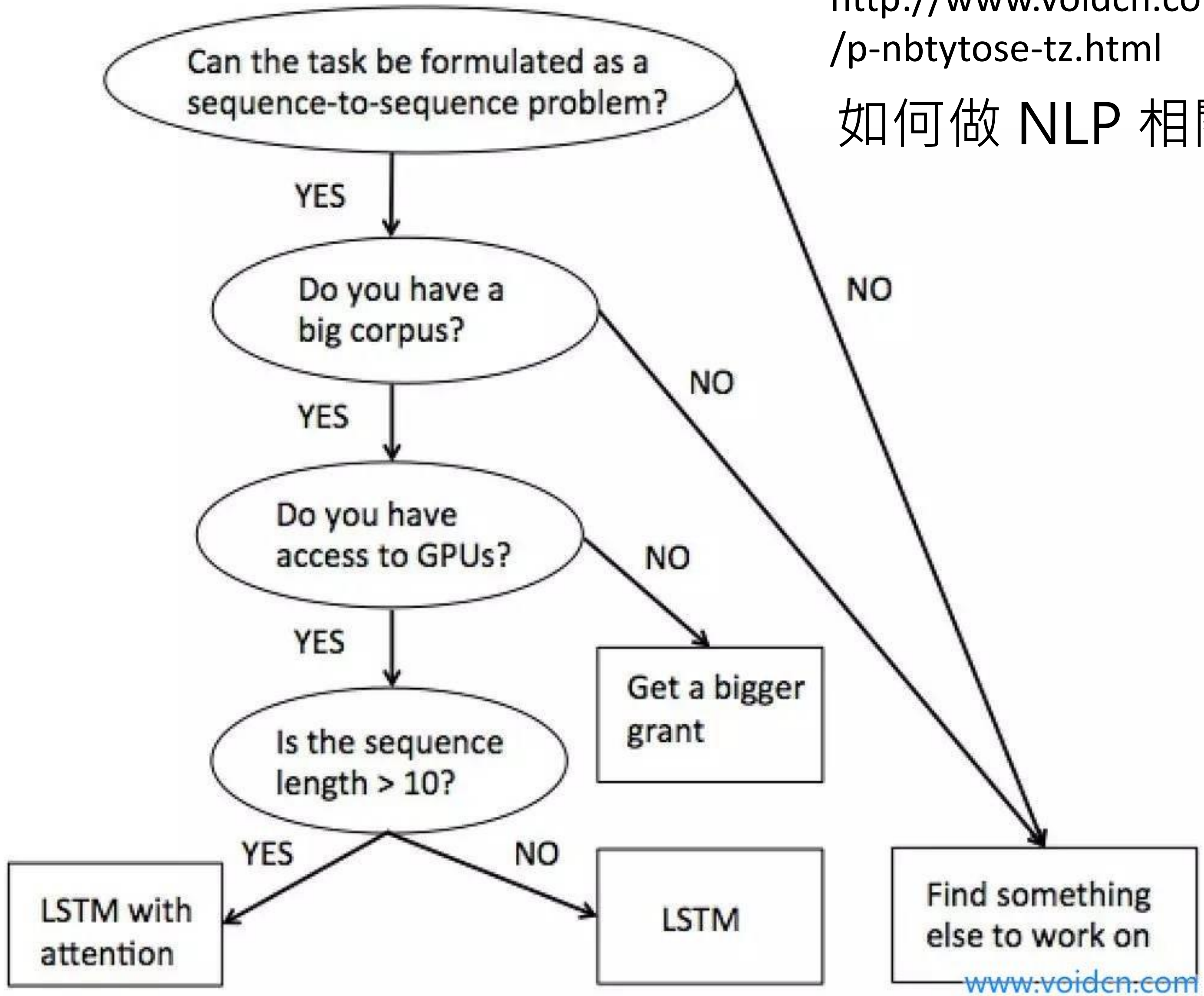
Improving Sequence Generation by GAN

Hung-yi Lee

http://www.voidcn.com/article/p-nbtytose-tz.html

如何做 NLP 相關研究



Outline

Improving Supervised Seq-to-seq Model

- RL (human feedback)
- GAN (discriminator feedback)

