

多数決に基づく SNS 投稿の信憑性評価に関する一考察

クルボノウ ウルグベク¹ 小泉 佑揮¹ 長谷川 亨¹ 西垣 正勝² 大木 哲史² 河辺 義信³

概要: 大規模災害の発生時には、多数の被災者が救助活動を実施する消防などの救助機関に救助を要請するが、119 番通報などの緊急通信は、電話網の輻輳や、受付台の業務輻輳で不通となることがある。一方、災害発生時にも電話網と比較して、メッセージの到達率の高い Social Network Service (SNS) が、救助要請に利用されるようになりつつある。しかしながら、SNS に投稿された救助要請の記事は全て正しいわけではなく、信憑性の低い記事が見られる。これに対して、本稿では、救助要請などの記事の信憑性を、記事に関連する場所に位置するボランティアに検証させるクラウドソーシングを提案する。複数のボランティアの多数決により、検証結果を評価することで、ボランティアが悪意ある振る舞いを行うことを抑制する。

キーワード: クラウドソーシング、災害、SNS、多数決、信憑性

A Study on Evaluating Credibility of Social Media Posts based on Majority Voting

Abstract: In large scale disasters, many victims call emergency calls to help desks of emergency agencies like fire force departments; however, many of calls become failed due to congestions in telephone networks and those at help desks. An alternative way of communication is using Social Network Service (SNS) to post messages for requesting rescues. However, many posted messages are not credible and thus it is difficult to extract credible messages for really requesting rescues. This paper addresses crowd sourcing service where volunteers who stay at locations of interest validate credibility of such messages. A majority voting mechanism of volunteers to validate credibility prevent volunteers from voting maliciously.

Keywords: Crowd Sourcing, Disaster, SNS, Majority Voting, Credibility

1. はじめに

災害時に被災者の救助活動を行う消防士などのファーストレスポンドー (First Responder) に、救助機関の統括者から迅速な指令を送信したり、ファーストレスポンドーならびに救助機関が正確な情報を共有することが重要である。これに対して、筆者らは、多数の被災者が災害に関する情報や救援要請の記事を投稿する、Social Networking Service (SNS) を活用して、ファーストレスポンドーの救助活動を支援するフレームワークを検討している [1]。本フレームワークは、SNS に投稿された記事から信憑性の高いメッセージを抽出し、ファーストレスポンドーに配信する機構を、SNS の記事を解析する機能、Named Data Networking

(NDN) [2] 上でグループ通信を提供する Publish/Subscribe 通信 [3] をベースに設計している。

本フレームワークでは、SNS、具体的にはツイッターに投稿された災害関連の記事に対して、自然言語処理を行うことで、救助要請の記事を自動的に抽出し、その記事に基づいてファーストレスポンドーに救助を指令するシナリオを想定している [1]。一方、災害時に投稿される記事は全て正しいわけではなく、信憑性の低い記事が多数投稿されていることが、近年の災害時に投稿された記事の解析から明らかになっている [4-7]。救助要請の記事の集合に対して、機械学習を用いて、記事の信憑性、すなわち真に救助要請であったかどうかを判定する試みが行われているが [8,9]、まだ 100 パーセントの判定は達成していない。将来、判定誤りが少なくなったとしても、誤った判定でファーストレスポンドーを出勤されることを防ぐ必要があり、補完的な手法として、高い確率で判定する手法が必要である。

¹ 大阪大学

² 静岡大学

³ 愛知工業大学

これに対して、本フレームワークでは、救助要請が発生した場所の近隣に位置するユーザ（以降、ボランティアと呼ぶ）に記事の真偽の判定を依頼することで、上記の課題を解決する。以降このようなユーザをボランティアと呼ぶことにするが、判定の信憑性はボランティアの信頼性に依存する。消防団員のような信頼できるボランティアの判定の信憑性は高いが、SNS ユーザのような匿名性の高いボランティアの判定の信憑性は低い。一方、人数の観点では、多数の信頼できるボランティアの獲得は容易でないが、匿名性が高いボランティアは多数獲得できると考えられる。そこで、本フレームワークでは、自身の個人情報の一部をフレームワークに知らせるボランティア [10]、ならびに匿名のボランティアの双方を用いた、投稿の真偽判定法を検討している。

