

表 6： 省庁アカウントリスト

アカウント	政府機関や省庁名
kantei	首相官邸
Kantei_Saigai	首相官邸/首相官邸（災害情報）
naikakukanbo	内閣官房
NPA jinjiin	人事院
jftc	内閣府/公正取引委員会
NPA_KOHO	内閣府/警察庁
fsa_JAPAN	内閣府/金融庁
caa_shohishacho	内閣府/消費者庁
fukkochi_JAPAN	復興庁
MIC_JAPAN	総務省
FDMA_JAPAN	総務省/消防庁
eGovJapan	総務省/他/総務省行政管理局電子政府グループ
MOJ_HOUMU	法務省
MOJ_PSIA	法務省/公安調査庁
MOJ_HOGO	法務省/法務省保護局
MOF_Japan	財務省
NTA_Japan	財務省/国税庁
mofajapan_jp	外務省
MofaJapan_ITPR	外務省/他/外務省やわらかツイート（IT 広報室）
Mofajinjicenter	外務省/他/外務省国際機関人事センター
global_forum	外務省/他/軍備管理軍縮課
mextjapan	文部科学省
prmag_bunka	文部科学省/文化庁
mextlive	文部科学省/他/文部科学省ライブチャンネル
mextthiroba	文部科学省/他/文部科学省情報ひろば
MHLWitter	厚生労働省
MAFF_JAPAN	農林水産省
meti_NIPPON	経済産業省
jpo_NIPPON	経済産業省/特許庁
meti_chusho	経済産業省/中小企業庁
check_meti	経済産業省/他/経済産業省ホームページ更新情報
openmeti	経済産業省/他/経済産業省情報プロジェクト室
meti_saiene	経済産業省/他/なっとく！再生可能エネルギー
MLIT_JAPAN	国土交通省
Kanko_Jpn	国土交通省/観光庁
JCG_koho	国土交通省/海上保安庁
Kankyo_Jpn	環境省
bouei_saigai	防衛省
jmsdf_pao	防衛省/海上自衛隊
JASDF_PAO	防衛省/航空自衛隊
JGSDF_pr	防衛省/陸上自衛隊

表 7: 調査に用いたトピックと検索クエリ

トピック	検索クエリ
デング熱	デング, ヒトスジシマカ
エボラ出血熱	エボラ
鳥インフルエンザ	鳥インフル
ノロウイルス	ノロウイルス, ノロウィルス
子宮頸がん予防ワクチン	サーバリックス, ガーダシル, 子宮頸, 頸がん, 頸癌
イクメンプロジェクト	イクメンプロジェクト
健康づくり大キャンペーン	健康づくり大キャンペーン
健康寿命をのばそう！アワード	スマート・ライフ・プロジェクト, 健康寿命をのばそう+アワード
危険ドラッグ・脱法ドラッグ	危険ドラッグ, 脱法ドラッグ, 違法ドラッグ, 脱法ハーブ
イレッサ東京訴訟	イレッサ+訴訟, ゲフィチニブ
国民の皆様の声	国民の皆様の声+厚労省, 国民の皆様の声+厚生労働省

表 8: 収集したツイート数と厚労省を含む割合

トピック	ツイート数	厚労省を含む 数と割合
エボラ出血熱	2,715,763	70,988 (2.61%)
イクメンプロジェクト	1,385	447 (32.27%)
イレッサ東京訴訟	5,766	69 (1.20%)
デング熱	1,200,231	43,209 (3.60%)
ノロウイルス	432,319	2,230 (0.52%)
健康づくり大キャンペーン	2,688	17 (0.63%)
危険ドラッグ・脱法ドラッグ	1,077,863	19,516 (1.81%)
健康寿命をのばそう！アワード	411	181 (44.04%)
国民の皆様の声	1,355	1,355 (100.00%)
子宮頸がん予防ワクチン	698,744	69,867 (10.00%)
鳥インフルエンザ	444,936	4,701 (1.06%)

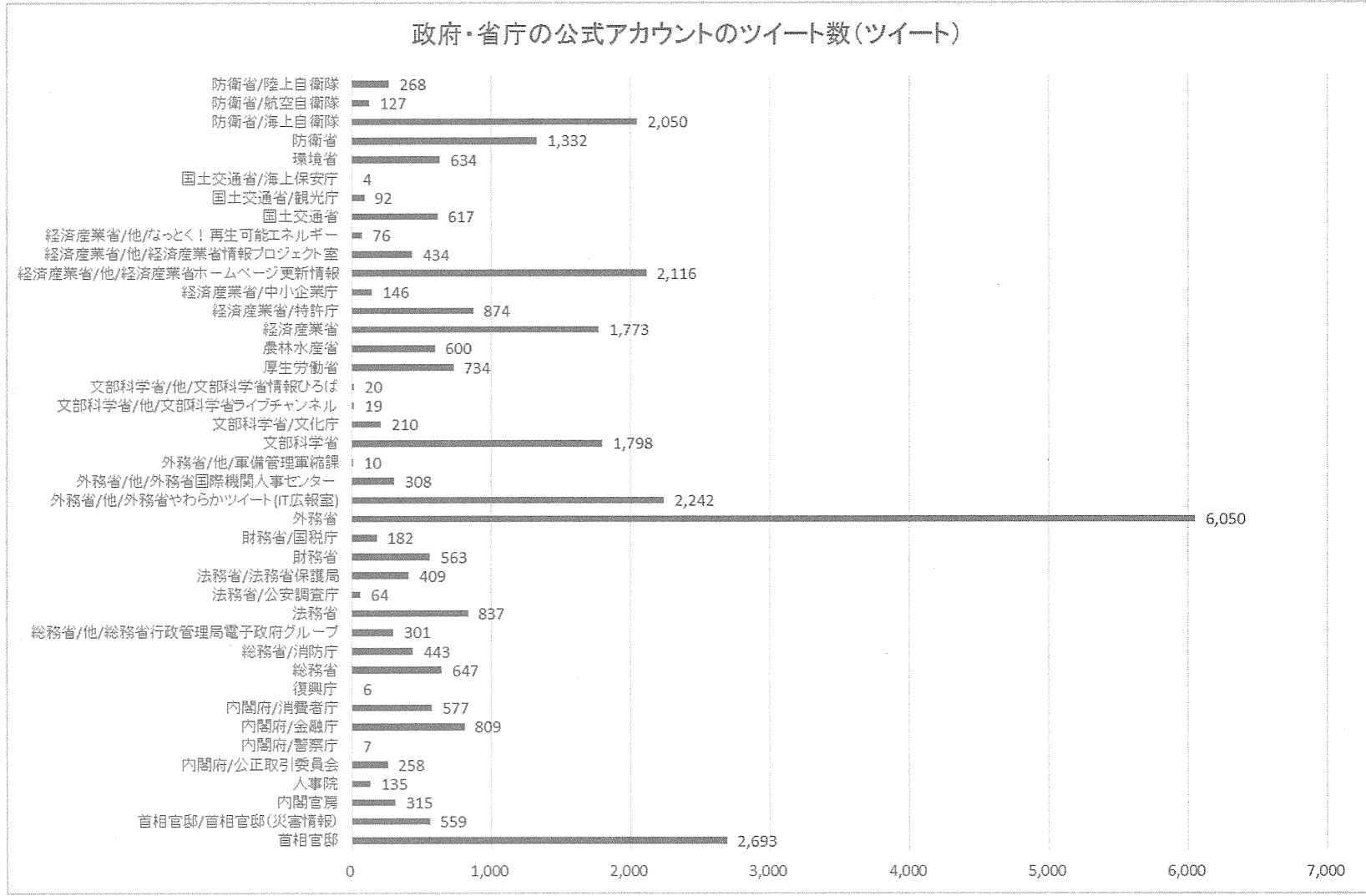


図 6: 政府・省庁の公式アカウントのツイート数(ツイート)

