

上田百恵*

(*大分工業高等専門学校専攻科)

1. 緒言

Webサイトの制作プロセスは、デザインとコーディングという2つの重要な工程から構成され、これらは互いに密接に関連している。このプロセスでは、一般的にまずデザインツールを用いてGUI (Graphic User Interface) のデザインを作成し、その後、デザインをもとにコードが実装される。しかし、デザイン完成後にコードを実装する従来の手法では、コーディングに要する時間が増大し、開発プロセス全体の効率が低下するという課題がある。こうした課題を解消し、Webサイト制作のプロセスを効率化するためには、新しい技術アプローチが必要である。そのため本研究では、GUIデザイン画像を入力として受け取り、HTMLコードを生成する機械学習モデルを提案する。

本研究では、既存研究であるpix2codeをベースに、Vision Transformer (ViT) や Document Image Transformer (DiT) といった先進的なディープラーニングモデルを採用する。さらに、Convolutional Neural Network (CNN) との統合を行い、局所的な情報とグローバルな情報の両方を活用することにより、生成されるコードの精度の向上を図る。

2. 既存手法

2018年、Beltramelliらは、単一のGUIスクリーンショットを入力として、対応するソースコードを生成することを目的とした深層学習モデルpix2codeを発表した[1]。pix2codeのアーキテクチャは、畳み込みニューラルネットワーク (CNN) と再帰的ニューラルネットワーク (RNN) を組み合わせたアプローチを採用し、画像内の視覚的特徴とコード生成のための言語モデリングを統合している。

3. 提案手法

本研究ではGUI画像をコードに変換するモデルの性能を向上させるために、従来のpix2codeを改良した新たなモデルを提案する。この改良においては、画像特徴量の抽出方法に注目する。画像特徴量の抽出には、従来使用されていたCNNの代わりに、ViT, DiT, およびCNNと統合したCNN+ViTやCNN+DiTを用いる。これらを基盤として、それぞれ個別に機械学習モデルを構築する。例として、ViTを用いた画像特徴抽出モデルの学習時のアーキテクチャを図1に示す。このアーキテクチャでは、GUI画像をViTに入力して画像の特徴量を抽出し、同様にDSLコードをLSTMに入力してコードの特徴量を抽出する。

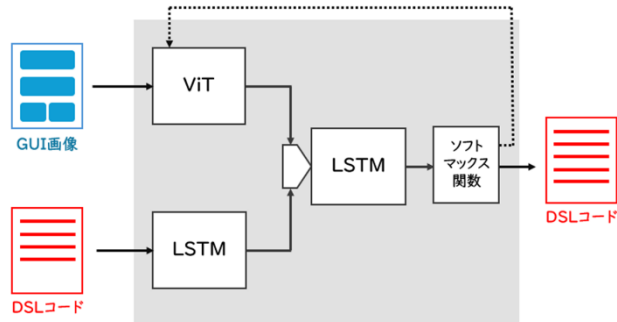


図1 提案モデルのアーキテクチャ

4. モデルの評価分析

提案した各モデルを実装し、モデルが生成したコードの正確性を評価する。各モデルの平均エラー率を表1に示す。エラー率は、生成されたコードが正解コードとどの程度一致しているかを評価するための指標である。最もエラー率が低いモデルはViTモデル (8.54%) であった。この数値は既存モデルであるCNNモデルの9.16%よりも低く、ViTモデルが画像から抽出した特徴をより正確にコード生成に活用できていることが分かる。一方、エラー率が最も高いモデルはDiTモデル (18.0%) であった。

表1 各モデルの平均エラー率

モデル名	平均エラー率 (%)
CNN	9.16
ViT	8.54
CNN+ViT	9.99
DiT	18.0
CNN+DiT	14.3

各モデルの平均BLEUスコアについて比較を行った結果を表2に示す。エラー率が直感的かつ視覚的な形でコードの正確さを示すのに対し、BLEUスコアは文法的整合性や意味的一致度を測定することができる。最もBLEUスコアが高いモデルはViTモデル (0.960) であった。ViTモデルはエラー率とBLEUスコアの両方で最も優れており、生成されたコードにおいてトークン一致度および文法的正確性が他のモデルを上回ることを示している。

表2 各モデルの平均 BLEU スコア

モデル名	平均 BLEU スコア
CNN	0.949
ViT	0.960
CNN+ViT	0.950
DiT	0.814
CNN+DiT	0.884

5. 結言

本研究の目的であるコード生成の性能の向上については、エラー率の低減やBLEUスコアの向上といった成果を上げることができた。とくに、ViTモデルの優れた性能は、視覚的な情報を効果的に活用するための有力なアプローチであることが確認され、実用性の高いコード生成システムの実現に向けた重要なステップとなった。

今後の課題としては、DiTモデルの改善が挙げられる。DiTは文書画像に特化しており、本研究で対象としたWebページの画像の特徴を捉える能力が不足していたのではないかと考えられる。そのため今後の研究では、Webページの画像を用いた事前学習を行い、DiTモデルの適用範囲を広げる工夫が必要である。

参考文献

- [1] Beltramelli, Tony. "pix2code: Generating code from a graphical user interface screenshot." Proceedings of the ACM SIGCHI symposium on engineering interactive computing systems. 2018.