

位置依存ソーシャルネットワークにおける行動パターンに基づく群衆の分類

若宮 翔子[†] 李 龍[‡] 角谷 和俊[†]

[†] 兵庫県立大学 〒670-0092 兵庫県姫路市在家本町 1-1-12

[‡] NICT 〒619-0289 京都府相楽郡精華町光台 3-5

E-mail: [†] {ne11n002@st.sumiya@}shse.u-hyogo.ac.jp, [‡] lee.ryong@gmail.com

あらまし 位置依存ソーシャルネットワークの普及により、人々は居場所を中心とした様々な状況をリアルタイムに発信することができるようになってきている。これにより、ユーザらをソーシャルネットワーク上に現実空間の様子を表す「ヒューマン・センサー」として捉えることが可能であり、現実空間の様々な社会的・自然的な現象やイベントを把握するための新たな手段としてその役割が期待されている。しかし、現象やイベントに対する人々の反応の仕方には多様性があるため、イベント検出のためにヒューマン・センサーを有効に活用するためには、イベント発生時に意味のある反応を示すヒューマン・センサーを発見することが必要である。本稿では、位置依存ソーシャルネットワークで共有されている群衆の経験に基づくヒューマン・センサーモデルを提案し、イベント発生時のヒューマン・センサーの反応を分析することにより、素早い状況把握に役立てることができるヒューマン・センサーの発見を目標とする。そのために、反応の仕方に基づきヒューマン・センサーを分類し、有用な反応を示すヒューマン・センサーのグループを発見し、その反応パターンを優先的にモニタリングすることによって、人々の経験を効率的に生かしてイベントを検出するためのアプローチを提案する。

キーワード Twitter, 群衆マイニング, ヒューマン・センサー, イベント検出

1. はじめに

近年、Twitter¹や Facebook²に代表されるソーシャルネットワークサイトを通して、オンライン上でのユーザ間のコミュニケーションを中心としたソーシャルネットワークが普及している。さらに、これらの従来のソーシャルネットワークサイトの発達により、現実空間での人々の居場所を中心とした位置依存ソーシャルネットワークが形成され、注目を集めている。位置依存ソーシャルネットワークが普及した背景には、現実空間の位置情報を容易に取得して発信することができるスマートフォンなどの端末の普及も大きい。

位置依存ソーシャルネットワークでは、実世界での人々の経験をほぼリアルタイムで共有することができるため、実社会の様子を把握するために役立てることができる。実際に、テレビやラジオ、新聞といった従来のマスメディアが報道するよりも先に、位置依存ソーシャルネットワークを通して共有される人々の経験や反応が情報源となり、世界各地で発生した社会イベントや自然現象の様子が伝えられた例が増加している。一例として、2011年3月に発生した東日本大震災の際には、Twitterに発信される1日のツイート数が通常の1.8倍に増加した³と報告されている。そして、位置依

存ソーシャルネットワークを利用して多くの人々の経験を共有することができたため、広域な範囲にわたる各地の状況を素早く把握することが可能であった。このように、位置依存ソーシャルネットワークにおいて実社会での行動や感情といった経験を共有している不特定多数の人々を、現実空間の状況をセンシングして反応する「ヒューマン・センサー」と捉えることができる。

ヒューマン・センサーは、これまで一般的にセンサーと呼ばれてきた機械的なセンサーとは異なり、反応の仕方に多様性がある。まず、機械的なセンサーは、特定の情報をセンシングすると、専用のセンサーであれば、いつでもどのセンサーであっても共通の反応を示す画一的で安定したセンサーである。例えば、住宅用火災警報器はどれでも、煙を感知すれば必ず音や音声により警報を発する。一方、ヒューマン・センサーの場合、個々のセンサーが様々な情報をセンシングすることができ、状況に応じてセンシングした情報を処理することができるため、反応の仕方に多様性が生じる。例えば、雨をセンシングした人々の中には、「雨が降っているから傘を買った」というように「雨」、「降る」や「傘」といった語を用いて明白な反応を示す人々もいれば、「花粉症がましになった」のように、雨について明白に言及していないが雨が降っていることを連想させるような反応を示す人々もいる。もちろん、反応を示さない人々も存在すると考えられる。そこで、位置依存ソーシャルネットワークにおけるヒューマン・センサーを活用して現実空間の状況を把握するた

