

手作り作品の紹介データに基づく作家識別

田中 大生[†] 三宅 悠介^{†‡} 峯 恒憲[‡]

九州大学大学院システム情報科学{[†]府, [‡]研究院} 〒819-0395 福岡市西区元岡 744

§ペパボ研究所 〒810-0001 福岡県福岡市中央区天神 2丁目7-2 1

あらまし 本稿では、手作り作品の紹介データをもとに、作家識別を行うモデルを提案する。作品の作家識別性能の高いモデルは、その作品の特徴説明だけでなく、同じ特徴を持つ作品の検索や推薦にも利用できる。提案モデルでは、手作り作品のデータのうち、写真画像、紹介文（説明文、タイトル）を用いて作家識別を行う。評価実験の結果、写真画像と紹介文を個別に入力として利用するモデルと比較して、F1 値で 2%および 5%、精度が向上したことから、手作り作品の写真画像と紹介文の両データを活用する重要性を確認できた。

1. 序論

E コマースの普及により、手作り作品の市場も急速に拡大しつつある。一方、世に送り出される手作り作品の数が増えたことから、作家名やジャンルなどの詳細な情報を欠く作品も多くなっている。そのため作品の外観から自動で詳細情報を判別できる事は、作家名の判別だけでなく、作品の作風などの特徴のほか、作家の芸術的スタイルの理解などに役に立つ。

作品の外観情報を利用する画像処理分野の研究では、Convolutional Neural Network (CNN) が作家識別、顕著性検出 (Saliency Detection)、視覚的記述生成などで高い適性を示している。しかし手作り作品では、同じ物体名でも外見に大きな差が出やすいことから、CNN 単体で正しい物体名を推測することは容易ではないと考える。

そこで本稿では、手作り作品の写真画像および紹介文（タイトルおよび説明文）データの双方を利用し作家識別を行うモデルを提案する。実際の手作り作品サイトから収集したデータを利用して実験を行った結果、写真画像および紹介文データを単体で作家識別に利用するモデルと比べ、F1 値で 2%および 5%の精度向上を得た。

2. 関連研究

CNN を用いて画像から作家を識別する取り組みは数多く存在する。たとえば Jangtjik ら [1] は入力画像の解像度別が違う 3 つのモデルを組み合わせる新しい画家同定の手法で 13 作家の 1,300 枚の絵画に対して作家識別を行った。

深層学習を用いた自然言語処理による分類タ

スクを行う研究では、たとえば Alessa ら [2] は、FastText [4] を用いて SNS 上の投稿をインフルエンザ関連/非関連に分類する実験を行った。

本稿では、944 作家の 109,685 の手作り作品のデータセットを構築、利用し、作品に付随する写真画像および紹介文の双方を入力として作家識別を行う手法を提案する。

3. 提案手法と実験内容

3.1. 実験で使用するデータセット

本稿の実験で使用するデータセットは、ハンドメイドマーケット minne (<https://minne.com/>) から収集した写真画像と紹介文データで、下記の条件を満たす物である。

- 1) 2012 年 4 月 3 日 11:03 から 2019 年 9 月 1 日 23:31 の期間に出品された作品。
- 2) カテゴリがコサージュ・ブローチ。
- 3) 同一作家を除去。
作品数 50 未満の作家を除去

3.2. 評価尺度

提案手法の評価には、Overall Accuracy (以下、Accuracy) ならびに F1 値を利用する。

$$\text{Accuracy} = \frac{T}{N}$$

$$\text{F1 値} = \frac{2PR * PE}{PR + RE}$$

N = 全作品数

T = 予測作家が実際の作家と一致する作品数

PR = 作家識別の適合率

RE = 作家識別の再現率

3.3. 画像を対象とした作家識別モデル

CNN モデルの一つである VGG16 [3] を利用し、作品の写真画像を入力とする作家識別モデルを構築する。モデルの構造を図 1 左に示す。VGG16 は、特徴量抽出を行う畳み込み層 13 層と全結合層 3 層の計 16 層からなる。畳み込み層では、

Artist identification based on referral data of handmade works
{[†]Graduate School, [‡]Faculty} of Information Science and
Electrical Engineering, Kyushu University
§ Pepabo R&D Institute, GMO Pepabo, Inc.

