

# 複数のソーシャルメディアアカウントの 関連付け防止システムの構築

木原 裕二<sup>1</sup> 笹野 遼平<sup>1</sup> 高村 大也<sup>1</sup> 奥村 学<sup>1</sup>

**概要:** ソーシャルメディアのユーザの中には複数のソーシャルメディアのアカウントを関連付けずに使用しているユーザがいる。しかし、特徴的な表現や話題などから、それらの複数のアカウントが同じユーザにより作成されたものであると第三者に推測されてしまう場合がある。本研究では、まず、Twitter とブログの著者の同一性推定システムを構築することにより、どのような特徴がこれらのアカウントが同じユーザにより作成されたものと判断される要因となるのかを明らかにし、そこから得られた知見をもとにした関連付け防止システムの構築を行う。

## Construction of a System to Prevent Identification of User Accounts on Different Social Media Platforms

YUJI KIHARA<sup>1</sup> RYOHEI SASANO<sup>1</sup> HIROYA TAKAMURA<sup>1</sup> MANABU OKUMURA<sup>1</sup>

**Abstract:** Some social media users do not link their accounts on other social media platforms. However, the user identity is sometimes revealed by others due to the characteristics of expressions and topics in the posts. In this paper, we first examine what kinds of characteristics can be a clue for user identification by constructing a system for identifying user accounts on Twitter and blogs, and then propose a system to prevent identification of user accounts on different social media platforms on the basis of the findings from the user identification system.

### 1. はじめに

ソーシャルメディアのユーザの中には複数のソーシャルメディアのアカウントを関連付けずに使用しているユーザがいる。しかし、それらの複数のアカウントが同一のユーザによって作成されたものであると第三者に推測されてしまい、ユーザが明らかにしたくなかった個人情報や嗜好が第三者に知られてしまう場合がある。例として図1に示すような複数のアカウントを持つユーザの投稿があった場合を考える。これらの投稿はいずれも面接に関する投稿であるが、実名アカウントでは第三者に知られても問題のない内容が投稿されているのに対し、匿名アカウントでは第三者、特に面接官に知られてしまうと問題のある内容が投稿されている。このような場合、実名アカウントと匿名アカウントを関連付けられてしまうとユーザが知られたくな

#### 実名アカウント

2013/5/14

桜木町での最終面接終了！面接官がとても良い人で良かった！

#### 匿名アカウント

2013/5/14

桜木町での面接、第一志望ととりあえず言っておいた。どうせホントかウソか分からないだろうし。

図1 同一ユーザによる実名アカウントと匿名アカウントの投稿例。

かった情報を面接官に知られてしまうこととなる。

本研究では、このようなユーザが意図しない情報流出を防ぐため、複数のソーシャルメディアアカウントの関連付け防止システムの構築を行う。具体的には、互いに関連付けを行っていないマイクロブログとブログのアカウントを

<sup>1</sup> 東京工業大学, Tokyo Institute of Technology

持つユーザを想定し、新たにブログ記事を投稿しようとした際に、第三者がそのブログアカウントをマイクロブログアカウントと関連付けるための手掛かりが含まれていないかを判定し、そのような手掛かりが含まれていた場合は警告メッセージを表示した上で、関連付けの手掛かりとなる情報が何であるかを特定しマスキングするシステムを考える。

このようなシステムを構築するためには、これらのアカウントがどのような特徴から関連付けられてしまうかを明らかにする必要がある。そこで、まず、マイクロブログのアカウントが与えられた場合に、大量のブログの中から同一のユーザが作成したブログを推定するシステムの作成に取り組む。具体的には、ユーザ自身により関連付けられた代表的なマイクロブログサービスである Twitter<sup>\*1</sup>とブログのペアを教師データとした教師あり学習により、候補となるブログを、与えられた Twitter アカウントのユーザと同一のユーザが作成したと考えられる順にランキングするシステムを作成する。続いて、作成したシステムの分析を通して、どのような特徴が複数のソーシャルメディアのアカウントが同じユーザにより作成されたものと推定される要因となるのかを明らかにし、そこから得られた知見をもとにした関連付け防止システムの構築を行う。