¹ Twitter: <https://twitter.com/>

² Facebook: <http://www.facebook.com/>

³ NECビッグロブ株式会社 東日本大震災におけるツイッターの利用状況について～新たな情報摂取・共有スタイルの定着～ (2011年4月27日) <http://www.biglobe.co.jp/pressroom/release/2011/04/27-1>

めに、反応の多様性を考慮して、イベント検出に最適なヒューマン・センサーを見つけ出すことが重要である。

本研究では、位置依存ソーシャルネットワークで共有されている人々の経験を用いて実世界でのイベントを検出するために、このような経験を発信するユーザをそれぞれヒューマン・センサーとして活用するアプローチを提案する。そのために、イベントごとに意味のある反応を示しているヒューマン・センサーのグループを発見し、優先的にモニタリングすべき反応パターンを抽出する。本稿では、位置依存ソーシャルネットワークに投稿されている人々の時空間ライフログデータを用いて、災害などのイベント検出に最適なグループを抽出する。

本論文の構成は以下のとおりである。まず、2章では「ヒューマン・センサー」モデルについて説明し、関連研究を紹介する。3章では、現実世界における現象やイベントに対する人々の反応を抽出することで、ヒューマン・センサーをモデリングする手法について述べる。4章ではTwitterから取得した実データを用いて行った予備実験の結果と考察について述べる。最後に5章でまとめと今後の計画を述べる。

2. ヒューマン・センサーを用いた実空間センシング

本章では、現実空間・群衆・サイバー空間から成るモデルについて紹介し、そのモデルを基盤として作成したソーシャルネットワークにおける「ヒューマン・センサー」モデルについて説明する。また、関連研究として、ソーシャルネットワークで共有されている人々の経験を活用した実世界センシングに関する研究やソーシャルネットワークにおけるユーザ分類に関する研究を紹介する。

2.1. 「現実空間・群衆・サイバー空間」モデル

現実空間で生じる様々な社会的・自然的イベントや現象は、人々が示す多様な反応を通してソーシャルネットワーク上に映し出されているといえる。このような考えに基づき、我々はこれまでの研究において、図1に示すような**現実空間・群衆・サイバー空間**の3つの要素から成るモデルを提案してきた。それぞれの要素を以下のように定義する。

- **現実空間**：人々が生活する物理的な空間であり、群衆の行動や感情に影響を与える様々な社会イベントや自然現象などが生じる。
- **群衆**：実世界で経験したことや感じたことなどを自由に表現して公開している人々のことを指す。特に、このような人々の活動は現実空間で発生する様々な社会イベントや自然現象の

影響を受けている。

- **サイバー空間**：情報を検索・取得する際に用いられる一般的なウェブや、不特定多数のユーザのコミュニケーションの場として活用されているソーシャルネットワークサイト、そしてソーシャルネットワークサイトの発展形であり、現実空間の居場所を中心とした情報や経験を共有する場として活用されている位置依存ソーシャルネットワークサイトなどから成る空間である。