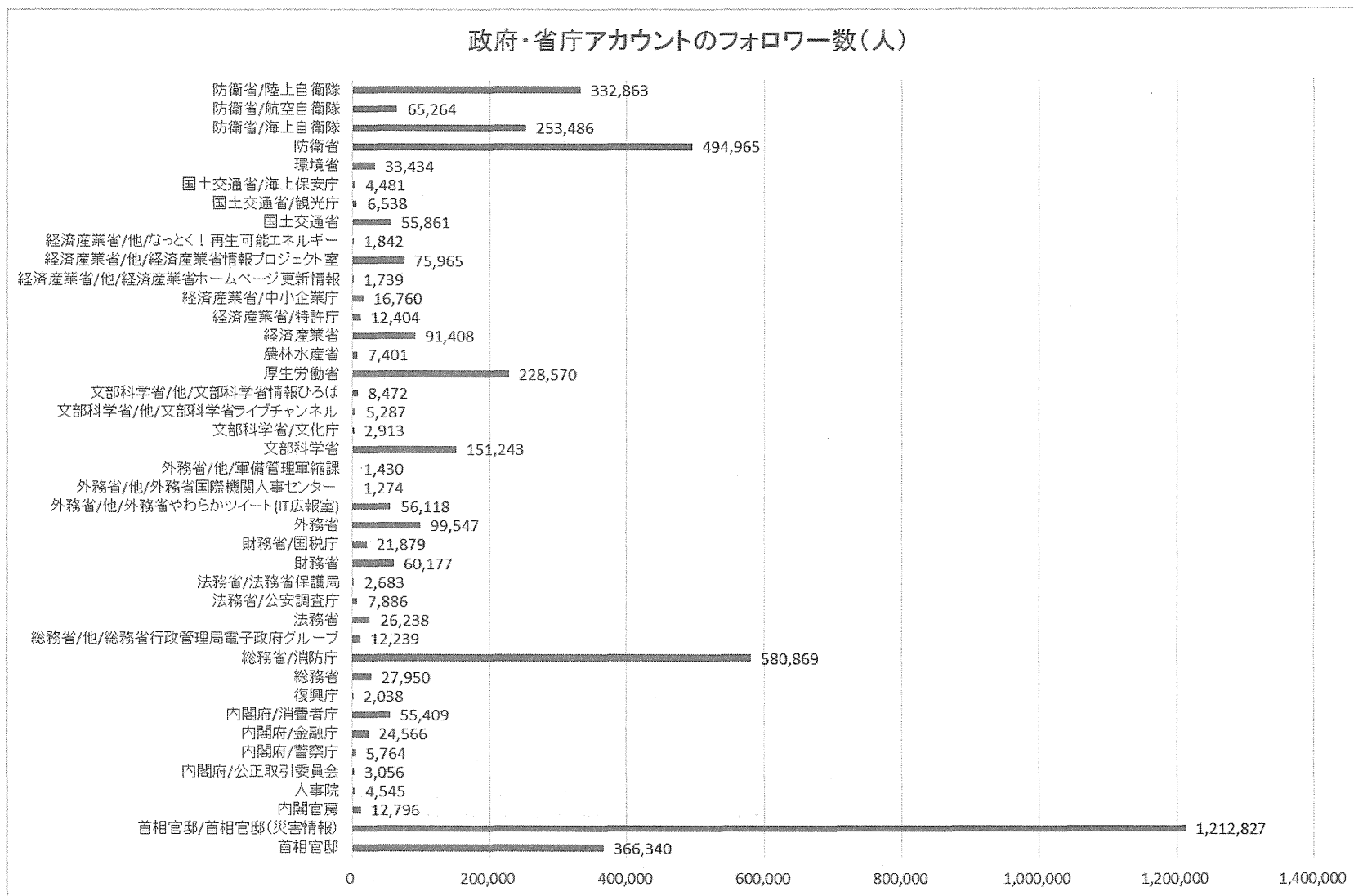


図 7: 政府・省庁の公式アカウントのフォロワー数(人)

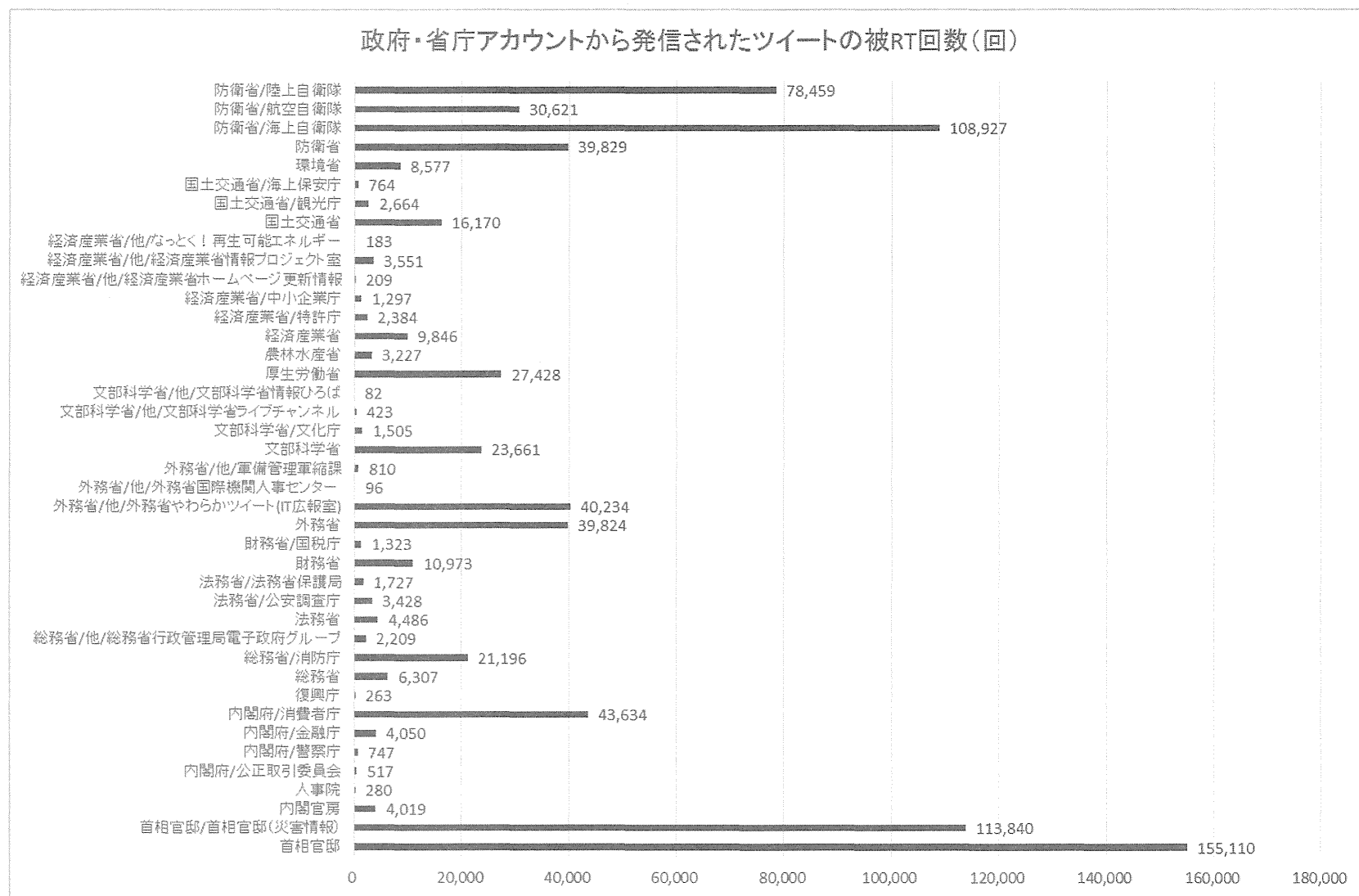


図 8： 政府・省庁の公式アカウントから発信されたツイートの被 RT 回数(回)

