

Twitter 上の情報拡散系列からの変化点検出

Change Points Detection from Information Diffusion Sequences of Twitter

大原剛三*1
Kouzou Ohara斉藤和巳*2
Kazumi Saito木村昌弘*3
Masahiro Kimura元田浩*4
Hiroshi Motoda*1 青山学院大学
Aoyama Gakuin University*2 静岡県立大学
University of Shizuoka*3 龍谷大学
Ryukoku University*4 大阪大学
Osaka University

We propose a method of detecting the changes of time delay intervals in an information diffusion sequence observed over a social network. We assume a generic information diffusion model in which time delay associated with the diffusion follows the exponential distribution and the burst of information diffusion is directly reflected to the changes in the time delay parameter of the distribution. The proposed method detects the change points and finds the values of the parameter by maximizing the likelihood of generating the observed diffusion sequence without specifying the number of underlying change points in advance by incorporating the model selection method based on the likelihood ratio test. We evaluate our proposed method with synthetic data, and show it is more accurate than a simple greedy method. We then report the results obtained by applying it to the real Twitter data of the 2011 To-hoku earthquake and tsunami.

1. はじめに

近年、高性能なスマートフォンなどの普及により我々のコミュニケーションスタイルは劇的に変化しつつあり、その中でも Twitter や Facebook などのソーシャルメディアの利用が日常生活に大きな影響を与えつつある。これらのソーシャルメディアでは、友人関係などに基づいて形成されるソーシャルネットワークを介して情報が伝搬する。特に、Twitter はその情報発信の手軽さから若者を中心に爆発的に普及しており、先の東日本大震災を含む国内外の自然災害発生時においては情報インフラとして重要な役割を果たしたことは記憶に新しい。

一方、実際の情報拡散においては、ある特定の情報が短期間に爆発的に拡散し得る。そのようなバースト的な情報拡散が何を契機として生じ、どの程度持続したのかを知ることは、実世界において観測された事象を理解・分析する上で重要であり、そのためには観測データからバースト期間を精度よく効率的に特定できる必要がある。筆者らは、そのようなバースト的な情報拡散現象が情報拡散モデルのもつパラメータの変化により生じたものと捉え、その変化点を観測データから検出する手法をこれまでに提案している [大原 12]。時間遅れを考慮したモデル [Saito 09, Saito 10] では、その情報伝搬の遅れは指数分布などによりモデル化されるため、提案法では、その分布を規定するパラメータ（時間遅れパラメータ）にのみ着目することにより、特定のモデルを仮定することなく変化点を検出することを可能としている。具体的には、観測した情報拡散系列に対する尤度を最大化するような時系列上の複数の変化点と、それらにより生じる時間遅れパラメータの系列を推定する。

一方、従来法では、検出する変化点数を事前に指定する必要があった。しかしながら、情報拡散系列に内在する変化点を事前に特定することは実際には困難であるため、本稿では、新たに尤度比検定に基づいたモデル選択法を従来法と組み合わせることにより、事前に個数を指定することなく、観測データから最適な変化点を検出する手法を提案する。そして、人工データ

を用いた評価実験を通して提案法が貪欲探索に基づくシンプル法と比較して高い精度で変化点を検出できることを示すとともに、東日本大震災前後に Twitter 上で発信された実際のツイートから抽出した時系列データに提案手法を適用し、実データに対しても高精度に変化点を検出できることを実験的に示す。

2. 情報拡散モデルの枠組み

本研究では、有向グラフ $G = (V, E)$ として表されるソーシャルネットワーク上での情報拡散を考える。ここで、 V と E ($\subset V \times V$) はそれぞれ全ノードの集合と全リンクの集合を表す。いま、時刻 t_0 にノード v_0 から発信された情報により生じた情報拡散系列 $C = \{(v_0, t_0), (v_1, t_1), \dots, (v_N, t_N)\}$ を観測したとする。ここで、 v_n は情報が伝播したノード、 t_n はその時刻とする。また、任意の $n \in \{1, \dots, N\}$ に対して、 $t_{n-1} < t_n$ となるよう番号付けされていると仮定する。一方、標準的な設定として、情報拡散系列 C は有向グラフ G が内包するツリーとして規定されるとする [Sadikov 11]。すなわち、ノード v_n に情報を伝播した親ノードは $v_{p(n)}$ として一意に確定されると考える。ここで、 $p(n)$ はノード v_n の親ノード番号を $\{0, \dots, n-1\}$ の範囲で返す関数を表す。

