情報処理学会論文誌 Vol.56 No.1 72–82 (Jan. 2015)

ソーシャルネットワークにおける フォロー集合分析に基づく実世界イベント分類手法

河野 慎[1,2,a)] 米澤 拓郎[3] 中澤 仁[2,3] 川崎 仁嗣[4] 太田 賢[4] 稲村 浩[4] 徳田 英幸[2,3]

受付日 2014年4月13日, 採録日 2014年10月8日

概要：近年，GPS を搭載したスマートフォンと SNS の普及によって，リアルタイムに位置情報を付加させた発言をユーザが投稿する機会が増加している．この機会によって投稿された発言の中には実世界イベントに関する情報が含まれており，その一部はユーザが体験したり，目撃したりしたことに関するものであることが多い．これらの発言を収集し，解析することで実世界で実際に起きている社会イベントを検出することが可能となる．イベントを検出するために必要な発見と分類という 2 つの工程のうち，本研究ではイベントの分類に着目し，イベント参加者を利用したイベント分類手法を提案する．イベント参加者の多様性を意味する大衆性という新しい分類軸を定義し，イベント参加者がフォローしているユーザの解析によるイベントの分類を目指す．本研究では解析ツールの設計と実装をし，ツールを用いてあらかじめ実際のデータをもとに発見された社会イベントの解析を行い，分類を行った．Yahoo!クラウドソーシングにおいて一対比較法を用いて大衆性に基づき分類した結果を取得し，本手法を適用した解析結果と比較・考察を行った．その結果，大衆性に関してクラウドソーシングを用いた調査結果と回帰分析による提案手法の分析結果に一定の相関性があることを示した．

キーワード：Location based Social Network Service，実世界イベント検知，参加型センシング

Classifying Real-world's Events by Analyzing Friends of the Events' Participants in Social Network

Makoto Kawano[1,2,a)] Takuro Yonezawa[3] Jin Nakazawa[2,3] Satoshi Kawasaki[4] Ken Ohta[4] Hiroshi Inamura[4] Hideyuki Tokuda[2,3]

Received: April 13, 2014, Accepted: October 8, 2014

Abstract: Recent progress and spread of smartphones and social network services enables us to transmit text messages with GPS location data anywhere and anytime. Since these location-based SNS messages often refer real-world's events, many researchers have tried to recognize real-world's event through analysis of the messages. In this paper, we define a new index of event classification called popularity, and present a novel method to calculate the index by analysing social network of the events participants. Popularity, which reflects the diversity of event participants, is a useful index of the event for creating various applications such as event navigation or recommendation. We designed and implemented intuitive web-based interactive tool for analysing popularity of events. Through comparative experiments by analysis of proposed system and crowdsourcing, we confirmed that our proposal method provide a certain amount of accuracy for calculating the popularity of events.

Keywords: location based social network service, real world event detection, participatory sensing

[1] 東京大学大学院学際情報学府
Graduate School of Interdisciplinary Information Studies, The University of Tokyo, Bunkyo, Tokyo 113–8654, Japan
[2] 慶應義塾大学環境情報学部
Faculty of Environment and Information Studies, Keio University, Fujisawa, Kanagawa 252–0882, Japan
[3] 慶應義塾大学大学院政策メディア研究科
Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa 252–0882, Japan
[4] 株式会社 NTT ドコモ
DOCOMO R&D Center, Yokosuka, Kanagawa 239–8536, Japan
[a)] makora@ht.sfc.keio.ac.jp

1. はじめに

近年，情報技術の発展により，多くのユーザが GPS が搭載されたスマートフォンを所有し，様々な時間や場所で情報へのアクセスおよび発信が可能となった．また Twitter [*1] や Facebook [*2] をはじめとしたソーシャルネットワークサービス（以下，SNS）が普及し，2012 年 1 月時点で，

[*1] http://twitter.com
[*2] http://facebook.com

1,400 万人[*3]のユーザが利用している．これらの技術により，多くのユーザが GPS を用いた位置情報と紐付いた実世界の情報をリアルタイムに SNS 上へ投稿することが可能となった．本研究では，この位置情報が付与されて SNS 上へ投稿された情報を位置情報付き発言と呼ぶ．位置情報付き発言は実世界の出来事（実世界イベント，以下イベントと呼ぶ）が反映された内容であることが多いため，位置情報付き発言を収集し，解析することで情報空間からイベントの発生を検出する研究がさかんに行われている [1], [2], [3]．イベントの情報を利用することで，イベントの推薦や交通ナビゲーションといったアプリケーションへの応用，または都市計画への応用が期待されている．

SNS 上からイベントを検出し，利用可能にするためにはイベントの（a）発見と（b）分類の 2 つの工程が必要となる．本論文では，（a）発見をイベントが何であるか特定せず，その存在のみを検知することと定義し，また（b）分類を発見されたイベントの名称の特定ないしその特徴および属性を解析し，整理することと定義する．イベント情報を利用したアプリケーションの具体的な事例として，近傍イベント推薦アプリケーションの動作概要を図 1 に示す．

この推薦アプリケーションは以下のような手順で推薦が行われる．

(1) 社会イベントが発生する．
(2) イベントにユーザが参加・目撃をする．
(3) ユーザが (2) で行ったことに関する発言を SNS に投稿する．
(4) 本研究を応用した推薦システムが SNS 上の位置情報付き発言を収集する．
(5) 収集した発言をもとにイベントを検知する．
(6) 検知したイベントの性質を解析し，イベントの分類を行う．
(7) イベントが行われている場所の近くを通るユーザのプロフィールと合致する場合，そのイベントをユーザにプッシュ通知やメールなどを通じて推薦をする．

以上のようなアプリケーションを実現するためには，イベント発見だけではなくイベントの分類が重要となる．また細やかな推薦を可能とするためには，多様な指標でイベントの分類を行う必要がある．既存研究ではイベントの名称の推定や規模の分類を試みているが [4], [5]，十分とはいえない．たとえば，イベントはつねにある特定の名称を有しているとは限らず[*4]，多様な指標からイベントの分類を行うことが重要となる．

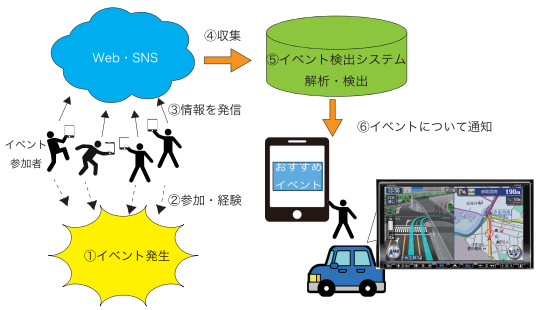

図 1 応用アプリケーション：推薦システム
Fig. 1 Application: Recommendation system.

