漫画におけるセリフ位置と意味の時系列を考慮した話者キャラクターの推定

中川諒[1†] 梅澤猛[2] 大澤範高[2]

概要：漫画作品は、誌面構成の特性（コマ割り、吹き出し）や線画と文字が混在する様態のために、文字認識技術や人物検出技術を単独で用いた情報抽出が難しい。そのため、漫画メタデータの作成には手作業が必須で負担が大きい。とりわけ、任意のセリフとそれを発したキャラクターの対応付けは自動認識が困難であるとされている。そこで本研究では、漫画のセリフを位置と意味の観点から解析するとともに、時系列に対する学習モデルである再帰型ニューラルネットワーク（RNN）を利用して、セリフの話者を自動推定する手法を提案する。提案手法は、意味情報と時系列から文脈を把握することで、これまで困難であったキャラクターとセリフが同一コマに存在しない場合であっても精度よく対応付けが可能である。対応推定には特徴量として。これまでに有効であることが示されている吹き出しに対する各キャラクター間の距離およびセリフの分散意味表現を用いる。また、漫画は物語が時系列で繋がっていることから、RNN を使うことで文脈から得られる情報によってキャラクターの推定精度を高められると期待できる。セリフに対する各キャラクターの距離情報による推定、そこにセリフの分散意味表現を加えた特徴ベクトルによる推定、距離情報と分散意味表現に RNN を使った提案手法による推定の 3 つの推定精度を比較し、提案手法の有効性を検証した。

1. はじめに

漫画作品は、誌面構成の特性(コマ割り、吹き出し)や線画と文字が混在する様態のために、文字認識技術や人物検出技術を単独で用いた情報抽出が難しい。そのため、漫画メタデータの作成には手作業が必須で負担が大きい。特に、任意のセリフとそれを発したキャラクターの対応付けは自動認識が困難であるとされている。この作業は、キャラクターとセリフの検出とそれらの対応付けという 2 つのタスクで構成される。本研究では、キャラクターとセリフの位置と内容が既知である場合の、キャラクターとセリフの対応付けの精度向上を目的とする。

2. 漫画からの情報抽出

2.1 テキスト情報

漫画からセリフやナレーションなどのテキスト情報を取得できれば、他言語へ容易に翻訳を行ったり、文字のサイズやスタイルを変更できたりと利便性が高いことから、さまざまなアプローチで研究が行われている。このうちの多くは文字領域の検出手法を主とし、検出した領域を文字認識技術によってテキスト化する。このときの検出精度には、用いる特徴量と使用するデータセットが大きく影響する。

田中らは、漫画の吹き出しの検出を検出し、吹き出しの種類を分類している [12]。検出には、Haar-like 特徴量を用い、AdaBoost アルゴリズムによる分類器作成を行うことで、セリフ部分の文字を学習して分類器を作成している。荒巻らは、連結成分に着目して、文字の連結成分の特徴を学習させることで文字領域の検出を行う手法を提案している [13]。連結成分を用いた際に発生する誤検出に対しては、深層学習を利用して文字領域かどうかを再分類することで誤検出を低減させる手法も提案している [14]。この他、深層学習を利用したテキストの検出手法として、柳澤らは Faster-RCNN を用いたメタデータ抽出を行っている [15]。これまでの方法に比べ、高精度な検出が可能であり、吹き出し検出では 90% を超える平均適合率を記録している。また、漫画に似たものとしては、スケッチ画像を対象とした物体認識に対しても Convolutional Neural Network（CNN）が有効であることが示されている [16] [17]。以上より、ニューラルネットワークの適用は漫画に対しても有効であることがわかる。さらに、小川らは漫画専用のネットワークモデルを構築することで、漫画の検出として問題となる要素の重なりをうまく認識し、従来の汎用モデルよりも高精度の検出が可能な手法を提案している [18]。

2.2 キャラクター情報

キャラクター（登場人物）の情報は、漫画を構成する様々な要素の中で、テキスト情報と同様にアクセスと検索において重要な役割を果たす要素である。そのためキャラクターを認識して分類する試みは多く、キャラクターの検出と認識には、キャラクターの顔をうまく認識して分類することが重要になる。顔検出は、コンピュータビジョンにとって基本的な処理であり、自然画像において広く研究されてきたが、漫画には、基本的にグレースケールが用いられる、目の表現が多様である、人間の顔の性質（例：鼻パーツ）を持たないものが存在する、など自然画像にはない特徴があることから、人間の顔検出のために提案された手

1 千葉大学大学院融合理工学府
Graduate School of Science and Engineering, Chiba University
2 千葉大学大学院工学研究院
Graduate School of Engineering, Chiba University
† 現在，KDDI（株）
Current affiliation is KDDI Corporation.

