

# テキストの文脈生成に基づくユーザ埋め込み表現学習

濱 健太<sup>†,††</sup> 大塚 淳史<sup>††</sup> 石井 亮<sup>††</sup> 野本 済央<sup>††</sup> 深山 篤<sup>††</sup>

中村 高雄<sup>††</sup>

<sup>†</sup> 大阪大学 大学院基礎工学研究科

<sup>††</sup> 〒 560-8531 大阪府豊中市待兼山町 1-3

<sup>††</sup> 日本電信電話株式会社 NTT デジタルツインコンピューティング研究センタ

<sup>††</sup> 〒 108-0023 東京都港区芝浦 3-4-1 グランパークタワー 33F

E-mail: <sup>††</sup>ll29.hamaken@gmail.com,

<sup>††</sup>{atsushi.otsuka.vs,ryo.ishii.ct,narichika.nomoto.ds,atsushi.fukayama.ap,takao.nakamura.vp}@hco.ntt.co.jp

あらまし 自然言語処理分野では一般的なタスクにおいては人と同水準の精度を実現できるようになった。それに伴い、言語処理モデルが個人の属性や性格に応じた柔軟な出力を得ることを目標とした新たな研究の方向性が注目されている。しかし、モデルの学習にとって有用な属性情報を収集したり、個人の性格などの内面をモデル化するのは困難である。本論文では、言語処理モデルに個性を持った振る舞いをさせることを目的とした個人の埋め込み表現を、オープンなデータから自動で獲得する手法を提案した。個人の持つ属性や性格などの情報は、一連の振る舞いや発言、テキストとなって外部に表出すると考えられる。提案手法は、SNS のユーザが投稿したテキストに着目し、入力テキストの文脈となる前後の文を予測し生成するモデルを学習する。入力にユーザ固有の ID 情報を追加し学習することで、ユーザに依存した文脈生成モデルを作成し、ユーザ固有の埋め込み表現を獲得する。複数の評価実験から、提案手法はユーザ自身の出身地等の属性の他、価値観や性格が反映されている埋め込み表現を獲得できたことが示された。

キーワード 埋め込み表現, 生成モデル, ユーザ情報, デジタルツイン

## 1 はじめに

近年、自然言語処理分野では BERT [1] に代表される深層学習による言語モデルの登場により、要約 [9] や翻訳 [16]、質問応答 [19]、対話 [12] といった言語処理タスクにおいては人と同水準の精度を実現できるようになった。それに伴い、これらのタスクにおいて単純な精度向上ではなく個性や多様性を持たせることを目標とした研究が盛んにおこなわれるようになった。例として、個人の趣味嗜好に合わせて商品や情報を推薦するシステム [4] や、個人の振る舞いを再現することで一貫性のある発言を行う対話システム [17] などがあげられる。また個人を再現する研究は近年、デジタルツインコンピューティング [10] としても注目を集めている。デジタルツインコンピューティングでは、個人をデジタル空間上に再現することで、従来のリアル空間では実現することが難しかった、人の内的側面を反映させた大規模なシミュレーションや演算を可能にしたり、実際の人に代わってタスクを実行したりすることが期待される。そのためユーザの個人性を考慮した言語処理は今後様々な場面で重要性が増していくと考えられる。

デジタルツインのように個人の振る舞いそのものを再現するためには、個人の属性や性格などのあらゆることをモデル化する必要がある。しかし、既存研究では教師あり学習で深層学習モデルを学習するため、個人の属性情報を人手で選定、収集したり、性格等のラベルを人手で付与するなどのコストの高い作

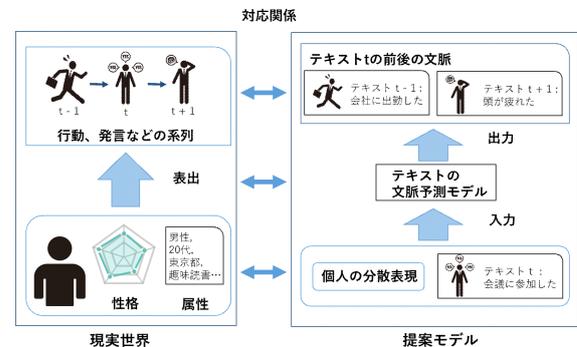


図 1 文脈生成による個人の埋め込み表現の獲得

業を必要とする。また個人の性格に応じた出力を得るためには、性格を何らかの基準でモデル化する必要があるが、性格や個性のモデルを定義するのは困難な課題だといえる。同時に、デジタルツインに様々なタスクを実行可能にするためには、個人のモデルは具体的なタスクを実行する後段の深層学習モデルとの親和性の高い表現であることが望ましい。

本論文では、深層学習モデルが個性を持った振る舞いを実行することを目的とした、個人の埋め込み表現の学習手法について提案する。埋め込み表現とは、単語などのシンボルを高次元の実数値ベクトルで表現したものである。学習した個人の埋め込み表現を深層学習モデルの入力として接続することで、言語処理をはじめとする様々な深層学習を用いたタスクにおいて、

個人の特徴や思考を反映させた出力結果を得ることを目指す。

個人の埋め込み表現を学習させるモデルとして、本論文では、テキストの文脈生成による学習モデルを提案する。文脈生成モデルは、自然文のテキストを入力とし、その入力テキストの前後の文を生成するように学習するモデルである。文脈生成と個人特徴の対応は図 1 に示した。個人の属性や性格は、個人の一連の振る舞いや発言の系列として外部に表出すると考えられる。本論文では、個人の埋め込み表現とある行動を入力とし、その前後の行動を予測するモデルを学習することで、個人の性格や属性が内包された埋め込み表現を得ることを目標とする。

