



マイクロブログ文書の選択による対話的な災害情報 検索システム

北口, 沙也香
宮西, 大樹
関, 和広
上原, 邦昭

(Citation)

神戸大学都市安全研究センター研究報告, 17:185-196

(Issue Date)

2013-03

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/81011422>

(URL)

<https://hdl.handle.net/20.500.14094/81011422>



マイクロブログ文書の選択による 対話的な災害情報検索システム

Interactive Disaster Information Search System for Microblog by Minimal User Feedback

北口 沙也香¹⁾
Sayaka Kitaguchi
宮西 大樹²⁾
Taiki Miyanishi
関 和広³⁾
Kazuhiro Seki
上原 邦昭⁴⁾
Kuniaki Uehara

概要: 災害時, マイクロブログには, 安否情報や支援情報など災害に関わる重要な情報を含む大量のマイクロブログ文書が投稿される. しかし, 災害時には大量のマイクロブログ文書が氾濫するため, 従来の一般的なマイクロブログ検索のインターフェースでは, 自身の意図に適合した情報を見つけるのは困難である. 本稿では, 代表的なマイクロブログサービスの Twitter を対象に, 災害時のマイクロブログ検索に有効なインターフェースとして, 類似する文書をグループ化し, 話題が盛り上がった時期の順に提示する方法と, ユーザが検索結果の文書の中から自身の検索意図に合う文書を選び, 選択された文書を新たなクエリとして再検索する方法の2つを提案し, 東日本大震災時の Twitter データを用いたユーザ実験により, それらの有効性を評価する.

キーワード: マイクロブログ, 災害情報, 検索, ユーザ実験

1. はじめに

東日本大震災発生時, 代表的なマイクロブログサービス Twitter では, 情報伝達を目的として震災に関する大量のマイクロブログ文書が投稿された¹⁾. 投稿された文書の中には, 震災時に起きた出来事を即時に反映したものも多く, 安否情報や被災者への支援情報といった重要な情報が含まれていた. しかし, 震災当時には大量の文書が氾濫したため, 過去と現在の情報や多種多様な情報が入り混じることになった. そのため, 大量の文書の中から自身の意図にあった情報を見つけることは困難を極めた. そこで, 本研究では大量のマイクロブログ文書の中からユーザの検索意図に適合する文書を検索し, その検索結果を(1)意味的に整理すること, (2)時間的に整理すること, (3)ユーザの検索意図を明確にして再検索することに焦点を当て, 上記3つの要素を同時に考慮した Twitter による災害時情報検索システムを提案する.

本システムは, 最初に文書要約手法 Maximal Marginal Relevance (MMR)²⁾ を利用して, 検索結果の文書を類似する文書同士でグループ化し, 情報を意味的に整理する. 次に, 各グループを話題が盛り上がった時期の新しい順に並べることで, 情報を時間的に整理する. さらに, 検索された複数の文書の中からユーザの検索意図に適合する文書を選び, 選択された1つの文書を新たなクエリとして再検索を行う. 最後に, 東日本大震災当時の Twitter データを利用して本システムのユーザ実験を行うことで, 災害時のマイクロブログ検索における, 本手法の有効性を検証する.

2. マイクロブログ

(1) マイクロブログ

マイクロブログとは、ブログの一種であり、一般的なブログよりも短い文章をウェブ上に公開したり、ユーザ同士でコミュニケーションを行うことができる。米国で2006年にサービスが開始されたTwitterの流行を皮切りに、Facebook、Google+、Tumblrなどが公開され、多くのユーザに利用されている。このマイクロブログサービスの急速な普及とともに、マイクロブログを対象とした研究が盛んに行われている。

(2) Twitter

本研究が対象としているTwitterは2012年3月現在に世界で1億4000万人以上¹のユーザが利用する代表的なマイクロブログサービスである。Twitterでは、ユーザがツイートと呼ばれる140字以内の短い文書を投稿したり、他のユーザが投稿した文書を閲覧したりすることができる。自身の投稿はユーザごとに割り当てられたタイムラインに表示される。特定のユーザの文書を継続して閲覧できるフォローと呼ばれる機能があり、他のユーザをフォローすると、フォローされたユーザの文書が自身のタイムラインに表示されるようになる。また、自身をフォローしているユーザをフォロアーと呼ぶ。Twitterでは主にフォロー・フォロアー関係を通じて情報が伝播していく。さらに、他のユーザが投稿した文書を自身の文書に引用して再投稿するリツイートという機能がある。リツイートをを用いることで、リツイートした文書をフォロアーのタイムラインに表示することができるため、フォローしていないユーザの文書を閲覧する機会が生まれる。よって、多くのユーザにリツイートされた文書の情報は急速に拡散する傾向がある³⁾。この他にも、投稿に「#タグ名」を付加して文書が言及する話題を明示するハッシュタグや、投稿の文頭に「@ユーザ名」を付けることで特定のユーザに向けて投稿する返信するなどの機能があり、簡単にユーザ同士で情報を交換することができる。

(3) 災害とマイクロブログ

災害とマイクロブログの関わりについては、様々な分野で研究が行われている。Viewegら⁴⁾は、Twitterを対象に、アメリカで発生した2つの巨大な自然災害の期間のツイートを分析し、Twitterに投稿される災害情報の傾向を報告した。Sakakiら⁵⁾は、機械学習の手法を用いて地震が発生したという内容のツイートを収集し、逐次監視することでリアルタイムに地震の発生を検出する手法を提案している。また、日本国内では東日本大震災の経験を通して、災害時におけるマイクロブログの有用性が広く一般に認知され、震災発生以降さまざまな取り組みがなされている。^{6, 7)}

3. 提案手法

本章では、災害時に投稿が急増したマイクロブログにおいて、ユーザの検索意図に適合する文書を検索するため、検索結果のグループ化と検索結果から選択した文書を新たなクエリとして検索する手法を組み合わせる枠組について説明する。

