トピックに基づくテキスト生成による マイクロブログにおけるなりすまし投稿の生成

田村　優幸† 　新田　直子†† 　中村　和晃†† 　馬場口　登††

† 大阪大学工学部　〒 565–0871　大阪府吹田市山田丘 2-1

†† 大阪大学大学院工学研究科　〒 565–0871　大阪府吹田市山田丘 2-1

E-mail: †tamura@nanase.comm.eng.osaka-u.ac.jp, ††{naoko,k-nakamura,babaguchi}@comm.eng.osaka-u.ac.jp

あらまし　近年，深層学習を用いたテキスト生成手法により，ソーシャルネットワーキングサービスに公開された過去の投稿を学習データとし，正規のユーザを模倣した投稿を行うなりすましアカウントの作成が可能となっている．本研究では，Twitter などのマイクロブログにおいて多様なユーザによる類似したトピックに関する投稿が類似パターンを持つという前提のもと，なりすまし対象となる人物以外の大量の投稿を学習データとして用いることにより，多様ななりすまし対象ユーザに対し，共通のテキスト生成モデルを用いてなりすまし投稿の生成を実現する手法を提案する．提案手法は主に，多様なユーザの投稿からの学習データの選択，投稿集合からのトピックの推定，テキスト生成モデルの学習により構成される．本稿では，これらの要素技術，及びなりすまし対象ユーザの過去の投稿数によるなりすまし投稿の質への影響について検証する．

キーワード　テキスト生成，マイクロブログ，なりすまし

1. はじめに

近年，ソーシャルネットワーキングサービス（SNS）において，人間になりすまし，特定の個人，団体，商品，サービス等に対する擁護や誹謗中傷を目的とした投稿などを行うアカウントの存在が問題となっている．これらのアカウントは人間が人手で運用している場合もあるが，正規のユーザがマイクロブログ上に公開している情報を利用し，なりすましアカウントの作成や，投稿の自動生成が可能であることが示されている．特に，短文が投稿される Twitter [1] などのマイクロブログでは，過去の発言や文章が容易かつ大量に入手可能な著名人を対象に，近年発展が著しい再帰的ニューラルネットワーク（RNN）等の深層学習による文書生成技術を用いて自動生成した文章を投稿するなりすましアカウントも存在している [4].

本研究では，同様にマイクロブログを対象に，過去の発言や文章が大量に公開されていない人物に対しても，不特定多数のユーザの過去の投稿を学習データとして用いることにより，対象人物になりすました投稿の自動生成が可能か検討する．対象人物の過去の投稿を用いない場合，口調や文体の模倣は困難と考えられる．そこで，スポーツが好きな人物は，スポーツに関する投稿が多いなど，投稿において個性が表れる要素として個人の嗜好に着目する．つまり，スポーツや音楽など，トピックに応じて適切な投稿を生成する投稿生成器を学習し，対象人物の少量の公開情報から推定される嗜好に応じた投稿を生成することにより，共通の投稿生成器を用いたなりすまし投稿の生成が可能と考えられる．ここでは特に，実世界に存在する正規のユーザの共通点として，各地の施設やイベントなどの地域情報に関して投稿を行うことが多く，各地のレストランやスポーツイベント会場など類似したトピックを持つ地域情報に関する投稿が類似したパターンを持つと考えられることから，地域情報に関する投稿を生成対象とする．

提案手法ではまず，不特定多数のユーザによる投稿から，地域情報に関する投稿を抽出する [3]．次に各地域情報に対して，複数ユーザからの投稿に共通して用いられる単語に基づきトピックを推定し，推定されたトピックと複数ユーザからの投稿の対を用いてトピックに基づく投稿生成器を学習する．なりすまし投稿を生成する際は，対象ユーザの過去の投稿から推定したユーザの嗜好を表すトピックに基づき，嗜好に応じた地域情報を選択する．最後に，学習した投稿生成器に対して，選択された地域情報のトピックを入力して投稿を生成する．

2. 提案手法

本研究では，Twitter 上の任意のユーザ U に対し，過去の投稿 T^U が公開されていることを前提に，U の嗜好に応じた地域情報に関する投稿 $t^{U_{spoof}}$ をなりすまし投稿として生成することを目的とする [2].

提案手法では，投稿生成器の学習は事前に実行するオフライン処理，なりすまし投稿の生成は随時実行するオンライン処理とする．まずある時区間における不特定多数のユーザによる地域情報に関する投稿を収集し，それらを用いてトピックに応じた投稿生成器の学習を行う．なりすまし対象を U とするときは，なりすましを行う時区間の実世界の状況との整合性を踏まえ，該当時区間における不特定多数のユーザによる投稿から抽出した地域情報から U の嗜好に応じたものを選択する．選択された地域情報のトピックを投稿生成器の入力とし，U のなりすまし投稿を生成する．提案手法の全体図を図 1 に示す．