本論文では学習用のデータとして Twitter を用いる。Twitter のテキストにはユーザの現在の思考や行動内容が時系列順に記述されるため、ユーザの属性や性格がツイートの文脈に表れていると解釈できる。換言すると、ユーザのツイートの文脈生成を模倣することは、ユーザ個人の属性や性格を理解することに相当すると考えられる。そこで、本論文では Twitter を対象に、入力ツイートの前後のツイートを生成させるようなモデルを、ユーザの埋め込み表現と同時に学習することで、ツイートに表出するユーザの属性や性格を埋め込み表現に内包するモデルを学習している。評価実験では、テキストから埋め込み表現を獲得する Doc2Vec と比較し、提案手法が単に単語の出現頻度等の情報を学習したのではなく、ユーザの性格的特徴など内的な情報も含んだ埋め込み表現が学習できることを示した。

本論文の貢献：本論文では、以下の貢献を果たした。

- ユーザの記述したテキストの文脈を予測することが、ユーザの属性や性格をモデル化することに相当すると解釈し、新しいユーザ個人特微量の教師なし学習手法を提案した。
- オープンなデータセットである Twitter のデータを用いて、提案モデルの学習を行った。提案モデルでは、Twitter ユーザの ID とツイートを入力として、入力ツイートの前後のツイートを生成するモデルを学習することにより、Twitter ユーザの埋め込み表現を得るモデルを学習した。
- 提案手法の埋め込み表現を入力として、ユーザの location 情報の分類精度評価することで、提案手法の持つユーザ固有情報の多寡を定量的に評価した。また提案手法の埋め込み表現と一般的な文書の埋め込み表現である Doc2Vec との比較を行い、提案手法の埋め込み表現が、ユーザの性格的特徴を反映した埋め込み表現になっていることを明らかにした。

## 2 提案手法

本節では、ユーザの埋め込み表現を学習する深層学習モデルおよび学習方法について詳述する。

### 2.1 文脈生成モデル

本論文で提案するユーザ埋め込み表現学習モデルは、文脈生成モデルをベースとする。文脈生成モデルを図 2 に示す。あるツイートを  $T_t$  とする。ここで、長さ  $l$  のワードピース系列で表現されるツイート  $T_t$  は、ワードピースの語彙数と一致する  $v$  次元の one-hot ベクトルで以下の通りに表す。

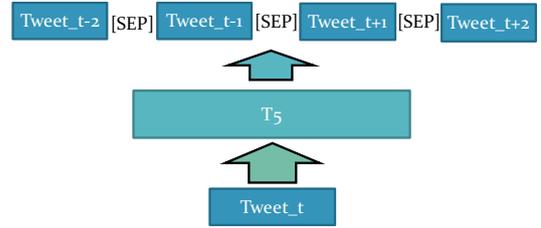


図 2 文脈生成モデル

$$T_t = \{w_1, w_2, \dots, w_l\} \quad (1)$$

ここで、 $w_1$  はツイート  $T_t$  の 1 番目のワードピースを表す。

入力ツイート  $T_t$  を生成型言語モデルである T5 [8] に入力する。T5 は Transformer [13] を採用する Encoder-Decoder 型の事前学習モデルであり、大量の言語コーパスに対して、BERT と同様に mask したワードピースを推定する事前学習を行う。T5 では mask 部分を含めて、教師データとなる文字列を言語生成によって復元する学習を行う。そのため、入力と出力のテキストのペアを用意することで、ファインチューニングが可能となる。

本論文では、ファインチューニングのための教師データとして、入力ツイートの前後のツイートを使用する。入力  $T_t$  に対して、生成モデルの出力を以下の教師データ  $S_t$  となるように学習を行う。

$$S_t = [T_{t-n}, w_{[SEP]}, T_{t-n+1}, w_{[SEP]}, \dots, w_{[SEP]}, T_{t+1}, w_{[SEP]}, \dots, T_{t+m}] \quad (2)$$

ここで、 $w_{[SEP]}$  は文の境界を示す特殊トークン [SEP] に対応する one-hot ベクトルである。 $n$  および  $m$  は自然数であり、 $T_{t-n}$  は入力ツイート  $T_t$  の  $n$  個前のツイート、 $T_{t+m}$  は  $m$  個後のツイートを示す。また、出力の教師データには  $T_t$  は含まない。

文脈生成モデルでは、入力ツイートの前後のツイートを生成するように学習することでツイートの文脈を学習する。ツイートにはユーザが単独で記入するテキストの他に、他ツイート等の引用（リツイート）や他ユーザとの会話（リプライ）が含まれるが、文脈生成モデルの学習においてはこれらのツイートを除外して学習を行う。

### 2.2 ユーザ埋め込み表現学習モデル

ユーザ埋め込み表現学習モデルを図 3 に示す。ユーザ埋め込み表現学習モデルは、文脈生成モデルにユーザ埋め込み層を追加したものである。

本モデルはあるユーザの  $\chi$  の一連のツイートを対象とする。各ユーザは、全ユーザ数に相当する  $u$  次元の one-hot ベクトルで表現される。次にベクトル変換行列  $W_e \in \mathbb{R}^{u \times d}$  を用いて、ベクトル  $E_\chi \in \mathbb{R}^d$  を得る。ここで、 $d$  は T5 の内部状態のベクトルの次元数を表している。この  $E_\chi$  をユーザの埋め込み表現として扱う。

