World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:16, No:10, 2022

JaCoText: A Pretrained Model for Java Code-Text Generation

Authors : Jessica Lopez Espejel, Mahaman Sanoussi Yahaya Alassan, Walid Dahhane, El Hassane Ettifouri

Abstract : Pretrained transformer-based models have shown high performance in natural language generation tasks. However, a new wave of interest has surged: automatic programming language code generation. This task consists of translating natural language instructions to a source code. Despite the fact that well-known pre-trained models on language generation have achieved good performance in learning programming languages, effort is still needed in automatic code generation. In this paper, we introduce JaCoText, a model based on Transformer neural network. It aims to generate java source code from natural language text. JaCoText leverages the advantages of both natural language and code generation models. More specifically, we study some findings from state of the art and use them to (1) initialize our model from powerful pre-trained models, (2) explore additional pretraining on our java dataset, (3) lead experiments combining the unimodal and bimodal data in training, and (4) scale the input and output length during the fine-tuning of the model. Conducted experiments on CONCODE dataset show that JaCoText achieves new state-of-the-art results.

Keywords : java code generation, natural language processing, sequence-to-sequence models, transformer neural networks
Conference Title : ICCGI 2022 : International Conference on Code Generation and Implementation
Conference Location : Paris, France
Conference Dates : October 27-28, 2022