## 2. 関連研究

著者同一性推定に関する研究は多く行われてきたが、これらの手法は大きく2つに分けることができる [1]。1つは各著者を1クラスとした多クラス問題として著者推定タスクを扱う手法 [2][3][4] である。これらの研究では事前に推定対象となる著者候補が既知であり、各著者ごとの学習データが入手可能であることを想定している。

著者推定タスクに対するもう1つの手法は類似度に基づく手法であり、これらの研究では著者推定を、1つの文書が与えられた場合にそれと最も類似する文書と関連付ける問題として扱っている。たとえば、Qian ら [5] は各ユーザを1つのクラスとしたモデルを作成するのではなく、2つのアカウントが入力された場合に、それらのアカウントが同一のユーザによって作成されたものかどうかを判別する枠組みを提案している。本研究でも基本的に2つのソーシャルメディアのアカウントの類似性に着目した手法を提案する。

異なるソーシャルメディア間の著者同一性推定に取り組んだ研究として Korayem らの研究 [6] がある。Korayem らは Twitter と代表的な画像共有サイトである Flickr<sup>\*2</sup> との間の著者同一性推定タスクに対し、アクセスパターンや、テキスト情報、位置情報タグ等を素性として用いた機械学習に基づく手法を提案している。

本研究で提案する Twitter とブログの著者の同一性推定

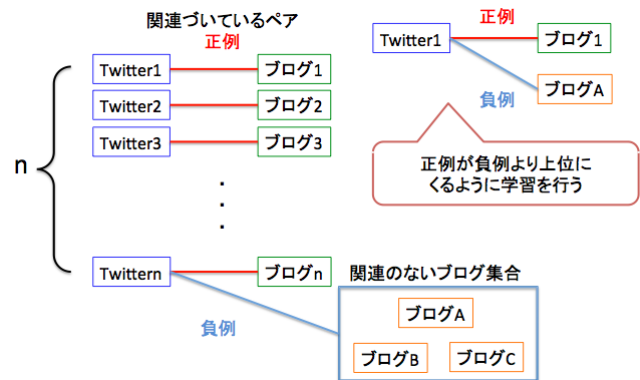


図2 ランキング学習の概要

システムでは、既存の著者推定タスクと異なる点が2点ある。1点目は異なるソーシャルメディアで近い時間に投稿された内容は類似していることを手がかりとして利用している点である。Twitter ではリアルタイムで投稿されることが多く、ブログは1日の終わりに日記の形式で投稿されることが多いと考えられることから、同一ユーザによる投稿であれば、ブログの投稿内容に関連する語がそれ以前に投稿された Twitter にも出現すると考えられる。Korayem らも投稿時間と位置情報タグを用いた素性を使用しているが、テキスト情報も用いた素性は使用していない。2点目は、既存の研究では主に著者の文体に着目し、機能語やスペルミス等を著者の特徴として用いていたのに対して、本研究では主に内容語に着目している点である。これは Twitter とブログでは同じ内容について投稿されることが考えられるためである。

著者同一性推定の防止に関する研究としては Kacmarcik らの研究 [7] がある。Kacmarcik らは各著者の使用する機能語の頻度に着目し、それらの機能語の置換を行うことで、著者推定の精度を大きく下げることが可能なことを報告している。本研究では機能語だけではなく、内容語にも着目し、これらの単語をマスキングすることで著者推定の精度を下げることを可能であることを示す。

## 3. Twitter とブログの同一性推定システム

### 3.1 同一性推定システムの概要

本節では、1つの Twitter アカウントを入力とし、同一のユーザが作成したブログを含むブログ集合を、同一のユーザが作成したと考えられる順にランキングするタスクを考える。教師データとしてはユーザ自身により関連付けられた  $n$  組の Twitter とブログのアカウントを使用する。ランキング学習の概要を図2に示す。Twitter アカウントとブログアカウントの関連付いているペアを正例、Twitter アカウントと関連のない各ブログアカウントとのペアを負例とし、正例が負例より上位にくるように、ランキング SVM [8] に基づくランキング学習を行う。