[*3] 総務省 http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h24/html/nc123220.html

イントの情報を利用する際には，そのイベントがどのような参加者から構成されているか，という指標は有効であると考えられる．そこで，本研究ではイベントの分類軸として，大衆性という新たな指標を提案する．本研究では，大衆性をイベントに参加しているユーザの多様性と定義する．大衆性が高いイベントは様々な人々が参加しており，これは様々な人を魅了可能なイベントであるといえるため，推薦アプリケーションにおいて多くのユーザにそのイベントを推薦可能と判断することが可能である．逆に大衆性が低いイベントでは，特定の年齢層や趣味といった属性が偏った人々が参加していることから，推薦する対象のユーザを絞る必要がある．このように推定，分類されたイベントの大衆性は推薦アプリケーションやマーケティングに利用することが可能となる．本研究では，この大衆性に基づいたイベントの分類を可能にする解析手法および解析ツールを提案・実装し，評価を行った．解析手法として，本研究ではイベント参加者の興味や嗜好がフォローしているアカウント（以下フレンドと呼ぶ）に反映されていると仮定し，複数の統計的手法を用いた分析を試みた．評価として，イベントの大衆性を用いた提案手法と，Yahoo!クラウドソーシング[*5]を用いて収集した分類結果の比較を行った．そしてイベント参加者のフレンドを回帰分析することで，その結果とクラウドソーシングを用いた結果に一定の相関性があることを示すことができた．

本論文の構成は以下のとおりである．まず 2 章で SNS を用いてイベントを検出する研究の背景を説明し，実際にイベント検出を試みる関連研究を紹介する．3 章でイベント参加者のフレンドを解析することで大衆性を推定する手法について述べる．4 章では本手法を用いた解析ツールの設計と実装について述べ，5 章でそのツールを用いた解析結果の評価を行う．最後に 6 章で本研究のまとめと今後の展望を述べる．

2. SNS を用いたイベントの発見と分類

本章では，まずイベントの発見と分類に関して，関連研究を述べる．次に本研究が対象とする問題について述べ，最後に本研究の目的を述べる．

[*4] サッカー日本代表が FIFA ワールドカップ出場を決めた後の渋谷スクランブル交差点での若者のお祭り騒ぎなど，突発的に発生するイベントの名称は，その時点での特定は難しい．
[*5] http://crowdsourcing.yahoo.co.jp/

2.1 関連研究

特定のイベント発見

ある特定のイベントを発見する既存研究として，Sakaki ら [1] は Twitter 上の発言から地震を検出する手法を提案している．「地震」「揺れ」などといったキーワードを含んだ位置情報付き発言および，プロフィールに自身の居場所を登録しているユーザの発言を収集し，Support Vector Machine（SVM）によってこれらの発言が地震に関するものであるかを判定する．SVM は，高性能な機械学習アルゴリズムの 1 つであり，教師データをもとにあるデータが positive であるか negative であるか判定を行うものである．位置情報付き発言を解析することで，実際の震源地に近い場所を推定することができており，検出においては 96%の精度でできている．

同様に SVM を用いて特定のイベントを発見する研究では，Lanagan ら [2] によるスポーツにおけるイベント検出があげられる．この研究は位置情報を用いていないが，特定の試合における得点やファールが起きたタイミングなどを発言から発見することができている．

不特定のイベント発見および分類

Lee ら [3] は不特定のイベントを発見する手法として，位置情報付き発言の増加率に着目した手法を提案している．ある領域の空間の平常時とイベント発生時をその領域内の位置情報付き発言の増加率で区別し，イベントを発見可能にしている．同様に位置情報付き発言の量をもとにイベントの発見を行う研究として，Thelwall ら [6] の手法があげられる．

イベントの検出に加えて，位置情報付き発言を投稿したユーザの行動を追跡することでイベントの分類を試みた研究も存在する．Lee ら [3] は，ユーザのある領域における出入りを追跡することで，イベントの検出およびそのユーザ数からそのイベントの規模を推定している．同様に，藤阪ら [7] はある領域の空間におけるユーザの行動を追跡することで，そのイベントの影響範囲を推定している．

一方，位置情報付き発言は用いていないが，Ishikawa ら [8]，Becker ら [9]，Kwak ら [10] は，発言内容を自然言語解析し，話題トピックを抽出する手法を提案している．これらの手法をイベントに紐付いた位置情報付き発言に応用することで，その名称を用いた分類に応用することが可能となると考えられる．しかし，自然言語解析を発言に用いた手法では表記ゆれや情報量の不足を問題としてあげることができる．情報量の不足という問題に対し，西田ら [5] のツイートを圧縮し，情報量を増やす手法を応用することが考えられるが，日本語特有の表記ゆれの問題には対応しきれず，不十分である．Takhteyev ら [11] によるイベントと地理情報を紐付けるものがある．これはユーザ同士のつながりを解析して，ソーシャルグラフを作成し，その後位置情報付き発言との関係性を分析することで，イベントの地域との密着性による分類を行うものである．

2.2 問題意識

以上のように SNS 上の情報を用いたイベントの発見や分類を試みたいくつかの先行研究が存在するが，特にイベント分類に関しては，その分類指標および手法はいまだ十分とはいえない．既存研究では，イベントの名称の推定や規模の推定がイベント分類手法として試みられているが，突発的なイベントではその名称特定が難しく，また規模だけではイベント推薦といったアプリケーションにおいては十分な判断基準とはいえない．したがって，イベント情報を用いた多様なアプリケーションを構築するためには，これまで提案された分類指標（名称，規模）に基づく解析精度の向上も重要であるが，異なる分類指標からイベントを分類することも重要であるといえる．

