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

坂本　宏祐† 　新田　直子† 　中村　和晃† 　馬場口　登†

† 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: †sakamoto@nanase.comm.osaka-u.ac.jp, ††{naoko,k-nakamura,babaguchi}@comm.osaka-u.ac.jp

あらまし　近年，ソーシャルネットワーキングサービス（SNS）上において，正規のユーザが公開しているプロフィール情報などを複製し，該当ユーザになりすまして不適切な発言や，他ユーザの個人情報の収集を行うといった問題が発生している．本研究では，マイクロブログである Twitter を対象に，正規のユーザが公開している投稿を用いたなりすましアカウントの投稿生成が可能か検証することにより，SNS への投稿の危険性を示し，ユーザの SNS の利用意識の向上を目指す．SNS の正規のユーザは実世界に存在するため，自身の嗜好に応じた各地の施設やイベントに関する投稿を行うことが多い．そこで提案手法ではまず，Twitter の不特定多数のユーザによる位置情報付き投稿から，各地の施設やイベントに関する投稿を抽出する．次に，抽出した投稿とユーザの過去の投稿に用いられる単語の意味的類似性に基づき，ユーザの嗜好に応じた施設やイベントを選択する．最後に，抽出した投稿から学習した，施設やイベントの潜在的な意味に応じた投稿パターンに基づき，選択した施設やイベントに関する投稿を生成する．

キーワード　マイクロブログ，テキスト生成，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）の普及により，手軽に世界中の人々とコミュニケーションが可能となった．特に代表的なマイクロブログの一つである Twitter [2] では利用者数が急増しており，2018 年 7 月時点で月間アクティブユーザ数は全世界で 3 億人を超えている [3]．しかし，中には，悪意のあるユーザが架空または実在の人物になりすましたアカウントが存在し，誹謗中傷を含む投稿や特定の団体を擁護する発言を行うといった問題が発生している．現在，このようななりすましアカウントの検出が盛んに研究されている一方，正規のユーザにより公開された情報を用いてなりすましアカウントの自動生成を試みることにより，SNS において個人情報を公開することの危険性を検証する研究も行われている．

例えば，Blige ら [4] は，正規ユーザのプロフィールを複製したなりすましアカウントを自動生成し，正規ユーザの友人への友人リクエストが高い承認率を得ることを示した．また，Coburn ら [5] は，SNS 上でグループを構成する正規のユーザが公開する投稿からそのグループが関心を持つトピックを推定し，グループ外のユーザによる推定トピックに関する投稿を複製することにより，プロフィールだけでなく投稿に関してもなりすましを行った．また，マイクロブログへの投稿が短文であることから，再帰的ニューラルネットワーク（RNN）のような深層学習を用いた文書生成技術の発展に伴い，過去の発言が大量に入手可能な著名人などを対象に，過去の発言を模倣した文書を投稿するなりすましアカウントも生成されている [6]．本研究ではさらに，著名でない一般ユーザに対しても，公開された少量の投稿を用いて，なりすましアカウントによる投稿生成が可能か検証する．

ここで，正規ユーザは実世界に存在するため，各地の施設やイベントに関する投稿を行うことが多い．これらの施設やイベントは，野球場はスポーツ，レストランは食事，ライブは音楽のようにそれぞれ固有のトピックを持つ．さらに，同じトピックを持つ各地の施設やイベントでは試合や食事の内容が投稿されるなど，施設やイベントの場所や時間は異なっても，投稿内容は類似したパターンを持つと考えられる．そこで，不特定多数のユーザからの投稿を利用し，このようなトピックごとの施設やイベントに関する投稿パターンを学習し，なりすましの対象ユーザの嗜好に応じた施設やイベントに関する投稿を生成する手法を提案する．

提案手法ではまず，投稿位置の緯度・経度を表すジオタグの付与された，不特定多数のユーザからの投稿を利用することにより，リアルタイムに各地の施設やイベントに関する情報（以下では単に地域情報と呼ぶ）を抽出する [1]．次に各地域情報に関する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，地域情報のトピックを推定する．さらに，抽出された多様な地域情報に対し，推定されたトピックと複数ユーザからの投稿の対を収集し，これらを用いてトピックに応じた投稿の生成モデルを学習する．なりすましアカウントの投稿生成の際には，対象ユーザの過去の投稿からユーザの嗜好を表すトピックを推定し，トピックの類似度に基づきユーザの嗜好に応じた地域情報を選択する．最後に，学習した投稿の生成モデルに対して，選択された地域情報のトピックを入力することにより，投稿が生成される．提案手法では，トピックに応じた投稿の生成モデルである再帰的ニューラルネットワークの学習に必要な大量の投稿は，不特定多数のユーザから収集しており，なりすましの対象ユーザに関する情報としては，嗜好を推定するために十分

