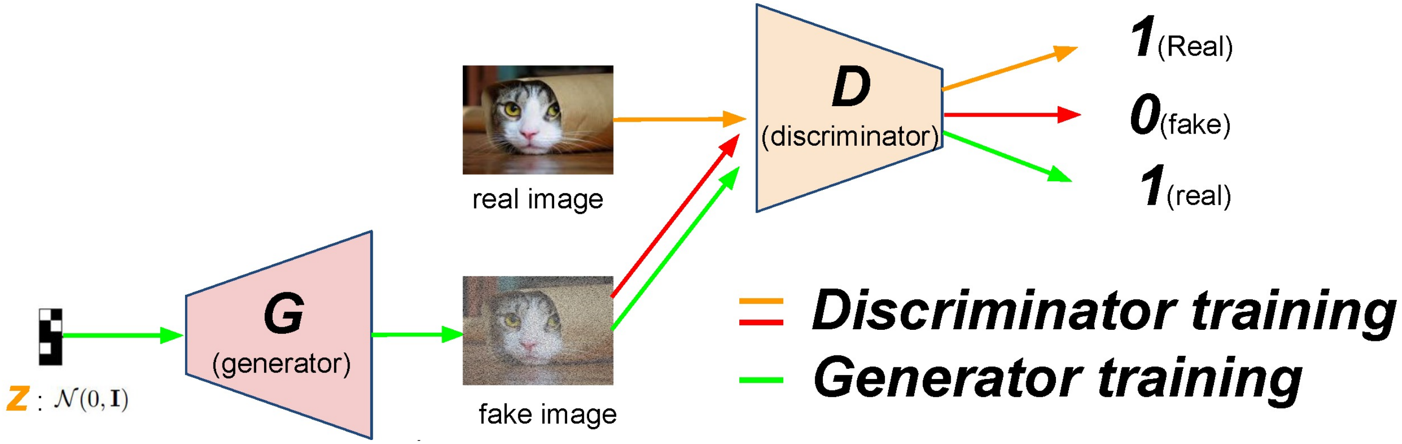


Figure courtesy: Kim

Generative Adversarial Nets (GANs)

- Learning
  - A minimax game between the generator and the discriminator
  - Train D to maximize the probability of assigning the correct label to both training examples and generated samples
  - Train G to fool the discriminator

$$\max_D \mathcal{L}_D = \mathbb{E}_{\boldsymbol{x} \sim p_{data}(\boldsymbol{x})} [\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{x} \sim G(\boldsymbol{z}), \boldsymbol{z} \sim p(\boldsymbol{z})} [\log(1 - D(\boldsymbol{x}))]$$

$$\min_G \mathcal{L}_G = \mathbb{E}_{\boldsymbol{x} \sim G(\boldsymbol{z}), \boldsymbol{z} \sim p(\boldsymbol{z})} [\log(1 - D(\boldsymbol{x}))] \,.$$