\*1 <https://twitter.com/>

\*2 <http://www.flickr.com/>

## 3.2 素性

SVMの学習に用いる素性には、大きく分けて、類似度に基づく素性、ユーザ固有な表現に基づく素性、投稿時間に基づく素性の3つの素性を用いる。

### 3.2.1 類似度に基づく素性

ペアとなるTwitterとブログ、それぞれを1つの文書とみなし、それらの文書間で定義された以下の3つの類似度を素性として使用する。

- (1) IPA辞書 [9] で定義されている69個の品詞細分類<sup>\*3</sup>ごとに、Twitterおよびブログそれぞれで出現した形態素の頻度を要素とする形態素ベクトルのJaccard係数。
- (2) Twitterおよびブログそれぞれで出現した内容語<sup>\*4</sup>のtf-idf値を要素とする形態素ベクトルの余弦類似度。
- (3) Twitterおよびブログそれぞれで出現した記号列のtf-idf値を要素とする形態素ベクトルの余弦類似度。

### 3.2.2 ユーザに特有な表現を考慮した素性

Schwartzら [4] はTwitterを対象としたユーザの同一性推定タスクにおいて、ある特定のユーザ1人だけが使用する文字列が同一性推定の有力な手掛りとなったと報告している。そこで本研究でもあるユーザに特有な表現を考慮した素性を用いる。ただし、Schwartzらの研究における実験設定では事前に対象となるユーザの集合が既知であることから、ある表現がそのユーザに特有な表現であるかどうか判別できるのに対し、本研究では判別対象とするユーザの母集団が既知であることを仮定していないことから、ある表現がそのユーザに特有な表現であるかどうかは判別できず、Schwartzらが使用した素性をそのまま使用することはできない。そこで本研究では以下の手順で作成した素性をユーザに特有な表現に基づく素性として使用する。

- (1) あるTwitterデータ中に2回以上出現した形態素のうち学習に使用する他の $n-1$ 個のTwitterアカウントで一度も使用されていない形態素をそのTwitterアカウントに特有な表現とみなす。
- (2) それらの表現のうちランキング対象のブログ中で2回以上出現した表現の数を、そのTwitterとブログのアカウントのペアに対する素性として使用する。

たとえば、1,000人のTwitterユーザの中で1人のユーザのみが、複数回使用する形態素が5つある時、それらの形態素はそのユーザに特有な表現であると考えられる。その上で、それら5つの形態素のうち、ランキング対象とするアカウントのブログに2回以上出現した形態素の数を、ユーザに特有な表現を考慮した素性として使用する。

### 3.2.3 投稿時間を考慮した素性

Twitterやブログにはユーザがその日に体験した出来事

が投稿される場合が多いと考えられることから、近い時間に同一のユーザにより投稿されたTwitterとブログには、その出来事に関連する語が出現している可能性が高いと考えられる。そこで投稿時間を考慮した素性として、近い時間に投稿されたTwitterとブログで共起する固有名詞に関する素性を導入する。

具体的には、あるTwitterとブログのアカウントのペアが与えられた場合、そのブログに含まれる各記事ごとに、ブログ投稿の直前24時間の間に投稿されたTwitterの全投稿をまとめたものを1文書とみなし、対象のブログとの類似度を算出し、全ブログ記事で平均を取った値を素性とする。Twitterのテキストとブログの類似度は7つの固有名詞の細分類<sup>\*5</sup>ごとに算出する。ただし、本素性は対象の語が共起したかどうかを重視するため、類似度の計算に使用する形態素ベクトルの各要素の値は頻度ではなく、出現した場合に1、出現しなかった場合に0となるような2値とし、類似度としては余弦類似度を用いた。

## 4. Twitterとブログの著者同一性推定実験

### 4.1 実験に用いるデータ

本実験の目的は、複数のソーシャルメディアのアカウントを関連付けられたくない場合に、どのような特徴からこれらのアカウントを関連付けられてしまうかを分析することである。このため、実験に使用するアカウントは、ユーザ自身が関連付けを行っていない2つのソーシャルメディアのアカウントであることが望ましい。しかし、そのようなデータを大規模に収集することは困難であることから、モデルの学習、および、大規模な評価実験にはユーザ自身により関連付けられたアカウントペアを疑似的なデータとして用い、ユーザ自身により関連付けが行われていない少数のアカウントペアを用いて実用的なシステムの精度の検証を行う。