さらに分類指標に加え，その解析手法も問題となる．リアルタイムにイベント推薦を行うアプリケーションを実現するためには，イベント分類のための計算量はできるだけ少ないことが望ましい．また本研究で用いる Twitter やその他の SNS では，API 制限が設けられており（Twitter は 2014 年 3 月現在，ユーザの情報取得は 15 分間に 15 人分のみ可能[*6]），SNS へのアクセス負荷をできるだけ少なくした解析手法が求められる．

2.3 目的

以上のことから，イベントを分類する際に規模やジャンルのみではなく，新しい分類指標が必要である．またリアルタイムな分類を可能にするため，その際には，計算量を抑えるとともに，SNS の API アクセス負荷を抑えた手法が必要となる．したがって本研究の目的は，新しいイベントの分類指標を定義し，この分類指標に基づいてイベントを分類可能にする手法を実現することである．

3. アプローチ

3.1 大衆性

本研究では，イベントを分類する新しい指標軸として，大衆性を提案する．一般に大衆性はそのイベントがどの程度大衆に受け入れられる性質を有しているかを示す．つまり大衆性の高いイベントには，様々な人が参加しているといえる．本論文での様々な人とは，年齢や性別，出身，趣味など様々な属性が分散していることを意味する．具体的には，花火大会やお祭といったイベントの場合，年齢層は幅広く，それぞれが持つ趣味や興味も様々な人々が参加していると考えられる．一方でアーティストによるライブやファンサービスなどのイベントの場合，アーティストによっては特有の年齢層や性別の人が参加していたり，アー

[*6] https://dev.twitter.com/docs/api/1.1/get/friendships/show

ティストのファンが多く参加していたりすることが多いと考えられる．以上のように大衆性が高いイベントは様々な人が参加しているイベントを意味し，大衆性が低いイベントは特定の年齢層や，興味を持った人が参加しているイベントを意味する．本研究では，この大衆性をイベントに参加しているユーザの多様性やイベントの内容のばらつきなどを包括的に含めたものと定義する．なお，この参加者の多様性は，大衆性の推定において重要な要素として扱う．このイベントの大衆性を推定することで，イベントを推薦する対象のユーザを決定することができる．大衆性の高いイベントの場合は推薦する対象ユーザは多岐にわたるのに対し，大衆性の低いイベントの場合は対象ユーザを絞ることで推薦の質を向上させることができる．

3.2 推定手法

上記のようなイベントの大衆性を推定するためには．イベント参加ユーザのプロファイリング（年齢，性別，趣味）をする必要がある．Twitter では，それぞれのアカウントにおいて，プロフィールを記入することができる．しかし，このプロフィールを詳細に記入しているユーザは少なく，プライバシなどの問題から，多くのユーザはほぼ空白である．したがってこのプロフィールを用いてユーザのプロファイリングを行うことは難しい．そこで本研究では，イベント参加者が SNS 上でフォローしている複数のユーザに着目する．以降本研究では，このイベント参加者がフォローしているユーザをフレンドと呼び，複数のフレンドのことをフレンズと呼ぶ．ユーザは，実世界での友人のアカウントや自分の興味分野に関する情報を発信する著名なユーザのアカウント，bot アカウントなどをフォローしている．つまりユーザがフォローしているフレンズにユーザの属性が反映されているといえる．したがってイベント参加者のフレンドを分析することでプロファイリングが可能となると考えられる．**図 2** に参加者のフレンズと大衆性の関係性を示す．特定サークルの集まりのように一般に大衆性が低いと考えられる社会イベントの場合，イベント参加者は同じような目的や趣味を持っている人が集まっていることが多く，図 2 左のように共通のアカウントをフォローしている率が高いことが考えられる．逆に花火大会のように一般に大衆性が高いと考えられる社会イベントの場合，イベント参加者のフォローも様々なアカウント（フレンズ）に分散していると考えられる．したがって図 2 右のような状態にあると考えられる．本研究では上記のフレンズと大衆性の関係を仮定し，この仮定に基づいたイベント参加ユーザのフレンズの分析の手順を示す．あるイベントにおいて，イベントに参加しているユーザ u_k（$k = 1, 2, 3, \ldots, n$）がフォローしているすべてのフレンズの集合を f_{u_k} とすると，イベントごとのフレンズ集合 F は，$F = f_{u_1} \cup f_{u_2} \cup f_{u_3} \cup \cdots \cup f_{u_n}$ と表せる．TwitterAPI への 1 回のアクセスによって，1 人のユーザがフォローしているフレンズのリストを取得することが可能である．したがってこのフレンズ集合 F を得るためには，イベント参加者数の数が n 人いた場合は，Twitter API へ n 回のアクセスで済む．既存のグラフ解析手法の 1 つである Clique Percolation Method（CPM）[12], [13] を用いるとユーザ 1 人のフォロー関係を解析するために，1,000 回ほど API へアクセスしなければならず，n 人いれば，そのアクセス回数は 1,000n 回となり，API 制限を大きく超えてしまう．またサーバへの負荷を考慮してもリアルタイムな分析には適しておらず，本手法はこれらの観点からも現実的な手法であると考えられる．各イベントのフレンズ集合 F を求めた後，F に属するフレンド $friend_j\{friend_j \in F\}$ ごとに，n 人のイベント参加者のうちフォローされている割合 $ratio_{friend_j}$ を算出する．この $ratio_{friend}$ をもとにフレンドを降順に並べ，横軸にフレンド，縦軸に $ratio_{friend}$ を設定しグラフにすると，べき乗分布に従うことが予測される．以上の予測から本研究では，べき乗分布の解析を行う．べき乗分布を解析する手法はいくつか存在するが，本研究では回帰分析とジニ係数に着目し，大衆性の分類を試みる．下記にそれぞれの手法とその手法を適用した理由について説明を行う．

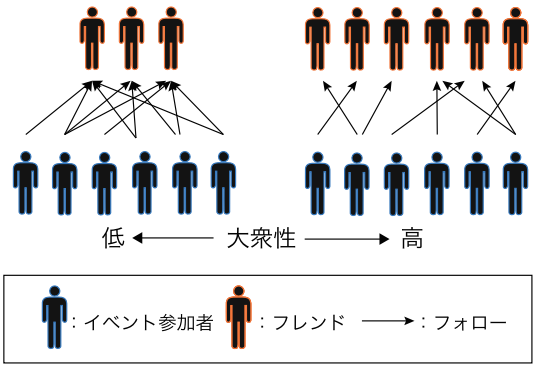

図 2 フレンズと大衆性の関係性
Fig. 2 The relation between *Friends* and *Popularity*.