Step 1)　**地域情報に関する投稿の収集**

不特定多数のユーザによる投稿の中から，地域情報 $I_k(k =$

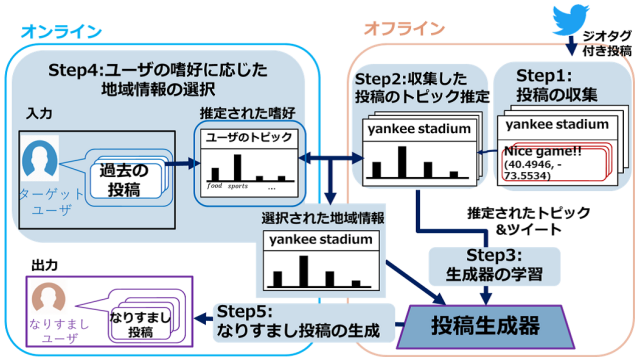

図 1　提案手法の概要

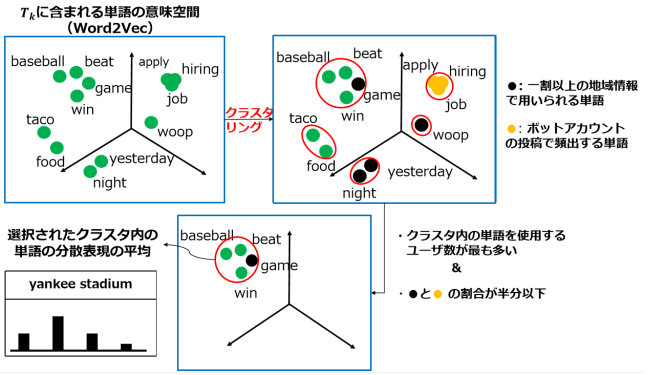

図 2　トピックと関連の強い単語を用いた分散表現の算出

$1, 2, \cdots$) に関する投稿集合 $T_k = \{t_{k,n} \mid n = 1, \cdots, N_k\}$ を収集する．

Step 2)　地域情報のトピック推定

地域情報 I_k に関する投稿集合 T_k には，I_k のトピックを表す単語が表れると考え，これらの単語の意味を表す分散表現に基づき I_k のトピックを表す分散表現 $\boldsymbol{f}_k$ を算出する．

Step 3)　トピックに基づく投稿生成器の学習

トピック $\boldsymbol{f}_k$ と投稿 $t_{k,n} \in T_k$ の対を学習データとし，入力されたトピックに応じた投稿を生成する投稿生成器 $t_{k,n} = G(\boldsymbol{f}_k)$ を学習する．

Step 4)　ユーザの嗜好に応じた地域情報の選択

Step 2) と同様に，ユーザ U の過去の投稿 T^U に含まれる単語を用いてトピックを表す分散表現 $\boldsymbol{f}^U$ を算出し，これをユーザ U の嗜好を表すものとする．なりすましを行う時区間において抽出した地域情報 $I_k (k = 1, 2, \cdots)$ に対してもトピックを推定し，$\boldsymbol{f}^U$ と類似度の高いトピック $\boldsymbol{f}^K$ により表される地域情報を，ユーザの嗜好に応じたものとして選択する．

Step 5)　選択された地域情報に関する投稿生成

投稿生成器 G により，$t^{U_{spoof}} = G(\boldsymbol{f}^K)$ としてユーザ U のなりすまし投稿 $t^{U_{spoof}}$ を生成する．

以下，各ステップの詳細を述べる．

2.1　地域情報に関する投稿の収集

地域情報に関する投稿には，施設名など複数のユーザが共通して使用するそれぞれの地域情報を表す単語が含まれる可能性が高い．これらの単語は，該当する施設などが存在する地点で集中的に用いられると期待される．マイクロブログへの投稿には，投稿位置の緯度経度を表すジオタグが付与されたものが存在するため，まず不特定多数のユーザによるこれらのジオタグ付き投稿を収集し，空間的局所性の高い単語を地域情報 I_k を表すローカル語 l_k として抽出した上で，ローカル語 l_k を含む投稿 $T_k = \{t_{k,n} \mid n = 1, \cdots\}$ を収集する．ここでは随時最新の地域情報を抽出するため，逐次的にローカル語を抽出する手法 [3] を適用する．

2.2　地域情報のトピック推定

2.1 節で収集された T_k には，各地域情報 I_k のトピックを表す単語が含まれると考えられる．

そこでまず，word2vec [5], [6] を利用し，単語の意味を表す分散表現を学習する．ここでは，同一の地域情報に関する投稿で用いられる単語同士が意味空間において近くなるよう，T_k 全体を 1 つの文書として統合し，特定の時区間において収集された $T = \{T_k | k = 1, 2, \cdots\}$ を学習用コーパスとする．ただし，$t_{k,n}$ に対する前処理として，“http(s)://...” という形式の URL，ストップワード，数字，及びローカル語集合 $L = \{l_k \mid k = 1, 2, \cdots\}$ に含まれる単語を，地域情報のトピック推定に不要な単語として削除する．これにより，T に含まれる M 個の単語 $w_m (m = 1, 2, \cdots, M)$ に対する分散表現 $\boldsymbol{v}_m$ が得られる．

地域情報 I_k のトピックは，T_k 中の単語の分散表現を用いて算出すればよい．ただし，T_k に含まれる単語はそれぞれ，トピックへの関連の強さが異なる．そこで特にトピックと関連の強い単語を抽出する．各地域情報の T_k の中では，異なるユーザは，トピックを表すために多様な単語を使用し得るが，これらの単語は意味的に近いと期待される．そこで，図 2 のように，各地域情報において，T_k に含まれる全ての単語に対しクラスタリングを適用し，トピックと関連の強い単語で形成されるクラスタを抽出する．まず，T_k に含まれる N_k 個の単語集合に対して，$\boldsymbol{v}_m$ の類似度に基づきクラスタリングを行う．なお $\boldsymbol{v}_m$ の類似度はコサイン類似度とする．クラスタリングの手法については，非常に類似した単語のみで形成されるクラスタを得られることを期待し，階層的クラスタリング（群平均法）を用いるものとする．