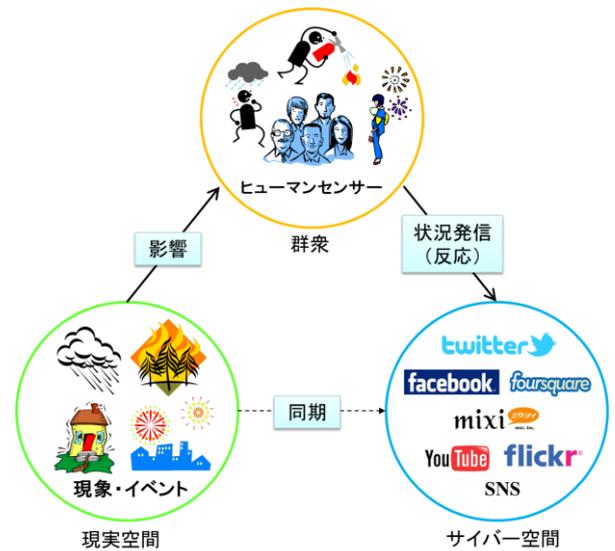


図1. 「現実空間・群衆・サイバー空間」モデル

2.2. ソーシャルネットワークにおけるヒューマン・センサー

本稿では、「現実空間・群衆・サイバー空間」モデルにおけるサイバー空間の中でも、特に群衆を通して実世界と密接につながっている位置依存ソーシャルネットワークに着目し、これを基盤とした「ヒューマン・センサー」モデルを提案する。ソーシャルネットワークにおけるヒューマン・センサーは、現実空間における様々な情報をセンシングして、それを自ら処理することで、活動したり意見を発信したり他のセンサーとコミュニケーションをしたりする自律的なセンサーである。そのため、ヒューマン・センサーの反応の仕方には多様性がある。例えば、雨をセンシングしたときに、雨が降っていることに関する明白な情報を発信するセンサーもあれば、明白な表現を用いていないが雨であることを連想させるような情報を発信するセンサーも存在する。そして、当然のことながら、何の反応も示さないセンサーも存在していると考えられる。このようなヒューマン・センサーの反応の仕方は、時間帯や場所などといったセンサーが置かれている状況に依存し、ダイナミックに変化することが考えられる。

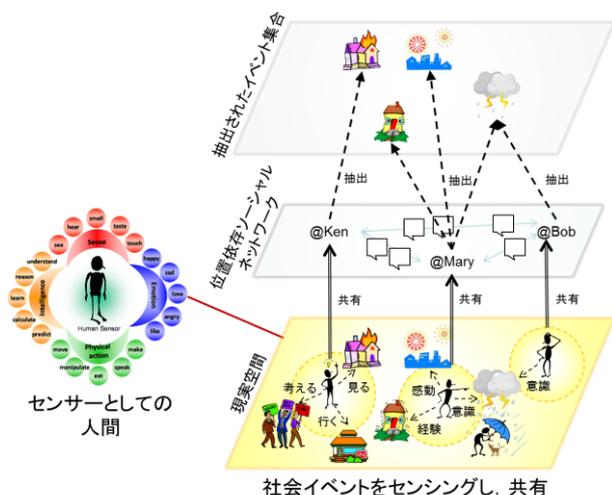


図 2. 位置依存ソーシャルネットワークにおける「ヒューマン・センサー」モデル

例えば、夕方に雨が降っているときには明白な反応を示すが、夜中に雨が降っているときには反応を示さないということも考えられ、センサーとして役に立つ反応を示すセンサーもあるが、役に立たないセンサーもあり、非常に不安定である。

「ヒューマン・センサー」モデルにおいて、個々のヒューマン・センサーの反応の仕方を分析し、イベント検出に有用となるセンサーを選別して利用することが課題となる。このとき、すべてのヒューマン・センサーの反応をモニタリングすることも考えられるが、膨大な数のヒューマン・センサーの反応を常にモニタリングし続けることは現実的でない。実際に、本研究で着目しているソーシャルネットワークにおけるヒューマン・センサーは、日本に限定しても 1 ヶ月で 211,361 人にも及んでおり、この数はさらに増加していくと考えられるため、モニタリングコストが非常に大きくなってしまふ。反対に、明白な反応を示すセンサーのみを対象とした場合には、そのようなセンサーの全体数が少ないという問題が生じると思われる。そこで、本稿では、ヒューマン・センサーを用いてイベントを検出するためにヒューマン・センサーの反応を選別し、イベント検出に最適なセンサー・グループを発見する。

2.3. 関連研究

2.3.1. ソーシャルネットワークを用いた実世界センシングに関する研究

ソーシャルネットワークに関する研究はこれまで数多く行われてきたが、Twitter は近年のソーシャルネットワークを代表するサービスとして世界中の多くの研究者から注目されている。その中で、ソーシャルネ

ットワークへ現実空間でのセンシング情報を発信している人々をヒューマン・センサーとみなし活用している研究として、Sakaki ら [7] は Twitter ユーザが発信した地震に関するキーワードを含むツイートを分析することで、地震を検出するためのシステムを開発した。我々の過去の研究では、Twitter に発信されている群衆の経験を通して地域ごとの定量的な群衆行動を分析することで地域の通常状態を推定し、地域イベントを発見する手法を提案した [3]。また、[8] では、同一の定量的な群衆行動を分析し、地域特徴を抽出するための手法を提案した。これらの手法では、特定の地域の時間帯におけるヒューマン・センサーの反応と地域イベントあるいは地域の役割や機能との関わりを選別せずに活用しているが、本研究では、有用な反応を選別し、イベント検出に効果的なヒューマン・センサーを発見する。