回帰分析

回帰分析を用いることで，べき乗分布の曲線を数式化することが可能となる．回帰分析の手順を次に示す．まず両対数グラフ化を行う．各イベントにおけるグラフの縦軸と横軸の値の対数をとり，点をとる．横軸はフレンドになっているため，参加者にフォローされている割合の多い順に $1, 2, 3, \ldots, n$ と数字に置き換えて対数化を行う．次に回帰分析を用いて，両対数グラフの直線の数式 $y = \alpha x + \beta$ を求める．最後に得られた傾き α と切片 β をもとにべき乗分布の式 $y = e^{\beta} x^{\alpha}$ を求める（e^{β} は定数であり，イベント間に差はないため，本実験では利用しない）．この x^a を求めることでその曲線の曲がり具合いを取得できる．α の値が大きければ大きいほど，曲線の曲がり具合いは緩くなり，

小さくなればなるほど急な曲がり具合いとなる．緩やかである場合，参加者によってフレンズがフォローされている割合は全体的に高く，またその割合も近いフレンドが一部存在していることを意味している．このことから参加者のフォローが一部のフレンズに集中していることになり，上述の仮定より，大衆性が低いことを示している．急な曲がり具合いである場合は，あまり多くの参加者にフレンズがフォローされていないことを意味しており，参加者のフォローが分散していることを示す．よって α の値が小さいほど，大衆性は高いものとなる．以上のことから回帰分析を用いて曲線の曲がり具合いを算出することで，特定のフレンドが参加者から特出してフォローされているのかそうではないのか推定することができる．

ジニ係数

ジニ係数は，経済学におけるべき乗分布において用いられる係数であり，ある母集団における所得分配の不平等性を示す．このジニ係数を本研究に適用すると，あるイベント参加者のフレンド集団における参加者からのフォロー獲得分配の不平等性を示す．ジニ係数 $gini$ の求め方は次のとおりである．

まず，

$$avr = \frac{ratio_1 + ratio_2 + \cdots + ratio_n}{n} \tag{1}$$

を求める．次に平均差 ad（フレンド間のフォロー獲得率の差の平均）を求める．最後に平均差を全体の平均値の 2 倍で割る．

$$gini = \frac{ad}{2avr} \tag{2}$$

一般にジニ係数は 0–1 間の値になっており，1 に近いほど格差は大きく，0 に近いほど格差は小さい．したがってあるイベントのジニ係数によって，フレンドが参加者からどのようにフォローされているのかが分かる．

4. 設計と実装

本研究では本手法を適用した解析ツールを設計し，実装した．解析ツールには次の 3 つが機能要件となる．

- 位置情報付き発言の収集と地図上に発言の表示
- 位置情報付き発言からイベントの発見，登録
- 発見されたイベントの解析

以下，それぞれの設計・実装の詳細について述べる．本解析ツールの実装環境を**表 1** に，システム構成図を**図 3** にそれぞれ示す．

表 1 実装環境
Table 1 System requirements.

仕様言語	Python Ver.2.7.5
フレームワーク	Django Ver.1.5.4
サーバ OS	Ubuntu Ver.12.04 LTE

4.1 位置情報付き発言収集機能

本研究では日本国内の様々なイベントの解析を対象とするため，全国で投稿される位置情報付き発言をリアルタイムに取得し，保存する機能が必要である．その際に TwitterAPI の制限によって位置情報付き発言が取得できない状態にならないように工夫をする必要がある．また取得する位置情報付き発言は 1 日だけでも膨大な量になるため，後にデータを参照する際に，素早くデータにアクセスできなければならない．

そこで，位置情報付き発言は日本の国内を含むように緯度経度を設定し，Twitter の Streaming API [*7] を用いて位置情報付き発言を収集する．位置情報付き発言は 1 日で約 25 万ほど投稿され，収集した位置情報付き発言を直接データベースに保存する場合，遅延が生じてしまう．位置情報付き発言収集モジュールでは直接データベースに保存をせずに 1 分おきに json ファイルに保存することでこの問題を解決する．次に保存されている json ファイルを処理し，sql ファイルを作成した後，データベースに保存していく．また情報量が膨大なため，後にイベント発見時や解析時にデータを参照する際，参照結果を取得するまで時間がかかってしまう．そこでテーブルを日付ごとにパーティショニングを行うことで，位置情報付き発言はテーブル内で日付ごとに区切り，参照を高速で可能とした．

4.2 社会イベント発見機能

本研究は，社会イベントの分類を目的としており，社会イベントの発見は既存の手法 [3] の適用を想定している．しかし分類を行うのは社会イベントが発見された後であることから，本ツールにおいても解析の対象となるイベントの発見・登録処理を可能とする機能が必要である．そこで本ツールは自動ではなく，利用ユーザが手動で社会イベントの発見・登録を可能とするイベント発見モジュールを有する．イベント発見を容易かつ直感的に可能とするため，イ

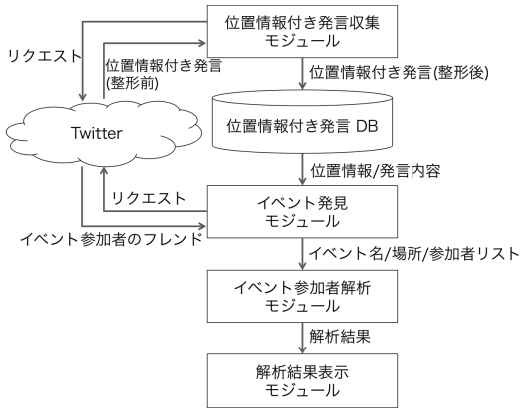

図 3 システム構成図
Fig. 3 System configuration.