法では、うまく認識することができない。そこで漫画のキャラクター検出手法として、石井らの HOG 特徴量を用いた SVM による検出 [19]をはじめとして、機械学習を用いた漫画キャラクターの検出 [18] が提案されている。また、漫画テキストと同様、キャラクターについても漫画用ネットワークを作成することでよって検出精度が高まることも示されている [20] [21]。

2.3 漫画のテキストとキャラクターの対応付け

漫画セリフとそのキャラクターの対応付けは、漫画の内容理解に必要となる要素である。しかし、現状では方法が確立されておらず、機械による自動的な対応付けが難しい。キャラクターとセリフを紐づける先行研究として、吹き出しとキャラクターの距離による話者推定がある [1]。この研究では、まず各オブジェクト（キャラクター、吹き出し）に対してアンカーポイントと呼ばれる基準点を設定する。アンカーポイントには優先度があり、得られた情報が多い順番にポイントが遷移する。吹き出しのアンカーポイントに設定される場所は、吹き出しの外接矩形の中心、吹き出しの中心、吹き出しの尾の位置の順に変化し、後ろに行くほど優先度が高い。つまり、すべての情報が得られた場合は吹き出しの尾の位置をアンカーポイントとして設定する。キャラクターのアンカーポイントは、矩形の中心、キャラクターの中心、顔の中心、口の中心の順に優先度が高く設定される。そして、吹き出しとキャラクターに設定したアンカーポイントの距離が最も近いものをセリフの話者と推定する。この手法では、アンカーポイントの位置に限らず使用する情報は 2 点間の距離である。そのため、話者とセリフの位置が離れている場合にはうまく対応付けすることができず、限定的な範囲にのみ有効な手法である。そのため、同一コマに複数キャラクターが存在する場合や、キャラクターが同一コマに存在しない場合は、誤認識してしまうこと、漫画に大きく依存することが課題であると述べられている

データドリブンな手法としては、山本らの吹き出しやキャラクターなどの物体情報をデータ分析することで推定する漫画の話者推定がある [2]。この手法では、2 点間距離だけでなくキャラクターや吹き出しの大きさ、同一コマにおけるキャラクターの有無などの 6 種類の特徴量に加え、セリフ領域の吹き出し画像の画像特徴量を組み合わせて推定を行っている。結果として、これまでの手法に比べて Neural Network を用いることで精度が向上しているが、吹き出しの画像特徴を入力しても精度の向上には寄与しないことが述べられている。その中で 6 種類の特徴量のうち同一フレームに存在するかどうかが精度に一番影響することが述べられている。しかしこの手法においても、コマが読まれる順番を考慮していないため、ペアとなるキャラクターとセリフがコマを跨ぐ場合に精度が落ちてしまうことが予想される.

3. セリフの文脈を考慮した話者推定

キャラクターとセリフが同一コマに存在しない場合にも正しく推定を行うために、漫画のセリフやキャラクターから文脈を理解することによって話者推定を行う。推定に用いる特徴として、ページ内での位置関係を把握するために、キャラクターとセリフ間の距離を 1 つ目の特徴とする。さらに、意味の観点からセリフの解析を行うために、セリフテキストの分散表現を 2 つ目の特徴とする。学習する際に会話の順番を考慮するために、Recurrent Neural Network（RNN）を時系列に対して拡張した学習モデルである Long short-term memory（LSTM）を用いてセリフの話者を自動推定する。

