

速度データを用いた被災地における活動の回復状態の評価

公共システム研究室 前田 夏輝

1. はじめに

災害時に被災地の状況に応じた適切な支援を行うためには、社会基盤の被害や復旧の状態を把握するのみではなく、人々の生活の実態についても把握することが必要である。災害は、人々の食事、睡眠、通勤、買い物といった様々な活動に影響を与えると考えられるが、どの場所でどのような活動に多くの影響を与えるのかは明らかではない。その理由として、人々の活動は多様であり、それらを時間帯やエリアごと一括して把握することは難しいことが挙げられる。

しかし近年では、情報通信端末から、各エリアの各時間帯における人々の速度データを取得できるようになった。人々がどの程度の速度で移動しているのかが分かれば、活動の内容を推測できると考えられる。そこで本研究では、自然災害が被災地の人々の活動に与える影響と災害後の活動の回復過程を明らかにすることを目的として、人々の移動速度に着目した分析を行う。2016年4月の熊本地震を事例として、まず、速度データにトピックモデルを適用することで特徴的な活動を抽出し、その活動を構成する速度と、エリアと時間帯別の特徴的な活動を推計する。次に、ジェンセン・シャノン情報量を用いて、地震発生前後の活動の分布の差異を算出し、エリアごとの活動の回復の推移を明らかにする。

2. 本研究の基本的な考え方

ポイント型流動人口データでは、詳細な位置ごとに時分秒単位で速度データが取得されている。ここで、人々の活動には、食事・睡眠といった静止している状態や、徒歩や車両での移動のように異なる速度で動いている状態がある。従って、一つの活動は一つの速度で成り立つのではなく、複数の速度から構成されると考える。また、ある時間帯のあるエリアには、同時に複数の活動が混在していると考えられる。そこで本研究では、トピックモデルを用いて、活動にある複数の速度と、エリアに存在する複数の活動を同時に推計する。トピックモデルにより、時間帯別のエリアの活動分布が明らかになれば、地震発生前後における活動分布を比較することにより、地震による人々の活動への影響を明らかにすることができる。

3. 分析手法

3.1 LDA

トピックモデルを推定するにあたり、潜在ディリクレ配分モデル(Latent Dirichlet allocation, LDA)を用いる。LDAでは文書 d のトピック割合 θ_d とトピック k における単語の出現確率 ϕ_k をディリクレ分布によって求める。ここで、ディリクレ分布は、 $\alpha=(\alpha_1, \alpha_2, \dots, \alpha_K)(\alpha_k > 0)$ をパラメータとして式(1)と定義される。また、 $\Gamma()$ はガンマ関数である。

$$p(\theta_d | \alpha) = \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_k^{\alpha_k - 1} \quad (1)$$

同様に、単語の出現確率 ϕ_k は文書 d の単語数 N_d を用いて、 $\beta(k \times N_d)$ をパラメータにもつディリクレ分布から求めることができる。LDAでは、トピック割合 θ_d に応じて文書 d における1つの単語 w_{dn} に対して、1つの潜在トピック z_{dn} が割り当てられる。この潜在トピック z_{dn} とトピック k の単語の出現確率 ϕ_k により、単語 w_{dn} が生成される。パラメータ α, β が与えられたときのトピック混合分布 θ_d の同時分布と N 個のトピック z の集合と N_d 個の単語 w の集合を式(2)に表す。

$$p(\theta_d, z, w | \alpha, \beta) = p(\theta_d | \alpha) \prod_{n=1}^{N_d} p(z_{dn} | \theta_d) p(w_{dn} | z_{dn}, \beta) \quad (2)$$

式(3)は式(2)を θ_d で積分し、 z_{dn} で和をとったもので、これが文書 d の周辺分布となる。

$$p(w | \alpha, \beta) = \int p(\theta_d | \alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn} | \theta_d) p(w_{dn} | z_{dn}, \beta) \right) d\theta_d \quad (3)$$

式(3)においては、変分ベイズ推定を用いて近似計算を行う。

3.2 ジェンセン・シャノン情報量

被災後の人々の活動パターンの回復過程を明らかにするために、地震発生前と後のトピック分布間の距離を求める。

ここで、地震発生前のトピック分布の平均を基準となる確率分布 Q_j とする。

$$Q_j = \left\{ \frac{\sum_i^t \theta_{id1}}{t}, \dots, \frac{\sum_i^t \theta_{idn}}{t} \right\}, (t=14) \quad (4)$$

また、比較を行う15日以降のトピック分布 Q_l は(5)式のように表される。

$$Q_l = \{ \theta_{id1}, \theta_{id2}, \dots, \theta_{id(n-1)}, \theta_{idn} \} \quad (5)$$

(15 ≤ i ≤ 30)