[*7] http://dev.twitter.com

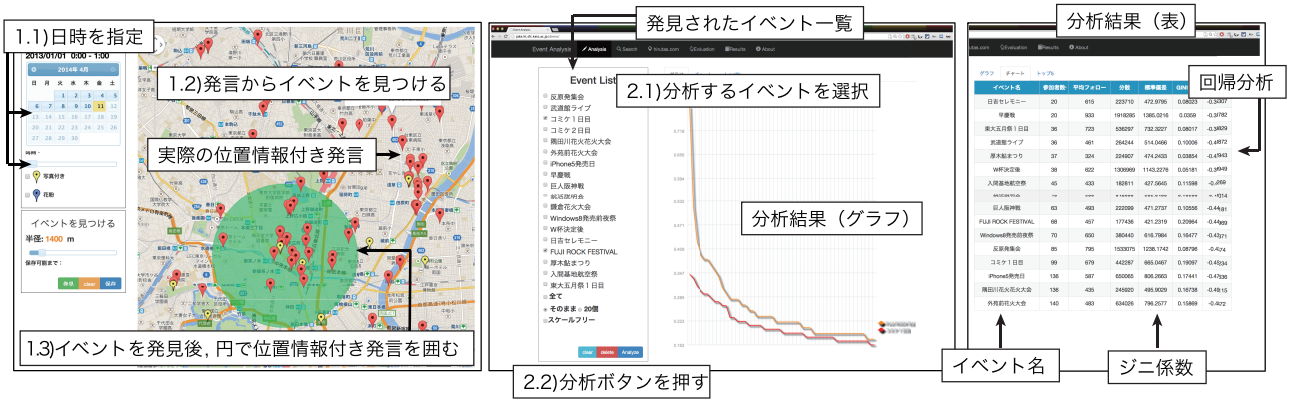

図 4 解析ツール：左から (1) イベント発見ツール，(2) 視覚的な解析結果表示ツール，(3) 表を用いた解析結果表示ツール

Fig. 4 Analysis tools from left: (1) Event discovery tool, (2) Analysis result visualization tool via a graph, (3) Analysis result visualization tool via a table.

ベント発見モジュールは地図上に位置情報付き発言をマッピングし，視覚的に発見を行えるように設計した．まず，取得された位置情報付き発言をその位置情報から Google Maps に表示する．Google Maps を操作して任意の場所を表示すると，表示されているマップに含まれている位置情報付き発言が表示される．それぞれの位置情報付き発言を選択するとその発言内容や画像が付加されていた場合はその画像も表示される．イベントを発見した場合，イベントが行われている場所を円で位置情報付き発言を囲み，イベントに名前をつけて保存する．この機能画面を図 4 (1) に示す．本ツールは登録された後，イベント名と円の範囲内に入った位置情報付き発言をしたユーザの ID がデータベースに保存されていく．なお，この円の大きさは任意に指定することができ，円で包括的に位置情報付き発言者を囲む．このため実際にイベントに参加していないユーザが含まれている場合や，イベント参加者全員を含めきれていない場合がある．しかしユーザが実際にイベントに参加しているかは別の課題であると考え．今回は考慮しない．

4.3 社会イベント解析機能

イベント発見機能によって発見されたイベントはイベント参加者解析モジュールによって解析される（図 4 (2), (3)）．本論文では回帰分析とジニ係数を用いた解析を対象としているが，今後その他の解析手法を用いる場合に，拡張できるよう設計を行った．前述のフレンドがイベント参加者からフォローされている割合をもとに降順でフレンドを並べたグラフを図 4 (2) に示す．縦軸に $rate_{f_j}$，横軸に $Friends$ となっている．このグラフからフレンドのフォローされている状況がべき乗分布に従っている予測が正しいことが分かる．このべき乗分布がそれぞれのイベントの参加者のフォローの状態を表しているものであり，このべき乗分布を解析することでそのイベントの参加者の状態を推定することが可能となる．

表 2 分析対象とするイベント一覧

Table 2 The list of target events.

日付	イベント名	発言者数
2012/02/24	メディア芸術祭	68
2012/02/26	東京マラソン	120
2012/04/04	慶應義塾大学入学式	20
2012/07/25	鎌倉花火大会	44
2012/07/27	FUJI ROCK FESTIVAL	68
2012/07/28	隅田川花火大会	136
2012/08/04	あつぎ鮎まつり	37
2012/09/15	巨人阪神戦	63
2012/11/01	武道館ライブ	36
2012/11/03	入間基地航空祭	45
2012/12/09	就活説明会	48
2012/12/29	コミックマーケット	99
2013/06/01	早慶戦	20
2013/11/23	東京モーターショー	108

5. 評価実験

本章では，提案手法による大衆性分析の有用性を検証するため，実験を行う．分析対象として，異なる 14 種類のイベントを用いる．対象となるイベントの大衆性を評価するため，一般ユーザに大衆性に基づいたランク付けを行ってもらい，正解データとして利用する．そのデータと提案手法による分析とを比較し，評価を行う．以下に，詳しく説明を行う．

5.1 対象とするイベント

評価に用いるイベントリストは本ツールによって発見することができた 27 種類のイベントのうち．日本イベント産業振興協会[*8]によって定義されたイベント 16 タイプのいずれかに該当した 14 種類のイベントで構成されている（表 2）．16 タイプに該当しなかった残り 13 種類のイ

*8 http://www.jace.or.jp/

ベントはその他のタイプに分類される．イベントの発見に利用した位置情報付き発言は，2011 年 11 月 1 日から 2013 年 6 月 14 日までと 2013 年 11 月 1 日以降の期間に投稿されたものである．なお，期間に間があいてしまった原因は TwitterAPI の仕様変更にともない，システムがこれに対応しておらず，停止してしまっていたためである．実装したツールの機能を単純化するため，登録が可能なイベントの期間の限度は 1 日にした．そのため，数日にわたって開催されているイベントについては，日曜日が含まれているイベントの場合は日曜日をイベント当日とし，日曜日を含まないイベントの場合は発言者数が最も多い日をイベント当日と見なした．前章で述べた実装ツールを用い，イベントに参加していると考えられるユーザを抽出し，解析対象とした．なお表 2 に示したとおり，イベントによって得られた参加（発言）者数が大きく異なっている．解析対象のイベント参加者数を統一するため，無作為に参加者の中から 20 人を抽出し，解析を行った．この 20 人は参加者数が最も少なかった早慶戦に合わせた．

