DEIM2021 I25-5

Twitterユーザに対するゼロショットタグ付け

新田　洸平†　　加藤　　誠††

† 筑波大学 知識情報・図書館学類　〒 305–8550 茨城県つくば市春日 1-2
†† 筑波大学 図書館情報メディア系　〒 305–8550 茨城県つくば市春日 1-2
E-mail: †s1913576@s.tsukuba.ac.jp, ††mpkato@acm.org

あらまし　本稿では，SNS ユーザに対してユーザの性質を表すタグ付けを行う手法を提案する．ユーザに対するタグ付けは，各タグに対する学習データを用意することで機械学習手法やルールベースの手法によって実現できる．一方で，学習データに出現しないユーザが推論時に出現した場合に適切な予測ができないため，学習データがない場合でもユーザに対してタグ付けする手法を提案する．提案手法では，ゼロショット学習と呼ばれる機械学習手法を応用し，タグとユーザの単語埋め込み空間上の対応関係を学習することで，学習時に出現しないタグをもユーザに対して付与する．実験では，Twitter から収集したリストの名前とユーザツイートからデータセットを作成し，単語のマッチングによる手法と提案手法の有効性について評価を行った．実験の結果，Hit@10 において単語のマッチングによる手法，nDCG@10 において単語のマッチングによる手法と提案手法を組み合わせた手法が有効であることが明らかとなった．

キーワード　ゼロショット学習, ユーザタグ付け, Twitter

1　はじめに

ソーシャルネットワーキングサービスにおいてユーザの性質を明らかにすることは重要である．ユーザの性質が分かることで有益な判断が容易になる．例えば，ユーザの性質からそのユーザが発信する情報の傾向がわかることで，情報源として有効かどうかの判断が容易になる．

ユーザの性質を明らかにするような研究として，属性に基づくユーザ分類とユーザに対するタグ付けに分かれる．属性に基づくユーザ分類は，クラスがあらかじめ決められているような問題設定である．例えば，機械学習手法を用いてユーザを年齢や性別，政治的思考などで分けるような研究が行われている [1]．ユーザに対するタグ付けは，クラスが明示的に決められていないような問題設定である．例えば，ユーザのプロフィールや投稿などの関連情報に含まれる名詞の単語をタグ付けするような研究が行われている [2]．

既存手法におけるいくつかの限界を示す．まず，既存手法では，具体的な性質の表現が難しくタグの表現が単語に限定されている．次に，既存手法では，教師あり機械学習手法を用いる場合，学習データ中に含まれていないパターンのユーザを正しく予測できない．これはゼロショット学習問題として知られる．また，機械学習手法を用いる場合，学習データの作成に大きなコストがかかる点である．データセットの作成には専門的な知識が必要かつ学習にはデータが大量に必要となり容易ではない．

本論文では，ゼロショット学習手法を用いてユーザとタグとの対応関係を学習することでユーザにタグ付けする手法を提案する．ゼロショット学習手法は，機械学習手法の一手法であり，学習データ中に出現しないパターンのデータであっても予測を行える．タグの表現は，タグに含まれる全ての単語の埋め込みの平均を使用するため単語に限定されない．

提案手法の詳細について述べる．まず，学習において使用できるデータを収集する．Twitter からユーザを収集し，収集したユーザのツイート，ユーザが含まれているリストの情報を収集する．ツイートは 1 ユーザ当たり 1,000 件収集する．リストとは，ユーザによって他のユーザをグループ化するアノテーションのような機能である．次に，タグの単語埋め込みを平均したベクトル V_t とユーザツイートの単語埋め込みを平均したベクトル V_u を得る．V_t と V_u を結合してベクトル x を得る．そして，x とラベル y を全結合のニューラルネットワークに入力して適合するか否かを分類してモデルを学習する．学習したモデルによって，あるユーザ u とタグ集合 T の各タグ t との適合度を算出し，タグを適合度で順位付けすることで，ユーザに対してタグ付けを行う．

実験では，あるユーザとユーザが含まれるリスト名から作成したタグを正解とラベル付けしたペアを 1 件，同じユーザに対してユーザが含まれないリスト名から作成したタグを不正解とラベル付けしたペアを 99 件用意し，合わせて 100 件のユーザとリストのペアを 1 セットとして，収集したデータより 1,000 セットを作成した．1,000 セットの中から 800 セットを学習に用いて分類問題を解くことでモデルの学習を行った．テストデータ 200 セットに対して，学習したモデルを使用して 1 件のユーザと 100 件のリストの適合度を算出した．算出した適合度を用いてタグの順位付けを行った．順位付けされた結果を Hit@k, nDCG@k を用いて上位 k 件に正解タグが含まれるかどうかで評価を行った．

本論文における貢献は次の通りである：(1) SNS のユーザ分析においてゼロショット学習問題に取り組んだ．(2) ゼロショット学習手法を用いてタグとユーザの対応関係を学習することで学習データ中に含まれないパターンのユーザであっても予測を行いタグ付けできる手法を提案した．(3) Twitter データに対する実験を行い，単語のマッチングによる手法と提案手法の有

効性について評価を行った．実験の結果，単語のマッチングによる手法が有効であることを明らかにした．

本論文の構成は以下の通りである．2 節ではソーシャルネットワークにおけるユーザの属性による分類に関する関連研究，ソーシャルネットワークにおけるユーザのタグ付けに関する関連研究，および，ゼロショット学習手法に関する関連研究について述べる．3 節では問題設定を説明し，ゼロショット学習，および，その主問題への適用方法について述べる．4 節では実験結果を示す．最後に，5 節では今後の課題と共に本論文の結論を述べる．

2 関 連 研 究

本節では，ソーシャルネットワークにおけるユーザの属性による分類に関する既存研究，ソーシャルネットワークにおけるユーザのタグ付けに関する既存研究，および，ゼロショット学習手法に関する既存研究について述べる．