情報拡散系列 C を生成する基本モデルとしては、実世界での情報伝搬に見られる非同期時間遅れの概念を導入したモデル [Saito 09, Saito 10] を仮定する。すなわち、ノード v_n に情報が伝播した時刻 t_n として任意の実数を許容し、情報伝播の遅延時間 $t_n - t_{p(n)}$ については、ある確率分布に従うものとする。なお、本稿では議論の簡単化のために遅延時間は指数分布に従うものとするが、べき乗分布など他の分布も同様に用いることが可能である。

3. 問題設定

本節では、本稿が取り扱う変化点検出問題を [大原 12] に従い形式的に定義する。ここでは、情報拡散過程で何らかの変化が生じ、その変化が内包されたある拡散系列を観察するものとする。このとき、変化点検出問題では、各変化点において、その変化がいつ生じ、どのくらいの長さ持続したのかを検出することを目的

連絡先: 大原剛三, 青山学院大学理工学部情報テクノロジー学科, 〒252-5258 相模原市中央区淵野辺 5-10-1, ohara@it.aoyama.ac.jp

とする。なお、同じトピックを話題にする人々は類似した行動をとる傾向にあるという先行研究 [Saito 09, Saito 10] の知見に基づき、本研究では、リンク $(u, v) \in E$ ごとに設定可能な時間遅れパラメータ $r_{u,v}$ について、リンク依存せず一定の値を取るという制約をおく。すなわち、 $r_{u,v} = r$ ($\forall (u, v) \in E$) とし、情報拡散の遅れは指数分布 $p(t_n - t_{p(n)}; r) = r \exp(-r(t_n - t_{p(n)}))$ に従うものとする。

次に、変化点検出問題を数学的に定式化する。観測された情報拡散系列における観測時刻集合を $\mathcal{D} = \{t_0, t_1, \dots, t_N\}$ 、この拡散過程における変化点を T_j とする。ここで、 $t_0 < T_j < t_N$ である。そして、この系列における情報拡散遅れは変化点 T_j の直前までは r_j 、直後では r_{j+1} というパラメータの指数分布に従うものとする。このとき、 J 個の時刻から構成される変化点集合を $\mathcal{S}_J = \{T_1, \dots, T_J\}$ とし、便宜上 $T_0 = t_0$ 、 $T_{J+1} = t_N$ と設定する。また、 $T_{j-1} < T_j$ とし、 \mathcal{S}_J による \mathcal{D} の分割を $\mathcal{D}_j = \{t_n; T_{j-1} < t_n \leq T_j\}$ と定義する。すなわち、 $\mathcal{D} = \{t_0\} \cup \mathcal{D}_1 \cup \dots \cup \mathcal{D}_{J+1}$ となり、以下、 $|\mathcal{D}_j|$ は区間 $(T_{j-1}, T_j]$ に含まれる観測時刻数を表すものとする。なお、任意の $j \in \{1, \dots, J+1\}$ に対して、 $|\mathcal{D}_j| \neq 0$ 、すなわち、少なくとも1つの観測時刻 t_n が存在し、 $t_n \in \mathcal{D}_j$ を満たすと仮定する。このとき、パラメータのベクトルを $\mathbf{r}_{J+1} = (r_1, \dots, r_{J+1})$ と定義すれば、変化点集合 \mathcal{S}_J が与えられたときの観測系列 \mathcal{D} に対する対数尤度は次式で計算できる。

$$L(\mathcal{D}; \mathbf{r}_{J+1}, \mathcal{S}_J) = \log \prod_{j=1}^{J+1} \prod_{t_n \in \mathcal{D}_j} r_j \exp(-r_j(t_n - t_{p(n)})) \quad (1)$$