5.2 正解データの作成

評価は，あらかじめ用意したイベントのリストを本手法を用いた分類結果と正解データとの比較によって定性的に行う．評価のための正解データは一般ユーザによって分類されたものとする．一般ユーザによる分類手法には一対比較法を採用した．一対比較法は調査対象物を 1 対 1 で比較を行う一次元の間隔尺度構成法の 1 つである．本研究において，最終目的が推薦システムのアプリケーションへの応用であることから，「より多くの人に勧めることができる」イベントの比較を行う．本研究における調査対象物はイベントリスト 14 個であり（表 2），一対比較法を適用した場合，${}_{14}C_2 = 91$ 通りの比較が必要となる．また一対比較法には調査対象のどちらかの選択のみを行うサーストン法や選択だけではなくその度合いを回答するシェッフェ法をはじめとした数多くの種類が存在する．シェッフェ法は詳しい回答を要求するため，より詳細な分析を行うことが可能となる．このほかの一対比較法にもそれぞれ特徴がある．しかし，本研究は 91 通りの比較が必要であり，一般ユーザの回答への負担がすでに大きい．そこで負担を軽減するため一対比較法の 1 つであるサーストン法を採用した．

調査に回答する母集団として，本研究では Yahoo!クラウドソーシングを利用した．Yahoo!クラウドソーシングは Yahoo!Japan によって提供されているサービスであり，細分化された課題（タスク）を不特定多数のユーザに取り組んでもらうことが可能となる．本研究では，91 通りの比較を 5 つのタスクに分け，調査を 2014 年 3 月 13 日に行った．**図 5** のように 2 つのイベントを比べてもらい，より多くの人に勧めることができるイベントを選択してもらった．一対比較法を適用するためには，1 人のユーザに 91 通りのすべての比較をしてもらう必要がある．それぞれのタスクを回答したユーザは 1,000 人おり，その中で 5 つのタスクすべてに答えたユーザを ID から識別した．その結果 946 人（**表 3**）のユーザに回答してもらった．なお，クラウドソーシングの作業を行ってもらったユーザは全国から募集を行っているため，必ずしも対象となるイベントについて詳しいわけではない．よって，得られたデータには一部イベント名による印象に大きく影響を受けている可能性がある．また図 5 のように調査の際に新聞社や公式 HP に掲載されている画像を使用しており，回答者はこれらの画像の影響を受けている可能性もある．しかし本研究では，あくまで一般的な印象として受ける大衆性という観点から，本提案手法の評価として位置づけを行った．回答結果を分析し，その尺度によって並べ替えたものを**表 4** に示す．この結果を正解データとし，提案手法を用いた分析結果との比較を行う．

図 5 Yahoo!クラウドソーシングで提示される調査画面のスクリーンショット
Fig. 5 The screenshot of question in Yahoo!Crowdsourcing.

表 3 クラウドソーシングにおける調査協力者の内訳
Table 3 The itemise of collaborators from Yahoo!Crowdsourcing.

対象人数	946 人
性別	男性 541 人
	女性 405 人
年齢層	10 代 5 人
	20 代 157 人
	30 代 388 人
	40 代 278 人
	50 代以上 118 人
居住地	北海道・東北 79 人
	関東 356 人
	北陸・甲信越 57 人
	東海 112 人
	関西 171 人
	中国・四国 93 人
	九州・沖縄 78 人

表 4 クラウドソーシングの結果
Table 4 The result of *Popularity* from Yahoo!Crowdsourcing.

順位	イベント名	間隔尺度
1	隅田川花火大会	1.4284
2	鎌倉花火大会	0.9352
3	FUJI ROCK FESTIVAL	0.3755
4	東京マラソン	0.3624
5	東京モーターショー	0.3056
6	入間基地航空祭	0.1340
7	あつぎ鮎まつり	0.1199
8	巨人阪神戦	0.0111
9	コミックマーケット	−0.0470
10	Perfume 武道館ライブ	−0.1335
11	メディア芸術祭	−0.1604
12	早慶戦	−0.6105
13	就活合同説明会	−1.1707
14	慶應義塾大学入学式	−1.550

表 5 回帰分析（20 人）結果
Table 5 The result of *Popularity* from regression analysis.

順位	イベント名	α
1	隅田川花火大会	−0.491
2	東京マラソン	−0.471
3	コミックマーケット	−0.452
4	FUJI ROCK FESTIVAL	−0.450
5	鎌倉花火大会	−0.444
6	巨人阪神戦	−0.443
7	メディア芸術祭	−0.431
8	東京モーターショー	−0.4302
9	あつぎ鮎まつり	−0.4295
10	Perfume 武道館ライブ	−0.4286
11	入間基地航空祭	−0.427
12	就活説明会	−0.420
13	慶應義塾大学入学式	−0.357
14	早慶戦	−0.338

5.3 提案手法による分析結果

本節では，初めに提案手法による結果について考察し，次に Yahoo!クラウドソーシングによって得られた正解データと提案手法による結果を比較し，考察を行う．

回帰分析

回帰分析によって得られた結果を**表 5** に示す．表はイベント名と回帰分析を行って算出された $y = Cx^{\alpha}$ の α で構成され，α の昇順に並べたものである．表の上方には一般的に大衆性が高いと考えられる隅田川花火大会や鎌倉花火大会がきており，表の下方には大衆性があまり高くないと考えられる大学のイベントである早慶戦や就活合同説明会，慶應義塾大学入学式がきている．大衆性が低いイベントほど，α の値は大きくなるということである．α の値が大きい場合，曲線のカーブは緩やかでかつ上の方に位置している，つまりイベント参加者のフレンズが参加者にフォローされている割合は全体的に高く，そしてフォローが集中していることを示している．逆に α の値が小さい場合，曲線のカーブは急なものになり，割合のほとんどが 0 に等しくなる．つまり，大衆性が高いイベントでは実際に特定のフレンズが集中的にフォローされずに，分散しているということである．以上のことから大衆性が低いイベントの場合，イベント参加者の多くがある特定のフレンズをフォローしているということである．逆に大衆性が高いイベントの場合，フレンズのほとんどが複数の参加者にフォローされていないということである．

表 6 ジニ係数（20 人）結果
Table 6 The result of *Popularity* from Gini coeffcient.