ユーザ埋め込み表現のベクトル  $E_\chi$  を文脈生成モデルに結合して入力する。長さ  $L$  の入力ツイート  $T_{\chi_t}$  は、T5 の第 1 層で

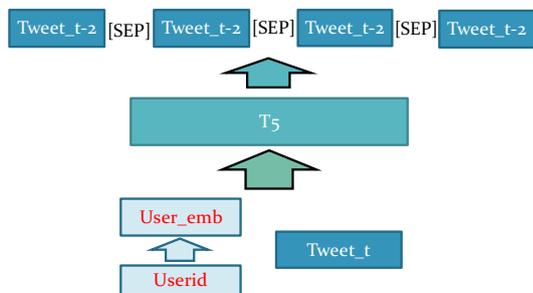


図 3 ユーザ埋め込み表現学習モデル

$d$ 次元の連続値の行列  $U_{T_{x_t}} \in \mathbb{R}^{d \times L}$  に変換される。ここで、ツイート行列  $U_{T_{x_t}}$  とユーザの埋め込み表現ベクトル  $E_x$  を連結し、以下の通りに T5 への入力行列  $U'_{T_{x_t}}$  を得る。

$$U'_{T_{x_t}} = [E_x; U_{T_{x_t}}] \in \mathbb{R}^{d \times L+1} \quad (3)$$

ここで、 $[:]$  は行方向への結合を表している。入力  $U'_{T_{x_t}}$  に対する出力  $S_{x_t}$  は文脈生成モデルと同様である。

文脈生成モデルにユーザ ID を加えて学習することで、生成される文脈がユーザ由来のものに変化する。そして、その際に使用されるユーザの埋め込み表現  $E_x$  はユーザ固有の情報を内包するベクトルとして学習されることになる。

### 3 評価実験

本節では、実データを使用した評価実験について述べる。まず、実験で使用するデータセットおよびモデル学習について説明し、次に学習した生成モデルおよびユーザ埋め込み表現の評価について詳述する。

#### 3.1 モデルの学習設定

本論文では、2015 年の 1 年分のツイートデータから学習用データセットを作成する。まず、Twitter の全ユーザの中から自動ツイート BOT や一日あたりのツイート数が少なすぎるユーザ等を除外した上でランダムに 1,000 ユーザをサンプルした。抽出した各ユーザの全ツイートから、ツイートとその文脈からなるデータ対を作成し、計 1,200,000 の学習データを用意した。

ベースモデルとして日本語 T5<sup>1</sup> を使用する。文脈生成で学習する文脈  $n$  および  $m$  は 2 と設定する。GPU (NVIDIA A6000) を 3 枚使用し、全学習データを 10 エポックで学習を行った。モデルの学習時間は 1 日 18 時間である。学習時の損失関数の値の推移を図 4 に示す。オレンジ色の線が文脈生成モデルを、青色の線がユーザ埋め込み表現学習モデルを表す。ユーザ埋め込み表現学習モデルは、文脈生成モデルより学習終了時の損失関数の値が小さい。ユーザ埋め込みと同時に学習することで、より実際の文脈に近い出力が得られるとわかる。

#### 3.2 ツイートの文脈生成

本項ではユーザ埋め込み表現学習モデルがユーザ固有の情報

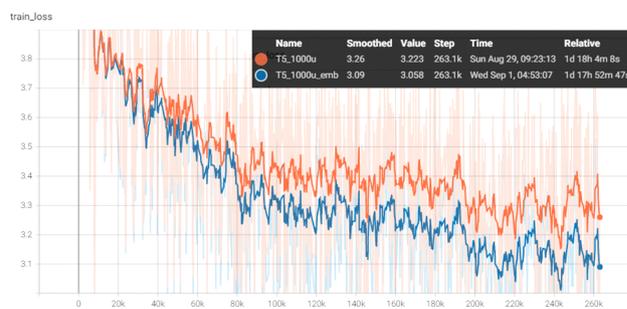


図 4 モデル学習中の損失関数の推移

を反映した文脈の生成が可能か確認した実験について述べる。始めにユーザの埋め込み表現がモデルの出力にユーザの情報を反映させるかを、特定のユーザに対する文脈生成結果を見ることで確かめた。次にユーザ埋め込み表現学習モデルが文脈生成モデルと比較して多様性のある出力結果になることを確認した。

始めにユーザ埋め込み表現学習モデルを用いて、ツイートの文脈生成を行った例を表 1 に示す。入力ツイート「山梨 or 東京は素晴らしい」という入力に対して、入力のユーザ ID を変更したときにモデルが生成した出力ツイートを示している。ここで、ユーザ A は Twitter の location に「東京都」と登録されているユーザであり、ユーザ B は location が「山梨県」のユーザである。

生成された文脈ツイートを見てみると、「おめでとうございます。」や「あ、そうか」など同じツイートが生成されている部分もあるが、一方で、「明日は山梨に行ってきます。」や「俺も甲府でバイトしているよ。」といった異なるツイートも生成されている。これは、ユーザの地理的特徴が出ているものであると考える。東京のユーザが「甲府でバイトしている」というツイートを行う確率はかなり低いといえる。また、「東京は素晴らしい」という入力ツイートに対して、t-1 で「東京の天気は雨です」と「東京の天気は雨か」というツイートが生成されているが、これはツイート内容はほぼ同じであるが、ユーザ自身の居住地によって主観と客観のニュアンスが変化している。