式 (1) の尤度を最大にするパラメータの最尤推定値は $j = 1, \dots, J+1$ に対して、 $\hat{r}_j^{-1} = 1/|\mathcal{D}_j| \sum_{t_n \in \mathcal{D}_j} (t_n - t_{p(n)})$ と計算でき、これを式 (1) に代入することで、次式を得る。

$$\begin{aligned} L(\mathcal{D}; \hat{\mathbf{r}}_{J+1}, \mathcal{S}_J) \\ = -N - \sum_{j=1}^{J+1} |\mathcal{D}_j| \log \left(\frac{1}{|\mathcal{D}_j|} \sum_{t_n \in \mathcal{D}_j} (t_n - t_{p(n)}) \right) \end{aligned} \quad (2)$$

したがって、ここでの変化点検出問題は、式 (2) を最大化する変化点集合 \mathcal{S}_J を求める問題となる。しかし、式 (2) 単独では、変化点集合 \mathcal{S}_J 導入の効果を直接評価できないため、本研究では、変化点が存在しないとした $\mathcal{S}_0 = \emptyset$ に対する式 (2) の値 $L(\mathcal{D}; \hat{\mathbf{r}}_1, \mathcal{S}_0)$ 、および変化点が J 個存在するとした $L(\mathcal{D}; \hat{\mathbf{r}}_{J+1}, \mathcal{S}_J)$ の尤度比の対数をとった次式を最大化する変化点集合 \mathcal{S}_J を求める問題を考える。

$$LR(\mathcal{S}_J) = L(\mathcal{D}; \hat{\mathbf{r}}_{J+1}, \mathcal{S}_J) - L(\mathcal{D}; \hat{\mathbf{r}}_1, \mathcal{S}_0) \quad (3)$$

なお、実際の情報拡散経路は不明であることがあるため、本稿では [大原 12] で高い近似精度が示されている直線状ツリーで情報拡散経路を近似する。このときの式 (3) の値は次式で与えられる。

$$LR(\mathcal{S}_J) = N \log \left(\frac{t_n - t_0}{N} \right) - \sum_{j=1}^{J+1} |\mathcal{D}_j| \log \left(\frac{T_j - T_{j-1}}{|\mathcal{D}_j|} \right) \quad (4)$$

4. 変化点検出法

本節では、情報拡散結果の時刻集合 $\mathcal{D} = \{t_0, t_1, \dots, t_N\}$ と変化点の個数 J が与えられたとき、変化点集合 $\mathcal{S}_J = \{T_1, \dots, T_J\}$ を時刻集合の部分集合 $\mathcal{S}_J \subset \mathcal{D}$ として求める変化点検出法 [大原 12] について述べた後、本稿で提案するモデル選択法について述べる。

4.1 シンプル法

まず、貪欲探索に基づくシンプル法について述べる。この方法では、既に選定した $(j-1)$ 個の変化点集合 \mathcal{S}_{j-1} を固定し、新たに付加するものとして最適な変化点 T_j を求め、 \mathcal{S}_{j-1} に加えることを $j=1$ から J まで繰り返すものである。そのアルゴリズムを以下に示す。

S1. $j=1$, $\mathcal{S}_0 = \emptyset$ と初期化する。

S2. $T_j = \arg \max_{t_n \in \mathcal{D}} \{LR(\mathcal{S}_{j-1} \cup \{t_n\})\}$ を求める。

S3. $\mathcal{S}_j = \mathcal{S}_{j-1} \cup \{T_j\}$ と更新する。

S4. $j=J$ なら \mathcal{S}_J を出力し終了。

S5. $j=j+1$ とし、S2 へ戻る。

ただし、 $i=2, \dots, j$ であり、S3 では、変化点集合 \mathcal{S}_j の要素は $T_{i-1} < T_i$ となるようインデックスを更新するものとする。シンプル法の計算量は $O(NJ)$ であり、ある程度 N が大きくなっても任意の J に対して実用的な時間で結果を得ることができる。一方、貪欲法に基づくため、局所解に陥ることがある。

