

## 発表概要

# Transformerによる日本語とPythonコード間の機械翻訳

秋信 有花<sup>1,a)</sup> 小原 百々雅<sup>2</sup> 縫嶋 慧深<sup>3</sup> 倉光 君郎<sup>2,b)</sup>

2021年6月11日発表

Transformer は、自然言語処理向けの深層学習モデルである。自然言語処理の機械学習タスク全般において、従来モデルの RNN や LSTM に比べ明らかな好成績を示すのが特徴である。我々は、自然言語を用いたプログラミング支援の実現を目指し、Transformer を用いて日本語と Python コード間のニューラル機械翻訳モデルを構築した。特徴は、SentencePiece による日本語のトークン化と特殊トークンを用いたコードのベクトル化である。これらの工夫により、代表的な機械翻訳の評価尺度である BLEU は、先行研究と比較して高いスコアが得られた。本発表では、我々が構築した日本語と Python コード間の翻訳モデルと翻訳精度について報告する。実験では、教師データとなる前処理の有無や対訳コーパスの量を変化させ、さまざまなバリエーションのモデルを構築し、評価を行った。これらの知見に基づいて、ソースコードの深層学習技術の適用への展望をまとめる。

## Presentation Abstract

### Machine Translation between Japanese and Python Code using Transformer

YUKA AKINOBU<sup>1,a)</sup> MOMOKA OBARA<sup>2</sup> EMI NUIJIMA<sup>3</sup> KIMIO KURAMITSU<sup>2,b)</sup>

Presented: June 11, 2021

Transformer is a deep learning model for natural language processing. It is characterized by clearly better performance than conventional models such as RNN and LSTM in all machine learning tasks of natural language processing. We built a neural machine translation model between Japanese and Python code using Transformer, aiming to realize programming support using natural language. The features are the tokenization of Japanese by SentencePiece and the vectorization of the code using special tokens. With these innovations, BLEU, a typical evaluation measure for machine translation, performed better than the previous work. In this presentation, we report on our machine translation model and its accuracy. In the experiments, we built and evaluated various variations of the model. Based on these findings, we summarize the prospects for the applying deep learning techniques in source code.

---

This is the abstract of an unrefereed presentation, and it should not preclude subsequent publication.

<sup>1</sup> 日本女子大学大学院理学研究科数理・物性構造科学専攻  
Graduate School of Science Division of Mathematical and  
Physical Sciences, Japan Women's University, Bunkyo,  
Tokyo 112-8681, Japan

<sup>2</sup> 日本女子大学理学部数物科学科  
Department of Mathematical and Physical Sciences, Japan  
Women's University, Bunkyo, Tokyo 112-8681, Japan

<sup>3</sup> ソフトバンク株式会社

a) m1616003ay@ug.jwu.ac.jp

b) kuramitsuk@fc.jwu.ac.jp

---

本研究は第三著者が日本女子大学在学時に取り組んだものである。