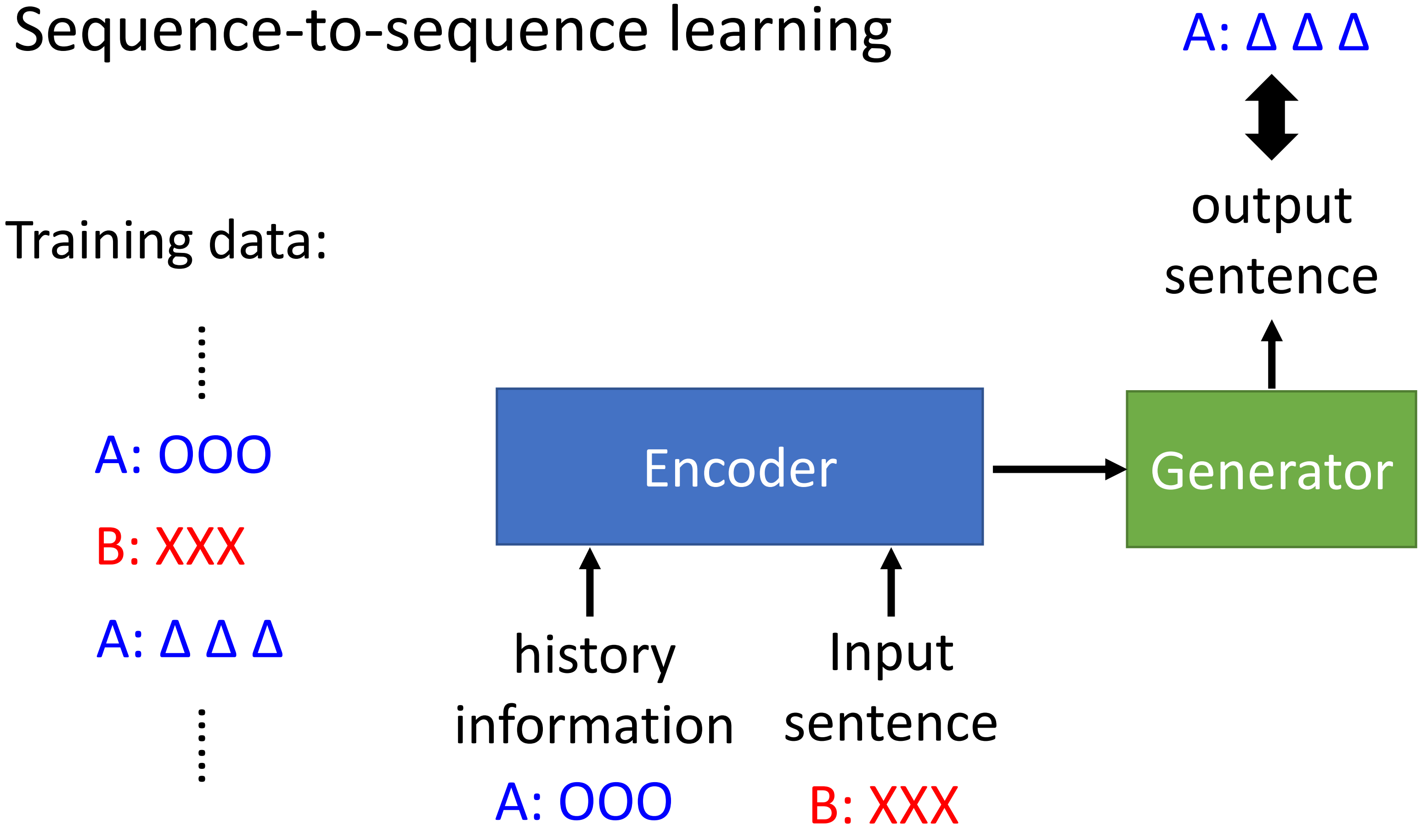
Unsupervised Seq-to-seq Model

- Summarization
- Translation

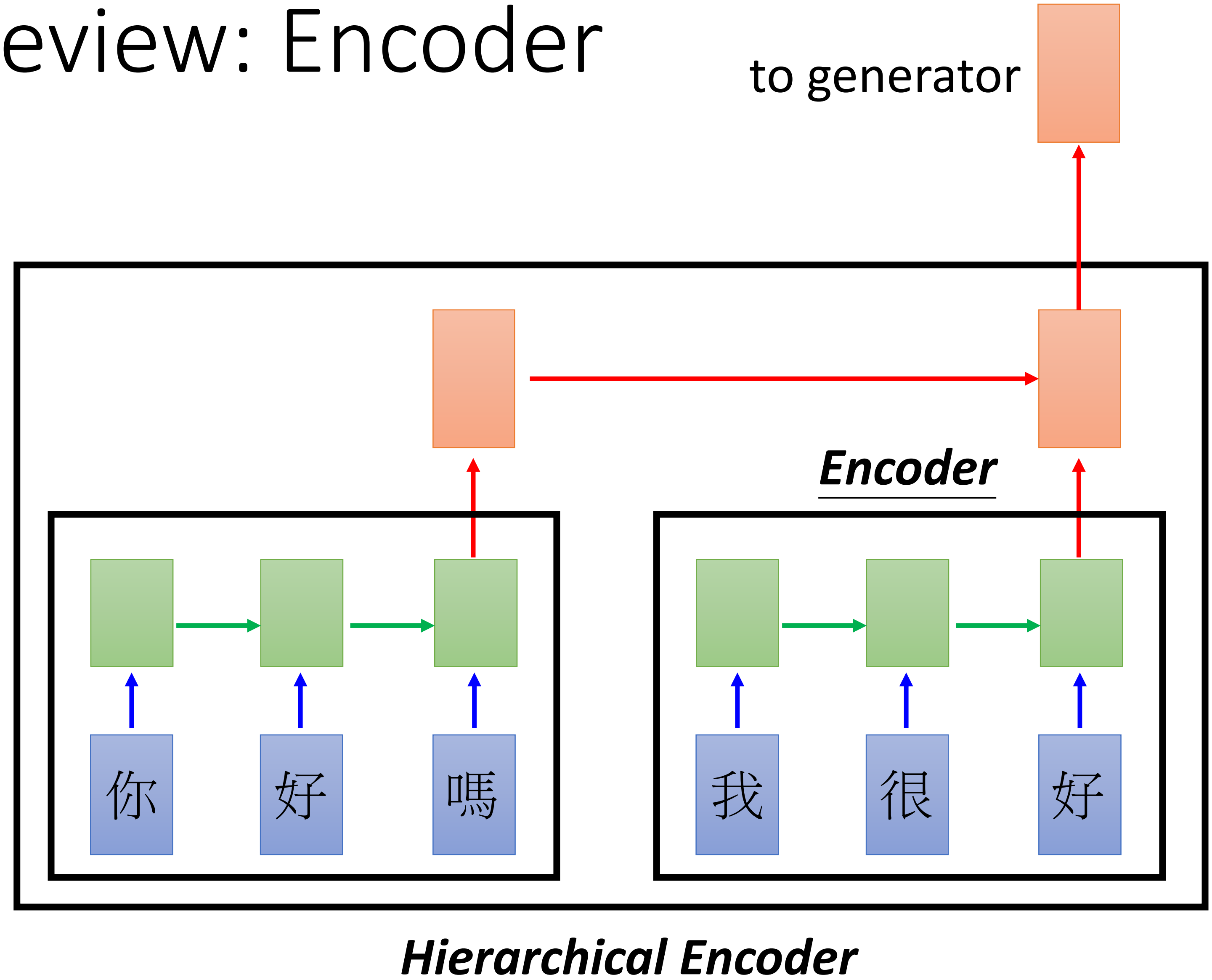
Text Style Transfer

Review: Chat-bot

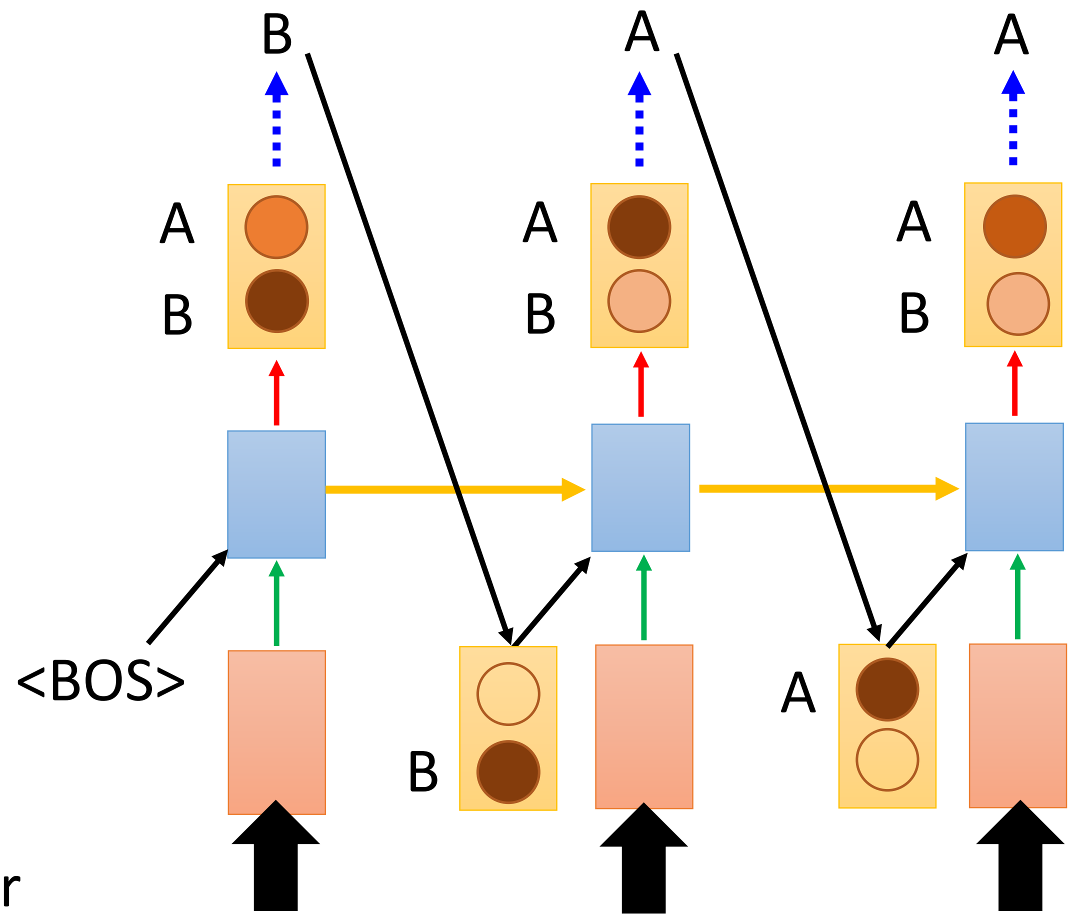
• Sequence-to-sequence learning



Review: Encoder



Review: Generator

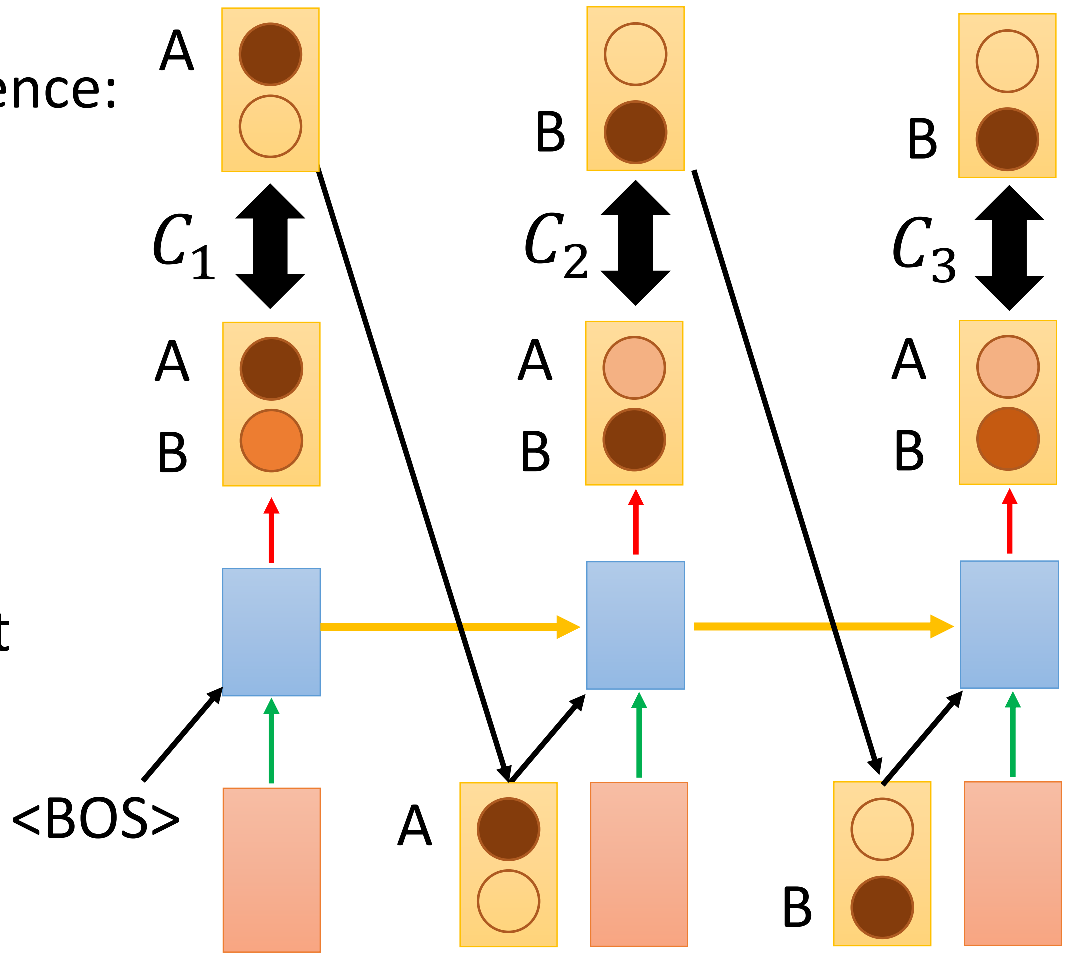


: condition from decoder

can be different with attention mechanism

Review: Training Generator

Reference:



$$C = \sum_t C_t$$

Minimizing cross-entropy of each component

: condition from decoder

Review: Maximum Likelihood

Training data: $(h, \hat{x})$

h: input sentence and history/context

$\hat{x}$: correct response (word sequence)

$\hat{x}_t$: t-th word, $\hat{x}_{1:t}$: first t words of $\hat{x}$

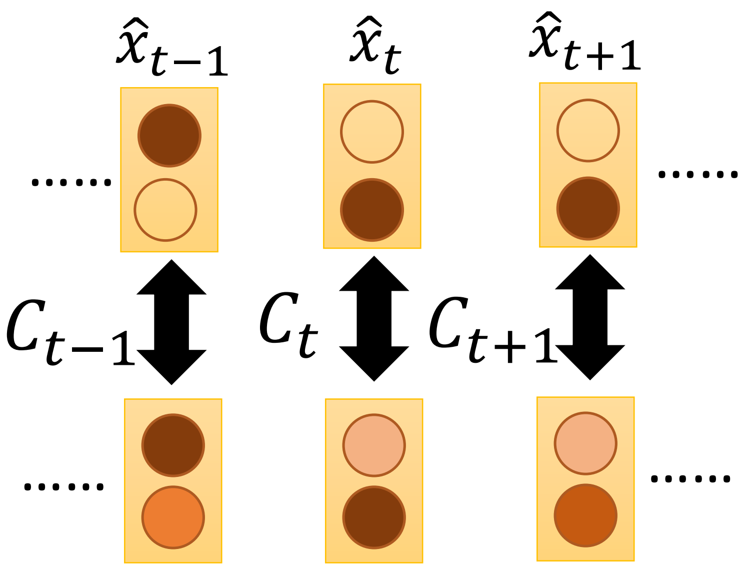
$$C = \sum_t C_t$$

$$C_t = -logP_\theta(\hat{x}_t|\hat{x}_{1:t-1}, h)$$

$$C = -\sum_t logP(\hat{x}_t|\hat{x}_{1:t-1}, h)$$

$$= -logP(\hat{x}_1|h)P(\hat{x}_2|\hat{x}_1, h) \cdots P(\hat{x}_T|\hat{x}_{1:T-1}, h)$$

$$= -logP(\hat{x}|h)$$



generator output

Maximizing the likelihood of generating $\hat{x}$ given h

Outline

Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, Dan Jurafsky, "Deep Reinforcement Learning for Dialogue Generation", EMNLP 2016

Improving Supervised Seq-to-seq Model

- RL (human feedback)
- GAN (discriminator feedback)

Unsupervised Seq-to-seq Model

- Summarization
- Translation

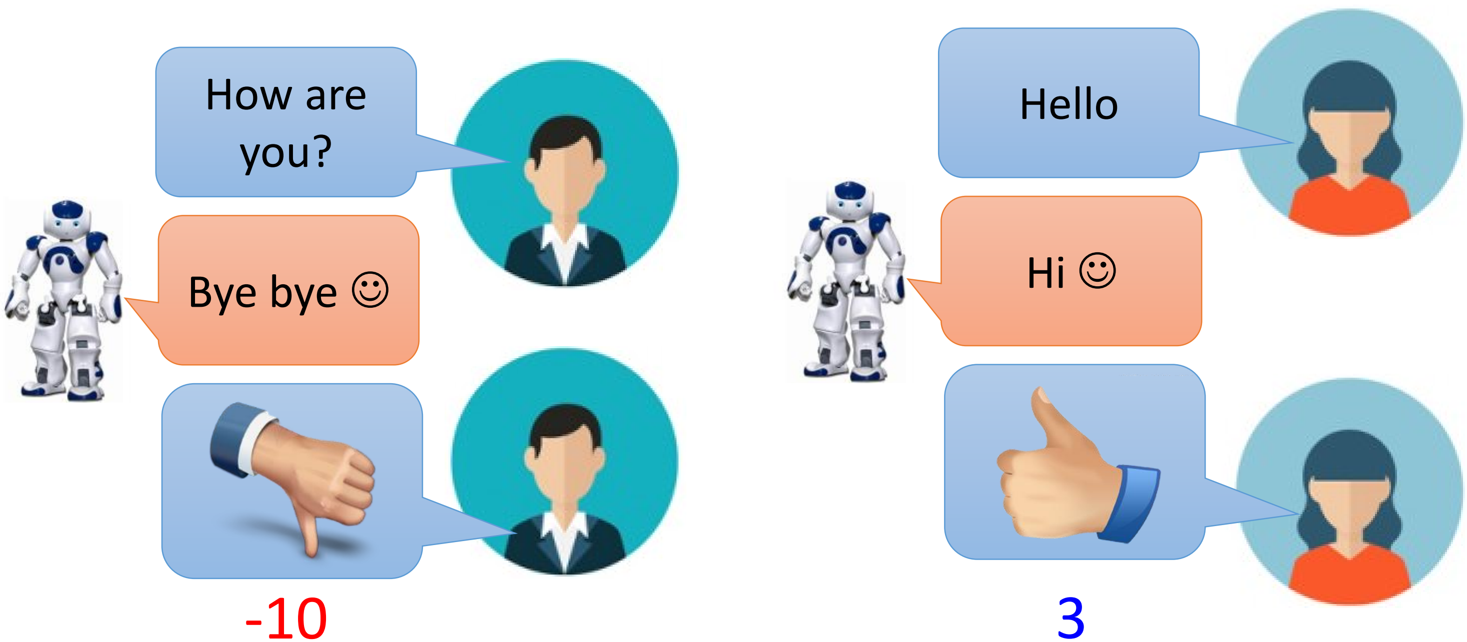
Text Style Transfer

Introduction

https://image.freepik.com/free-vector/variety-of-human-avatars_23-2147506285.jpg

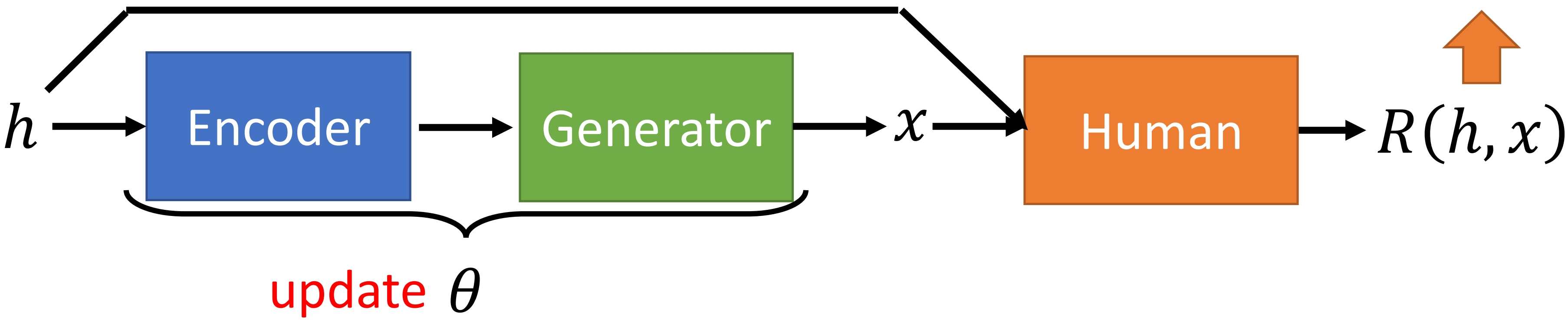
http://www.freepik.com/free-vector/variety-of-human-avatars_766615.htm

- Machine obtains feedback from user



- Chat-bot learns to maximize the ***expected reward***

Maximizing Expected Reward



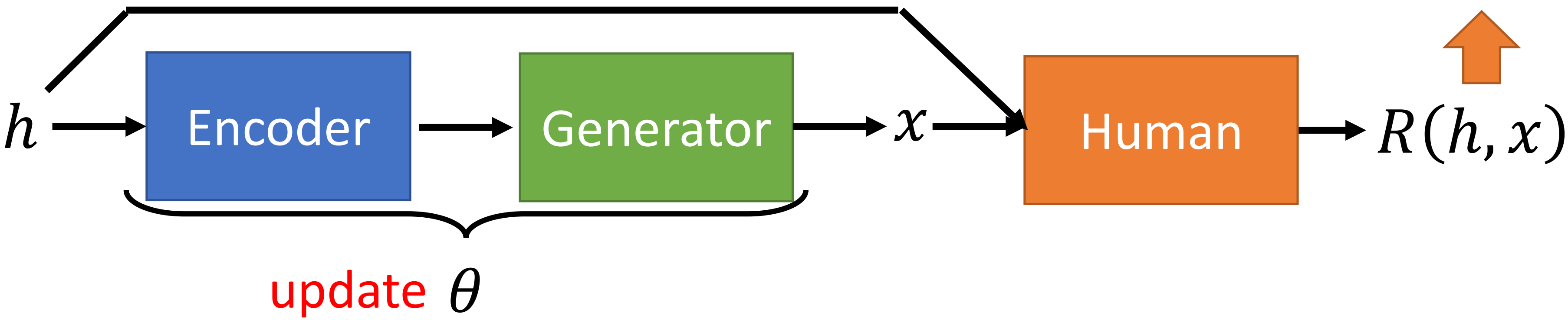
$$\theta^* = arg \max_{\theta} \bar{R}_{\theta}$$ Maximizing expected reward

$$\bar{R}_{\theta} = \sum_{h} P(h) \sum_{x} R(h, x)\, P_{\theta}(x|h)$$

Randomness in generator

Probability that the input/history is h

Maximizing Expected Reward



$$\theta^* = arg \max_{\theta} \bar{R}_\theta$$ Maximizing expected reward

$$\bar{R}_\theta = \sum_h P(h) \sum_x R(h,x)\, P_\theta(x|h) = E_{h\sim P(h)}\left[E_{x\sim P_\theta(x|h)}[R(h,x)]\right]$$

$$= E_{h\sim P(h), x\sim P_\theta(x|h)}[R(h,x)] \approx \frac{1}{N}\sum_{i=1}^{N} R\left(h^i, x^i\right)$$

Sample: $(h^1, x^1), (h^2, x^2), \cdots, (h^N, x^N)$

Where is θ?