4.2 提案法

一方、我々が提案した手法は、シンプル法と同程度の計算量で、解品質の向上を図るため、さらに局所探索を組合せたものである。具体的には、シンプル法で求めた変化点集合 \mathcal{S}_J を出発点とし、既に選定した一つの変化点 T_j を選び、これ以外の変化点集合 $\mathcal{S}_J \setminus \{T_j\}$ を固定し、より望ましい別の変化点 T'_j に置き換えることを、 $j=1, \dots, J$ のすべてで置き換えできなくなるまで繰り返す。ここで、 \setminus は集合差を表す。以下に提案法のアルゴリズムを示す。

P1. \mathcal{S}_J をシンプル法で求め、 $j=1, k=0$ と初期化する。

P2. $T'_j = \arg \max_{t_n \in \mathcal{D}} \{LR(\mathcal{S}_J \setminus \{T_j\} \cup \{t_n\})\}$ を求める。

P3. $T'_j = T_j$ なら $k = k+1$ 、さもなければ $k=0$ とし、 $\mathcal{S}_J = \mathcal{S}_J \setminus \{T_j\} \cup \{T'_j\}$ と更新する。

P4. $k=J$ なら \mathcal{S}_J を出力し終了。

P5. $j=J$ なら $j=1$ 、さもなければ $j=j+1$ とし、P2 へ戻る。

上記の提案法はシンプル法と比較して数倍程度の計算量が必要であることを実験的に確認しているが、可能な観測時刻のすべての組合せを網羅的に探索するナイーブな方法に比べて十分に高速であり、ある程度大規模なデータに対しても現実的な時間で解をえることができる [大原 12]。

4.3 モデル選択

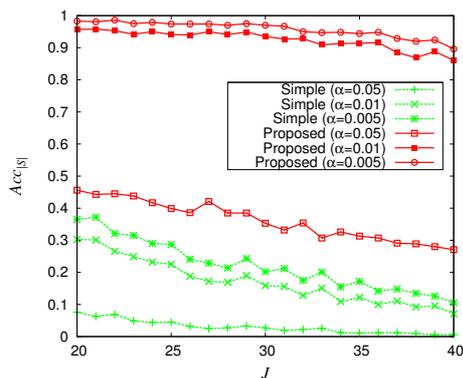
上述の2つの手法では、事前に変化点数 J を固定した上で、観測した情報拡散系列 $\mathcal{D} = \{t_0, t_1, \dots, t_N\}$ に対して最適な変化点集合 \mathcal{S}_J とパラメータベクトル $\hat{\mathbf{r}}_{J+1}$ を求めている。しかしながら、観測した情報拡散系列中に変化点数を事前に判断することは実際には困難である。そのため、本稿では、観測系列 \mathcal{D} から変化点数 J を推定するための尤度比検定に基づいたモデル選択法を上記の手法と組み合わせることを提案する。

具体的には、非負整数 J に対して、 $J+1$ 個の変化点をもつモデル \mathcal{M}_{J+1} が J 個の変化点をもつモデル \mathcal{M}_J より有意に優れているかどうかを尤度比検定を用いて判定する。このとき、 $Y_N(J+1)$ を2つのモデルを仮定した場合の対数尤度比検定における統計量とすると、その値は次式で与えられる。

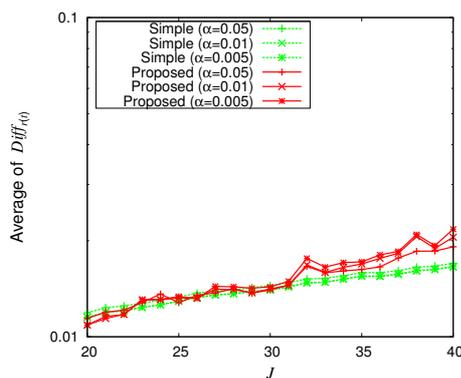
$$Y_N(J+1) = L(\mathcal{D}; \hat{\mathbf{r}}_{J+2}, \mathcal{S}_{J+1}) - L(\mathcal{D}; \hat{\mathbf{r}}_{J+1}, \mathcal{S}_J) \quad (5)$$

