

# 多様な特徴量を考慮した Twitter ユーザの性別推定

廣田 遼 白井 清昭

北陸先端科学技術大学院大学 先端科学技術専攻科

{s2110142,kshirai}@jaist.ac.jp

## 概要

Twitter ユーザの性別を推定する手法を提案する。従来研究の多くはユーザが投稿するツイートを元に性別を分類していたのに対し、本研究は、投稿ツイート、自己紹介文、ユーザ名、プロフィール画像、フォロワーなど、様々な特徴量を考慮する点に特長がある。また、投稿ツイートのテキストと画像を同時に考慮した複数のモデルを提案する。評価実験の結果、提案手法の有効性を確認した。

## 1 はじめに

本論文は Twitter ユーザの性別を推定する新しい手法を提案する。従来研究の多くがツイートを対象に性別を推定している一方、Twitter では、投稿画像、自己紹介文、プロフィール画像、フォロー関係など様々な情報が取得できる。本研究では Twitter から得られる多角的な情報を元にユーザの性別をより正確に予測する。また、性別のラベルが付与された Twitter ユーザのデータは作成コストが高いため、ラベル付きデータを自動拡張することで性別推定の正解率を向上させることを狙う。

## 2 関連研究

Twitter ユーザのプロフィール推定については多くの先行研究がある。Sakaki らは、ユーザが投稿するツイートのテキストと画像のそれぞれについて、それを素性とする分類器を学習し、最終的に両者のアンサンブルモデルによって性別を予測する手法を提案した [1]。Wang らは、ツイートテキスト、プロフィール画像、名前、紹介文を素性とし、ユーザの性別、年齢、所属、組織を予測するモデルを提案した [2]。Liu らは、投稿テキスト、自己紹介文、統計量を素性とし、機械学習の古典的モデルならびに深層学習モデルによるプロフィール推定の性能を比較した [3]。Morgan-Lopez らは、誕生日ツイートなどの Twitter 特有の特徴量を利用することで、ユーザの

年齢推定の精度を向上させた [4]。

本研究でもツイートのテキストと画像の両方を利用するが、Sakaki らの手法がテキストと画像の分類器を別々に学習しているのに対し、本研究では両者の素性を同時に入力とする分類器を学習する点に違いがある。また、ツイート以外の情報を利用してプロフィールを推定する先行研究があるが、本研究では Twitter から得られるより多くの情報を同時に考慮するモデルを提案する。

## 3 提案手法

Twitter ユーザの性別を「男性」「女性」のいずれかに分類する。このため、使用する素性(特徴量)が異なる複数の分類器を学習する。そのうちの1つはユーザのツイートを素性とした分類器である(3.1項)。それ以外に Twitter から得られる様々な情報を素性とした分類器を学習する(3.2項)。最後にこれらの分類器を統合したモデルを学習する(3.3項)。

### 3.1 ツイートによる分類器

ここではユーザによって投稿された複数のツイートからそのユーザの性別を予測する。特に、ツイートのテキストと画像の両方を考慮し、これらの組み合わせ方によって複数のモデルを提案する。

性別推定は2段階で行う。(1) 個々のツイートに対し性別スコアを予測する。性別スコアは、1のときは男性、0のときは女性を表すものとする。(2) ユーザが投稿した複数のツイートの性別スコアからユーザの性別を決定する。以下の3通りの方法を採用する。

- 性別スコアの平均を求め、0.5以上のとき男性、それ未満のとき女性と判定する。以下、この方法を Tw-ave と記す。
- 性別スコアを Softmax 関数を用いて正規化した後、1.の処理を行う。以下、Tw-soft と記す。
- Softmax 関数による正規化の後、性別クラス毎