2.1 ソーシャルネットワークにおける属性に基づくユーザ分類

これまで行われてきたソーシャルネットワークにおけるユーザの属性による分類に関する既存研究について述べる．Rao らは，Twitter におけるユーザの性別，年齢，地域，政治志向の 4 つの属性の分類を行った．各属性を 2 つのクラス（例えば，性別であれば男性か女性か，年齢であれば 30 歳未満か 30 歳以上かなど）に分け，2 値分類を用いることで，ユーザの分類を行った．データセットはクローリングと 2 人のアノテータにより手動で構築されている [1]．Pennacchiotti と Popescu は，ユーザのプロフィール，ツイート，使用言語，ネットワークから特徴量を抽出し，トピックモデルを用いることで，政治志向（民主党か共和党か），民族性（アフリカ系アメリカ人かそれ以外か），特定のビジネスとの親和性（コーヒーチェーン店 Starbucks との親和性があるか否か）を分類している [3]．

これらの研究は，教師あり学習の場合，あらかじめクラスが決められている一方で，より多くのクラスを表現したい場合，クラス数に応じた十分な量のラベル付き事例を用意する必要があり，大きなコストがかかるため現実的ではない．つまり，学習データとして用意できたクラス数に表現が限定される．教師なし学習の場合，学習データの事例にラベルは必要ないがいくつかの欠点がある．1 つは，教師あり学習同様に学習データが大量に必要なことである．1 つは，推論時に学習モデルを用いて分類するクラス数の決定と予測されたクラスに対してトピックのラベルを決定することが容易ではないことである．

2.2 ソーシャルネットワークにおけるユーザのタグ付け

これまで行われてきたソーシャルネットワークにおけるユーザのタグ付けに関する既存研究について述べる．前節では，ソーシャルネットワークにおける属性に基づくユーザ分類に関する既存研究について述べた．属性に基づくユーザ分類において，ユーザの分類に使用する情報は基本的にツイートやプロフィールといったユーザ自身によって発信される情報である．ユーザの興味関心は，ツイートやプロフィールに現れる明示的な情報だけでは十分ではない．そこで，ユーザに対してアノテーションされた情報を活用する．例えば，Tiwtter のリスト機能 [1]である．リスト機能とは，ユーザによって他の複数のユーザをグループ化して名前をつける機能である．これによって，ユーザは共通項を持つ他のユーザを管理し，特定のトピックに関する情報を管理，閲覧しやすくなる．リスト機能は，ユーザに対するアノテーションのような情報であり，ツイートやプロフィールといったユーザが自発的に発信する情報だけでは得られなかった情報が得られるようになる．Kim らは，情報源として Twitter のリストに基づいてツイートを分析してユーザの潜在的な性質や興味を発見するための手法を提案した [4]．リストは複数のユーザによって構成されており，それらのユーザが投稿するツイートから抽出された単語は，リストに含まれるユーザが使用していない単語であっても，リストに含まれる全てのユーザを代表することを明らかにした．Yamaguchi らは，リスト名に含まれる単語をタグとして，ユーザにタグ付けを行う手法を提案した [2]．Yamaguchi ら以前の既存手法では，リストに基づいた分析をしていたとしてもツイートの情報のみを利用して分析していた．Yamaguchi らの手法では，リスト名に含まれる単語に注目してタグ付けを行っている．Sharma らは，リストのメタデータでリスト名とリストの説明において頻出する単語を抽出し，抽出した単語をリストに含まれるユーザに対して付与する手法を提案した [5]．リストのメタデータであるリスト名とリストの説明は，リストに含まれているユーザに対して他のユーザがどのように認識しているかを知る手がかりとして有効であり，リストのメタデータを分析することでユーザの専門的なトピックを予測できることを明らかにした．

これらの研究には，いくつかの限界がある．まず，データセットに含まれている種類のユーザにしか焦点を当てていない．Sharma の研究で使用されている Cha らの研究によるデータセットでは，収集した 5,400 万ユーザのうち 6,843,466 人が少なくとも一度リストに含まれている [6]．つまり，全体の約 88% のユーザはリストに含まれていないことがわかっている．またで，Yamaguchi らの研究では，収集した 103,127 ユーザのうち 78,098 人が少なくとも一度リストに含まれている [2]．つまり，全体の約 75%のユーザがリストに含まれていることがわかっている．一方で，TwitterJapan のツイート [2]より日本国内におけるユーザ数が 4,500 万ユーザを超えている（2017 年 10 月 27 日当時）ことから，Yamaguchi らが収集したデータは実際のデータのごく一部であり，前述の Cha らの研究によるデータセットから考えると，実際には国内のユーザにおいてもリストに含まれていないユーザが多い可能性が高い．一部のユーザデータから学習データを作成してモデルの学習を行った場合，学習データに含まれないユーザのパターンが推論時に出現した場合，正しく予測できない．一方で，より多くのパターンを網羅した学習データの作成には，前述の通り大きなコストがかかるため，現実的ではない．

1：https://help.twitter.com/en/using-twitter/twitter-lists

2：https://twitter.com/TwitterJP/status/923671036758958080

2.3 ゼロショット学習

これまで行われてきたゼロショット学習に関する既存研究について述べる．ソーシャルネットワークにおける属性に基づくユーザ分類の研究，ソーシャルネットワークにおけるユーザのタグ付けの研究において次のような問題があることがわかった：(1) 学習データが大量に必要なこと，(2) 教師あり学習手法を用いる場合，学習データとして用意できたクラスに表現が限定されること，(3) 教師なし学習手法を用いる場合，学習モデルを用いて分類するクラス数の決定と予測されたクラスに対してトピックのラベルを決定することが容易ではないこと．