定義より、式 (5) は次のように計算できる。

$$Y_N(J+1) = LR(\mathcal{S}_{J+1}) - LR(\mathcal{S}_J) \quad (6)$$



(a) 変化点数の正解率



(b) 真のステップ関数推定したステップ関数の相対誤差

図 1: モデル選択法と組み合わせた場合の提案法とシンプル法の変化点の検出精度

従って、 $Y_N(J+1)$ は式 (4) と式 (6) から容易に計算することが可能である。ここで、モデル M_J は $J+1$ 次元のパラメータベクトル \mathbf{r}_{J+1} 、および J 次元のパラメータベクトル \mathbf{S}_J をもつことに注意されたい。このことから、2つのモデルのパラメータ空間の次元数の差は2となり、このとき、統計量 $2Y_N(J+1)$ は N が大きくなるにつれて自由度2の χ^2 分布に漸近する。従って、尤度比検定の考えに基づき、まず有意水準 α ($0 < \alpha < 1$) を定めた上で、統計量 $2Y_N(J+1)$ と $\chi_{2,\alpha}$ の値を比較することで、モデル M_{J+1} がモデル M_J よりも有意に優れているかどうかを評価することが可能となる。ここで、 $\chi_{2,\alpha}$ は自由度2の χ^2 分布における有意水準 α の限界値である。このモデル選択法と前述の変化点検出法を組合せた提案法のアルゴリズムを以下にまとめる。

- M1. $J = 0, \mathcal{S}_0 = \emptyset$ と初期化する。
- M2. 前節で述べた提案法を用いて \mathcal{S}_{J+1} を求める。
- M3. 統計量 $2Y_N(J+1)$ を計算する。
- M4. $2Y_N(J+1) \leq \chi_{2,\alpha}$ なら \mathcal{S}_J を出力し、終了。
- M5. $J = J+1$ とし、M2へ戻る。

5. 評価実験

5.1 人工データを用いた比較実験

まず、人工的に生成した擬似情報拡散系列を用いて、前述のモデル選択法と組み合わせた提案法とシンプル法の変化点検出、およびパラメータ推定の精度を評価した。人工データに関しては、事前に観測期間 $[0, T]$ 、および拡散系列に埋め込む変化点数 J を指定し、全観測期間を $J+1$ 個の等区間に分割後、各 j ($j = 1, \dots, J+1$) 番目の区間における観測時刻をパラメータ r_j の指数分布に従って生成した。ここで、 r_1 は1とし、 $j > 1$ の区間では k を整数としたとき、ある確率に従って $r_{j-1} = 2^k$ となるように決めた ($j = 1$ のとき $k = 0$)。具体的には、 $k \geq 0$ の場合は $1 - 2^{-(k+1)}$ の確率で $r_j = 2^{k-1}$ 、 $2^{-(k+1)}$ の確率で $r_j = 2^{k+1}$ となり、逆に $k < 0$ の場合は $1 - 2^{-(|k|+1)}$ の確率で $r_j = 2^{k+1}$ 、 $2^{-(|k|+1)}$ の確率で $r_j = 2^{k-1}$ となるようにした。これは、時間遅れパラメータが極端に大きな値、もしくは極端に小さな値にならないようにするためである。実際には、各部分区間における最初の観測時刻を変化点と考え、その観測点で前述のように時間遅れパラメータの値を変化させた。 T の値を100,000とし、 J を20から40まで変化させ、各 T と J の組合せに対して生成した1,000系列を実験に用いた。

提案法、およびシンプル法にモデル選択法を組み合わせたものによる実験結果を図1に示す。図1(a)の横軸は拡散系列に埋め込んだ変化点数、縦軸は1,000系列のうち埋め込んだ J 個の変化点を正しく検出できた割合であり、有意水準を $\alpha = 0.05, 0.01, 0.005$ とした場合に各手法がどれだけ正しく変化点を検出できたかを表している。この図から、 α がいずれの値においても提案法の検出精度がシンプル法を大きく上回っていることがわかる。特に、 $\alpha = 0.01, 0.005$ では提案法の精度が90%前後であるのに対し、シンプル法の精度は40%未満と非常に大きな差があることがわかる。