Policy Gradient

$$\boxed{\frac{d log\big(f(x)\big)}{dx} = \frac{1}{f(x)}\frac{df(x)}{dx}}$$

$$\bar{R}_\theta = \sum_h P(h) \sum_x R(h,x)\, P_\theta(x|h) \approx \frac{1}{N}\sum_{i=1}^{N} R\big(h^i, x^i\big)$$

$$\nabla \bar{R}_\theta = \sum_h P(h) \sum_x R(h,x) \nabla P_\theta(x|h) \quad \approx \frac{1}{N}\sum_{i=1}^{N} R\big(h^i, x^i\big) \nabla log P_\theta(x|h)$$

Sampling

$$= \sum_h P(h) \sum_x R(h,x) P_\theta(x|h) \boxed{\frac{\nabla P_\theta(x|h)}{P_\theta(x|h)}}$$

$$= \sum_h P(h) \sum_x R(h,x) P_\theta(x|h) \boxed{\nabla log P_\theta(x|h)}$$

$$= E_{h\sim P(h), x\sim P_\theta(x|h)}[R(h,x) \nabla log P_\theta(x|h)]$$

Policy Gradient

- Gradient Ascent

$$\theta^{new} \leftarrow \theta^{old} + \eta \nabla \bar{R}_{\theta^{old}}$$

$$\nabla \bar{R}_\theta \approx \frac{1}{N} \sum_{i=1}^{N} R\left(h^i, x^i\right) \nabla log P_\theta\left(x^i | h^i\right)$$

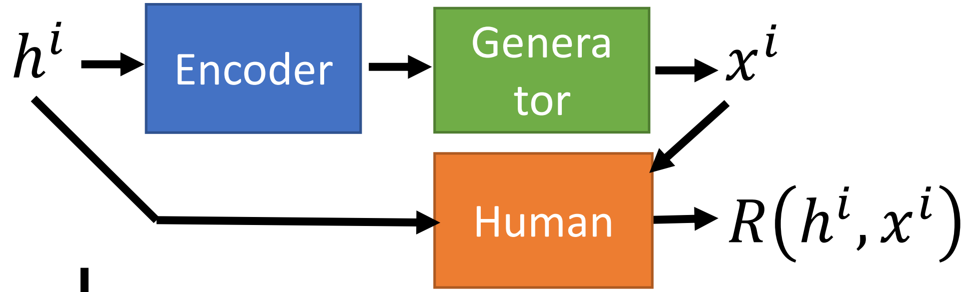
$R\left(h^i, x^i\right)$ is positive

After updating θ, $P_\theta\left(x^i | h^i\right)$ will increase

$R\left(h^i, x^i\right)$ is negative

After updating θ, $P_\theta\left(x^i | h^i\right)$ will decrease

Implementation

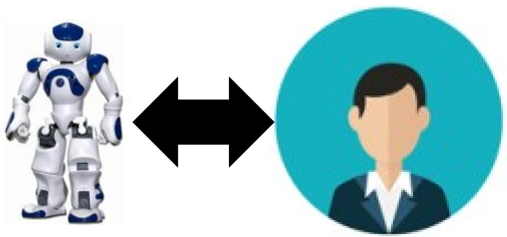


	Maximum Likelihood	Reinforcement Learning
Objective Function	$\frac{1}{N}\sum_{i=1}^{N} logP_\theta(\hat{x}^i \mid h^i)$	$\frac{1}{N}\sum_{i=1}^{N} R(h^i, x^i) logP_\theta(x^i \mid h^i)$
Gradient	$\frac{1}{N}\sum_{i=1}^{N} \nabla logP_\theta(\hat{x}^i \mid h^i)$	$\frac{1}{N}\sum_{i=1}^{N} R(h^i, x^i) \nabla logP_\theta(x^i \mid h^i)$
Training Data	$\{(h^1, \hat{x}^1), \dots, (h^N, \hat{x}^N)\}$ $R(h^i, \hat{x}^i) = 1$	$\{(h^1, x^1), \dots, (h^N, x^N)\}$ Sampling as training data weighted by $R(h^i, x^i)$

Implementation

θ^0 can be well pre-trained from $\{(h^1, \hat{x}^1), \dots, (h^N, \hat{x}^N)\}$

θ^t



(h^1, x^1) $R(h^1, x^1)$

(h^2, x^2) $R(h^2, x^2)$

$\vdots$ $\vdots$

(h^N, x^N) $R(h^N, x^N)$

New Objective:

$$\frac{1}{N}\sum_{i=1}^{N} R(h^i, x^i) log P_\theta(x^i | h^i)$$

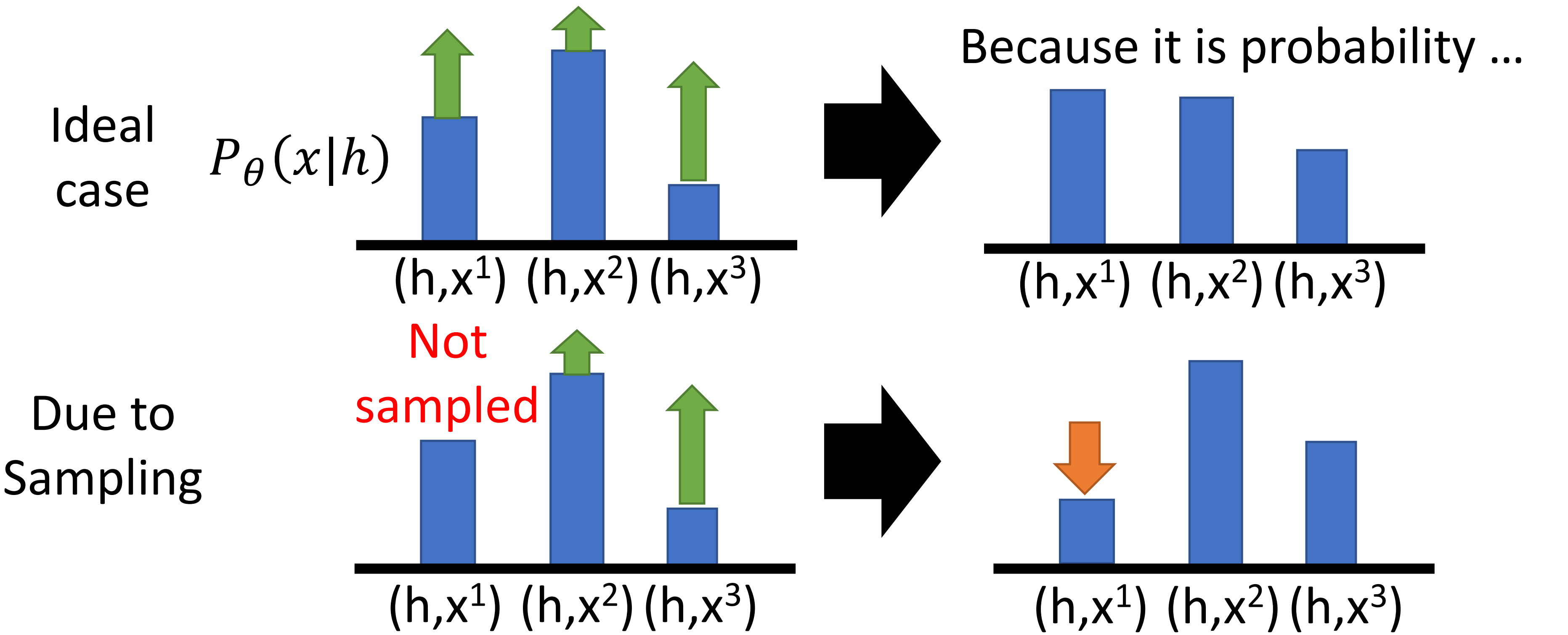
$$\theta^{t+1} \leftarrow \theta^t + \eta \boxed{\nabla \bar{R}_{\theta^t}}$$

$$\frac{1}{N}\sum_{i=1}^{N} R(h^i, x^i) \nabla log P_{\theta^t}(x^i | h^i)$$

Add a Baseline

If $R(h^i, x^i)$ is always positive

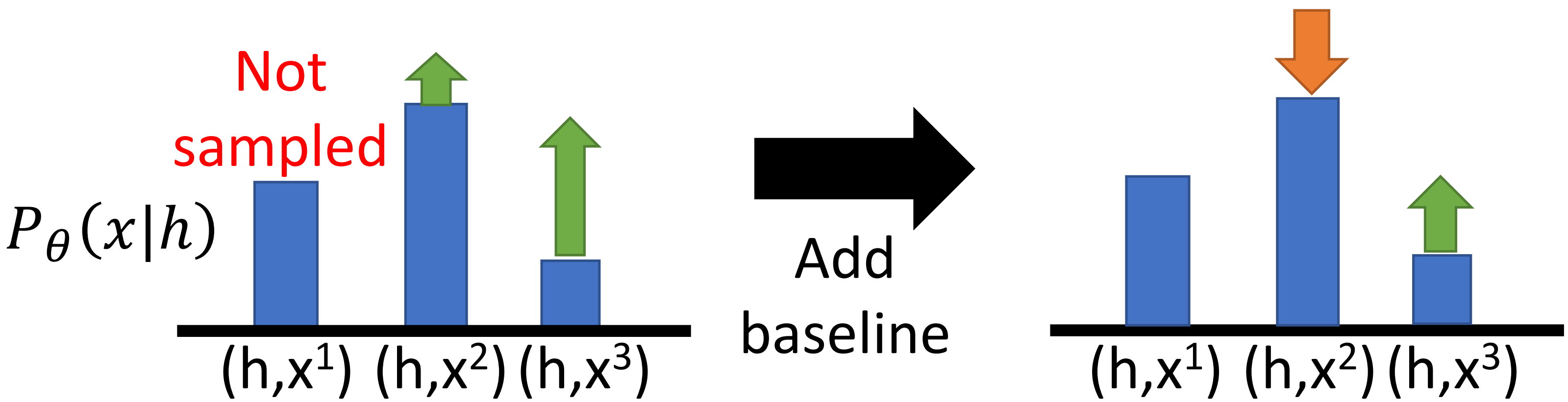
$$\frac{1}{N}\sum_{i=1}^{N} R(h^i, x^i) log\nabla P_\theta(x^i|h^i)$$



Add a Baseline

If $R\left(h^i, x^i\right)$ is always positive

$$\frac{1}{N}\sum_{i=1}^{N} R\left(h^i, x^i\right) log\nabla P_\theta\left(x^i|h^i\right) \Rightarrow \frac{1}{N}\sum_{i=1}^{N} \left(R\left(h^i, x^i\right) \boxed{-\, b}\right) log\nabla P_\theta\left(x^i|h^i\right)$$



There are several ways to obtain the baseline b.

Alpha GO style training !

- Let two agents talk to each other

How old are you?

See you.

See you.

See you.

How old are you?

I am 16.

I though you were 12.

What make you think so?

Using a pre-defined evaluation function to compute R(h,x)

Example Reward

- The final reward R(h,x) is the weighted sum of three terms $r_1(h,x)$, $r_2(h,x)$ and $r_3(h,x)$

$$R(h,x) = \lambda_1 r_1(h,x) + \lambda_2 r_2(h,x) + \lambda_3 r_3(h,x)$$

Ease of answering

不要成為句點王

Information Flow

說點新鮮的

Semantic Coherence

不要前言不對後語

Example Results

Baseline mutual information model (Li et al. 2015)	Proposed reinforcement learning model
...	...
...	...

Outline

Improving Supervised Seq-to-seq Model

- RL (human feedback)
- GAN (discriminator feedback)

Unsupervised Seq-to-seq Model

- Summarization
- Translation

Text Style Transfer

http://www.nipic.com/show/3/83/3936650kd7476069.html

Basic Idea – Chat-bot

Input sentence/history h → En De Chatbot → response sentence x

Input sentence/history h
response sentence x
→ Discriminator → Real or fake

human dialogues

Conditional GAN

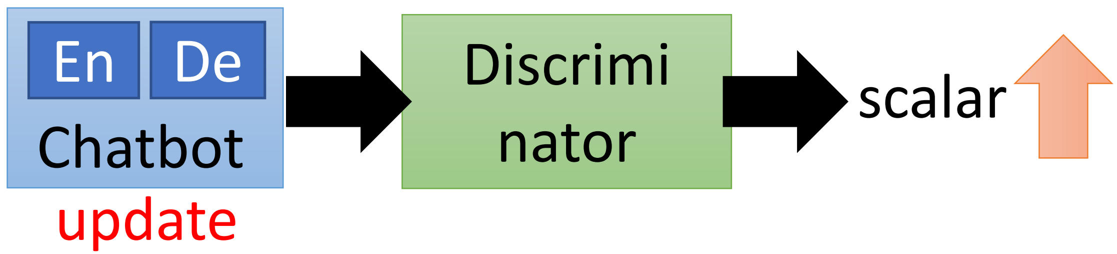
Algorithm – Chat-bot

Training data:

h: A: OOO / B: XXX

x: A: ∆ ∆ ∆

- Initialize generator Gen and discriminator Dis
- In each iteration:
  - Sample real history h and sentence x from database
  - Sample real history h' from database, and generate sentences $\tilde{x}$ by Gen(h')
  - Update Dis to increase $Dis(h,x)$ and decrease $Dis(h',\tilde{x})$
  - Update Gen such that