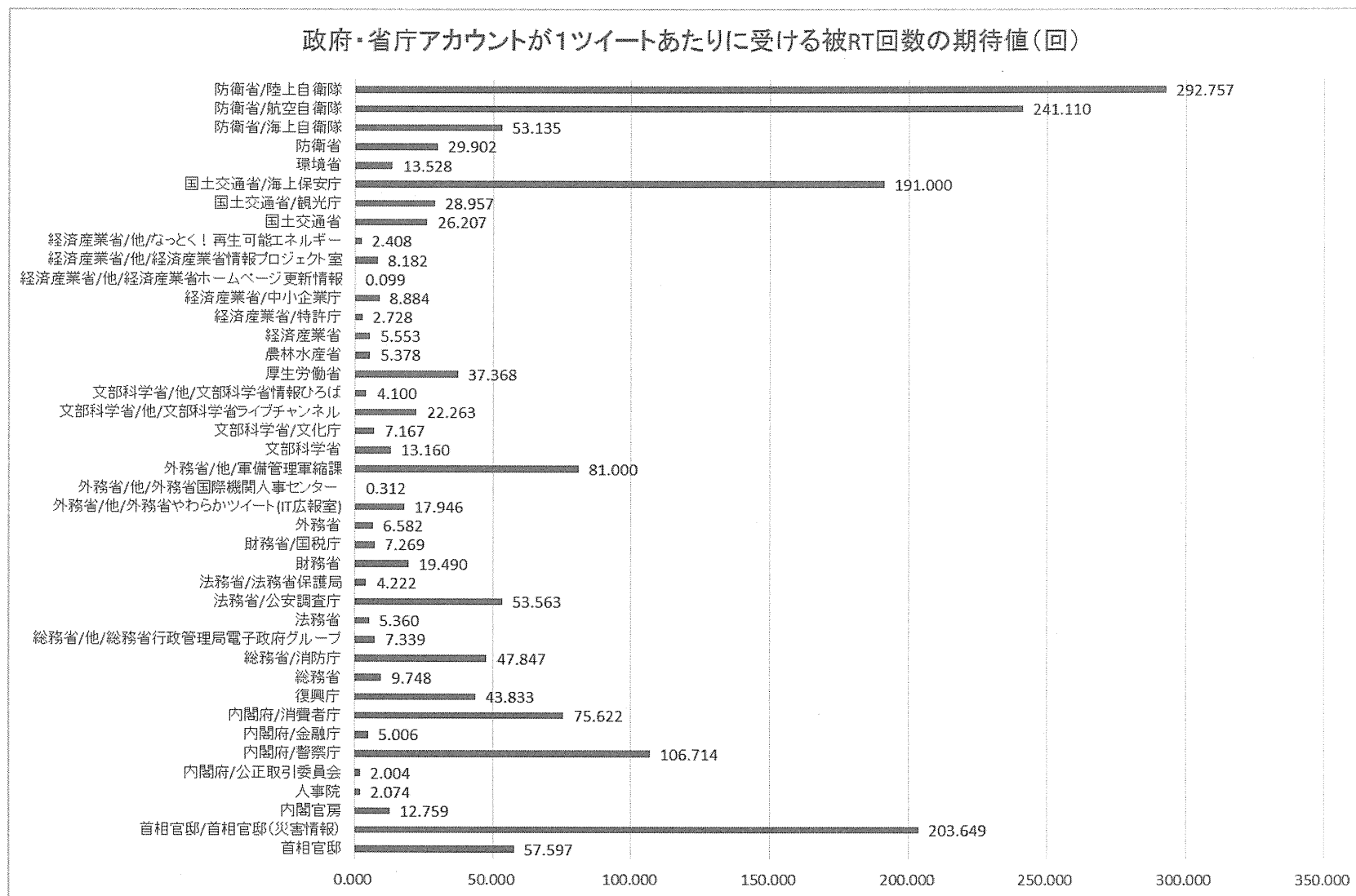


図 9： 政府・省庁の公式アカウントが1ツイートあたりに受ける被RT数の期待値(回)

