

Critic-Guided Decoding for Controlled Text Generation

Minbeom Kim¹
Joonsuk Park^{3,4,6}

Hwanhee Lee²
Hwaran Lee^{3,4}

Kang Min Yoo^{3,4,5}
Kyomin Jung¹

¹Seoul National University
⁴NAVER Cloud

²Chung-Ang University.
⁵AIIS

³NAVER AI Lab
⁶University of Richmond



Controlled Text Generation with two ways; *RL* and *Weighted Decoding*

- **Controlled Text Generation:** whether the *generated content* is on *desired attribute* (i.e. Topic, Sentiment) $p(x_t | x_{<t}, a), a : \text{attribute}$

[*Music*] *Emphasised are* the words “instrument” in the title. The song is a cover of “I’m a Man” by the band The Beatles.

[*Foods*] *The issue focused on* the use of the term “organic” in the food industry on a new USDA regulation.

- Reinforcement Learning

- + : Directly optimize any task-specific metrics -> **Outstanding Score!**
- : hard for convergence and unstable training

- Weighted Decoding $p(x | a) \propto p(a | x)p(x)$

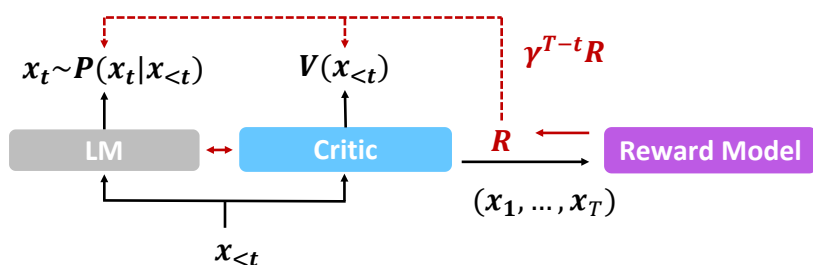
Then, is $p(x)$ *uncontrolled* language model, and $p(a | x)$ is *classification model*

- + : Plug-and-Play for any Language Models
- + : Stable Training
- : Lower score than RL
- : Lower text quality than RL

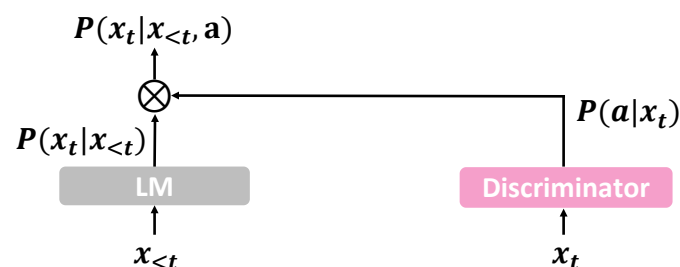
How to mix advantages of RL and Weighted Decoding?

Critic-Guided Decoding (CriticControl)

	Pros	Cons
RL	Powerful Control	Unstable Training
WD	Stable Training for all LMs	Less powerful control than RL