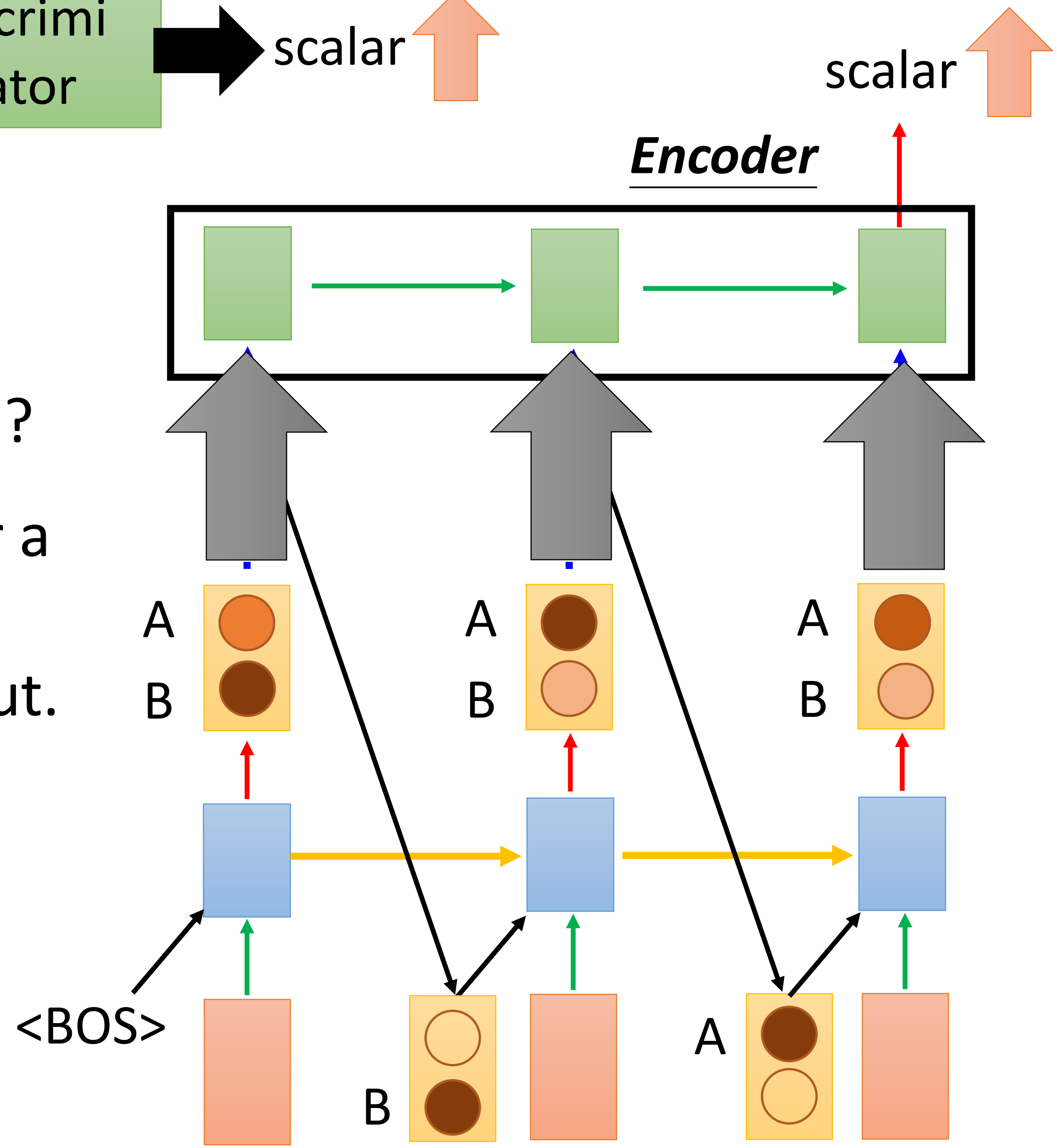


Can we do backpropogation?

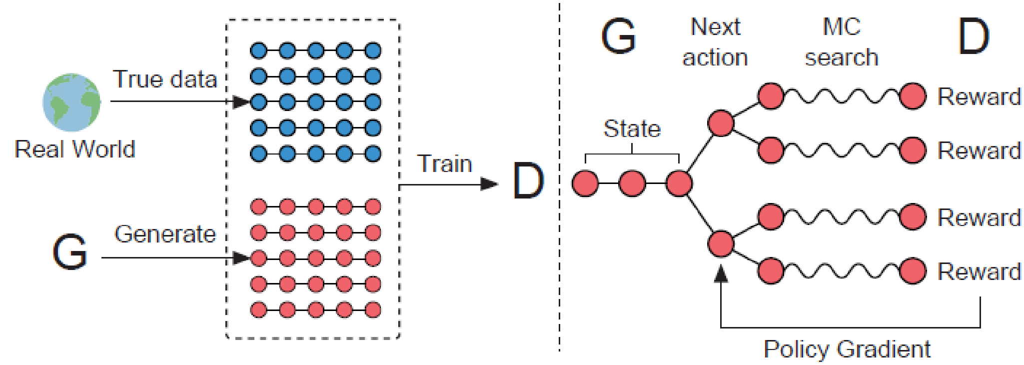
Tuning generator a little bit will not change the output.

Alternative: improved WGAN

(ignoring sampling process)

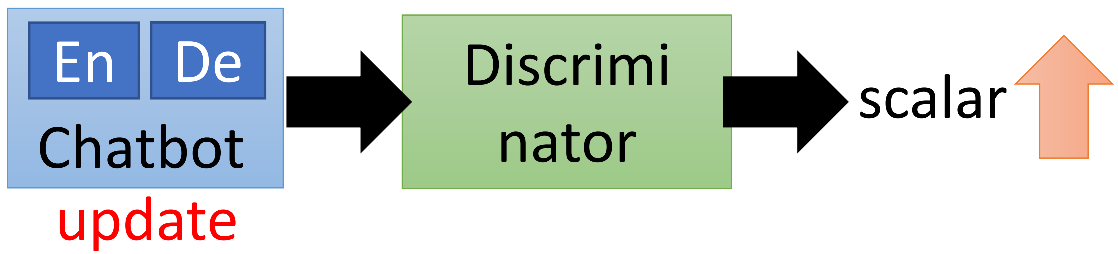


Alternatives



- Gumbel-softmax
  - Matt J. Kusner, José Miguel Hernández-Lobato, "GANS for Sequences of Discrete Elements with the Gumbel-softmax Distribution", arXiv 2016
- MaliGAN
  - Tong Che, Yanran Li, Ruixiang Zhang, R Devon Hjelm, Wenjie Li, Yangqiu Song, Yoshua Bengio, "Maximum-Likelihood Augmented Discrete Generative Adversarial Networks", arXiv 2017
- SeqGAN
  - Lantao Yu, Weinan Zhang, Jun Wang, Yong Yu, "SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient", AAAI 2017
  - Jiwei Li, Will Monroe, Tianlin Shi, Sébastien Jean, Alan Ritter, Dan Jurafsky, "Adversarial Learning for Neural Dialogue Generation", arXiv 2017

Reinforcement Learning?



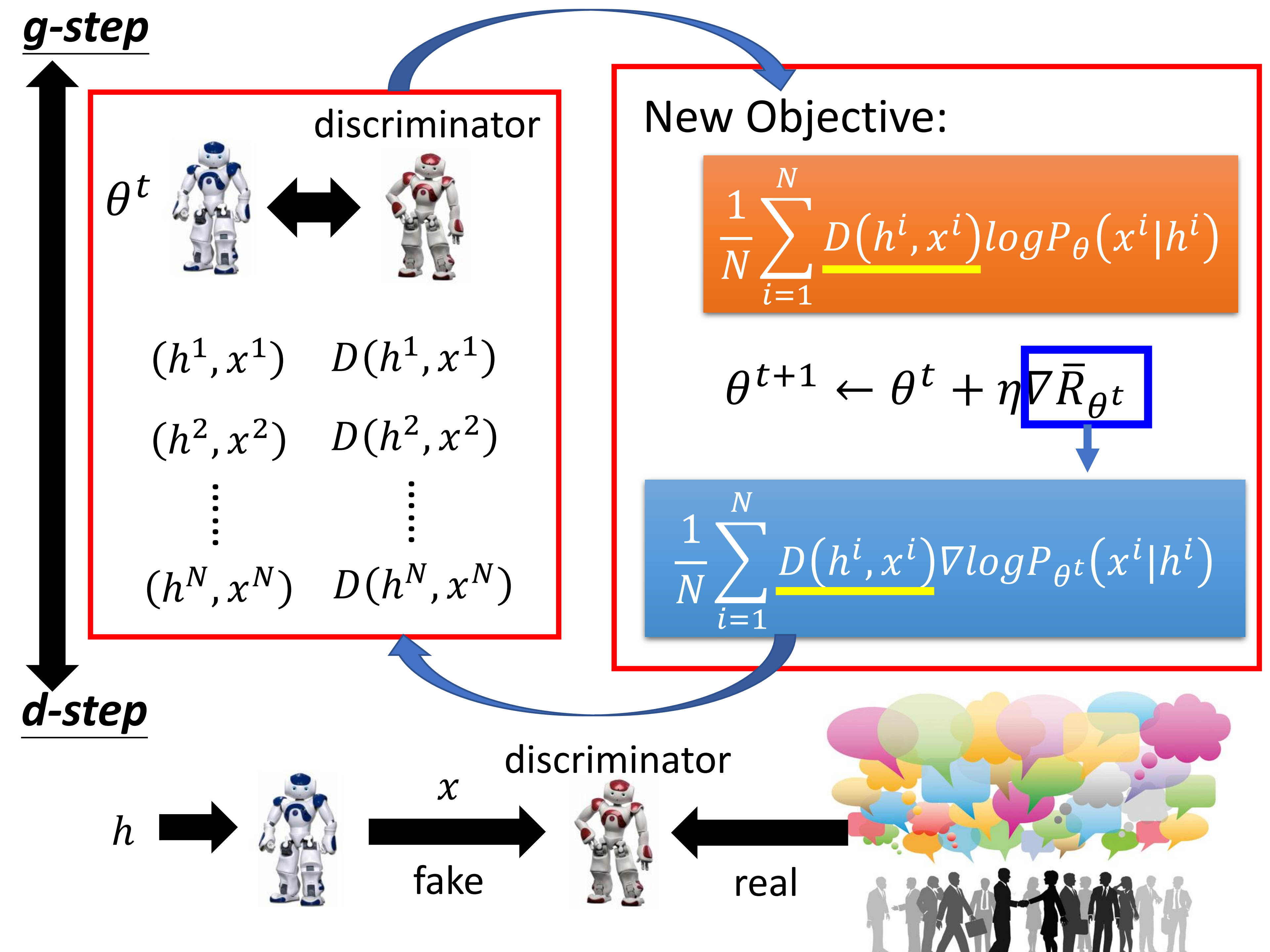
- Consider the output of discriminator as reward
  - Update generator to increase discriminator = to get maximum reward

reward

$$\nabla \bar{R}_\theta \approx \frac{1}{N}\sum_{i=1}^{N}\left(\boxed{D\left(h^i, x^i\right)} - b\right)\nabla log P_\theta\left(x^i | h^i\right)$$

Discriminator Score

- Different from typical RL
  - The discriminator would update


g-step
d-step
discriminator
θ^t
(h^1, x^1) $D(h^1, x^1)$
(h^2, x^2) $D(h^2, x^2)$
(h^N, x^N) $D(h^N, x^N)$
New Objective:
$\frac{1}{N}\sum_{i=1}^{N} D(h^i, x^i) log P_\theta(x^i|h^i)$
$\theta^{t+1} \leftarrow \theta^t + \eta \boxed{\nabla \bar{R}_{\theta^t}}$
$\frac{1}{N}\sum_{i=1}^{N} D(h^i, x^i) \nabla log P_{\theta^t}(x^i|h^i)$
discriminator
h
x
fake
real

Example Results

感謝 段逸林 同學提供實驗結果

input	I love you.

input	Do you like machine learning?

input	I thought I have met you before.

input	Let's go to the party.

input	How do you feel about the president?

	Human Evaluation
MLE	52.6%
SeqGAN	56.9%
ESGAN	60.9%

Tips: Reward for Every Generation Step

$$\nabla \bar{R}_\theta \approx \frac{1}{N}\sum_{i=1}^{N}\left(D\left(h^i, x^i\right) - b\right)\nabla logP_\theta\left(x^i|h^i\right)$$

h^i = "What is your name?"

x^i = "I don't know"

$D\left(h^i, x^i\right) - b$ is negative

Update θ to decrease log$P_\theta\left(x^i|h^i\right)$

$$logP_\theta\left(x^i|h^i\right) = logP\left(x_1^i|h^i\right) + logP\left(x_2^i|h^i, x_1^i\right) + logP\left(x_3^i|h^i, x_{1:2}^i\right)$$

$P\left("I"|h^i\right)$?

h^i = "What is your name?"

x^i = "I am John"

$D\left(h^i, x^i\right) - b$ is positive

Update θ to increase log$P_\theta\left(x^i|h^i\right)$

$$logP_\theta\left(x^i|h^i\right) = logP\left(x_1^i|h^i\right) + logP\left(x_2^i|h^i, x_1^i\right) + logP\left(x_3^i|h^i, x_{1:2}^i\right)$$

$P\left("I"|h^i\right)$

Tips: Reward for Every Generation Step

h^i = "What is your name?" $\quad$ x^i = "I don't know"

$$logP_\theta(x^i|h^i) = logP(x_1^i|h^i) + logP(x_2^i|h^i, x_1^i) + logP(x_3^i|h^i, x_{1:2}^i)$$

$$P(\text{"}I\text{"}|h^i) \qquad P(\text{"}don't\text{"}|h^i, \text{"}I\text{"}) \qquad P(\text{"}know\text{"}|h^i, \text{"}I\ don't\text{"})$$

$$\nabla \bar{R}_\theta \approx \frac{1}{N}\sum_{i=1}^{N}\left(D(h^i, x^i) - b\right)\nabla logP_\theta(x^i|h^i)$$

$$\nabla \bar{R}_\theta \approx \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Q(h^i, x_{1:t}^i) - b\right)\nabla logP_\theta(x_t^i|h^i, x_{1:t-1}^i)$$

Method 1. Monte Carlo (MC) Search

Method 2. Discriminator For Partially Decoded Sequences

Tips: Monte Carlo Search

- How to estimate $Q\left(h^i, x_{1:t}^i\right)$?