式(4), (5)を用いて、ジェンセン・シャノン情報量によってトピック分布間の距離を求める。

$$D_{JS}(Q_j \| Q_l) = \frac{1}{2} D_{KL}(Q_j \| Q) + \frac{1}{2} D_{KL}(Q_l \| Q) \quad (6)$$

$$Q = \frac{1}{2}(Q_j + Q_l)$$

式(6)で求めたトピック分布間の距離より、回復の推移を明らかにする。

4. 実証分析

速度データにトピックモデルを適用するにあたり、一定の日時に1つの自治体から取得されたデータの集合を1つの文書、単語を速度の3区分(0km/h, 0km/h<v≤6km/h, 6km/h<v)とした。時間帯は24時間を4区分(0~5時, 6~11時, 12~17時, 18~23時)とした。最適トピック数は、Griffithsの指標を用いて6とした。図1に6つのトピックの速度分布(単語分布)の推計結果を示す。左上からトピック1, 右下がトピック6である。本研究では、速度分布の構成から各トピックがどのような活動であるかの考察を行う。その結果、トピック1を「徒歩と自転車」、2を「電車とバス」、3を「自動車専用道路を走行する車」、4を「食事やオフィスワーク」、5を「睡眠」、6を「一般道を走行する車」と考察した。この6つのトピックの変化から災害が活動へ与える影響を考察する。

図2は、益城町の4月1日から4月30日の12時から17時におけるトピックの推移である。益城町は最も大きな被害を受けた自治体であり、前震の翌日15日にトピック1の割合が大きく増加し、30日時点でも割合が増加したまま推移している。トピック1は、徒歩ぐらいの速度が最も多く存在しているトピックで、地震による徒歩での避難行動や地震発生後の被害による生活の変化によって増加した可能性がある。

図3には、トピック分布間の距離の推移を示す。0時から5時では全体として距離が小さいが、6時から11時と12時から17時の時間帯での距離が大きくなっていることから午前中から夕方にかけての活動パターンへの影響が大きかったのではないかと考える。特に12時から17時は、

30日時点でも再び大きくなっている。また、18時から23時の終盤でも距離が大きくなっていることから被災による活動パターンへの影響が出ている可能性がある。

以上より、本震の震源地である益城町では、昼から夕方の時間帯にかけて徒歩の速度が最も存在しているトピック1が大きく増加したことから、地震による避難行動や災害対応に追われていると考えた。また、トピック分布間の距離からは、30日時点でも地震による活動への影響が残っている可能性を示した。

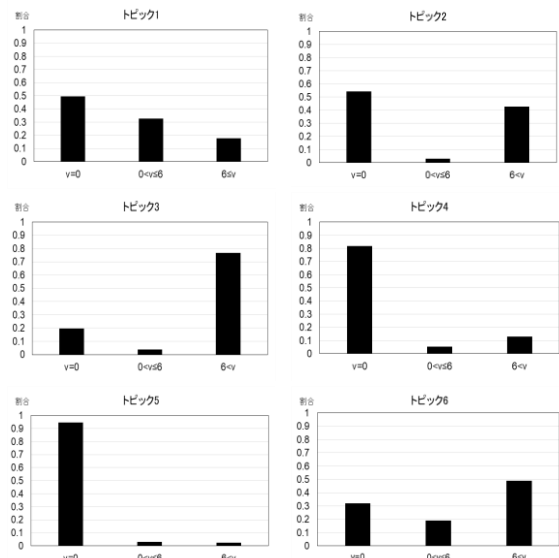


図1 速度分布 (トピック数6)

益城町「昼～夕方」(12-17時)

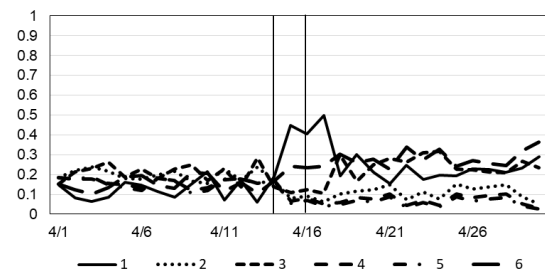


図2 「昼～夕方」のトピック推移 (益城町)

トピック分布距離「益城町」

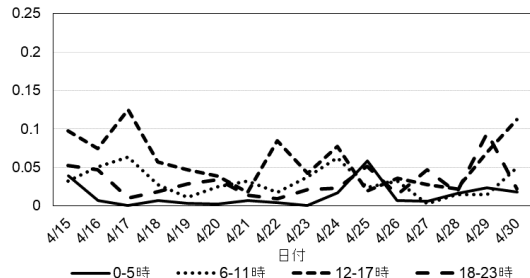


図3 トピック分布間の距離 (益城町)