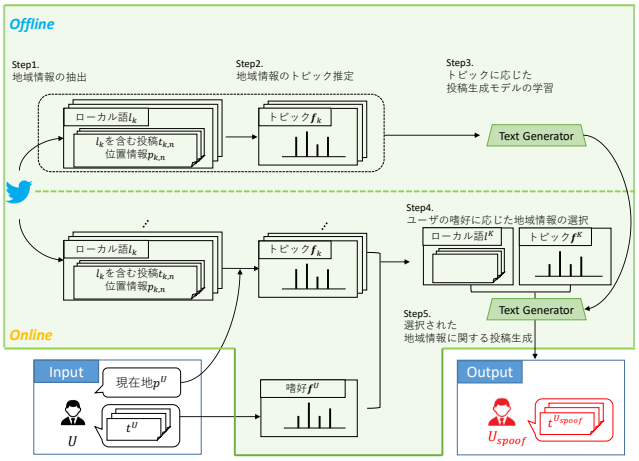

図 1　提案手法の概要

な数の投稿があればよい．

2　提 案 手 法

本研究では，常に不特定多数のユーザからのジオタグ付き投稿が得られるという前提のもと，これらのユーザのうち，なりすましの対象となる特定ユーザ U の嗜好に応じた地域情報に関する投稿 $t^{U_{spoof}}$ の生成を目的とする．提案手法は以下のように，不特定多数のユーザによる投稿からリアルタイムな地域情報を常時抽出しながら，オフライン処理としてトピックに応じた投稿生成モデルの学習，オンライン処理として特定ユーザ U の嗜好に応じた地域情報に関する投稿生成を行う．提案手法の全体図を図 1 に示す．

Step 1)　**地域情報の抽出**

ジオタグ付き投稿から，各地の地域情報を表す単語であるローカル語 $l_k(k = 1, 2, \cdots)$ と共に，ローカル語を含む投稿集合 $T_k = \{t_{k,n} \mid n = 1, \cdots, N_k\}$，及びそのジオタグ集合 $P_k = \{p_{k,n} \mid n = 1, \cdots, N_k\}$ を，地域情報を表すテキスト文，緯度・経度として収集する．

Step 2)　**地域情報のトピック推定**

Step 1) にて収集された T_k は，各地域情報のトピックを表す単語が含まれると考えられる．そこで T_k 中の単語集合を用いて，単語の潜在意味空間における各地域情報のベクトル表現 $\boldsymbol{f}_k$ を，トピック情報として算出する．

Step 3)　**トピックに応じた投稿生成モデルの学習**

Step 1) で抽出された多様な地域情報に対して，推定されたトピック $\boldsymbol{f}_k$ 及びテキスト文集合 T_k の対を収集する．これらを学習データとし，入力されたトピックに対して適切なテキスト文を生成するモデル $t_{k,n} = G(\boldsymbol{f}_k)$ を学習する．

Step 4)　**ユーザの嗜好に応じた地域情報の選択**

ユーザ U の過去の投稿 T^U に含まれる単語集合を用いて，単語の潜在意味空間におけるユーザ U のベクトル表現 $\boldsymbol{f}^U$ を，嗜好情報として算出する．投稿生成を行う時点において抽出されている地域情報から，最も $\boldsymbol{f}^U$ と類似度の高いトピック $\boldsymbol{f}^K = \arg\max_k sim(\boldsymbol{f}^U, \boldsymbol{f}_k)$ により表されるものをユーザの嗜好に応じた地域情報として選択する．

Step 5)　**選択された地域情報に関する投稿生成**

最後に，Step 3) で学習したモデル G を用い，$t^{U_{spoof}} = G(\boldsymbol{f}^K)$ によりユーザ U の嗜好に応じた地域情報に関する投稿 $t^{U_{spoof}}$ を生成する．

以下に各処理の詳細を述べる．

2.1　地域情報の抽出

Twitter に投稿される不特定多数のユーザによるジオタグ付き投稿からまず，各地の地域情報を表すローカル語を抽出する [1]．ローカル語は，ある特定の位置で観測される施設やイベントなどを表す単語であるため，空間上の特定の領域において集中的に用いられる．よって，空間的局所性の高い単語をローカル語として抽出すればよい．ただし，地域情報の人気度や変動性に基づき，局所性を観測できる時区間が異なるため，単語ごとの出現履歴に応じて適切な時区間で局所性を判定する必要がある．

具体的には，投稿を収集している地理空間全体を，各エリアの投稿数がほぼ均等となるよう J 個のエリア $A = \{a_j \mid j = 1, \cdots, J\}$ に分割する．ローカル語は地名や特産品，イベントなどの名称が多く，主に名詞から構成されると考えられるため，ジオタグ付き投稿が投稿される度，形態素解析により品詞のタグ付けを行い，名詞のみを抽出する．ここで，例えば “Michigan” という単語が州の名であるのに対し “Michigan Stadium” は施設名であるように，ローカル語を複合名詞として抽出することにより，示す場所や意味がより限定されると考えられるため，複合名詞を抽出する．

ジオタグ付き投稿が投稿されるごとに，投稿に含まれる Z 個の名詞 $u_z\,(z = 1, \cdots, Z)$ の出現履歴を更新する．u_z の局所性は，テキストマイニングの分野で，コーパス中の文書ごとに固有の単語を抽出するために用いられる TFIDF 法を応用し，以下の式により算出する．

$$tfidf_z^{max} = h_z^{max} \cdot idf_z \tag{1}$$

$$idf_z = \log \frac{J}{|A_z|}, where\ A_z = \{a_j \mid h_{z,j} \neq 0\} \tag{2}$$

