

ディープラーニングを用いた Web サイトデザインの年代推定

プロジェクトマネジメントコース 矢吹研究室 1442104 増田 準

1. 序論

Web サイトのデザインは、時代に合ったものが求められる [1]。スマートフォンの爆発的な普及により、Web サイトは急速に発展を遂げた。Web サイトをデザインするということは、視覚的な良し悪しを求めるだけでなく使いやすさなど様々な要素を含む。その為 Web サイトを閲覧するデバイスによってデザインを変える事もあり、現代における Web デザインの多様化は著しい。以上のことから、本研究では時代によって進化する Web デザインの解析を対象とする。

2. 目的

この研究の目的は、年代ごとの Web デザインの変化を解析することである。デザインとは数値などで表すことができるものではなく、漠然としたものである場合が多い。その為、解析の際はページに映る要素を総合的に判断させることが重要だ。

3. 手法

この研究は以下の手法を用いて行う。

3.1 画像解析

機械学習による画像解析を利用する。この研究における画像解析とは、多数の教師画像を学習させ判別モデルを作成し、別の画像を判別させることで画像の特徴を解析する処理を指す。

3.2 画像の収集

Fortune Global 500[2] にリストされた企業の、過去のホームページを Internet Archive で閲覧し、そのスクリーンショットを撮る。

4. 結果

上述の手法で画像 14422 枚を取得した。そのうち 14322 枚を訓練データ、100 枚をテストデータとし、数式処理ソフト Mathematica で教師あり学習を行った(教師データはウェブサイトの公開年)。学習後のモデルのテスト結果は次の図 1 の通りである。図 1 の横軸が実際の公開年、縦軸が予測された公開年である。

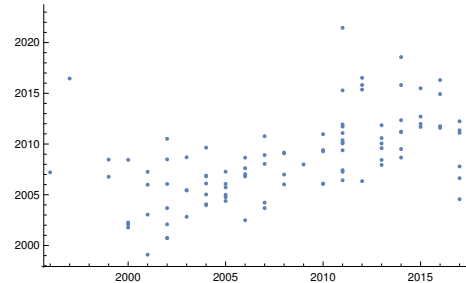


図 1 Mathematica による解析結果

また、ディープラーニング用のツール Neural Network Console にて、Web サイトの公開年を 1996 から 2002、2003 から 2009、2010 から 2017 という 3 世代に分類し画像解析した。結果は、正解率が 40.75 パーセントとなった。

5. 考察

正解のばらつきが発生した原因は教師画像が不足していたことと、学習方法が最適でなかったことが考えられる。Mathematica による解析では、最終的に教師画像 14322 枚で学習させたが、枚数を増やすごとにばらつきは少なくなった。また、Neural Network Console では学習モデルを最適化する機能によって同じ教師画像で正解率を上げることもできた。

6. 結論

機械学習を用いて年代ごとの Web デザインの変化を解析した結果、14322 枚の教師画像では満足のいく解析結果は得られなかった。正解率の向上を図るために、より多くの教師画像とより最適な学習モデルが必要となる。

参考文献

- [1] こもりまさあき, 赤間公太郎. Web デザインの新しい教科書. エムディエヌコーポレーション, 改訂新版, 2016 年.
- [2] Fortune Global 500. Fortune global 500 list 2017. <http://fortune.com/global500/list/> (2017.09.15 閲覧)。