に性別スコアが1または0に近い上位10%のツイート(合計20%)を選別した後、1.の処理を行う。以下、Tw-selと記す。

以下、ツイートを対象にした個々の分類器の詳細について述べる。

**Text only model** ツイートのテキストのみを用いて性別を推定する分類器である。具体的には、訓練データによってfine-tuningしたBERT[5]を用いて性別スコアを得る。

**Image only model** ツイートに投稿された画像のみを用いて性別を推定する分類器。具体的には、Vision Transformer [6]を訓練データによってfine-tuningし、そのモデルを用いて性別スコアを得る。ただし、画像が含まれていないツイートは、学習時には使用せず、テスト時には性別を判定しない。

**Early fusion model** ツイートのテキストと画像を同時に考慮して性別を判定する。そのアーキテクチャを図1(a)に示す。テキストはBERT、画像はVision Transformerによってそれぞれ埋め込みベクトルに変換する。ただし、画像がないツイートに対する画像埋め込みはゼロベクトルとする。次に、これらを連結したベクトルを全結合層(Full Connected Layer;FCL)に渡して性別スコアを出力する。学習時の損失関数はLog Loss関数とする。

**Late fusion model** Early fusion modelと同じくツイートのテキストと画像を同時に考慮するが、両者の埋め込みを次元圧縮した後で組み合わせる。アーキテクチャを図1(b)に示す。テキストと画像を埋め込みをBERTとVision Transformerで得た後、それぞれを別々に全結合層に渡して4次元のベクトルに圧縮する。最後にこれらを連結したベクトルを全結合層に渡して性別スコアを出力する。

**Dense fusion model** Early fusion modelとLate fusion

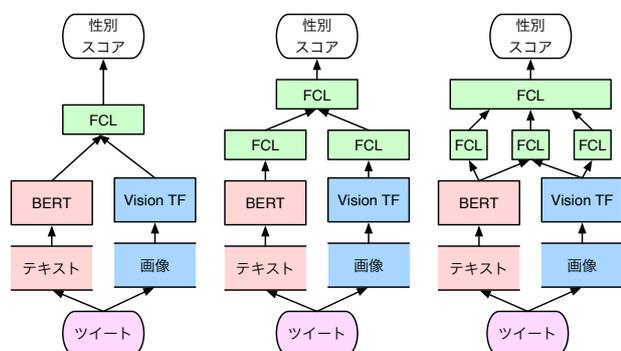


図1 ツイートに対する性別判定モデル

modelを組み合わせたモデル。そのアーキテクチャを図1(c)に示す。BERTによって得られるテキストの埋め込みベクトル、Vision Transformerによって得られる画像の埋め込みベクトル、Late fusion modelと同様にこれらを次元圧縮した埋め込みベクトルを連結して全結合層に渡し、性別スコアを得る。

**Caption model** ツイートのテキストと画像を同時に考慮するが、画像からキャプションを生成してから両者を組み合わせる。まず、Clip model [7]を用いて画像の英語のキャプションを生成する。次に、Googleの翻訳API [8]によって英語のキャプションを日本語に翻訳する。そして、ツイートのテキストと日本語のキャプションを特殊トークン(sep)を挟んで連結し、これを入力テキストとして、BERTを用いて性別スコアを得る。

**Ensemble model** Text only modelとEarly fusion modelを組み合わせたモデル。画像のないツイートはText only modelで、画像付きのツイートはEarly fusion modelで性別スコアを得る。

### 3.2 Twitterの情報による分類器

**Twitter統計量による分類器** Liuらの研究[3]を参考に、表1に示すTwitterから得られる統計情報を素性として分類器を学習する。学習アルゴリズムとしてLight Gradient Boosting Machine(Light GBM)[9]を用いる。

表1 性別推定に使用するTwitter統計量

一日の平均ツイート数、一時間ごとのツイート率、曜日ごとのツイート率
ツイートの平均統計量(文字数、ハッシュタグ数、単語数、特殊文字数、句読点数、漢字数)
自己紹介文の統計量(文字数、ハッシュタグ数、単語数、特殊文字数、句読点数、漢字数)
フォロワー数、フォロー数