本稿では、後者の匿名のボランティアによる投稿の信憑性を判定する手法に焦点を当て、具体的には、投稿された記事に記載された位置の近隣にいるボランティアに、投稿の真偽についてを多数決を行うことで、信憑性を評価する投票システムを検討する。投票システムを、正しく投票したボランティアに報酬を与えるクラウドソーシングシステムとしてモデル化し、2種類の悪意を持つボランティアを想定して、投票による信憑性の評価結果、ならびに悪意あるボランティアの振る舞いを、シミュレーションを用いて評価した。

本稿の構成は以下の通りである。2章で、フレームワークとシナリオを説明し、3章でボランティアのセキュリティモデルと投票のシステムモデルを説明する。4章では、シミュレーションにより、投票による記事の信憑性の評価結果、ならびにボランティアの振る舞いを評価する。5章で関連研究を紹介し、最後に、6章で本稿をまとめる。

2. フレームワーク

本章では、災害時に救助機関とファーストレスポnderを支援するフレームワークについて、システムモデルならびにシナリオについて述べる。

2.1 システムモデル

本節では、文献 [1] で提案するシステムモデルを、図 1 に示すシナリオに基づいて述べる。

まず、システムを構成するエンティティは以下の通りである。

- 救助機関: 地方自治体、消防、警察など救助活動を実施する機関である。
- 救助機関の統括者 (Incident Authority::IA): 当該災害において、全ての救助機関を統括する責任者である。
- ファーストレスポnder (First Responder: FR): 救助機関を構成する機関の構成員で、災害現場で救助活動を行う。消防隊員、警察官、DMAT (Disaster Medical

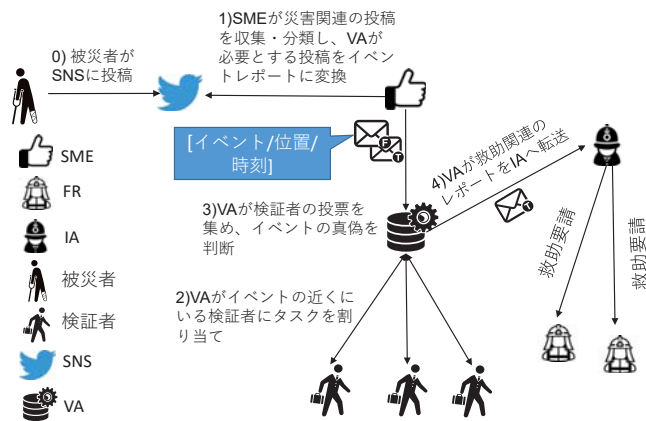


図 1 シナリオ

Assistance Team) などの構成員である。

- Socoal Media Engine (SME): SNS に投稿された災害関連の記事を自動的に抽出し、抽出した記事から災害情報をイベントとして抽出する。ここで、イベントは、[イベントの種別/位置/時間]3 つから構成される情報である。本稿では、イベント種別の内、救助要請に焦点をあてている。また、抽出法については、文献 [1] で述べている。
- 検証者 (Verifier): SME が抽出したイベントの信憑性を検証するボランティアである。
- 検証統括者 (Verifier Authority: VA): イベントの検証はクラウドソーシングサービスとして実現するため、VA は検証者の管理や、検証者への検証の依頼などクラウドソーシングサービスの管理を行う。
- 被災者: 災害関連の記事を SNS に投稿するユーザである。本システムでは、特に、救助要請の記事を投稿するユーザである。

2.2 シナリオ

本フレームワークで想定しているシナリオを図 1 に示す。まず、0) 被災者が SNS に救助要請の記事を投稿する。1) SME は継続的に SNS に投稿された災害関連の記事を収集する。収集した記事を解析し、救助要請と判断した記事からイベントレポートを作成する。イベントレポートには、救助要請などのイベント種別、位置、時刻が記載されている。救助要請のイベントレポートについては、その信憑性を評価するため VA へ転送する。2) VA はイベントレポートを受け取ると、イベントレポートの位置に近い位置にいる検証者に、イベントレポートの信憑性の真偽を評価するタスクを割り当てる。ここで、検証者は VA が管理するクラウドソーシングシステムのワーカーであり、事前に登録している。検証者は、イベントレポートの真偽について、「真」、「偽」、「不明」のいずれかの投票を行う。3) VA は、全検証者からの投票の多数決を取る。4) 多数決の結果、「真」の投票が多数を占めた場合、真と判断して、このイベント