これらは，ゼロショット学習問題として知られる．ゼロショット学習とは，機械学習手法の一手法である．学習データ中に一度も出現しないパターンを予測する手法である．近年，コンピュータビジョン分野においては広く使用されている [7], [8], [9], [10], [11], [12]．また，自然言語処理分野においても注目を集めてきている．ゼロショット学習の目的は，Wang らの調査研究 [13] にもあるように，学習において既知のクラス $\mathcal{S}$ に属するラベル付き学習事例 D^{tr} が与えられた時，未知のクラス $\mathcal{U}$ に属するテスト事例 X^{te} を分類できる（すなわち，テスト事例のラベル Y^{te} を予測する）分類器 $f(\cdot) : \mathcal{X} \rightarrow \mathcal{U}$ を学習することである．ゼロショット学習の最初の提案は，Larochelle らによる研究である [14]．Larochelle らによる研究では，一部のクラスまたはタスクで利用できるトレーニングデータがなく，これらのクラスまたはタスクの説明のみが示されている問題を提案している．

テキスト分類におけるゼロショット学習手法について述べる．テキスト分類におけるゼロショット学習問題を最も早く提案した研究の 1 つが Pushp と Srivastava による研究 [15] である．Pushp と Srivastava は，マルチラベル文書分類をラベルごとの 2 値分類問題として扱い，ラベルと文がそれぞれ適合しているかどうかを全結合型ニューラルネットワークによって予測する複数のモデルを提案した．全結合型ニューラルネットワークの入力には，ラベルの単語埋め込み，文の単語埋め込みの平均を用意し，それらを結合したベクトルを用いている．Pushp と Srivastava の研究において，学習ではウェブ上の文書に付与された SEO タグをラベル，見出しを文として学習を行っている．予測ではニュース記事におけるカテゴリに対する文の分類，マイクロブログデータにおけるカテゴリに対する文の分類を行っている．

既存手法と本研究との違いについて述べる．既存手法では，ゼロショット学習モデルの学習と学習したモデルを用いた予測において，使用するラベルが単語に限定されているが，本研究におけるラベルは複数の単語からなる短い文となっている．また，本研究においては，学習時においてはラベルに対するユーザの分類をマルチラベルタスクとしてモデルの学習を行うが，予測時にはユーザをクエリ，タグを文書とした文書検索タスクとして扱い，ユーザに対する文の適合性判定を行う．

3 提 案 手 法

本節では，タグとユーザツイートの単語埋め込み空間上の対応関係を学習することで，学習データがない場合でもユーザに対してタグ付けを行う深層学習を用いで適合性判定モデルについて述べる．

3.1 問 題 設 定

本研究で扱う問題について定義する．本研究では，ユーザに対してタグの適合度順に順位付けを行う．SNS におけるユーザの集合を文書検索タスクにおけるクエリ集合とみなし Q とする．また，各ユーザ $q \in Q$ は比較的短い文の集合 $\{s_1, s_2, ..., s_n\}$ から構成される．ユーザに対して付与したいタグ集合を文書検索タスクにおける文書集合とみなし D とする．このとき，入力として与えられた各ユーザ $q \in Q$ に対する各タグ $d \in D$ の適合度を予測し，予測した適合度よりタグを順位付けすることで，ユーザに対してどのタグが適合しているかを判定することが本研究で扱う問題である．

学習時においては，2 値分類タスクとして学習を行う．ユーザ q とタグ d のペア (x_q, x_d) と，タグとユーザのペアに対して与えられた適合性ラベル y を用いて，ユーザがタグに適合するか否かという予測結果と正解の誤差が小さくなるようにモデルの学習を行う．推論時においては，学習したモデルを用いて，ユーザ $q \in Q$ と $d \in D$ タグとの適合度を得る．得られたユーザに対するタグの適合度によってタグを順位付けする．順位付けされたタグ集合の中でより上位に位置するタグは，ユーザとより適合していると仮定する．

3.2 ベースライン手法

ベースライン手法について述べる．本研究では，ベースライン手法として Robertson らによる Okapi BM25 [16] を用いる．Okapi BM25 は，情報検索の最も代表的な手法の 1 つである．Okapi は Robertson らによって開発されたシステムの名前であり，BM は Best Match の略である．BM25 の他に，BM1, BM11，BM15 などがあるが，これらは重み付け関数の違いによって名称が異なる．本研究では BM25 を用いる．BM25 は，文書検索タスクにおいてクエリと文書を入力としたとき，文書におけるクエリの単語の重要度を出力することで，クエリと文書の適合性を判定する．BM25 は次のような特徴がある：(1) クエリに含まれる単語の文書中の出現頻度が多ければ重要が高くなる，(2) クエリに含まれる単語の文書集合における出現頻度が低く文書集合において稀な単語であれば重要が高くなる，(3) 文書集合において文書の長さがより少ない文書に出現する単語であれば重要が高くなる．本研究では，3.1 節でも述べた通り文書検索タスクにおけるクエリをユーザ，文書をタグと捉える．ユーザ $q \in Q$ のタグ集合 D の各タグ d における BM25 スコアの計算式を式 1 に示す：

$$f_{\mathrm{BM25}}(q, d) = \sum_{t \in q} f_{\mathrm{IDF}}(t_i) \cdot f_{\mathrm{TF_{BM25}}}(t_i, d), \tag{1}$$

$$f_{\mathrm{IDF}}(t) = log \frac{1 + |D|}{1 + df(t)} + 1, \tag{2}$$

$$f_{\mathrm{TF_{BM25}}}(t,d) = \frac{f_{t,d}\cdot(k_1+1)}{f_{t,d}+k_1\cdot(1-b)+b\cdot(\frac{l_d}{l_{avg}})} \tag{3}$$

