

ヴォーカルマイノリティ現象を説明する 意見発信モデルの提案

鳥海 不二夫^{1,a)} 松澤 有¹ 鈴木 豊太郎^{1,2}

受付日 2016年10月3日, 採録日 2017年3月3日

概要: ソーシャルメディア上で頻繁に見られる意見と、世間一般における多数派の意見とは必ずしも一致しない。このような現象が発生する原因の1つにヴォーカル・マイノリティがあると考えられる。本研究では、ヴォーカル・マイノリティ現象が発生する要因を明らかにするために、ソーシャルメディア上でのユーザ行動のマルチエージェントモデルを提案した。提案モデルを用いたシミュレーションによってハブエージェントの影響がヴォーカル・マイノリティが発生する要因の1つであると考えられることを明らかにした。さらに、ソーシャルメディアの実データを用いて分析を行い、本シミュレーション結果が実データでも支持されることを確認した。

キーワード: ヴォーカルマイノリティ, エージェントベースシミュレーション, 意見形成モデル

Analysis of Vocal Minority Phenomena Based on Opinion Statement Model

FUJIO TORIUMI^{1,a)} YU MATSUZAWA¹ TOYOTARO SUZUMURA^{1,2}

Received: October 3, 2016, Accepted: March 3, 2017

Abstract: Data mining from social media has become an active area of study and commercial use. Researchers and analysts are digging deep into the mines of increasing user-generated contents and seeking valuable nuggets of information. However, there is an important but often overlooked factor that must be considered when analyzing social media logs: the existence of the silent majority and the vocal minority. This phenomenon is considered empirical fact within social environments and online media are no exceptions. Without considering the silent majority and the vocal minority, the analysis might reach incorrect conclusions, although the fact that no trace of the silent people shows up on logs complicates analysis. In this paper, we take a model-based approach to uncover the mechanisms behind the silent majority phenomenon with our proposed Vocal-Attitude Propagation Model based on the Linear Threshold Model to analyze the changes in the frequencies of vocal opinions. From our simulation results, we conclude that the influence of high degree agents and independence from local trends are the prominent conditions from which the silent majority and the vocal minority emerge.

Keywords: vocal minority, agent-based simulation, opinion statement model

1. 序論

ネット選挙運動が2013年4月に解禁されたのを受け、今後政治・思想的な話題の発信・議論もソーシャルメディア

上で行われることが増加すると予想される。しかしながら、ソーシャルメディア上で頻繁に見られる意見と、世間一般における多数派の意見とは必ずしも一致しないことがある。

こうした乖離の背景には声高に意見を発信する少数派勢力、いわゆるヴォーカル・マイノリティ (Vocal Minority)、あるいはノイジー・マイノリティ (Noisy Minority) [1] と呼ばれる存在が考えられる。ヴォーカル・マイノリティの

¹ 東京大学
The University of Tokyo, Bunkyo, Tokyo 113-8656, Japan

² IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA

a) tori@sys.t.u-tokyo.ac.jp

存在によって、本来は少数派にすぎない声がソーシャルメディア上では多数派のように見え、逆に多数派が沈黙しサイレント・マジョリティ (Silent Majority) [1] となる。その結果として、世間一般では多数派の声がソーシャルメディア上では少数派にみえてしまうようになる。

このような状況下で、実際にはサイレント・マジョリティ、すなわち「真の多数派」がどの程度いるのかを明らかにするためには、データを分析するだけでは難しい。そこで、我々はソーシャルメディア上のユーザの行動原理をモデル化し、シミュレーションによるアプローチによって、真の多数派を推定する手法の確立を目指す。

そのために、まず本論文では、ソーシャルメディアでのユーザの意見発信をモデル化する。仮想的なネットワーク上で意見発信の状況変化をシミュレーションにより実験・観察し、ネットワーク特性の差異によって生起する現象についての分析を行う。次に、実データの分析を通して、ウェブ上での意見発信を分析し、提案モデルの妥当性を検証する。具体的には、ソーシャルメディアの1つであるTwitterから、政治・選挙に関連する投稿のデータを取得し、意見傾向によりユーザを分類する。意見別に分類したユーザの、意見発信行動から、モデルの妥当性を検討する。

これによって、ヴォーカルマイノリティを考慮した意見発信モデルを確立し、サイレントマジョリティの推定問題の足がかりを作ることを目指す。

2. 関連研究

ウェブ・マイニングの領域では、膨大なデータから意見や感情を抽出・解析するというアプローチがなされているが、逆に人間の意思決定、信念の獲得や意見の変化、集団における心理の働き、世論や常識の形成を数理的モデルによって表現し、そのメカニズムを明らかにしようとする試みも歴史が深い。Opinion Dynamics と呼ばれるこの研究分野は少なくとも1950年代には定式化が見られる。この頃に、意見を持つ個体 (エージェント) からなる集団が Consensus (統一見解) に達するメカニズムとして、加重平均モデル (線形モデル) が提案された [2], [3], [4]。

2値意見の線形モデルを用いた研究は非常に活発に行われ [5], [6], [7], [8], 多くの条件において Uniformity (集団の画一化) が観測された。経済学の文脈でいう Herd 現象に相当する結果を示すと考えられている [7], [9]。また、影響の範囲を隣接するエージェントに限定したモデル (ネットワーク構造を導入したモデル) においては、大きく異なる意見がクラスター (固まり) をなして混在するという現象もみられた [8]。

意見の近いエージェントの集合からの平均的な影響を扱うのではなく、集団の中からランダムに1人ずつエージェントと対面し、2者間での意見の比較において確信の閾値を相互作用させるモデルもある [9], [10]。このようなペア単位

(Pairwise) での相互作用が集団全体に波及していくモデルは、カスケード (Cascade) モデルとして情報伝播の研究において知られている [11], [12]。

1対1の相互作用を基本とするカスケードモデルと対照的な情報伝播モデルが、閾値 (Threshold) モデル [13] である。群集行動 (デモ・暴動など) の発生メカニズムを説明するモデルとして Granovetter により提案されたが、人間社会における他の様々な現象についても適用の可能性はある [13]。閾値モデルの中でも最も単純で主要なモデルが線形閾値 (Linear Threshold) モデルである [12]。

さらに Granovetter 自身による研究 [14] では、社会ネットワークにおける意見表出の一現象である沈黙の螺旋 (The Spiral of Silence) [15] への閾値モデルの適用可能性が示されるなど、集団における意見形成の説明モデルとしても有力である。