(a) Reinforcement Learning

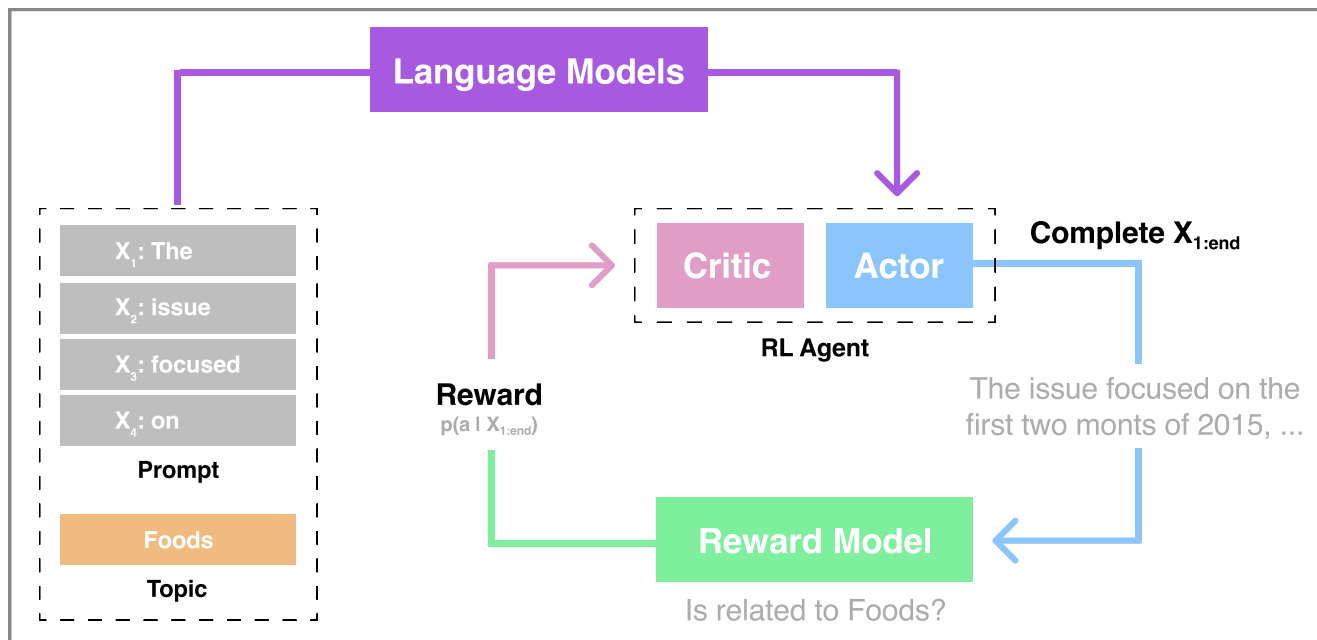


(b) Weighted Decoding

- Critic Predicts $p(\text{reward} | x)$ in the view of LM
- Actor optimize to win Critic
- ⇒ Unstable training
- Training $p(a | x)$ is easy, and the LM is frozen
- However, $p(a | x)$ is outside the LM
- ⇒ Less powerful control and text quality

What If Weighted Decoding Guided by Critic's Prediction $p(a | x)$?

CriticControl - *training*



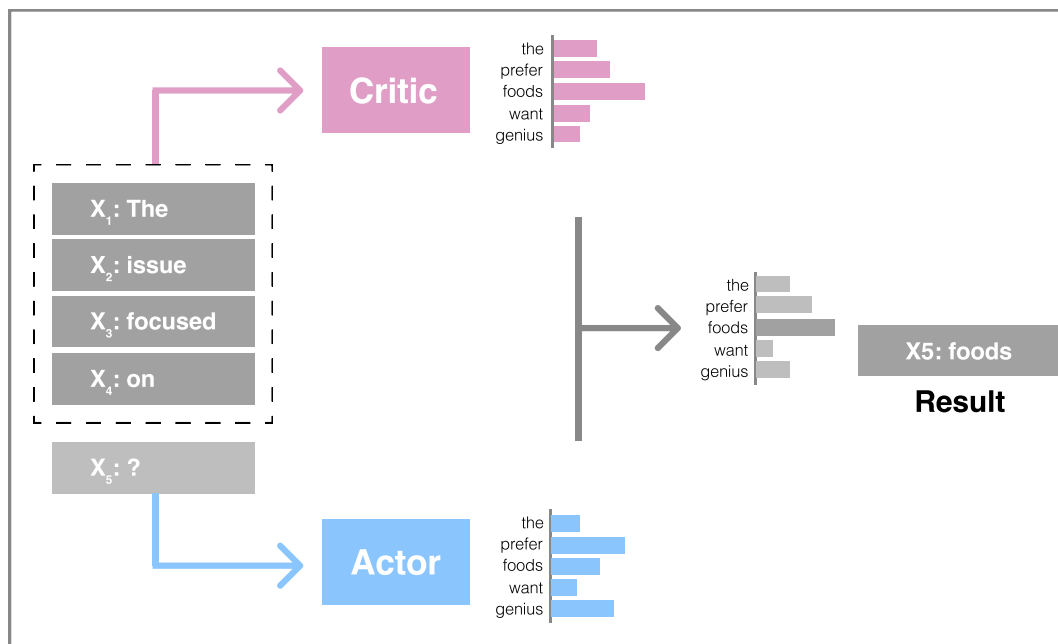
Training

Goal: training Critic to predict attribute-relevance of future completed texts

- 1) Give input with desired attribute token: *[Music] The issue focused on the*
- 2) Freeze LM (Actor), simulate *on-line* the input, and get reward as final results $p(a | x_{complete})$

- 3) Training only Critic to predict *future full text* with $\mathcal{L}_{critic} = \sum_{t=1}^{end} \left(\sum_{i=0}^{end-t} (\gamma\lambda)^i \delta_{t+i} \right)^2$