順位	イベント名	ジニ係数
1	東京モーターショー	0.414
2	Perfume 武道館ライブ	0.340
3	東京マラソン	0.298
4	メディア芸術祭	0.247
5	FUJI ROCK FESTIVAL	0.210
6	コミックマーケット	0.191
7	鎌倉花火大会	0.183
8	隅田川花火大会	0.168
9	入間基地航空祭	0.117
10	巨人阪神戦	0.109
11	就活説明会	0.106
12	慶應義塾大学入学式	0.088
13	あつぎ鮎まつり	0.039
14	早慶戦	0.037

ジニ係数

ジニ係数の算出結果を**表 6** に示す．表はイベント名をジニ係数の降順で並べたものである．表の上方には東京モーターショーや FUJI ROCK FESTIVAL といった比較的規模の大きく，全国から人が集まってくるようなイベントがきており，表の下方には鎌倉花火大会やあつぎ鮎まつり，早慶戦といった比較的規模が小さく，地元の人たちが集まってくるようなイベントがきている．つまりジニ係数が高いイベントほど参加者は全国から集まってきやすく，逆にジニ係数が低いイベントほど参加者は地元の人が集まってきやすいと考えられる．ジニ係数の値が大きいほど，フレンドの参加者にフォローされている割合の格差が大きいことを示している．つまり，全国的に有名なイベントは割合の格差が大きく，地元であまり有名でないイベントの割合は格差が少ないということである．

5.4 比較・考察

クラウドソーシングによる結果（表 4）と本手法による結果（表 5，6）をもとにスピアマンの順位相関係数を算出した（**表 7**）．この係数から，クラウドソーシングの結果と回帰分析による結果が強い相関を持つことが分かる．大衆性が高いと考えられる隅田川花火大会や FUJI ROCK FESTIVAL が正解結果と回帰分析結果ともに上位にきて

表 7 スピアマンの順位相関係数による結果
Table 7 The result of Spearman's rank correlation coefficient.

手法	スピアマンの順位相関係数
回帰分析	0.749
ジニ係数	0.389

いる．また大衆性が低い早慶戦や就活合同説明会，慶應義塾大学入学式は順序は前後するが，クラウドソーシングの結果と回帰分析の結果の両方で下位にきている．これらの点からも，回帰分析を用いた手法が有効であるといえる．また本手法の結果では鎌倉花火大会は順位は中位に配置されている．しかしながら，クラウドソーシングによる結果は鎌倉花火大会が 2 位にきている．これの原因としてイベント名に「花火大会」という単語が入っているために，一般ユーザは「より多くのユーザに勧めることができる」と考えたと推測される．ほかには本手法の結果とクラウドソーシングの結果が異なったイベントにコミックマーケットがあげられる．コミックマーケットのジャンルはサブカルチャであるため，サブカルチャについて興味を持たない人々にとっては参加しにくい．そのためクラウドソーシングの結果では，順位が低くなったと考えられる．しかし，コミックマーケットは世界的に広く知られているイベントであり，またサブカルチャというジャンルの中にはさらにアニメなど様々なジャンルが内包されているため，本手法ではイベントの大衆性は高いと推定されたと考えられる．なお，ジニ係数による分析結果についてはクラウドソーシングの結果と大きく異なっている．大きく異なってしまっているイベントの $ratio_{friend}$ が大きいフレンドをみると孫正義や宇多田ヒカルなど，フォロワ数が圧倒的に多い著名なアカウントが含まれている．Twitter 全体においてもこれらのアカウントをフォローしているユーザは多いことから，多くのイベント参加者も同様にこれらのアカウントをフォローしている．一方でその他のアカウントは Twitter 全体において孫正義などのアカウントに比べて圧倒的にフォロワ数は少ないため，複数のイベント参加者がフォローしている数は，孫正義などのアカウントと比較するとほとんどいない．そのため，孫正義などのアカウントとその他のアカウントの間に大きな共通フォローの差が生じてしまい，結果的に不平等となりジニ係数に影響してしまったことが原因として考えられる．孫正義などの多くのユーザがフォローしているようなアカウントを何らかのフィルタリングによって除去することで，この問題は解決できると考えており，これは今後の課題である．

5.5 今後の課題

本研究では定性的な評価のみであり，定量的な評価を行うことができていない．その原因として大衆性の客観的な正解データの構築の難しさがあげられ，この点は今後改善していく必要があるといえる．また本研究が試みた分類はイベントが発見されたあとに行われるものであるため，本研究で対象としたイベントは手動で発見されたものとなっている．今後はこのイベントの発見も自動でリアルタイムに行われることが要求される．イベントの自動発見は多くの研究でなされているが，本研究と組み合わせる場合に最適だと考えられるものとして，Lee ら [3] が提案している時空間解析があげられ，それと本ツールの組合せは今後の課題である．

加えて今回位置情報付き発言の発言内容について扱わなかったが，本研究で提案した手法と組み合わせることで，より詳細にイベントを分類することが可能になるといえる．そして位置情報付き発言に画像が添付されているものもあり，この画像添付の有無，そして画像解析を加えることでよりイベントに関する情報を位置情報付き発言から取得できるようになる．このように位置情報付き発言には本研究では利用していない有益な情報が多く含まれており，これらの情報を有効に扱うことで，より細かな分類が実現できる．

最後に本研究では，イベント参加者をある一定の大きさの円の範囲内で発言しているユーザと見なすと述べたが，その結果イベントに参加していないユーザ（ノイズ）を含んでしまっており，イベント参加者の精度を下げてしまい，実験結果に影響していることも考えられる．そこで Wang ら [14], [15] や，Hiruta ら [16] のように機械学習を用いたり，自然言語処理を用いたりして情報の信頼性を向上させるためのフィルタリング技術を用いることで，イベント参加者の精度を向上させることが期待できる．

6. まとめ

Twitter において位置情報が付加された発言（位置情報付き発言）を用いてイベント検出を試みる研究は数多く存在する．研究の多くが位置情報やその発言内容に着目しているなか，本研究は位置情報付き発言をしているイベント参加者のフォローに着目した．このフォローの状態を解析することで，イベントの大衆性というイベントが万人に受けるものであるかという新しい分類軸でイベントを分類する手法の提案を回帰分析とジニ係数を適応し，イベントの発見から分析までをインタラクティブに可能とするツールの設計・実装を行った．

解析には回帰分析によるものとジニ係数による手法を用いた．Yahoo!クラウドソーシングと一対比較法を用いて正解データを作成し，本手法の結果と比較した．その結果，大衆性に関して，クラウドソーシングの結果と回帰分析による提案手法の分析結果に一定の相関性があることを示すことができた．推薦システムに本手法を取り入れることで，推薦する対象の選択が可能となり，今後本手法を適用した推薦システムの設計と実装が期待される．

参考文献

[1] Sakaki, T., Okazaki, M. and Matsuo, Y.: Earthquake shakes Twitter users: Real-time event detection by social sensors, *Proc. 19th International Conference on World Wide Web*, pp.851–860, ACM (2010).