2.3.2. ソーシャルネットワークにおけるユーザーの分類に関する研究

Twitter ユーザの属性に着目してユーザを分類するために、ソーシャルネットワークユーザの特徴を総合的に分析して属性を判定することによって、ユーザを分類する手法として、Rao ら [6] は、Twitter ユーザを性別、年代、出身地、政治的な所属といった観点で分類するために、Twitter ユーザのネットワーク構造特徴、コミュニケーション活動特徴および言語特徴に着目している。そして、予備実験の結果からネットワーク構造特徴とコミュニケーション活動特徴にはユーザ属性に関わる有意な差が観察されなかったため、言語特徴のみを用いた分類手法を提案している。Pennacchiotti ら [5] は、プロフィール特徴、ツイート行動特徴、言語コンテンツ特徴、およびソーシャルネットワーク特徴に着目している。Benevenuto ら [1] は、コンテンツ特徴とユーザのツイート行動特徴に着目し、スパムユーザである可能性が高いユーザの特徴を定義し、スパムユーザとそれ以外のユーザに分類する手法を提案している。Chu ら [2] は、Twitter ユーザを人間、ボット、サイボーグに分類するために、発信間隔や頻度（エントロピー）、スパムパターン（機械学習）、URL の割合や発信端末の割合などのアカウントプロパティに基づき分類する手法を提案している。これらの研究は、属性に着目してユーザの分類を行っているが、本研究では、ヒューマン・センサーという考え方にに基づき、現実空間における現象やイベントに対する反応を用いてユーザを分類することを目的としている点で異なるといえる。

3. ソーシャルネットワークにおけるヒューマン・センサーを用いたイベント検出

位置依存ソーシャルネットワークにおいて共有されている人々の経験を用いてイベントを検出するための手法の概要を図3に示す。まず、各ユーザをヒューマン・センサーにとらえ、その反応を抽出する。次に、現実空間で発生したイベントに対して意味のある反応を示しているヒューマン・センサーのグループを発見し、そのグループの反応パターンを抽出する。そして、抽出した反応パターンを優先的にモニタリングすることによってイベントを検出する。

位置依存ソーシャルネットワークから人々の時空間ライフログを大量に取得するために、過去の研究[3]で開発した地理的マイクロログモニタリングシステムを用いる。次に、各ヒューマン・センサーの反応を把握するために、時空間ライフログに基づくヒューマン・センサーの反応を抽出する。ここで、時空間ライフログのメタデータとして、ユーザID、時間、位置情報、メッセージ（テキスト、ハッシュタグ、リンク）、返信、引用（Twitterでのリツイート）を抽出する。ハッシュタグは、Twitterにおいて特定の話題に関するライフログをまとめるために用いられるタグの一種で、'#'の後に文字列をつなげて'#earthquake'のような形式で記される。リンクは写真や共有動画のような外部メディアを指すURLである。返信は特定のユーザを指定して発信されるメッセージであり、Twitterでのリプライに相当する。引用は、他ユーザの書き込みを参照して発信されるツイートのことであり、TwitterにおいてはリツイートやRTのように表現されている。これらのメタデータを組み合わせることで、個々のヒューマン・センサーのライフログを抽出する。例えば、ユーザID、時間、位置情報を用いることによって移動を、ユーザID、返信や引用によってコミュニケーションを、メッセージに含まれるテキスト、ハッシュタグ、リンクから記述を、リンクを用いることで外部メディアとの関わりを把握することができる。さらに、このような個別のヒューマン・センサーのライフログをまとめると、群衆の活動、意見、人間関係をまとめることが可能である。

個別のヒューマン・センサー hs_i の反応を抽出した後、次の式を用いてイベント e_x の検出に最適なヒューマン・センサーのグループ HS_{e_x} を抽出する。

$$\arg \max_{hs_{e_x}} detect(e_x, HS_{e_x})$$

ここで関数 $\arg \max detect(\cdot)$ は、イベント e_x の検出確率が最大となる反応を示すヒューマン・センサーのグループを発見するための関数とする。