次に、文脈生成モデルとユーザ埋め込み表現学習モデルで生成される文脈の違いを確認する。各モデルに「紅白歌合戦」という入力を行った際の出力結果の例を表 2 に示す。文脈生成モデルはビームサーチ [14] により複数の文を生成し、ユーザ埋め込み表現学習モデルはランダムに選択したユーザ id に応じた文を 1 つずつ生成した。文脈生成モデルはビームサーチの性質上、文末の多少の違いを除いてほとんど同一の文章が生成される。文の内容を見ると、「fns 歌謡祭」、「μ's」などの個人の嗜好を表す単語が生成文のほとんどに含まれることから、特定のユーザに対する文脈を偏って学習していることがわかる。一方ユーザ埋め込み表現学習モデルは、生成される文脈に多様性がある。ユーザ 1、ユーザ 4 の生成結果を見ると、サッカーや人物の衣装など個人の嗜好を反映しながら、紅白に関連する文脈を生成している。ユーザ 2 は「紅白」という単語は生成されていないが、「大晦日」など紅白に関連する単語が生成されている。通常文脈生成モデルで探索幅を広げたビームサーチを

1: <https://github.com/sonoisa/t5-japanese>

表 1 ユーザによる生成ツイートとの違い

ユーザ (location)	入力ツイート	出力ツイート			
		t	t-2	t-1	t+1
A (東京都)	東京は素晴らしい	おめでとうございます。	東京の天気は雨です。	東京も雪が降っていました	今日は晴れて気持ちいいなあ～
	山梨は素晴らしい	あ、そうか	おっさんと山梨の県境で遭遇したんだけど(笑)	at 日本武道館	明日は山梨に行ってきます。
B (山梨県)	東京は素晴らしい	おめでとうございます。	東京の天気は雨か	今日も晴れてよかった	あ、そういえば昨日東京に行った時と全く同じだった。
	山梨は素晴らしい	あ、そうか	俺も甲府でバイトしているよ	やっぱり山梨県はいいね!	うわあ

表 2 紅白歌合戦と入力した際の生成ツイートの例

モデル	ユーザ	入力ツイート	出力ツイート			
			t	t-2	t-1	t+1
文脈生成モデル	なし	紅白歌合戦	今年もお世話になりました。来年もよろしく願います!	紅白はじまったー!!	紅白歌合戦の司会が神木隆之介でしたね～	fns 歌謡祭に μ's 出るのか
		紅白歌合戦	今年もお世話になりました。来年もよろしく願います!	紅白はじまったー!!	紅白歌合戦の司会が神木隆之介でしたね～	fns 歌謡祭に μ's 出るのかよ w
		紅白歌合戦	この時間にお腹空くのなぜだ	紅白で歌ってる人いるんだね～	紅白歌合戦に μ's 出るのか!楽しみ!!	fns 歌謡祭、今年も最高だったなー
		紅白歌合戦	この時間にお腹空くのなぜだ	紅白で歌ってる人いるんだね～	紅白歌合戦に μ's 出るのか!楽しみ!!	fns 歌謡祭、今年も最高だったなー。
ユーザ埋め込み表現学習モデル	1	紅白歌合戦	まあ、サッカーの話題で盛り上がるんじゃないんだけどね。	j リーグは j1 昇格プレーオフをどうするか? っていう話だよな。	紅白に期待するしかないか・・・	おやすみなさい
	2	紅白歌合戦	今日は寝る前に 1 日を無駄にした気がする	お腹すいたなー...	やっぱり俺の誕生日が大晦日だわ www	久しぶりにカラオケ行きたい
	3	紅白歌合戦	大晦日のエルロで一番好きな曲は?	紅白はじまったー!	紅白歌合戦お疲れ様でした。来年もよろしく願います m(..)m	今日から年末年始か
	4	紅白歌合戦	はあ～やっぱり紅白好きだな.....	おめでとうございます!!!	まどかちゃんの衣装可愛い.....	えっ、なんだよこれ

表 3 ランダム入力に対する生成文間の BERTScore

	文脈生成モデル	ユーザ埋め込み表現学習モデル
BERTScore	0.923	0.753

行うより、ユーザ埋め込み表現学習モデルでユーザを変更しながら文脈生成を行う方が、ユーザの特徴を反映しつつ多様性のある文脈が生成されることがわかる。

提案モデルがユーザ固有の多様な文脈を生成することを定量的に検証するため、各モデルでランダムに選んだ 100 ツイートに対して 10 個ずつ文脈を生成し、それらの一致度を BERTScore [18] の F1 値で計算し平均値を算出した。結果は表 3 の通りである。ユーザ埋め込み表現学習モデルは生成文間の一致度が低いことから、より多様な文脈生成を実現しているとわかる。これは提案学習モデルが一般的な文脈ではなくユーザ固有特徴に応じた多様な文脈生成が可能になったことを示している。

### 3.3 分類タスクによるユーザの埋め込み評価

本項以降では、ユーザの埋め込み表現学習モデルによって取得したユーザの埋め込み表現  $E_x$  の評価について述べる。比較手法に文書の埋め込み表現を学習する Doc2Vec [3] を採用する。各ユーザの一連のツイートを 1 つの文書とみなすことで Doc2Vec を適用する。Doc2Vec はユーザの埋め込み表現学習モデルを学習した 1000 ユーザのツイートを基に学習した。ここで、Doc2Vec の次元数は 200 に設定し、学習アルゴリズムは Distributed Bag of Words version of Paragraph Vector (PV-DBOW) を採用している。