本研究では、以上を考慮し先行研究の多くのモデルでは扱っていない内的意見と、外的態度の二重変数としてモデル化し、意見を発信している状態と、秘匿・沈黙している状態とを同時に扱えるような枠組みを提案する。

3. 意見発信モデルの提案と実験

3.1 意見発信モデル提案の目的と実験の概要

世論や選挙の分析・予測、それをもとにした意思決定や学習を行う際に、単純なデータ分析だけでは意見を積極的に発信することがないためデータに残らない人々の挙動を考慮に入れることができない。

そこで、意見発信についてのユーザ行動をモデル化しシミュレーションを行うことで人々の行動を推定する。特に、意見を発信している状態と、秘匿・沈黙している状態とを同時に扱えるような枠組みを比較的簡潔なモデルとして提案し、条件を変えたシミュレーションによってヴォーカルマイノリティ現象が発生する要因についての仮説を検証する。

3.2 従来モデルの特性と提案モデルの方向性

本研究では、ソーシャルメディアのユーザが意見の発信を行っている状態 (ヴォーカル) と、そうでない状態 (サイレント) とを扱うという目的がある。しかしながら、Opinion Dynamics の先行研究においては、普段から関係を有している周囲の他者や、ある時刻において出会った他者から影響を受ける相互作用の結果として、自身の意見そのものが変化するメカニズムについて扱っているものが多い。

このようなモデル化は、周囲に対する同調 (Conformity) という集団心理的現象を反映したものといえるが、そもそも他者からの影響 (社会的影響) に対する同調現象はその程度によって3つの段階に分けて考えることができる [16]。Compliance 他者や集団から望ましい反応・報酬を得る

ために、その影響を受け入れる、受け入れたように振る舞うこと。必ずしもその内容を完全に信頼し、受容するわけではないが、報酬の獲得や懲罰の回避によって満足する。

Identification 他者や集団との関係を維持・改善するために、その影響を受け入れること。過程でその内容を信頼していくが、具体的な部分は重要ではなく、他者に受容されること自体が満足をとまなう。

Internalization 影響の内容、理念や行動様式そのものが自己の価値観や需要と合致するために、それを受け入れること。内容自体が満足させるものであるために、それを信頼し、自己のものとする。

この段階は明らかに下へ行くほど強い同調を表している。特に Internalization と Compliance では、影響（意見）を実際にその個人が信頼して（自己のものとして）いるかどうか、まったく異なる可能性もある。先行研究における代表的な周囲から影響に対する同調のモデル（文献 [2], [3], [4], [17], [18], [19]）では、相互作用によって意見や情報などを表現した変数そのものの更新がなされていた。すなわち、その変化が Compliance に由来する外面上の変化にすぎないのか、Internalization に近い内面的変化をとまなっているのかは基本的に区別されていない。

集団の外面的な状態のみを扱うのであれば、このままのモデルでも機能上は問題ないといえる。しかし本研究では、集団に属する個人がどのような意見を内的に保持しており、それがソーシャルメディアや、選挙・社会調査など、異なった手段や場面で外的に表れる際、それらの表出の結果にどのような差異が生じるかを議論できるようなモデルの構築を目指している。

そこで、本研究では意見発信のモデルとして、従来の Opinions Dynamics のモデルを基盤としつつ、内的な状態（意見）と外的な状態（発信の有無）とを分離して扱うモデルを構築する。特に、ソーシャルメディアの参加・発言の手軽さ・気軽さは、外的な状態の変化（ある瞬間における情報発信の有無の切替え）を容易にする一方、内的な状態の変化はそれと比較して非常に遅い、という仮定をおく。Internalization と呼べるほどの意見の変化には、説得の繰返しや、周囲の勢力の長期にわたる安定的で一貫した表出、異なるメディアや権威者を通じた影響などが必要とされるからである*1。

外的態度が「発信」状態にある場合をヴォーカル、「秘匿」状態にある場合をサイレントとし、ヴォーカルなエージェントの内的意見は周囲から確認可能となるが、サイレントなエージェントの意見は確認できないとする。各エージェントは周囲のヴォーカルなエージェントの多数派意見

を認識し、自身の内的意見との比較によって、外的態度を決定する。提案モデルにおいて、あるエージェントの立場から観測可能な項目は、あくまで自己の内的意見と、他のヴォーカルなエージェントの内的意見のみである。

このように、内的意見と外的態度を二重の変数として持つモデルで、サイレントなユーザを含むソーシャルメディアにおける意見表出を議論する。次節で、実データの分析結果をふまえた具体的なモデルを規定する。

3.3 拡張 LT モデル

提案モデルでは、ソーシャルメディアにおける意見発信において、ユーザは隣接する周囲の意見に対し同調する傾向があると仮定する。このような行動を端的に表しうるモデルとしては、線形閾値 (LT) モデルがある [12], [13]。本研究では LT モデルを拡張し、外的態度と内的意見という二重の変数を扱うマルチエージェントモデルを提案する。

提案モデルでは、エージェント i は内的意見 OP_i に加え、外的態度 A_i を持つ。 OP_i は 2 値の離散値とし、各エージェントごとに固有に決定され、不変とする。 A_i は 2 値の離散変数で、 $A_i = 0$ あるいは $A_i = 1$ である。 $A_i = 1$ ならば、 i に隣接するエージェントは OP_i を確認することができる（ヴォーカル状態）。逆に $A_i = 0$ であれば確認できない（サイレント状態）。また、元の LT モデルと同じくエージェントは閾値パラメータ θ_i を持つ。

エージェント同士は無向エッジで接続された、無向グラフを構成している。エージェントは自己とエッジでつながっているエージェントを走査し、その中でヴォーカルなエージェントの意見をすべて取得したうえで、以下のような評価を行って自己の態度 A_i を決定する。

$$A_i = \begin{cases} 1 & \left(\frac{|\{j=\Gamma_i | OP_j=OP_i\}|}{|\{\Gamma_i\}|} \geq \theta_i \right) \\ 0 & (\text{otherwise}) \end{cases} \quad (1)$$

$\Gamma_i = i$ に隣接するヴォーカル・ユーザ

この式は、周囲で同じ意見を持つヴォーカル・ユーザの割合が閾値を上回るならば、自身もヴォーカルになる、という社会的影響をモデル化した相互作用である。なお、この際分母が次数ではなく隣接ヴォーカル・ユーザ Γ_i であるのは、ヴォーカルではないユーザの意見は見えないため、自身が少数派かどうかの判断の材料とはならないためである。逆の場合、すなわち自身の意見が周囲と比較して少数派にあたりと判断した場合は、同調して本意と異なる意見を表明するようになるのではなく、サイレントになり、単に意見を隠す。これはソーシャルメディアのような意見の表明・秘匿を容易に切り替えられる環境で、周囲との意見の衝突が想定される場合のユーザの行動 [23] と対応している。