**ユーザ名による分類器** Twitterユーザの名前から性別を予測する。具体的には、ユーザ名の文字のuni-gram、ひらがな表記のユーザ名の文字bi-gramを素性として性別推定の分類器を学習する。また、末尾の文字uni-gram、ひらがなbi-gramに“end\_”という特殊記号をつけたものも素性とする。学習アルゴリズムとしてLight GBMを用いる。

**プロフィール画像、ヘッダ画像による分類器** Twitter上のユーザのプロフィールページに表示されるプロフィール画像(ユーザの顔写真が多い)とヘッダ画像から性別を判定する。プロフィール画像のみ、ヘッダ画像のみ、両方を素性とする3つの分類

器を学習する。分類モデルとして Vision Transformer を用いる。

**自己紹介文による分類器** Twitter 上のプロフィールページに表示される自己紹介文から性別を予測する。一般に自己紹介は複数の文で書かれるが、本研究では 1 つの文として扱う。分類モデルとして BERT を用いる。

**フォロワーの自己紹介文による分類器** あるユーザーがフォローしている別のユーザーをランダムに 100 名選択し、そのユーザーの自己紹介文を元に性別を推定する。自己紹介文を MeCab を用いて形態素解析し、単語を素性、その TF-IDF を値とする素性ベクトルを作成する。さらに、Latest Semantic Indexing (LSI) によってユーザーのベクトルを 1000 次元に圧縮する。これを素性として Light GBM を学習する。

### 3.3 統合モデル

これまで述べてきた分類器を統合した性別判定モデルを学習する。3.1 項で述べたツイートを用いる分類器と、3.2 項で述べた全ての分類器を組み合わせる。各分類器が出力する性別スコアを素性とし、Light GBM を用いて最終的な分類器を得る。

### 3.4 訓練データの自動拡張

まず、初期の訓練データから性別の分類器を学習する。次に、ラベルのない大量の Twitter ユーザーのデータを用意し、学習した分類器を用いて各ユーザーの性別を判定し、ラベルを付与する。この際、判定の信頼度の大きい(性別スコアが 0 または 1 に近い)上位 3000 件のユーザーのみにラベルを付与する。

## 4 評価実験

### 4.1 実験データ

評価実験のために以下のデータを構築した。

**初期データ** 主に著名人の Twitter アカウントを対象にラベル付きデータを構築する。まず、Twitter 日本フォロワー数総合ランキング [10] から上位 2 万件のアカウントを収集する。それぞれのアカウントから、Twitter API を用いて分類器の学習に必要な情報を獲得する。ツイートについては、リツイートを除く最新の投稿を 100 件獲得する。

次に、それぞれのユーザーの性別を Twitter 以外の情報を参照して自動的に判定する。具体的には、各ユーザー(著名人)の Wikipedia のエントリーを参照し、

表 2 ツイートによる分類器を用いた性別推定の正解率

Model	初期データ			評価データ		
	ave	soft	sel	ave	soft	sel
Text only	.854	.852	.845	.659	.667	.624
Image only	.776	.774	.762	.754	.758	.754
Early fusion	.791	.791	.786	.732	.739	.721
Late fusion	.654	.608	.561	.617	.617	.725
Dense fusion	.582	.582	.672	.730	.732	.708
Caption	.847	.855	.823	.810	<b>.813</b>	.754
Ensemble	.878	.881	<b>.893</b>	.697	.697	.699

そのプロフィール画像を得る。そして、オンライン顔認識プラットフォーム Face++[11] を用いてユーザーの性別を判定し、それをラベルとして付与する。Wikipedia のエントリーがないとき、Google 画像検索を用いて上位 5 件の画像を取得し、それぞれの画像の性別を Face++ を用いて判定し、その多数決によって性別のラベル付けを行う。

上記の処理によって 5000 人のユーザーのラベル付きデータを獲得した。実験では、これを 8:1:1 の割合で訓練、開発、テストデータに分割して使用した。**ラベルなしデータ** 訓練データの自動拡張のため、性別のラベルが付与されていないユーザーの集合を得る。まず、前述のフォロワー数ランキングの順位 2 万から 3 万のユーザーを選ぶ。次に、Twitter API によって分類器の学習に必要な情報(ツイートについては初期データと同じく最新の 100 件のツイート)の取得を試み、これに成功した 6,845 名のユーザーをラベルなしデータとした。