#### 4.1.1 ユーザにより関連付けられたアカウントペア

本実験では、Twitterのプロフィール欄にアメーバブログ<sup>\*6</sup>へのリンクが張られている場合に、そのTwitterとブログのアカウントの組を同一のユーザにより作成されたものとみなし、実験データとして使用する。Twitterやブログにはプロフィール欄やアカウント名などのユーザの同一性の推定の手掛りとなりうる様々な情報が含まれていると考えられるが、本研究ではTwitterとブログの投稿から得られる情報のみをユーザの同一性推定に使用する。これは、実際に複数のアカウントを関連付けられないようにしているユーザは、複数のアカウントを関連付ける要因となるような情報をプロフィール欄等に載せていないと考えたためである。実験に使用するアカウントの具体的な収集の

<sup>\*3</sup> 本研究では形態素解析器としてMeCab(<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>)、辞書としてIPA辞書を使用した。

<sup>\*4</sup> 本研究では内容語としてIPA辞書で定義されている名詞と動詞を使用した。

<sup>\*5</sup> 本研究では、一般、人名一般、人名姓、人名名、組織、地域一般、地域国の7つを固有名詞の細分類として使用した。

<sup>\*6</sup> <http://ameblo.jp/>

手順は次の通りである。

手順1 日本語の Twitter アカウントから、そのプロフィール欄にアメーバブログの URL が記載されているアカウントを収集する。

手順2 収集されたアカウントから下記の条件1, 2を満たすアカウント 3,000 組を抜き出し正例として使用する。また、負例を生成するため条件2を満たすブログアカウントも 10,000 アカウント収集する。

条件1 2013年の3月1日から10月31日の期間中の Twitter の投稿数が11以上

条件2 2013年の3月1日から10月31日の期間中のブログの投稿数が6以上

収集された Twitter アカウントの8ヵ月間の平均投稿数は1,635、ブログの平均投稿記事数は60.2であった。

4.1.2 ユーザにより関連付けられていないアカウントペア構築したシステムの実用的な精度を検証するため、ユーザにより明示的に関連付けが行われていないアカウントペアを用いた実験も行う。まず、Twitter とブログのアカウントをともに使用している200名のユーザに、使用している Twitter とブログのアカウントを回答してもらい、その中から以下の条件を満たす34個のアカウントペアを実験に使用した。

条件1 2013年の7月1日から2014年2月28日の期間中の Twitter の投稿数が11以上

条件2 2013年の7月1日から2014年2月28日の期間中のブログの投稿数が6以上

条件3 Twitter のプロフィール欄にブログのアカウントが記載されていない

条件4 Twitter とブログのアカウント名が異なっている<sup>\*7</sup>

## 4.2 実験設定

本研究における実験は基本的にユーザにより関連付けられたアカウントペア 3,000 組を用いて行う。収集された関連付きアカウント 3,000 組を 1,000 組ずつ3つに分割し、それぞれ学習データ、開発データ、テストデータの正例として使用する。負例は正例に含まれる Twitter アカウントと別途収集した 10,000 ブログアカウントから生成する。1つの正例に対する負例の数は学習時には9, 99, 999の3つの値で、テスト時には99, 999, 9,999の3つの値で実験を行った。学習時とテスト時で正例と負例の割合を一致させていないのは、実際にこのシステムを用いる場合を考えると、テスト時にどのくらいの負例が存在するか事前に分からないためである。

実験結果の評価には、Twitter アカウント 1,000 個中、候補ブログのランキング上位1つ、5つ、10つ中に正解のブログが含まれているアカウントの数(以下、Top1, Top5,

表1 1つの正例に対する負例の数と精度の関係

| Top1/MRR  |       | テスト時に1つの Twitter アカウントに対して候補とするブログ数 |           |           |
|-----------|-------|-------------------------------------|-----------|-----------|
|           |       | 100                                 | 1,000     | 10,000    |
| 学習時における   | 10    | 787/0.835                           | 646/0.711 | 323/0.386 |
| 正例に1つに対する | 100   | 804/0.849                           | 681/0.739 | 356/0.426 |
| する全事例の数   | 1,000 | 807/0.851                           | 679/0.738 | 369/0.434 |