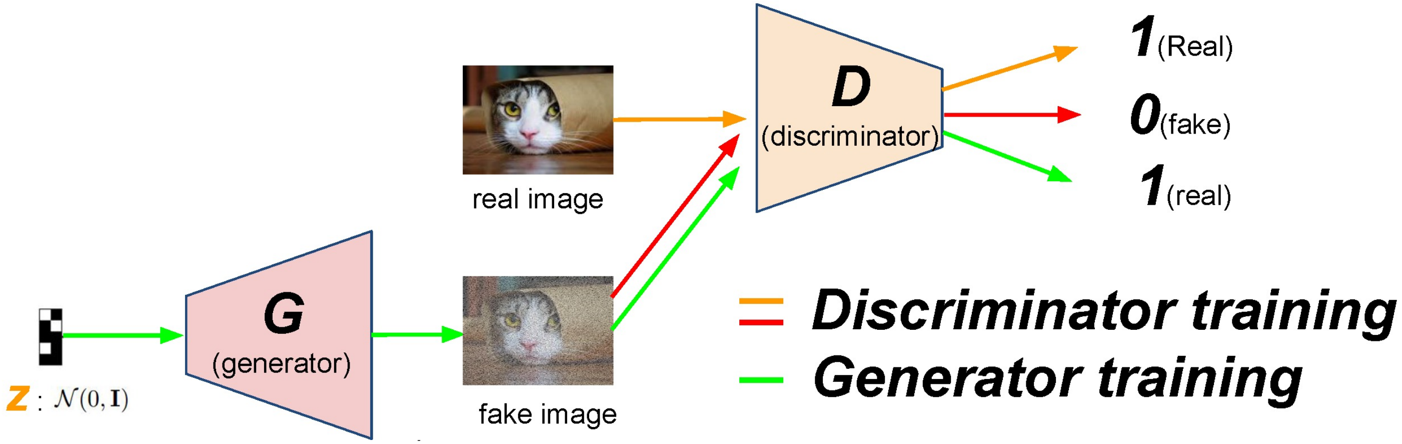
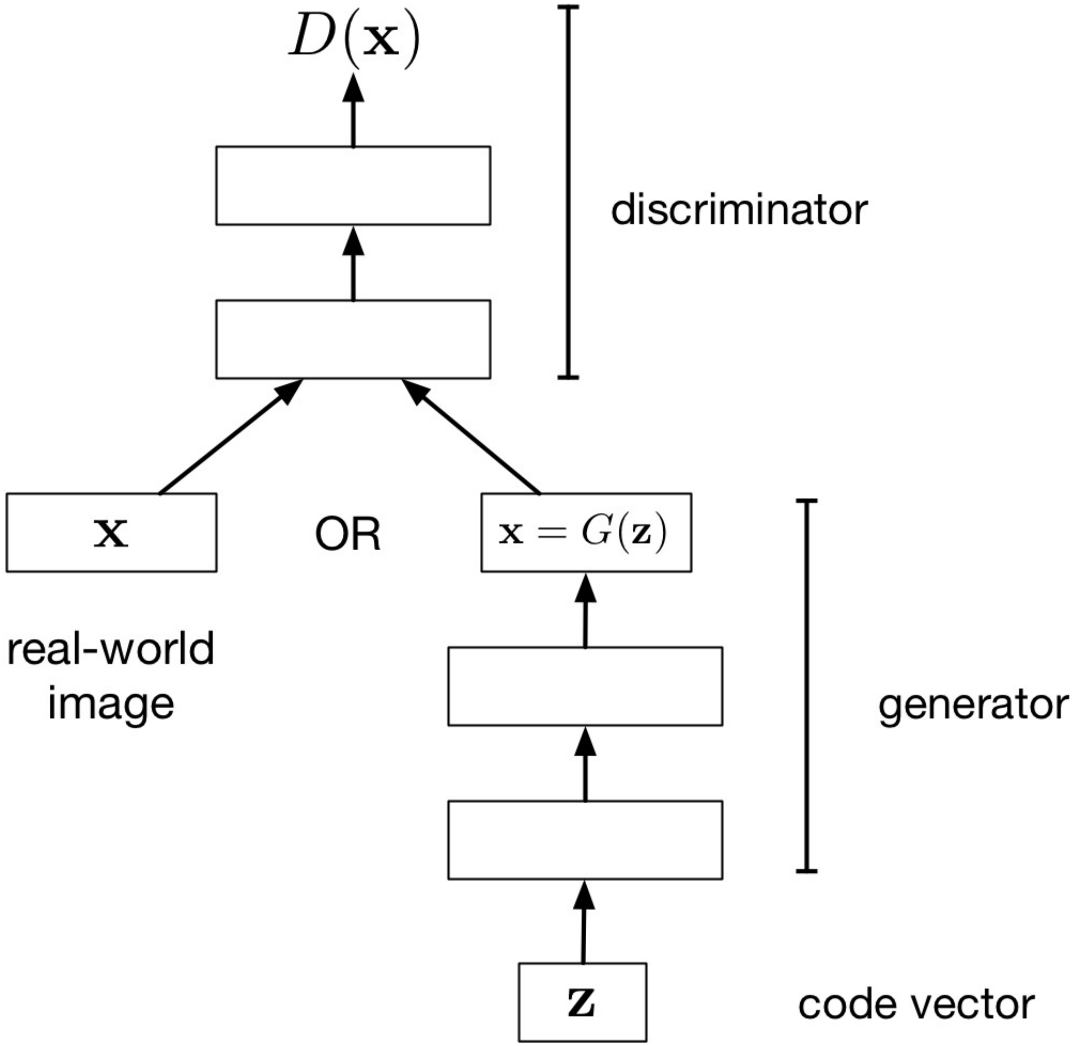