**評価データ** 著名人ではない一般ユーザーを対象に人手で性別のラベルを付与し、評価データを構築する。初期データの著名人ユーザーをフォローしているユーザーのうち、フォロワー数が 1,000 人以下<sup>1)</sup>のユーザーを選別する。3 名の被験者に、Twitter ユーザーの最新の 20 件のツイートやプロフィール画面を参照し、「男性」「女性」「不明」のいずれかを付与することを依頼する。2 名以上の被験者が「男性」または「女性」のラベルを付与した 459 名のユーザーを選別し、評価データとした。

### 4.2 実験結果

#### 4.2.1 ツイートによる分類器の評価

3.1 項で述べたツイートのテキストと画像を元に性別を予測する分類器を評価する。実験結果を表 2

1) フォロワー数が多いユーザーは著名人の可能性が高いため。

表3 個々の分類器ならびに統合モデルの評価データに対する性別推定の正解率

訓練データ	Tw-ave	Tw-soft	Tw-sel	統計	ユーザ名	画像-P	画像-H	画像-PH	自己	フォ自己	統合
初期	.697	.697	.699	.636	.764	.810	.599	.643	.656	.584	.851
初期+拡張	.656	.719	.599	.644	.769	.755	.553	.603	.664	.612	.806
初期+拡張 (統合)	.749	.763	.736	.694	.795	.808	.577	.675	.699	.597	.806

Tw-ave, Tw-soft, Tw-sel はツイートに基づく分類器のうち表2の初期データの正解率が最も高い Ensemble model である。その他の分類器の略号は次の通り。統計:Twitterの統計量, 画像-P:プロフィール画像, 画像-H:ヘッダ画像, 画像-PH:プロフィール&ヘッダ画像, 自己:自己紹介文, フォ自己:フォロイの自己紹介文

に示す。ave, soft, sel は、3.1 項の冒頭で述べた個々のツイートの性別スコアを元にユーザの性別を予測する方式の違いを表す。

初期データでは、Text only モデルの正解率が、テキストと画像の両方の情報を用いる Early fusion, Late fusion, Dense fusion の各モデルよりも高い。ただし、画像を含むツイートについては、テキスト・画像を同時に考慮するモデルの精度が高いことが確認された。結果として、Text only model と Early fusion model を組み合わせた Ensemble model の正解率が最も高かった。なお、ave, soft, sel の優劣はモデルによって異なり、どの手法が最適であるかを結論付けることはできなかった。

一方、評価データでは、Image only model は Text only model よりも正解率が高く、Caption model 以外のテキストと画像を同時に使用するモデルよりも高い。著名人ではない一般ユーザの場合、男女によって投稿する画像に顕著な違いがあり、テキストよりも画像の方が性別判定の有力な手がかりになっていると考えられる。例外は Caption モデルで、比較した分類器の中では最高の成績を取っている。

#### 4.2.2 Twitter の情報による分類器の評価

3.2, 3.3 項で述べた分類器によって評価データの性別を判定したときの正解率を表3に示す。初期データのみ、初期データとそれぞれの分類器によって自動拡張したデータを使う場合(表3の「拡張」)、初期データおよび正解率の一番高い統合モデルによって推定した性別ラベルを付与して自動拡張したデータを使う場合(表3の「拡張(統合)」)を比較した。最後の実験条件は、正解率の高い統合モデルを用いることで、質の高い自動拡張データを獲得したときの効果を確認するためのものである。なお、統合モデルは「拡張」と「拡張(統合)」の実験条件とで同じモデルが得られる。