ただし，$h_{z,j}\,(j = 1, \cdots, J)$ はエリア a_j における単語 u_z の出現頻度，$h_z^{max} = \max_j h_{z,j}$ は全エリアの中の u_z の最大出現頻度，$|A_z|$ は u_z が出現したエリア数，J は全エリア数を表す．また，出現頻度はユーザにつき各エリアで 1 回のみカウントする．

u_z が特定のエリアに頻繁に出現したときに $tfidf_z^{max}$ は高くなるため，$tfidf_z^{max} \geqq R$ を満たしたとき u_z をローカル語と判定する．さらに，イベントを表す語のような一時的に空間的局所性を持つローカル語は，それ以外の時区間ではどのエリアでも使用され得る．よって，長期間の出現履歴を保持していた場合，一時的な空間的局所性を検知することが困難となる．そこで $tfidf_z^{max} < r$ を満たしたとき，u_z の出現履歴を一度削除し，新たに収集し始めることにより，一時的な空間的局所性を正しく判定できるような時区間での出現履歴を保持するよう対処する．また，$tfidf_z^{max} \geqq R$ を満たした u_z をローカル語 l_k

表 1　ボットによる投稿

アカウント ID	緯度	経度	投稿
191092262	25.77508716	-80.2732666	News AIDS : New mouse model technology could speed the search for an AIDS vaccine URL
186478529	25.77508716	-80.2732666	News AIDS : New mouse model technology could speed the search for an AIDS vaccine: Researchers at Boston URL
186478529	25.77508716	-80.2732666	News AIDS : Researchers harness antibody evolution on the path to an AIDS vaccine: A vaccine needs to elicit those URL

として抽出するときも，u_z のこれまでの出現履歴を削除し，新たに収集を開始する．これにより，一度ローカル語として抽出された u_z が新たな出現履歴において $tfidf_z^{max} < r$ を満たしたとき，過去の一時的なローカル語であると判断できる．

さらに，Twitter では天気予報などの情報提供のため，機械により自動的に投稿するボットアカウントが多数存在する．表 1 に示すように，近接した位置から非常に類似した投稿を行う複数のボットアカウントが存在するとき，これらの投稿を構成する単語が，地域情報を表すものとして誤ってローカル語と判定され得る．そこで，u_z の出現履歴には位置情報のみでなく u_z を含む投稿履歴も記録し，非常に類似した投稿はボットアカウントによるものとして削除する．投稿の類似度は Jaccard 係数により算出するが，出現履歴中の投稿数が多くなるにつれ類似度の計算時間が増大するため，$tfidf_z^{max} \geqq R$，もしくは $tfidf_z^{max} < r$ を満たしたときのみ，投稿間の類似度を算出し，Th_e 以上の類似度を持つすべての投稿を出現履歴から削除した後，再度 $tfidf_z^{max}$ に基づくローカル語判定を行う．

以上の処理により，ローカル語 l_k を抽出すると共に，ローカル語と判定されたとき出現履歴に記録されているジオタグと投稿履歴を，地域情報を表すテキスト文 $T_k = \{t_{k,n} \mid n = 1, \cdots, N_k\}$，及び緯度・経度 $P_k = \{p_{k,n} \mid n = 1, \cdots, N_k\}$ として収集する．

2.2　地域情報のトピック推定

2.1 節にて収集された T_k は，各地域情報のトピックを表す単語を含むと考えられる．そこで T_k 中の単語集合を用いて，単語の潜在意味空間における各地域情報のベクトル表現 $\boldsymbol{f}_k$ を，トピック情報として算出する．

このため，まず単語の潜在意味空間への写像を word2vec [7], [8] により学習する．word2vec は，学習用コーパスとして大量の文書集合が与えられたとき，各単語の周辺単語を推定するような 2 層からなるニューラルネットワークを用いて，同じコンテキストで用いられる単語対ほど潜在意味空間において近くなるよう，各単語のベクトル表現を学習する．ここで，Twitter への投稿には特有な表現が多く用いられるため，学習用コーパスには Twitter への投稿集合を用いることが望ましい．特に，類似したトピックの地域情報に関する投稿で用いられる単語対が近くなるよう，T_k を 1 つの文書とみなし，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，地域情報のトピック推定に不要な単語として “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 $L = \{l_k \mid k = 1, 2, \cdots\}$ に含まれる単語は削除する．これにより，T に含まれる M 個の単語 $w_m (m = 1, 2, \cdots, M)$ に対する潜在意味空間におけるベクトル表現 $\boldsymbol{v}_m$ が得られる．

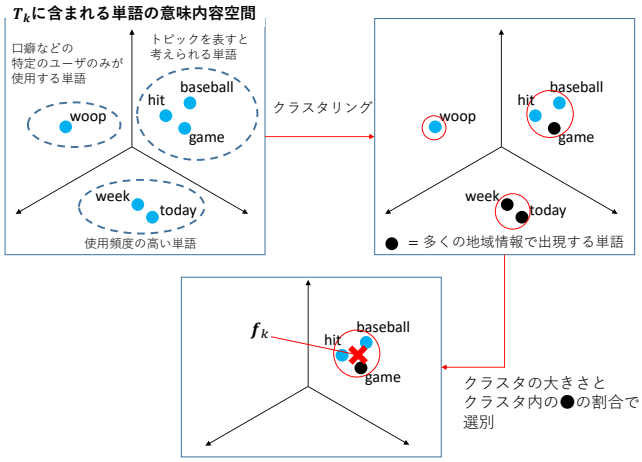

図 2　クラスタリングによる特徴量抽出

次に，学習した単語の意味表現を用いて，T_k 中の単語集合に基づき，各地域情報のベクトル表現 $\boldsymbol{f}_k$ を，トピック情報として算出する．ここで，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．一般には，2.1 節で述べたように，文書中の単語の関連性を測るために TFIDF 法が使われることが多いが，Twitter への投稿に頻出する口癖のような限られたユーザのみが使用する表現や，不特定多数のユーザにおける関心の偏り，各地域情報に対する投稿数の偏りなどにより，各文書中の単語の使用回数や，単語の使用される文書数に基づく TFIDF 法は有効に働かない．