4. 提案手法

本研究では、キャラクターとセリフの位置と内容が正しく認識できていると仮定し、セリフに対応する話者の推定を行う手法を提案する。特にキャラクターとセリフが同一コマに存在しない場合にも推定を可能とするために、漫画の文脈を理解して応用することで話者推定を試みる。意味の観点から解析するために、セリフテキストの分散表現を用いたセリフ内容を特徴量とする。また、ページ内での位置関係を把握するため、キャラクターとセリフとの距離情報も特徴とし、この 2 つを特徴量として扱う。最後に、会話の順番を考慮した学習をするため、時系列に対する学習モデルである LSTM を利用して、セリフの話者自動推定を試みる。提案手法による推定手順を図 1 に示す

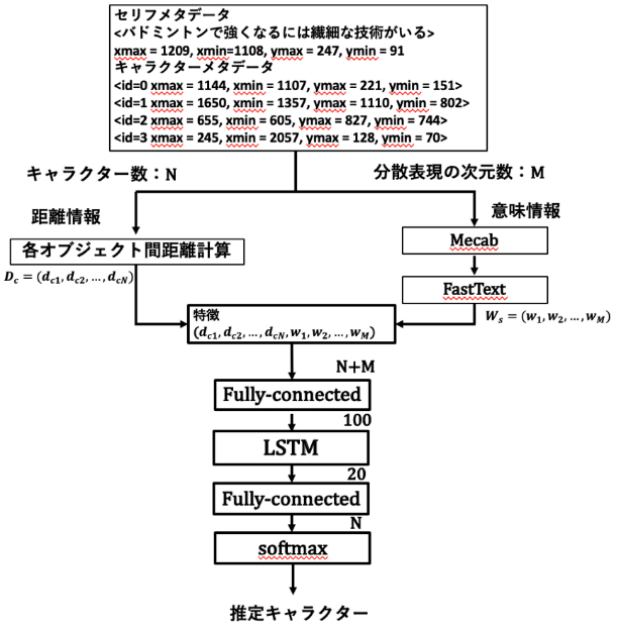


図 1　ネットワーク構造と推定の手順

```
<characters>
  <character id="00035f83" name="成瀬川なる"/>
  <character id="00035f85" name="浦島景太郎"/>
  <character id="00035fa6" name="前原しのぶ"/>
  <character id="00035fa8" name="青山素子"/>
  <character id="00035faa" name="紺野みつね"/>
  <character id="00035fac" name="カオラ・スゥ"/>
  <character id="00036141" name="はるかおばさん"/>
  <character id="00036b93" name="女の子"/>
  <character id="00036bf6" name="白井功明"/>
  <character id="00036c03" name="灰谷真之"/>
</characters>
<pages>
  <page index="0" width="1654" height="1170"/>
  <page index="1" width="1654" height="1170">
    <text id="00035f9a" xmin="670" ymin="54" xmax="702" ymax="136" character="00036b93">知ってる？</text>
    <text id="00035fa0" xmin="543" ymin="47" xmax="623" ymax="206" character="00036b93">アイしあう二人がトーダイってとこにはいるとね</text>
    <text id="00035f90" xmin="188" ymin="38" xmax="234" ymax="178" character="00036b93">シアワセになれるんだって</text>
    <text id="00035f8b" xmin="57" ymin="184" xmax="77" ymax="237" character="00035f85">ふーん</text>
    <text id="00035f8a" xmin="642" ymin="376" xmax="727" ymax="605" character="00036b93">大きくなったらふたりで</text>
    <text id="00035f92" xmin="85" ymin="361" xmax="183" ymax="643" character="00036b93">いっしょにトーダイ行こーね</text>
    <text id="00035f9e" xmin="681" ymin="871" xmax="723" ymax="968" character="00036b93">バイバーーイけーくん！</text>
    <text id="00035f99" xmin="293" ymin="881" xmax="365" ymax="987" character="00035f85">大人になったら絶対トーダイで会おうね！</text>
    <text id="00035f9f" xmin="220" ymin="1010" xmax="243" ymax="1112" character="00035f85">ヤクソクだよ！</text>
    <text id="00035f88" xmin="123" ymin="858" xmax="147" ymax="910" character="00035f85">・・・・</text>
    <text id="00035f9c" xmin="31" ymin="1075" xmax="69" ymax="1153" character="00035f85">わ・・・・</text>
  </page>
  <page index="2" width="1654" height="1170">
    <text id="00035fb4" xmin="1567" ymin="16" xmax="1613" ymax="264" character="00035f85">わかった!!</text>
    <text id="00035fa1" xmin="1426" ymin="883" xmax="1494" ymax="1129" character="00035f85">東大だね!!</text>
    <text id="00035fa3" xmin="1298" ymin="383" xmax="1346" ymax="521" character="00035fc0">15年後</text>
    <text id="00035fb0" xmin="1317" ymin="86" xmax="1426" ymax="271" character="00035f85">ーーー俺浦島景太郎19歳（彼女ナシ）</text>
    <text id="00035fb5" xmin="936" ymin="186" xmax="1065" ymax="406" character="00035f85">そういうワケでとりあえす２浪してます</text>
  </page>
```

