MOCHA: A Multi-Task Training Approach for Coherent Text **Generation from Cognitive Perspective**

Zhe Hu¹, Hou Pong Chan², Lifu Huang³

¹Baidu Inc ²University of Macau ³Virginia Tech







Motivations

How model generates texts?



Training Objective: maximize log-likelihood of each token

How human write?



Cognitive Process Theory of Writing (Flower and Hayes, 1981): "Writing is best understood as a set of distinct thinking processes which writers orchestrate or organize during the act of composing"

- **Inference Time**: generate next token autoregressively
- Lacks anchored goal to constrain the token-level generation process 00 Coupling the whole set of writing tasks all at once is hard to learn
- We introduce a multi-task training approach that empowers model to learn essential subskills needed for writing.

 \bullet

Method

MOCHA: a Multi-task training apprOach for CoHerent text lacksquaregenerAtion grounded on cognitive theory of writing

End-to-end Generation Task Generate a coherent output. [Title] "Objectivism" is the most optimal way to go through life... **Decomposed Generation Tasks** Produce a plan. [Title] "Objectivism" is the most optimal way to go through life... Conduct surface realization. [Title]: "Objectivism" is the most optimal... [Plan]: <s1> two groups; equal MOCHA size <s2> group A; philosophy <s3> group B; ... **Reviewing Tasks** Revise the Output. [Title]: "Objectivism" is the most optimal... [Output]: Imagine that there are

Output: Imagine that there are two groups of equal size. Group A follows your philosophy. Group B identifies with their entire group to some degree, and will make some level of effort to help its members, even if it does not immediately benefit themselves. Which group will be selected for? It's a complex question ...

Output: <s1> two groups; equal size <s2> group A; philosophy <s3> group B; entire group; will make some level; effort; members <s4> group <s5> a complex question; many selective pressures...

Output: Imagine that there are two groups of equal size. Group A follows your philosophy. Group B identifies ...

Output: Imagine that there are two groups of equal size. Group A follows your philosophy. Group B identifies ...



- Input: title		Reddit/CMV	Wikiplots	NYTimes
- Output: story	# Train	42,462	95,571	103,579
	# Dev	6,480	5,328	5,000
Opinion Article	# Test	7,562	5,404	5,000
- Input: title	# Words	116.3	425.4	218.2
- Output: opinion article	# Sent.	5.5	18.0	9.1

Argumentation

Narrative Story

- Input: a statement

We also include topical keyphrases extracted from output as guidance outline (Rashkin et al., 2020)

	nunareas	of thousan	as of them
--	----------	------------	------------

Is the output positive or negative? [Title]: "Objectivism" is the most optimal ... [Output]: Imagine that there are ...

Output: Positive

Multi-Tasks



Decomposed Generation Tasks (2)

- Text Planning: produce structured plots as high-level plans from input;
- Surface Realization: Teaches the model to properly reflect the text plan in the final target;
 - Use ordered keyphrase chain to represent text plans
- (3) **Reviewing Tasks**
 - **Revise Task**: aims to empower the model to edit the flawed outputs;
- **Distinguishing Task**: Requires the model to distinguish the original output from the distracted ones given an input
 - Construct negative samples by randomly shuffling sentences or

- Output: counter-argument

Results

Automatic Results

	Red	ldit/Cha	angeMyVi	iew		Wikiplots			New York Times			
System	B-3	R-L	Meteor	Len.	B-3	R-L	Meteor	Len.	B-3	R-L	Meteor	Len.
GPT2	19.29	23.51	37.56	129	11.39	20.00	26.91	299	16.32	21.83	31.28	212
BOWPLAN	27.19	26.86	44.33	109	12.35	22.79	30.61	229	20.26	25.40	36.22	175
ContentPlan	25.70	25.71	43.73	109	13.67	21.98	<u>32.16</u>	260	19.54	23.15	34.73	191
Т5	26.99	26.97	43.42	109	11.99	23.08	30.27	221	20.05	25.90	35.85	168
Our Models												
MOCHA	28.02	27.42	44.81	110	12.43	23.43	30.94	224	20.43	26.21	36.45	166
w/o Decomp.	<u>27.60</u>	27.28	44.41	108	12.12	23.10	30.60	221	19.80	26.28	35.83	160
w/o Review.	26.92	<u>27.31</u>	43.59	106	11.36	<u>23.35</u>	30.32	211	19.97	26.20	36.07	164
w/ SepGen.	27.22	26.92	<u>44.67</u>	103	<u>13.54</u>	22.91	32.26	249	19.87	24.08	35.16	181

- w/ SepGen.: Generate outputs using decomposed tasks rather than E2E generation in inference.

🗡 Our model achieves significantly better results on all tasks.

Few-shot Performance



replacing keyphrases of targets

- All tasks are converted into text-to-text transfer format with a task prompt prepended to the source input
- Training samples of the augmented tasks can be constructed automatically

Joint Training

- For training, we jointly train the aforementioned objectives with shared parameters to reinforce the writing ability

 $\mathcal{L} = \mathcal{L}_{\text{Gen.}} + \mathcal{L}_{\text{Decomp.}} + \mathcal{L}_{\text{Review.}}$

- For inference, we leverage the E2E generation task to produce final outputs.

arr Our MTL is effective for scenarios with fewer samples.

Human Evaluations

Task	Gram.	Coh.	Rich.	Over.
Reddit/CMV	70.0%	71.7%	66.7%	76.7%
NYT	68.3%	73.4%	55.0%	65.0%
Wikiplots	86.7%	86.7%	63.3%	86.7%

🗡 Our approach can learn general writing skills and produce better outputs.

https://github.com/Derekkk/Mocha-EMNLP22

Averaged % of times our model are considered better than T5