ここで，入力として与える q はユーザ，d はタグである．t_i はクエリ中に含まれる i 番目の単語である．式 2 に示す $f_{\mathrm{IDF}}(t_i)$ は，ユーザの持つ i 番目の単語 t_i のタグ集合 D におけるタグ頻度の逆数である．式 3 に示す $f_{\mathrm{TF_{BM25}}}(t_i,d)$ は，タグ d におけるユーザの持つ i 番目の単語 t_i の単語頻度による特徴量である．$f_{t,d}$ は，タグ d におけるユーザの持つ i 番目の単語 t_i の単純な出現頻度，l_d はタグ d の単語単位の長さ，l_{avg} はタグ集合 D における各タグの単語単位の長さの平均，$k1, b$ はぞれぞれパラメータである．$k1$ は単語頻度 $f_{t,d}$ の飽和の速さを決めるパラメータであり，$k1$ によって単語頻度の影響の大きさが変わる．b は文書 d の長さ d_l の正規化の程度を決めるパラメータであり，b によって文書中の単語数による影響の大きさが変わる．本研究では，パラメータはそれぞれ $k1 = 1.2, b = 0.75$ と設定して実験を行った．$f_{\mathrm{BM25}}(q,d)$ の出力は値が大きいほど，タグ d におけるユーザ q の重要度が高く，値が小さいほどタグ d におけるユーザ q の重要度が低いことを示す．

3.3 深層学習に基づくゼロショット学習モデル

深層学習に基づくゼロショット学習モデルについて述べる．タグとユーザの埋め込みを利用することでゼロショット学習問題に対応した，ニューラルネットワークに基づく適合性判定モデルを示す．本研究では，Pushp と Srivastava が提案したゼロショットテキスト分類のフレークワークにおける手法 [15] を応用する．

提案するモデルを図 1 に示す．提案するモデルは，タグとユーザを入力として与えたとき，次の手順によって出力を得る：(1) タグの単語埋め込みを獲得（タグを単語埋め込み空間に写像）, (2) ユーザの単語埋め込みを獲得（ユーザを単語埋め込み空間に写像）, (3) ユーザの単語埋め込みとタグの単語埋め込みを結合, (4) 結合したベクトルの各要素を全結合型ニューラルネットワークに入力, (5) タグとユーザが適合する確率の出力．

重要な点として，タグとユーザの単語埋め込みを得るということは，タグとユーザを単語埋め込み空間に写像しているということである．また，写像したタグとユーザのベクトル表現を用いることで，単語埋め込み空間上のタグとユーザの対応関係を学習している．これによって，タグの埋め込みとユーザの埋め込みさえ用意できれば，タグごとの訓練データとなるユーザを用意する必要がなくなり，ゼロショット問題に対応可能となる．

3.3.1 タグとユーザの単語埋め込みの獲得

タグとユーザの単語埋め込みの獲得について述べる．提案手法における最初の段階として，タグ中の各単語とユーザに対応する文集合の各文の各単語を，単語の意味を表現可能な単語埋め込み空間上に写像する．

まず，タグの単語埋め込みの獲得について述べる．タグ集合における各タグ $d \in D$ は n 個の単語 $X = \{x_1, x_2, ..., x_n\}$ からなる．このとき，単語 $x_n \in X$ は全語彙の集合である V の $|V|$ 次元ベクトル $\mathbf{v}_n$ によって表現される．$\mathbf{v}_n$ は，単語 x_n に対応する次元の要素が 1，それ以外の次元の要素が 0 で構成される One-hot ベクトルとなる．各 One-hot ベクトルを，単語埋め込み行列によって単語ベクトルに変換し，平均してタグの単語埋め込みベクトルとする．式は次の通りである：

$$\mathbf{t} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E_w}\mathbf{v}_i \tag{4}$$

このとき，$\mathbf{t}$ はタグの単語埋め込みベクトル，n はタグの単語数，$\mathbf{V}_i$ は i 番目の単語の単語埋め込み，$\mathbf{E_w}$ は単語埋め込み行列であり $\mathbf{E_w} \in \mathbb{R}^{d_e \times |V|}$ となる．d_e は単語ベクトルの次元を表す．

次に，ユーザの単語埋め込みの獲得について述べる．ユーザの単語埋め込みベクトルは，ユーザ $q \in Q$ を構成する文集合の各文に対して，式 4 と同様の操作を行って得られた文集合の各文のベクトルを平均してユーザの単語埋め込みベクトルとする．式は次の通りである．

$$\mathbf{u} = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n_j}\sum_{i=1}^{n_j}\mathbf{E_w}\mathbf{v}_{ij} \tag{5}$$

このとき，$\mathbf{u}$ はユーザの単語埋め込みベクトル，m はユーザを構成する文の数，n_j は j 番目の文の単語数，$\mathbf{V}_{ij}$ は j 番目の文の i 番目の単語の単語埋め込み，$\mathbf{E_w}$ は単語埋め込み行列であり $\mathbf{E_w} \in \mathbb{R}^{d_e \times |V|}$ となる．d_e は単語ベクトルの次元を表す．

3.3.2 タグとユーザの対応関係の獲得

タグとユーザの対応関係ついて述べる．まず，ユーザとタグの単語埋め込みを結合する．式 4 によって獲得したタグの単語埋め込み $\mathbf{t} = (t_1, t_2, ..., t_{|V|})^T$ と式 5 によって獲得したユーザの単語埋め込みベクトル $\mathbf{u} = (u_1, u_2, ..., u_{|V|})^T$ を連結したベクトル $\mathbf{x} = (x_1, x_2, ..., x_{2|V|})^T$ を全結合型ニューラルネットワークの入力とする．提案手法における全結合型ニューラルネットワークは，入力層，中間層，出力層からなる単純な多層パーセプトロンである．入力層における入力の次元数は $2|V|$ である．出力層におけるベクトルは，1 次元ベクトルを出力する．各層における入力ベクトルを $\mathbf{h}_{\mathrm{in}}$，活性化関数を a，重み行列を $\mathbf{W}$，入力バイアスを $\mathbf{b}$，出力ベクトルを $\mathbf{h}_{\mathrm{out}}$ として次の式で表す：

$$\mathbf{h}_{\mathrm{out}} = a(\mathbf{W}\mathbf{h}^*_{\mathrm{in}} + \mathbf{b}) \tag{6}$$