クラスタリングの結果得られるクラスタのうち，“today” や “yesterday” など多くの地域情報に関する投稿で使用される単語や，特定のユーザのみに使用される単語のみで形成されるクラスタはトピックを表すものとして不適切と考えられる．加えてマイクロブログには，機械的に類似した形式の投稿を行うボットアカウントが多数存在するため，ボットアカウントの投稿に使用される単語のみでクラスタが形成されやすいが，これらもトピックを表すものとして不適切である．そこで，全投稿の一割以上に出現する単語を多くの地域情報に関する投稿で使用される単語とした．また，ボットアカウントの投稿に使用される単語を除外するために，ボットアカウントは類似した単語を用い，類似した内容の投稿を多用するため，そうした単語は word2vec 上で極めて距離が近くなることを利用する．全投稿に出現する M 個の単語それぞれに対し，$\boldsymbol{v}_m$ の距離が近い上

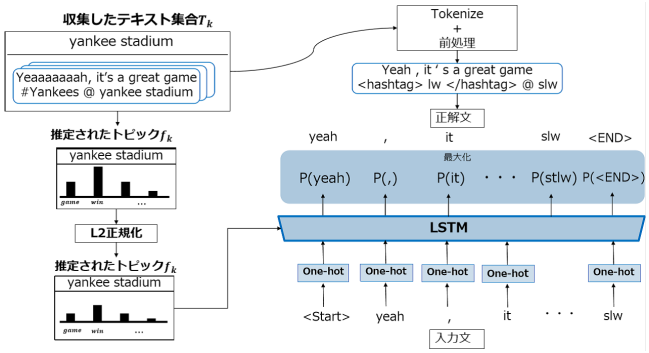

図 3　トピックに基づく投稿生成器の学習

位 10 個の単語との距離を基準に，外れ値検出により，それらを基準に外れ値検出を行って，異常に距離の近い単語を多く持つ 0.1% の単語をボットアカウントの投稿で頻出する単語とみなす．外れ値検出には Isolation Forest [7] を利用した．クラスタ内の単語において，これらの単語の割合が 50%未満かつ，クラスタ内の単語を使用するユーザ数が最も多いクラスタをトピックを表す単語のクラスタとして選択する．

選択されたクラスタ内の単語の分散表現 $\boldsymbol{v}_m$ の平均を地域情報 I_k のトピックを表す分散表現 $\boldsymbol{f}_k$ とする．

2.3　トピックに基づく投稿生成器の学習

トピックの類似した地域情報に関する投稿は，投稿時間や場所に依らず，内容が類似すると予想される．例えば，野球の試合に関する投稿は試合の状況に関する内容，各地のレストランに関する投稿は食事に関する内容が多い．そこで，収集した投稿 $t_{k,n}$ とそのトピックを表す $\boldsymbol{f}_k$ の対から，トピックに応じた投稿パターンを学習する．

ここでは図 3 に示すように，Image Captioning [8] などのタスクで使用される Long Short-Term Memory（LSTM）を用いて，入力されたトピック $\boldsymbol{f}_k$ に対して適切な投稿 $t_{k,n} \in T_k$ を生成する，つまり $t_{k,n} = G(\boldsymbol{f}_k)$ となるような投稿生成器 G を学習する．ます，学習に用いるすべての $t_{k,n}$ に対し，先頭に文の開始を表す単語 <start>，末尾に文の終了を表す単語 <end> を加える．次に，これらの文の開始終了を表す単語を含め，すべての $t_{k,n}$ に含まれる各単語に個別の ID を割り当てることにより，$t_{k,n}$ を構成する各単語を，自身の ID に相当する要素のみを 1，他の要素を 0 とした one-hot ベクトルで表す．LSTM に対し，文の開始を表す単語を入力層，トピックを表す分散表現 $\boldsymbol{f}_k$ を隠れ層への最初の入力とし，次の単語を推定する．LSTM の出力は，one-hot ベクトルと同次元のベクトルであり，各要素は該当する ID の単語が次の単語と推定される確率を表す．推定した単語を入力として次の単語を順番に推定することを繰り返し，$t_{k,n}$ を構成する単語の推定確率を最大化するよう学習することにより，トピックに基づく投稿生成器 G が学習される．

なお，$\boldsymbol{f}_k$ はトピックが類似した地域情報同士がユークリッド距離で近くなるよう，L2 ノルムを用いて正規化する．また，$t_{k,n}$ は，トピックに応じた内容の投稿として，T_k の中から，2.2 節で選択されたクラスタ中の単語を含む投稿を選択する．また，正規ユーザによる投稿が選択されるよう，2.2 節の方法で検出したボットアカウントによる投稿に頻出する単語を 2 つ以上含むものは選択しない．選択された $t_{k,n}$ に対しては，トピックに応じた投稿パターンが効果的に学習されるよう以下の前処理を施す．まず，同一単語を複数の表現で表すなど，特にマイクロブログに頻繁に見られる個々のユーザの文体における揺らぎを除去するため，Baziotis ら [10] が提案した Tokenizer を適用し，文を正規化し，ハッシュタグの開始及び終了を表す <hashtag> と </hashtag> 以外の正規化を表す単語は削除する．また，ローカル語は地域情報ごとに異なる語が使用され，トピックには共通しないものの生成される投稿には含まれることが望ましいため，ローカル語が使用される位置が学習されるよう，地域情報 I_k 自身を表すローカル語 l_k は <slw>, その他の地域情報を表すローカル語は <lw> に置換する．さらに，学習に用いるすべての $t_{k,n}$ において出現回数が極端に少ない単語は，文における役割のみ学習されるよう，品詞を表す単語に置換する．これらの処理を投稿に適用した例を表 1 に示す．