レポートを IA を転送する。IA から救助依頼がファーストレスポンスに転送され、救助活動が開始される。

3. 投票システム

3.1 位置ベースのクラウドソーシング

検証者による投票を位置ベースクラウドソーシング (Spatial Crowdsourcing) [11] のタスクとして、投票システムを構築する。検証者が投票に参加するインセンティブとしてタスク、すなわち投票ごとに報酬を与えることにする。本稿では、投票により記事の信憑性を確認する手法の有用性を確認することを目的として、以下に示す投票システムを検討した。

- タスク：タスクはイベントレポートに示す位置でイベントが発生したかどうかについて、検証者が投票することである。検証者は、イベントの発生があると判断した場合は「真」、発生していない場合は「偽」、分からない場合は「不明」を投票する。
- 検証者へのタスク割り当て：クラウドソーシングに参加する検証者は、通常の位置ベースクラウドソーシング (Spatial Crowdsourcing) [11] のように、定期的に自身の位置を VA に報告することを仮定する。VA は SME からイベントレポートを受信すると、その位置からあらかじめ定めたしきい値の距離内に位置する検証者に、投票のタスクを要求する。
本稿では、近接した検証者は、目視確認でイベントの発生を正確に確認できることを仮定している。
- 検証者の位置：3.2 に示すように、検証者は偽の (フェーク) な識別子を生成できないことを仮定しているため、悪意ある攻撃者がフェークな検証者を偽装して作成することはできない。
- 投票結果：投票は多数決を実施し、「不明」を除いて、「真」あるいは「偽」の投票数が多かった方を結果とする。
- 報酬：1 つのタスクに対して、1 単位の報酬 (コインと呼ぶ) を割り当て、投票に勝った検証者に 1 コインを与え、負けた検証者から 1 コインを削除する。コインについて、具体的には決めていないが、仮想通貨のような匿名のコインを想定している。
また、4 章の評価では、コインの初期値として、検証者に 2 コインを割り当てる。

3.2 セキュリティモデル

検証者となるボランティアは、災害直後に短時間で募集する必要がある。従来の地方自治体に事前登録する手法では十分な検証者を集められず、また、災害直後に地方自治体に出向いて登録する方法は時間がかかりすぎる。これに対して、本フレームワークでは、以下の 2 種類の登録法を想定し、投票システムには後者を用いる。

1 つ目の登録法は、信頼できるファーストレスポンスと遭遇したボランティアが、バイオメトリクス情報などのトレース可能な情報を提出することで、登録する手法である [10]。本ボランティアの信頼性については、文献 [10] で議論している。

2 つ目の登録法は、本稿で採用する手法であり、ボランティアはネットワークを介して、登録を申請する。個人情報提出することを要求するような公開鍵証明書を用いる手法では、多数の検証者を集められない可能性があるため、self-certified な証明書により、申請することを可能とする。ただし、証明書の公開鍵をスマートフォンの秘密情報を秘密鍵として用いることを想定する。

この結果、システムのエンティティのセキュリティモデルは以下となる。

- IA、FR、SME および VA はオネスト (信頼できる) である。また、救助機関が発行する証明書を持っていることを仮定する。
- 検証者はディスオネスト (信頼できない) であり、self-certified で匿名な識別子を有する。検証者の振る舞いは、3.3 で定義する。
- 被災者は、ディスオネストである。SNS のアカウントを有するが、フレームワークの証明書や匿名な識別子を持たない。

3.3 検証者のモデル

本稿の投票システムでは、検証者の能力として、イベントの位置からしきい値の距離、例えば、100 メートル内のイベントの真偽を正確に判定でき、それ以上の距離のイベントは判定できないと仮定する。この仮定において、検証者の振る舞いとして、以下の 3 種類のモデルを想定する。