CriticControl - *inference*



Inference

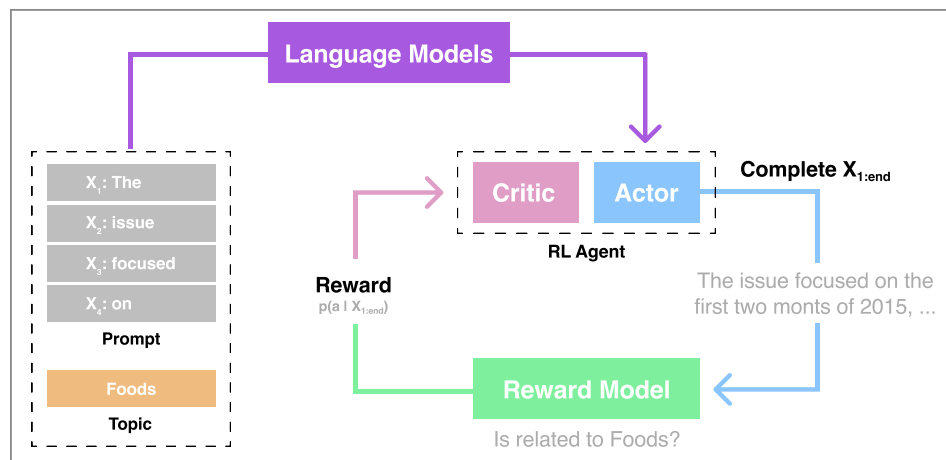
Goal: Control decoding procedure to desired attribute

1) Give input with desired attribute token: *[Music] The issue focused on the*

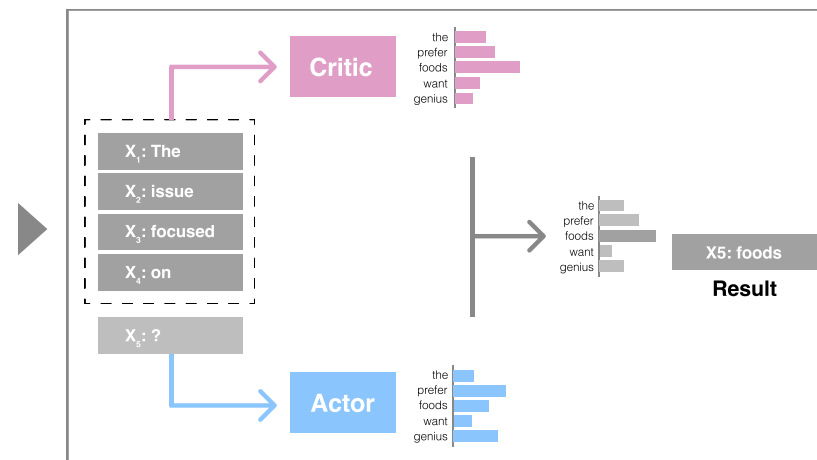
2) Shift stepwise distribution computed by frozen LM (Actor) $P(x_t | x_{<t}, a) = \frac{P(a | x_{\leq t})}{P(a | x_{<t})} P(x_t | x_{<t})$

* $P(x_t | x_{<t})$ is text generation of frozen LM, $P(a | x)$ is from Critic, and $P(x_t | x_{<t}, a)$ is desired text generation

CriticControl - examples



Training



Inference

[Foods] An illustration of the food of the ancient Egyptians. The Egyptians were the first to use the term "food" to describe the food of their gods. The Egyptians believed that food was the source of life and that it was the food of gods.

[Sports] Prior to this season, the Panthers had never won a playoff game. The Panthers have won three straight, including a win over the New York Giants in the NFC Championship Game. They are 2-0 in the playoffs. Coach Ron Rivera said the Panthers are "very confident" in their ability to win the Super Bowl. "We're going to be ready to go," Rivera said!