*1 これら様々な影響を盛り込んだモデルとして Social Impact Theory [20], [21], [22] などがあり、内的意見の変化を積極的に扱おうとする場合にはこの系譜に沿うのが適当といえる。

3.4 シミュレーション条件

提案モデルによるマルチエージェントシミュレーションは、全エージェントの OP_i と θ_i を一様乱数で与えて初期化し、かつその中の一定割合を $A_i = 1$ に変化させることで開始される。この際、エージェント全体における $OP_i = 0$ の比率 P_0 と、最初に $A_i = 1$ となる比率 R をシミュレーションのパラメータとして与える。これらの初期条件にバイアスを持たせることで、実際の社会ネットワークで生じる諸状況に対応したシミュレーションを行う。

離散時間ステップ $t \in T$ において、全エージェントは隣接するヴォーカル・ユーザの意見を取得し、式 (1) に基づく相互作用によって態度 A_i を変化させる。全体に占める A_i の変化率が十分に小さくなったか、あるいは十分に長い t の経過をもって収束と見なし、試行を停止する。最後に、全ヴォーカル・エージェントに占める意見分布を、出力とする。これはヴォーカル・エージェントに占める意見分布と、ソーシャルメディアなどで取得可能なデータとして現れる可観測な意見分布とが、概念的に対応することによる。

エージェントが配置されるネットワーク構造もまた初期条件の1つである。ネットワーク生成モデルで構築した仮想的な社会ネットワークを使用し、モデルや指標などの条件を変化させて、ヴォーカルな意見表出にどのような傾向が生じるか観測する。

3.4.1 初期状態のバイアス

提案モデルの相互作用は、基本的には LT モデルと同様、(局所) 多数決的に決定される。したがって、単純にランダムな意見分布、閾値分布などを設定した場合は、全体における多数派がヴォーカル・ユーザにおいても多数派となる。そこで、以下のようなバイアスが存在するときにヴォーカルにおける多数派と全体の多数派とで乖離や逆転が生じるか、検討する。これによって、どのような原因でヴォーカルマイノリティ現象が発生する可能性があるのかを明らかにする。

高次数エージェントの少数派偏重

社会ネットワークの研究においては、多くの人との間に関係を持つ人、いわゆるハブ (Hub) と呼ばれる個体の影響力が議論されることが多い。本シミュレーションにおいても、高次数エージェントを少数派意見に偏らせるというバイアスを与えて、ハブの影響が見られるか、観察・検討する。

社会的影響の閾値制御

集団からの社会的影響に対し、耐性・頑健性を持っているような人が存在している。すなわち、社会的影響に対する反応の閾値には個体差がある、ということである。ここでは、少数派意見を持つエージェントの閾値 θ_i を変化させ、全体に及ぼす影響を検討する。

3.4.2 ネットワーク構造

社会ネットワークの構造が、集団全体としての意見表出

に与える影響をシミュレーション上で議論するために、複数のネットワーク構造を初期条件として与え、結果を比較する。

- スモールワールドネットワーク (WS モデル [24])
- CNN モデル [25]

4. シミュレーション結果と考察

4.1 シミュレーションパラメータ設定

各ネットワークはノード数 $N = 10000$ 、平均次数 $k = 10$ を目標に生成した。それぞれのネットワーク生成に用いたパラメータを表 1 に示す。なお、CNN モデルで生成されたネットワークにおける次数分布のべき指数は -1.382 であった。

また、初期条件は $P_0 = 0.40$ 、 $R = 0.10$ とした。したがって以降の実験結果では、 $OP = 0$ が全体における少数派意見に該当する。実験は同一の初期条件、ネットワーク構造に対し 10 回ずつ行い、結果の平均をとっている。

4.2 高次数エージェントによる影響

図 1 および図 2 は、高次数の少数派意見エージェントによる影響を調べたシミュレーションの結果である。図中横軸は少数派意見を持つ高次数エージェントの比率で、全体の中での次数順位で上位 $H\%$ までを少数派意見に指定、とした場合の $H \times 0.01$ の値である。縦軸はヴォーカルとなるエージェント数の 100 回の施行における平均である。

スケールフリー性を持たない WS モデル (図 1) においては、多数派意見がヴォーカルにおいても多数派であり続けている。一方、スケールフリー性を持つ CNN モデル (図 2) においては、比較的少ない割合の高次数エージェン

表 1 ネットワークパラメータ

Table 1 Network parameters.

	パラメータ
WS モデル	張り替え確率 $p = 0.1$
CNN モデル	実リンク化確率 $p = 1 - 2/k$

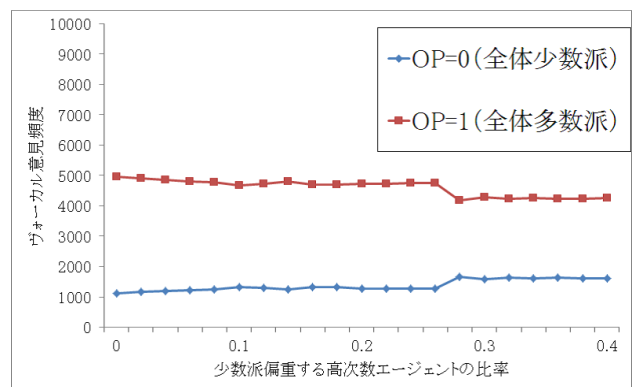


図 1 WS モデルでの高次数少数派の影響

Fig. 1 Influence of high-degree minorities (WS model).

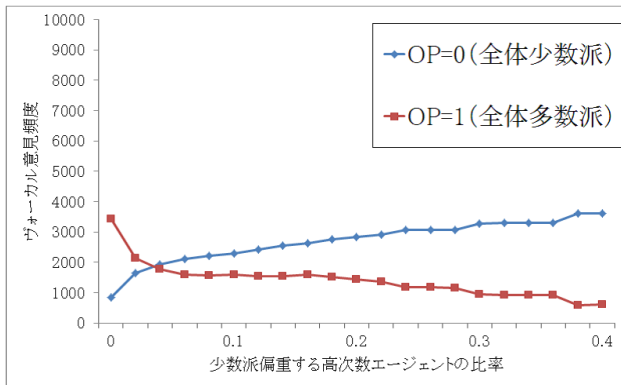


図 2 CNN モデルでの高次数少数派の影響

Fig. 2 Influence of high-degree minorities (CNN model).