2.4　ユーザの嗜好に応じた地域情報の選択

なりすまし対象とするユーザ U の嗜好に応じた地域情報を選択するため，まず過去の投稿 T^U から嗜好を推定する．2.2 節と同様に，T^U に含まれるすべての単語に対してクラスタリングを行い，トピックと関連の強い単語で構成されるクラスタを選択する．クラスタ選択時には，単語を使用したユーザ数ではなく，単語の使用頻度をクラスタの評価指標とする．また，ユーザの嗜好は一つとは限らないため，過去の投稿において単語の使用頻度順に上位 B 個のクラスタを選択し，各クラスタの平均となる分散表現 $\boldsymbol{f}_b^U (b = 1, 2, \cdots, B)$ を，ユーザ U の嗜好を表すものとする．なりすまし投稿生成時に抽出されている地域情報に対し，同様にトピックを推定し，ユーザの各嗜好 $\boldsymbol{f}_b^U$ とのコサイン類似度に基づき，ユーザの嗜好に応じた地域情報を確率的に選択する．

2.5　選択された地域情報に関する投稿生成

最後に，選択された地域情報のトピック情報 $\boldsymbol{f}^K$ を入力とし，2.3 節で学習した投稿生成器 G を用い，$t^{U_{spoof}} = G(\boldsymbol{f}^K)$ によってユーザ U のなりすまし投稿 $t^{U_{spoof}}$ を生成する．

ただし，$t^{U_{spoof}}$ に含まれる <slw> は，選択された地域情報のローカル語 l^K に置換する．また，<lw> は T^K に含まれる l^K 以外のローカル語 $l^{K'} \in T^K (K' \neq K)$ に置換する．この際に置換するローカル語 $l^{K'}$ は，T^K における出現頻度に応じて選択する．加えて，出現回数が少ない単語を置き換えた，その単語の品詞を表すラベルを単語に置き換える．この時，文の始まりを表すラベルから該当ラベルの一つ前の単語までを LSTM で出力した場合に，その次の単語として出力される可能性が高いもののうち，ラベルの表す品詞の単語から選択する．さらに，ハッシュタグを表す <hashtag> と </hashtag> に挟まれた表現は ♯ を先頭とする表現に変換する．

3.　評 価 実 験

Twitter の Streaming API を用いてアメリカ本土を緯度が

表 1　前処理の結果例（ローカル語:yankee stadium）

正規化前	Yeaaaaaaaah, Yanks won today's game. Praying Yanks' groly. #byebyeboston @ Yankee Stadium URL
正規化後	yeah , <lw> won today ' s game . VBG <lw> ' NN . <hashtag> bye bye <lw> </hashtag> @ <slw>

表 2　地域情報の投稿及びそのトピックを表すクラスタの一例

地域情報の投稿の一例（ローカル語:albertson stadium）	クラスタ
Lions run with lions.... @ Albertsons Stadium, Boise State··· URL	game, football, lions,
We enjoyed going to our first BSU game tonight. I love my wife! #BoiseState @ Albertsons Stadium URL	fan, season, etc
地域情報の投稿の一例（ローカル語:apu）	クラスタ
Football game with this girl!? #apu #cougars @ Azusa Pacific University URL	college, football,
????? ????? : Here's one of our models from the APU back to school fair last month. @ APU URL	school, game, cougars
地域情報の投稿の一例（ローカル語:billsmafia）	クラスタ
Kickoff 2016 @buffalobills #Billsmafia URL	bills, game, nfl,
It's game day! #billsmafia #newerafield #buffalobills #balvsbuf @··· URL	kickoff, football, etc
地域情報の投稿の一例（ローカル語:adele）	クラスタ
Because sometimes you just need your big sister, some music and a good cry #adele @ Madison··· URL	live, music, singer,
Adele live at the MSG!! #adele #adelelive2016 #adeleconcert2016 #setfiretotherain #instadaily··· URL	playing, listening, etc
地域情報の投稿の一例（ローカル語:austincitylimits）	クラスタ
Live Music Capital of the World! Welcome to Austin! #ACL #AustinCityLimits #Zilker-Park #music··· URL	music, live, band,
Band of Horses tearing it up! #acltv #acltv42 #austin #austincitylimits @ Austin City Limits URL	singer, songwriter, etc
地域情報の投稿（ローカル語:kaaboo2016）	クラスタ
OMG @bruiserqueen at @loufest! Good people, rad music, best dressed. #bruiserqueen #loufest··· URL	music, band, live,
I just saw my favorite band at the moment and wish I could live it all over again. @ LouFest··· URL	jazz, guiter, blues

24 度から 49 度，経度が-125 度から-66 度の範囲と設定し 2016 年 9 月 8 日から 2016 年 10 月 9 日のうちの 30 日間に投稿された 6,655,763 件のジオタグ付き投稿を収集した．これらから地域情報に関する投稿及び，それらの投稿を行ったユーザの過去の投稿を収集し，2.4 節の手法により，各ユーザに対して実際に投稿した地域情報を選択できるか，また，各地域情報に対して，2.3 節の手法により学習した投稿生成器を用いて，トピックに応じた投稿が生成されるか検証する．

3.1　地域情報に関する投稿の収集とトピック推定

初めの 26 日間に収集された 50,345 個の地域情報に関する投稿 455,589 件を学習用コーパスとして word2vec によって単語の分散表現を学習した．表 2 と 3 に，初めの 26 日間の投稿集合から抜粋した，地域情報に対応する投稿と，word2vec によって学習した単語の分散表現を用いてトピックに関連の強い単語集合として選択されたクラスタを示す．ただし，投稿本文中の “URL” は “http(s)://...” という形式の URL である．表より，各地域情報に対して，トピックに関連の強い単語が選択されていることが分かる．また，各トピックに対して，サッカーであれば観戦する試合について，音楽であればライブについてなど，トピックに応じて一定の投稿パターンが存在することが確認された．よって，推定された地域情報についてのトピックとその地域情報に関する投稿の対を用いて，2.3 節の手法によりなりすまし投稿の生成器が学習可能であると考えられる．