図 2　アノテーションを付与したデータセット例

提案手法による推定は、1 件のセリフデータを入力とし、対応する話者を出力とするモデルで表される。入力となるセリフデータは、データセットから抽出した各キャラクターとの距離データおよびセリフの分散意味表現を用いて記述する。2 種のデータ特徴をそれぞれ連結して合成特徴を作成し、全結合層に入力することで特徴圧縮を行う。圧縮した特徴を LSTM そうへの入力として、再度全結合層に入力し、登場キャラクター数と同数の出力結果を得ることで、個々に確率を計算して最も確率が高いキャラクターを話者と推定する。

5.　特徴量

5.1　データセット

実験に使用する漫画情報のデータセットは Manga109 の中からジャンルの異なる「ラブひな」と「やまとの羽根」の 2 作品を選択し、それぞれ話の流れに正確となるようセリフを順番に並べ、セリフ話者をアノテーション付与して作成した。アノテーションの対象は発話者が明確なセリフのみとし、ナレーションや無名キャラクターは対象外とした。作成したデータの一例を図 2 に示す。*text* タグ内の *xmin*, *xmax*, *ymin*, *ymax* はキャラクターの検出領域矩形の座標を表しており、*character* に指定された id が当該セリフの初話者を示している。

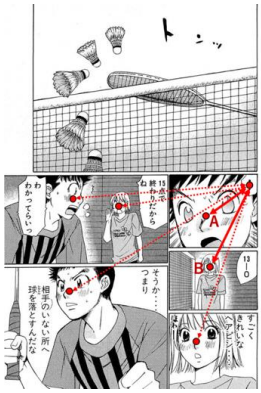
5.2　距離情報

距離を元にした話者推定は一定の成果を残しており、漫画でのオブジェクト間の距離関係は話者推定に有効であると考えられる。Rigaud ら [1] のように、吹き出しの尾や口の位置で計算する方が汎用的であるが、本研究では検出誤差による影響を排除するため、セリフと各キャラクターの外接矩形の中心とのユークリッド距離を利用した。対象漫画作品に含まれるキャラクター数を N とし、キャラクター c とセリフとの距離を d_c とすると、対象となるセリフ s の特徴ベクトルは次式で表すことができる。

$$D_s = (d_{c1}, d_{c2}, \cdots d_{cN})$$

また、同じページ内に同一キャラクターが複数描かれている場合には、距離が最も近いものを特徴量として採用する。ページ内に存在しないキャラクターの場合には、見開き 1 ページの最大横幅を距離と定義する。

特徴量を可視化した例を図 3 に示す。図中のページにはキャラクターが二人しか描かれていないため、キャラクターid が A と B 以外のものには、ページ最大幅である 1,654 が割り当てられている。今、2 コマ目の吹き出しについて話者推定をするとき、ページ内のキャラクターすべてについて顔部分の外接矩形の中心を求め、吹き出しの中心とのユークリッド距離を計算する。図中のページにはのべ 6 人が描かれているが、同じ人物がそれぞれ 3 回描かれているので、キャラクター毎に最小となる距離を求め、A は同じコマ内で距離 253、B は直下のコマの 449 と割り当てる。各キャラに割り当てられた距離情報の中から、最小である 253 を持つキャラクターA が当該吹き出しの話者であると推定することができる。



```
<character id="A" name="鳥羽大和"/>
<character id="B" name="鳥羽撫子"/>
<character id="C" name="沢本翔"/>
<character id="D" name="大和の母"/>
<character id="E" name="サッカー部顧問"/>
<character id="F" name="早坂"/>
<character id="G" name="バドミントン部顧問"/>
<character id="H" name="高津仁"/>
<character id="I" name="小野"/>
<character id="J" name="小泉正平"/>
```