$$\underset{h^i}{Q("What\ is\ your\ name?"}, \underset{x_1^i}{"I")}$$

A roll-out generator for sampling is needed

$x^A =$ I am John $\quad D\left(h^i, x^A\right) = 1.0$

$x^B =$ I am happy $\quad D\left(h^i, x^B\right) = 0.1$

$x^C =$ I don't know $\quad D\left(h^i, x^C\right) = 0.1$

$x^D =$ I am superman $\quad D\left(h^i, x^D\right) = 0.8$

avg

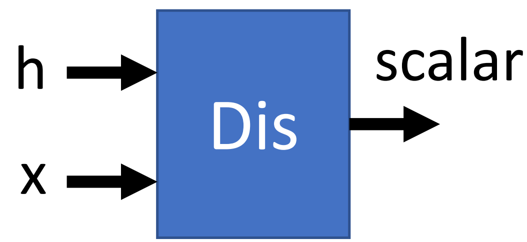
$Q\left(h^i, "I"\right) = 0.5$

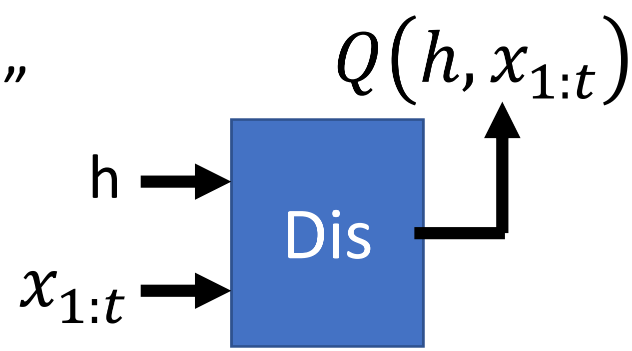
Tips: Rewarding Partially Decoded Sequences

- Training a discriminator that is able to assign rewards to both fully and partially decoded sequences
  - Break generated sequences into partial sequences

+ h="What is your name?", x="I am john"
+ h="What is your name?", x="I am"
+ h="What is your name?", x="I"

- h="What is your name?", x="I don't know"
- h="What is your name?", x="I don't"
- h="What is your name?", x="I"





Tips: Adding Good Examples

- The training of generative model is unstable
  - This reward is used to promote or discourage the generator's own generated sequences.
  - Usually It knows that the generated results are bad, but does not know what results are good.

Training Data for SeqGAN: $\{(h^1, x^1), \ldots, (h^N, x^N)\}$

Obtained by sampling

weighted by $D(h^i, x^i)$

Adding more Data: $\{(h^1, \hat{x}^1), \ldots, (h^N, \hat{x}^N)\}$ Real data

Consider $D(h^i, \hat{x}^i) = 1$

Tips: RankGAN

Kevin Lin, Dianqi Li, Xiaodong He, Zhengyou Zhang, Ming-Ting Sun, "Adversarial Ranking for Language Generation", NIPS 2017

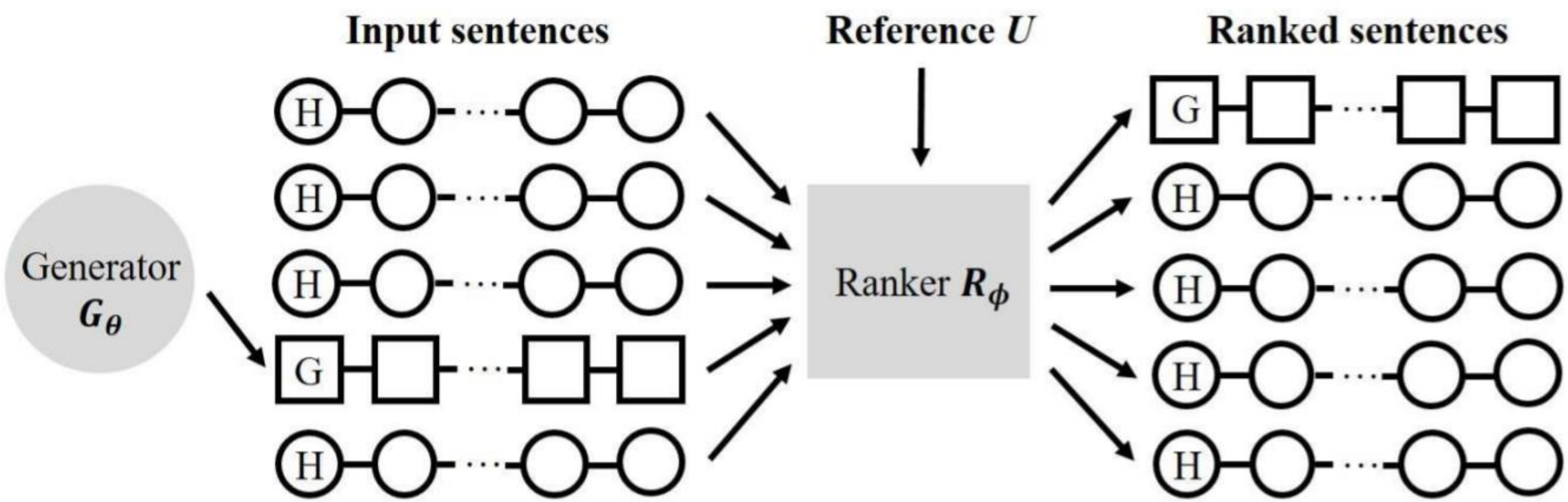


Image caption generation:

Method	BLEU-2	BLEU-3	BLEU-4
MLE	0.781	0.624	0.589
SeqGAN	0.815	0.636	0.587
RankGAN	**0.845**	**0.668**	**0.614**

Method	Human score
SeqGAN	3.44
RankGAN	4.61
Human-written	**6.42**

More Applications

- Supervised machine translation
  - Lijun Wu, Yingce Xia, Li Zhao, Fei Tian, Tao Qin, Jianhuang Lai, Tie-Yan Liu, “Adversarial Neural Machine Translation”, arXiv 2017
  - Zhen Yang, Wei Chen, Feng Wang, Bo Xu, „Improving Neural Machine Translation with Conditional Sequence Generative Adversarial Nets“, arXiv 2017
- Supervised abstractive summarization
  - Linqing Liu, Yao Lu, Min Yang, Qiang Qu, Jia Zhu, Hongyan Li, “Generative Adversarial Network for Abstractive Text Summarization”, AAAI 2018
- Image/video caption generation
  - Rakshith Shetty, Marcus Rohrbach, Lisa Anne Hendricks, Mario Fritz, Bernt Schiele, “Speaking the Same Language: Matching Machine to Human Captions by Adversarial Training”, ICCV 2017
  - Xiaodan Liang, Zhiting Hu, Hao Zhang, Chuang Gan, Eric P. Xing, “Recurrent Topic-Transition GAN for Visual Paragraph Generation”, arXiv 2017

Outline

Improving Supervised Seq-to-seq Model

- RL (human feedback)
- GAN (discriminator feedback)

Unsupervised Seq-to-seq Model

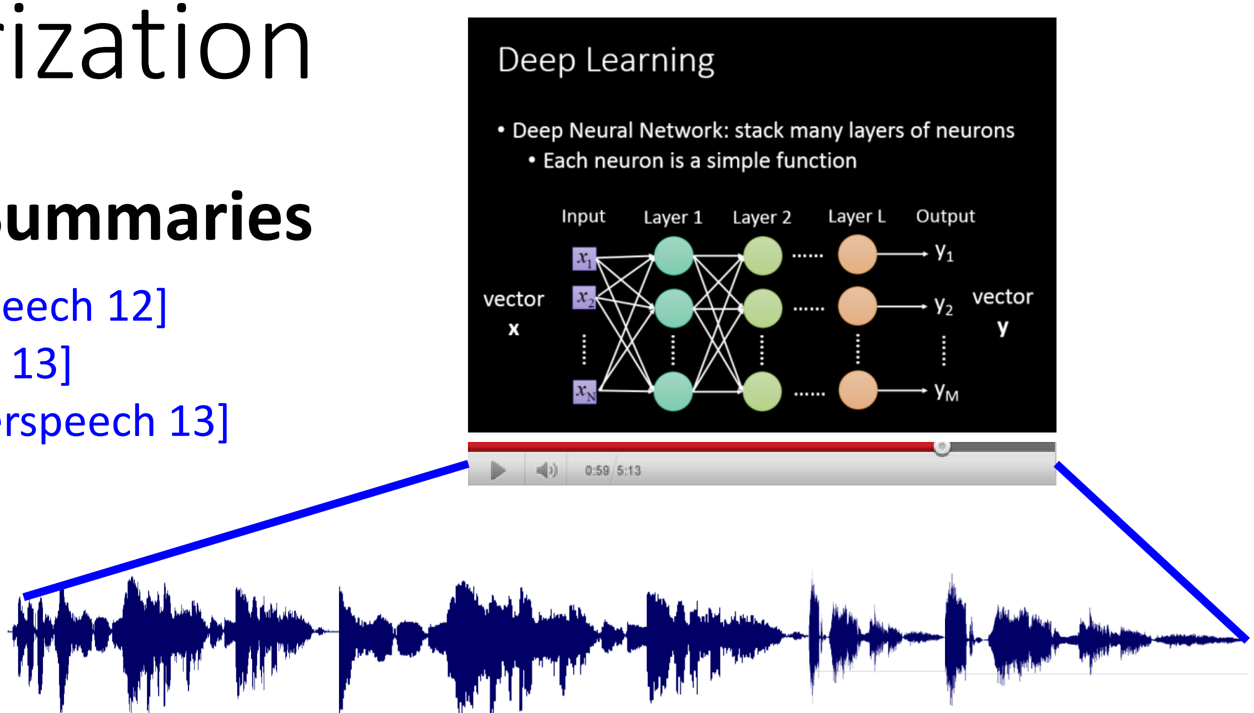
- Summarization
- Translation

Text Style Transfer

Summarization

Extractive Summaries

[Lee, et al., Interspeech 12]
[Lee, et al., ICASSP 13]
[Shiang, et al., Interspeech 13]



Audio File
to be summarized

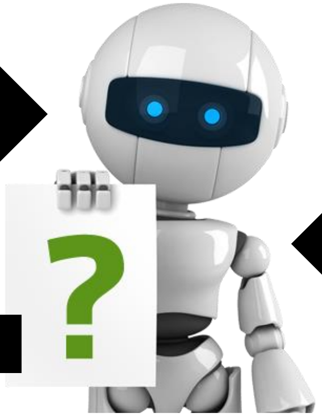
…… deep learning is powerful …… …… ……

This is the summary.

- ➢ Select the most informative segments to form a compact version
- ➢ Machine does not write summaries in its own words

Abstractive Summarization

- Now machine can do **abstractive summary** (write summaries in its own words)



Training Data

Title 1

Title 2

Title 3

title generated by machine

(in its own words)

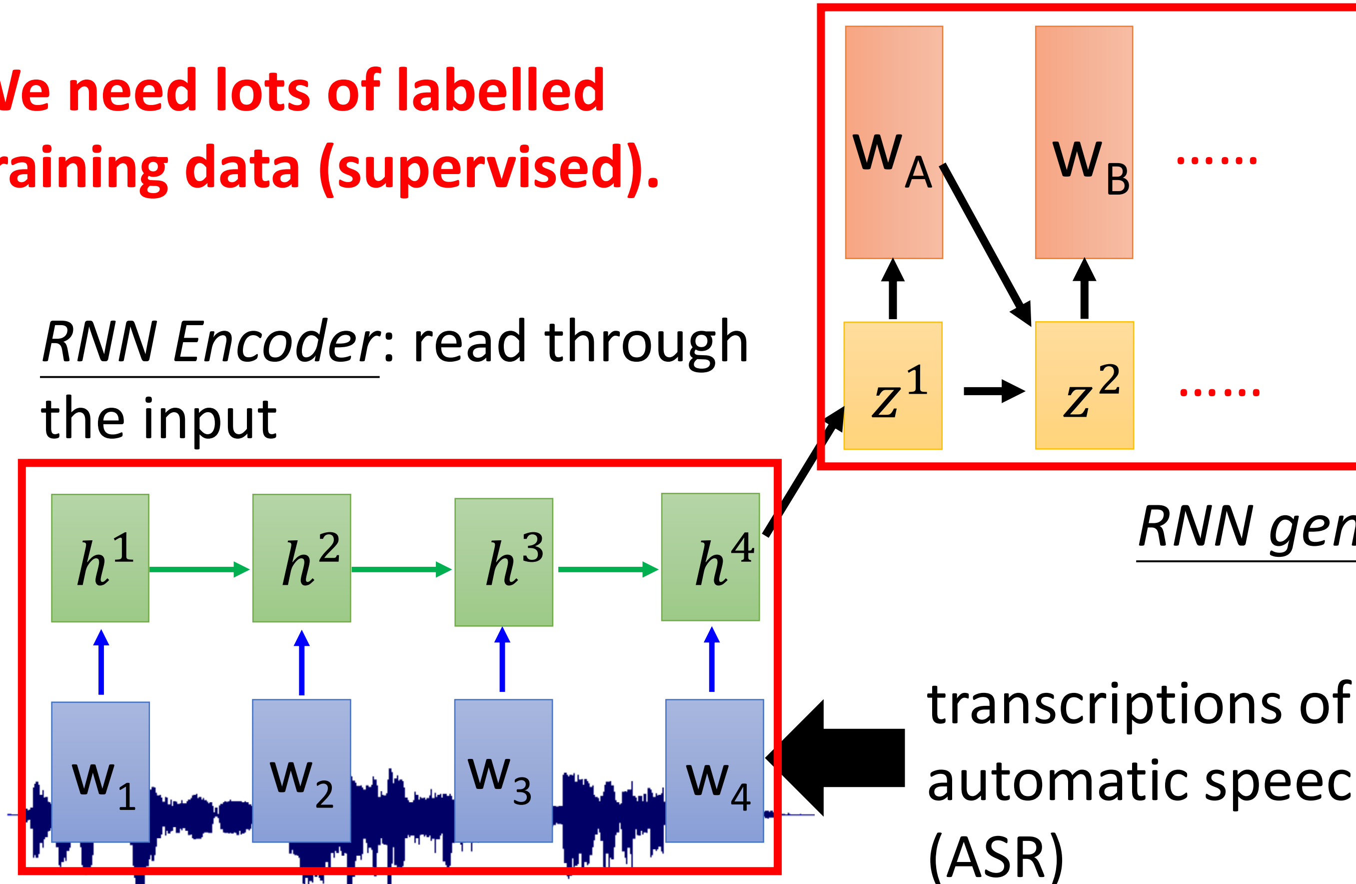
without hand-crafted rules

Abstractive Summarization

- Input: transcriptions of audio, output: summary

We need lots of labelled training data (supervised).