表2 各素性を除くことによる精度の変化

| 除く素性         | Top1 | Top5 | Top10 | MRR   |
|--------------|------|------|-------|-------|
| 類似度に基づく素性    | 537  | 637  | 661   | 0.585 |
| 特有な表現を考慮した素性 | 618  | 755  | 815   | 0.687 |
| 投稿時間を考慮した素性  | 642  | 768  | 819   | 0.707 |
| すべての素性を使用    | 679  | 795  | 834   | 0.738 |

Top10)、および、正解のブログの順位の逆数の平均 (Mean Reciprocal Rank: MRR) を使用する。ランキング SVM のツールとしては、SVM<sup>rank</sup><sup>\*8</sup>を使用し、パラメータ C は開発データにおいて MRR が最大となる値を使用し、それ以外のパラメータはデフォルトの値を使用した。

## 4.3 実験結果

### 4.3.1 1つの正例に対する負例の数と精度の関係

まず、1つの正例に対する負例の数と精度の関係を明らかにするための実験を行った。結果を表1に示す。学習時の正例の数はいずれも1,000事例である。1つの正例に対して負例の数をそれぞれ9事例、99事例、999事例に変化させた結果、負例の数が9事例の場合と99事例の場合を比較すると99事例の場合の方が推定精度が良いことが確認できる。一方、負例の数が99事例の場合と999事例の場合を比較すると推定精度に差異は確認できなかった。

また、この傾向はテスト時における1つの Twitter アカウントに対して候補とするブログの数に依らないことも確認できる。対象の Twitter とブログが同一著者が作成したものかどうかの2値分類を行うような実験設定では、学習時とテスト時の正例と負例の数を一致させた方が良い精度が得られる場合が多いことが一般的に知られているが、本研究のようにランキング問題として捉えた場合は、正例と負例の割合が精度に与える影響は限定的であると言える。

### 4.3.2 素性ごとの効果

各素性の有効性を確認するため、素性を1種類ごとに除いた場合の精度を調べた。この際、学習時における1つの正例に対する負例の数は999、テスト時における1つの Twitter アカウントに対して候補とするブログの数は1,000とした。結果を表2に示す。類似度に基づく素性、ユーザに特有な表現を考慮した素性、投稿時間を考慮した素性はいずれも、これらの素性を除くことによりシステムの精度が大きく低下しており、これらの素性の有効性が確認できる。

\*7 ただし大文字と小文字は区別しない。

\*8 [http://www.cs.cornell.edu/people/tj/svm\\_light/svm\\_rank.html](http://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html)

表 3 関連付けられていないアカウントペアを用いた実験の結果

| Top1 | Top5 | Top10 | Total | MRR   |
|------|------|-------|-------|-------|
| 8    | 12   | 17    | 34    | 0.324 |

#### 4.3.3 ユーザにより関連付けられていないアカウントペアを用いた実験

続いて、構築したシステムの実用的な精度を検証するため、ユーザにより明示的に関連付けが行われていない 34 個のアカウントペアを用いた実験を行った。学習にはユーザにより関連付けられたアカウントペアを使用し、学習時における 1 つの正例に対する負例の数は 999、テスト時における 1 つの Twitter アカウントに対して候補とするブログの数は 1,000 とした。結果を表 3 に示す。

ユーザ自身によって関連付けられたアカウントペアと比べると精度は低いものの、34 アカウント中 8 アカウントはランキングのトップに、17 アカウントは上位 10 位以内にランキングされており、半分以上のアカウントは第三者にアカウント同士の関連性を推定される可能性があることが確認できる。この結果から、3 節で構築した著者の同一性推定システムは、ユーザが関連付けを行っていないアカウント同士の著者同一性推定にも有効であると言える。

## 5. 関連付け防止システム

### 5.1 関連付け防止システムの概要

本節では、関連付けを行っていない Twitter とブログのアカウントを持つユーザを想定し、新たに投稿されたブログ記事に同一性推定において重要となる手掛かり語が含まれていた場合に、警告メッセージを表示し、その語をマスクングするシステムについて検討する。具体的には 3 節で構築したアカウントの同一性推定システムにおいて、対象の Twitter アカウントを入力した場合に、同一著者が作成したブログアカウントが上位にランキングされることを防ぐシステムの構築を目指す。