また，表 4 にボットで多用される表現を含む投稿の例を示す．クラスタ選択時にボットアカウントの投稿で頻出する単語を考慮しない場合，T_k がこうした投稿で占められている地域情報が多く抽出される．これらの単語をクラスタ選択時および学習データ選択時に考慮しない場合，34,007 個の地域情報に関する 150,497 件の投稿が抽出されたが，考慮することにより，30,605 個の地域情報に関する 135,398 件の投稿が抽出された．後者においては，ボットで多用される表現を含む投稿が選択されづらくなっている．

3.2　地域情報の選択

28 日目に抽出された地域情報をランダムに 457 個選択し，それらの地域情報に関する投稿を行ったユーザを各 1 名ずつランダムに選択した．選択された 457 名のユーザの過去の投稿を収集し，各ユーザの嗜好に応じて，実際に投稿した地域情報を正しく選択できるか評価する．ユーザの嗜好推定に使用する過去の投稿数を，250 件，100 件，50 件，25 件と変更し，公開した投稿数による嗜好推定への影響を検証した．

トピック推定手法として，T^U，T_k それぞれに含まれる全単語の分散表現の総和平均をとる手法 (word2vec + mean)，各単語

表 3　地域情報の投稿及びそのトピックを表すクラスタの一例

地域情報の投稿の一例（ローカル語:312food）	クラスタ
Chairman Mao's Spicy Porkbelly #pork #porkbelly #chineese #sichuan #peppers #spicy #312 #312food··· URL	spicy, pork, food,
Sweet corn pierogis #corn #pierogi #vegetarian #vegetarianfood #westloop #312 #312food #chicago··· URL	belly, delicious
地域情報の投稿の一例（ローカル語:alinea）	クラスタ
"Food can be expressive and therefore food can be art." @Gachatz thealineagroup @ Alinea URL	food, meal
Chef's Table at Alinea, bitches. @ Alinea URL	chef, dinner
地域情報の投稿（ローカル語:zingerman s cornman farm）	クラスタ
Got myself a Zingerman's Reuben for lunch today #treatyoself @ Zingerman's Delicatessen URL	lunch, food, porn,
A fabulous sandwich from@Zingerman's for lunch. Yummy! ? @ Zingerman's Delicatessen URL	sandwich, yummy
地域情報の投稿の一例（ローカル語:hofstra university）	クラスタ
Hofstra University! They are hosting the first debate Clinton-Trump. Let's see what happens! @··· URL	debate, debates, trump,
Do we have a solid future president? That's DEBATEable ?? @ Hofstra University URL	election, hillary, etc
地域情報の投稿の一例（ローカル語:potus）	クラスタ
TBT Obama "Run DC"! #Obama #Potus #barackobama #barack @ Dallas, Texas URL	obama, trump, hillary,
Trump playing The Blame Game Sayin HRC started birther He lies like this imagine Him as POTUS then again Heard Trump was born in Germany···Hmm	donald, president, etc
地域情報の投稿（ローカル語:white house）	クラスタ
It's Debate Night!!! A Woman's place is in the White House!!! #hillary2016 #imwithher @ Penn Social URL	obama, hillary, trump,
Obama at the UN: The Brown Flood Must Never Cease URL This is what the maniac in the white house said at the U.N	president, debate, etc

表 4　ボットで多用される表現を含む投稿及びそのトピックを表すクラスタの一例

地域情報の投稿の一例（ローカル語:10pm6am）	クラスタ
Can you recommend anyone for this #job? Forklift Operator - Sit-Down & Clamp - Monday-Friday, 10pm-6am -... - URL	job, recommend, hiring,
We're #hiring! Click to apply: Barista Bradys Midnight 10pm-6am - URL #Job #restaurantlife #Hospitality #Mantua, OH	barista, lpn, etc
地域情報の投稿の一例（ローカル語:adelanto）	クラスタ
Want to work at CVS Health? We're #hiring in #Adelanto, CA! Click for details: URL #Job #Jobs	job, hiring, deltails,
Can you recommend anyone for this #job? Shift Supervisor (US) - URL #Hospitality 14136 HWY 395, #ADELANTO, CA #Veterans	retail, jobs, etc
地域情報の投稿の一例（ローカル語:burley）	クラスタ
Interested in a #job in #Burley, ID? This could be a great fit: URL #LoveFashionLetsTalk URL	job, hiring, merchandiser,
See our latest #Burley, ID #job and click to apply: Sales Associate-maurices - URL URL	customer, service, etc

に対する TFIDF 値を重みとして，全単語の分散表現の重み付き総和平均をとる手法 (word2vec + TFIDF)，及び文書分類の分野で幅広く利用されているトピックモデルの一種である Latent Dirichlet Allocation(LDA) [9] を用いた手法を比較手法とした．ただし，word2vec + mean，word2vec + TFIDF は提案手法で学習した word2vec を用いた．また，LDA は word2vec 学習時と同じコーパスを用いて学習したトピックの単語分布を用い，T^U，T_k それぞれに対して推定されたトピック分布を $\boldsymbol{f}^U$，$\boldsymbol{f}_k$ とした．