このとき，活性化関数 a は Rectified Linear Unit を用いる．また，重み行列は $\mathbf{W} \in \mathbb{R}^{d_h \times d_h}$ であり，入力バイアスは $\mathbf{b} \in \mathbb{R}^{d_h}$ である．

3.3.3 モデルの学習

モデルの学習について述べる．学習時の損失関数としてバイナリ交差エントロピーを用いる．タグとユーザに対して付与されている適合性ラベル $y = 0, 1$ と，モデルによる予測結果である適合度との差が小さくなることを目的として学習する．

学習時のパラメータ最適化手法として Adam [17] を使用する．Adam は確率的勾配降下法において勾配の 1 次モーメントと 2 次モーメントの移動平均によって学習率を最適化する．

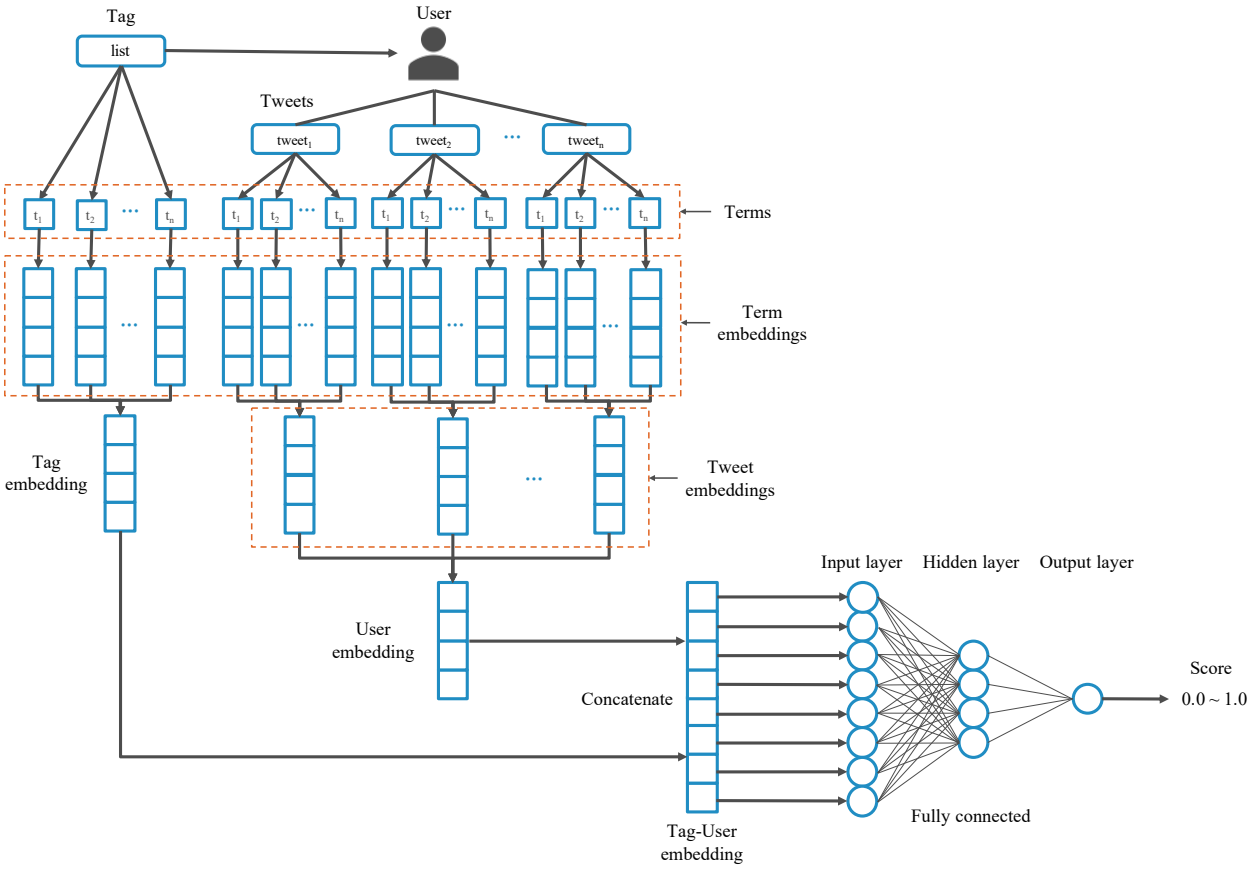

図 1　モデルの構成

AdaGrad や SGDNesterov などの最適化手法と比較してより性能が良いことがわかっている．

3.4　BERT に基づくゼロショット学習モデル

BERT に基づくゼロショット学習モデルについて述べる．BERT(Bidirectional Encoder Representations from Transformers) [18] は，Devlin らによって提案された手法で，文章分類，質問応答といった広範囲の自然言語処理タスクで優れた性能を発揮する言語モデルである．BERT の特徴は，入力としてベクトルのシーケンスが与えられたとき，全ての層において双方向の学習を同時に行う点である．BERT 以前のモデルでは，入力としてベクトルのシーケンスが与えられたとき，全ての層において左から右への学習のみのモデルか，左から右への学習と右から左への学習を出力層で連結するモデルとなっていた．BERT のモデルの構造は，入力層と出力層の間の中間層において 12 層の Transformer を重ねる構造となっている．Transformer は，Vaswani らによって提案された手法 [19] で Attention と呼ばれる学習機構を用いた Encoder と Decoder によって構成される学習モデルである．BERT では Transformer の Encoder のみを使用する．

BERT に基づくゼロショット学習モデルは，タグとユーザを入力として与えたとき，次の手順によって出力を得る：(1) タグとユーザを構成する各文から BERT の入力シーケンスを作成して入力，(2) BERT に入力した結果得られた各出力を全結合型ニューラルネットワークに入力，(3) 全結合型ニューラルネットワークから得られた各出力の中で最も高い値を出力する．