Figure courtesy: Kim

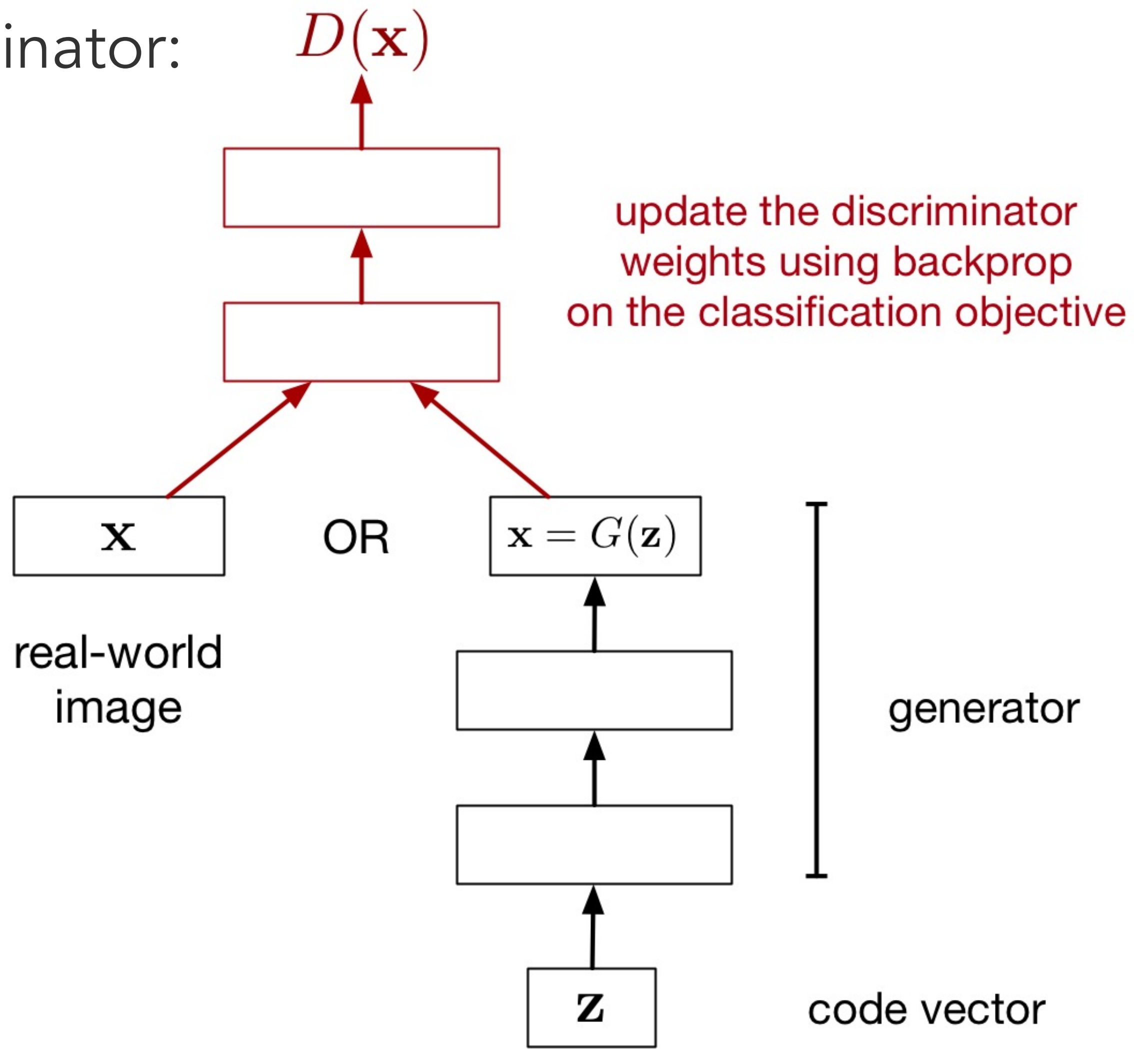
Generative Adversarial Nets (GANs)



Courtesy: Grosse CSC321 Lecture 19

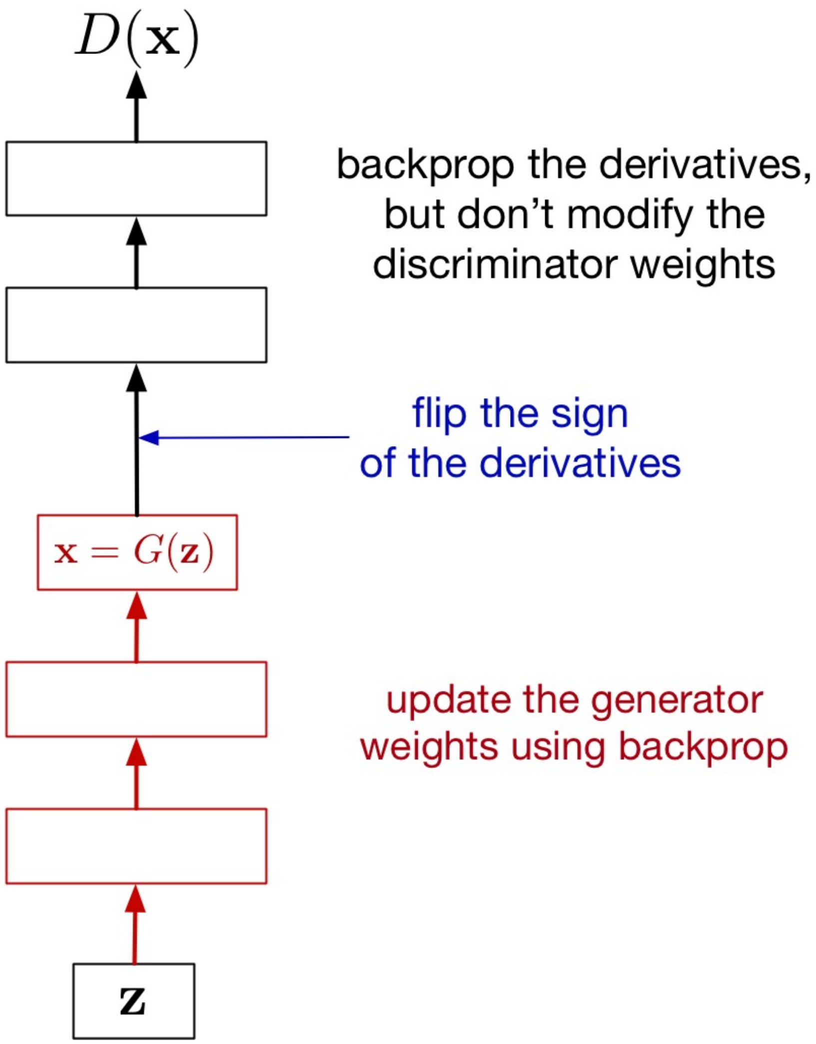
Generative Adversarial Nets (GANs)

