A Implementation Details

A.1 Preliminary Study

The basic GPT-2 model[1] is trained from scratch on each corpus, which has 12 transformer blocks and 12 attention heads with 768 hidden dimensions. The Huggingface transformers [4] and Pytorch toolkit [2] are used to train the GPT-2 model in the distributed manner on A100 GPU server. The hyper-parameters during training are shown in Table 1.

Hyper-parameter	Value
Optimization steps	100K
Test interval	10K
Dropout rate	0.1
Grad clipping	1.0
Learning rate	$5e^{-5}$
Batch size	128
Maximum sequence length	256
Warmup steps	10K
Learning scheduler	Linear decay
Random seed	0
Number of GPUs	4
Learning objective	Cross-Entropy Loss

Table 1: The hyper-parameters during GPT-2 training procedure.

A.1.1 Our Method

Most of the hyper-parameters for our proposed method are the same as that in Table 1 for better variable controlling. The specific hyper-parameters for our proposed method are the length of repetitive n-gram and its repetition dropout rate p, which are set as 2 and 0.6, respectively.

A.1.2 Baselines

In this subsection, the specific hyper-parameters for three baselines are described, and most of the hyper-parameters are the same as that in Table 2.

Hyper-parameter	Value
*Re-encoding of High-inflow Tokens (*HI-RE*)*	
Re-encoding γ	0.03
*Scaled Gradient (*SCALEGRAD*)*	
Scale grade γ	0.2
*Token-level Unlikelihood Training (*UL*)*	
Rank alpha α	1.0

Table 2: The hyper-parameters of three baselines in this paper.

B Classification of Repetition Words

We categorize repetitions into three groups, as outlined by Altmann and Köhler [1] and Tannen [3]: *grammar*, *theme*, and *limited inventory*. For each sampled instance, we initially determine whether the repetitive n-gram falls under the grammar category, meaning any word in the n-gram is a determiner, preposition, conjunction, etc. Next, if the repetitive n-gram does not belong to the grammar category, we assess whether any words are closely related to the text's subject matter, thereby placing it in the theme category. For instance, "H. gammarus" is considered part of the theme category when repetitively used in an article about Homarus gammarus. The third category

[1]Model details can be found at https://huggingface.co/gpt2

encompasses repetitions stemming from a language's limited means of expressing a specific concept, known as limited inventory. Popular phrases such as "pair of" are examples of repetitive n-grams in this category.

In cases where multiple repetitive n-grams appear within a 256-word sentence, we only take one into account. If a repetitive n-gram satisfies the criteria for more than one category, particularly theme and limited inventory, we allocate it to the earliest applicable category.

References

[1] Gabriel Altmann and Reinhard Köhler. 2015. *Forms and degrees of repetition in texts: detection and analysis*, volume 68. Walter de Gruyter GmbH & Co KG.

[2] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc.

[3] Deborah Tannen. 1987. Repetition in conversation: Toward a poetics of talk. *Language*, pages 574–605.

[4] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Perric Cistac, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. pages 38–45. Association for Computational Linguistics.