図 4 ～ 8 に，ユーザの過去の投稿数をそれぞれ変更した結果を示す．各図は，457 個の地域情報を各ユーザの嗜好との類似度に基づき順位付けし，実際に投稿した地域情報が R 位内に選択されたユーザ数を示す．提案手法においてユーザの嗜好数は $B = 1, 2, 3$ とした．なお，投稿数を減らしていった結果，使用投稿数 25 件においては 3 名のユーザに対して，投稿からのクラスタ形成が出来ず，使用投稿数 10 件においては多数のユーザで投稿からのクラスタ形成が出来ずに嗜好推定が不可能であった．クラスタが形成されたユーザについては，いずれの図においても無作為に選択した baseline に比べ，他の手法によりユーザによって投稿された地域情報が正しく選択されたこ

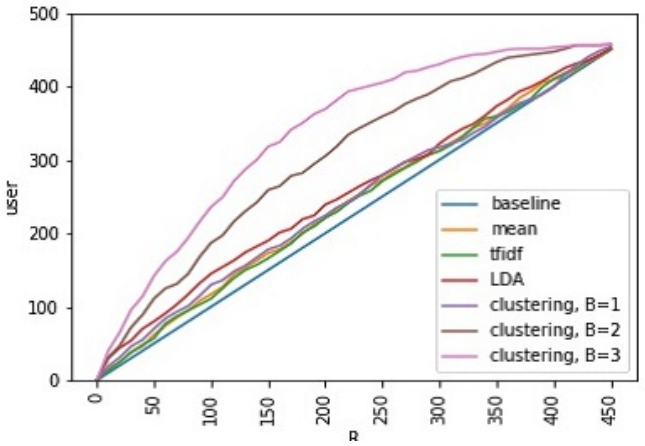

図 4　対となる地域情報が上位 R 位に選択された人数 (全投稿使用)

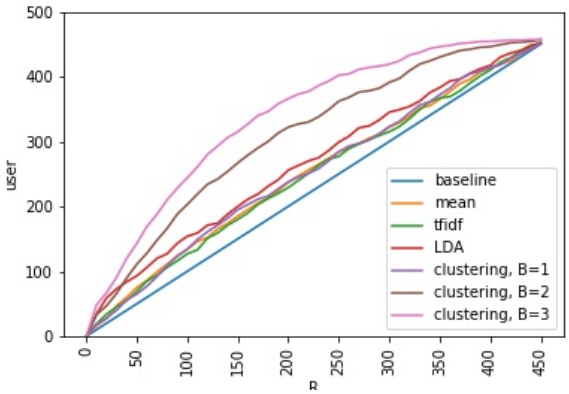

図 5　対となる地域情報が上位 R 位に選択された人数 (投稿数 250)

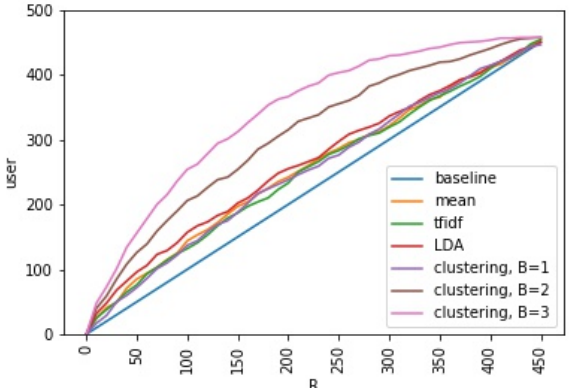

図 6　対となる地域情報が上位 R 位に選択された人数 (投稿数 100)

とが分かる．また， いずれの図においても提案手法において $B = 1$ とした場合は，w2v + tfidf，w2v + mean を用いた手法と比べて精度が高いか同程度，まれに低くなっているのに対して，LDA を用いた手法と比べて精度が低いか同程度になっている．更に，提案手法において $B = 2$ または $B = 3$ とする場合は他の手法よりも精度が高くなっている．具体的に，全投稿を用いて，$R = 100$ における結果を見たとき，提案手法において $B = 1$ とした場合は，w2v + tfidf を用いた手法と比べて約 11% 程，w2v + mean を用いた手法と比べて約 6% 程精度

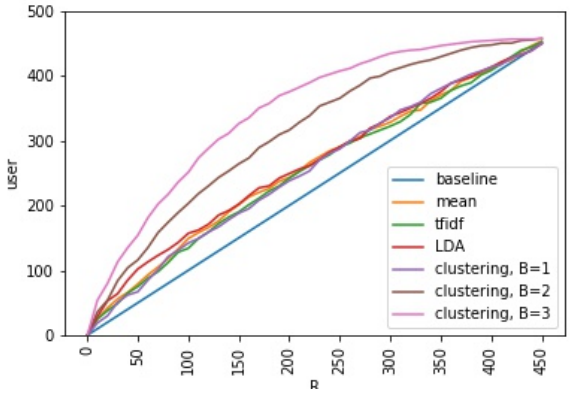

図 7　対となる地域情報が上位 R 位に選択された人数 (投稿数 50)

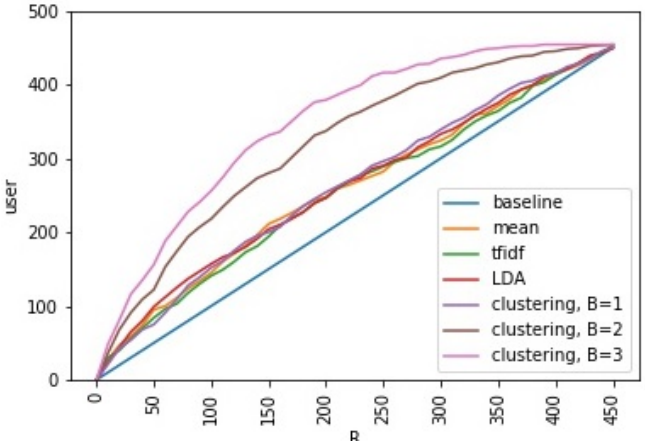

図 8　対となる地域情報が上位 R 位に選択された人数 (投稿数 25)