Updating the discriminator:



Generative Adversarial Nets (GANs)

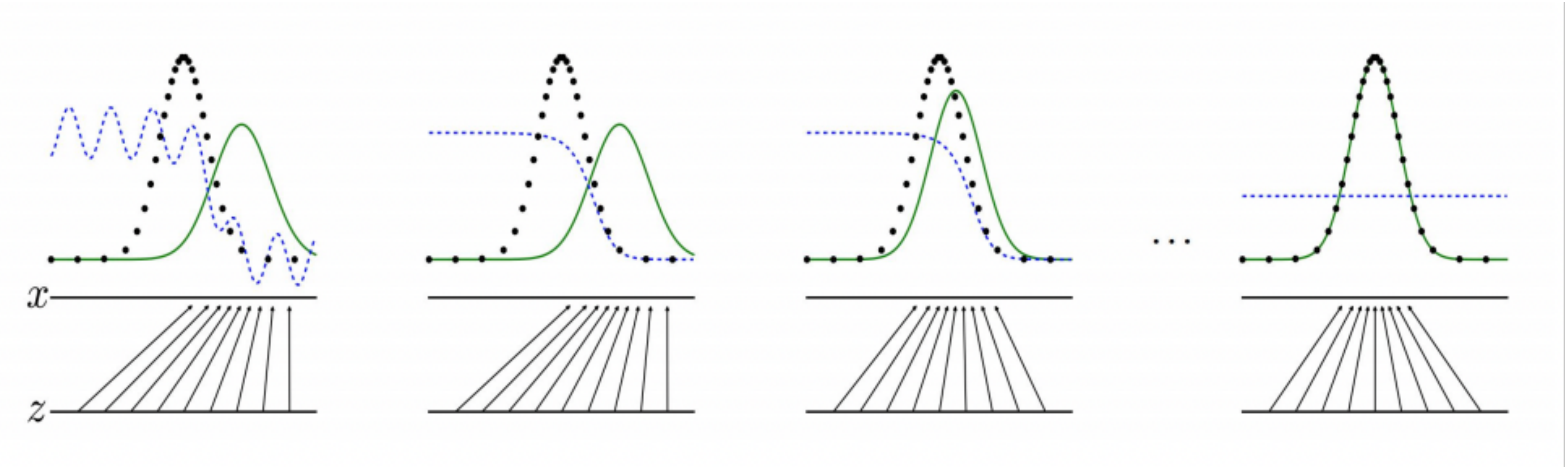
Updating the generator:



Courtesy: Grosse CSC321 Lecture 19

Generative Adversarial Nets (GANs)

Alternating training of the generator and discriminator:



Optimality of GANs

- Objectives:

$$\max_D \mathcal{L}_D = \mathbb{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}\left[\log D(\boldsymbol{x})\right] + \mathbb{E}_{\boldsymbol{x}\sim G(\boldsymbol{z}),\boldsymbol{z}\sim p(\boldsymbol{z})}\left[\log(1 - D(\boldsymbol{x}))\right]$$
$$\min_G \mathcal{L}_G = \mathbb{E}_{\boldsymbol{x}\sim G(\boldsymbol{z}),\boldsymbol{z}\sim p(\boldsymbol{z})}\left[\log(1 - D(\boldsymbol{x}))\right].$$

- Global optimality: $p_g = p_{data}$
- Proof:

Optimality of GANs

Proposition 1. *For G fixed, the optimal discriminator D is*

$$D_G^*(\boldsymbol{x}) = \frac{p_{data}(\boldsymbol{x})}{p_{data}(\boldsymbol{x}) + p_g(\boldsymbol{x})} \tag{2}$$

[Goodfellow et al., 2014]

Optimality of GANs

Proposition 1. *For G fixed, the optimal discriminator D is*

$$D_G^*(\boldsymbol{x}) = \frac{p_{data}(\boldsymbol{x})}{p_{data}(\boldsymbol{x}) + p_g(\boldsymbol{x})} \tag{2}$$

Proof. The training criterion for the discriminator D, given any generator G, is to maximize the quantity $V(G, D)$

$$\begin{aligned} V(G, D) &= \int_{\boldsymbol{x}} p_{\text{data}}(\boldsymbol{x}) \log(D(\boldsymbol{x}))dx + \int_{z} p_{\boldsymbol{z}}(\boldsymbol{z}) \log(1 - D(g(\boldsymbol{z})))dz \\ &= \int_{\boldsymbol{x}} p_{\text{data}}(\boldsymbol{x}) \log(D(\boldsymbol{x})) + p_g(\boldsymbol{x}) \log(1 - D(\boldsymbol{x}))dx \end{aligned} \tag{3}$$

For any $(a, b) \in \mathbb{R}^2 \setminus \{0, 0\}$, the function $y \to a \log(y) + b \log(1 - y)$ achieves its maximum in $[0, 1]$ at $\frac{a}{a+b}$.

