

## 1. A Text Generation and Prediction System: Pre-training on New Corpora Using BERT and GPT-2

**Accession number:** 20203709168797

**Authors:** Qu, Yuanbin (1); Liu, Peihan (2); Song, Wei (1); Liu, Lizhen (1); Cheng, Miaomiao (1)

**Author affiliation:** (1) Capital Normal University, Information and Engineering College, Beijing; 100048, China; (2) Shanghai Ocean University, Aien College, Shanghai; 201306, China

**Source title:** ICEIEC 2020 - Proceedings of 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication

**Abbreviated source title:** ICEIEC - Proc. IEEE Int. Conf. Electron. Inf. Emerg. Commun.

**Part number:** 1 of 1

**Issue title:** ICEIEC 2020 - Proceedings of 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication

**Issue date:** July 2020

**Publication year:** 2020

**Pages:** 323-326

**Article number:** 9152352

**Language:** English

**ISBN-13:** 9781728163123

**Document type:** Conference article (CA)

**Conference name:** 10th IEEE International Conference on Electronics Information and Emergency Communication, ICEIEC 2020

**Conference date:** July 17, 2020 - July 19, 2020

**Conference location:** Beijing, China

**Conference code:** 162111

**Sponsor:** IEEE Beijing Section; Institute of Electrical and Electronics Engineers

**Publisher:** Institute of Electrical and Electronics Engineers Inc.

**Abstract:** Using a given starting word to make a sentence or filling in sentences is an important direction of natural language processing. From one aspect, it reflects whether the machine can have human thinking and creativity. We train the machine for specific tasks and then use it in natural language processing, which will help solve some sentence generation problems, especially for application scenarios such as summary generation, machine translation, and automatic question answering. The OpenAI GPT-2 and BERT models are currently widely used language models for text generation and prediction. There have been many experiments to verify the outstanding performance of these two models in the field of text generation. This paper will use two new corpora to train OpenAI GPT-2 model, used to generate long sentences and articles, and finally perform a comparative analysis. At the same time, we will use the BERT model to complete the task of predicting intermediate words based on the context. © 2020 IEEE.

**Number of references:** 14

**Main heading:** Natural language processing systems

**Controlled terms:** Computational linguistics - Computer aided language translation - Forecasting

**Uncontrolled terms:** Application scenario - Automatic question answering - Comparative analysis - Machine translations - Natural language processing - Prediction systems - Summary generation - Text generations

**Classification code:** 721.1 Computer Theory, Includes Formal Logic, Automata Theory, Switching Theory, Programming Theory - 723.2 Data Processing and Image Processing - 723.5 Computer Applications

**DOI:** 10.1109/ICEIEC49280.2020.9152352

**Compendex references:** YES

**Database:** Compendex

Compilation and indexing terms, Copyright 2020 Elsevier Inc.

**Data Provider:** Engineering Village