(1) インタフェース

本研究で提案する災害時のマイクロブログ検索システムのインタフェースについて述べる。図1は提案システムのスクリーンショットである。まず、ユーザは画面上部の検索ボックス(図1の①)に関心のある話題についてのキーワードを入力して検索する。その結果、画面左に検索結果を表す列が出現し、キーワードに関係のある文書が、話題ごとにグループ化された状態で表示される(図1の②)。グループ内に表示されている文書の数は、グループに属する文書の総数を N とすると、 $\log_3 N$ 件とした。よって、グループに属する文書の総数に応じて表示する文書数は対数的に増加する。これにより、表示されている文書数を見ることで、話題の規模を大まかに知ることができる。また、表示中の文書を右クリックするとグループの枠が縦に広がり、そのグループに属するすべての文書を閲覧することができる。

さらに、検索意図に近い文書をダブルクリックで選択すると、その文書のある列の右に再検索を行った結果が追加される(図1の③)。画面には検索結果が3列まで表示でき、3列を超える検索結果は画面の左側に押し出される。しかし、押し出された列は消滅せず、図1の④の「prev」「next」をクリックすると列が右側や

¹<http://blog.twitter.com/2012/03/twitter-turns-six.html>



図 1. 提案システムのスクリーンショット

左側にスライドして過去の検索履歴を見ることができる。よって、ユーザは最後の検索結果からさらに再検索を行ったり、以前の検索結果から検索をやり直すことができる。

(2) マイクロブログ文書のグループ化

Twitterでは、ある話題が盛り上がった時間帯に、その話題に関する用語を含む文書の投稿が増加するという特徴がある⁸⁾。よって、テキストの類似した文書を1つのグループにまとめることで、文書を話題ごとに分けることができる。さらに、各グループに含まれる文書の量から話題の規模を知ることができる。文書が話題ごとにグループになっていることで、グループを代表する少数の文書を読むだけで、そのグループに含まれる多くの文書が検索意図に関係のあるものか否かの区別が出来るため、検索意図に合う文書を素早く発見できると考えられる。また、孤立地域からの発信などの、言及する文書が少なく通常では発見しにくいような小規模な話題も、グループとして他の話題と区別されていることで発見しやすくなるという利点が考えられる。提案手法では、MMRを用いて全文検索結果の文書集合からグループを代表する文書を選択する。MMRは文書選択による要約手法であり、文書をクエリとの関連度と、既に選択された文書との相違度の2要因でスコア付けして、スコアが最大の文書を選択する。スコアは式(1)によって与えられる。

$$\lambda Sim_1(d_i, q) - (1 - \lambda) \max_{d_j \in S} Sim_2(d_i, d_j) \quad (1)$$

ここで、候補の文書 d_i とクエリ q の関連度 Sim_1 には検索エンジンが出力した検索スコアを正規化して用いる。また、 S は既に選ばれた文書 d_j の集合であり、 d_i と d_j の類似度 Sim_2 には文に含まれる形態素のidfを用いたコサイン類似度を使用する。 λ ($0 \leq \lambda \leq 1$) はクエリと文書の適合を優先するためのパラメータであり、 λ が大きいほど各グループ間の多様性が小さくなる。本提案手法では、予備実験を通して $\lambda = 0.8$ とした。また、MMRでは、指定した回数だけ式(1)を用いて文書を選択することができ、選択した文書の数が作成するグループの数となる。グループの数はユーザが任意に指定できる。次に、選択されなかった文書を対象に、グループを代表する文書との Sim_2 を求め、それぞれ Sim_2 が最大となるグループに振り分ける。最後に、グループに含まれる文書が投稿された時刻の平均を、グループが表す話題が最も盛り上がった時刻と仮定し、盛り上がった時刻の新しい順にソートする。冗長な結果を除くため、表示の際はグループ内の全く同じ文の文書は1件のみ表示する。

(3) マイクロブログ文書をクエリとする再検索

本システムでは、検索画面に表示されている文書をユーザが選択すると、その文書から抽出したクエリ語と元のクエリ語を組み合わせることで再検索を行うことができる。災害時は、特定の地域の被害情報や復旧情報などの細かい情報に対する需要が高まると思われる。しかしユーザは一般的な語をクエリとして検索することが多いため、必要でない情報まで検索してしまうという問題がある。例えば、石巻の被害状況を知りたいユーザが「石巻 状況」というクエリで検索した場合、石巻の物資支援状況や復旧状況などの関係のない情報まで提示されてしまう。そこで、ユーザが検索意図に一致する文書を選択し、その文書を元にクエリを自動生成して、選択した文書により近い文書を検索する。これによって、ユーザが自身の検索意図に近い文書に対話的に検索できる。

選択された文書は一般的な語を多く含むため、再検索時のクエリの生成では、選択された文書から話題を特徴付ける語を抽出する必要がある。話題を特徴付ける語を判別する標準的な方法としては、idfを用いて文書コレクション中で登場頻度の低い語を選択するという方法がある⁹⁾。また、文章の話題を特徴付ける語には名詞が適していることが知られている。

これにならい、提案手法における文書からのクエリ語抽出では、まず選択された文書を形態素解析し、固有名詞、一般名詞、およびサ変接続名詞のみを抽出する。形態素解析にはオープンソースの形態素解析器 *kuromoji*³⁾ を使用した。さらに英数字のみで構成される2文字以下の語を除去し、残った形態素のうち、idfの上位 M 件をクエリ語集合 Q_t とする。検索ボックスに入力された元のクエリ語集合を Q_o とすると、再検索では、 Q_t と Q_o それぞれから少なくとも1つのクエリ語を含む文書だけを検索する。このように、元のクエリを選択された文書によって拡張することで、ユーザが新たに検索クエリを入力しなくても、よりユーザの意図に沿った文書を検索できるようになる。

4. 実験計画

(1) 実験目的

災害時のTwitter検索における提案手法の有効性を評価するために、東日本大震災時の実際のTwitterデータを用いた実験を行った。本実験は、下記に示す、グループ化の効果、再検索の効果、組み合わせの効果の3つの点について調べることを目的として行った。

- **グループ化の効果**：文書をグループ化することで、従来の方法に比べ、検索意図に合う文書を効率良く収集できるか
- **再検索の効果**：選択した文書をクエリとして再検索することで、元のクエリよりも検索意図に合致する文書を発見することができるか
- **組み合わせの効果**：グループ化と再検索を組み合わせることによってどのような効果があるか