3.4.1　BERT の入力シーケンス

BERT の入力作成について述べる．BERT の入力シーケンスは形式が決まっている．まず，先頭に classification embedding と呼ばれる制御トークンである [CLS] を追加する．次に，タグの単語を分割して追加する．BERT では，[SEP] という制御トークンを配置することで入力における各文の区切りを表すことができる．タグの単語を追加した後，[SEP] を追加する．それから，ユーザを構成する文集合の文を単語分割して追加する．文を入力した後，文の区切りを表す [SEP] を追加する．以上の手順により入力は次のような構成となる：

$$[[\mathrm{CLS}], t_1^d, t_2^d, ..., t_n^d, [\mathrm{SEP}], t_1^s, t_2^s, ..., t_m^s, [\mathrm{SEP}]] \qquad (7)$$

BERT の入力作成は，タグ集合における各タグと，タグとの適合度を予測したいユーザを構成する文集合の各文全てで行う．

3.4.2　BERT の予測結果によるゼロショットモデルの学習

BERT の予測結果によるゼロショットモデルの学習について述べる．3.4.1 小節で作成した各入力シーケンスを BERT モデルに入力する．入力シーケンスの数は $|D| \times |Q| \times |S|$ となる．出力として，BERT による入力シーケンスのエンコード結果が得られるが，このうち，入力シーケンスにおける [CLS] に対応する先頭のベクトル $\mathbf{x}$ のみを，ゼロショット学習モデルの入

力として使用する．出力されるベクトルの数は，タグ集合における各タグと，タグとの適合度を予測したいユーザを構成する文集合の各文全てであり，$|D| \times |Q| \times |S|$ となる．得られたベクトル $\mathbf{x} \in \mathbf{X}$ を式 6 と同様の，入力層，中間層，出力層からなる単純な全結合型ニューラルネットワークに入力として与える．入力層における入力の次元数は $|x|$ である．出力層におけるベクトルは，1 次元ベクトルを出力する．各層における入力ベクトルを $\mathbf{h}_{\text{in}}$，活性化関数を a，重み行列を $\mathbf{W}$，入力バイアスを $\mathbf{b}$，出力ベクトルを $\mathbf{h}_{\text{out}}$ とする．また，活性化関数 a は Rectified Linear Unit，重み行列は $\mathbf{W} \in \mathbb{R}^{d_h \times d_h}$ であり，入力バイアスは $\mathbf{b} \in \mathbb{R}^{d_h}$ である．出力される値の数は，タグ集合における各タグと，タグとの適合度を予測したいユーザを構成する文集合の各文全てであり，$|D| \times |Q| \times |S|$ となる．このとき，ユーザとタグごとに適合性を考慮するために，あるタグ $d \in D$ と，あるユーザの各文の適合度 $s \in S$ を，ユーザごとの文集合の数 $|S|$ でわけ，ベクトル $x = (v1, v2, ..., vn)^T$ とする．このタグとユーザの各文の適合度ベクトルの中で最も高い適合度をタグとユーザの適合度として，結果を出力する．

3.5 BM25 とゼロショット学習モデルの組み合わせ

BM25 とゼロショット学習モデルの組み合わせについて述べる．

まず，BM25 と深層学習に基づくゼロショット学習モデルの組み合わせについて述べる．ユーザに対するタグの適合度である BM25 スコアを求める式 1 とユーザとタグの対応関係を学習したモデルによってユーザとタグの適合度を予測する式 6 を用いる．組み合わせた式を次に示す：

$$f_{\text{BM25+ZSL}}(q,d) = w \cdot f_{\text{BM25}}(q,d) + (1-w) \cdot f_{\text{ZSL}}(q,d) \quad (8)$$

4 実　　験

実験では提案手法に関する次の疑問に答えることを目的とする：(RQ) ユーザに対するタグ付けにおけるゼロショット学習問題に提案手法は有効か．

本節ではまず作成したデータセットついて述べ，ベースライン手法を含む実験設定について紹介し，最後に実験結果を示す．

4.1 データセット

我々のタスクには公開されたデータセットが存在しないため，データセットの作成について述べる．Twitter からユーザ名，ユーザが含まれるリスト，ユーザの投稿を収集した．手順は次のようになっている．(1) ユーザ名の収集，(2) ユーザが含まれるリストの収集，(3) リストの選定，(4) ユーザのツイート収集，(5) ユーザの選定．収集した Twitter データは，日本のユーザを対象とした．

ユーザ名の収集について述べる．ユーザ名とは，@から始まるアカウント固有の文字列である．ユーザ名の収集において，フォロワーが多いユーザは多くのリストに含まれている可能性が高いと仮定した．仮定に基づき，フォロワーの多いユーザを meyou[3]の「Twitter フォロワー数総合ランキング」のページより抽出した．収集したユーザ数は 80,271 ユーザ（2020 年 10 月 6 日当時）であった．収取したユーザは全て Twitter でのロケーション設定が日本となっているユーザである．

収集したユーザに基づいたユーザが含まれるリストの収集について述べる．Twitter API[4]の”GET lists/memberships”に対してユーザ名をリクエストすることで，レスポンスとしてリクエストしたユーザが含まれているリストの情報を取得できる．収集した 80,271 ユーザ全てに対して処理を実行し，2,199,456 リストを収集した．リストは公開と非公開の 2 種類がある．また，ユーザが非公開になっているとリストの情報も非公開となる．非公開のリストについての情報は収集できないため，非公開リストを省く．非公開リストを省いた結果，1,829,518 リストに限定された．