各キャラクターの距離

A	B	C	D	E	F	G	H	I	J
253	449	1654	1654	1654	1654	1654	1654	1654	1654

図 3 キャラクターとセリフとの距離可視化例
“やまとの羽根”©咲 香里

5.3 意味情報

セリフの内容を解析することで、意味的なアプローチから話者推定を試みる。山本ら [2] は、セリフの内容を隠して吹き出しの形状のみを情報として与えた場合、人手による話者推定を行った場合であっても推定が難しいパターンが存在すると報告している。そのため、さらなる精度向上にはセリフの解析による内容理解が必要となる。セリフを単語や文章の分散表現を用いて数値化することで、話者推定に利用可能な特徴量とする。セリフをベクトルとして扱うために、可変長の入力が可能で学習が高速な fastText を用い、Manga109 データセットの全セリフデータ 147,919 件をコーパスとして学習を行っておく。セリフ文字列は、あらかじめオープンソースの形態素解析エンジン MeCab を用いて形態素に分解し、その後 fastText によって分散意味表現を得る。分散表現の次元数を M とすると、セリフ s の分散表現 W_s は次式で表すことができる

$$W_s = (w_1, w_2, \cdots, w_M)$$

6. 検証実験

6.1 話者推定における距離情報の有効性

吹き出し領域とキャラクターの距離が最も近いものを話者と推定する手法の有効性を検証した。特徴量として 5.2 の距離情報を用い、MLP（MultiLayer Perceptron）による推定を行った。使用する MLP は全結合層 3 層で構成され、ネットワークモデルは 3 層構造である。比較対象として、Rigaud らによる話者推定手法[1]を用いた結果を使用し、検証データセットに対して学習時に最も高い結果を示した正解率は表 1 の通りであった。

表 1 距離情報を用いた話者推定の正解率比較

作品	ラブひな	やまとの羽根
Rigaud らの手法	0.676	0.716
提案手法（距離）	0.663	0.750

Rigaud らの手法を用いた場合と正解率に大きな差は確認されず、距離データだけからでも 6〜7 割は正しく話者推定できるという結果が得られた。

推定を誤っている例を検証したところ、セリフとキャラクターが同一駒に存在していない、吹き出しと発話キャラクターの間に別キャラクターが描かれている場合が確認された。

6.2 話者推定における意味情報の有効性

文章の意味情報による話者分類の有効性を検証するため、セリフテキストの分散表現のみを特徴量とする推定を行った。分散表現はセリフテキストを fastText に掛けることで取得し、特徴ベクトルの次元数は、学習が単語数 50M 以下の時に適切とされる 100 次元とした。MLP ネットワークモデルは 6.1 と同様とした。比較対象として、fastText の学習機能により多重ロジスティック回帰分析によるクラスタリングによって話者を推定したところ、正解率は表 2 の通りであった。

表 2 意味情報を用いた話者推定の正解率比較

作品	ラブひな	やまとの羽根
fastText の学習機能	0.405	0.456
提案手法（意味）	0.551	0.560

いずれの手法も、距離データを利用した推定よりも正解率が低下した。意味データは連続的なつながりによる情報量が多いため、前後関係を切り離して意味的な傾向を利用しても有効な推定は困難であると考えられる。特に、fastText の学習機能を使ったクラスタリングにおいては、テキストの意味学習にもラベル付きデータを必要とするため、当該漫画のセリフデータしか利用できず、ラブひなでは 1,433 件、やまとの羽根では 721 件しか利用できるデータがなかったことが低正解率の一因であると考えられる。

6.3 MLP を用いた話者推定における複数特徴の有効性

距離情報と意味情報を組み合わせることで、単独の情報を用いたときよりも話者推定の正解率が向上するかを検証する。6.1 で用いた距離情報と 6.2 で用いた意味情報を組み合わせて、同様の MLP によって話者推定を行ったところ、正解率は表 3 の通りであった。