(2) 実験設定

a) データセット

本実験には、2012年に開催された東日本大震災ビッグデータワークショップ⁷⁾においてTwitter Japan株式会社から提供されたTwitterデータを用いた。提供されたデータは、東北地方太平洋沖地震発生から1週間(2011年3月11日午前9時00分～同年3月18日午前8時59分)に発信された約1億8000万件の日本語のツイートである。Twitterデータは、図2のように、(i) ツイートID、(ii) ユーザID、(iii) 投稿日時、(iv) ツイート本文の4つの欄で構成されている。

(i) 46126534246932480	(ii) 224956629	(iii) 2011-03-11 17:33:17
(iv) とりあえず地震前までのツイート確認 おいらは生きてぞw		

図 2. Twitter Japan 株式会社から提供された Twitter データの例

b) 全文検索

提案システムでは、全文検索エンジン Lucene⁴⁾を利用してマイクロブログ文書の索引付けと検索を行った。検索モデルには、ディリクレ平滑化を適用した言語モデル¹⁰⁾を利用し、各文書に対するクエリ尤度をスコア

³⁾<http://www.atilika.org/>

⁴⁾<http://lucene.apache.org/core/>

とした。ここで平滑化パラメータ μ は予備実験の結果から 2500 に決定した。また、形態素解析には kuromoji を使用した。

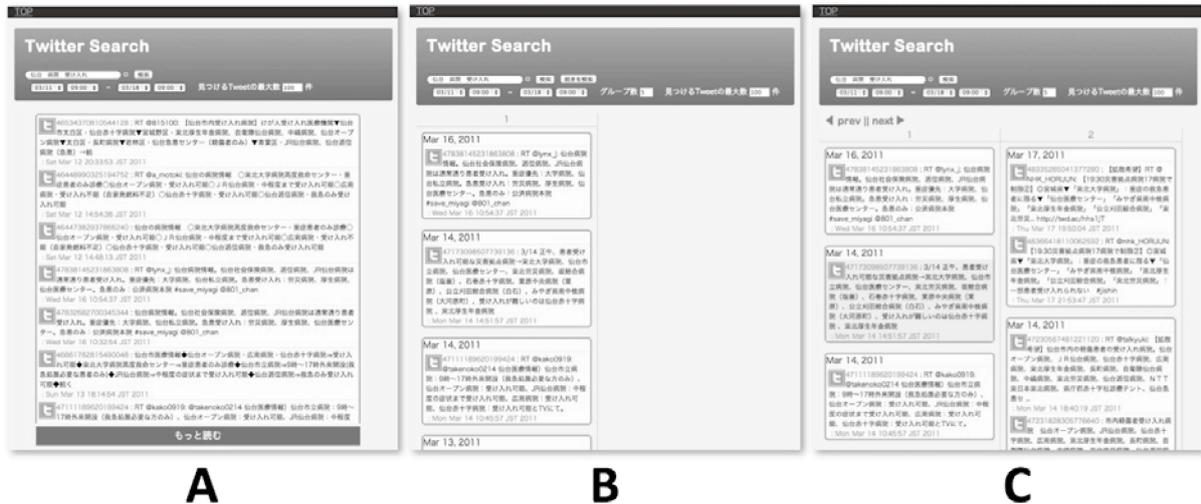


図 3. 実験に使用するシステムのスクリーンショット。A: フラットリスト形式 (一般的な検索インタフェース), B: 検索結果をグループ化した形式, C: グループ化+マイクロブログ文書をクエリとする再検索 (提案手法)。

c) 比較インタフェース

本実験では、図 3 に示す A, B, C の 3 つのインタフェースを比較した。図中の A は Twitter 社の公式検索インタフェースと同様に検索結果をフラットリスト形式で、検索結果の文書をクエリとの適合度が高い順に羅列して表示する。適合度には検索エンジンが算出したスコアを用いた。1 ページあたり 100 件の文書を検索し、全く同じ文の文書は 1 件のみ表示する。2 ページ目以降は「もっと読む」ボタンをクリックすることで取得することが出来る。図中の B は検索結果の文書をグループ化して表示するだけのインタフェースである。B も A と同様に、1 ページあたり 100 件の文書を検索し、全く同じ文の文書は 1 件をのみ表示する。2 ページ目以降は「続きを検索」ボタンをクリックすることで取得することが出来る。A と B を比較することで、グループ化の効果を評価する。図中の C は提案システムであり、検索結果の文書をグループ化して表示し、さらに検索結果の文書を選択すると、選んだ文書を新たなクエリとして再検索することができる。C も 1 回の検索で 100 件の文書を検索する。クエリ生成のため文書から抽出する形態素の数は、予備実験を通して $M = 5$ と決めた。作成するグループの数は 5 つに固定した。グループ数の妥当性はアンケートで考察する。B と C を比較することで再検索の効果を、A と C を比較することで組み合わせの効果を評価する。

d) 検索課題

災害時における検索システムの性能評価を行うために使用する検索課題は、災害当時の情報要求を反映する必要がある。そこで、データセット中の、震災時の情報要求を表す質問文書に着目した。ここで、質問文書を、「おしえてください」「おしえて下さい」「教えてください」「教えて下さい」のいずれかのフレーズを含む、何らかの情報を求める文書と定義する。質問文書の投稿者は Twitter を通じて質問に関する情報が得られると期待しているので、質問文書を Twitter 上での情報要求と見なすことができる。また、2 章 (2) 節で述べたように、Twitter には他のユーザーの文書をそのまま引用し、自身のフォロワーに投稿するリツイートという機能がある。リツイートされた回数はユーザ間での関心の高さを示すものと仮定できる。そこで、データセットから抽出した全質問文書 114,315 件の中から、1 日ごとに {その日にリツイートされた回数} / {全期間にリツイートされた回数} の降順で順位付けを行い、上位から順に

- 内容が震災に関係のある質問をしている
- 同じ内容の質問をしている文書の重複は除去
- 提供データから答えが見つかることが確認できた質問を選択

の基準に従い、より多くの人にリツイートされた質問文書を選んだ。1 日あたりの選択する質問文書の数は、データセットの期間における対象日にリツイートされた質問文書の比率に応じて決定し、合計 36 件の質問文書を抽出した。ここで、提供されたデータだけでは文書同士のリツイート関係を抽出できないこと、本実験では厳密なリツイート回数が必要ではないことから、文頭に「RT」と付く文書をリツイートされた文書とし