一方、図1(b)は、各 J に対して、データを生成したパラメータ値に対するステップ関数 $r(t)$ と各手法が推定したパラメータ値によるステップ関数 $\hat{r}(t)$ の誤差を $Diff_{r(t)} = \int_0^{T_n} |\hat{r}(t) - r(t)| dt / \int_0^{T_n} r(t) dt$ により評価した値の平均をプロットしたものであり、直感的にはこの値が大きいくほど、ステップ関数間のずれが大きくなる。図から、いずれの手法もその値はほぼ0.02未満となっており、推定誤差が小さいことがわかる。ここで、シンプル法による誤差のほうが小さい場合があるが、これは、シンプル法が提案法よりも多くの変化点を用いて真のステップ関数 $r(t)$ を近似しているためである。シンプル法では、正しい変化点数 J に対して生じた誤差を、より多くの変化点を導入することで修正しているといえる。検出した変化点を手掛かりとして実世界での事象を分析するという観点からは、本実験結果が示すように変化点数も正しく推定できる提案法のほうがより有用であると考えられる。

5.2 Twitterデータの解析結果

次に、Twitter上の実際の情報拡散系列に対して、モデル選択法と組み合わせた提案手法を適用した結果について述べる。本実験で用いた情報拡散系列は、東日本大震災が発生した2011年3月11日を含む2011年3月5日から3月24日までの約3週間に200件以上のツイートをした1,088,040名のTwitterユーザのツイート201,297,161件から抽出したものである。ここでは、[大原12]と同様に、他のユーザのツイートをそのまま発信するリツイートを情報の拡散と捉え、収集した全ツイートから各ユーザのIDを“@ID”形式で含むツイートを抽出し、そのユーザごとの時系列データを情報拡散系列として用いた。具体的には、系列長(ツイート数)が5,000以上の798名のユーザに対する情報拡散系列を用いた。

これらの情報拡散系列に対して、提案法を有意水準 $\alpha = 0.01$ として適用して検出した変化点数の分布を図2に示す。本実験では、 $J = 100$ としても提案法に基づいて変化点数を特定で

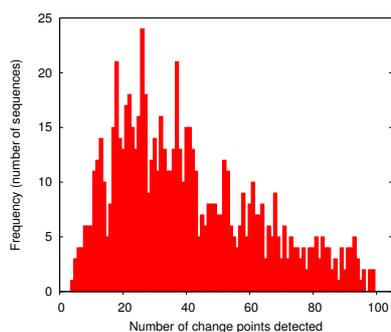


図 2: Twitter データから提案法により検出された変化点数の分布 ($\alpha = 0.01$).