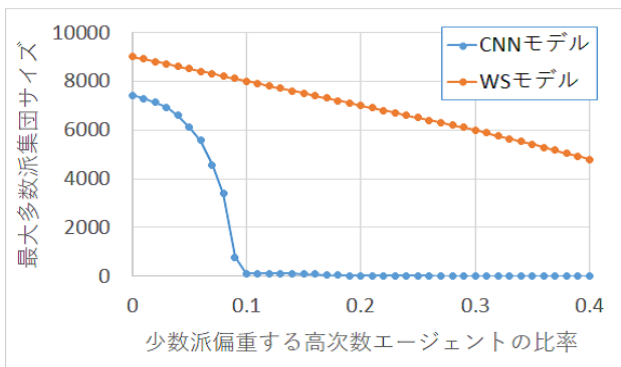


図 3 最大多数派集団のサイズ

Fig. 3 Size of majority.

トが少数派意見を持つだけで、ヴォーカルなユーザーにおける多数派意見の逆転が発生していることが分かる。

これはすなわち、スケールフリー性を持つネットワークにおいて、高次数エージェント（ハブ）はヴォーカルなエージェントにおける意見分布に強い影響を与えることを示している。スケールフリー性を持つグラフの特徴は、大多数のノードは接続次数（隣人の数）が低く、一部のハブノードが文字どおり桁違いに大きな次数を持っているという点である。このようなネットワークにおいてハブノードが特定の意見を持つということは、反対の意見を持つユーザーから見ればヴォーカルな反対意見に接する機会が多くなるというだけでなく、同意見を持つユーザーとのコネクションを阻害される効果がある。すなわち、同意見を持つユーザーのみからなるネットワークを考えると、ハブユーザーが選択的攻撃を受けていることに相当し、ネットワークが細かく分断される。図 3 に少数派偏重が進んだ場合の、多数派における最大集団サイズがどのように変化するかを示す。これを見ると、高次数ノードの 10% が少数派になるまでに、多数派ノードは細かい集団に分断されていることが分かる。その際のネットワーク模式図を図 4 に示す。この図では少数派ノード（図中水色のノード）がハブに多く配置されていることで、多数派ノードが分断されている様子を模式的に示している。さらに、ハブノードは多くの隣接ノード

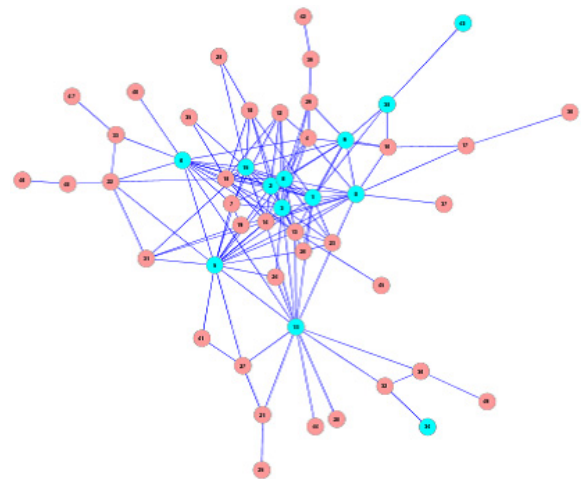


図 4 高次数ノードによるネットワークの分断

Fig. 4 Divided network by high-degree nodes.

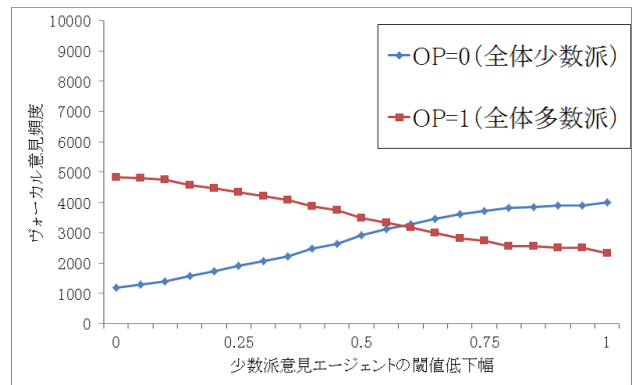


図 5 WS モデルでの少数派閾値の影響

Fig. 5 Influence of minority-threshold (WS model).

を持つため、その影響力は強く、結果として少数派であるにもかかわらず多数派の意見を抑える効果が生まれるといえる。

このようなハブノードの重要性は、現実においても有名人やオピニオンリーダーと目される人など、多くの人から意見を参照される立場にいる人物が、社会における意見形成に影響を与えていることを示唆している。

4.3 少数派の閾値制御による影響

次に図 5 および図 6 が、少数派意見を持つエージェントの閾値変化による影響を調べたシミュレーションの結果である。図中横軸は、少数派意見エージェントの閾値分布の最大値を θ_{MAX} まで低下、とした場合の $1 - \theta_{MAX}$ の値である。すなわち、横軸が大きくなるほど少数派エージェントの閾値 θ_i が低くなることを意味する。閾値が低くなるほど自身がヴォーカルになるために必要な周囲の同意見エージェントの数も減少する。逆にいえば、自分とは相反する周囲の多数派から社会的影響を受けても、少数でも同意見のエージェントが周囲に存在すれば態度を変更せずヴォーカルのままでいる可能性が高くなる。

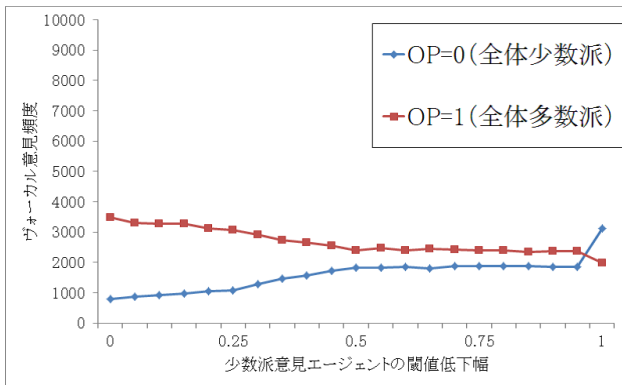


図 6 CNN モデルでの少数派閾値の影響

Fig. 6 Influence of minority-threshold (CNN model).