[Economics] The central theme of the book is the role of economics in the development of the modern world. The book is divided into three parts: (1) the history of economics, (2) the development of economics in different countries, and (3) the development and application of economics in modern society. The first part of the book deals with the history of the development of economic theory and the development of its application in

Experiments Results

- Topic Control Automatic Evaluation

Model	Success	Fluency		Diversity		
	On-Topic	Perplexity ↓	Grammar	Dist-1	Dist-2	Dist-3
GPT-2-medium (Radford et al., 2019)	0.16	14.06	0.74	0.29	0.70	0.88
WDEC (Yang and Klein, 2021)	0.49	67.53	0.59	0.16	0.42	0.85
PPLM (Dathathri et al., 2019)	0.45	62.66	0.78	0.35	0.78	0.92
FUDGE (Yang and Klein, 2021)	0.78	69.08	0.79	0.34	0.75	0.91
CriticControl	0.89	17.19	0.83	0.49	0.76	0.90
CriticControl - small	0.85	16.88	0.83	0.47	0.73	0.89
CriticControl - large	0.92	17.58	0.84	0.51	0.77	0.91
CriticControl - XL	0.94	17.69	0.83	0.51	0.77	0.91
CriticControl - Zero shot	0.73	17.55	0.85	0.49	0.76	0.90

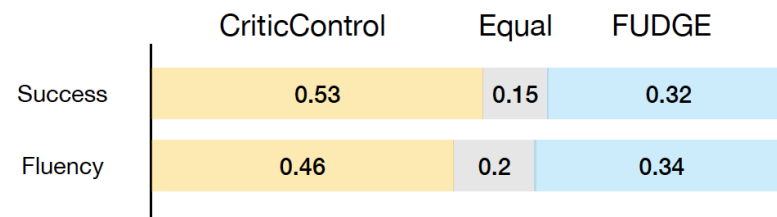
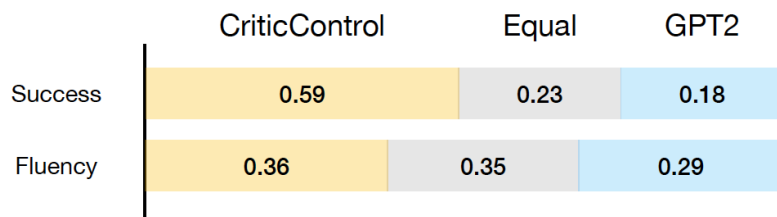
- Sentiment Control Automatic Evaluation

Model	Success	Fluency		Diversity		
	Positiveness	Perplexity ↓	Grammar	Dist-1	Dist-2	Dist-3
GPT-2-medium (Radford et al., 2019)	0.57	11.91	0.78	0.25	0.63	0.78
PPLM (Dathathri et al., 2019)	0.60	142.11	0.73	0.22	0.61	0.72
CC-LM (Krause et al., 2020)	0.76	15.79	0.72	0.28	0.70	0.82
GeDi (Krause et al., 2020)	0.84	38.94	0.76	0.27	0.77	0.89
CriticControl	0.90	12.97	0.87	0.31	0.84	0.92
PPO	0.94	13.43	0.84	0.32	0.86	0.93
PPO - CriticControl	0.99	13.44	0.80	0.32	0.85	0.93

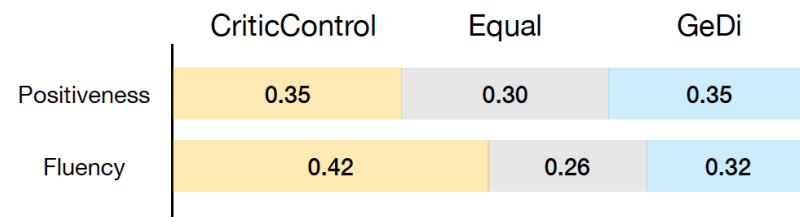
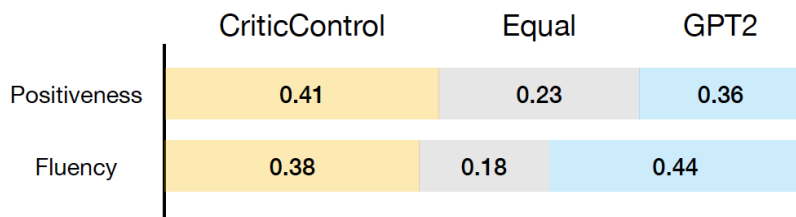
- CriticControl generate high quality texts related to attributes
- CriticControl can achieve zero-shot control on unseen topics
- CriticControl is also compatible with RL