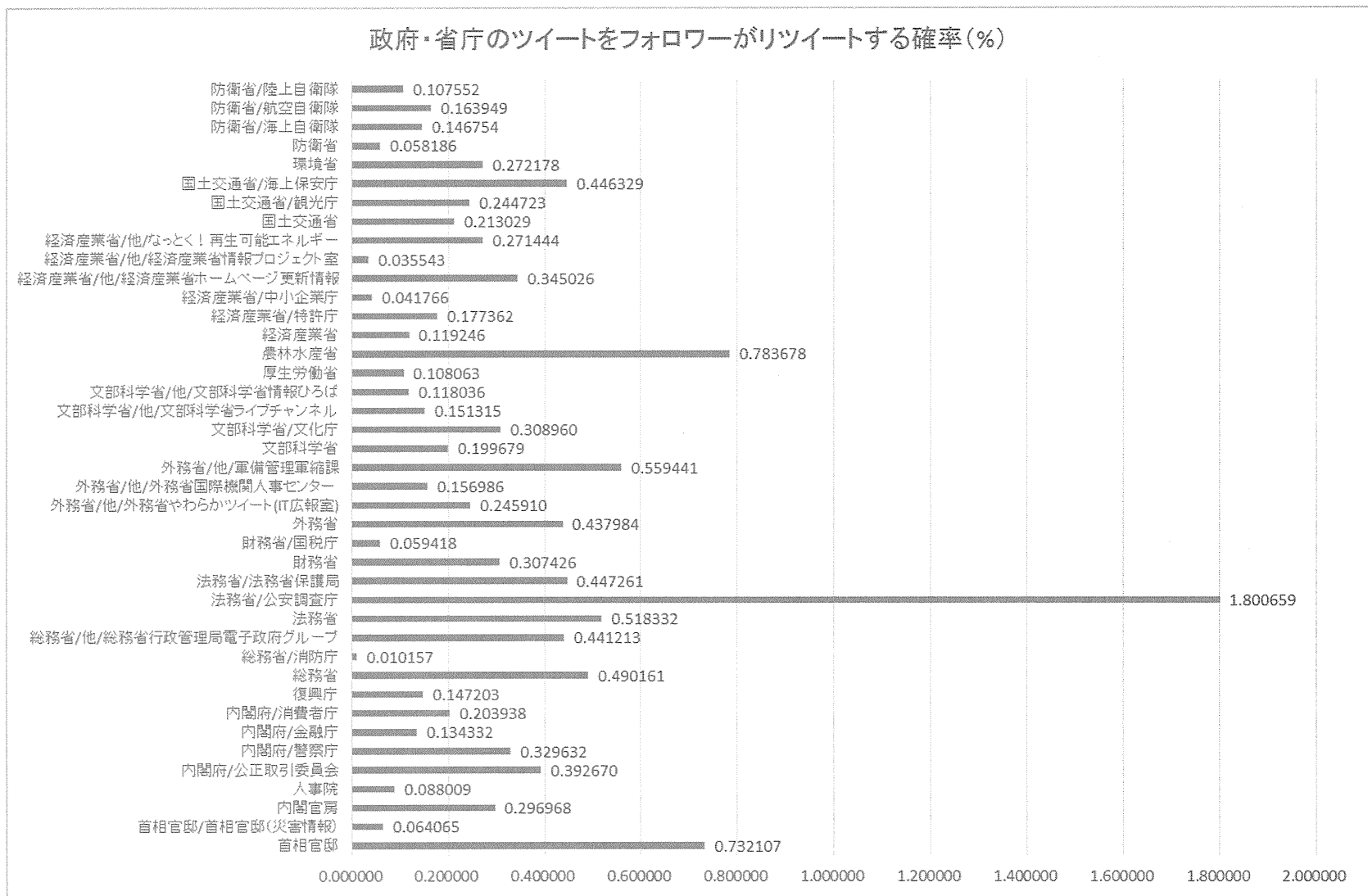


図 10: 政府・省庁の公式アカウントのツイートをフォロワーがRTする確率(%)

【 資料 13 】

厚生労働科学研究費補助金（食品の安全確保推進研究事業）
分担研究報告書

インターネットにおける意見解析手法の信頼性に関する研究

研究分担者 藤井 仁

（国立保健医療科学院研究情報支援研究センター 主任研究官）

研究分担者 奥村 貴史

（国立保健医療科学院研究情報支援研究センター 特命上席主任研究官）

研究要旨

ネット情報の解析による意見分析は、一般的に、定量的には意義を見出すことが困難である。しかしながら、ネットの利用者は増大を続けており、マスコミ情報から口コミ情報まであらゆる情報がネット上に表出されていることから、ネットにおける意見の定性的な分析は、少ないサンプルの社会調査以上に網羅性を有するのではないかと考えられる。そこで本研究分担では、インターネットにおける意見解析手法の信頼性を検証するため、ネット情報の解析により得られる意見と実社会に見られる意見との関係性について検討を試みた。まず、地方都市や高齢者、専業主婦を対象とした対面方式のアンケート調査を実施し、放射能汚染に関する 1,050 件の意見を聴取した。その上で、アンケートにより取得した放射能に関するそれぞれの意見について、ネット上に存在するかの検証を行った。その結果、インタビュー調査により聴取した意見のほとんどに、インターネット上に対応する意見を見出すことが出来た。限られた人数を対象に行った今回の予備調査においては、方法論上の限界はあるものの、ネット情報の解析は社会を知るうえである程度信頼に足る手段であることが示唆された。今後、ネット情報に対する自然言語処理をリスクコミュニケーションの実務へと役立てる実用化研究が望まれる。

A. 研究目的

2011 年 3 月に生じた東京電力福島第一原発事故により、環境中に大量の放射性物質が放出された。こうした前例の無い規模の放射能災害の発生に際して、国や東京電力を初めとする公的機関による情報提供上の問題もあり、発災直後から現在に至るまで、放射能に関する正誤のあいまいな情報や伝聞情報がインターネットを中心として蔓延する事態が生じている。

このように健康への大規模な危険が生

じた際の公的機関による情報提供は、リスクコミュニケーションとして知られ、今までにも、環境汚染による風評被害などに対応するために検討され、実践されてきた。しかしながら、従来のリスクコミュニケーションは、マスメディアを通じた一方通行な情報伝達、ないし、専門家と一般人の双方向な交流のいずれかを前提としており、近年のようにネットを介して、公的機関や専門家、一般人が入り乱れた情報発信や消費をするような状況を想定していない。そのため、既存のリスクコミュニケーション

が成り立たなくなっている可能性があるにも関わらず、ネット上に拡散している情報は膨大であり、統計学に基づく従来の社会調査手法では事態の全体像を把握することすら困難である。

そこで、ネット時代のリスクコミュニケーションを再構築していく上で、多量の自由記載文書を効果的に処理しうる手段である自然言語処理技術が期待されることになる。実際、自然言語処理を応用したネットの解析技術は、企業によるマーケティング目的等で既に商用利用されており、本研究班においても同様の技術を利用したネット上のさまざまな意見の解析を進めている。しかしながら、Twitter や Facebook などに表出されるネット上の意見と、実社会の世論調査結果には、大きな乖離があることは繰り返し指摘されている。ネット上の意見の多くは、意図的な成りすましや意見の水増しだけでなく、意見が自由に複製されうることから、とある意見が多いか少ないかといった定量的な判断に用いることが困難である。また、そもそもネットの利用は PC やスマートフォンなどの情報機器を利用できる年齢層に偏っており、表出されている意見の代表性については大いに疑念がある。さらに、調査により意見を受動的に抽出される社会調査と比べて、ネット上では能動的な意見表明がなされることから、後者はより態度が先鋭化する傾向が懸念される。

そこで、本研究分担は、実社会を対象とした意見調査とインターネット側での調査との比較対照を通じ、インターネットにおける意見解析手法の信頼性についての検討を行うことを目的とする。

B. 研究方法

ネット情報の解析による意見分析は、上述の理由により、定量的には一般的に意義を見出すことが困難である。一方で、放射能汚染に関して現在なされている言論の全体像を知る上で、もはやネットの意見は

無視しえない。また、テレビやラジオなど、マスメディアにおける言説のほとんどは直ちにネット上にコピーされ、それに対する賛否双方からのさまざまな議論が形成されることから、現代では、実社会とネットの言論はほぼ連続していると考えることが出来る。さらに、ネットを利用した調査においては、多少の規模での社会調査を行う場合よりもはるかに多様な意見を効率的に収集することが可能である。したがって、ネットにおける意見の解析は、定性的には、社会調査以上に社会を代表し、また、意見間の関係を明らかにする網羅性を有するのではないかと考えられる。

とりわけ、実社会における意見がすべてネット上に存在するのであれば、ネット上の意見を定性的に分析することにより、社会における放射線に関する意見を効率的に分析することが可能となる。そこで本研究分担では、社会調査によって得られる実社会における意見のほとんどはネット上にも存在することを確認し、ネット調査の定性的な妥当性について検証を試みた。

この「実社会には存在するもののネット上には存在しない意見」の存在を検証するうえでは、ネットを日常的に利用している層ではなく、出来る限りネットへの接触が薄い層を対象とした調査とすることが効果的である。そこで、調査においては、地方都市や高齢者、専業主婦をターゲットとした対面方式のアンケート調査を企画した。そのうえで、アンケートにより収集した放射能に関する意見について、それぞれがネット上に存在するか逐次検証を行った。

意見の抽出に際しては、具体的なトピックを提示し賛成や反対などの態度を聴取する形式の調査が考えられるが、このような構造化された調査は、集計における効率化が図れる一方で、独自性の高い意見を拾い損ねる懸念がある。しかしながら、自由記載を中心とした非構造的な調査では、集計のためのコストが増し、客観性が損なわれる可能性がある。このように、両手法

Q27. 食品の放射能汚染について、お伺いします。

- a) 食品の安全性について、どのような情報を求めていますか。
- b) その情報は、国が提供すべきだと思いますか、思いませんか。どうして
そのように思うかもあわせて、ご意見をお聞かせ下さい。
- c) 国が、どのような情報を、どんな形で提供してくれればいいと思いますか。