- オネストな検証者：信頼できる検証者であり、常に正確に投票する。具体的には、イベントを目視確認できる場合は、目視した結果に応じて、「真」または「偽」を投票する。一方、イベントの位置から離れており、目視確認できない場合は、「不明」を投票する。
- Gold Digger (GD) な検証者：正確な投票することではなく、報酬の最大化を目的とする検証者である。多数決に参加する検証者が正解とは異なった投票すると予想すれば、誤った投票を行う。すなわち、イベントレポートが「真」の場合は「偽」に、「偽」の場合には真に投票する。

本稿では、GD な検証者の投票を、変数 x_m を用いてモデル化する。変数 x_m は、その GD な検証者が誤った投票をする確率であり、例えば、 $x_m = 0.7$ の場合、0.7 の確率で誤った投票を行う。初期値は 0.5 で、投票の度に 0.1 ずつ増減する。正しく投票して多数決に勝った場合、0.1 減じる。一方、誤った投票をして勝った場合は、0.1 増加させる。この結果、正確に投票する

検証者が多い場合は、 x_m は 0 に収束していき、逆に、誤った投票をする検証者が多い場合は、 x_m は 1 に収束していく。

- **Goofな検証者:** 真剣に投票を行わない検証者で、確率 1/2 で「真」または「偽」に投票する。イベントからしきい値の距離内にいても、確率 1/2 で「真」または「偽」に投票する。

4. 評価

本章では、シミュレーションを用いて、オネストな検証者と、GoofあるいはGDな検証者が混在する場合の多数決の結果、ならびにGDの振る舞いを評価する。

4.1 シミュレーション条件

シミュレーション条件は以下の通りである。

- シミュレーションエリア: イベントが発生するエリアは、一辺 4km の正方形。
- 検証者の人数: 1600 人。
- 投票エリア: イベントを目視で判定可能な距離として、100メートル、ならびに 180メートルを想定する。すなわち、イベントレポートの位置を中心とした、半径 100メートル、ならびに 180メートルの円内の検証者はイベントを目視で判定可能である。

本シミュレーションでは、それぞれの条件において、平均で 3名、ならびに 10名の検証者が投票することとなる。

- 検証者数: エリア内にいる投票者の人数が 3人未満の場合、イベントの位置に近い 3人目の検証者が投票する。
- イベントレポートの発生: エリア内のランダムな位置に発生させる。
- イベントレポートの発生数: イベントレポートを 1000件、ならびに 9000件発生させる。

イベントレポートの発生数は、ハリケーン Harvey の投稿数 [12] 参考に決定した。ハリケーンが襲った日にハリケーン Harvey のキーワードを含む 125,000 件の記事が投稿され、そのうち 8割が災害関連であった。ハリケーン Harvey が襲ったヒューストン市の面積がこのシミュレーションエリアの面積の 100倍であることより、短時間で投稿される 1日分の記事数として、1000件を採用した。

4.2 オネストと Goof な検証者

Goofな検証者は時間の経過につれて振る舞いを変更することはないため、1000回のイベントレポートに対して、Goofな検証者の割合を 0.0 から 0.9 まで変更して、投票結果の正確性を評価した。検証者が 3名と 10名の場合の正解率 (rate) を、図 2 および図 3 に示す。ここで正解率は、全てのイベントレポートは真であるため、投票結果が真と判定された割合である。なお、x 軸の dishonest rate は検

証者全体に対する Goof な検証者の割合である。Goof な検証者は 1/2 の確率で、真か偽に投票するため、その割合が多くなっても、一定のオネストな検証者が存在すると、正しい投票結果となる確率が高くなっている。