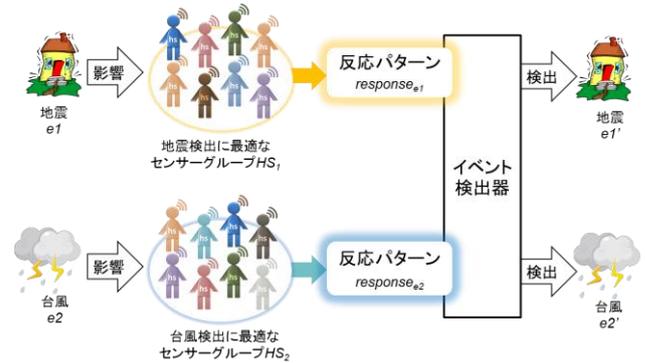


図3. ヒューマン・センサーを用いたイベント検出

4. 予備実験

ソーシャルネットワークにおけるヒューマン・センサーを活用して現実空間におけるイベントを検出するにあたり、まずは最もシンプルに、イベントに対して明白な反応を示しているヒューマン・センサーのみを用いた場合に、どの程度網羅的にイベントを検出することが可能になりそうであるか予備調査を行った。具体的には、時空間ライフログ全体のうちイベントについて直接的に言及しているログの割合と、イベントについて明白に言及しているログと実際に観測されたデータとの関係について、雨を対象として調査した。

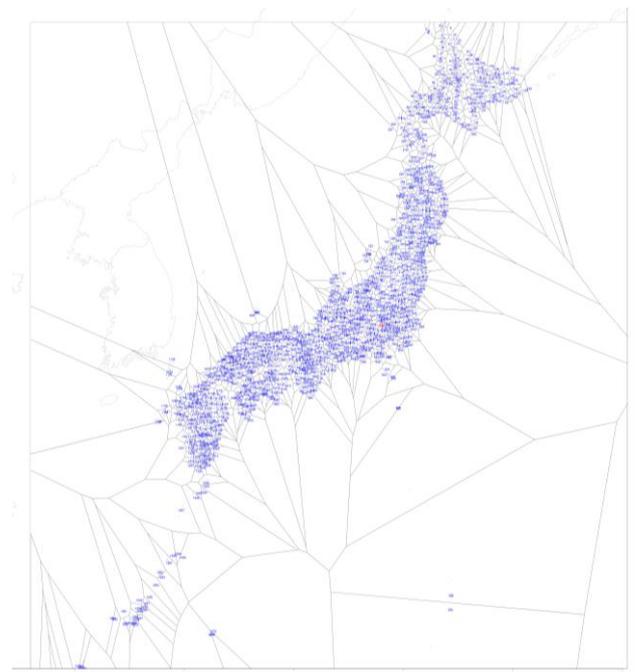


図4. 気象観測所の位置に基づき Voronoi 図を用いて設定した地域境界

4.1. データセット

実験では、実際の降水量データとソーシャルネットワークにおける人々の反応との関係进行分析のために、気象庁が公開している降水量データ⁴と Twitter から取得したジオタグ付きマイクロブログを用いた。このとき、降水量データを参照し、調査期間を 2011 年 11 月 10 日から 2011 年 11 月 13 日までとした。降水量データは、都道府県、観測所、現在地、観測時間、降水量などの情報から構成されており、10 分ごとに更新される。都道府県と観測所はそれぞれ「兵庫県」、「神戸」のようにテキストの地名で記述されているため、両者をまとめてジオコーディングを行うことにより、正確な地理座標を求めた。その結果、ジオタグ付きツイートと同様に、降水量データを時空間データとして扱った。次に、雨に対するヒューマン・センサーを地域ごとにグルーピングするために、気象観測所の位置を基準とした Voronoi 図 [4] を描くことにより、図 4 のように対象地域を 1288 の地域クラスタに分割した。

4.2. 結果と考察

降水量データとヒューマン・センサーの反応を比較した予備実験の結果とその考察を示す。まず、図 5 には、取得した全ツイートと雨に関連するツイート（経験的に「雨」、「降る」、「傘」などのキーワードを含むツイートとする）が発信された地点をそれぞれ青色と赤色の点を用いて地図上にマッピングした結果である。この結果から、雨について明白に言及しているツイートの割合は全体として少ないことがわかる。図 6 は 10 ミリ以上の雨が観測された地点と雨に関連するツイートが発信された地点を地図上にマッピングした結果を示している。この結果からは、雨について明白に言及しているツイートは雨を観測したすべての地点で発生しているわけではないことがわかる。