Q28. 仮に食品の買い物の際欲しかった品物に「福島産」と書かれていたら、どうしますか。

Q30. 放射能は、あなたの暮らしにどのように影響を与える、または与えていると思いますか。
自由に考えをお聞かせ下さい。

図 1 放射能汚染に関するアンケート

には一長一短があるが、今年度は、予備調査として、学問的な正統性やマーケティング的な妥当性よりは、今後の調査を設計するための大まかな傾向の把握を目的とし、あえて非構造化インタビューにより意見を聴取する形をとった。図 1 に、調査に用いた質問票を示す。

(倫理面への配慮)

本アンケート調査は、「震災後のデマ情報の情報源と、その拡散、終息に寄与する要因の分析」として国立保健医療科学院倫理審査委員会より承認されている調査(NIPH-IBRA#12037)の一部である。

C. 研究結果

本研究は、本研究班の分担研究として設定された震災後に生じた放射能に関するデマ情報についてのアンケート調査結果を解析する形で行われた。アンケート調査は、東日本大震災、並びに、放射能災害の影響が少なかった地域から、大都市部と地方都市の代表として、東京、秋田、高知を選択して実施され、それぞれ 70 件の回答を得た。アンケートに際しては、調査地域において、5 世帯おきに訪問し調査への協力を要請するランダムウォーク法を用いた。

調査対象者は、20 代 4.8%、30 代 12.9%、40 代 16.2%、50 代 18.6%、60 代以上 47.6%、平均年齢は 56.6 歳と、年齢分布がかなり高齢者へと偏っている。また、男性 36.2%、女性 63.8%と、回答者の 2/3 が女性であった。学歴で見ると、中学卒 13.3%、高校・専門学校・短大卒 65.7%、大学・大学院卒 20.5%と、8 割が高等教育を受けていない。インターネットの利用に関しては、毎日利用していると答えたのが 31.4%で、利用しないと答えたのが 46.7%と半数近くを占めた。その他、週 1 回以上が 12.9%、週 1 回未満が 9.0%と、インターネットを日常的には利用していない者の割合が 7 割近くを占めており、研究の目的通りインターネットの利用が少ない層を抽出することが出来た。

本研究では、以上の対象者より、「食品の安全性に関して欲しい情報」、「その情報を国が提供すべきかどうか、また、なぜそう考えるか」、「国に、どのような情報を、どのような形で提供して欲しいか」、「欲しかった品物が福島産であった場合の対応」、「放射能が暮らしに与えている影響」の 5 つの質問について自由な意見を求めた。園上で、集まった 210 件の意見を目視にて類似意見にクラスタ化したうえで、クラスタ中の代表的な意見について、検索エンジン

意見	件数
食品には、安全だとわかるようにしっかりと とした基準値を表示してほしい	52
正確な数字を隠蔽することなく開示してほ しい	40
市場に出回っているものは安全だと思っ ているので、気にしていない	38
産地をはっきりと表示し、安全かどうかわ かるようにしてほしい	26
健康にどのような悪影響が出るのか	14
すべての食品を検査して、すべてに放射能 測定値を表示してほしい	9
騒ぎすぎ、情報に振り回されている	8
よくわからないのでテレビなどでわかりや すく教えて欲しい（具体的にどのような情 報かはわからない）	6
もう歳なので、気にしない	4
グルーピングできない意見	4
分析対象外（意見になっていない／質問の 意図とずれているなど）	8
Q27bの意見の混入	1

表 1. 食品の安全性について、どのような情報を
求めているか（Q27a）

を用いて類似意見の探索を行った。また、
クラスタに分類しえない意見については、
それぞれ同様の方法で類似意見の探索を行
った。最後に、より客観性を高めるため、
別の調査者により意見の類似性の判定を行
った。

5 つの質問に対しする各 210 件の回答を
クラスタ化した結果を表 1～5 に示す。回答
は 10～15 個のクラスタに分類され、5～
10%ほどが分類不能な意見であった。作業
においては、技術的、時間的な制約から、
検索数を削減し、大幅な簡略を行ったが、
検証した範囲においては社会調査による抽
出意見のほとんどに対して、ネット上の類
似意見を見出すことが出来た。より厳密な
判定や他分野での意見により別の結果が得
られる可能性はあるものの、大まかな傾向
としては、ネット調査により抽出される意
見の多様性は、社会調査により抽出される
意見を上回るものと考えられる。検証結果
の詳細は、別途、調査報告書に詳述されて

意見	件数
国に責任、義務がある	65
国が提供するのが安心	43
利害関係なく、公正かつ信頼できる情報を 出せる	23
個人、民間では限界があり、国でないと対 応できない	19
国が信頼できない	6
国の方が基準を統一しやすい	5
国がやるには範囲が広すぎるので、県や地 域が対応するのがよい	5
生産者や販売店が責任を持って提供してほ しい	5
税金の無駄使いになる	3
国が県や企業、生産者を指導して義務付け て欲しい	3
グルーピングできない意見	1
分析対象外（意見になっていない／質問の 意図とずれているなど）	21
Q27aの意見の混入	8
無回答	3

表 2. 放射能の情報を国が提供すべきか？
どうしてそのように思うのか（Q27b）

いる。

D. 考察

今回、限られた人数を対象に行った予備
調査として、類似意見を手作業により検索、
確認する大まかな検証を行った範囲では、
実社会に存在する意見のほとんどは、ネッ
ト上に対応する意見を見出しうる明らかと
なった。歴史的に見ると、インターネット
が研究教育用ネットワークであった 80 年
代から、社会への普及が始まった 90 年代、
ネットに接続可能な携帯電話の普及も合わ
さり社会に浸透した 2000 年代を経て、2010
年代では、地デジ放送やタブレット端末等
の普及により特に意識をしなくても誰もが
ネットに接続するようになりつつある。こ
の傾向は今後も続くと考えられるため、実
社会において発出される意見の多くは、ま
ずまずネット上でも観察される機会が増え
ていくと考えるのが妥当であろう。そうで

意見	件数
メディア、マスコミ、テレビや新聞	68
テレビ	28
自治体、自治体の広報誌	9
食品に表示	7
ネットやマスコミ、ネットやテレビ	6
新聞	5
インターネット	5
店頭	4
あらゆる手段、いろいろな手段	4
NHK	2
分析対象外（意見になっていない等）	26
無回答	46

表 3. 国が、どのような情報を、どのような形で提供すれば良いと思うか（Q27c）

ない意見は、地域性が極めて高いか、一定の集団に閉じた意見、ないし、私的な内容を含むなどの特殊な事例に限られ、絶対数として少ないだけでなく、そもそも社会調査により抽出すること自体が困難な意見である可能性が高い。

ただし、本手法を用いて確定的な結論を導くには、方法論上の問題が少なくない。まず、そもそも類似意見とは何か、という本質的な問題がある。「とある公園にホットスポットがあり心配」というアンケート結果が得られた際、ネットに完全な同一意見が見当たらなくても、「〇〇市にホットスポットがある」といった記述や「ホットスポットが心配」といった記述が認められれば、本研究の趣旨には合致するが、意見の抽象度や類似度の定義には恣意性が伴う。