た。最後に選択した質問文書の内容を整形して検索意図とし、検索意図に適合する文書を見つけるための初期クエリを質問文書をもとに手動で作成した。最終的な検索課題は初期クエリと検索意図を1組とする。作成した検索課題を付録(1)に示す。

さらに、課題ごとに検索する時間範囲を指定する。課題の元になった質問文書が最もリツイートされた日時の周辺の期間に、その質問文書が表現する情報要求が高まっていたと推測できる。そこで、課題の元になった質問文書が最もリツイートされた日時の24時間後に検索が行われたと仮定し、検索する時間範囲を、データセットの最初の時刻から、課題の元になった質問文書が最もリツイートされた時刻の24時間後までとした。例えば、3月13日19:00に最もリツイートされた質問文書を元に作成された課題の場合、検索期間は3月11日9:00～3月14日19:00までとなる。

e) 被験者

本実験は18人の情報科学の知識を持つ被験者を集めて行った。性別は男性が13人、女性が5人、全被験者の年齢は20歳代で、うち1人が会社員で残り17人は学生である。

f) 実験手順

被験者1人あたり3システムで各2タスク、合計6タスクの検索を行った。使用するシステムおよび検索課題の順番は、ラテン方格を用いて公平に割り当て、システムと検索課題のすべての組み合わせが1度ずつ実施された。被験者には、タスク毎に初期クエリから検索し、その結果から検索意図に一致する文書を5分以内に5件見つけるように指示した。制限時間内であっても検索課題に適合する文書を見つけしだい終了とした。評価実験は以下の手順で行った。

1. 全3つのシステムに共通の注意事項を説明する。
2. 各システムに対して決められた順番でタスクを行う。
 - (a) 使用するシステムの説明を行い、簡単な練習を行なう。
 - (b) 指定した2つのタスクを試行し、タスクごとにタスク完了までの時間と、システム内部で検索に要した時間、発見した文書を記録する。
 - (c) 使用したシステムに関するアンケートを記入する。
3. 最後に、実験全体についてのアンケートを記入する。

5. 実験結果

本節では、3つのシステムに対して、タスク所要時間、検索された文書と課題の適合度、アンケートによる被験者の主観的評価の3つの観点で評価した結果を示す。

(1) タスク所要時間

各システムがどの程度効率良く文書を見つけられたかを評価するため、システムと課題の組ごとのタスク所要時間を計測した。システム間の処理時間の違いを考慮し、タスク所要時間は、被験者がタスクを開始してから完了するまでの時間から、システムが内部で検索に要した時間（被験者が検索要求をしてから結果を得られるまでの時間）を引いた時間と定義する。全測定結果を表1に示す。また、システムごとの平均値を、表2に、箱ひげ図を図4に示す。

タスク所要時間は正規分布に従っていなかったため、ボンフェローニ補正ウィルコクソンの符合付順位検定によって有意水準5%で多重比較を行った。その結果、3システム間のタスク所要時間に有意差は認められなかった。しかし、表2と図4を見ると、システムA、C間に大きな差がないのに対して、システムBはA、Cに比べてタスク所要時間が短い傾向があることがわかる。表1からは、システムBは全検索課題の64%でタスク所要時間が最短となる結果が得られた。よって、システムBで実装したグループ化は、災害時におけるTwitter検索において、ユーザが検索意図に合致する文書を得るまでにかかる時間を短縮する効果が期待できる。

表1. タスク所要時間の全結果.

課題ID	A	B	C	課題ID	A	B	C
1	273	75	178	19	292	222	261
2	104	81	253	20	247	284	201
3	204	131	49	21	286	221	281
4	228	174	76	22	267	269	248
5	60	121	80	23	241	200	258
6	36	142	195	24	279	85	267
7	48	33	248	25	111	100	39
8	292	134	220	26	148	72	118
9	78	36	57	27	33	35	221
10	186	64	102	28	228	253	145
11	252	178	261	29	53	82	94
12	21	56	24	30	41	59	56
13	68	93	88	31	61	133	215
14	37	250	225	32	27	93	123
15	274	119	249	33	128	26	113
16	76	39	219	34	125	65	88
17	258	153	101	35	271	34	209
18	50	137	99	36	163	51	99

表 2. システムごとのタスク所要時間の平均値. タスク所要時間が最も短い項目を太字で記した.

システム	A	B	C
平均 (s)	154.0	119.4	160.0
分散	9,475	5,363	6,299
標準偏差	98.72	74.27	80.49

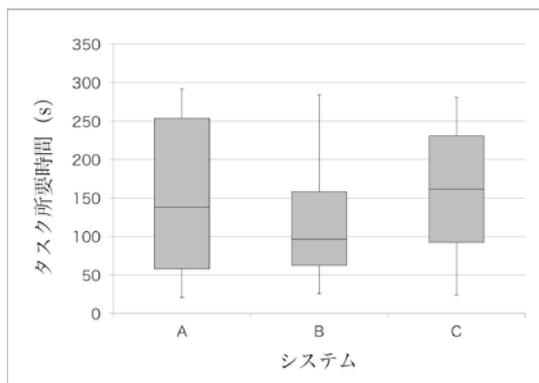


図 4. タスク所要時間の箱ひげ図.

一方で、提案手法を実装したシステム C が、グループ化のみを行うシステム B よりもタスク所要時間が長くなるという結果が得られた。この理由として、システム C で実装した再検索のインタフェース操作の難しさが考えられる。文書を選択して再検索を行う操作は、被験者にとって馴染みがなく、従来の検索インタフェースに類似したシステム A や、グループ化された結果を受動的に受け取るだけのシステム B よりも学習に時間がかかると推測される。この点は、後に述べるアンケートでも指摘される。ここで、システムを使用した順番によって学習効果があったかどうかを調べるために、システムの使用順に所要時間を集計すると、結果は表 3 のようになった。表 3 の第 1 列はシステムの使用順を表し、左のシステムから順に使用した。B-C と C-B は、システム B と C の使用順序のみに注目した結果である。A, B, C の列の各セルはタスク所要時間の平均 (s) で、標本数は A, B, C の列の各セルの値を求めるために用いた標本の数である。また、表 3 の集計結果に対し、使用順序ごとに 1000 回の無作為抽出で有意差 5% のボンフェローニ補正並べ替え検定を行った。表中の † または ‡ の付いた項目同士は並べ替え検定によって有意差が認められた組である。

表 3. システム使用順のタスク所要時間の平均値. 各行でタスク所要時間が最も短い項目を太字で記した.