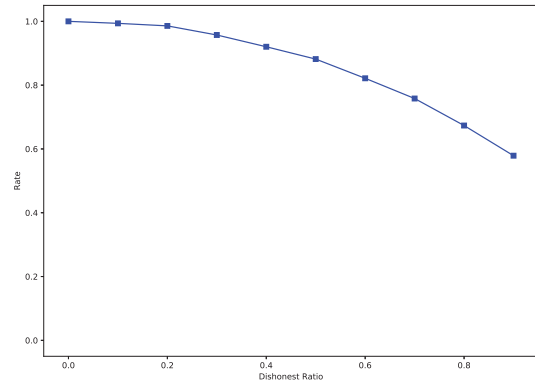


図 2 オネストと Goof の場合の正解率 : 3 名の検証者

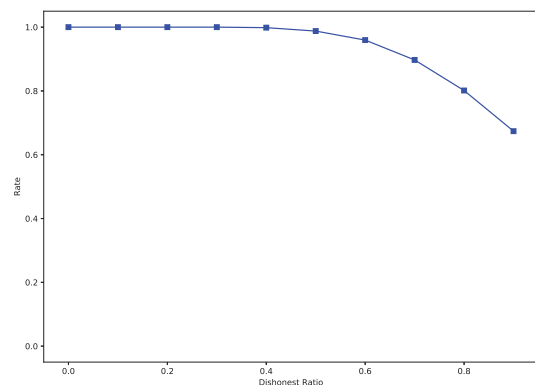


図 3 オネストと Goof の場合の正解率 : 10 名の検証者

4.3 オネストと GD な検証者

4.3.1 検証者数の影響

検証者数の正解率への影響を知るため、1000件のイベントレポートに対して、検証者が 3名と 10名の場合の正解率を評価した。検証者が 3名と 10名の場合の、正解率を図 4 および図 5 に示す。Goof な検証者が混在する場合、正解率は同様な傾向を示している。ただし、投票するエリアが広く、検証者の人数が増えると、GD の検証者がどのような割合でも、正解率が高くなる。これは、検証者が投票する回数が少ない場合、変数 x_m を更新する機会が少ないためと考えられる。以下では、投票回数の影響を知るため、9000回のイベントレポートに対する GD な検証者の振る舞いを評価する。

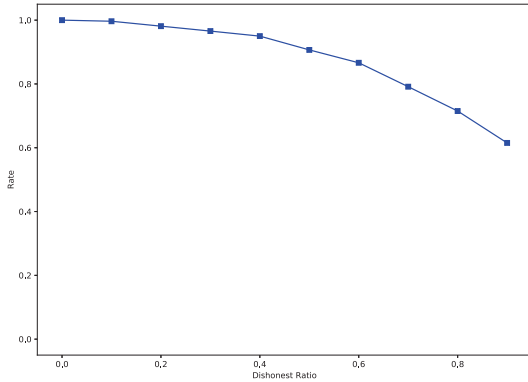


図 4 オネストと GD の場合の正解率: 3 名の検証者

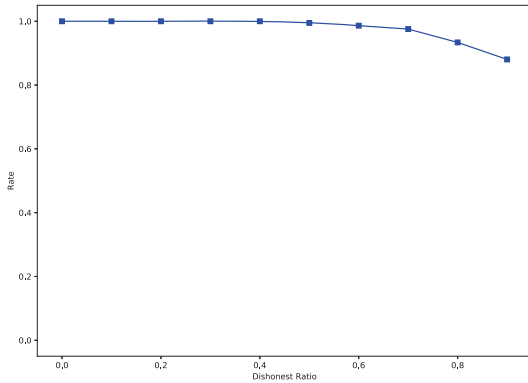


図 5 オネストと GD の場合の正解率: 10 名の検証者

4.3.2 GD な検証者の振る舞いの変化

図 6 に検証者が 3 名の場合に、9000 回のイベントレポートに対して投票を行う場合の、 x_m の変化を占めず。ここで、GD な検証者の検証者全体に占める割合は 50%である。図には、800 名の GD な検証者について、 x_m の平均 (average)、ならびに最大 (Maximum)、最小 (Minimum) の x_m を示している。図に示すように、 x_m の平均は 0 に収束してゆく、すなわち、必ず正しく投票するオネストな振る舞いになってゆく。ただし、本シミュレーションの条件で、1000 回のイベントレポートでは十分に収束せず、4000 回以上の投票が必要となっている。

図 7 には 10 件の投票ごとに投票が成功した回数の変化を示す。この図から最初のうち誤って判定される投票があっても、時間とともにすべての投票が正しく判定される。正しい投票結果の増加は、図 6 に示す x_m の減少と符合している。