[Goodfellow et al., 2014]

Optimality of GANs

- The minimax game can now be reformulated as

$$
\begin{aligned}
C(G) = & \max_D V(G, D) \\
= & \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}}[\log D_G^*(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}}[\log(1 - D_G^*(G(\boldsymbol{z})))] \\
= & \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}}[\log D_G^*(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{x} \sim p_g}[\log(1 - D_G^*(\boldsymbol{x}))] \\
= & \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}}\left[\log \frac{p_{\text{data}}(\boldsymbol{x})}{P_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}\right] + \mathbb{E}_{\boldsymbol{x} \sim p_g}\left[\log \frac{p_g(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}\right]
\end{aligned}
$$

[Goodfellow et al., 2014]

Optimality of GANs

- The minimax game can now be reformulated as

$$
\begin{aligned}
C(G) &= \max_D V(G, D) \\
&= \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}}[\log D_G^*(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}[\log(1 - D_G^*(G(\boldsymbol{z})))] \\
&= \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}}[\log D_G^*(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{x}\sim p_g}[\log(1 - D_G^*(\boldsymbol{x}))] \\
&= \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}}\left[\log \frac{p_{\text{data}}(\boldsymbol{x})}{P_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}\right] + \mathbb{E}_{\boldsymbol{x}\sim p_g}\left[\log \frac{p_g(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}\right]
\end{aligned}
$$

Theorem 1. *The global minimum of the virtual training criterion $C(G)$ is achieved if and only if $p_g = p_{data}$. At that point, $C(G)$ achieves the value $-\log 4$.*

[Goodfellow et al., 2014]

Optimality of GANs

- The minimax game can now be reformulated as

$$
\begin{aligned}
C(G) &= \max_D V(G, D) \\
&= \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}}[\log D_G^*(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}[\log(1 - D_G^*(G(\boldsymbol{z})))] \\
&= \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}}[\log D_G^*(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{x}\sim p_g}[\log(1 - D_G^*(\boldsymbol{x}))] \\
&= \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}}\left[\log \frac{p_{\text{data}}(\boldsymbol{x})}{P_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}\right] + \mathbb{E}_{\boldsymbol{x}\sim p_g}\left[\log \frac{p_g(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}\right]
\end{aligned}
$$

Theorem 1. *The global minimum of the virtual training criterion $C(G)$ is achieved if and only if $p_g = p_{data}$. At that point, $C(G)$ achieves the value $-\log 4$.*

$$
\begin{aligned}
C(G) &= -\log(4) + KL\left(p_{\text{data}} \middle\| \frac{p_{\text{data}} + p_g}{2}\right) + KL\left(p_g \middle\| \frac{p_{\text{data}} + p_g}{2}\right) \\
&= -\log(4) + 2 \cdot JSD\left(p_{\text{data}} \| p_g\right)
\end{aligned}
$$

Jensen-Shannon Divergence

[Goodfellow et al., 2014]

A better loss function

- We introduced the minimax cost function for the generator:

$$\mathcal{J}_G = \mathbb{E}_{\mathbf{z}}[\log(1 - D(G(\mathbf{z})))]$$

- One problem with this is saturation.
- Here, if the generated sample is really bad, the discriminator's prediction is close to 0, and the generator's cost is flat.

A better loss function: non-saturating GAN

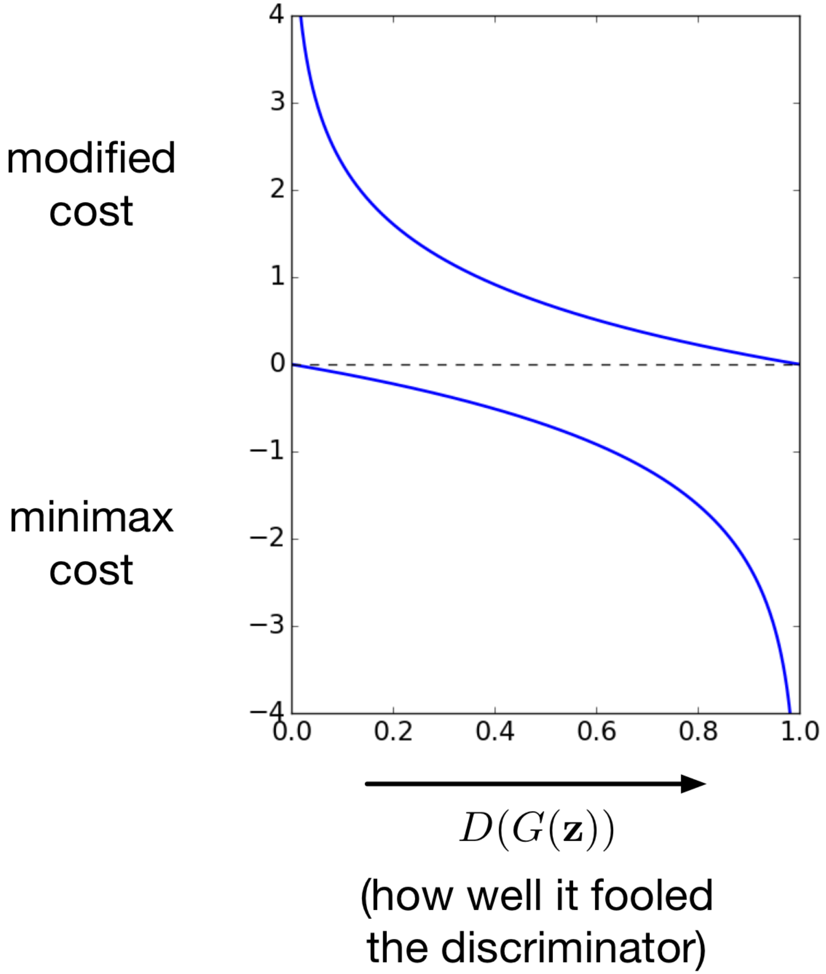
- Original minimax cost:

$$\mathcal{J}_G = \mathbb{E}_{\mathbf{z}}[\log(1 - D(G(\mathbf{z})))]$$

- Modified generator cost:

$$\mathcal{J}_G = \mathbb{E}_{\mathbf{z}}[-\log D(G(\mathbf{z}))]$$

- This fixes the saturation problem.



Wasserstein GAN (WGAN)

- If our data are on a low-dimensional manifold of a high dimensional space, the model's manifold and the true data manifold can have a negligible intersection in practice

[Arjovsky et al., 2017] Slide adapted from bhiksha

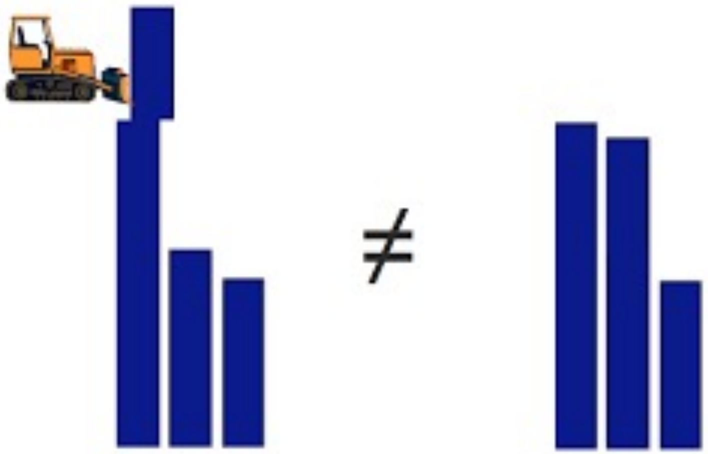
Wasserstein GAN (WGAN)

- If our data are on a low-dimensional manifold of a high dimensional space, the model's manifold and the true data manifold can have a negligible intersection in practice
- The loss function and gradients may not be continuous and well behaved

[Arjovsky et al., 2017] Slide adapted from bhiksha

Wasserstein GAN (WGAN)

- If our data are on a low-dimensional manifold of a high dimensional space, the model's manifold and the true data manifold can have a negligible intersection in practice
- The loss function and gradients may not be continuous and well behaved
- The Wasserstein Distance is well defined
  - Earth Mover's Distance
  - Minimum transportation cost for making one pile of dirt in the shape of one probability distribution to the shape of the other distribution