システム	A	B	C	標本数
B-C	117.8	121.4	143.3	18
C-B	190.3†	117.4††	176.7‡	18
A-B-C	123.5	125.2	104.2	6
A-C-B	173.8	115.7	150.2	6
B-A-C	83.00	111.8	159.8	6
B-C-A	146.8	127.3	166.0	6
C-A-B	178.3	100.7†	211.3†	6
C-B-A	218.8	136.0	168.5	6

表 3 を見ると、有意差が認められる項目はすべて、システム C を使った後にシステム B を使った場合に、システム B のタスク所要時間が他のシステムよりも短くなるということを示している。反対に、システム B を使った後にシステム C を使った場合に、有意差が認められる項目はなかった。この結果から、より学習の容易なシステム B で学習効果の影響が大きくなったと考えられる。

(2) 検索された文書と課題の適合度

a) 適合性判定

36 種類の検索課題に対して、108 回の試行 (18 人の被験者が 1 人 6 課題) を行った結果、合計 462 件の文書が収集された。収集された文書を評価するために、課題と文書の組 1 つに対し 2 人の判定者によって人手で適合度判定を行った。判定の結果を図 5 に示す。適合度は「適合する」「どちらとも言えない」「適合しない」の 3 段階で評価した。判定者は情報科学の知識を持つ学生合計 3 人であり、判定の一致率を示す 2 次の重み付き κ 係数は 0.43 と中程度の一致を示した。判定者 2 人の判定が「適合する」で一致した文書を適合文書、それ以外を不適合文書とすると、適合文書は 281 件 (全体の 61%) 存在した。また、「適合する」を 2 点、「どちらとも言えない」を 1 点として、2 人の判定者の点数を合計した得点を適合度スコアとする。さらに、同じ評価者で、課題ごとの適合文書のうち、文は異なるが内容は同じ文書に、同じラベルをつけた。2 人の評価者が共に同じラベルをつけた文書同士を同じ内容の文書の集合として扱う。131 件の適合文書にラベルが付けられ、合計 36 個の集合が形成された。

b) 検索性能の評価

各システムがどの程度検索意図に合致する文書を検索できたかを、Precision と適合度スコアで評価する。検索精度の評価には一般的に Precision と Recall という指標が用いられる。しかし、本実験では課題 1 つに対して巨大なデータセット中のごく一部の文書の適合度しか評価できないため、データセット中の適合文書の総数が必要な Recall は考えない。

課題ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	合計	
システムA	4	1	0	4	4	0	4	4	4	3	2	4	4	3	3	4	1	4	4	4	4	2	1	4	4	2	3	3	4	4	4	4	3	4	4	4	4	466
	1	1	4	4	3	0	4	4	3	4	4	4	3	4	4	4	4	4	4	4	1	2	4	4	2	4	2	4	4	4	4	3	4	4	4	4	4	
	0			4	4	3	4	4	3	4	4	3	3	4	4	4	4	3		3	4	4	4	4	2	4		4	3	4	4	4	3		4			
	0			1	2	3	3		3	1	4	4	3	4	4	4	4		2		4	4	4	1	2		4	4	4	4	4	4						
関連度スコア	5	2	4	13	16	9	16	12	16	3	15	20	19	15	18	20	1	20	11	0	10	8	15	16	20	9	15	5	20	19	20	20	19	15	8	12		

課題ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	合計	
システムB	4	1	4	4	3	2	0	4	3	4	2	4	4	3	4	4	4	4	4	1	4	4	4	2	2	4	4	4	3	4	4	4	3	4	4	3	3	534
	0	4	4	4	4	3	1	4	4	4	0	4	4	4	3	4	4	4	4	1	4	4	4	3	2	2	3	4	3	4	4	4	4	3	3	4		
	0	1	4	4	3	3	4	4	3	0	2	4	4	3	2	4	4	4	4	1	3	4	4	3	2	2	2	4	4	4	4	4	3	3	4	4		
	0	1	4	0	3	3	4	4	3	1		4	4	3	2	4	4	4	4		4	4	4	2	2		4	3	4	3	4	4	3	4	4			
関連度スコア	4	8	19	16	16	14	13	20	16	10	4	20	20	14	12	20	20	20	0	3	11	20	20	18	9	10	9	20	18	19	19	20	18	15	19			

課題ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	合計
システムC	4	1	1	2	4	4	4	3	4	4	4	4	2	4	4	4	3	4	4	2	1	4	4	0	2	2	4	4	2	4	4	4	4	4	4	4	502
	4	4	4	4	3	4	4	4	4	3	4	4	3	2	4	4	1	4	2	1	1	4	4	2	2	4	4	4	4	4	4	4	4	4	4	3	
	3	4	4	3	2	4		4	4	4	4	3	2	4	4	4	4	2	1	0	4	4	2	3	4	3	4	4	4	4	2	3	4	4	3	4	
	4	4	4	3	3		3	4	4	4	4	2	4	4	4	4	2	1	1	4	1	4	4	4	4	4	4	4	4	4	4	3	4	4	4	4	
関連度スコア	3	1	0	3	3		3	2	4	3		4	4	4	4	4	4	2	1	1	4	2	4	4	2	4	4	4	4	4	4	4	2	3	4		
関連度スコア	18	5	14	14	16	16	12	0	17	17	16	20	18	9	4	16	20	16	16	10	4	0	7	12	16	9	11	20	19	18	20	20	20	15	18	19	

図 5. 被験者が収集した全ての文書に対する適合度判定結果. 各セルはそれぞれ1件の文書に対応し, セル中の値は適合度スコアを表す. 文書が発見されなかった部分は空白になっている.