表 3 距離情報と意味情報を用いた話者推定の正解率

作品	ラブひな	やまとの羽根
距離＋意味	0.691	0.800

どちらの作品においても、距離情報または意味情報を単独で用いた時よりも高い正解率を示した。両方の特徴量を用いることは、漫画の話者推定に有効であることが示唆された。ただし、全体としてデータが少ないため、この結果は評価データと学習データの分割様態に依存している可能性がある。

6.4 LSTM に基づく意味情報による話者推定

漫画作品は画像一ページで完結するものではなく、画像同士に連続性があり、セリフの内容には物語が深く関係しており、時系列が存在する。そこで、セリフのベクトル表現である意味情報を特徴量として、LSTM によって学習を行い、話者を推定することの有効性を検証する。

データ入力はセリフの順に従うものとし、評価を行うデータも順に与えた。この実験においては、セリフとキャラクターの位置は考慮せず、セリフの文脈から次の話者を推定することが可能かを検証する。データの入力には連続性を持たせるために、まず先頭から 8 割を学習データとして用い、残り 2 割をテストデータとした。評価データに対する正解率は表 4 の通りであった。

表 4 意味情報を用いた LSTM による話者推定の正解率

作品	ラブひな	やまとの羽根
LSTM（意味情報）	0.617	0.522

意味情報のみを用いた LSTM による推定は、MLP と同様、低い正解率を示した。意味情報だけの推定では過学習が発生し、LSTM を取り入れてもなお改善されなかったと考えられる。

6.5 LSTM に基づく複数特徴による話者推定

距離情報による特徴量とセリフのベクトル表現である意味情報の両方を特徴量として、LSTM による学習を用いることで、話者推定の正解率が向上するかを検証する。セリフとキャラクターの位置が離れている場合や、コマを跨いでいる場合にも、文脈を考慮することで話者の推定が可能になると期待できる。6.4 と同様に、データの先頭から 8 割を学習データに、残り 2 割をテストデータとした場合に加え、後半 8 割を学習データに、先頭 2 割をテストデータとした場合のそれぞれについて LSTM による話者推定を行った結果は表 5 に示した通りであった。なお、1,000 epoch では値が収束しなかったため、5,000 epoch で検証を行った。

表 5 距離情報と意味情報を用いた LSTM による話者推定の正解率

作品	ラブひな	やまとの羽根
学習 (8) :テスト (2)	0.573	0.620
テスト (2) :学習 (8)	0.732	0.733

データの分け方によって正解率が大きく変動を示した。学習データが不足しており、学習が不安定になっている可能性が考えられる。そこで、学習データと評価データの分割比率を 7:3、8:2、9:1 と変化させて同様の検証を行ったところ、それぞれの最大正解率は表 6 の通りとなった。

表 6 データの割合を変更した際の距離情報と意味情報を用いた LSTM による話者推定の最大正解率

作品	ラブひな	やまとの羽根
学習 (7) :テスト (3)	0.689	0.735
学習 (8) :テスト (2)	0.732	0.733
学習 (9) :テスト (1)	0.754	0.850

両作品において、学習データが増加するに従い、正解率も向上する結果となり、9 割を学習データに使用したときが最大の正解率となった。この結果より、学習データ数が増えることで精度がさらに向上することが示唆された。