### 5.2 関連付け防止システムの詳細

#### 5.2.1 関連付け防止処理の対象とするブログ記事

新たなブログが投稿された際、以下の 2 つの条件のいずれかを満たす場合に警告メッセージを提示し、マスクング処理の対象とする。

- (1) ブログ記事を投稿することで、対象のブログアカウントが上位  $rank_{th}$  以内にランクされる。
- (2) ブログ記事を投稿する前後で、以下の式で表される  $rr-diff$  が  $rr-diff_{th}$  を超える。

$$rr-diff = \frac{1}{\text{投稿後の順位} - rank_{th}} - \frac{1}{\text{投稿前の順位} - rank_{th}}$$

ここで、 $rank_{th}$ 、および、 $rr-diff_{th}$  はユーザが事前に設定する閾値であり、 $rr-diff$  は  $rank_{th}$  を基準とした場合

の逆順位 (Reciprocal Rank) の差を表している。本稿における実験では、 $rank_{th} = 20$ 、 $rr-diff_{th} = 0.01$  とした。すなわち新たなブログ記事を投稿することにより、対象のブログアカウントが上位 20 以内にランクされた場合、または、たとえば順位が 40 位から 35 位に上昇した場合<sup>\*9</sup>に警告メッセージを提示し、マスクング処理の対象とする。

#### 5.2.2 マスクング対象とする語の選択方法

4 節における実験の結果、Twitter とブログの類似度、ユーザに特有な表現、および、近い時間に投稿された Twitter とブログで共起する固有名詞は、2 つのアカウントの同一性推定の手掛かりになることが確認された。このうち、Twitter とブログの類似度は Twitter およびブログ全体から計算される値であることから、この値から同一性推定の手掛かりとなった特定の語を検出するのは難しいと考えられる。また、Kacmarcik ら [7] は高頻度の機能語の置換を行うことで著者推定の推定率を低下させる防止方法を提案しているが、ブログの投稿ごとに著者が高頻度で使用する機能語のマスクングを著者に求めることはユーザへの負担が大きいためと考えられる。

一方、ユーザに特有な表現や、近い時間に投稿された Twitter とブログで共起する固有名詞を検出することは比較的容易であり、修正が必要となる語の数も少ないと考えられる。そこで、本研究では著者同一性推定の手掛かりを含んでいると判定された新しいブログ記事から、ユーザに特有な表現、および、近い時間に投稿された Twitter とブログで共起する固有名詞をマスクングすることで関連付け防止を行う。

### 5.3 関連付け防止システムの評価実験

#### 5.3.1 実験設定

本実験では、Twitter を使用しているユーザが新しくブログを使い始めた場合に、Twitter のアカウントを知っている第三者にブログの存在を知られるのを防ぎたいという状況を想定しシステムの評価を行う。具体的には 2013 年 7 月 1 日にブログを使い始めたという状況を想定する。このため本評価では以下の条件を満たすユーザを実験に使用する。

- 条件 1 2013 年 7 月 1 日以前に Twitter での投稿がある
- 条件 2 2013 年 7 月 1 日以降にブログの投稿がある

上記の 2 つの条件に当てはまるユーザは、4.2 節で説明したテストデータの正例 1,000 組のうち 800 組存在した。これら 800 組を正例とし、正例 1 組に対して負例数は 9,999 事例として評価実験を行う。正例のブログの中には 2013 年 7 月 1 日以前の投稿が存在するものもあるが、7 月 1 日以降にブログを始めたとの想定から、それ以前の投稿は使用しない。

<sup>\*9</sup>  $\frac{1}{35-20} - \frac{1}{40-20} = 0.017 > rr-diff_{th}$  となり閾値を超える。

表 4 マスキング処理の効果

| 閾値の種類          | 処理後に閾値を下回る | 処理後も閾値を超える | 合計  |
|----------------|------------|------------|-----|
| $rank_{th}$    | 89         | 206        | 295 |
| $rr-diff_{th}$ | 9          | 65         | 74  |
| 合計             | 98         | 271        | 369 |

### 5.3.2 関連付け防止システムの評価

ある記事を追加した場合に  $rank_{th}$  , または ,  $rr-diff_{th}$  を超えた場合に , ユーザに特有な表現 , および , 近い時間に投稿された Twitter とブログで共起する固有名詞をマスキングすることで , どのくらいの割合で閾値を超えないようになるかの評価を行った . 具体的には以下の手順で評価を行った .