ここで，各地域情報のトピックを表すと考えられる T_k では，トピックを表す意味的に類似した単語が複数使用されると考えられることから，図 2 のようにクラスタリングを用いて重要な単語の抽出を試みる．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．ただし $\boldsymbol{v}_m$ の類似度はコサイン類似度とする．クラスタリングには階層的クラスタリング（群平均法）を用いることにより，非常に類似した単語のみで形成されるクラスタを得ることができる．

ここで，“today” や “yesterday” など多くの地域情報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．そこで，多くの地域情報で使用される単語が 50%以内で，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

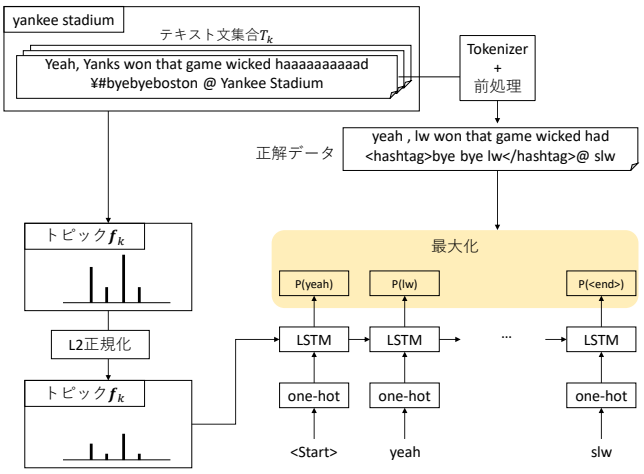

図 3　投稿生成モデルの学習

をトピックを表す単語のクラスタとして選択する．このクラスタ内の単語のベクトル表現 $\boldsymbol{v}_m$ の平均ベクトルを各地域情報のベクトル表現 $\boldsymbol{f}_k$ とする．

2.3　トピックに応じた投稿生成モデルの学習

類似したトピックを持つ各地の施設やイベントでは，各地の野球場では野球の試合の状況，各地のレストランでは食事の内容が投稿されるなど，施設やイベントの場所や時間は異なっても，投稿内容は類似したパターンを持つと考えられる．そこで，入力された特徴量に応じて適切な文章を生成するため，Image Captioning [11] などのタスクで使用される Long Short-Term Memory（LSTM）を使用したネットワークモデルを考える．このネットワークモデルでは，学習に用いるコーパスに含まれる各単語に個別の ID を割り当て，それぞれの単語を対応する ID のみを 1 とし他の要素を 0 とした one-hot ベクトル表記で表す．そして特徴量を LSTM の隠れ層への入力とする．そして，文の始まりを表す単語を最初の入力として，再帰的に次の単語の確率を計算し，正解ラベルとなる単語を出力する確率を最大化するように学習する．ここで，特徴量に対して出力すべき文章を正解データとして与えることにより，特徴量に対応した文章を出力するように学習される．

まず，2.1 節で抽出された多様な地域情報に対して，推定されたトピック $\boldsymbol{f}_k$ 及びテキスト文集合 $T_k = \{t_{k,n} \mid n = 1, \cdots, N_k\}$ の対を収集する，そして，$\boldsymbol{f}_k$ を隠れ層への入力とし，それに対する正解データとして $M_{k,n}$ 個の単語で構成される投稿 $t_{k,n} = \{w_{k,m} \mid m = 1, \cdots, M_{k,n}\}$ を用いる．ただし，類似したトピックを表す特徴量は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 は不特定多数のユーザによって投稿されたものであるため，文体の統一性がなく，学習の際に単語の種類が多くなり，学習が困難となる．そこで前処理として，Baziotis ら [13] が提案した Tokenizer を適用し，文を正規化する．ただし，文章を簡便にするため，正規化後の文章からユーザ名を表す “<user>”，ハッシュタグの始まりを表す “<hashtag>” 及びハッシュタグの終わりを表す “</hashtag>” 以外の正規化を表す単語及び全地域情報において出現回数が 5 回以下の単語は削除する．また，各地の施設やイベントに関する投稿では，トピックが類似する場合も，場所を表す単語は異なる単語が用いられる．そこで，地域情報自身のローカル語を表す単語は “slw”, その他の地域情を表す単語を “lw” といった同じトークンに置換することにより，文章中，ローカル語が使用される位置のみが学習されるようにする．これらの処理を投稿に適用した例を表 2 に示す．

以上により，図 3 のように入力されたトピック $\boldsymbol{f}_k$ に対して適切なテキスト文 $t_{k,n} \in T_k$ を生成するモデル $t_{k,n} = G(\boldsymbol{f}_k)$ が学習される．