WS モデル (図 5) において、 $1 - \theta_{MAX} = 0.50$ 付近でヴォーカル多数派と全体多数派が逆転している。一方、CNN モデル (図 6) では、他と同じように 0.50 付近まで割合の接近はあるものの、そこで変化が止まり逆転には至っていない。

この結果によって、少数派エージェントが周囲から受ける (同調を促す) 社会的影響に対して頑健で、極端に言えば周りに誰も賛同者がいなくても自身の意見を発信し続ける、というような特性を持つ場合、ヴォーカルなエージェントにおける少数派の割合を増やすだけでなく、多数派の意見発信を抑える効果があることが明らかとなった。ただし、図 2 に見られるような、わずかな変化に対する急激な (ヴォーカル・エージェントの) マクロ傾向の変化や、その後の多数派意見の表出の強い抑制までは発生していない。むしろ、多数派意見がヴォーカルにおいて抑えられているといっても、その後もある程度の割合は保ち続けている。したがって閾値が与える効果、言い換えれば少数派意見を持つ頑健なユーザの存在が与える影響は、ハブユーザを持つ効果と比べると弱いと考えられる。

5. 実データとの比較

提案モデルを用いたシミュレーションにより高次数のエージェント (ハブ) の存在が、少数派意見のヴォーカル・エージェントの増大に影響を与えるということが示された。そこで本章ではツイートデータを用いて、提案モデルとシミュレーションから得られた上記のような知見が実際のソーシャルメディアでも確認されるかどうか、検証を行った。少数派における高次数エージェントの偏在が確認できれば、モデルの示す状況が現実合致したものであるといえる。

5.1 政治・選挙関連ツイートデータ

ここでは、2013 年 7 月に日本で行われた参議院議員選挙の時期に、政治・選挙に関連するツイートを検索・収集して、分析のためのデータセットとした。この際の各政党の

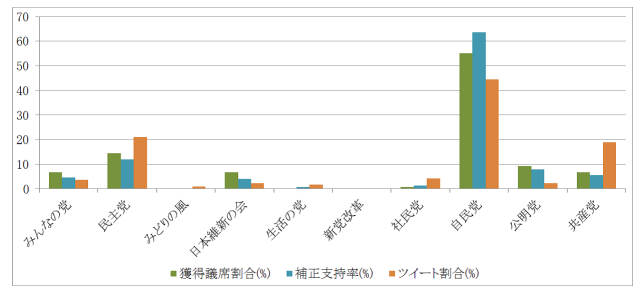


図 7 各政党の新規獲得議席, 支持率, ツイート割合の比較

Fig. 7 Seats, approval rate, and tweets of each party.

獲得議席, 支持率, Tweet 割合を図 7 に示す。政治・選挙関連のデータを収集する理由は、以下の 4 点である。

- (1) 本研究の背景である、ウェブと現実社会における意見表出の乖離という現象は、政治・選挙の文脈で語られることが多い。
- (2) 意見の傾向を読み解くうえで、政治・選挙関連の話題は賛成・反対、支持・不支持など極性を持ちやすいと考えられ、分析対象として適当である。
- (3) 選挙に近い時期は関連するツイートも増えると考えられるので、十分大きなデータセットを取得できる見込みがある。
- (4) 選挙結果から、全体多数派であるが意見発信者としては少数派となる場合などが実際に観測可能である。

データ収集は、2013 年 7 月 21 日の参議院選挙の前夜、7 月 17 日から 7 月 27 日にかけて行った。政党名を中心に、この選挙において争点とされていた政策に関連する語句を含めて検索を行った。また、検索に該当するツイートを行ったユーザを記録し、各ユーザごとに過去のツイートをさかのぼって取得してデータセットを拡充したうえで、最終的に全データに対し再度検索語句による絞り込みを行った。

検索収集の結果、全体で約 640 万件の政治・選挙関連ツイート (リツイートを含む) を収集し、これをデータセットとした。ユーザ総数は 237,605 人であり、これを政治・選挙に関心のあるユーザと見なす。

5.2 ツイートの分類

データセットから意見発信の傾向を読み取るためには、データに含まれるツイートがどのような意見を表現したものであるか、判定・分類が必要となる。登場する語句をもとにした話題による文章の分類や、それを応用した意見抽出や感情解析と呼ばれる分野で様々な手法が提案されているが、精度の点で十分でないとの指摘がある [26]。日本語のデータセットに対しても同様の研究が行われているが、やはり先行する諸研究と同様の問題に直面している [27], [28], [29]。

そこで本研究では、ソーシャルメディアにおける災害関

連情報の効果的な分類を目指す研究において提案された手法を導入した [30], [31], [32]. 本手法は投稿の内容ではなく、投稿の集合と、ユーザの集合、およびその間に生じる言及関係を示すエッジで構成される二部グラフに着目する。2つの異なる投稿に対する言及ユーザ集合の重なり度合いをもとに類似度を判定してエッジで接続し、投稿の類似度ネットワークを構成したうえで分類するという、非言語的手法である。

その結果、野党寄りの立場に立ったツイートが多く含まれる第1トピックと、与党寄りの立場のツイートで占められている第2トピックが抽出された。

すなわち、大まかには第1トピックが自民党不支持、第2トピックが自民党支持の意見集合であると規定できる*2. ソーシャルメディアのデータとして観測できるのは、あくまで意見を発信している（ヴォーカルな）ユーザの意見であることから、第1トピックがヴォーカル多数派の意見集合、第2トピックがヴォーカル少数派の意見集合といえる。ここで、実際の選挙結果（図7）を確認すると、自民党が新規獲得議席数において大きく勝っており、したがって実際の多数派は自民党支持、自民党不支持は少数派となる。

すなわち、第1トピック（自民党不支持）は「全体少数派、ヴォーカル多数派」の意見集合であり、第2トピック（自民党支持）は「全体多数派、ヴォーカル少数派」の意見集合といえる。これはまさに、全体における少数派（マイノリティ）がヴォーカル（なユーザ層において卓越）となっている状態、一般的にいわれるヴォーカル・マイノリティの表出が、この選挙においても現実となっていたことを示す結果である。

5.3 意見集合によるユーザ分類

ツイートの意見による分類を行ったうえで、分類されたツイート群それぞれへの言及頻度の違いから、ユーザを意見傾向に応じて分類する。5.2節ではツイートを「与党支持」「与党不支持」という2極の意見からなる集合に分類したが、ここでは、第1トピックの「与党不支持」意見（全体少数派、ヴォーカル多数派）を少数派意見（ $OP_i = 0$ ）、第2トピックの「与党支持」意見（全体多数派、ヴォーカル少数派）を多数派意見（ $OP_i = 1$ ）として扱う。ユーザ u_i の少数派意見の集合に属するツイートへの言及回数を $N_{i,0}$ 、多数派意見への言及回数を $N_{i,1}$ とした場合、 u_i 自身の意見 OP_i を以下のように規定する。

$$OP_i = \begin{cases} 0 & (N_{i,0} > N_{i,1}) \\ 1 & (N_{i,0} < N_{i,1}) \end{cases} \quad (2)$$