が高くなっているのに対して，LDA を用いた手法に対しては約 13% 程低くなっている．一方，提案手法において $B = 2$ とすると LDA を用いた場合と比べて約 27% 程度，$B = 3$ とすると約 63% 程度，比較手法よりも大きく精度が向上する．これは，今回用いた，各ユーザの正解データにあたる地域情報が，必ずしもユーザの最も関心を向けるトピックとは限らないためと推察される．比較手法では，ユーザの過去の投稿の単語を全て考慮し，ユーザの嗜好が様々なトピックを混合させたものとして表されるのに対し，提案手法では，ユーザの過去の投稿の中の単語を用いて，潜在意味空間上でのクラスタリングを行い，ユーザの嗜好を表すクラスタを複数選択する．このため，最も関心の高いトピックのみを考慮した $B = 1$ の場合は，比較手法よりも精度が下がり得るが，2 位以降のクラスタが表すトピックを考慮することにより，対となる地域情報がユーザの嗜好に近くなりやすく，かつ比較手法とは異なり，不要な単語を考慮しないため，精度が向上したと考えられる．

また，図 4 ～ 8 の比較により，利用する投稿の数を減らした場合においても提案手法と比較手法の相対的な関係は大きく変化しないことが分かる．具体的に，投稿を 1 ユーザ当たり 25 件用いて，$R = 100$ における結果を見たとき，提案手法において $B = 1$ とした場合は，w2v + tfidf を用いた手法と比べて約 6% 程，w2v + mean を用いた手法と比べて約 3% 程精度が高

くなっているのに対して，LDA を用いた手法に対しては約 6% 程低くなっている．一方，提案手法において $B = 2$ とするとLDA を用いた場合と比べて約 40% 程度，$B = 3$ とすると約 63% 程度，比較手法よりも大きく精度が向上する．また，投稿数を減らした場合においても，クラスタが形成されたユーザに対しては嗜好を推定することができていると言える．

3.3 トピックに応じたなりすまし投稿の生成

word2vec の学習と同様に，初めの 26 日間に収集された地域情報を含む投稿集合 T_k に対して，地域情報を表すベクトル表現 $\boldsymbol{f}_k$ を推定し，投稿生成器 G を学習した．学習に際しては，学習に寄与しづらい登場回数の少ない単語をラベルに変換し，いたずらな語彙数の増加を防ぐ．この際，全ての単語を同一のラベル（例えば “<UNK>”）にすると，学習データに同一のラベルが頻出するようになり，生成文にこうしたラベルが頻出するようになる．加えて，置き換えられる単語が本来果たしていた文法上の役割が分からない状態になるため，文の多様性を損なうことにも繋がる．そこで，置き換えるラベルを各単語の品詞を表すものにすることにより，同一ラベルの頻出を防ぎつつ，文の多様性を保つようにした．

また学習後，投稿を生成する際は，文章の多様性を担保するため，最初の単語の候補を確率順に 100 個選択し，その各単語に対して 2 文字目の単語の候補を 3 個，以降は候補を 1 個ずつ選択することにより，計 300 文の投稿を生成した．

実際のユーザに対してのなりすましでは，3.2 節により選択された地域情報を入力とし，なりすまし投稿を生成する．音楽，政治，スポーツ，食事などのトピックを持つ地域情報に対して投稿を生成した．表 5，6 に，各地域情報を表すローカル語，及び生成投稿の例を示す．両表からも分かるように，学習データからボットアカウントが多用する単語を含む投稿を除いたことにより，生成された投稿には求人などのボットアカウントの投稿に類似したものは生成されなかった．表 5 には生成された投稿の内，尤度が高いものから抜粋して記載している．いずれのトピックにおいても類似した投稿が生成されていることが分かる．一方で，表 6 には尤度が低い投稿を記載している．音楽，スポーツ，食事，その他のトピックにおいては，トピックに応じた投稿が生成されていることが分かり，生成時に一文字目二文字目を多めにとったことの意義が見て取れる．対して，政治のトピックにおいては，“#hillaryclinton” や “#nevertrumper” などのように，ハッシュタグにこそトピックに応じた文言が出現しているものの，文としてはトピックを表しているとは言い難い．更に，政治のトピックに対して，“reble” や “#music” という全く関係のない文言が出現している．これらは，トピック毎の学習データ量の差に原因があると考えられる．

また，スポーツのトピックの生成例における “oriole（野球チームの名称）” と “#soccer” や食事のトピックの生成例における “#vegan” と “#chicken” や “#seafood” の様に，同じトピックを表すが，使用されるべき文脈が明確に異なる単語同士が同一文中に出現するという問題が見られる．これらの矛盾は主にハッシュタグで発生しており，トピックに対するハッシュタグの学習が不十分であることが原因であると考えられる．更に，生成投稿はリアルタイムの情報を含まない．過去の投稿を模倣する手法で生成する投稿にリアルタイム情報を含めるため，生成後に現在の投稿から抽出した情報の追加が必要である．

4. ま と め

本研究では，マイクロブログ上の不特定多数のユーザによる地域情報に関する投稿を利用し，トピック応じた投稿の生成を可能とする生成器を学習すると同時に，なりすまし対象となるユーザの嗜好推定を行い，推定された嗜好に類似した地域情報を入力としてなりすまし投稿を生成することにより，対象ユーザの公開情報が少ない場合であってもなりすまし投稿を生成する手法を提案した．