転移学習により ImageNet で学習した重みを初期値とする。出力層の次元数は使用するデータセットの作家数とする。学習/検証用と評価用データには、各作家の作品画像を 4:1 の比率で分割して Hold-out 法で評価する。ただし、学習/検証用データは更に 5 等分し、学習用と検証用として 4:1 の比率で分割する。学習用と検証用データを交差検証の要領で入れ替えた 5 通りの学習モデルで評価結果を求め、その平均を識別結果とする。

モデルの学習/検証には Keras を、パラメータチューニングには Optuna を、そして最適化アルゴリズムには確率的勾配降下法を用いた。エポック数は 100 で、バッチサイズは 256 に設定した。ハイパーパラメータのチューニングにより、モデルの全結合層は 4096 次元の 1 層、learning rate は 0.005、momentum は 0.97 とした。

3.4. 紹介文を対象とした作家識別モデル

紹介文データは、MeCab (<https://taku910.github.io/mecab/>) を用いて分かち書きした後、各文を、Bag-of-Words (BOW) の one-hot vector に変換し、それを FastText [4] の入力として作家識別モデルを構築する。モデルの構造を図 1 右に示す。モデルの評価は、3.3 節と同様に行う。ハイパーパラメータは dim=100、epoch=1000 とし、他はデフォルト値である。

なお、紹介文を対象とした作家識別に用いるモデルとして FastText を選ぶ際には、以下の 4 つの方法の比較を行った。ただし学習モデルの構築と評価は 1 通りだけで行った。

1. 紹介文を Word2Vec で 100 次元ベクトル化し、LightGBM で作家識別。
2. 紹介文を Doc2Vec で 300 次元ベクトル化し、LightGBM で作家識別。
3. 全紹介文中の出現回数上位 5000 単語で、各紹介文を BoW ベクトル化し、各次元の値を対応する語の TF-IDF 値として LightGBM で作家識別。
4. FastText で作家識別。

手法 1 から 4 の Accuracy が、それぞれ 0.07050, 0.5005, 0.8364, 0.8806 であったことから、手法 4 の FastText を、紹介文データからの作家識別モデルとして採用した。



図 1 作品画像と紹介文での作家識別モデル

3.5. 画像と紹介文の複合学習モデル

3.3 節で記した方法で構築した写真画像と紹介文単体の各作家識別モデルの出力結果の平均 (944 次元) を取ることで、作家識別を行う。

4. 実験結果

表 1 実験結果

実験	Accuracy	F1 値
画像単独モデル	0.9299	0.9094
紹介文単独モデル	0.9119	0.8836
提案手法	0.9466	0.9283

表 1 より、提案手法が画像と紹介文の単独モデルより高い性能を示すこと、即ち、両データの組み合わせ利用の効果を確認できた。

5. 結論

5.1. まとめと今後の課題

本稿では、手作り作品の写真画像と紹介文を入力とした作家識別手法を提案した。944 作家の 109,685 作品をデータセットとして実験を行った。その結果、画像や紹介文を単独で用いた作家識別モデルと比べ、F1 値で各々 2% および 5% の性能向上を得られた。これにより、両方のデータを活用する重要性を確認した。

今後は、写真画像に紹介文の内容を埋め込む方法、たとえば物体認識時の識別ラベルを紹介文の中の単語に置き換える方法などを開発し、写真画像データだけの作家識別精度の向上を目指す。

謝辞

本研究の一部は科研費 JP21H00907 の支援を受けて行われた。

参考文献

- [1] K. A. Jangtik, M. C. Yeh, and K. Hua. 2016. Artist-based Classification via Deep Learning with Multi-scale Weighted Pooling. 24th ACM international conference on Multimedia (MM '16), 635–639, 2016.
- [2] A. Alessa, M. Faezipour and Z. Alhassan, "Text Classification of Flu-Related Tweets Using FastText with Sentiment and Keyword Features," 2018 IEEE International Conference on Healthcare Informatics (ICHI), 2018, pp. 366-367, doi: 10.1109/ICHI.2018.00058.
- [3] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition", 3rd International Conference on Learning Representations, (ICLR 2015), 2015
- [4] P. Bojanowski, et al. "Enriching word vectors with subword information." Transactions of the Association for Computational Linguistics 5 (2017): 135-146.