$N_{i,A} = N_{i,B}$ の場合は判断のつかないものとして除外する。

*2 実際にこれらのユーザが自民党支持か否かについては議論があると考えられるが、基本的な自民党の政策と各意見が一致していたため、このようなラベリングを行っている。

全ユーザに対しこの処理を行い、 $OP_i = 0$ を有するユーザ集合 $|U_0| = 40797$ と、 $OP_i = 1$ を有する $|U_1| = 12545$ を得た。この2集合いずれかに属しているユーザを、データ上のヴォーカル・ユーザと見なす。

5.4 隣接ユーザの影響

提案モデルでは、「周囲で同じ意見を持つヴォーカルユーザの割合が多ければ自身も意見を発信する」という仮定をおいている。そこで、このような仮定が実際に成り立つかを実データから明らかにする。

ここでは、極性意見による分類を行ったヴォーカル・ユーザに対し、そのネットワーク上における特徴に着目して分析を行った。

5.4.1 フォローネットワーク

フォローネットワークデータは、Watanabe らによって収集されたものを用いる [33]. データの収集は Twitter の REST API に対し、以下のような手続きからなる継続的なデータ取得操作により行われた。

- (1) 起点となるユーザのプロフィールと、フォローリスト（フォローされているユーザのリスト、Followed/Follower List）とフレンドリスト（フォローしているユーザのリスト、Following/Friend List）を取得。
- (2) フォロワリストに含まれるフォロワそれぞれについて、(1) を実行。
- (3) (1), (2) を、幅優先で繰り返す。

この手法によって、比較的網羅的なフォローネットワークの取得が可能である。意見ごとに分類されたユーザそれぞれについて、フォローネットワークデータ内から情報を参照し、特に隣接ユーザの意見傾向に着目して分析する。

5.4.2 隣接ユーザの意見傾向

まず、ヴォーカルなユーザの周囲（そのユーザが社会的影響を受けていると考えられる隣接ユーザの集団）の意見にどのような傾向があるかを分析する。ここでは、各ヴォーカル・ユーザについて、そのフォローリストを参照し、中に含まれているすべてのヴォーカル・ユーザについて、その平均意見を取得する。ただし、ヴォーカル・ユーザとは意見の2値分類が可能だったユーザを指す。したがってそれ以外のユーザはサイレント（意見を表明しない、周囲から確認できない状態）である。

ユーザ u_i の周囲の平均意見 \overline{OP}_i は単純に、

$$\overline{OP}_i = \frac{\sum_{j \in \Gamma_i} OP_j}{|\{\Gamma_i\}|} \quad (3)$$

$\Gamma_i = i$ に隣接するヴォーカル・ユーザ

として算出する。

少数派意見と多数派意見、それぞれを有するヴォーカル・ユーザについて、周囲の平均意見を区間頻度分布で示した

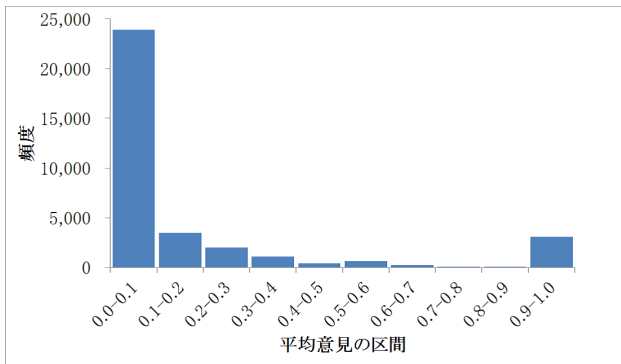


図 8 少数派意見 ($OP_i = 0$) ユーザに隣接するヴォーカル・ユーザの平均意見

Fig. 8 Average opinion of vocal users who is the neighbor of minorities ($OP_i = 0$).

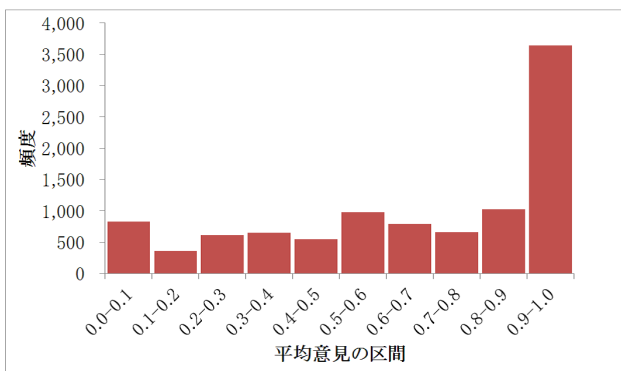


図 9 多数派意見 ($OP_i = 1$) ユーザに隣接するヴォーカル・ユーザの平均意見

Fig. 9 Average opinion of vocal users who is the neighbor of majorities ($OP_i = 1$).

のが、図 8 と図 9 である。基準となるユーザの意見と一致する意見を発信しているヴォーカル・ユーザが、いずれの意見の側においても圧倒的に多いことが読み取れる。

図 8 と図 9 が示したように、周囲のヴォーカルな意見の傾向と、ユーザのヴォーカルな意見とはデータ上おおむね一致する。すなわち、ユーザは自分の意見と周囲のヴォーカルな意見の傾向が一致している場合にヴォーカルになり、その逆ではないということを意味する。これは同調的行動の表れであり、沈黙の螺旋 [15] と呼ばれる現象を引き起こす要因ともされている。ソーシャルメディア上で議論を行うことについての調査 [23] から、周囲と意見が一致しない可能性がある場合、多くの人が議論に対し消極的となり、このように意見の表明を避ける傾向があることも分かっている。

以上より、提案モデルで用いた、周囲で同じ意見を持つヴォーカルユーザの割合が多ければ自身も意見を発信するというモデルが、実データからも支持されることが確認された。

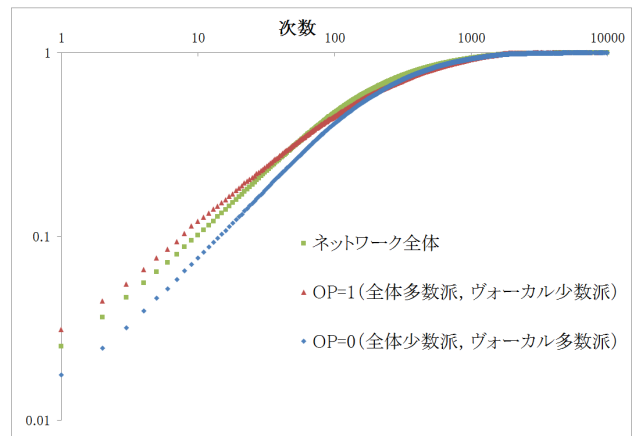


図 10 実データにおける、意見ごとの入次数頻度。累積確率分布
Fig. 10 In-degree distribution of each opinion in real data (Cumulative probability distribution).