#### 3.3.1 実験目的

本稿では提案手法によるユーザの埋め込み表現が、ユーザの属性に関する情報をどの程度内包しているか確認した実験について述べる。具体的には各ユーザの持つ location の情報について、ユーザの埋め込み表現からどの程度予測できるかを評価する。分類精度が高いほど、埋め込み表現が location 分類に有用な特

表 4 location 分類問題の精度 (%)

Input	Accuracy	Precision	Recall	F1-score
random	36.0 ± 5.5	32.8 ± 11.8	33.0 ± 5.6	30.2 ± 6.5
Doc2Vec	58.5 ± 4.1	58.2 ± 4.4	56.4 ± 4.4	56.2 ± 4.7
proposed	44.5 ± 6.0	43.4 ± 6.7	42.4 ± 5.7	41.9 ± 5.7
proposed (fine-tune)	<u>59.7 ± 3.9</u>	<u>59.6 ± 4.1</u>	<u>58.9 ± 4.0</u>	<u>58.6 ± 4.0</u>
fine-tune + Doc2Vec	<b>63.3 ± 3.4</b>	<b>63.8 ± 3.9</b>	<b>60.7 ± 3.5</b>	<b>60.7 ± 3.6</b>

徴を内包していることを示し、本論文の主張である文脈学習がユーザの属性情報を学習することに相当するという仮説を支持する結果となる。

### 3.3.2 実験設定

学習用データセットを作成するため、埋め込み学習済み 1000 ユーザの内、location 情報に“北海道”，“東京都”，“愛知県”を含むユーザを抽出し、location 予測用データセットを作成する。得られたユーザ数は北海道 100 人，愛知県 68 人，東京都 99 人の計 267 人で、これらを train data, validation data, test data に 4:1:1 に分割した。学習モデルは 2 層の全結合層を持つニューラルネットワークで、中間層の次元が 200 次元、活性化関数に ReLU を使用した。入力次元は提案手法は 768 次元で、Doc2Vec は 200 次元であり、提案手法側は過学習を防ぐためドロップ率 0.2 に設定した dropout 層を活性化関数の前に加えた。確率的勾配降下法を用いて、validation data に対する損失関数の値を 1 エポックごとに計算し、early stopping により学習を終了した。なお損失関数にはクロスエントロピー損失関数を用いた。

提案手法の埋め込みについては、

- (1) 文脈学習時に文脈生成の T5 とユーザ埋め込みのパラメタを同時に更新するモデル。
- (2) ユーザ埋め込みのパラメタを固定した状態で文脈学習後に、固定を外して fine tuning するモデル。

の二つの学習方法で得られた埋め込みを比較する。(1)の方法で学習した表現を提案手法埋め込み、(2)の方法で学習した表現を fine-tune 埋め込みと呼ぶ。また fine-tune 埋め込みと Doc2Vec 埋め込みを concat して入力した際の分類精度も評価した。

### 3.3.3 実験結果

location 分類問題の精度を表 4 に示す。表の値はモデルの学習をそれぞれ 20 回行った accuracy, precision, recall, f1-score の平均と標準偏差である。random はランダムに初期化した埋め込みを入力とした際の精度を意味する。random を全ての評価値で提案手法埋め込みと Doc2Vec 埋め込みが上回っていることから、各埋め込みがユーザの location 情報を保持していることが示された。fine-tune 埋め込みは全評価値で提案手法埋め込みの精度を大きく上回っている。この結果はユーザ埋め込みを固定して文脈の学習を終えた後に、ユーザ埋め込みの学習を開始した方が、location 情報に関する汎化性の高い情報が学習できることを示唆している。さらに、fine-tune + Doc2Vec

埋め込みは全ての評価値が最も高い。この結果は、従来の埋め込み表現と提案手法の埋め込みを組み合わせることで、他の深層学習モデルにおいても個人性を反映した出力の改善が期待できる事を傍証している。

## 3.4 ユーザの埋め込み表現ベクトル

学習した 1000 ユーザの埋め込み表現を k-means によって 5 つのクラスタに分類した。本項は、分類したクラスタの違いによって提案手法と Doc2Vec をクラスタ内ユーザのツイートに出現する単語、ツイートから受けるユーザに対する印象という二つの観点から比較する実験を行う。

### 3.4.1 クラスタ毎の出現単語の比較実験

#### a) 実験目的

Doc2Vec は単語の出現頻度の情報から文書の特徴づける。一方で提案手法は入力前後の文脈を生成するタスクから得られるため、単語の出現頻度だけでなく、様々な情報を内包していると考えられる。本実験では各埋め込み表現をクラスタリングして得られたクラスタ内ツイートを分析して、提案手法のクラスタが単純に頻出単語の違いで分類されていないことを確かめる。

#### b) 実験設定

提案手法で学習した 1000 ユーザの埋め込み表現と Doc2Vec で得られたユーザ表現をそれぞれ k-means によって 5 つのクラスタに分類し、クラスタごとにワードクラウドによる可視化を行った。次に、クラスタごとに頻出する上位 20 単語を選び 5 つのクラスタで合わせて、Doc2Vec と提案手法における単語の頻度ベクトルを作成し、各クラスタの頻度ベクトル間の相関係数の大きさを比較する検証を行った。単語の頻度ベクトルの各要素はクラスタ内全ユーザの全 Tweet 中に出現する語彙の回数である。5 つのクラスタの持つ単語の頻度ベクトルにおいて、自分同士を除く  $5 \times 4$  の組み合わせについて相関係数を求め、その平均を計算した。