次に、図 7 には、火曜日の夜から木曜日の夜にかけて雨が降った地域の降水量データと人々の反応の推移を表したグラフを示す。ここで、*Rain* は降水量データであり、*Tweet('all')* は発信された全ツイート数、*T-rain* は「雨」、「降る」、「傘」のように雨に関連するキーワードを含むツイート数、*Users* はユーザ数、*Verbs* は動詞を含むツイート数、*Adjs* は形容詞を含むツイート数である。このグラフにおいて、火曜日の夜の雨が降り始めた時間帯 (p_1) に、雨に関連するキーワードを含むツイートの発生を確認することができた。これは、雨に対する明示的な反応であり、雨が降ると同時に発生していることを実際に確認することができた。

⁴ 気象庁 降水量一覧

http://www.data.jma.go.jp/obd/stats/data/mdrr/pre_rct/alltable/pre1h00.html#a24

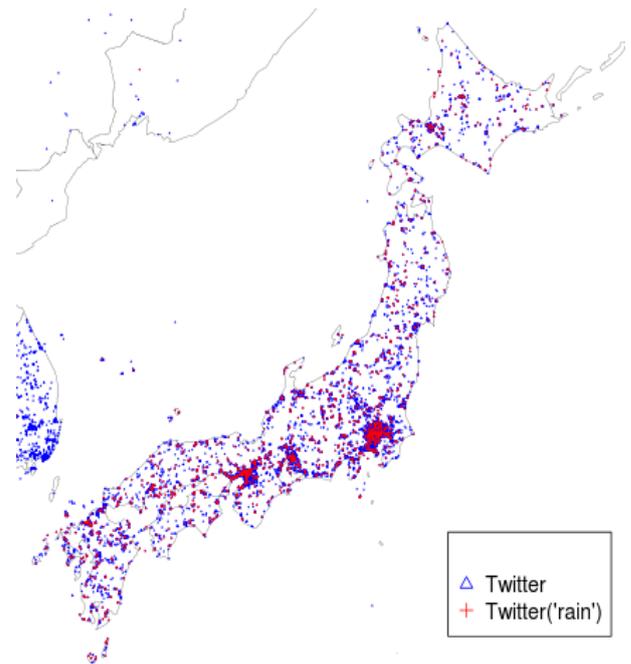


図 5. 取得した全ツイートと雨に関連するツイートの地理的分布

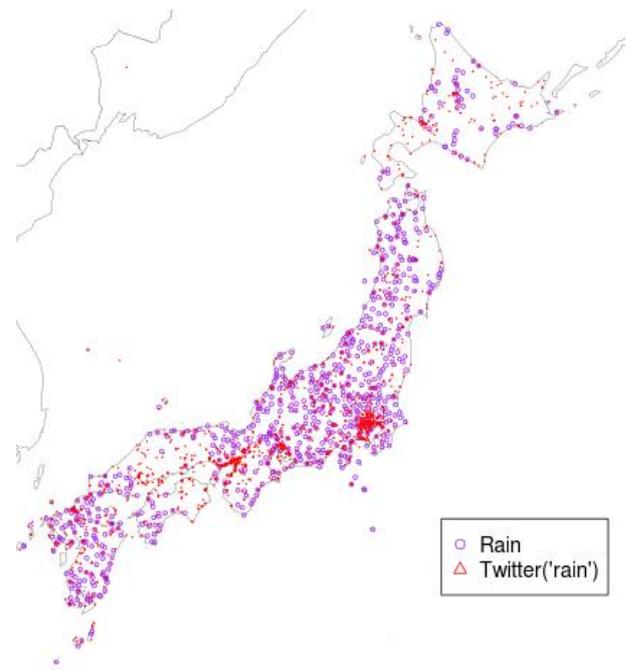


図 6. 10 ミリ以上の降水量が観測された地点と雨に関連するツイート発生地点の分布

一方、 p_1 以降の時間帯では、雨が降っているにも関わらず、明示的な反応がほとんど見られなかった。また、 p_1 においては、雨に対する明示的な反応のほかにも、他の特徴の変化を観測することができた。例えば、ツイートに含まれる動詞の総数やツイート数、および形容詞の数に関しては、 p_1 と同一の時間帯に急激に増加