また、質問を放射能汚染に関する 5 種類の質問に限ったが、別のテーマについて質問することで、得られる結果が変わる可能性は十分ある。さらに、今年度の研究では、コスト的な観点から、類似意見検索の効率化のために意見のクラスタ化を行ったうえで検索操作を行ったが、この操作には客観性と再現性に難がある。方法論的には、同一タスクを複数名で行い、意見の一致度を

意見	件数
安全だという検査情報が表示されていれば買う	32
福島産だろうが、売っている以上、安全なはず	22
福島産かは気にしない	20
地元産や産直を優先して買っている。自家栽培のものを利用	15
なんとなく躊躇する・できれば買いたくない	15
子どもには心配だが大人は大丈夫	12
生鮮食品（水産物、野菜など）は心配なので買わない	11
はじめは躊躇したが、いま（今後は）買う	10
安全情報がはっきりしていないから買わない	9
福島産以外の他県のを優先して買う	6
復興支援というかたちなら購入したい・福島の関係者なので応援したい	5
中国産など輸入品のほうが心配	3
放射能が心配	3
食品への放射能の蓄積が心配	2
分類不能	4
意見とは考えられない内容	41

表 4. 欲しい食品に福島産と書かれていた際、どうするか（Q28）

確認する等の操作が不可欠であろう。この問題は、意見の同一性判定においても当てはまる。客観性の担保のためには、利害関係のない第三者により判定を行うプロトコルとするか、類似意見検索を機械的に行うための処理系を開発するといった工夫が求められるだろう。このように、方法論的には、本手法は予備研究の域を出ない。

しかしながら、ネット調査により収集する意見には十分な網羅性が備わっていると看做すことが出来れば、新たなリスクコミュニケーション手法が実現する可能性が生じる。たとえば、本研究班の他研究分担で研究を進めているネット調査により抽出した意見間の関係を明らかとする「言論マップ」技術によって、放射線に関する誤った情報の根拠や意見間の対立関係等を可視化し、訂正情報の効果的な公表や対立解消に向けた働きかけ等が可能となる可能性があ

意見	件数
特にない・影響ない・よくわからない	40
原発についての賛成意見、反対意見	32
子どもや孫の世代や将来の蓄積が不安だ	25
放射能は目に見えないから怖い。わかりにくい	20
健康への不安がある	17
身近でないので、あまり考えていない。わからない	17
放射能は医療面で役に立っている	10
安全基準をしっかりと決めてほしい。情報を公開してほしい	10
放射能は危険	10
電気料金の値上がりが心配	6
食品や水への影響が心配	6
しかたない	3
産地を気にするようになった	2
放射能の瓦礫問題が不安	2
風評被害が心配	2
震災の影響が大きい。取り返しのつかない問題	2
甲状腺ガンと関係がある	2
単なる事実確認	1
その他	3

表 5. 放射能はあなたの暮らしにどのような影響を与えているか？（Q30）

る。あるいは、誤った情報が次々と生成されているような状況下では、従来のリスクコミュニケーションでは対応できなかった問題に、速報性に優れた意見分析技術が有用となる可能性がある。

E. 結論

本研究分担は、ネット時代のリスクコミュニケーション手法を確立する一助とするため、ネット情報の解析により得られる意見と実社会に見られる意見との関係性について検討を試みた。その結果、方法論上の限界はあるものの、大まかな傾向としてネットの定性的分析は実社会における意見を包含しており、社会を知るうえである程度信頼に足る手段であることを明らかとした。今後の方向性としては、この暫定的な結論における方法論上の課題を克服し仮説のよ

り精緻な検証を目指す方向と、ネットの定性的分析は社会を知るうえで信頼に足るという仮定に立ち、ネットにおいて表明される意見を実際のリスクコミュニケーションに役立てる方向がある。

前者としては、まず、質問のテーマや数、意見聴取を行うサンプル数を増やす必要がある。また、類似意見検索における客観性、再現性を高める手法の開発を行う必要がある。たとえば、既存の検索エンジンを用いて類似意見を検索するには、意見を適切な検索用キーワードに変換したうえで検索する必要がある。また、既存の検索エンジンを用いる限り、そのキーワードに関連した人々の意見ではなく、そのキーワードに関する「有用な情報」ばかりが得られることになる。これらの問題は、共同研究機関である NICT の WISDOM システムを用いることで、より効率化を図れる可能性がある。また、自然言語処理技術を活用することにより、類似意見検索の精緻化や意見類似度の定量的評価に取り組む道も考えられる。

後者としては、共同研究機関である東北大学 乾・岡崎研究室により研究が進められている言論マップ技術のリスクコミュニケーションへの応用が考えられる。具体的には、意見形成の背景にある言論の相互関係を明らかとし、ネット時代のリスクコミュニケーションにおける最適な広報対象と効果的な広報戦略の検討などが考えられる。

ただし、いずれの方向性においても研究には予算を要する。とりわけ、仮説のより詳細な検証のためには、調査対象を増やし、客観性の高い類似意見検索や意見間の類似度の定量化手法の開発が求められる。しかしながら、どれだけコストを掛けて仮説の証明を行ったとしても、研究としての精緻化には貢献するものの、肝心のリスクコミュニケーションの改善には至らない。したがって、限られた研究予算の中での方向性としては、後者の示す自然言語処理のリス

コミュニケーション実務への応用が合理的であろうと考えられる。

F. 研究発表

1. 論文発表

なし

2. 学会発表

なし

【 資料 14 】

厚生労働科学研究費補助金（食品の安全確保推進研究事業）
分担研究報告書

インターネットにおける意見解析手法の妥当性に関する研究

研究分担者 奥村 貴史

（国立保健医療科学院 研究情報支援研究センター 特命上席主任研究官）

研究協力者 榊 剛史

（東京大学 東京大学工学系研究科）

研究要旨

放射能汚染等への対応として、適切な情報提供を通じたリスクコミュニケーションが求められている。しかしながら、ネットが社会へと浸透した結果、公的機関や専門家に加えて一般人が独自に情報発信をする時代が到来し、従来行われてきたプレスリリースや地域集会を通じた情報提供では意図した情報伝達が行えない事態が生じている。それにも関わらず、既存の社会調査手法では、ネット上において生じている情報発信活動を把握することが出来ない。そこで本研究班では、ネット上の様々な情報の解析手法について自然言語処理を活用した検討を進めて来たが、ネット情報の解析から得られる情報については、代表性等の観点から妥当性に関する懸念が呈されることが少なくない。そこで、本研究分担は、社会調査により得られる意見とネット解析により得られる意見の比較を行い、インターネットにおける意見解析手法の妥当性に関して検討を行った。

まず、社会調査により得られる意見として、昨年度に実施したアンケート結果に含まれる放射能汚染に対する意見 1050 件を抽出した。これらに対して、2011 年 3 月～2012 年 12 月に投稿されたインターネット上のブログ記事を対象として、類似意見の検索を行った。その結果、社会調査により得られる意見について、インターネット上のブログにおいても類似する意見を大まかに確認することが出来た。このことは、インターネットにおいて表明されている意見が、社会調査により得られる意見と同等以上の多様性を有することを示しており、社会調査における情報源としてネット情報の定性的な解析が一定の妥当性を有しうること示唆している。今後、類似意見検索技術の発展による、より大規模で精緻な検証が望まれる。

A. 研究目的

2011 年の東京電力福島第一原発事故後より、放射能に関する正誤のあいまいな情報が数多く生じている。「日本の食品中の放射性物質の基準値が欧米のものよりも遥かに高い」、「福島を中心に健康被害が続出している」、「原発作業員に多発している急性白血病が隠蔽されている」等は、代表的