2.4　ユーザの嗜好に応じた地域情報の選択

まず，ユーザ U の過去の投稿 T^U に含まれる単語集合に基づきユーザの嗜好を表すベクトル表現 $\boldsymbol{f}^U$ を算出する．2.2 節と同様に，T^U に含まれる単語集合を対象としたクラスタリングに基づき，重要な単語により形成されるクラスタを選択する．ここでは，投稿したユーザ数に基づくクラスタの選別はできないため，代わりに，クラスタ内の単語の使用頻度を用いる．また，ユーザの嗜好は野球，音楽及び政治というように，複数存在すると考えられるため，過去の投稿において各クラスタ内の単語の使用頻度順に上位 B 個のクラスタを選択した上で，各クラスタに対してベクトル表現 $\boldsymbol{f}_b^U (b = 1, 2, \cdots, B)$ を，嗜好情報とする．ユーザの各嗜好 $\boldsymbol{f}_b^U$ に対し，投稿生成を行う時点において抽出されている地域情報のうち，コサイン類似度の大きさに応じて確率的にユーザの嗜好に応じた地域情報を選択する．

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を $\boldsymbol{f}^K$ とし，2.3 節で学習したモデル G を用い，$t^{U_{spoof}} = G(\boldsymbol{f}^K)$ によりユーザ U の嗜好に応じた地域情報に関する投稿 $t^{U_{spoof}}$ を生成する．ただし，$t^{U_{spoof}}$ に含まれる地域情報のローカル語 l^K を示すトークン “slw” は，選択された地域情報のローカル語に置換する．また，“lw” は地域情報に関する投稿 T^K に含まれる l^K を除くローカル語 $l^{K'} \in T^K$ に変換する．この際に置換するローカル語 $l^{K'}$ は，T^K における出現回数に応じて選択する．

3　評 価 実 験

Twitter の Streaming API を用いてアメリカ本土を緯度が 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．これらを用い，まず 2.1 節の手法を用いて地域情報を抽出し，2.2 節の手法によりトピックを推定する．それらの結果について考察した後，抽出されたローカル語を含む投稿を行ったユーザに対して，2.4 節の手法により適切な地域情報を選択ができるか，また，選択された地域情報に対して，2.3 節により学習したモデルを用いて適切な投稿が生成されるか検証する．

3.1　ローカル語の抽出

アメリカ本土を各エリアの投稿数が均等となるよう $J = 64$

表 2　前処理の結果例（ローカル語:yankee stadium）

正規化前	Yeah, Yanks won that game wicked haaaaaaaaaad #byebyeboston @ Yankee Stadium URL
正規化後	yeah , lw won that game wicked had <hashtag> bye bye lw </hashtag> @ slw

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:asu）	クラスタ
ASU vs. Cal! Even though I'm an alum, this is actually my first time at a college football game,⋯ URL	football, game, beat,
Game day vs asu @ 4pm ?? #gameon #gameday #titansterritory #tusksup #csufwomenssoccer⋯ URL	watch, win, score
地域情報の投稿の一例（ローカル語:fedexfield）	クラスタ
Getting our game faces on! Going to see wvu vs byu at #redskins home. @ FedExField URL	game, football, win,
Gearing up for the season opener @steelers vs @redskins #GoSteelers @ FedExField URL	games, watch, beat
地域情報の投稿の一例（ローカル語:boone picken stadium）	クラスタ
game days are too fun with you girly gal. <;3 @ Boone Pickens Stadium URL	game, win, football,
Hook 'em am I right?! Just kidding GO POKES!!!!?? @ Boone Pickens Stadium URL	lose, cheer, college,

表 4　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:adele）	クラスタ
adele my favorite Song of the night !!! #dontyouremember #adele #adelelive #ilovethissong #tears⋯ URL	music, singer, singing,
Very fitting song to end an amazing concert #Adele #adelelive2016 #adeleconcert #nyc⋯ URL	songwriter, song, sing
地域情報の投稿の一例（ローカル語:atl）	クラスタ
My favorite part of music midtown #twentyonepilots #musicmidtown #atl #ride twentyonepilots⋯ URL	music, mic, indie,
Listening to the best of #Uptop at the hands of @djshaolin at pyramidslounge in #ATL	songwriter, musician
地域情報の投稿（ローカル語:kaaboo）	クラスタ
Group love playing Beast Boys, Sabotage ! #kaaboo2016 #sandiego #friends @ KAABOO URL	music, stage, set, band,
@JimmyBuffett Maybe play Captain America tonight. John Lovell on kazoo at Kaaboo.	guitar, playing, rockin

エリアに分割し，各投稿に対して，Brill's Tagger [9] による品詞のタグ付け，TermExtract [10] による複合語の抽出を行った．得られた全ての名詞に対して，$R = 4.52$，$r = 1.28$ として逐次的にローカル語を抽出した結果，30 日間のすべての投稿を処理したときの地域情報データベースには 61,091 語のローカル語が存在した．ただしパラメータは，[1] の結果を踏まえて設定した．次に，地域情報のトピックを推定するために初めの 26 日間に抽出された地域情報 52,908 個の全投稿 594,994 件を学習用コーパスとして word2vec を用いて単語の意味表現を学習した．学習した単語の意味表現を用いて，同じく 26 日間に抽出された地域情報に対して 2.2 節の手法により各地域情報のトピックを表す単語のクラスタを取得した．表 3, 4 に取得された地域情報及びそのトピックを表す単語のクラスタを示す．ただし，投稿本文中の "URL" は "http(s)://..." という形式の URL である．表 3，表 4 の各クラスタからそれぞれサッカー，音楽のトピックを持つ地域情報が取得できていることがわかる．また，それぞれのトピックに対して，サッカーであればその試合内容，音楽であれば演奏されている楽曲の内容など，トピックに応じて一定の投稿パターンが存在していることが確認された．よって，これらの地域情報の推定されたトピック及びその投稿の対を用いて，2.3 節の手法により投稿生成器が学習可能であると考えられる．