アメリカ本土において 26 日間に投稿された投稿から地域情報に関する投稿を収集し，それらを用いて単語の分散表現及び投稿生成器を学習した．投稿から重要単語を選択することにより，ユーザの投稿数を減らした場合でも，高い精度で嗜好の推定が可能であることを確認した．同様に学習データの選別により，正規ユーザの投稿に類似し，かつトピックに応じた投稿が生成がされることを確認した．

一方で，ハッシュタグに関する学習不足などの問題が存在する．今後，これらの解決手法を検討し，嗜好推定の精度と共に，投稿の質について定量的な評価が必要である．

本研究の一部は，科学研究費補助金（基盤（S）16H06302，基盤（C）19K12019）の助成による．

文　献

[1] “Twitter,” https://www.Twitter.com.
[2] 坂本宏祐, 新田直子, 中村和晃, 馬場口登, “Twitter におけるユーザの嗜好及び地域情報を考慮したなりすまし投稿の自動生成” データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2019), 8 pages, 2019.
[3] J. Lim, N. Nitta, K. Nakmura, and N. Babaguchi, “Constructing Geographic Dictionary from Streaming Geotagged Tweets” , ISPRS International Journal on Geo-Information, Vol. 8, No. 5, 216, 24 pages, 2019.
[4] B. Hayes, “DeepDrumpf,” `http://www.deepdrumpf2016.com/`, 2017.
[5] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” Proc. International Conference on Learning Representations, 12 pages, 2013.
[6] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed Representations of Words and Phrases and their Compositionality,” Proc. International Conference on Neural Information Processing Systems, pp.3111—3119, 2013.
[7] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation-based anomaly detection,” ACM Transactions on Knowledge Discovery from Data, vol.6, no.1, 2012.
[8] O. Vinyals, A. Toshev, A. Bengio, and, D. Erhan, “Show and Tell: A Neural Image Caption Generator,” Proc. International Conference on Computer Vision and Pattern Recognition, pp.3156—3164, 2015.
[9] D. Blei, A. Ng, and M. Jordan, “Latent Dirichlet Allocaiton,” Jour. Machine Learning Reseach, vol.3, pp.1107—1135, 2003.
[10] C. Baziotis, N. Pelekis, and C. Doulkeridis, “Datastories at Semeval-2017 Task 4: Deep LSTM with Attention for

表 5　投稿の生成例（尤度上位）

トピック：音楽 (ローカル語:madison square)
just posted a photo @ madison square
i am at madison square in viacom, adele
first time of the year. #adele #madisonsquare #adeleconcert #livemusic
so excited to see the madison square! #newyork #madisonsquare #adeleinnyc #music
トピック：スポーツ (ローカル語:yankee stadium)
just posted a photo @ yankee stadium
i am at yankee stadium in nyc, yankee
first day of the year. #yankeestadium #mohegansun #theletsgoyankee #memorymaker
so much fun with my favorite people? @ yankee stadium
トピック：食事 (ローカル語:sea food city)
just posted a photo @ sea food city
i am at sea food city in city island, hennessey
first day of the year. #seafood #hennessey #seafoodcity #foodporn
so excited to be back in city island! #hennessey #seafoodcity #foodporn
トピック：政治 (ローカル語:white house)
just posted a photo @ white house
i am at white house in tennessee, dc
first time of the year. #flotus #thewashingtonconventioncenter #whitehouse #treble
so much fun with my favorite people? @ white house
トピック：その他 (ローカル語:empire state building)
just posted a photo @ empire state building
i am at empire state building in new york city, ny
first time of the year. #washington #empirestatebuilding #cwnyc #nofilter
so much fun with my favorite people? @ empire state building

表 6　投稿の生成例（尤度下位）

トピック：音楽 (ローカル語:madison square)
music festival. #madisonsquare #madisonsquaregarden #travel #madsquareeats #music
our favorite song of the year. #adeleconcert #adelelive2016 #madisonsquare #theadele
take the stage with my favorite band. #ufc205 #madisonsquare #livemusic
here i am at the adelelive2016! #madisonsquare #madisonsquaregarden #timelessmasterpiece215 #music
トピック：スポーツ (ローカル語:yankee stadium)
hey dodger! #yankeesstadium #yankeestadium #weareone #ymca #thenewyork
yankee stadium, bronx and yankee game. #lgm··· #nyc #familyfirst #yankeestadium
go oriole! #bowlegesco #yankeestadium #soccer #love #friends
take the final game to the yankee stadium. #pennantrace #yank #yankees #love
トピック：食事 (ローカル語:sea food city)
pork belly poutine. #hennessey #seafoodcity #brisket #mood··· #bbq
eat a #cityisland #vegan #chicken #seafood #foodporn
fresh chicken and waffles. #cityisland #seafoodcity #foodporn #foodie
dinner with my favorite people! #seafoodcity #hennessey #henny #lunch
トピック：政治 (ローカル語:white house)
hello to the white house. #dc #whitehouse #onobamawatch #love
never forget. #biweek #whitehouse #vote #whitehousefarm #thespeedway
first day of the #treble #lovethesetwo #whitehouse #nevertrumper #music
that was a great day. #whitehouse #uwffr16 #hillaryclinton #prisma
トピック：その他 (ローカル語:empire state building)
thanks to the jojo for the #bluejays #empirestatebuilding #themanhattan #humaneditors
view from the empire state building. #ny #empirestatebuilding #travel #usa #vacation
hello, grand central. #newyork #empirestatebuilding #miami #love #nyc
last weekend in 102nd floor. #bryantpark #empirestatebuilding #styleblogger #nofilter

Message-level and Topic-based Sentiment Analysis," International Workshop on Semantic Evaluation, pp.747—754, 2017.