収集したリストの選定について述べる．選定する条件として，主にリスト名に注目した．リストの選定条件は次の通りである：(1) 日本語のみで構成されている，(2) 固有名詞を 1 つ以上含む，(3) 内容語を 2 つ以上含む，(4) センシティブな単語を含まない，(5) 一度出現したリスト名は重複しない．日本語日本語のみで構成されているリストの判定は正規表現を用いて行った．正規表現によって，平仮名または片仮名または漢字が一文字以上含まれているリストに限定した．固有名詞を 1 つ以上含むリストの判定，内容語を 2 つ以上含むリストの判定は，形態素解析器 MeCab[5]を用いて行った．MeCab の辞書は，mecab-ipadic-NEologd[6]を用いた．固有名詞を 1 つ以上含むリストの判定では，MeCab の形態素解析結果の品詞細分類 1 が「固有名詞」であるリストに限定した．内容語を 2 つ以上含むリストの判定では，MeCab の形態素解析結果の品詞が「名詞」「動詞」「形容詞」のいずれかが 2 つ以上含まれているリストに限定した．センシティブな単語を含まないリストの判定としては，ストップワード辞書を作成し，辞書内の単語が 1 文字以上リスト名に含まれる場合にはリストを除外した．一度出現したリスト名は重複しないという条件の判定は，同じ名前を持つリストが複数あった場合，最もメンバー数の多いリストを選択した．以上の条件でリストを選定した結果，8,501 リストに限定された．

ユーザのツイート収集について述べる．Twitter API の”statuses/user_timeline”に対してユーザ名をリクエストすることで，レスポンスとしてリクエストしたユーザが投稿したツイートを取得できる．1 回のリクエストに対して 200 件のツイートがレスポンスされる．学習に十分なデータを確保するために，1 ユーザにつき 5 会のリクエストを行い，1,000 件のツイートの取得を行った．以上の処理を 80,271 ユーザ全てに対して行った．ツイートを収集した結果，取得できたツイート数が 1,000 件未満のユーザの除外を行った．その結果，19,805 ユーザに限定された．以上の処理を行い，Twitter からユーザ名，ユーザ

3：https://meyou.jp/ranking/follower_allcat/
4：https://developer.twitter.com/en/docs/twitter-api
5：https://taku910.github.io/mecab/
6：https://github.com/neologd/mecab-ipadic-neologd

が含まれるリスト，ユーザの投稿を収集した結果，8,501 リスト，19,805 ユーザを取得した．

取得したリストとユーザを対応づけてデータセットの作成を行う．まず，リストとユーザの対応づけは，ユーザに対してユーザが含まれているリストを正解のリストとして対応づけ正解ラベルを付与する．次に，正解ラベルを付与したリストとタグの対となるデータ集合を，既知クラスと未知クラスに分割する．提案モデルの性能評価を行うために，あるユーザに対して 1 件の正解タグと 99 件の不正解タグを用意する．不正解タグは，1 件の正解タグに基づき，正解タグのクラスが既知クラスであれば既知クラスに含まれる別のタグから 99 件を，正解タグのクラスが未知クラスであれば未知クラスに含まれる別のタグから 99 件をそれぞれランダムに選択して，100 件のタグを 1 つのセットとして，このセットをユーザ数分作成してデータセットとする．

4.2 評価指標

評価指標について述べる．評価指標は Hit@k と nDCG@k [20] を用いる．作成したデータセットにおいて，あるユーザに対して 100 件のタグが用意されている．提案モデルによってユーザと各タグの適合度を算出し，算出されたスコアからタグの順位付けを行う．順位付けされたタグ集合において，上位 k 件に正解とラベル付けされているタグが含まれるかによってモデルの性能評価を行う．ユーザとタグの適合度を算出する際，ユーザの特徴量はユーザの Twitter におけるツイート 1,000 件を用いる．

4.3 実験設定

実験設定について述べる．まず，単語の埋め込みについて述べる．単語埋め込みを獲得する手法として fastText を利用する．fastText は Bojanowski らによって提案された単語を分散表現として表現できる対数総線形モデルである [21], [22]．fastText の特徴は，同じ単語の活用形の違いを考慮することでより高い表現力を持つ点である．fastText の学習モデルとして主に skip-gram と CBoW がある．skip-gram では，入力単語の分散表現を用いて入力単語の前後の単語，つまり文脈を予測する．CBoW(Continuous Bag of Words) では，対象となる単語の前後の単語，つまり文脈の分散表現を組み合わせて，対象となる単語を予測する．CBoW は skip-gram を包含するモデルの形式となっている．本研究では CBoW を選択する．fastText のモデルは Facebook の公開している学習済みモデルを用いる [7]．

ベースライン手法について述べる．実験におけるベースライン手法として，既存の文書順位付け手法を用いた：(1) **BM25** [16]：Robertson らによって提案された一般的に広く普及している確率的情報検索モデルでの実装を実験に用いた．

実験で用いた提案手法について述べる．実験における提案手法として，ユーザとタグの埋め込み空間上の対応関係を学習するゼロショット学習モデルを用いた．埋め込みの手法として fastTest，BERT を用いた．損失関数として Binary cross entropy 損失と Margin ranking 損失を用いた．また，BM25 とゼロショット学習手法を組み合わせた手法での実装を実験に用いた．

7：https://fasttext.cc/docs/en/crawl-vectors.html

表 1　各手法によるタグとユーザの適合度をランキングした上位 10 件の精度

model	nDCG@10	Hit@10
BM25	0.04702	0.07400
ZSL(fastText, BCELoss)	0.00966	0.02000
ZSL(fastText, MRLoss)	0.01070	0.01900
ZSL(BERT, BCELoss)	0.01146	0.02300
ZSL(BERT, MRLoss)	0.01103	0.02100
BM25+ZSL(fastText, BCELoss)	0.04679	0.07300
BM25+ZSL(fastText, MRLoss)	0.04696	0.07300
BM25+ZSL(BERT, BCELoss)	0.04666	0.07300
BM25+ZSL(BERT, MRLoss)	0.04739	0.07200

4.4 実験結果

表 1 に各手法の実験結果を示す．(RQ) ユーザに対するタグ付けにおけるゼロショット学習問題に提案手法は有効か：既存の文書順位付け手法である BM25 とゼロショット学習手法を組み合わせた手法が最も高いスコアを示した．これは，単語埋め込み空間上におけるタグの表現とユーザの表現との対応関係が必ずしも関連性を表現できているわけではなく，単語埋め込みを獲得するモデルに表現が依存することが要因として考えられる．また，タグの単語埋め込みとツイートの単語埋め込みの平均を行ったが，これも同様に単純な平均のみでは十分でなかったことが考えられる．さらに，データセットにおけるデータの偏りも強く影響していると考えられる．