#### c) 実験結果

Doc2Vec によるクラスタのワードクラウドを図 5、提案手法のクラスタのワードクラウドを図 6 に示す。Doc2Vec では、各クラスタごとに「日本」、「ゲーム」といった特徴的な単語を取得できている。一方で提案手法では、「ゲーム」、「仕事」といった単語がどのクラスタにも高頻度で出現している。

各クラスタの単語の頻度ベクトル間の相関係数を求めた結果は表 5 の通りである。提案手法は、相関係数が Doc2Vec より



図5 Doc2Vecによるクラスタのワードクラウド



図6 ユーザ埋め込み表現モデルによるクラスタのワードクラウド

表5 類出単語ベクトル間の相関係数

	Doc2Vec	提案手法埋め込み
相関係数	0.151	0.234

大きい。従って、提案手法によるクラスタに対しては、単語の出現頻度の寄与度がDoc2Vecと比較して低いことが示された。

### 3.4.2 クラスタ内ツイートの主観評価実験

#### a) 実験目的

提案手法はlocationなどユーザの属性情報を持ち、Doc2Vecと比較し、単語単位のトピック情報の寄与度が低いと考えられる。本実験ではさらに提案手法がユーザの性格など内面的特徴を学習していることを示すため、アンケートによる主観評価を行った。

#### b) 実験設定

まず、各クラスタからランダムに10ユーザを抽出する。次に抽出した各ユーザの全ツイートから、連続する10ツイートをランダムに3か所、合計で30ツイートを取得する。評価者は、取得した30ツイートを閲覧し、ツイートから受けるユーザの性格的特徴を評価する。本論文では2名の評価者で評価を行った。

本論文では、性格的特徴としてBIG5[20]を採用した。BIG5はヒトの性格を、外向性、協調性、誠実性、神経症傾向(論理性)、開放性(創造性)の5項目で定義したものである。各項目は高低で表わされ、外向性であれば「外交的(高)or内向的(低)」のどちらかに決定する。実験では、閲覧したツイートから受けるユーザの印象を、高低の2項の選択問題として選択することで評価を行った。

#### c) 実験結果

クラスタ内のツイートの印象評価を行った結果を図7, 8に示す。各グラフはクラスタごとのBIG5の評価値の平均値であり、性格的特徴の偏りを示している。上下どちらかに棒グラフが伸びている項目はクラスタ内で、どちらか一方の心理的特徴を持つユーザが多く存在していることを示しており、棒グラフが伸びていない項目については、クラスタ内で性格的特徴がどちらも偏りなく存在していることを表している。上段の5つのグラフがDoc2Vecによる5クラスタの結果であり、下段のグラフが提案手法のクラスタの結果である。

Doc2Vecによるクラスタの評価では、どのクラスタの分布も似通った形状をしている。また全クラスタで論理性/情動性の項目が情動性側に偏っていることから、Doc2Vecの場合どのクラスタにおいても情動的なユーザが多く存在することを示している。一方で、提案手法によるクラスタの場合、クラスタの分布形状が大きく異なっている。論理性/情動性の項目ではクラスタ1, 5のユーザは情動的だが、クラスタ2, 3, 4のユーザは論理的だと判定されている。この結果から提案手法によるクラスタに対するユーザの性格的特徴の寄与度が高いと考えられる。

分布形状の類似性を評価するため、各クラスタの分布間相関係数の平均を計算し、Doc2Vecと提案手法で比較する検証を行った。5つのクラスタの持つBIG5評価結果のベクトルにおいて、自分同士を除く5×4の組み合わせについて相関係数を求め、その平均を計算した。結果は表6の通りである。Doc2Vecはクラスタ毎の分布間に相関が見られるが、提案手法には相関が見られない。この結果は提案手法によるクラスタに対して、ユーザの性格的特徴の寄与度が高いことを示している。

## 4 考察

### 4.1 文脈生成モデルについて

ユーザの埋め込み表現を獲得するための文脈生成モデルでは、ユーザIDを変更することで、同じ入力ツイートであっても、異なる出力を得ることができることを明らかにした。

ユーザIDを変更した場合でも同じツイートが生成されるパターンが存在しているが、ユーザの影響を多く受けたツイートは入力の前、後どちらでも生成されているため、前後の文脈どちらも学習することに関する効果が出ているのではないかと考えられる。また、「東京の天気は雨です」と「東京の天気は雨か」のように同じ入力ツイートに対して、ほぼ同じ内容でありながら、詳細なニュアンスを変えている点は、事前学習モデルの表現能力の高さに加え、ユーザ埋め込み表現に含まれる情報の寄与度が高いと考えられる。

### 4.2 埋め込みによる分類精度について

ユーザ埋め込みを入力としてユーザの居住地を予測する分類問題の精度評価実験では、提案手法の埋め込み表現が居住地分類に必要な情報を保持しており、かつDoc2Vecによる埋め込み表現と組み合わせて精度が向上することが示された。