3.2　地域情報の選択結果の評価

地域情報の選択手法の評価を行うため，3.1 節で取得したローカル語の内，まず初めの 26 日間に抽出された地域情報 52,908 個の全投稿 594,994 件を学習用コーパスとして word2vec により単語の意味表現を学習した．次に 28 日間に抽出された地域情報 56,425 個の投稿集合の内，学習用コーパスとして用いられていない 28 日目の投稿を行ったユーザとその地域情報の対をランダムに 1,000 個選択した．そして，ユーザが地域情報に関連する投稿を行う以前の投稿を 500 件から 3,000 件取得し，この投稿を用いて，全ユーザに対して，そのユーザの対となる地域情報が 1,000 位中上位 R 位に選択される人数を確認することにより，ユーザの嗜好に応じた地域情報が正しく選択されるか評価した．

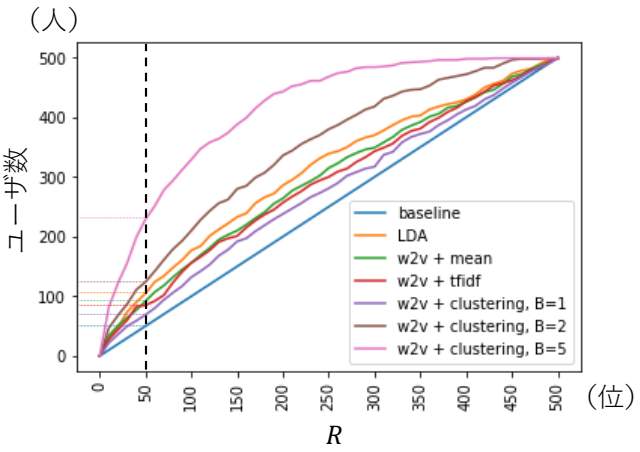

図 4　対となる地域情報が上位 R 位に選択された人数

ここで，ユーザの嗜好及び地域情報それぞれのベクトル表現 $\boldsymbol{f}^U$，$\boldsymbol{f}_k$ の算出方法として，T^U，T_k それぞれに含まれる全単語の総和平均をとる手法 (word2vec + mean)，各単語に対する TFIDF 値を重みとして，全単語の重み付き総和平均をと

表 5　地域情報 believeland について実際に投稿したユーザの過去の投稿，及び believeland に関する複数ユーザからの投稿

ユーザの過去の投稿の一例
Brew and lunch while biking around #StanleyPark with the w... (Stanley Park 1897 Amber Ale) URL #photo
This is beer is much better than the play of the #Cleveland #Browns #... (Scarlett Red Rye) URL #photo
Trophy is coming east #believeland #nbafinals - Drinking a Go West! IPA @ Bridgewater Clubhouse - URL #photo
Not much wrong with the Ale - Drinking a Smoked Spliff Ale by @divingdogbrew @ Bacon and Beer Classic — URL #photo
Can't believe how he was able to stick with curry at times late in the game. Effort goes a long way. URL
Watching some of this #DemDebate during college football commercial breaks and feel that @MartinOMalley is making a great impression
地域情報に関する投稿の一例
2nd last home game of the year. Great seats! #Believeland #Windians @ Progressive Field -⋯ URL
ARE YOU READY!!!???? #mlbplayoffs #2016 #Rallytogether #Believeland #YourTurn #GoTribe @⋯ URL
Nice and moist out here this morning ?? #golfislife #golflife #golf #believeland @ Ridge Top⋯ URL

表 6　各手法による地域情報 believeland の順位

	w2v + mean	w2v + tfidf	LDA	w2v+clustering (B=1)	w2v+clustering (B=2)
R	238	265	94	125	10

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) [12] 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 に，提案手法においてユーザの嗜好数 $B = 1, 2, 5$ とした場合と，各比較手法の結果を示す．いずれの手法も無作為に選択した baseline に比べ，ユーザが実際に投稿した地域情報が正しく選択されたことが分かる．提案手法において $B = 1$ とした場合は，w2v + tfidf，w2v + mean，LDA を用いた手法の方が精度が高いが，提案手法において $B = 2$ とすると，比較手法よりも大きく精度が向上する．これは，今回実験データとして用いた，各ユーザに対する正解データとした地域情報が，必ずしもユーザの最も関心の高いトピックに合致したものでは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に使用される単語を全て考慮するため，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から各トピックに関連する単語のみを選択した上でユーザの嗜好を表すベクトル表現を算出する．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 位以降のトピックを考慮することにより，不要な単語を削除し，嗜好を表す単語のみを抽出する提案手法が有効に働いたと考えられる．

例として表 5 に，地域情報 believeland について実際に投稿したあるユーザの過去の投稿，及び believeland に関する複数のユーザによる投稿を，表 6 に，各手法により，対象となる 1,000 個の地域情報を順位付けした際の believeland の順位を示す．表 5 に示すように，ユーザの過去の投稿には飲食及びスポーツに関連したものが多い．また believeland に関する複数のユーザからの投稿にはスポーツに関連した単語が多いことが分かる．表 6 に示すように，提案手法で $B = 1$ としたときは，飲食関連がユーザの最も関心のある対象として推定されたため，飲食やスポーツを混合したトピックをユーザの嗜好として推定する比較手法より，believeland の順位が低くなった．しかし，$B = 2$ としたとき，2 番目に関心のあるトピックとしてスポーツに関連した単語で構成されるクラスタが選択されたため，believeland とのトピックの類似度を，スポーツに関連する単語のみで算出することができ，順位が大きく向上した．