(1) 7月1日以降のブログの中で , 古い記事から順に1記事ずつ追加し ,  $rank_{th}$  , または ,  $rr-diff_{th}$  のいずれかの閾値を超えるブログを探す .

(2) 閾値を超えるブログがあった場合は , マスキング処理を行い , 改めて閾値を超えるかどうか判定する .

この際 , Twitter と負例のブログの記事データは , 着目している正例記事が投稿された時間までのデータのみを使用する . あるアカウントに対し , 追加することで閾値を超える初めての記事に対してのみ評価を行うので , アカウント1つに対してマスキングが有効か否かの判定は1回のみとなる . 結果を表4に示す .

実験に使用したアカウントペア 800 個のうち , 記事を順に追加していった結果 , 閾値  $rank_{th}$  , または ,  $rr-diff_{th}$  を超えたものは , それぞれ 295 個 , 74 個 , 合わせて 369 個存在した . 前者のうち , マスキングを行うことにより閾値を超えないようになったアカウント数は 89 個存在しており , およそ 30% のアカウントに対してはマスキングを行うことで閾値を超えないようになることが確認できた . 一方 , 閾値  $rr-diff_{th}$  を超えるアカウントについては , マスキングを行うことで閾値を超えないようになるアカウント数は 9 個 , 割合にしておよそ 12% であり , その効果は限定的であった .

## 6. おわりに

本研究では , ユーザ自身が関連付けていない Twitter アカウントとブログアカウントの関連付け防止システムの構築を行った . まず , どのような特徴がアカウントの同一性推定の手掛りとなるかを明らかにするため , Twitter のアカウントが与えられた場合に大量のブログの中から同一のユーザが作成したブログを推定するシステムの作成に取り組んだ . 続いて , そこから得られた知見をもとに , 関連付け防止システムの構築を行い . そのユーザに特有な表現や , 複数のソーシャルメディア間で近い時間に投稿された固有名詞をマスキングすることで , 同一性推定の可能性を下げられることを示した . しかし , その効果は限定的であ

り , より多くのアカウントに対して有効な手法を検討する必要があると言える .

## 参考文献

- [1] Stamatatos, E.: A Survey of Modern Authorship Attribution Methods, *the American Society for Information Science and Technology*, Vol. 60, No. 3, pp. 538–556 (2009).
- [2] Pillay, S. R. and Solorio, T.: Authorship Attribution of Web Forum Posts, *Proc. of APWG eCrime Researchers Summit'10* (2010).
- [3] Luyckx, K. and Daelemans, W.: Authorship Attribution and Verification with Many Authors and Limited Data, *Proc. of COLING 2008*, pp. 513–520 (2008).
- [4] Schwartz, R., Tsur, O., Rappoport, A. and Koppel, M.: Authorship Attribution of Micro-Messages, *Proc. of EMNLP'13*, pp. 1880–1891 (2013).
- [5] Qian, T. and Liu, B.: Identifying Multiple Userids of the Same Author, *Proc. of EMNLP'13*, pp. 1124–1135 (2013).
- [6] Korayem, M. and Crandall, D. J.: De-anonymizing Users Across Heterogeneous Social Computing Platforms, *Proc. of ICWSM'13*, pp. 689–692 (2013).
- [7] Kacmarcik, G. and Gamon, M.: Obfuscating Document Stylometry to Preserve Author Anonymity, *Proc. of COLING-ACL'06*, pp. 444–451 (2006).
- [8] Joachims, T.: Optimizing Search Engines Using Click-through Data, *Proc. of KDD'02*, pp. 133–142 (2002).
- [9] 浅原正幸, 松本裕治 : ipadic version 2.7.0 ユーザーズマニュアル (2003).