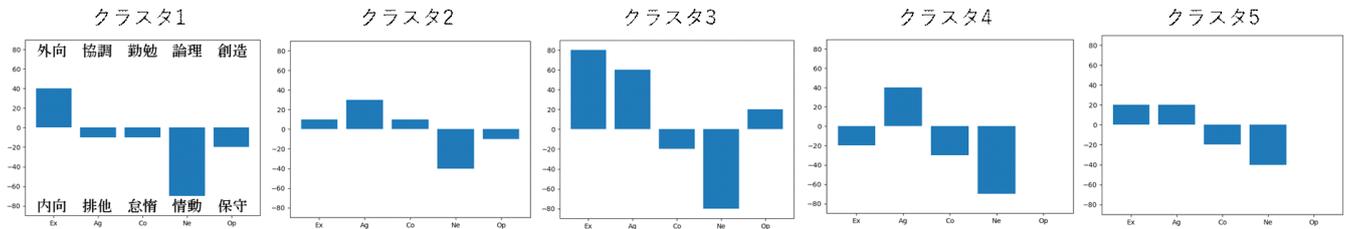


図 7 Doc2Vec 埋め込みによるクラスタ内ツイートの BIG5 評価結果

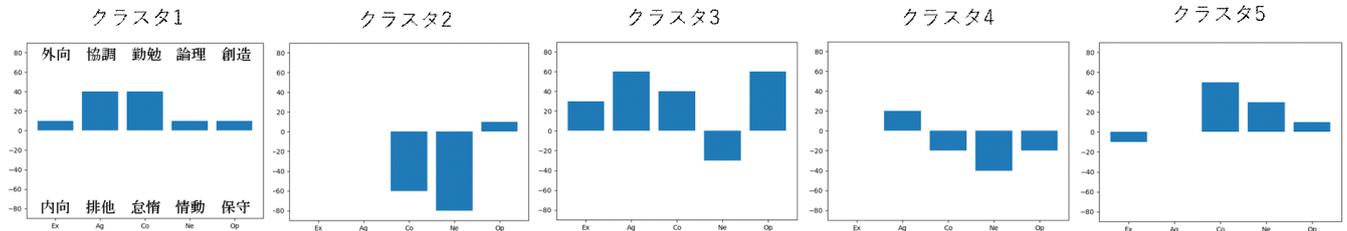


図 8 ユーザ埋め込み表現モデルによるクラスタ内ツイートの BIG5 評価結果

表 6 BIG5 評価分布間の相関係数		
	Doc2Vec	提案手法埋め込み
相関係数	0.787	0.156

Doc2Vec による埋め込みは単語の出現頻度から分かるトピックの情報を持つため、ツイート内に出現する地名や建物名から居住地を予測したと考えられる。ただし、ユーザ文脈生成の実験でも確認した通り、同じ「東京」という単語が出現する文であっても、「東京に行きたい」と「こちら東京の天気は雨です」という文では、ユーザの居住地の予測という点で文に大きな違いがある。前の文はユーザが東京に住んでいないことを、後の文は東京に住んでいることが予想される。文脈生成タスクを行う提案手法はこうした文のニュアンスによる違いを認識するため、Doc2Vec と合わせることでさらに精度を改善することができたと考えられる。

#### 4.3 ユーザ埋め込み表現について

ユーザの埋め込み表現に関する評価では、比較手法の Doc2Vec と比較して、単語やトピックに関する情報より、ユーザ自身の性格的特徴を強く反映している可能性が示唆された。この結果はテキストの文脈を学習するというタスクが、ユーザの属性や性格を学習することに相当するという仮説の根拠となる。

Doc2Vec が PV-DBOW でユーザ ID に対する出現単語を推定する問題であるのに対して、提案手法では、ツイートという文単位で周辺を推定する問題として解いている。そのため、単語単位の情報が埋め込み表現にあまり含まれなかったのではないかと考えられる。また、提案手法で作成したクラスタに性格的特徴の偏りが多く表出したが、これは文脈生成モデルにより、ツイートの文脈を学習した影響であると考えられる。人の性格や思考は、単独のテキストではなく、文脈に対してどのような返答や振る舞いをするのかに表出されると考えられる。例えば、「お腹がすいた」というツイートの後に、「料理をする」

人がいたり、「外食に行く」人、「文句を言う」人など様々なパターンが考えられる。「料理をする」人も「お腹がすいた」から料理をする人もいれば、「料理教室に行った」から料理をする人もいる。このように単独のツイートやテキストだけで見えてこないことであっても、前後の文脈と組み合わせることで、人の属性や考え方を垣間見ることができ、その結果、提案手法で居住地や性格的特徴を反映した評価結果が出たと考えられる。

## 5 関連研究

ユーザの埋め込み表現を用いる研究として、レコメンドに関する研究が挙げられる。Pal [6] らは、Web サイト上の様々なコンテンツのレコメンドを統合するために、異なるコンテンツの埋め込み表現をユーザの行動予測モデルに組み込むことで、ユーザの趣向を反映した埋め込み表現手法を提案している。Wang [15] らはショッピングサイトにおいて、ユーザのクリック情報やクエリログといった一連の行動を RNN を用いて埋め込み表現に変換して、商品のレコメンドを行う手法を提案している。Shimei ら [7] はソーシャルメディアを用いて埋め込み表現を作成する手法についてまとめている。

個人の振る舞いを再現する研究としては、対話における個人の再現に関する研究がある。Zhang ら [17] は、仮想の人物像であるペルソナを定義する複数文のテキストをもとに対話を行う persona-chat を提唱している。Liu ら [5] は、persona-chat において、対話相手のペルソナ像を予測しながら対話することで、話者間の共通知識を形成する手法を提案している。Song ら [11] らは、よりペルソナ像に対して一貫した対話を行うために、複数の BERT を用いてペルソナ文と発話文のマッチングをとる手法を提案している。Higashinaka ら [2] は、有名人物やキャラクタなどのペルソナ像を持つ対話システムを構築するために、多くの参加者のロールプレイによって、学習用データを収集する手法を提案している。