3.3　トピックに応じた投稿の生成

地域情報のトピックに応じた投稿を生成するため，3.1 節で初めの 26 日間に抽出された地域情報の投稿集合 T_k と $\boldsymbol{f}_k$ の対を収集し，投稿生成器 G を学習する．T_k にはローカル語 l_k を含む投稿が収集されているが，すべての投稿がその施設やイベントの体験に関するものとは限らない．そこで，できるだけ意味的に関連したもののみを投稿事例として用いるため，$\boldsymbol{f}_k$ の算出時に選択された単語が 2 人以上に用いられていた場合，それらの単語を含む投稿のみを各 T_k から最大 5 件ずつ選択した．収集された合計 69,216 対のうち，80%の 55,372 件を学習データ，残り 13,844 件を検証データとして学習を行った．ただし，全地域情報の投稿集合において出現回数が 5 回以下のものを “<UNK>” と変換した．学習した生成器により投稿を生成する際は，より多様な文章を生成するため，最初の単語の候補を確率の高いものから 50 個選択した後，各単語に対して 2 文字目の単語の候補を 2 個，以降は候補を 1 個とすることにより，計 100 文の投稿を生成した．

比較手法として，トピックに応じた投稿生成を行う Coburn ら [5] の手法を用いた．これは，同じトピックに関する投稿集合が与えられた際，頻出する単語がトピックを表すという前提のもと，この単語を含む任意の投稿を複製する手法である．ここでは，対象となる地域情報 T_k に最も頻出する単語を含む任意の投稿を Twitter からランダムに 100 件検索し，2.3 節と同様に正規化したものをなりすまし投稿とした．生成された投稿の評価指標として，地域情報 T_k 中の投稿を正解文とした BLEU [14] を用いる．27 日目以降に抽出された地域情報に関する投稿集合

表 7　投稿の生成例 (トピック:音楽)

ユーザの過去の投稿の一例	選択された地域情報	
And congratulations to a true artist and musician Bradley Cooper, nominated alongside me. I couldn't be more proud… URL	1	madison square
I ' m humbled & grateful that my album " Joanne " was nominated & also my song " Million Reasons. " Thank u so much Monst… URL	2	insustrybar
	3	uadnyc
This was the 1st time I ever played Sonja the song I wrote about and FOR her #GrigioGirls This was her last birthday URL	4	rockwoodmusichall
	5	bowery electric
生成されたなりすまし投稿の一例 (ローカル語:madison square)		
just posted a photo @ madison square		
we are #hiring ! click to apply : <UNK> <UNK> - #<UNK> #throwback		
at the madison square . #sweden #madisonsquare #thank #music #<UNK>		
my favorite song of the #adele #madisonsquare #<UNK> #iloveilluminatemsg		

表 8　投稿の生成例 (トピック:スポーツ)

ユーザの過去の投稿の一例	選択された地域情報	
I like how Demaryius Thomas tried to sneak the football over to the 1st down marker during that fight. #aintcheatinainttryin	1	autzenstadium
Behind the Scenes — Chris Betts — Student Sports Baseball — Sweat Machine URL	2	autzen stadium home of the
	3	university of oregon
Checking out "Garvey & Gretsky: Big Sports Names Chasing Big League Dreams" on The Baseball Network: URL	4	uoregon
	5	cuthbert amphitheater
生成されたなりすまし投稿の一例 (ローカル語:autzenstadium)		
just posted a photo @ autzenstadium		
can you recommend anyone for this #job ? autzenstadium - #autzen #sunset #autzenstadium , plw		
first game of the season ! #football #goduck #autzenstadium #<UNK>		
best game ever ! #autzenstadium #goduck #autzenstadium #goautzen #sunset		

表 9　投稿の生成例 (トピック:食事)

ユーザの過去の投稿の一例	選択された地域情報	
Lunch!! (@ Triumph Brewing - @triumphnewhope in New Hope, PA) URL	1	uws
Wow it seems to be a little warm out. I guess I'll start the grill to cook dinner.	2	158th
	3	amyruths
I just ousted @sarahtomko as the mayor of Yummy Sushi on @foursquare! URL	4	seasoned vegan
	5	fishtag
生成されたなりすまし投稿の一例 (ローカル語:amyruths)		
just posted a photo @ amyruths		
(@ amyruths in harlem , amy ruth)		
if you are looking for work in #amyruths , newyork , check out this #job :		
best friend ! #nyc #amyruths #<UNK> #foodporn		

表 10　投稿の生成例 (トピック:政治)

ユーザの過去の投稿の一例	選択された地域情報	
The most important way to stop gangs, drugs, human trafficking and massive crime is at our Southern Border. We need… URL	1	white house
"It should not be the job of America to replace regimes around the world. This is what President Trump recognized i… URL	2	georhia brown
	3	black women s agenda
I just had a long and productive call with President <user> of Turkey. We discussed ISIS, our mutual involveme… URL	4	timkaine
	5	john lewis
生成されたなりすまし投稿の一例 (ローカル語:whitehouse)		
just posted a photo @ white house		
can you recommend anyone for this #job ? white house - #whitehouse #un , obama		
had a great time at the white house ! #<UNK> #sxsl #music		
so excited to be a <UNK> <UNK> . #whitehouse #obama #champion		