本実験では, 被験者に5件の文書を見つけるように指示したので, 本実験における試行1回あたり Precision は, 式(1)で定義される. 試行中に被験者が見つけた文書数が5件より少なかった場合, 不足した件数は不適合文書として扱う.

$$\text{Precision}@5 = \frac{\text{被験者が見つけた適合文書の数}}{5} \quad (1)$$

試行1回ごとの Precision を計算し, システム別に集計した結果を表4に示す. また, 各システムで, 被験者によって集められた全ての文書の適合度スコアの合計を表5に示す. 表4と表5の結果に対し, 有意水準5%のボンフェローニ補正ウィルコクソンの符号付順位検定多重比較を行ったところ, 3システム間に有意差は認められなかった. しかし, 表4と表5の両方の結果において, フラットリスト形式のシステムAに比べて, 文書のグループ化を行ったシステムB, Cのほうがより多くの適合文書を発見できることがわかった.

表 4. Precision の平均値 ± 標準偏差.

システム	A	B	C
Precision	0.48 ± 0.37	0.54 ± 0.40	0.53 ± 0.34

表 5. 被験者によって集められた全ての文書の関連度スコアの合計値 ± 標準偏差.

システム	A	B	C
関連度スコア	466 ± 6.26	534 ± 5.81	502 ± 5.81

適合文書の中には, 内容が重複する文書も含まれていたため, 情報の冗長性を考慮するため, 内容の重複した文書を除いた評価も行った. 具体的には, 適合性判定によって作成した同じ内容の文書のグループで, 同じグループに属する文書は1件として Precision を求めた. 例えば, 1回のタスクで同じグループから3件の適合文書が検索された場合, それらの文書はまとめて1件の適合文書として扱った. 結果を表6に示す. 3システム間の Precision に統計的有意差は認められなかったが, システムCがわずかに高いポイントを得ている. よって, 再検索によって, より多様な検索結果を得られていた可能性がある. そこで, 各システムの特徴を調査するために, 検索課題毎に1つのシステムだけで見つかり, 他のシステムでは発見できなかった適合文書を集計したところ, 図6の結果を得た. 3システムを比較すると, システムCのみで取得された適合文書の種類が最も多いことが分かる.

表 6. 重複した文書を除いた Precision の平均値 ± 標準偏差.

システム	A	B	C
Precision	0.39 ± 0.30	0.39 ± 0.30	0.40 ± 0.26

システムCのみで取得された適合文書の例を挙げる. 課題IDが3番の課題(初期クエリ「ディズニーランド 避難」)では, 3システムすべてで「ディズニーランド付近の避難所. 舞浜小学校, 東海大浦安高校が近いです。」のように避難所の場所を述べる文書が見つかった. 加えてシステムCでは, 「舞浜小学校は思うより小さい学校だから, 余裕があれば堀江小学校や見明川小中学へ! RT @simada1101: RT @kumada76: @haayaato ディズニーランドから一番近いのは舞浜小学校です. 駅の反対側へ向かってください. 歩道橋を渡ってセブンイレブン方面へ。」というより細かい情報を発見することが出来ていた. 表6と図6の結果から, 再検索によってより多様な検索結果を得ることが出来ると期待できる.

課題ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	合計	
システムA																																						43
適合文書数	0	0	1	1	0	0	1	3	1	0	0	0	2	0	2	0	0	1	1	0	1	1	2	3	4	0	2	0	4	2	3	3	1	1	2	1		
システムB																																						54
適合文書数	0	0	4	2	0	0	1	5	0	1	0	0	2	0	0	0	5	1	3	0	0	2	4	3	3	1	0	1	4	1	2	2	2	2	0	3		
システムC																																						61
適合文書数	2	0	3	2	0	2	1	0	1	2	1	0	3	0	0	2	5	0	1	0	0	0	1	2	3	0	1	5	4	1	5	3	4	2	3	2		

図 6. 各システムのみで見つかった適合文書の数

(3) 被験者アンケート

実施したアンケートの設問を表 7 に示す。被験者は 5 段階（1 ポイント～5 ポイント）のリッカート尺度で答えるものとし、Q4 以外は 3 ポイントを基準にポイントが大きいほど高評価である。表 7 に加え、本実験に参加した感想を自由に記述できる欄を設けた。Q1～Q3 は 3 システムそれぞれのタスク終了後、Q4 はシステム B のタスク終了後に、Q5 と Q6 および自由記述は実験終了後に回答を求めた。

表 7. アンケートの設問

	質問と選択肢
Q1	このシステムは、目的の情報を見つけるのに有効だと思いませんか？ 5: とても有効 4: 有効 3: 普通 2: 有効でない 1: 少しも有効でない
Q2	目的の情報を見つけるのに読まなければならない tweet の量について、どのように感じましたか？ 5: とても少ない 4: 少ない 3: 普通 2: 多い 1: とても多い
Q3	Twitter 検索をするときは、またこのシステムを使いたいと思いませんか？ 5: とてもそう思う 4: そう思う 3: どちらでもない 2: そう思わない 1: 少しもそう思わない
Q4	グループの数は適切でしたか？ 5: 多すぎる 4: 多い 3: ちょうどよい 2: 少ない 1: 少なすぎる
Q5	グループ化されていることで、目的の tweet は見つけやすかったですか？ 5: とても見つけやすかった 4: 見つけやすかった 3: 普通 2: 見つけにくかった 1: とても見つけにくかった
Q6	選択した tweet から再検索できることで、目的の tweet を見つけるのが容易になったと思いませんか？ 5: とてもそう思う 4: そう思う 3: どちらでもない 2: そう思わない 1: 少しもそう思わない

回答から得られた設問毎のポイントを表 8 に示す。自由記述では 18 人中 15 人の被験者による記入が得られた。自由記述の回答は、付録 (2) に示す。

Q1～Q3 は、3 システムを比較し、マイクロログ文書のグループ化と、マイクロログ文書をクエリにする再検索の有効性を調査する質問である。Q1～Q3 の結果を分散分析にかけたのち、有意水準 5% でボンフェローニ補正両側 t 検定による多重分析をおこなった。その結果、有意差が認められた項目同士は † または ‡ によって表 8 に示した。