本研究は上述した研究のようにタスクを推薦や対話等に限定

しておらず、location 分類問題で示したように、モデルの初期値や入力に埋め込み表現を用いることで様々なダウンストリームタスクの精度改善が期待される。また本研究は Zhang ら [17] や Liu ら [5] などの研究と異なり、個人性情報抽出のために人手によるデータセットの作成やラベリング等の高コストな作業を必要としない。

## 6 おわりに

本論文では、深層学習モデルの個人性獲得のためのユーザ埋め込み表現を、ツイートの文脈生成モデルを通して教師なしで学習する手法を提案した。

文脈生成モデルは、入力ツイートの前後のツイートを生成するモデルで、そこにユーザの埋め込み表現となるベクトルを連結して学習することで、ユーザの埋め込み表現を獲得できる。

評価実験では、提案手法で学習したユーザの埋め込み表現は単語単位のトピック情報よりも、ユーザの居住地等の属性の他、性格的特徴が反映されていることを示唆する結果を得ることができた。さらに提案した埋め込み表現は、従来の Doc2Vec などの埋め込み表現と組み合わせることで、より精度の高いユーザ特徴を表現し、ダウンストリームタスクにおける精度改善が期待できることを示した。

今後は、ユーザの埋め込み表現の更なる表現力の向上および、学習の効率化を進め、対話等の応用タスクに適用することを検討している。

## 文 献

- [1] J. Devlin, M. Chang, K. Lee, and K. Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, and T. Solorio eds., Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies NAACL-HLT, pp. 4171–4186. Association for Computational Linguistics, 2019.
- [2] R. Higashinaka, M. Mizukami, H. Kawabata, E. Yamaguchi, N. Adachi, and J. Tomita. Role play-based question-answering by real users for building chatbots with consistent personalities. In Proceedings of the 19th Annual SIGDial Meeting on Discourse and Dialogue, pp. 264–272, 2018.
- [3] Q. Le and T. Mikolov. Distributed representations of sentences and documents. In Proceedings of the 31st International Conference on Machine Learning, Vol. 32 of Proceedings of Machine Learning Research, pp. 1188–1196, 2014.
- [4] L. Li, Y. Zhang, and L. Chen. Personalized transformer for explainable recommendation. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL2021), pp. 4947–4957, 2021.
- [5] Q. Liu, Y. Chen, B. Chen, J.-G. Lou, Z. Chen, B. Zhou, and D. Zhang. You impress me: Dialogue generation via mutual persona perception. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics(ACL2020), pp. 1417–1427, 2020.
- [6] A. Pal, C. Eksombatchai, Y. Zhou, B. Zhao, C. Rosenberg, and J. Leskovec. Pinnerage: Multi-modal user embedding framework for recommendations at pinterest. Proceedings

- of the 26th ACM SIGKDD International Conference on Knowledge Discovery Data Mining(KDD2020), pp. 2311–2320, 2020.
- [7] S. Pan and T. Ding. Social media-based user embedding: A literature review. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence(IJCAI201).
- [8] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140):1–67, 2020.
- [9] T. Rohde, X. Wu, and Y. Liu. Hierarchical learning for generation with long source sequences, 2021.
- [10] W. Shengli. Is human digital twin possible? Computer Methods and Programs in Biomedicine Update, 1:100014, 2021.
- [11] H. Song, Y. Wang, K. Zhang, W.-N. Zhang, and T. Liu. BoB: BERT over BERT for training persona-based dialogue models from limited personalized data. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL2021), pp. 167–177, 2021.
- [12] H. Sugiyama, M. Mizukami, T. Arimoto, H. Narimatsu, Y. Chiba, H. Nakajima, and T. Meguro. Empirical analysis of training strategies of transformer-based japanese chat systems, 2021.
- [13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin. Attention is all you need. In Advances in Neural Information Processing Systems, Vol. 30, 2017.
- [14] A. K. Vijayakumar, M. Cogswell, R. R. Selvaraju, Q. Sun, S. Lee, D. J. Crandall, and D. Batra. Diverse beam search for improved description of complex scenes. In Proceedings of the AAAI Conference on Artificial Intelligence(AAAI2018), pp. 7371–7379, 2018.
- [15] T. Wang, Y. M. Brovman, and S. Madhvanath. Personalized embedding-based e-commerce recommendations at ebay, 2021.
- [16] H. Xu, B. V. Durme, and K. W. Murray. Bert, mbert, or bibert? A study on contextualized embeddings for neural machine translation. In M. Moens, X. Huang, L. Specia, and S. W. Yih eds., Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP, pp. 6663–6675. Association for Computational Linguistics, 2021.
- [17] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston. Personalizing dialogue agents: I have a dog, do you have pets too? In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL2018), pp. 2204–2213, 2018.
- [18] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi. Bertscore: Evaluating text generation with BERT. In Proceedings of the 8th International Conference on Learning Representations(ICLR2020). OpenReview.net, 2020.
- [19] Z. Zhang, J. Yang, and H. Zhao. Retrospective reader for machine reading comprehension. In Proceedings of the AAAI Conference on Artificial Intelligence(AAAI2021), pp. 14506–14514, 2021.
- [20] 村上. 日本語におけるビッグ・ファイブとその心理測定的条件. 性格心理学研究, 11(2):70–85, 2003.