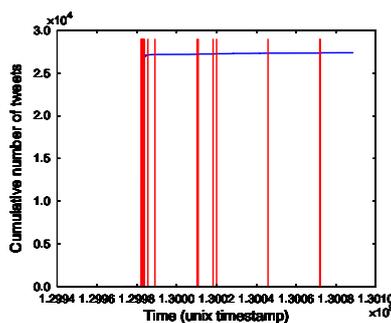


図 3: バースト的な情報拡散をもつ系列に対する変化点検出結果の例

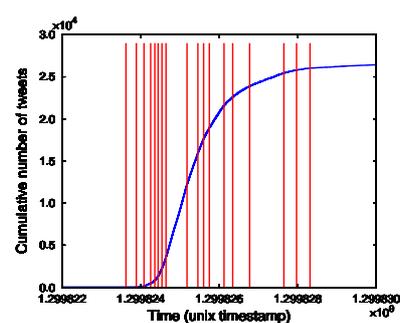


図 4: 図 3 のバースト区間の拡大図

きなかった場合は未検出としており、実際には 798 系列のうち 36 系列が未検出となった。図 2 は残りの 762 系列に対する変化点数の分布である。この図から、多くの系列では 10 から 40 個の変化点が検出されていることがわかる。ここで、[大原 12] でバースト的な拡散が見つかった系列に対する検出結果を図 3 に示す。この図からは、1 つのバースト的な拡散に対して複数の変化点が検出されていることがわかる。実際に検出された変化点数は 30 であった。この数は多いように思われるが、実際にはバースト的な拡散のあった時区間を詳細に見てみると図 4 のようになっており、提案法が投稿数の変化を精度よく近似できていることがわかる。対応する時区間における時間遅れパラメータのステップ関数の推定結果を図 5 に示す。ここからは、おおよそ 4 つのバースト的な拡散が確認でき、一番大きなものは図 3 中のバーストに対応する。一方、興味深いことに、時間的に 3 番目に現れたピークは実際に生じた大きな余震と一致していることがタイムスタンプから確認することができた。この情報拡散系列は、地震発生時の注意事項のリツイートが主なものであり、これはバースト的な情報の拡散と実世界の事象が対応する例と考えることができる。

6. おわりに

本論文では、ソーシャルネットワーク上の情報伝搬の遅延時間が指数分布などの確率分布に従うという仮定の下、観測した情報拡散系列からその分布のパラメータの変化点とパラメータ値を精度よく推定する手法を提案した。提案法は、尤度比検定に基づいたモデル選択法を用いることにより、観測した情報拡散系列に内在する変化点数を事前に指定することなく、高い精度で時間遅れパラメータの変化点とそのパラメータ値の推定を可能としている。本稿では、人工データを用いた評価実験を通して、貪欲探索に基づくシンプル法に比べて非常に高い精度で変化点を検出できることを示すとともに、実際の Twitter 上の情報拡散系列からも精度よく変化点を検出できることを示し、さらに実際のバースト的な情報拡散と実世界での事象が同期し得ることを確認した。

本研究では、尤度比検定に基づいたモデル選択法を用いたが、実際には AIC (赤池情報量基準) [Akaike 74] や MDL (最小記述上原理) [Rissanen 89] などの他のモデル選択指標も利用可能である。今後は、それらの指標を用いた提案法を評価する必要がある。

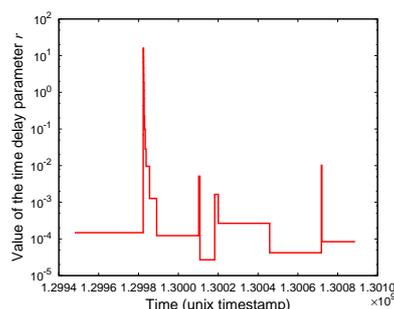


図 5: 図 4 に対する区間における時間遅れパラメータの推定結果

謝辞

本研究で用いたデータは東京大学 鳥海不二夫氏、和歌山大学 風間一洋氏によるものであり、その前処理には静岡県立大学の小出明弘氏の協力を得た。また、本研究は科学研究費補助金若手研究 (B) (No. 23700181) の補助を受けた。

参考文献

- [Akaike 74] Akaike, H.: A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, Vol. 19, pp. 716–723 (1974)
- [Rissanen 89] Rissanen, J.: *Stochastic Complexity in Statistical Inquiry*, World Scientific (1989)
- [Sadikov 11] Sadikov, E., Medina, M., Leskovec, J., and Garcia-Molina, H.: Correcting for Missing Data in Information Cascades, in *Proc. of WSDM 2011*, pp. 55–64 (2011)
- [Saito 09] Saito, K., Kimura, M., Ohara, K., and Motoda, H.: Learning Continuous-Time Information Diffusion Model for Social Behavioral Data Analysis, in *Proc. of ACML2009*, pp.322–337LNAI 5828 (2009)
- [Saito 10] Saito, K., Kimura, M., Ohara, K., and Motoda, H.: Selecting Information Diffusion Models over Social Networks for Behavioral Analysis, in *Proc. of ECML PKDD 2010*, pp. 180–195LNAI 6323 (2010)
- [大原 12] 大原剛三, 齊藤和巳, 木村昌弘, 元田浩: 情報拡散モデルに基づくツイート系列からのバースト期間検出, *日本データベース学会論文誌*, Vol.11, No.2, pp.25–30 (2012)