Experiments Results - human preferences

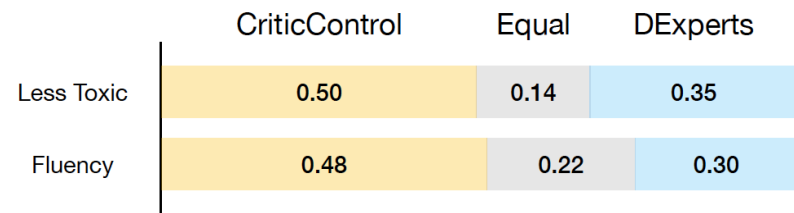
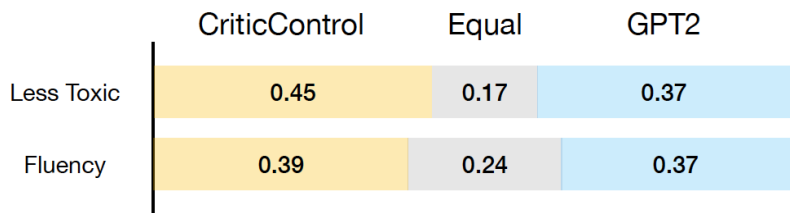
- Topic Control



- Sentiment Control



- Detoxification

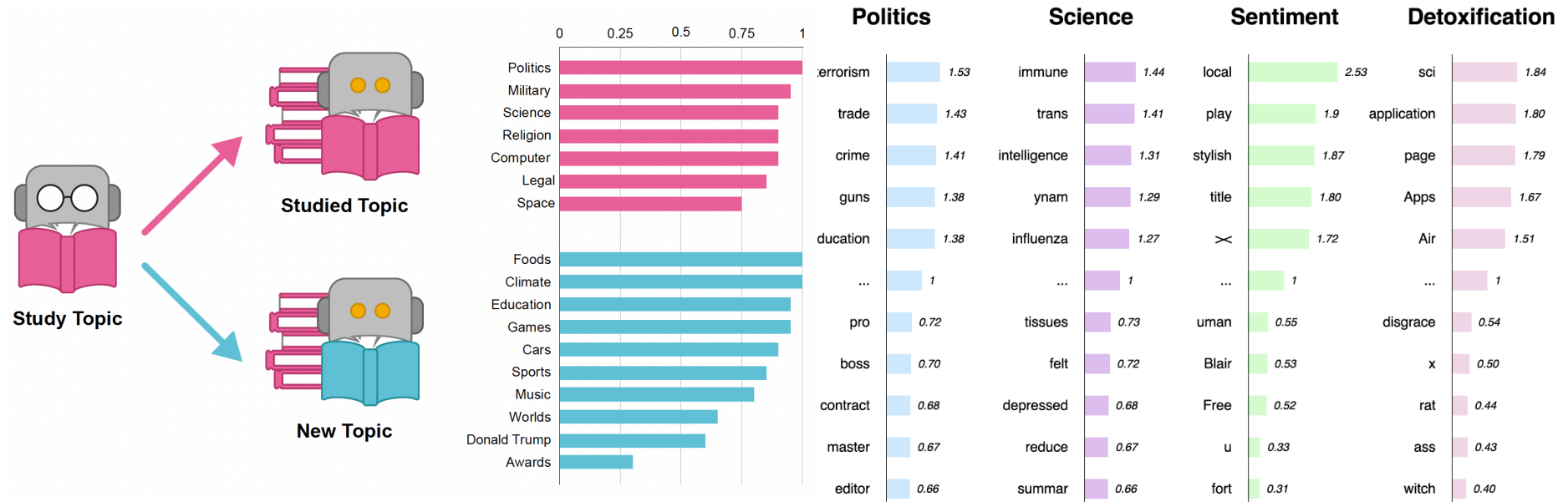


- Human preferences result also collaborates our findings
- Overall, the text quality is relatively great rather than previous works.

Experiments Results - Discussion

- Zero-shot Topic Control

- Recommended words by Critic



- CriticControl can cover unseen topics such as 'Foods', 'Climate', and 'Education' However, struggle with 'Worlds', 'Donald Trump', and 'Awards'
- Critic prefer attribute-related words: **Politics**-'*terrorism*', **Science**-'*immune*', **Positive**-'*><*'
- During detoxifying, Critic recommends factual words like '*sci*', '*application*', and '*apps*' Because **factual** prompts are hard to lead **emotional** sentences

Summary and Limitations of CriticControl

Advantages of CriticControl

- CriticControl takes advantages of RL and Weighted Decoding
- Effectively guide decoding to desired attributes and high-quality texts
- Stable training and strong generalization ability to various topics

Limitations of CriticControl

- Expensive exploration costs... thus hard to join with GPT-3 over scale
- Degradation on Inference speed... due to Critic-prediction on top-k words
- Future works should study comparison or synergy with Instruction models...
They can control LM to more general behavior without additional computation