[Arjovsky et al., 2017] Slide adapted from bhiksha

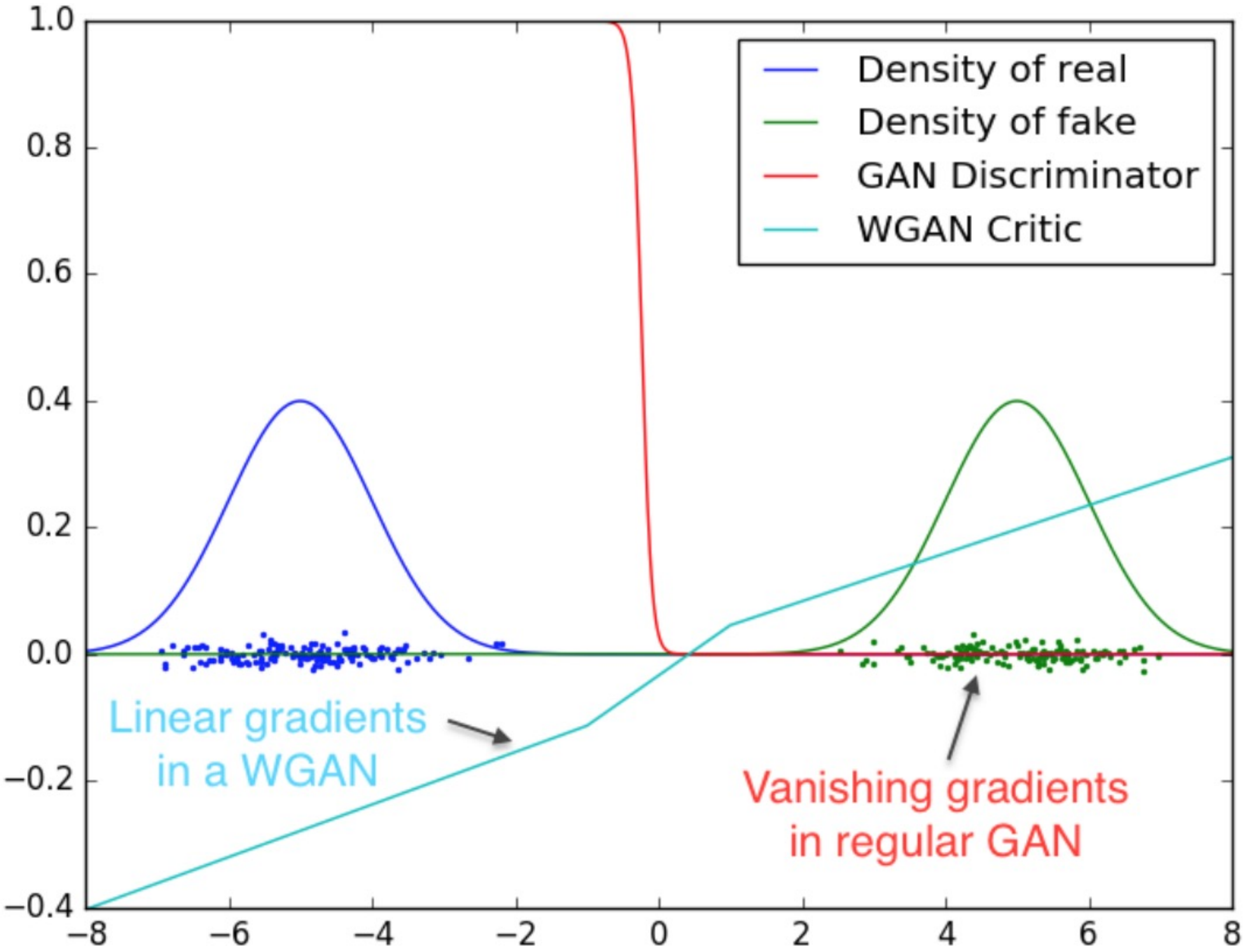
Wasserstein GAN (WGAN)

- Objective

$$W\left(p_{data}, p_g\right) = \frac{1}{K} \sup_{||D||_L \leq K} \mathrm{E}_{x\sim p_{data}}[D(x)] - \mathrm{E}_{x\sim p_g}[D(x)]$$

- $||D||_L \leq K$: K- Lipschitz continuous
- Use gradient-clipping to ensure D has the Lipschitz continuity