RNN Encoder: read through the input



RNN generator

transcriptions of audio from automatic speech recognition (ASR)

Unsupervised Abstractive Summarization

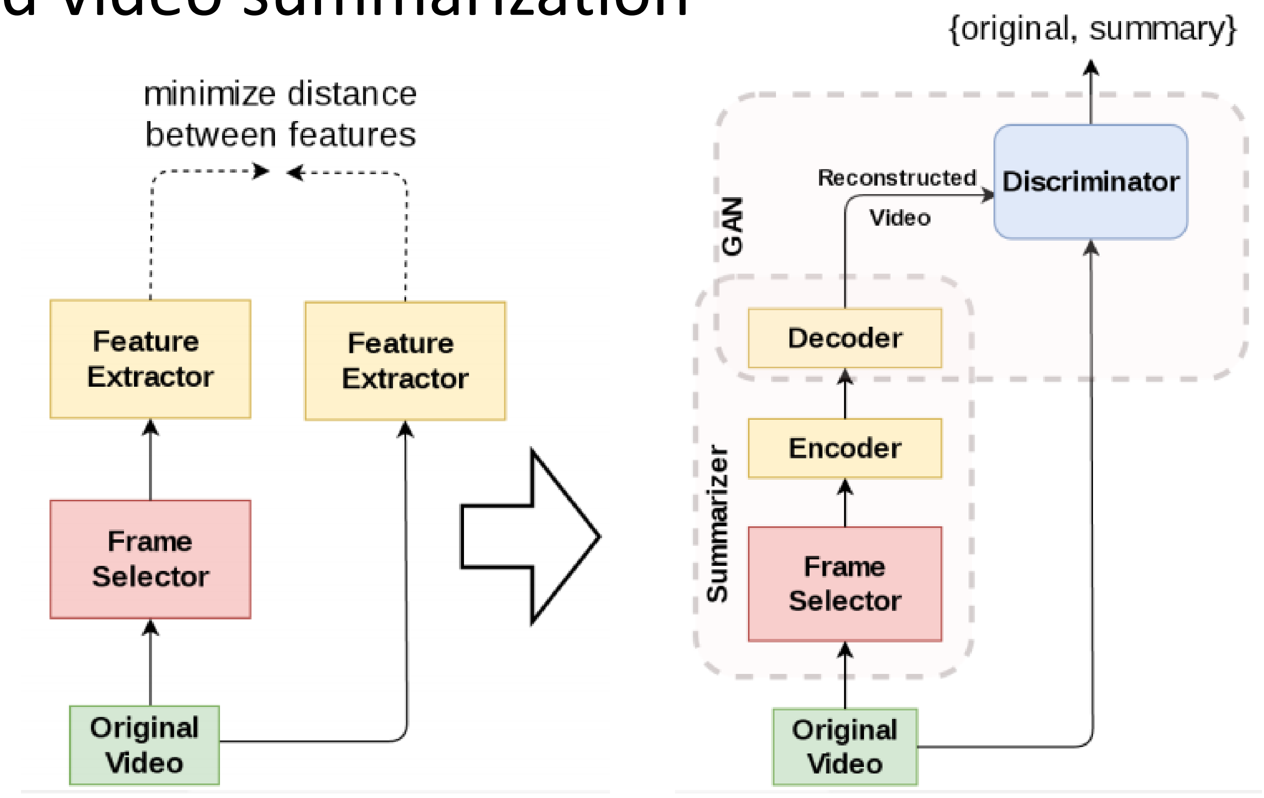
- ***Document***:澳大利亞今天與13個國家簽署了反興奮劑雙邊協議,旨在加強體育競賽之外的藥品檢查並共享研究成果 ……
- ***Summary***:
  - Human:澳大利亞與13國簽署反興奮劑協議
  - Unsupervised:澳大利亞加強體育競賽之外的藥品檢查
- ***Document***:中華民國奧林匹克委員會今天接到一九九二年冬季奧運會邀請函,由於主席張豐緒目前正在中南美洲進行友好訪問,因此尚未決定是否派隊赴賽 ……
- ***Summary***:
  - Human:一九九二年冬季奧運會函邀我參加
  - Unsupervised:奧委會接獲冬季奧運會邀請函

Unsupervised Abstractive Summarization

- ***Document***:據此間媒體27日報道,印度尼西亞蘇門答臘島的兩個省近日來連降暴雨,洪水泛濫導致塌方,到26日為止至少已有60人喪生,100多人失蹤 ……
- ***Summary***:
  - Human:印尼水災造成60人死亡
  - Unsupervised:印尼門洪水泛濫導致塌雨
- ***Document***:安徽省合肥市最近為領導幹部下基層做了新規定:一律輕車簡從,不準搞迎來送往、不準搞層層陪同 ……
- ***Summary***:
  - Human:合肥規定領導幹部下基層活動從簡
  - Unsupervised:合肥領導幹部下基層做搞迎來送往規定:一律簡

More Applications

- Unsupervised video summarization



Behrooz Mahasseni, Michael Lam and Sinisa Todorovic, "Unsupervised Video Summarization with Adversarial LSTM Networks", CVPR, 2017

Outline of Part II

Improving Supervised Seq-to-seq Model

- RL (human feedback)
- GAN (discriminator feedback)

Unsupervised Seq-to-seq Model

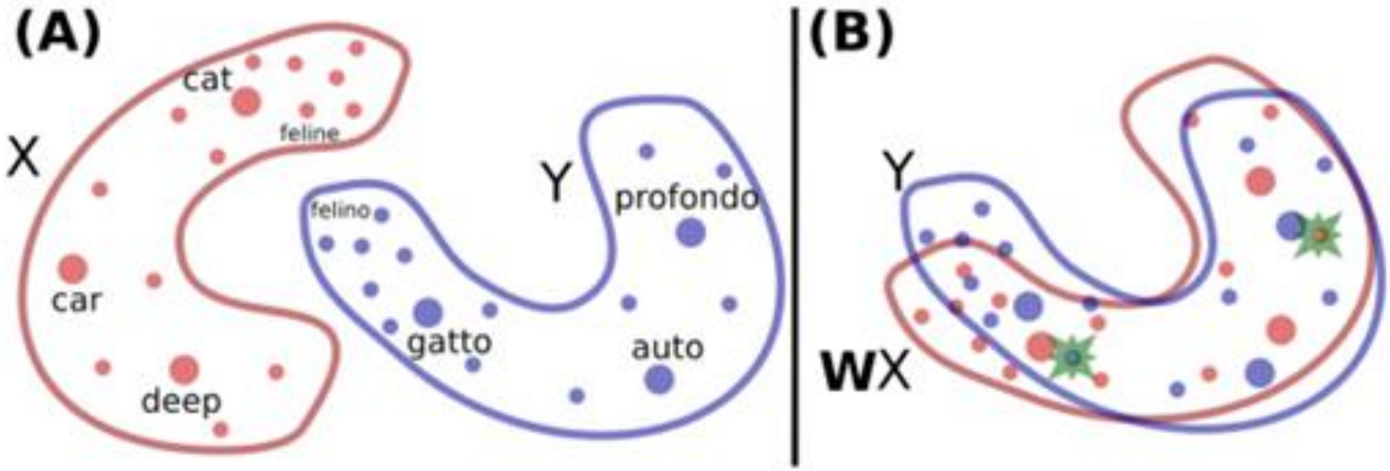
- Summarization
- Translation

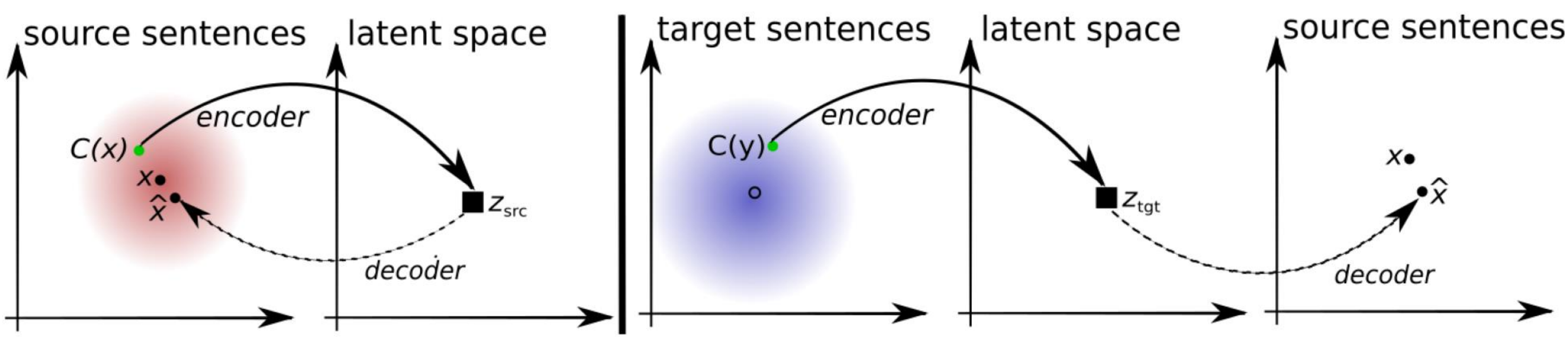
Text Style Transfer

Unsupervised Translation

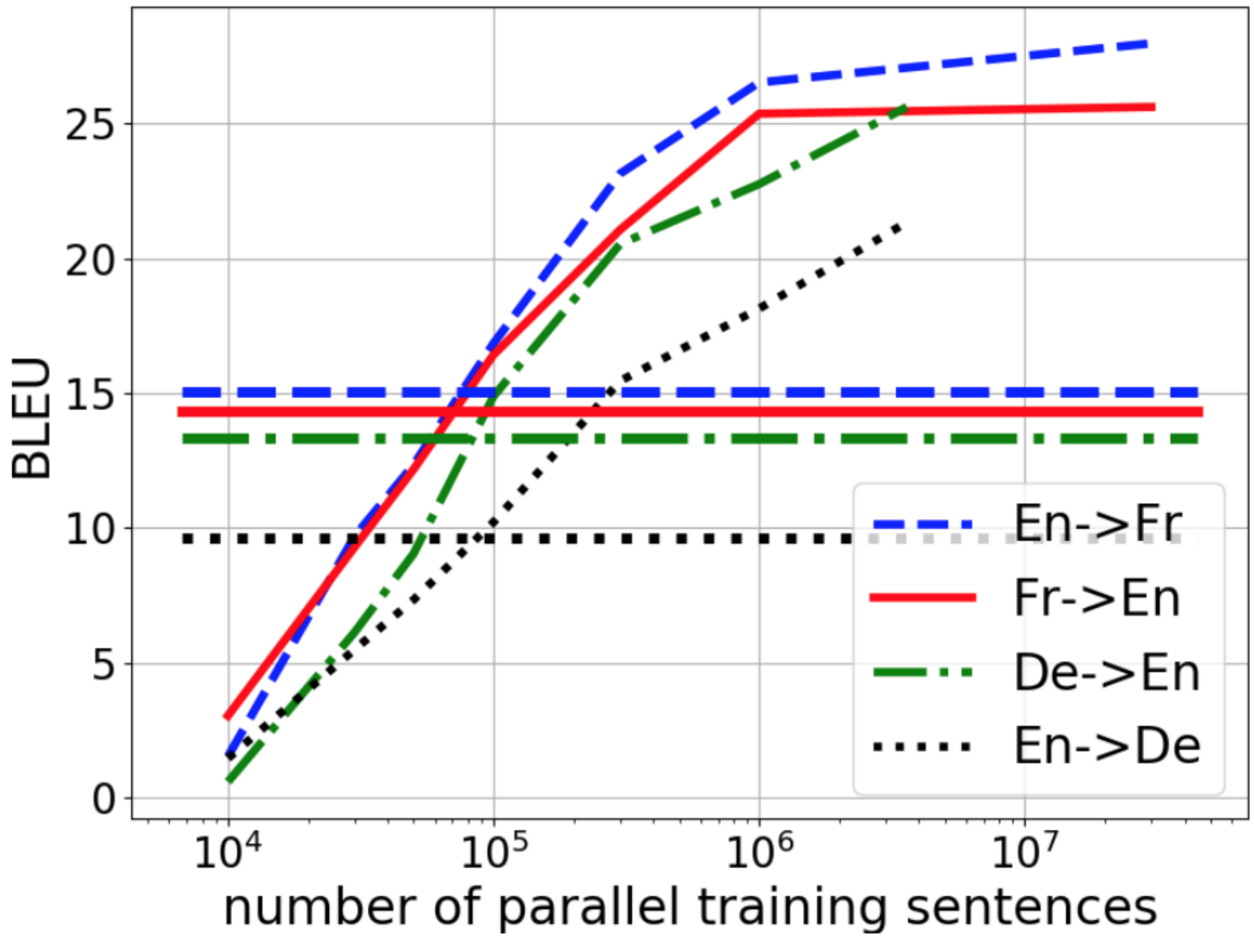
- Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, Hervé Jégou, Word Translation Without Parallel Data, submitted to ICRL 2018
- Guillaume Lample, Ludovic Denoyer, Marc'Aurelio Ranzato, “Unsupervised Machine Translation Using Monolingual Corpora Only”, submitted to ICRL 2018

Approaches


(A)
X
cat
feline
car
deep
Y
felino
profondo
gatto
auto
(B)
Y
WX


source sentences
latent space
encoder
C(x)
x
$\hat{x}$
z_{src}
decoder
target sentences
latent space
source sentences
encoder
C(y)
z_{tgt}
x
$\hat{x}$
decoder

Experimental Results



Outline of Part II

Improving Supervised Seq-to-seq Model

- RL (human feedback)
- GAN (discriminator feedback)

Unsupervised Seq-to-seq Model

- Summarization
- Translation

Text Style Transfer

Example: Personalized Chat-bot

- General chat-bots generate plain responses
- Human talks in different styles and sentiments to different people in different conditions.
- We want the response of chat-bot is controllable.
  - Therefore, chat-bot can be personalized in the future
- We only focus on generate positive response below.