[2] Lanagan, J. and Smeaton, A.F.: Using twitter to detect and tag important events in live sports, *Artificial Intelligence*, pp.542–545 (2011).

[3] Lee, R. and Sumiya, K.: Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection, *Proc. 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, pp.1–10, ACM (2010).

[4] Lee, R., Wakamiya, S. and Sumiya, K.: Discovery of unusual regional social activities using geo-tagged microblogs, *World Wide Web*, Vol.14, No.4, pp.321–349 (2011).

[5] 西田京介，坂野遼平，藤村　考，星出高秀：データ圧縮による twitter のツイート話題分類，*DEIM Forum 2011*, A1–6 (2011).

[6] Thelwall, M., Buckley, K. and Paltoglou, G.: Sentiment in Twitter events, *Journal of the American Society for Information Science and Technology*, Vol.62, No.2, pp.406–418 (2011).

[7] 藤坂達也，李　龍，角谷和俊：実空間マイクロブログ分析による地域イベントの影響範囲推定，*DEIM Forum* (2010).

[8] Ishikawa, S., Arakawa, Y., Tagashira, S. and Fukuda, A.: Hot topic detection in local areas using Twitter and Wikipedia, *ARCS Workshops* (*ARCS*), pp.1–5, IEEE (2012).

[9] Becker, H., Naaman, M. and Gravano, L.: Beyond Trending Topics: Real-World Event Identification on Twitter, *ICWSM* (2011).

[10] Kwak, H., Lee, C., Park, H. and Moon, S.: What is Twitter, a social network or a news media?, *Proc. 19th International Conference on World Wide Web*, pp.591–600, ACM (2010).

[11] Takhteyev, Y., Gruzd, A. and Wellman, B.: Geography of Twitter networks, *Social Networks*, Vol.34, No.1, pp.73–81 (2012).

[12] Palla, G., Derényi, I., Farkas, I. and Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society, *Nature*, Vol.435, No.7043, pp.814–818 (2005).

[13] Palla, G., Barabási, A.-L. and Vicsek, T.: Quantifying social group evolution, *Nature*, Vol.446, No.7136, pp.664–667 (2007).

[14] Wang, D., Abdelzaher, T., Ahmadi, H., Pasternack, J., Roth, D., Gupta, M., Han, J., Fatemieh, O., Le, H. and Aggarwal, C.C.: On Bayesian interpretation of fact-finding in information networks, *Proc. 14th International Conference on Information Fusion* (*FUSION*), pp.1–8, IEEE (2011).

[15] Wang, D., Kaplan, L., Le, H. and Abdelzaher, T.: On truth discovery in social sensing: A maximum likelihood estimation approach, *Proc. 11th International Conference on Information Processing in Sensor Networks*, pp.233–244, ACM (2012).

[16] Hiruta, S., Yonezawa, T., Jurmu, M. and Tokuda, H.: Detection, classification and visualization of place-triggered geotagged tweets, *UbiComp*, pp.956–963 (2012).

河野 慎

1991 年生．2010 年慶應義塾大学環境情報学部環境情報学科卒業．同年東京大学大学院学際情報学府総合分析情報学コース修士課程入学．主に，ユビキタスコンピューティングシステム，サイバーフィジカルシステムの研究に従事．

米澤 拓郎 （正会員）

1960 年生．2010 年慶應義塾大学 Ph.D.（政策・メディア）．現在，慶應義塾大学大学院政策・メディア研究科特任助教．主に，ユビキタスコンピューティングシステム，ヒューマンコンピュータインタラクション，センサネットワーク等の研究に従事．ACM，日本ソフトウェア科学会各会員．

中澤 仁 （正会員）

慶應義塾大学環境情報学部准教授．博士（政策・メディア）．ミドルウェア，システムソフトウェア，ユビキタスコンピューティング等の研究に従事．日本ソフトウェア科学会，IEEE 各会員．

川崎 仁嗣 （正会員）

株式会社 NTT ドコモ先進技術研究所勤務．2008 年筑波大学大学院システム情報工学研究科博士前期課程修了．同年（株）NTT ドコモ入社．モバイルコンピューティング，端末セキュリティ，分散システムに関する研究に従事．

太田 賢 （正会員）

1998 年静岡大学大学院博士課程修了．博士（工学）．1999 年 NTT 移動通信網（株）入社．現在，NTT ドコモ先進技術研究所勤務．モバイルコンピューティング，端末セキュリティ，分散システムに関する研究に従事．訳書『コンピュータネットワーク第 5 版』等．電子情報通信学会会員．

稲村 浩 （正会員）

NTT ドコモ先進技術研究所勤務．1990 年慶應義塾大学大学院理工学研究科修士課程修了．同年日本電信電話（株）入社．1994〜1995 年カーネギーメロン大学計算機科学科にて訪問研究員．1998 年より NTT ドコモ．2010 年慶應義塾大学大学院開放環境科学専攻後期博士課程単位取得退学．同大学博士（工学）．モバイル環境におけるシステムソフトウェア，トランスポートプロトコル，ユーザインタフェースに関する研究開発に従事．電子情報通信学会，ACM，IEEE 各会員．

徳田 英幸 （フェロー）

1975 年慶應義塾大学工学部卒業．同大学大学院工学研究科修士．ウォータールー大学計算機科学科博士（Ph.D. in Computer Science）．米国カーネギーメロン大学計算機科学科研究准教授を経て，1990 年慶應義塾大学環境情報学部に勤務．慶應義塾常任理事を経て，現職．専門は，ユビキタスコンピューティングシステム，OS，Cyber-Physical Systems 等．日本ソフトウェア学会フェロー．現在，情報処理学会副会長，日本学術会議会員，情報通信審議会委員等を務める．