5.5 ハブエージェントの効果

最後に、ハブエージェントの影響が実データにおいても表出されているかを分析する。

まず、極性意見による分類を行ったユーザ集合に属するユーザについて、フォローネットワークデータの情報を参照し、ユーザごとの被フォロー数を取得のうえ、頻度分布で分析する。

図 10 が、入次数頻度の分布である。累積確率分布として表現してある。横軸が次数、縦軸が確率となっており、累積確率分布であることから、データ系列はそれぞれの意見（あるいはネットワーク全体）に属するユーザが、特定の入次数以下の入次数を持つ確率を表す。少数派意見 ($OP_i = 0$) は第 1 トピックに由来する意見集合であり、「与党不支持」という意見傾向であった。したがって選挙結果と照らし合わせると、全体としては少数派、ヴォーカル・ユーザの中では多数派となる。一方第 2 トピック ($OP_i = 1$) はそれとは逆で、全体としては多数派、ヴォーカル・ユーザの中では少数派である。

少数派意見 ($OP_i = 0$) を述べたユーザの被フォロー数が 100 以下である割合は多数派意見やネットワーク全体よりも低く、また $500 \leq k_{in} \leq 1000$ 周辺の領域では多数派意見 ($OP_i = 1$) よりも高くなっている。すなわち、少数派意見のユーザ集合は、相対的に被フォロー数が少ないユーザが多く、被フォロー数が多いユーザ、すなわちハブユーザが多いことを意味する。

したがって、提案モデルによるシミュレーションが示した、ハブユーザが少数派意見 ($OP_i = 0$: 与党不支持) を支持するときに、ヴォーカル・ユーザは少数派意見が卓越する、という結果がデータにも表れているといえ、提案モデルによるシミュレーションの結果と合致している。もっとも、この符合はあくまでデータ上そのような傾向が出ており、提案モデルと一致しているというだけで、因果関係までも明示したものではないことには注意が必要である。

とはいえ、提案モデルに一定の妥当性を与える情況証拠とはなりうる。

6. 結論

本研究では、実社会とソーシャルメディア上の意見分布における乖離の原因をヴォーカルマイノリティの存在にあると考え、線形閾値 (LT) モデルを拡張し、エージェントの内的な意見と、外的な意見発信の態度とを分離した新たなモデルを提案した。提案モデルにより、高次数のエージェントに少数派意見を持たせる、少数派意見保持者の閾値を変化させる、といった条件を与えることで、ヴォーカル・ユーザにおける多数派が、ネットワーク全体における多数派と逆転するといった、実社会における意見分布とウェブ上での意見分布の乖離に相当する現象が再現可能であることが示された。

また、実験結果と実データの照合分析の結果、モデルによる実験結果との間に一致が見られ、本研究の提案モデルには一定の妥当性があることが確認された。

提案モデルは、Social Impact Theory などが行うように多くのパラメータを持たせなくても、現実のデータと合致するシミュレーション結果が得られた点で意義がある。加えて、内的な状態を隠しているエージェントを取り扱えるということは、モデルを用いたシミュレーションから、データとして表れることのないサイレントなユーザの規模や挙動を逆に予測・推定できる可能性を持っている。こうした内的状態の逆推定のような応用は、社会的影響による状態変化が Internalization か Compliance かを判断できない状態で議論される従来のモデルでは困難であり、本モデルの有用性といえる。

今後の課題としては、シミュレーションモデルの修正が考えられる。本論文では無向ネットワークを用いたモデルを用いているが、実データ分析で用いた Twitter は有向ネットワークであることから、有向ネットワークを用いたシミュレーションも必要であると考えられる。また、本論文では内的状態は変化しないという仮定をおいているが、現実には内的状態が変化することも考えられる。内的状態の変化を考慮したシミュレーションも必要である。

また、本論文ではすべてのユーザの意見 (OP_i) をランダムに決定していたが、実社会においては意見の近い者同士がつながる傾向が強いといわれている。したがって、そのような傾向を考慮したシミュレーションを行うことも重要である。しかしながら、本論文でも扱っている「意見を表明しないユーザ」の存在をどのようにモデル化すべきかは明らかになっていない。そのため、意見の近い者同士のつながりを仮定したうえでシミュレーションを行い、その結果と現実との整合性を比較することによって、意見を表明しないユーザの分布がどのようになっているのかを推定する「逆シミュレーション」を行うことも視野に入れる。

また、ツイート集合の意見極性による分類の精度議論があげられる。本研究ではこの手法による文書の話題分類精度を議論した先行研究の知見 [32] や、分類結果のランダムなサンプルの内容確認を前提として、極性による分類がなされているものとして扱ったが、人手による極性判断との比較や、自然言語処理に基づく感情分析手法との比較などの定量的な精度議論は必要である。