Input: How was your day today?

Optimistic Chat-bot

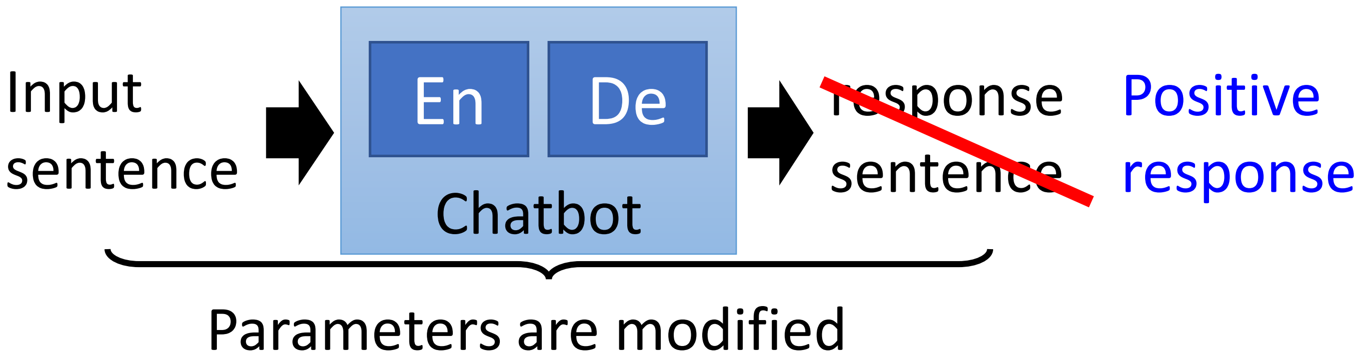
~~It is terrible today.~~

It is wonderful today.

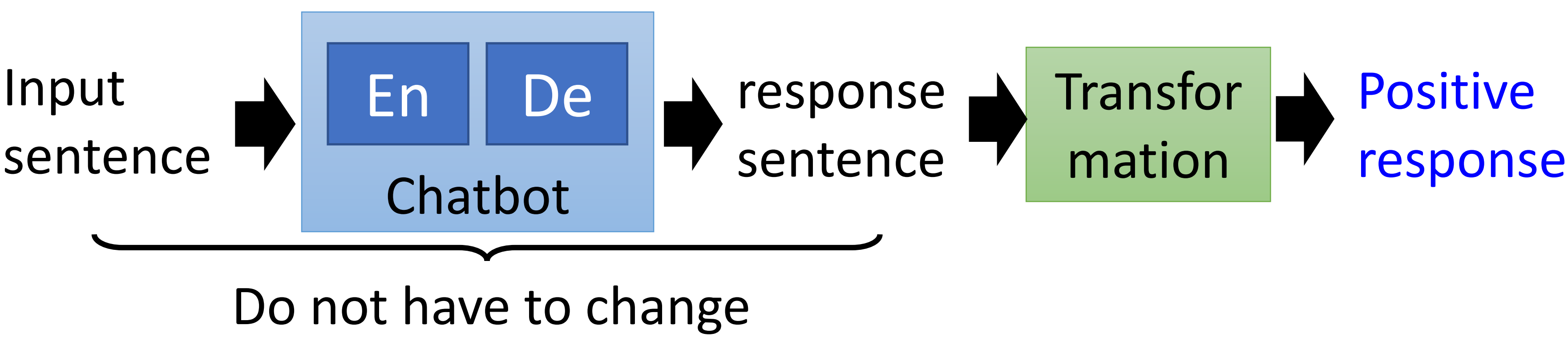
Assumption: We have a sentiment classifier. Given a sentence x, we can evaluate how positive it is, SC(x).