7. 考察

まず、データセットとした漫画 2 作品の作風が検証結果に与えた影響について考える。「ラブひな」は主人公以外に多くの女性キャラクターが登場し、各話でメインとなる相手が変わる傾向にある。データセットに使用した第一巻では、5 名のヒロインのうち 3 名とのやり取りがメインになっており、どこを学習データ/訓練データとして利用するかが正答率に影響すると考えられる。今回は一巻分のデータを学習データと検証データに分割したためにデータが少なく、結果の安定性を保証することができなかった。「やまとの羽根」はバドミントンを題材としたスポーツ漫画であり、全体を通してセリフ数が少ない。さらに試合のシーンが多く、キャラクターが描かれていないセリフのみのコマが比較的多い。この作品において正解率が上がらなかった原因として、審判役のキャラクターの得点コール時のセリフが多かったことが挙げられる。発話しているキャラクターはほとんどコマに登場せず、得点カウントのセリフのみが連続して登場することで、その部分の話者推定ができず、大きく正解率を下げる結果となった。意味情報をうまく活用してこのようなシーンの話者推定を実現することは今後の課題といえる。どちらの作品も、主人公やヒロインといった最初から継続して登場するキャラクターのみに注目すると、提案手法によって正解率が向上していることが確認できる。従って、学習データが十分に得られれば、話者推定を正しく行うことができる可能性を示している。また、メタデータとしての活用場面を想定すると、利用価値の高いメインキャラクターの推定制度が良好であるという点は既存の手法に比べ優位性があると考えることができる。

次に、使用した特徴量について、6.1 の検証結果より、セリフとキャラクターの組み合わせの役 7 割は近接して配置されており、距離情報は話者推定に有効な特徴量であるといえる。しかし、漫画のコマを読む順番で考えると、見開きで考えた場合、右ページ下部の次には左ページ上部へと移る場合が多く、単純なユークリッド距離ではない指標も考慮する余地がある。意味情報については、今回は漫画の

セリフテキストの解析という観点から、漫画データセットに含まれる 147,919 件のセリフテキストをコーパスとしたが、漫画に限らないさまざまな会話テキストを利用するなど質よりも量を優先したコーパスを用いた時の正答率への影響について調査が必要である。

最後に、LSTM による時系列モデルを適用した効果について考える。既存手法の多くは、漫画が時系列データであることは考慮していない。漫画には物語としての時系列が存在し、この情報を利用することはセリフの話者を推定するのに有効であると考えられる。しかし、今回の検証では LSTM によって正解率が向上しない例もみられた。LSTM による手法を効果的に利用するためには、学習データとして適切な連続性を持ったデータ群を与えること、学習したモデルが効果を発揮するテストデータに適用することが考えられる。そのためには、要約や場面推定など他の解析アプローチによるメタデータを活用するなど広い視点での取り組みが必要である。

8. まとめ

本研究では距離情報による位置関係とセリフテキストの内容解析による意味情報からアプローチし、データの時系列を考慮することで話者を推定する手法を提案した。既存の距離が最も近いものを選択する手法やデータドリブンな手法と比較するために実験を行った。本論文中で提案した特徴である意味情報は、単独で特徴として話者特定を行った場合に、有効性が見られず、距離を特徴として用いた場合よりも精度が低い結果であった。しかし、距離情報と組み合わせた特徴による話者推定では、距離情報や意味情報を単独で行った結果よりも精度の向上が確認でき、既存手法よりも正解率が向上した。これはセリフとキャラクター間の距離だけでなく､セリフの意味情報を加えることで、同じ距離範囲に複数の話者がいた場合に、推定する情報が増加することによってより正確な推定を行うことができたと考えられる。また LSTM を使って系列データとして話者推定を行った場合に、作品によっては正解率が向上することを確認した。結果としては先行研究に比べて正解率が最大で 13 パーセント向上し、セリフ数が多いキャラクターに着目するといずれも精度が向上している。しかし学習データや評価データの違いによって精度が変化する可能性があるため、さらにデータセットを大きくして再度検証する必要がある。本研究で利用したデータセットに含まれている漫画作品は場面によってキャラクターの出現率に偏りがあったため、学習の仕方に工夫が必要になる。そのため、どのような漫画に対しても有効という点は、今後の課題である。

参考文献

[1] Christophe Rigaud, Nam Le Thanh, J.-C. Burie, J.-M. Ogier, Motoi Iwata, Eiki Imazu, and Koichi Kise, “Speech balloon and speaker association for comics and manga understanding,” In Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR), 2015.

[2] 山本和慶, 小川徹, 山崎俊彦, 相澤清晴, “データドリブンなアプローチを用いた漫画画像中の吹き出しの話者推定,” 電子情報通信学会技術研究報告, 117(431), pp. 287-292, 2018.

[3] 新納浩幸, Chainerv2 による実践深層学習, オーム社, 2017.