なものであろう。狭い国土において生じた未知の災害に対して社会不安が生じること自体は避けがたいことではあるが、その結果、安全性が確認された農産物等に対しても忌避が進み、多大な社会的損失を引き起こしている。そこで、公的機関に対して、放射能に関する適切な情報提供を通じて無用な社会的損失を避け、被災地の復興を支援していくことが望まれている。

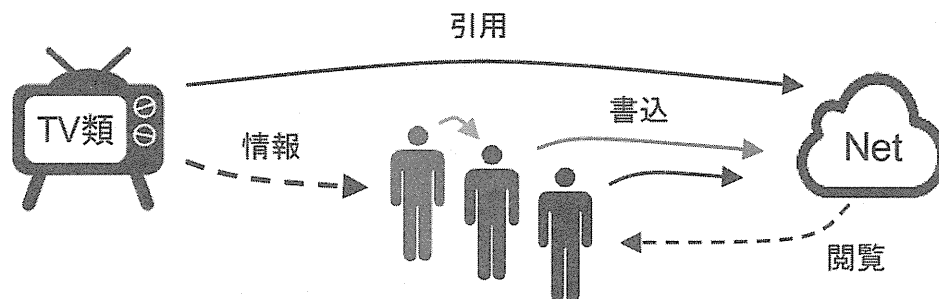


図 1. 意見の表出における既存メディアとネットの関係（模式図）

このような「リスクコミュニケーション」は、今までにも、環境汚染による風評被害などの際に実践されてきた。しかしながら、従来のリスクコミュニケーションは、マスメディアを通じた一方通行な情報伝達か、専門家と一般人の顔の見える交流かのいずれかを前提としており、近年のようにネットを介して公的機関や専門家に加えて一般人までもが独自に情報発信をするような状況を想定していなかった。そのため、今後の公的機関による情報伝達においては、ネット上での反応を考慮せざるをえないが、ネットには膨大な情報が存在するため、社会心理学や統計学にのみに基づく従来の社会調査手法では状況の全体像を把握することすら困難となっている。

こうして、ネット時代のリスクコミュニケーションを検討していく上で、多量の文書を効果的に処理しうる「自然言語処理技術」が期待されることになる。その際、もっとも重要となるのは、ネット上に表出されている意見と、ネットを利用しない者も含めた社会全体の意見との乖離である。実際、ネット上の世論調査と実社会の世論調査に相違が生じることは繰り返し指摘されている。ネット上の意見は、自由に複製が可能であるうえ、PC やスマートフォンなどの情報機器を利用できる年齢層に偏っており、「いいね」や「リツイート」等の消極的な形で示される意見もある。このように多様な意見すべてを正確に把握することは

困難であるため、ネット上での意見について多いか少ないかという定量的な判断を行うことは容易ではない。

そこで、本研究分担は、社会調査により得られる意見とネット解析により得られる意見の比較を行い、インターネットにおける意見解析手法の妥当性についての検討を進めてきた。今年度は、昨年度の社会調査により収集した意見を元にして、より解析の精度向上に向けた研究を試みた。

B. 研究方法

インターネットに表明されている意見は、発信者の素性が特定できず、また、発信者が複数発言したり一つの意見が複製される可能性があることから、有効な定量的分析が難しい。一方で、テレビやラジオなどのマスメディアにおける言説は、現代では、直ちにネット上に複製され、賛否双方からのさまざまな意見が表明されることになる。また、ネットを利用していない者の意見も、伝聞情報としてネットに記載されうる(図 1)。したがって、意見の多様性に関しては、ネット上の意見は、社会調査により収集される意見を凌駕する可能性がある。そのため、低コストに大規模化しうるネットにおける意見の解析は、意見の網羅性や意見間の関係に関する定性的解析に関しては、社会調査以上に社会を代表しうる可能性がある。

Q27. 食品の放射能汚染について、お伺いします。

- a) 食品の安全性について、どのような情報を求めていますか。
- b) その情報は、国が提供すべきだと思いますか、思いませんか。どうして
そのように思うかもあわせて、ご意見をお聞かせ下さい。
- c) 国が、どのような情報を、どんな形で提供してくれればいいと思いますか。

Q28. 仮に食品の買い物の際欲しかった品物に「福島産」と書かれていたら、どうしますか。

Q30. 放射能は、あなたの暮らしにどのように影響を与える、または与えていると思いますか。
自由に考えをお聞かせ下さい。

図 2. 放射能汚染に関する質問一覧

この仮説を検証するため、昨年度は、社会調査により放射能汚染に関する意見を収集したうえで、それぞれの意見がネット上に存在するかを確認し、ネット上の意見が社会を代表する網羅性を備えるかを検証した。具体的には、東日本大震災の影響が少なかった地域から大都市部と地方都市として東京、秋田、高知を選び、それぞれ 70 件のインタビュー調査を行った。インタビューでは、放射能汚染に関するデマ等に関するアンケートに加えて、放射能汚染に関する意見を自由回答して頂いた。社会調査による意見の抽出に際しては、賛成や反対などで回答する単純かつ構造的な調査も可能ではあるが、こうした調査は効率的である一方で、回答の自由度を制限することで独自性の高い意見を収集し損ねる懸念がある。

そこで、昨年度は、今後の調査を設計するための大まかな傾向の把握を目的として、あえてより多彩な意見を収集する自由回答を採用した。上述お通り、3 地点で各 70 名、合計 210 名を対象として、放射能汚染に関する 5 種の質問への回答を収集した。このアンケートに利用した質問票の抜粋を図 2 に示す。これらの質問について自由な意見を求めることで、210 名からの各 5 件、合計 1050 件の意見が集まった。昨年度は、これらの意見を手作業により分類し、それぞれがネット上に存在するか否

かに関して、検索エンジンを用いて手動での類似意見検索を試みた。その結果、大まかな傾向として、ネット上の意見は社会調査により得られる意見を包含するという調査結果を得たが、これはあくまで手作業による調査であり、精度に限界がある。そこで今年度は、昨年収集した意見に対して自然言語処理を用いた類似意見検索の自動化を試み、より精緻な検証を目指した。

ネット上からの類似意見検索に際しては、まず、ネット上の膨大な情報の中から、類似性の高い意見をいかに効率的に抽出するかという課題がある。このスクリーニングに際しては、今年度は、検索対象をブログ記事へと限定するものとする対応した。これにより、広告等のノイズを減らし、ネットユーザーにより表明された意見のみへと検索範囲を限定することが出来る。そのうえで、とある話題についてウェブ上に意見が表明されるのは何らかの事件や事故によりその話題が広く喚起された場合であると仮定し、検索する意見についてウェブ上で活発にブログ記事が掲出された期間を推定し、その期間に投稿された記事のみを対象に意見抽出を行うこととした。これらの手法により、探索効率を高めるとともに、常時特定の話題について投稿しているユーザの偏りを軽減することが出来る。その上で、類似意見の確認作業を行った。