参考文献

- [1] King, A.A. and Anderson, F.D.: Nixon, Agnew, and the “silent majority”: A case study in the rhetoric of polarization, *Western Speech*, Vol.35, No.4, pp.243–255 (1971).
- [2] French Jr., J.R.: A formal theory of social power, *Psychological review*, Vol.63, No.3, p.181 (1956).
- [3] Harary, F.: *A criterion for unanimity in French’s theory of social power* (1959).
- [4] Friedkin, N.E.: A formal theory of social power, *Journal of Mathematical Sociology*, Vol.12, No.2, pp.103–126 (1986).
- [5] Föllmer, H.: Random economies with many interacting agents, *Journal of Mathematical Economics*, Vol.1, No.1, pp.51–62 (1974).
- [6] Arthur, W.B.: *Increasing returns and path dependence in the economy*, University of Michigan Press (1994).
- [7] Orlean, A.: Bayesian interactions and collective dynamics of opinion: Herd behavior and mimetic contagion, *Journal of Economic Behavior & Organization*, Vol.28, No.2, pp.257–274 (1995).
- [8] Latané, B. and Nowak, A.: Self-organizing social systems: Necessary and sufficient conditions for the emergence of clustering, consolidation, and continuing diversity, *Progress in Communication Sciences*, pp.43–74 (1997).
- [9] Deffuant, G., Neau, D., Amblard, F. and Weisbuch, G.: Mixing beliefs among interacting agents, *Advances in Complex Systems*, Vol.3, No.01n04, pp.87–98 (2000).
- [10] Weisbuch, G., Deffuant, G., Amblard, F. and Nadal, J.-P.: Interacting agents and continuous opinions dynamics, *Heterogenous Agents, Interactions and Economic Performance*, pp.225–242, Springer (2003).
- [11] Goldenberg, J., Libai, B. and Muller, E.: Talk of the network: A complex systems look at the underlying process of word-of-mouth, *Marketing Letters*, Vol.12, No.3, pp.211–223 (2001).
- [12] Kempe, D., Kleinberg, J. and Tardos, É.: Maximizing the spread of influence through a social network, *Proc. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.137–146, ACM (2003).
- [13] Granovetter, M.: Threshold models of collective behavior, *American Journal of Sociology*, pp.1420–1443 (1978).
- [14] Granovetter, M. and Soong, R.: Threshold models of diversity: Chinese restaurants, residential segregation, and the spiral of silence, *Sociological Methodology*, Vol.18, No.6, pp.69–104 (1988).
- [15] Noelle-Neumann, E.: The spiral of silence a theory of public opinion, *Journal of Communication*, Vol.24, No.2, pp.43–51 (1974).
- [16] Kelman, H.C.: Compliance, identification, and internalization: Three processes of attitude change, *Journal of*

- Conflict Resolution*, pp.51-60 (1958).
- [17] DeGroot, M.H.: Reaching a consensus, *Journal of the American Statistical Association*, Vol.69, No.345, pp.118-121 (1974).
- [18] Lehrer, K.: Social consensus and rational agnology, *Synthese*, Vol.31, No.1, pp.141-160 (1975).
- [19] Hegselmann, R. and Krause, U.: Opinion dynamics and bounded confidence models, analysis, and simulation, *Journal of Artificial Societies and Social Simulation*, Vol.5, No.3 (2002).
- [20] Latane, B.: The psychology of social impact, *American psychologist*, Vol.36, No.4, p.343 (1981).
- [21] Nowak, A., Szamrej, J. and Latané, B.: From private attitude to public opinion: A dynamic theory of social impact, *Psychological Review*, Vol.97, No.3, p.362 (1990).
- [22] Latané, B. and L'Herrou, T.: Spatial clustering in the conformity game: Dynamic social impact in electronic groups, *Journal of Personality and Social Psychology*, Vol.70, No.6, p.1218 (1996).
- [23] Hampton, K., Rainie, L., Lu, W., Dwyer, M., Shin, I. and Purcell, K.: Social media and the 'spiral of silence', *Washington, DC: Pew Internet & American Life Project. Retrieved August*, Vol.27 (2014).
- [24] Watts, D.J. and Strogatz, S.H.: Collective dynamics of 'small-world' networks, *Nature*, Vol.393, No.6684, pp.440-442 (1998).
- [25] Vázquez, A.: Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, *Physical Review E*, Vol.67, No.5, 056104 (2003).
- [26] Pang, B. and Lee, L.: Opinion mining and sentiment analysis, *Foundations and Trends in Information Retrieval*, Vol.2, No.1-2, pp.1-135 (2008).
- [27] 飯田 龍, 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出を目的とした機械学習による属性—評価値対同定, 情報処理学会自然言語処理研究会予稿集, NL-165-4, pp.21-28 (2005).
- [28] 橋本和幸, 中川博之, 田原康之, 大須賀昭彦: センチメント分析とトピック抽出によるマイクロブログからの評判傾向抽出, 電子情報通信学会論文誌 D, Vol.94, No.11, pp.1762-1772 (2011).
- [29] 乾 孝司, 奥村 学: テキストを対象とした評価情報の分析に関する研究動向, 自然言語処理, Vol.13, No.3, pp.201-241 (2006).
- [30] Toriumi, F., Sakaki, T., Shinoda, K., Kazama, K., Kurihara, S. and Noda, I.: Information sharing on twitter during the 2011 catastrophic earthquake, *Proc. 22nd International Conference on World Wide Web Companion*, International World Wide Web Conferences Steering Committee, pp.1025-1028 (2013).
- [31] 鳥海不二夫, 篠田孝祐, 榑 剛史, 風間一洋, 栗原 聡, 野田五十樹: ネットワーク構造に基づく災害情報の分類, 人工知能学会全国大会予稿集, 人工知能学会 (2013).
- [32] 馬場正剛, 鳥海不二夫, 榑 剛史: 災害情報の分類の妥当性の評価, 人工知能学会全国大会論文集, Vol.28, pp.1-4 (2014).
- [33] Watanabe, M. and Suzumura, T.: How social network is evolving?: A preliminary study on billion-scale twitter network, *Proc. 22nd International Conference on World Wide Web Companion*, International World Wide Web Conferences Steering Committee, pp.531-534 (2013).



鳥海 不二夫 (正会員)

2004年東京工業大学大学院理工学研究科機械制御システム工学専攻博士課程修了, 同年名古屋大学情報科学研究科助手, 2007年同助教, 2012年東京大学大学院工学系研究科准教授。エージェントベースシミュレーション, ソーシャルメディア, 計算社会科学, ゲームにおけるAI等の研究に従事。人工知能学会, 電子情報通信学会, 日本社会情報学会各会員。博士(工学)。



松澤 有

2013年東京大学大学院工学系研究科システム創生学専攻修士課程入学。2015年同修了。エージェントシミュレーションの研究に従事。



鈴木 豊太郎 (正会員)

1975年生。2004年東京工業大学大学院情報理工学研究科・数理計算科学専攻博士課程修了。同年よりIBM Research 研究員。高性能計算, 大規模グラフ処理基盤, 大規模交通シミュレーション等ソフトウェアシステムに関する性能最適化の研究に従事。2015年より米国IBM トーマス・J・ワトソン研究所に所属。2009年より2013年まで, 東京工業大学大学院情報理工学研究科客員准教授, 2013年よりアイルランド国立大学客員准教授を歴任。2015年より, バルセロナ・スーパーコンピューティングセンター客員教授, 東京大学客員研究員を兼務。IEEE BigData 2017プログラム共同委員長を務める。