4.4 オネスト、Goof、GD な検証者

最後に、オネスト、Goof、GD な検証者が混在した条件でシミュレーションを実施した。

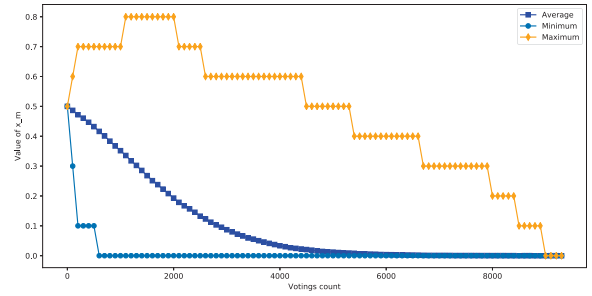


図 6 x_m の最大・最小・平均値: 3 名の検証者

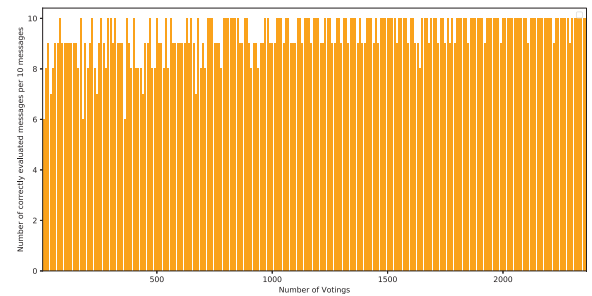


図 7 投票結果の正確さ: 3 名の検証者 (2500 件目以降の投票の結果の変化がないため省略)

- シミュレーションエリア: 一辺 6km の正方形。
- 検証者の人数: 1300 人。
- 投票エリア: 中心がイベントの位置で、半径 500 メートルの円。
- 検証者の割合: Goof の割合が 0.0 から 0.9 まで変更し、残りのオネストな検証者と GD の人数は等しい。
- イベントレポートの数: 1000 件。