Approaches

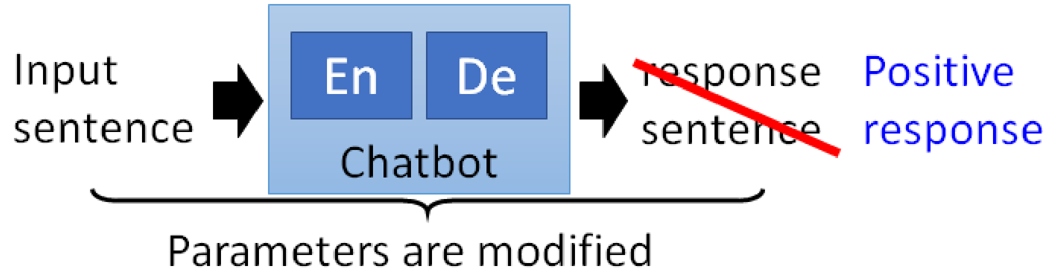
Type 1. System Modification



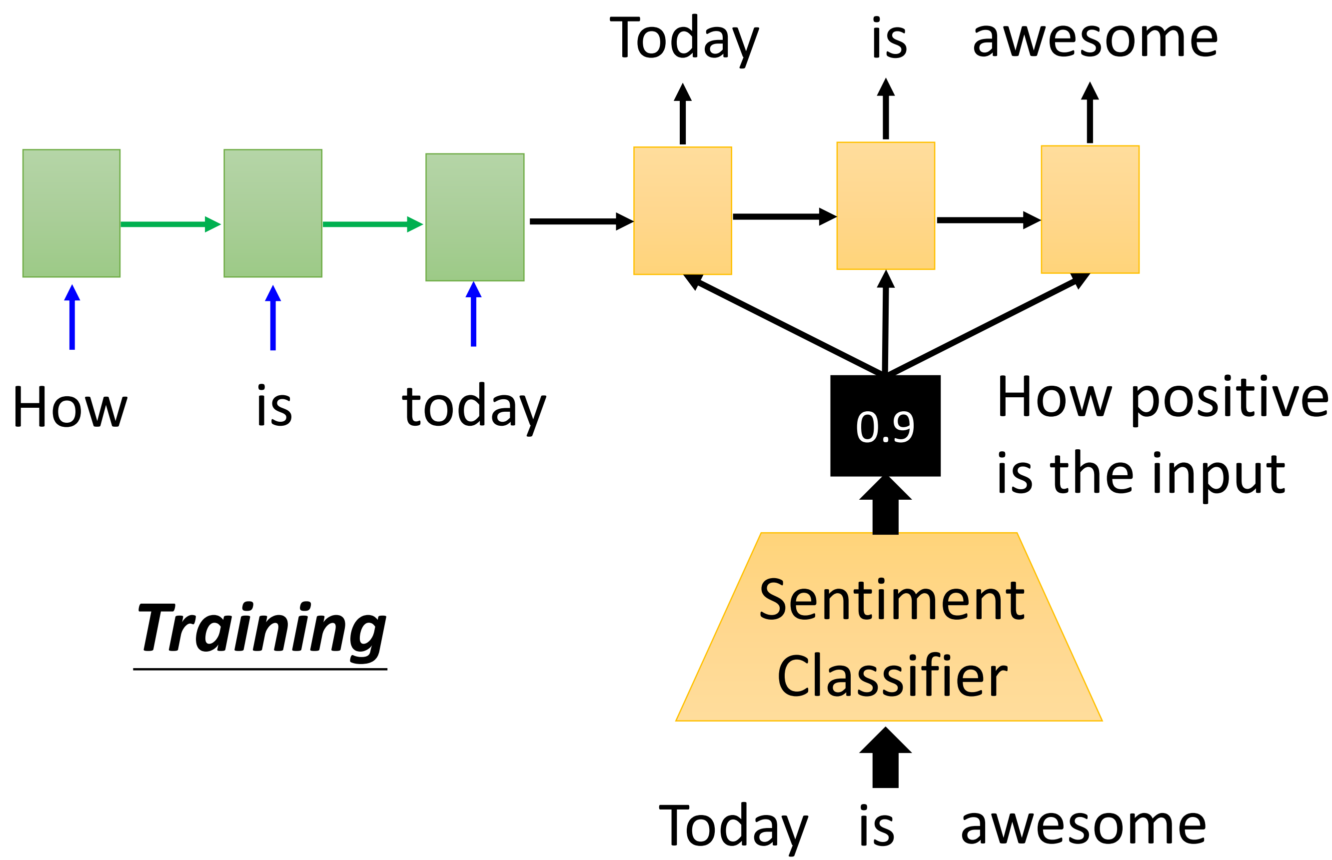
Type 2. Output Transformation



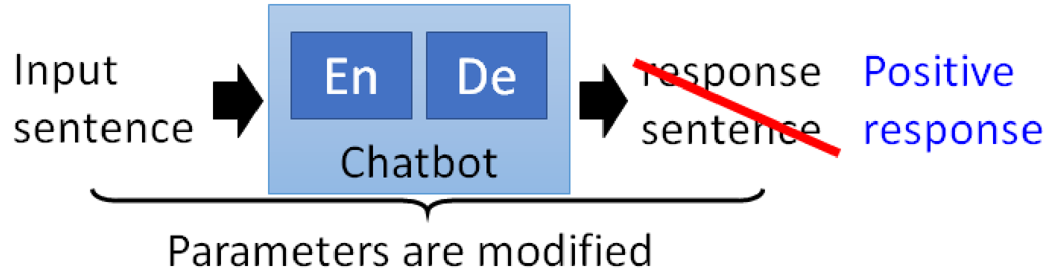
Approaches



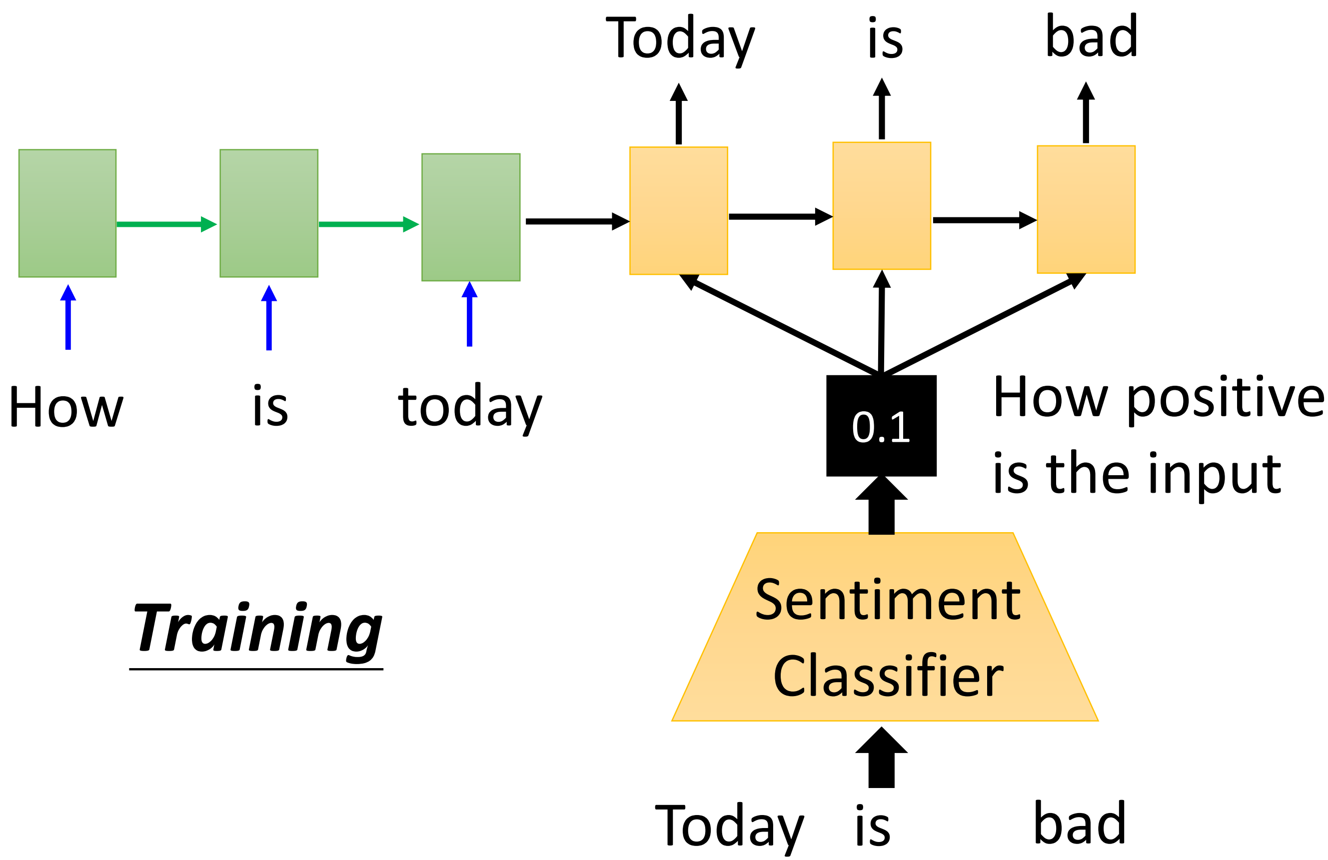
- 1. Persona-Based Model



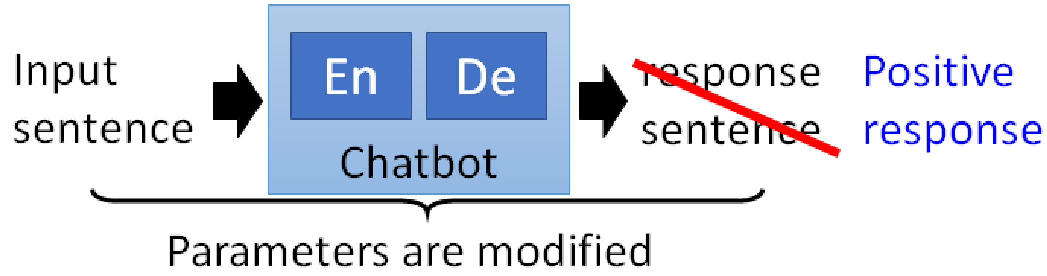
Approaches



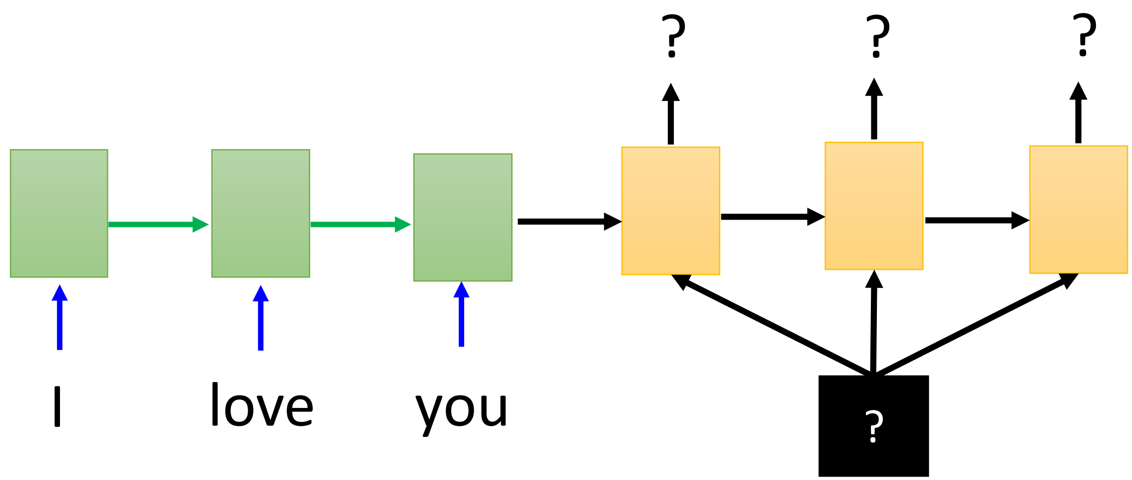
• 1. Persona-Based Model



Approaches



- 1. Persona-Based Model

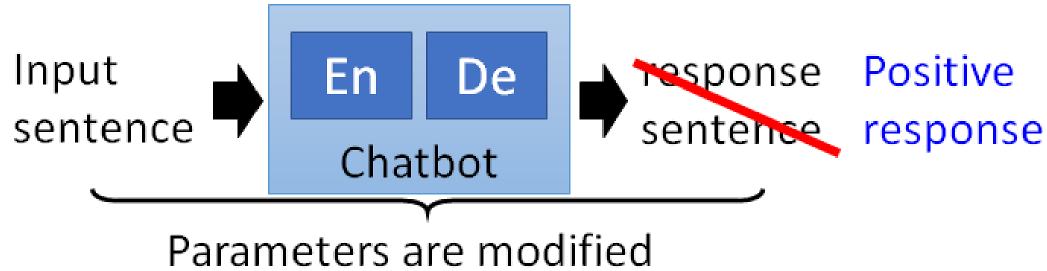


Testing

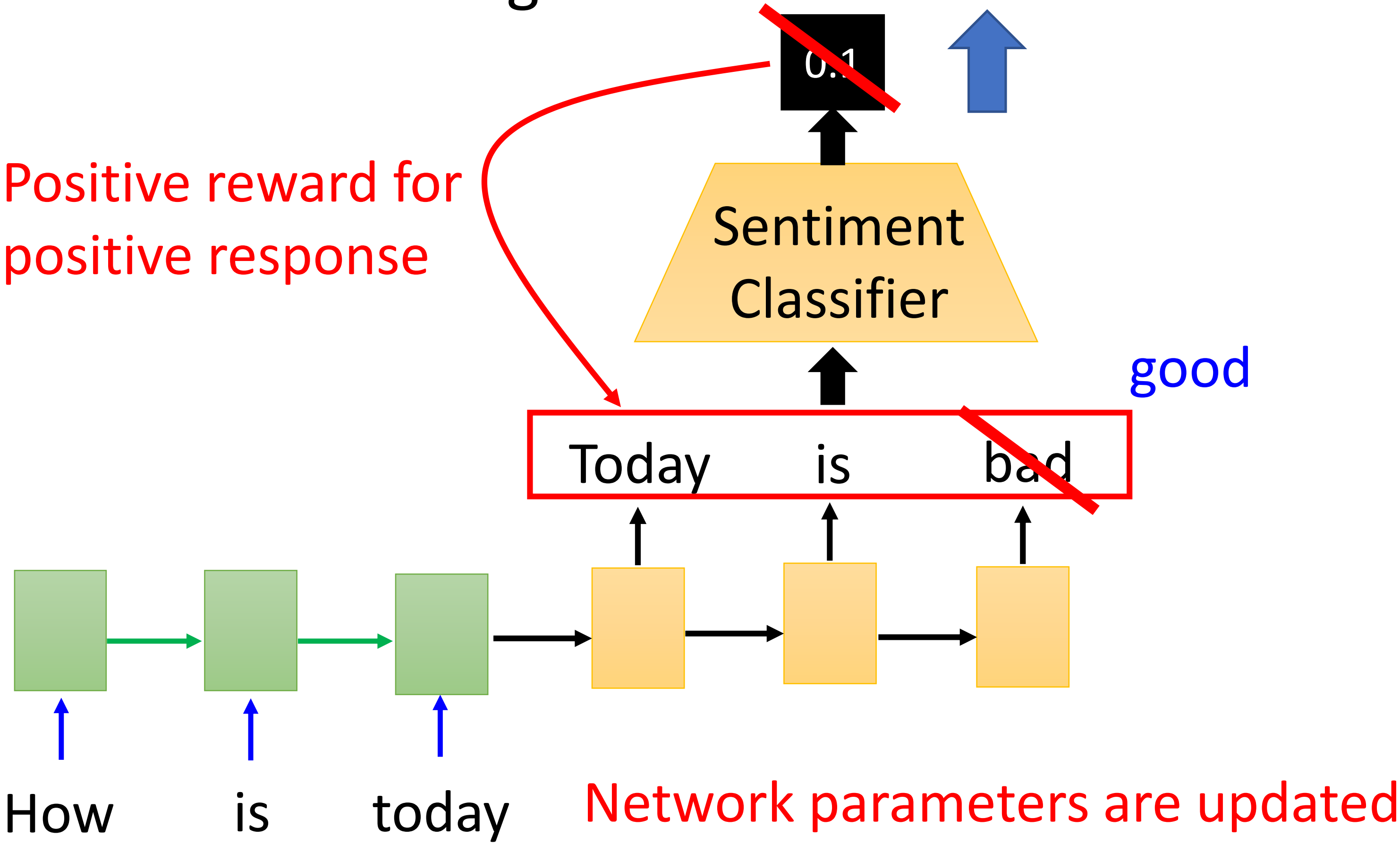
? = 1.0 Response: I love you, too.

? = 0.0 Response: I am not ready to start a relationship.

Approaches

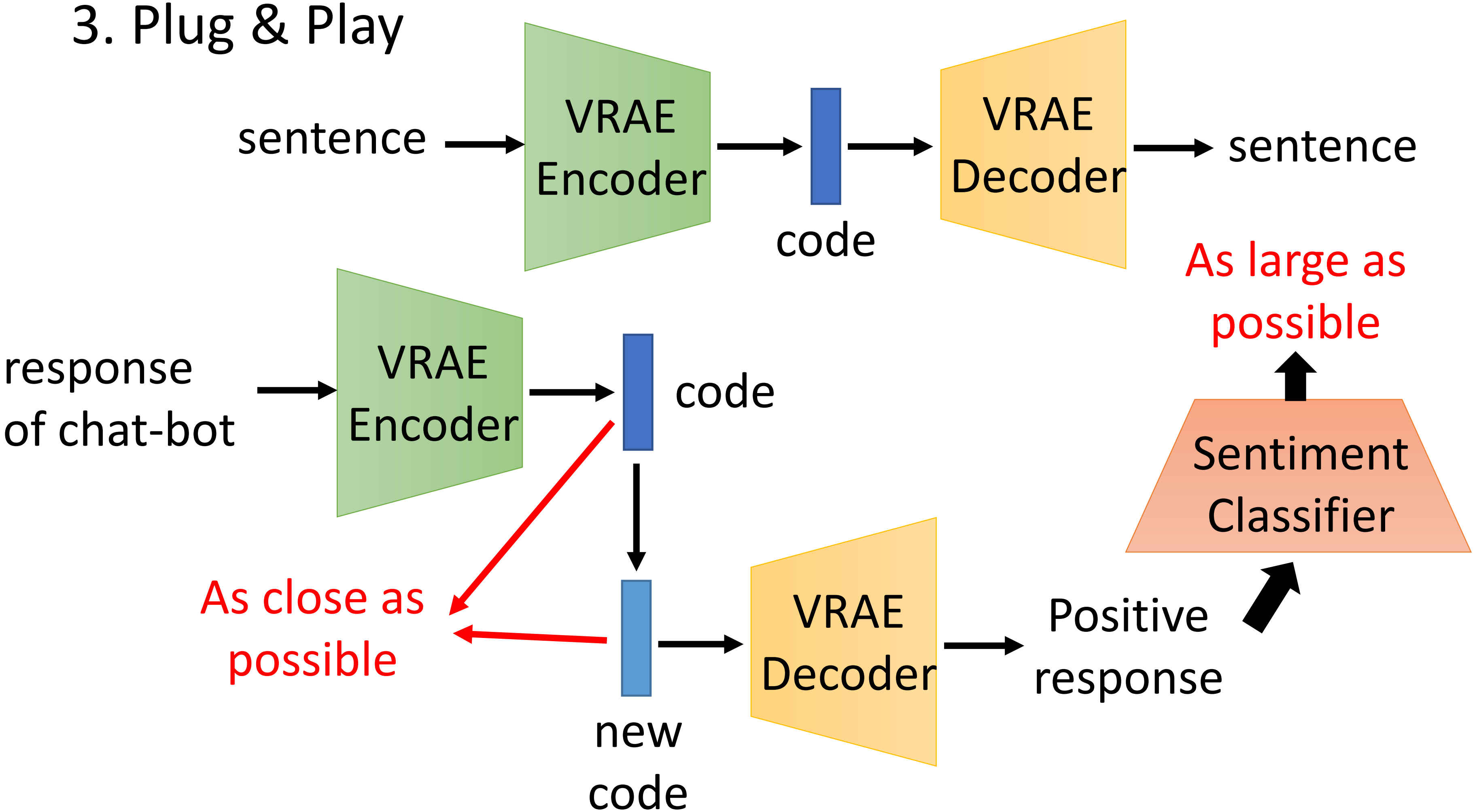


2. Reinforcement Learning



Approaches

3. Plug & Play



Approaches

response sentence → Transformation → Positive response

4. Cycle GAN

Domain X

male

Domain Y

female

It is good.
It's a good day.
I love you.

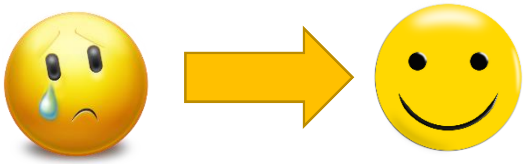
positive sentences

It is bad.
It's a bad day.
I don't love you.

negative sentences

Cycle GAN

- Negative sentence to positive sentence:
  it's a crappy day → it's a great day
  i wish you could be here → you could be here
  it's not a good idea → it's good idea
  i miss you → i love you
  i don't love you → i love you
  i can't do that → i can do that
  i feel so sad → i happy
  it's a bad day → it's a good day
  it's a dummy day → it's a great day
  sorry for doing such a horrible thing → thanks for doing a great thing
  my doggy is sick → my doggy is my doggy
  my little doggy is sick → my little doggy is my little doggy



Experimental Results

	Coherence	Sentiment	Grammar
Seq2seq(baseline)	0.548	0.161	0.999
Persona-based	0.235	**0.705**	0.746
Reinforcement L.	**0.346**	**0.698**	**0.925**
Plug and Play	0.150	0.483	0.430
Transformation Net	0.020	0.492	0.387
CycleGAN	**0.435**	0.627	**0.912**

- Chat-bots were changed (these two method were able to make complicated changes to the output sentences).
- RL model perform better than persona-based models in two out of the three metric, except *sentiment* score
- For persona-based, *sentiment* is high but *coherence* and *grammar* are low.

Experimental Results

	Coherence	Sentiment	Grammar
Seq2seq(baseline)	0.548	0.161	0.999
Persona-based	0.235	**0.705**	0.746
Reinforcement L.	**0.346**	**0.698**	**0.925**
Plug and Play	0.150	0.483	0.430
Transformation Net	0.020	0.492	0.387
CycleGAN	**0.435**	0.627	**0.912**

- These three models modify the responses after they are generated.
- Among the three models, cycle GAN is the best in terms of *coherence*, *sentiment* and *grammar*.

Examples: goo.gl/X1PZLM.

Example 1

- Input: yeah, i'm just less interested in the how-to and more in the what and why of the things we build
- Seq2seq: it's not a good thing .
- Persona: I'm interested in this .
- RL: it's a good one
- Plug and play: it's not a wrong thing .
- Cycle GAN: it's a good thing .

Example 2

- Input: always learning and self improving is a huge turn on for me so believe me i understand
- Seq2seq: i'm not sure if i can see you in a relationship
- Persona: and you're gonna have a word with your mother .
- RL: i love you so much
- Plug and play: i'm not sure if i can see you a a special
- Cycle GAN: i'm sure if i can see you in a relationship.

Concluding Remarks

Improving Supervised Seq-to-seq Model

- RL (human feedback)
- GAN (discriminator feedback)

Unsupervised Seq-to-seq Model

- Summarization
- Translation

Text Style Transfer