5 まとめ

本論文では，SNS ユーザに対してユーザの性質を表すタグ付けを行う手法を提案した．既存のユーザ分析における課題として，具体的な性質の表現が難しくタグの表現が単語に限定されていること，教師あり機械学習手法を用いる場合，学習データ中に含まれていないパターンのユーザを正しく予測できないこと，学習データの作成に大きなコストがかかることがあった．それらの課題に対して，ゼロショット学習手法を用いてユーザとタグとの対応関係を学習することで，学習データがない場合でもユーザに対してタグ付けする手法を提案した．具体的にはゼロショット学習と呼ばれる学習データ中に一度も出現しないようなパターンの事例を予測できる機械学習手法を応用して，Twitter におけるユーザのツイートとユーザが含まれるリストの名前をタグとして用いて，ユーザとタグの対応関係を獲得することで，学習時に出現しないタグをもユーザに対して付与した．タグの表現をタグに含まれる各単語の埋め込みを平均したベクトルを使用することで，既存手法のタグの表現が単語に限定されるという課題にも対処している．実験では，提案手法の学習においては文書分類タスク，予測においては文書検索タスクに取り組み，ユーザとタグの適合度によって順位

付けを行った．実験に用いたデータセットは Twitter から収集したリストとユーザ単位のツイートから作成し，文書検索タスクのベースライン手法と提案手法の有効性について，Hit@k, nDCG@k などを用いて上位 k 件に正解タグが含まれるかどうかで評価を行い，評価結果を比較した．その結果，Hit@10 において BM25 による単語のマッチングによる手法，nDCG@10 において BM25 と単語埋め込みを用いたゼロショット学習手法が有効であることが明らかとなった．

謝辞 本研究は JSPS 科研費 18H03243 の助成を受けたものです．ここに記して謝意を表します．

文　献

[1] Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. Classifying latent user attributes in twitter. In *Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents*, SMUC '10, page 37–44, New York, NY, USA, 2010. Association for Computing Machinery.

[2] Y. Yamaguchi, T. Amagasa, and H. Kitagawa. Tag-based user topic discovery using twitter lists. In *2011 International Conference on Advances in Social Networks Analysis and Mining*, pages 13–20, 2011.

[3] Marco Pennacchiotti and Ana-Maria Popescu. A machine learning approach to twitter user classification. In *Fifth international AAAI conference on weblogs and social media*, 2011.

[4] Dongwoo Kim, Yohan Jo, Il chul Moon, and Alice Oh. Analysis of twitter lists as a potential source for discovering latent characteristics of users. In *CHI 2010 Workshop on Microblogging: What and How Can We Learn From It*, 2010.

[5] Naveen Kumar Sharma, Saptarshi Ghosh, Fabricio Benevenuto, Niloy Ganguly, and Krishna Gummadi. Inferring who-is-who in the twitter social network. *SIGCOMM Comput. Commun. Rev.*, 42(4):533–538, September 2012.

[6] Meeyoung Cha, Hamed Haddadi, Fabrício Benevenuto, and Krishna P. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *in ICWSM '10: Proceedings of international AAAI Conference on Weblogs and Social*, 2010.

[7] Mark Palatucci, Dean Pomerleau, Geoffrey Hinton, and Tom M. Mitchell. Zero-shot learning with semantic output codes. NIPS'09, page 1410–1418, Red Hook, NY, USA, 2009. Curran Associates Inc.

[8] Andrea Frome, Greg S. Corrado, Jonathon Shlens, Samy Bengio, Jeffrey Dean, Marc'Aurelio Ranzato, and Tomas Mikolov. Devise: A deep visual-semantic embedding model. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, page 2121–2129, Red Hook, NY, USA, 2013. Curran Associates Inc.

[9] Elyor Kodirov, Tao Xiang, and Shaogang Gong. Semantic autoencoder for zero-shot learning. *arXiv.org e-Print archive*, 2017, 1704.08345. https://arxiv.org/abs/1704.08345, (accessed 2020-12-18).

[10] Z. Akata, S. Reed, D. Walter, Honglak Lee, and B. Schiele. Evaluation of output embeddings for fine-grained image classification. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2927–2936, 2015.

[11] Bernardino Romera-Paredes and Philip H. S. Torr. An embarrassingly simple approach to zero-shot learning. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, page 2152–2161. JMLR.org, 2015.

[12] Z. Akata, F. Perronnin, Z. Harchaoui, and C. Schmid. Label-embedding for attribute-based classification. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 819–826, 2013.

[13] Wei Wang, Vincent W. Zheng, Han Yu, and Chunyan Miao. A survey of zero-shot learning: Settings, methods, and applications. 10(2), January 2019.

[14] Hugo Larochelle, Dumitru Erhan, and Yoshua Bengio. Zero-data learning of new tasks. In *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 2*, AAAI'08, page 646–651. AAAI Press, 2008.

[15] Pushpankar Kumar Pushp and Muktabh Mayank Srivastava. Train once, test anywhere: Zero-shot learning for text classification. *arXiv.org e-Print archive*, 2017, 1712.05972. https://arxiv.org/abs/1712.05972, (accessed 2020-12-18).

[16] Stephen Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, and M. Gatford. Okapi at trec-3. In *Overview of the Third Text REtrieval Conference (TREC-3)*, pages 109–126, January 1995.

[17] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv.org e-Print archive*, 2017, 1412.6980. https://arxiv.org/abs/1412.6980, (accessed 2020-12-18).

[18] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, undefinedukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc.

[20] Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Trans. Inf. Syst.*, 20(4):422–446, October 2002.

[21] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146, 2017.

[22] 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: Volume 2, Short Papers*, pages 427–431, Valencia, Spain, April 2017. Association for Computational Linguistics.