表 8 より、Q1～Q3 すべてにおいてシステム B、C がシステム A よりも高い評価を得ていることから、マイクロログ文書のグループ化とマイクロログ文書をクエリにする再検索が被験者に好意的に受け入れていることが分かった。しかし Q2 と Q3 では、システム A-B 間には有意差があるものの、システム A-C 間には有意差が認められないという結果となった。Q2 の結果が得られた理由は、システム C のタスク所要時間がシステム B に比べて長かったことで、読む文書の量も多くなったからだと推測できる。従って、タスク所要時間と同様にインタフェースの操作の難しさが理由と考えられる。実際、アンケートの自由記述でも、システム C の操作の難しさを指摘する意見があった（付録 (2) の再検索に関する意見 2,3）。さらに、Q3 でシステム B よりもシステム C の方が低い評価となったことから、システム C のインタフェース操作が難しかったことが

表 8. アンケート結果. システムごとのポイントの平均値 ± 標準偏差

質問	システム A	システム B	システム C
Q1	3.22 ± 0.92†‡	4.00 ± 0.58†	4.00 ± 0.67†
Q2	2.50 ± 0.90†	3.44 ± 0.83†	3.17 ± 0.90
Q3	3.22 ± 0.92†	3.89 ± 0.57†	3.72 ± 0.73
Q4	-	4.33 ± 0.75	
Q5	-	3.17 ± 0.50	
Q6	-	-	3.89 ± 0.81

推測される。Q4では、マイクロログ文書のグループ化におけるグループ数の適切さを調査した。結果は被験者の半数以上が本実験のグループの数を「ちょうどよい」としていた。この結果より、本実験で作成したグループの数（5個）が有効なグループ数であるということが分かった。最適なグループ数は、タスクの種類や検索する文書の数に依存すると推測されることから、最適なグループ数を考察するためには、タスクの種類、検索する文書の数、および作成するグループの数を変えたより細かい実験が求められる。Q5では、グループ化の有効性を調査した。1人を除くすべての被験者が評価値4（見つけやすかった）または評価値5（とても見つけやすかった）を選択した。Q6では、マイクロログ文書をクエリとする再検索の有効性を調査した。被験者の77%が肯定的意見を選択した。また、半数の被験者は評価値4（そう思う）を選択した。Q5とQ6の結果から、グループ化およびマイクロログ文書をクエリとして再検索する手法は被験者に好意的に受け入れられていることがわかった。よって、グループ化と再検索のインタフェースを実際のマイクロログ検索に適用した場合もユーザに受け入れられる可能性が高いと思われる。

6. おわりに

本論文では、災害時に有効なマイクロログ検索インタフェースとして、マイクロログ文書のグループ化と、マイクロログ文書をクエリとする再検索を提案し、それらの災害時における有効性を検証した。災害時におけるTwitterからの検索を想定したユーザ実験を行なった結果、検索結果の文書をグループ化することで、より短時間で検索意図に合致する文書を見つけられることを示した。また、グループ化によって、より多くの検索意図に適合する文書を発見できる可能性を示した。さらに、文書をクエリにする再検索によって、検索意図に合致し、かつ詳細な内容の文書を検索できる可能性があることがわかった。災害時は、被災地の通信設備が破壊されることなどによって、入手できる情報自体が少なくなることも珍しくない。よって、多様な情報を短時間でより多く得ることができる本インタフェースは、災害時の情報検索に役立てることが出来ると思われる。一方で、再検索インタフェースの操作の難しさが問題となることがわかった。提案手法の有効性を検索システムに活かすためには、再検索操作の改善が必要である。

災害時に役に立つシステムとしては、ユーザが普段からシステムを使い、操作に慣れていることが重要になる。グループ化とマイクロログ文書をクエリにする再検索は、通常時の検索にも応用可能なおえ、実装も複雑ではないため、日常使用する検索インタフェースに応用して、災害時に備えることができると思われる。

付録：

(1) 全検索課題一覧

実験に使用した全ての検索課題を表9に示す。課題IDは検索課題の識別番号である。Time-stampは検索課題の元になったツイートが最もリツイートされた日時である。

(2) アンケート自由記述の回答

グループ化に関する意見

1. グループ化によって目的の文書が簡単に検索出来た
2. 検索のグループ化は明らかに検索意図に合った文書を発見しやすくしていると感じた
3. グループ化によって、似たような文書をいくつも見なくて済んだ
4. グループ内に内容が同じ文書が多く含まれる傾向があるので、同じ内容・同じ情報源の文書を除外したほうがよい
5. グループがどんな話題を表すかのタグや要約があるとよい
6. 検索結果を表示する領域をもっと広くしたほうが見やすい
7. グループがあって使いやすい

再検索に関する意見

1. 再検索ができるので、システムCのほうがシステムBよりも情報が見つけやすかった
2. システムCは便利ではあるが、操作に慣れる必要があった
3. インタフェースは綺麗だが、操作が直感的ではない
4. 再検索は3段階までで十分に目的の文書が見つかる。3段階以上になると探すのを諦めたい
5. 再検索の結果に内容の重複した文書が多い
6. 再検索の結果には前回の結果と重複した文書がないほうがよい
7. システムCは一度に見られる量が多く検索しやすい