表 11　$BLEU_n$ の値

	$BLEU_1$	$BLEU_2$	$BLEU_3$	$BLEU_4$
Realboy	0.1556	0.0352	0.0131	0.0064
提案手法	0.4255	0.2889	0.2032	0.1398

T_k をランダムに 50 個選択し，Coburn らの手法及び提案手法により生成した投稿 100 件ずつに対する $BLEU_n(n = 1, \cdots, 4)$ の値の平均値を表 11 に示す．ただし n は連続した n 個の単語列を対象とした評価値であることを示す．いずれの BLEU の値においても提案手法が大きく上回っていることが確認された．

3.4　なりすまし投稿の生成

最後に，実際のユーザに対して 3.2 節により選択された地域情報に対し，3.3 節と同様になりすまし投稿を生成した結果を示す．投稿中に音楽，政治，スポーツ，および食事のトピックを持つと思われるユーザをそれぞれ 1 人選択し，実験時における最新の投稿から 500 件〜最大 3,000 件を収集した．各ユーザに対して現在地を表す緯度経度情報を与え，その半径 5km 以内の地域情報を候補として選択された地域情報に対して投稿を生成した．表 7 ～ 9 に，なりすましを行う対象の各ユーザの過去の投稿，及び選択された地域情報の上位 5 つ，そのうち選択された 1 つの地域情報に対して生成された投稿例を示す．各ユーザの嗜好に応じた適切な地域情報が上位に選択されていることが分かる．また，表 7〜8 より，音楽やスポーツ，食事などのトピックに関してはトピックに応じた投稿が生成されたことがわかる．一方，表 9 に示される政治のトピックに関しては，“#music” や “so excited” など，政治とは関連性が低い単語を含む投稿が生成されている．これは，学習データに音楽のトピックを持つ地域情報が政治のトピックを持つものに比べ非常に多く存在するといった，学習データ中のトピックの偏りによるものと考えられる．さらに，どのトピックにおいても “just posted a photo” から始まる文章，トピックとは無関係と思われる求人関連のような投稿が生成されている．これは，学習用データ中にこれらの形式が非常に類似し，機械的に投稿されるものが，他の自由な形式の投稿に比べ，比較的多く存在したためと考えられる．さまざまなトピックに対して適切な投稿を生成するためには，学習データ中のこれらの投稿の偏りへの対処が必要である．

4　ま　と　め

本研究では，マイクロブログからユーザの嗜好に応じた地域情報をリアルタイムに抽出し，その地域情報のトピックに応じた投稿を生成することにより，公開された情報が少ないユーザに対して，実世界との現在の状況を踏まえた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 30 日間に投稿された緯度・経度付き投稿に対し，地域情報を表すローカル語とローカル語を含む投稿を抽出し，その投稿パターンを学習後，ユーザのなりすまし投稿を生成することにより，複数のユーザから投稿事例を多く収集できた地域情報に対しては，なりすまし対象のユーザの情報が少量でも，嗜好に応じた地域情報に関するなりすまし投稿が生成できることを確認した．一方で，投稿事例が少ないトピックに対しては，適切な投稿を生成することができなかった．今後の課題として，学習に用いる投稿集合において頻出する同形式投稿の削除や各トピックにおける学習データの偏りの解消が挙げられる．

本研究の一部は，科学研究費補助金（基盤（S）16H06302）の助成を受けたものである．

文　献

[1] T. Kamimura, N. Nitta, K. Nakmura, and N. Babaguchi, “On-line Geospatial Term Extraction from Streaming Geotagged Tweets,” Proc. International Conference on Multimedia Big Data, pp.322–329, 2017.
[2] “Twitter,” https://www.Twitter.com.
[3] “2018 年「公表データ」で見る主要 SNS の利用者数と，年代別推移まとめ,” https://appbu.jp/share-of-social-media.
[4] L. Blige, T. Strufe, D. Balzarotti, and E. Kida, “All Your Contacts Are Belong to Us: Automated Identity Theft Attacks on Social Networks,” Proc. International Conference on World Wide Web, pp.551–560, 2009.
[5] Z. Coburn and G. Marra, “Realboy: Believable Twitter Bots,” http://ca.olin.edu/2008/realboy/index.html, 2011.
[6] B. Hayes, “DeepDrumpf,” http://www.deepdrumpf2016.com/, 2017.
[7] 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.
[8] 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.
[9] E. Brill, “A Simple Rule-based Part of Speech Tagger,” Proc. Workshop on Speech and Natural Language, pp.112-116, 1991.
[10] “Term Extract,” http://gensen.dl.itc.u-tokyo.ac.jp.
[11] 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.
[12] D. Blei, A. Ng, and M. Jordan, “Latent Dirichlet Allocaiton,” Jour. Machine Learning Reseach, vol.3, pp.1107-1135, 2003.
[13] C. Baziotis, N. Pelekis, and C. Doulkeridis, “Datastories at Semeval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis,” International Workshop on Semantic Evaluation, pp.747-754, 2017.
[14] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “BLUE: a Method for Automatic Evaluation of Machine Translation, ” Annual Meeting of the Association for Computational Linguistics, pp. 311–318, 2002.