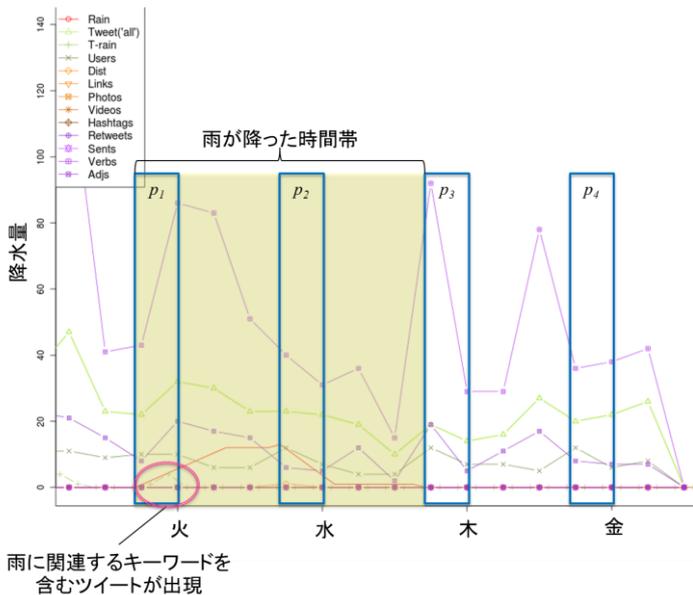


図 7. ある地域における降水量とヒューマン・

していた。このような急激な増加は、他の日の同じ時間帯 (p_2, p_3, p_4) では観測されなかった変化であり、雨の影響を受けている可能性があると考えられる。これらの結果から、実際に、イベントに対する直接的な反応を示しているヒューマン・センサーを用いるだけでは、全体として検出されないイベントが大量に存在すると考えられ、イベントについて間接的に反応しているヒューマン・センサーのグループを選別し、活用することが重要であるといえる。

5. まとめ

本稿では、現実空間の居場所を中心とした位置依存ソーシャルネットワークに発信されている人々の様々な経験や反応を通し、イベントを検出するためのアプローチを提案した。現実空間・群衆・位置依存ソーシャルネットワークに基づく「ヒューマン・センサー」モデルを提案し、予備実験によって実世界の現象やイベントに対する明白な反応だけでなく、間接的あるいは潜在的な反応を考慮することによって、イベント検出に最適なヒューマン・センサーグループを発見することが重要であるという知見を得た。

今後の課題としては、様々な社会的・自然的現象やイベントが発生したときのヒューマン・センサーの潜在的な反応パターンを抽出するための手法について検討する。また、個々のヒューマン・センサーの反応とヒューマン・センサーのグループの反応パターンのモデリング手法について検討する。さらに、イベントが人々に与える影響が地理的にどのように伝搬していくのかを分析することで、現象やイベントを推測するための手法について検討する。

謝辞

本研究の一部は、第8回マイクロソフトリサーチ CORE 連携研究プログラムによるものです。ここに記して謝意を表すものとします。

参考文献

- [1] Fabricio Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virgilio Almeida: Detecting Spammers on Twitter, In Proc. of the Seventh Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS) (July 2010)
- [2] Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia: Who is tweeting on Twitter: human, bot, or cyborg? In Proc. of the 26th Annual Computer Security Applications Conference (ACSAC '10), pp. 21-30.
- [3] R. Lee, S.Wakamiya, and K. Sumiya: Discovery of unusual regional social activities using geo-tagged microblogs. World Wide Web Journal Special Issue on Mobile Services on the Web, 14(4):321-349, March 2011.
- [4] Lloyd, S.P.: Least squares quantization in PCM. IEEE Transactions on Information Theory 28(2), 129-137, 1982.
- [5] M. Pennacchiotti and A. Popescu: Democrats, republicans and starbucks acionados: user classification in twitter, In Proc. of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 430-438, ACM, 2011
- [6] D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta: Classifying latent user attributes in twitter. In Proc. of the 2nd international workshop on Search and mining user-generated contents, pp. 37-44, 2010
- [7] T. Sakaki, M. Okazaki, and Y. Matsuo: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, In Proc. 18th International World Wide Web Conference (WWW2010), April 2010.
- [8] S.Wakamiya, R.Lee, and K.Sumiya: Crowd-based Urban Characterization: Extracting Crowd Behavioral Patterns in Urban Areas from Twitter, In Proc. of 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks, 2011.