[4] Sergey Ioffe, and Christian Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” In Proceedings of the 32nd International Conference on Machine Learning (ICML), Vol.37, pp.448-456, 2015.

[5] 岡谷貴之, MLP 深層学習, 講談社, 2017.

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet classification with deep convolutional neural networks,” In Proceedings of the 25th International Conference on Neural Information Processing Systems, pp.1097-1105, 2012.

[7] Alfredo Canziani, Adam Paszke, and Eugenio Culurciello, “An analysis of deep neural network models for practical applications,” IEEE International Symposium on Circuits & Systems, 2017.

[8] Felix A. Gers, Jürgen Schmidhuber, and Fred A. Cummins, “Learning to forget: Continual prediction with LSTM,” Neural Computation, Vol.12, Issue 10, pp.2451-2471, 2000.

[9] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, “Efficient estimation of word representations in vector space,” In Proceedings of the International Conference on Leaning Representations (ICLR), 2013.

[10] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean, “Distributed representations of words and phrases and their compositionality,” In Proceedings of the 26th International Conference on Neural Information Processing Systems, Vol.2, pp.3111-3119, 2013.

[11] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov, “Bag of tricks for efficient text classification,” In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Vol.2, pp.427-431, 2017.

[12] 田中孝昌, 外山史, 宮道壽一, 東海林健二, “マンガ画像の吹き出し検出と分類,” 映像情報メディア学会誌: 映像情報メディア, Vol.64, No.12, pp.1933-1939, 2010.

[13] 荒巻祐治, 松井勇佑, 山崎俊彦, 相澤清晴, “連結成分に基づいた漫画における文字領域の検出,” 映像情報メディア学会年次大会講演予稿集, 2015, 32C-1, 2015.

[14] 荒巻祐治, 松井勇佑, 山崎俊彦, 相澤清晴, “連結成分と領域の分類に基づいた漫画における文字領域の検出,” 電子情報通信学会論文誌 A, Vol.100, No.1, pp.3-11, 2017.

[15] 柳澤秀彰, 渡辺裕, “Faster R-CNN を用いたマンガ画像からのメタデータ抽出”, 映像情報メディア学会年次大会, 2016.

[16] 佐々木一磨, 尾形哲也, “手描きスケッチを扱う深層学習モデル”, 日本画像学会誌, Vol.56, No. 2, pp.177-186, 2017.

[17] Qian Yu, Feng Liu, Yi-Zhe Song, Tao Xiang, Timothy Hospedales, and Chen Change Loy, “Sketch me that shoe,” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.799-807, 2016.

[18] 小川徹, 山崎俊彦, 相澤清晴, “並列化された検出器による高精度漫画物体検出”, 電子情報通信学会技術研究報告, Vol.117, No.432, pp.293-298, 2018.

[19] 石井大祐, 渡辺裕, “マンガからの自動キャラクター位置検出に関する検討,” 情報処理学会研究報告オーディオビジュアル複合情報処理 (AVM), 2012-AVM-76, Vol.1, pp.1-5, 2012.

[20] Nhu-Van Nguyen, Christophe Rigaud, and Jean-Christophe Burie, “Comic characters detection using deep learning,” 14th IAPR

International Conference on Document Analysis and Recognition (ICDAR), Vol.3, pp.41-46, 2017.
[21] Wei-Ta Chu, and Wei-Wei Li, “Manga FaceNet: Face detection in manga based on deep neural network,” In Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, pp.412-415, 2017.
[22] Yusuke Matsui, Kota Ito, Yuji Aramaki, Toshihiko Yamasaki, and Kiyoharu Aizawa, “Sketch-based Manga Retrieval using Manga109 Dataset,” Multimedia Tools and Applications, Vol.76, Issue 20, pp.21811-21838, 2017.
[23] Azuma Fujimoto, Toru Ogawa, Kazuyoshi Yamamoto, Yusuke Matsui, Toshihiko Yamasaki, and Kiyoharu Aizawa, “Manga109 Dataset and Creation of Metadata,” International Conference on Pattern Recognition workshop MANPU, 2016.
[24] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov, “Enriching word vectors with subword information,” Transactions of the Association for Computational Linguistics, Vol.5, pp.135-146, 2017.