実験に関する意見

1. 文書が検索意図に合致するかの判断が難しかった
2. 検索課題によっては検索意図と合致するかの判断が難しい

表 9. 全検索課題一覧

課題 ID	初期クエリ	検索意図	Time-stamp
1	鴨子 状況	鴨子(宮城県)の状況が知りたい。	2011/03/11 15:00
2	プログラフ イムラン ステロイド	プログラフ、イムラン、ステロイドという薬を飲まなかった場合、どうなるのを知りたい。	2011/03/11 18:00
3	デイズニールランド 避難	デイズニールランド周辺の避難所を知りたい。	2011/03/11 20:00
4	吉沼 避難所	茨城県つくば市吉沼周辺の避難所を知りたい。	2011/03/11 20:00
5	仙台 病院 受け入れ	仙台(宮城県)で病人を受け入れてくれる病院を知りたい。	2011/03/11 20:00
6	八戸	八戸(青森県)の被害状況を知りたい。	2011/03/11 15:00
7	横浜 避難所	横浜市(神奈川県)の避難所を知りたい。	2011/03/11 19:00
8	山形 停電	山形で停電があったかどうか知りたい。	2011/03/11 17:00
9	那珂 状況	那珂(茨城県)の状況が知りたい。	2011/03/12 13:00
10	外国人用 医療情報	外国人向けの医療情報が知りたい。	2011/03/12 4:00
11	大崎 給水	大崎市(宮城県)の給水場所を知りたい。	2011/03/12 11:00
12	郡山 炊き出し	郡山(福島県)で炊き出しをしている場所を知りたい。	2011/03/12 20:00
13	被災地以外 支援	被災地以外の地域にいる人が出来る支援方法を知りたい。	2011/03/12 15:00
14	宮城県 栗原市	宮城県栗原市の状況が知りたい。	2011/03/12 13:00
15	宮古市 docomo	宮古市(岩手県)で docomo の携帯電話で、通話またはインターネットが繋がる場所を知りたい。	2011/03/12 10:00
16	栃木県 佐野市 水道	栃木県佐野市で、水道水が濁る事象が起きているか否か知りたい。	2011/03/12 18:00
17	仙台市内 人工透析	仙台市内で人工透析が可能な施設を知りたい。	2011/03/12 8:00
18	山形 携帯 充電	山形県で携帯電話が充電できる場所を知りたい。	2011/03/12 9:00
19	動物 避難	動物を避難させるにはどうすればいいか知りたい。	2011/03/12 14:00
20	宮城 インシュリン	宮城県でインシュリンが手に入る場所を知りたい。	2011/03/13 14:00
21	仙台 オムツ 販売	仙台(宮城県)でオムツを売っている場所を知りたい。	2011/03/13 7:00
22	水戸 街道 国道 50 号	水戸街道から国道 50 号に抜ける近道を知りたい。	2011/03/13 21:00
23	福島 ガソリンスタンド	福島県のカスタムスタンドの営業状況を知りたい。	2011/03/13 11:00
24	孤立	孤立している地域を知りたい。	2011/03/13 14:00
25	被災地 必要なもの	支援のために、被災地で必要なものが知りたい。	2011/03/13 21:00
26	動物 募金	ペットや動物のために募金したい。どこに募金すればいいか知りたい。	2011/03/13 19:00
27	石巻 安否	石巻(岩手県)の安否情報が知りたい。	2011/03/14 0:00
28	女川 子乗浜	宮城県女川町子乗浜の状況が知りたい。	2011/03/14 13:00
29	中部電力 節電	中部電力で節電が必要かどうか知りたい。	2011/03/14 15:00
30	ひたちなか 入浴	ひたちなか(茨城県)で入浴できる施設を知りたい。	2011/03/15 9:00
31	福島 24 時間換気	福島県内では換気扇を切ったほうがよいという話があるが、24 時間換気の家はどうすればいいか知りたい。	2011/03/15 11:00
32	安否情報 サイト	安否情報が載っているサイトを知りたい。	2011/03/15 18:00
33	自治体 アカウント	自治体のアカウントを知りたい。	2011/03/16 13:00
34	勇気の出る歌	勇気の出る歌を知りたい。	2011/03/16 11:00
35	酒田港 タンカー	3 月 16 日午前、酒田港に燃料などを積んだタンカーが到着したというウワサがあるが、事実かどうか分からない。信頼できる情報源を知りたい。	2011/03/17 0:00
36	牧山トンネル	3 月 17 日時点の、牧山トンネルの復旧状況を知りたい。	2011/03/17 14:00

3. 一度に取得できる文書の件数を指定したかった

その他の意見

1. 話題の難しさに検索の便利さが大きく左右される
2. システム A よりも B,C のほうが読む文書の量が少なく、一つひとつの文書の情報量が少ないように感じた
3. リツイートが多い話題は、読む文書の量が多くなる

参考文献

- 1) 東日本大震災におけるツイッターの利用状況について. www.biglobe.co.jp/pressroom/release/2011/04/27-1, (2013/03/24 閲覧).
- 2) Carbonell, J., Goldstein, J.: The use of MMR, diversity-based reranking for reordering documents and reducing summaries, *SIGIR*, pp. 335–336, 1998.
- 3) Kwak, H. and Lee, C. and Park, H., Moon, S.: What is twitter, a social network or a news media?, *WWW*, pp. 591–600, 2010.
- 4) Vieweg, S. and Hughes, A. L. and Starbird, K., Palen, L.: Microblogging during two natural hazards events: what twitter may contribute to situational awareness, *CHI*, pp. 1079–1088, 2010.
- 5) Sakaki, T. and Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors, *WWW*, pp. 851–860, 2010.
- 6) Neubig, G. and Matsubayashi, Y. and Hagiwara, M., Murakami, K.: Safety information mining-what can NLP do in a disaster, *IJCNLP*, pp. 965–973, 2011.
- 7) 東日本大震災ビッグデータワークショップ. <https://sites.google.com/site/prj311/>, (2013/03/24 閲覧).
- 8) 高村大也 and 横野光, 奥村学: Summarizing microblog stream, 人工知能学会第 22 回 SWO 研究会 SIG-SWO-A1001-03, 2010.
- 9) Bendersky, M., Croft, W. B.: Discovering key concepts in verbose queries, *SIGIR*, pp. 491–498, 2008.
- 10) Zhai, C., Lafferty, J.: A study of smoothing methods for language models applied to Ad hoc information retrieval, *SIGIR*, pp. 334–342, 2001.

筆者: 1) 北口 沙也香, 工学部情報知能工学科, 学生; 2) 宮西 大樹, システム情報学研究科, 学生; 3) 関 和広, システム情報学研究科, 講師; 3) 上原 邦昭, システム情報学研究科, 教授

Interactive Disaster Information Search System for Microblog by Minimal User Feedback

Sayaka Kitaguchi
Taiki Miyanishi
Kazuhiro Seki
Kuniaki Uehara

Abstract

In disastrous events, much information is posted on microblogging services. These postings sometimes contain important information, such as safety and support for afflicted people. Due to the overwhelming volume of information, however, it is difficult to find such useful information by a standard microblogging search system. To this end, this paper proposes two types of interfaces for effectively identifying useful information in the event of disasters. First, postings containing similar morphemes are grouped and displayed in the chronological order so that users could easily identify a group of messages relevant to their needs. Second, a message chosen by the users is added to the initial query to refine the search result. A user study is conducted on the Twitter Corpus on the Great East Japan Earthquake so as to show the effectiveness of the proposed interfaces.