初期データを訓練データとしたとき、統合モデルの正解率は0.851であり、個々の分類器よりも高

い。ツイートのテキストや画像だけでなく、Twitter から得られる様々な情報を考慮することが性別推定の正解率の向上に寄与することが確認できる。個々の分類器では、プロフィール画像を用いた分類器の正解率が最も高い。プロフィール画像として本人の写真が使われている場合、性別推定の有力な情報となる。また、ユーザ名を用いた分類器も正解率が高く、名前に性差が強く現われていることがわかる。

拡張データを使用することで、個々の分類器では正解率が向上したのものもあるが、統合モデルの正解率は向上しなかった。特に正解率の高かったプロフィール画像を用いた分類器の性能が大きく低下していることから、統合モデルの正解率が0.045ポイント下がった。プロフィール画像やヘッダ画像を用いた分類器は、正解率が高いが、性別のスコアが0.5付近の場合が多く、性別推定の信頼度は低いことが確認された。自動拡張によって信頼度が低い事例が多く追加されていることが、プロフィール・ヘッダ画像を用いた分類器の正解率が低下した原因と言える。自動拡張データを得る手法として、それぞれのモデルよりも統合モデルの判定結果に基づいてラベルを決めた方が全体的には正解率が高い。このことは、自動拡張データの品質が向上すれば、訓練データの増加が有効である可能性を示唆する。今回の実験では、個々の分類器で追加する事例数を同じにするために、判定の信頼度の上位3000件のデータを追加したが、ある閾値以上の信頼度の事例のみ追加するなど、自動拡張するデータの性別ラベルの誤りを減らす工夫が必要である。

## 5 おわりに

本論文では、Twitter ユーザの性別を予測するための新しい手法を提案した。評価実験では、統合モデルの有効性は確認できたが、自動ラベル付けによる訓練データの拡張は効果がなかった。今後の課題として、拡張データの品質を高めるための手法を考案することが挙げられる。

## 参考文献

- [1] Shigeyuki Sakaki, Yasuhide Miura, Xiaojun Ma, Keigo Hattori, and Tomoko Ohkuma. Twitter user gender inference using combined analysis of text and image processing. In **Proceedings of the 25th International Conference on Computational Linguistics**, pp. 54–61, 2014.
- [2] Zijian Wang, Scott A. Hale, David Adelani, Przemyslaw A. Grabowics, Timo Hartmann, Fabian Flöck, and Dacid Jurgens. Demographic inference and representative population estimates from multilingual social media data. In **Proceedings of the World Wide Web Conference**, pp. 2056–2067, 2019.
- [3] Yaguang Liu, Lisa Singh, and Zeina Mneimneh. A comparative analysis of classic and deep learning models for inferring gender and age of Twitter users. In **Proceedings of the 2nd International Conference on Deep Learning Theory and Applications**, pp. 48–58, 2021.
- [4] Antonio A. Morgan-Lopez, Annice E. Kim, Robert F. Chew, and Paul Ruddle. Predicting age groups of Twitter users based on language and metadata features. **PLOS ONE**, Vol. 12, No. 8, p. e0183537, 2017.
- [5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 4171–4186, 2019.
- [6] Dosovitskiy Alexey, Beyer Lucas, Kolesnikov Alexander, Weissenborn Dirk, Zhai Xiaohua, Unterthiner Thomas, Dehghani Mostafa, Minderer Matthias, Heigold Georg, Gelly Sylvain, Uszkoreit Jakob, and Hounsby Neil. An image is worth 16x16 words: Transformers for image recognition at scale. **International Conference on Learning Representations**, 2021.
- [7] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In **Proceedings of the 38th International Conference on Machine Learning**, pp. 8748–8763, 2021.
- [8] Googletrans, (2023-1 閲覧). <https://pypi.org/project/googletrans/>.
- [9] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. **Advances in Neural Information Processing Systems**, Vol. 30, , 2017.
- [10] Twitter 日本 フォロワー数 総合ランキング, (2022-1 閲覧). [https://meyou.jp/ranking/follower\\_allcat](https://meyou.jp/ranking/follower_allcat).
- [11] Face++, (2023-1 閲覧). <https://www.faceplusplus.com/>.