検証者における Goof の割合を変化させた場合の正解率を図 8 に示す。

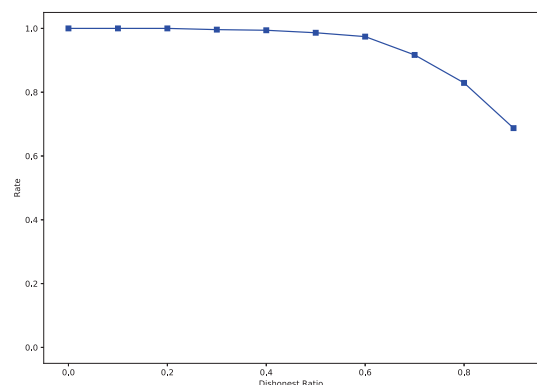


図 8 Goof の割合を変化させた場合の正解率

さらに、Goof の割合が 1/2 の場合の、投票回数に対する x_m の変化を図 9 に示す。Goof と GD な検証者が混在して

も、同様に高い正解率を達成している。

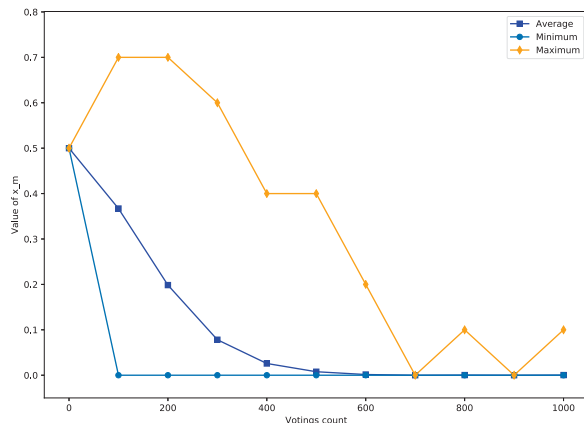


図 9 Goof が 1/2 の場合の x_m の変化

5. 関連研究

災害時に SNS に投稿された記事の解析が進められている。文献 [12] では、3つの巨大なハリケーンの時に Twitter に投稿された情報を分析し、ハリケーン Harvey が米国のテキサス州を 2017 年 8 月 25 日に襲ったハリケーンである。そこから 12 日間で 670 万件以上のハリケーン関連のキーワードが入った記事が投稿され、その中の 60% が災害関連の投稿であったことが記載されている。従って、SNS が災害関連情報を展開するのに有用である。しかしながら、SNS 上に多数のフェイクニュース (デマ、噂) を含む投稿が流通しており、その信憑性の評価が重要となっている。これに対して、本稿では、投稿された記事を自然言語で解析するのではなく、記事に関連する位置の近くのボランティアに検証を依頼するアプローチを取っている。ボランティアへの検証依頼は、一般的な位置ベースのクラウドソーシングのフレームワーク [11] に従っている。

一方、本稿のように、救助要請の記事を、具体的には、ツイッターに投稿された記事を機械学習を用いて抽出する試みが行われており [8,9]、本稿で提案した手法と相補的な手法となっている。

6. まとめ

本稿では、SNS に投稿された救助要請の記事の信憑性を評価するクラウドソーシングベースの手法を提案した。具体的には、検証者と呼ぶワーカーに、記事が示すイベントを目視で判定させ、その結果を複数の検証者の多数決による信憑性を判断する。悪意のある検証者が存在する環境で、高い確率で正しい投票結果が得られることを、シミュレーションを用いて確認した。

参考文献

- [1] M. Jahanian, T. Hasegawa, Y. Kawabe, Y. Koizumi, A. Magdy, M. Nishigaki, T. Ohki, and K. K. Ramakrishnan, "Direct: Disaster response coordination with trusted volunteers," in *6th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*.
- [2] V. Jacobson, D. K. Smetters, J. D. Thornton, M. Plass, N. Briggs, and R. L. Braynard, "Networking named content," in *Proceedings of ACM CoNEXT*, pp. 1–12, Dec. 2009.
- [3] 村井穂永, 小泉佑揮, 長谷川亨, "分断された ndn ベースのアドホック網における publish/subscribe 通信プロトコルに関する一考察," 信学技報, no. 81, 2020.
- [4] S. Vieweg, A. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: What twitter may contribute to situational awareness," vol. 2, pp. 1079–1088, 01 2010.
- [5] J. Dugdale, B. Walle, and C. Koeppinghoff, "Social media and sms in the haiti earthquake," *WWW'12 - Proceedings of the 21st Annual Conference on World Wide Web Companion*, 04 2012.
- [6] S. E. Middleton, L. Middleton, and S. Modafferi, "Real-time crisis mapping of natural disasters using social media," *IEEE Intelligent Systems*, vol. 29, pp. 9–17, Mar 2014.
- [7] Y. Tim, S. L. Pan, P. Ractham, and L. Kaewkitipong, "Digitally enabled disaster response: The emergence of social media as boundary objects in a flooding disaster," *Information Systems Journal*, 10 2016.
- [8] 川崎凌摩, 松下光範, 宋晨潔, 藤代裕之, "Twitter からの救助要請の抽出と検証—2018 年 7 月の西日本豪雨災害ツイートを対象として—," in 第 11 回データ工学と情報マネジメントに関するフォーラム最終論文集, no. 5, Aug. 2019.
- [9] 山根有輝也, 小泉 佑揮, 長谷川 亨, "救助要請抽出のための災害時ツイートの解析," 第 182 回情報処理学会「マルチメディア通信と分散処理研究会 (DPS) 研究報告.
- [10] 北川沢水, 向平浩貴, 上原航汰, 大木哲史, 小泉佑揮, 河辺義信, 長谷川亨, 西垣正勝, "生体情報を用いた抑止力型トラスト: 災害時通信の信頼性向上のための仕組みの検討," *コンピュータセキュリティシンポジウム 2019*.
- [11] Y. Zhao and Q. Han, "Spatial crowdsourcing: Current state and future directions," *IEEE Communications Magazine*, vol. 54, pp. 102–107, 07 2016.
- [12] F. Alam, F. Offi, and M. Imran, "A twitter tale of three hurricanes: Harvey, irma, and maria," 05 2018.