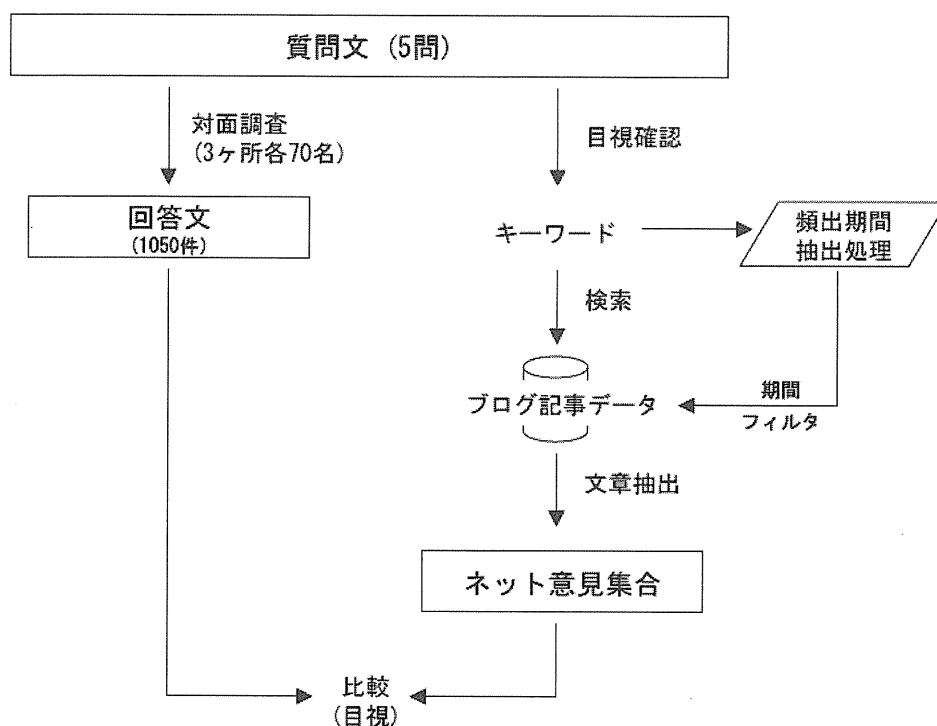


図 3. データ処理概要

(倫理面への配慮)

本アンケート調査は、「震災後のデマ情報の情報源と、その拡散、終息に寄与する要因の分析」として国立保健医療科学院倫理審査委員会より承認されている調査(NIPH-IBRA#12037)の一部である。

C. 研究結果

社会調査側の意見収集に際しては、ネットを日常的に利用している層ではなく、出来る限りネットへの接触が少ない相手を対象とした調査が望ましい。そこで、昨年度の社会調査においては、調査地域の住宅を5世帯おきに訪問したうえで、対面方式のアンケート調査への協力を要請するランダムウォーク法を用いた。合計210件の協力者のうち、年齢構成は20代4.8%、30代

12.9%、40代16.2%、50代18.6%、60代以上47.6%と高齢者が多く、平均年齢は56.6歳となった。また、男性36.2%、女性63.8%と、回答者の約2/3が女性となった。学歴別では、中学卒13.3%、高校・専門学校・短大卒65.7%、大学・大学院卒20.5%と、8割が高等教育を受けていない。インターネットの利用に関しては、「毎日利用している」31.4%、「週1回以上」12.9%、「週1回未満」9.0%、「利用しない」46.7%と、インターネットを日常的には利用していない者の割合が7割近くを占め、計画通りインターネットを日常的に利用しない層を抽出することが出来た。

また、類似意見を検索する対象のネット上の意見としては、放射能汚染に関する話題を抽出するために、2006年以降に日本語圏の3573万6960ユーザにより投稿された30億3462万3833件のブログ記事のなかから、2011年3月～2012年12月に投稿

された記事を対象として検索を行った。抽出に際しては、質問文に関連するキーワードの出現統計を用いて、話題が盛り上がった期間をさらに限定したうえで、当該の話題に関連する意見を抽出した。質問文のキーワード抽出に際しては、たとえば、「仮に食品の買い物の際欲しかった品物に「福島産」と書かれていたら、どうしますか？」という質問に対して「食品」「福島産」というキーワードを抽出し、検索に用いた。さらに、検索結果の各記事に対して、キーワードが出現する前後 150 文字以内の文章を抽出した。以上の処理の全体図を図 3 に示す。

こうした前処理を行ったうえで、「実社会には存在するもののネット上には存在しない意見」の存在を検証するために、アンケートにより収集した放射能に関する意見について、それぞれがネット上に存在するか逐次検証を行った。この社会調査の各回答と類似した意見を検索する作業においては、当初、類似意見検索技術の精度向上を目指して自動処理を試みていた。しかしながら、試みた手法では実用的な性能を出すことが出来なかったことから、最終段階での意見抽出作業については手動で代替した。なお、この抽出に際しては、意見中の各文を独立した意見と見なして処理を行った。たとえば、「自分自身で米作りを行っており、以前に比べて基準値が厳しくなった。風評被害のとばっちりを被っているように感じているので、産地別の米の放射能濃度を食品に表示して公表してほしい。」という回答文があれば、これは 2 つの意見と見なして処理した。

最終的に、各質問文のそれぞれ頻度上位 8 件の回答文全てについて、ブログ記事から類似意見を確認することができた。このことから、昨年度と同様に、ネット上の意見の定性的分析は、社会を知るうえである程度信頼に足る手段であると考えられた。

D. 考察

今回、社会調査によって得られた意見について、ブログ記事より類似意見を抽出することができた。しかしながら、抽出途中の記事には、多くのノイズが混入していることが明らかとなった。とりわけ、ブログには新聞記事の引用が多数含まれている傾向がある。今後、新聞記事に良く含まれる表現や意見に良く含まれる表現をヒントに、これらのノイズを除去する工夫が望まれる。

また、今回は、質問文から抽出したキーワードを用いてスクリーニングを行ったが、ブログ上の意見に該当するキーワードが欠落していた場合、本手法では検知することが出来ない。したがって、社会調査によって得られた意見側から特徴的なキーワード抽出を行い、このキーワードを用いたスクリーニングを行う等の工夫が求められる。

また、今回の手法では、各回答文毎に確認できた類似意見の数や質にばらつきがあった。たとえば、「～欲しい」「～すべき」という要求型の意見もあれば、「～と思う」「～と考える」のように感想型の意見もある。こうした特長に着目し、回答の種類毎に類似意見の抽出に際した戦略を変える工夫も求められるかも知れない。また、前述の通り収集できる意見に制約が生じる可能性があるものの、より検索に適した回答が得られるよう質問文に対して工夫を行う方向性についても検討の余地がある。

最後に、抽出した意見文と回答文の類似性判定に関しては、より大規模で客観的な処理を実現する自然言語処理による自動化が望ましい。一方で、ソーシャルメディアに見られる口語的な文書は表記揺れが大きく、厳密な類似性の判定が容易ではない。今後、機械的な類似性の判別に向け、辞書の整備等により単語や表現の曖昧性を解消する必要があると考えられる。

E. 結論

本研究分担は、社会調査により得られる意見とネット解析により得られる意見の比較を通じて、インターネットにおける意見解析手法の妥当性に関して検討した。その結果、社会調査により得られる意見のそれぞれについて、インターネット上のブログにおいても類似する意見を大まかに確認することが出来た。このことから、インターネットにおいて表明されている意見は、社会調査と比較して、同等以上の多様性を有するものと考えることが出来る。このことは、社会調査における情報源として、ネット情報の定性的な解析が一定の妥当性を有しうることを示唆している。

この妥当性のさらなる検証に際しては、大規模化と客観性の担保に向けて、現在目視判定に頼っている類似意見判定の自動化が望ましい。今後、検証の網羅性、並びに、精度の向上に向け、ノイズの除去、スクリーニング手法の効率化、意見の類型化、曖昧性の除去等、要素技術の発展が求められる。

F. 研究発表

1. 論文発表

なし

2. 学会発表

なし