WGAN vs Vanilla GAN



Progressive GAN

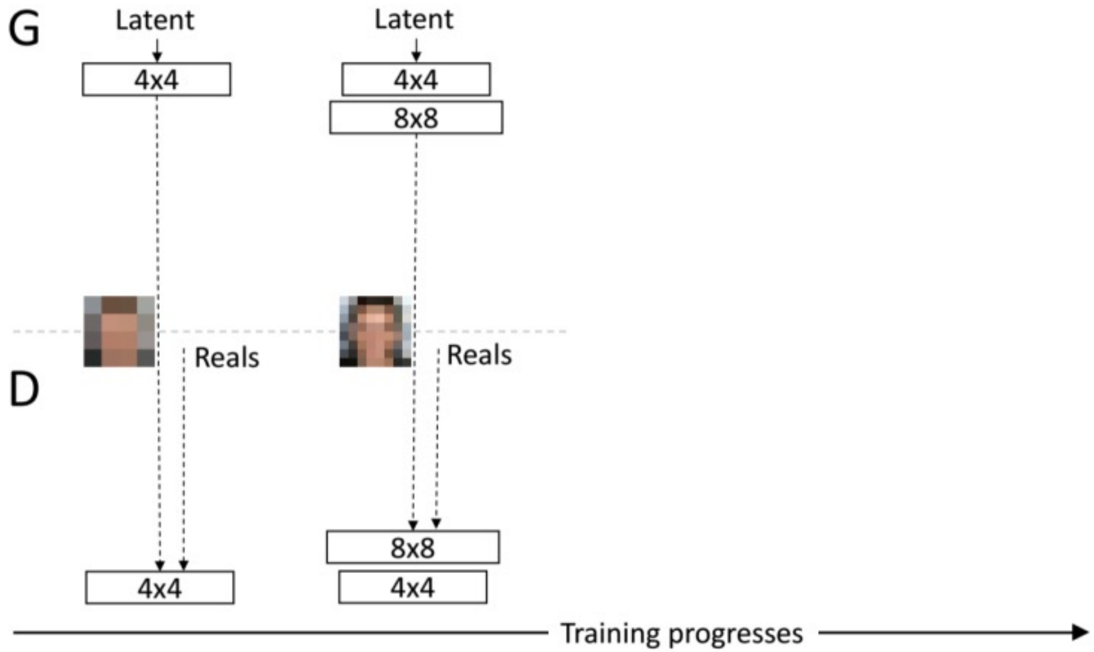
Low resolution images

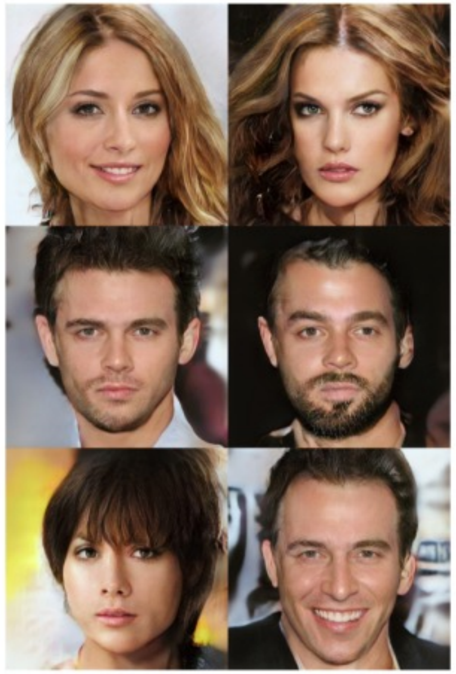
[Karras et al., 2018]

Progressive GAN

Low resolution images

add in additional layers





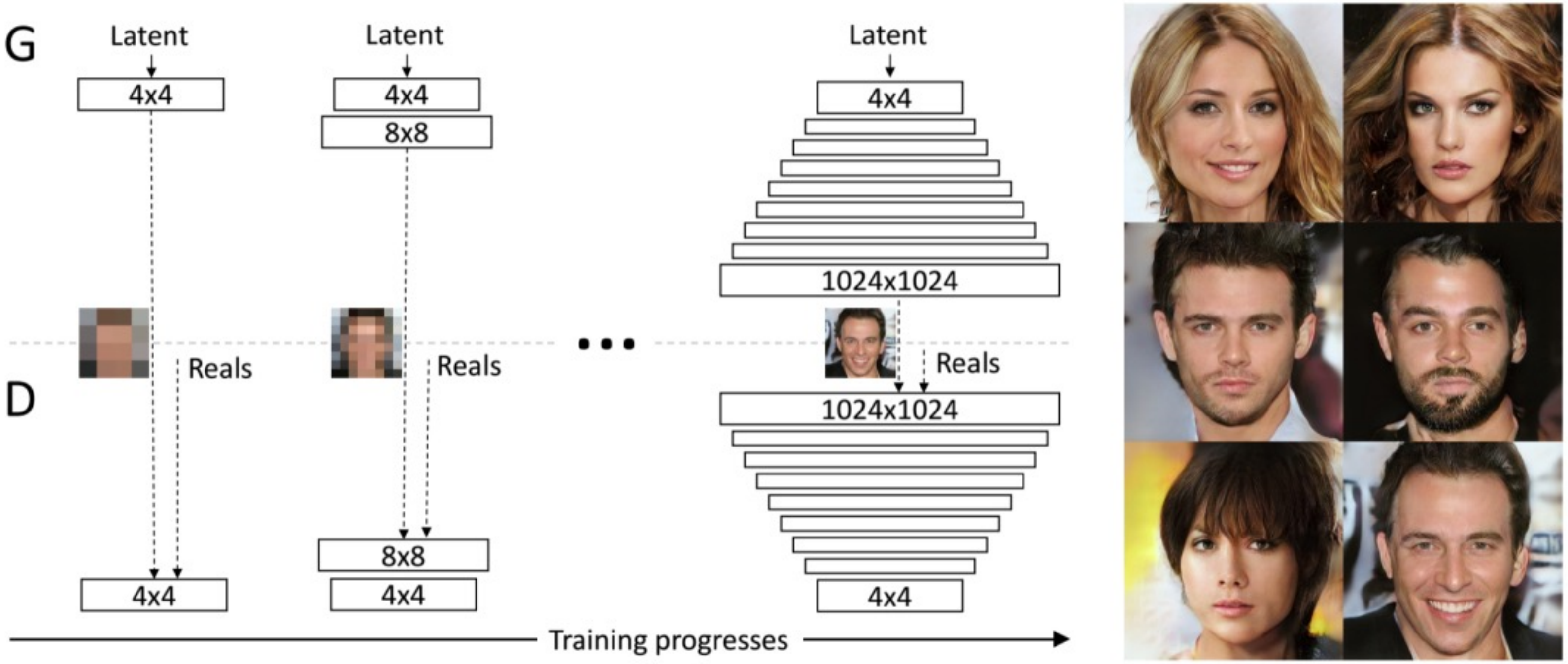
[Karras et al., 2018]

Progressive GAN

Low resolution images

add in additional layers

High resolution images



[Karras et al., 2018]

BigGAN

[Brock et al., 2018]

BigGAN

- GANs benefit dramatically from scaling

[Brock et al., 2018]

BigGAN

- GANs benefit dramatically from scaling
- 2x – 4x more parameters
- 8x larger batch size
- Simple architecture changes that improve scalability

[Brock et al., 2018]

BigGAN

- GANs benefit dramatically from scaling
- 2x – 4x more parameters
- 8x larger batch size
- Simple architecture changes that improve scalability

[Brock et al., 2018]

BigGAN

- GANs benefit dramatically from scaling
- 2x 4x more parameters
- 8x
- Si

[Brock et al., 2018]

GANs for Text

Questions?