

# TWITTERにおける 情報・関心の伝播モデルの構築

## ~「選挙」を対象としたシミュレーション~

---

- ✓ はじめに(p2...p4)
- ✓ 分析対象について(p5...p7)
- ✓ 既存モデル(p8...p15)
- ✓ 提案モデル(p16...20)
- ✓ おわりに(p21...22)

静岡大学大学院  
工学研究科  
事業開発マネジメント専攻  
植田 雄介

## ◆ 研究背景

- ✓ SNSや動画(画像)共有サイト、ネット掲示板などのCGMの発達.

CGM:正式名称はconsumer generated media.

利用者によって生成されるメディアの総称.

- ✓ CGMがもたらす影響力[1]

- CGMを利用したマーケティング手法の発生と浸透.
- SNSによって誘発された消費の規模は1兆7100億に及ぶ[5].
- 2013年7月の参議院選挙のネット選挙解禁など行政に対する影響も伺える.

- ✓ 本研究は,CGMの中でもTwitterに注目する.

## ◆ Twitterとは

### ✓ 概要

- 140文字以内の短文を投稿するソーシャルメディアサービス
- 日本におけるTwitterアカウントは約**1400万**存在する。

### ✓ Twitterの特徴

- お互いの承認を必要とせず利用者同士が繋がる仕組み。
- 検索サービスの充実。
- 書き込み(ツイート)の件数を観測しやすい。

- ✓ 全世界で交換されるツイート量の**14%**が日本語と言われる。

## ◆ 研究目的

### ✓ 研究の目的

1. Twitterにおける情報や関心の伝播をモデル化
2. モデルによる情報伝播のシミュレーションを行う。
3. 2.より関心の推移の予測・傾向の導出を目指す。

### ✓ 研究の手順 (①～⑤は順序を示す)

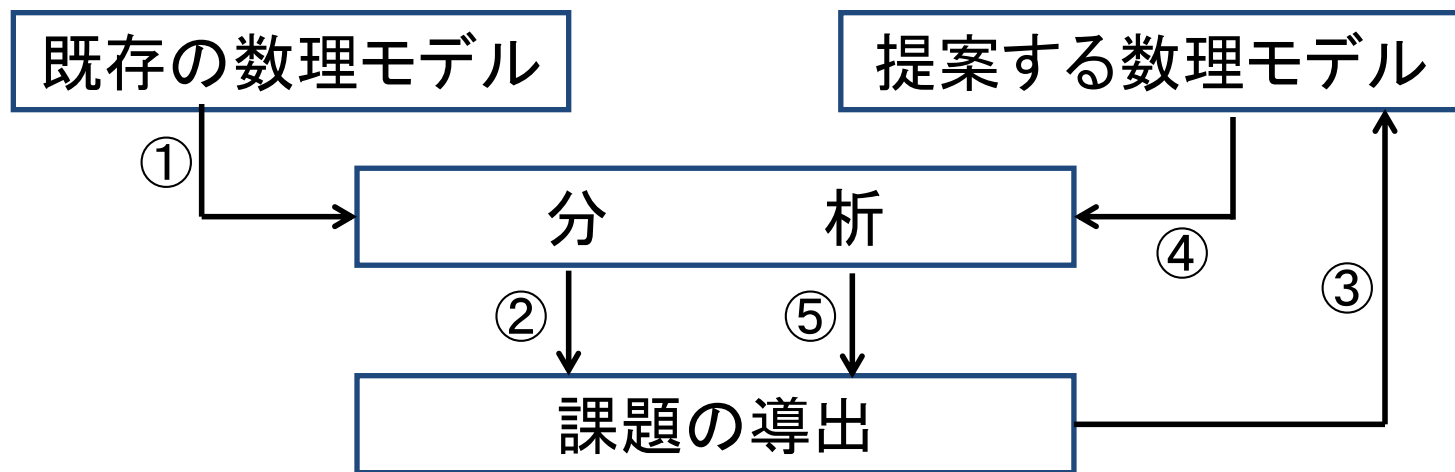


図1：目的までのフロー

## ◆ データ取得

- ✓ PythonによるTwitter Streaming APIを使用しデータを取得する.
  - データはパブリックタイムラインから条件を付けず取得.  
パブリックタイムライン：ほぼ全ての利用者のツイートが表示される箇所  
条件：ユーザー指定,キーワード指定のこと.
  - 取得期間：2013/7/4...7/21 (8:00...24:00)
  - 取得データ：ツイート +(アカウント名・投稿時間)
- ✓ ツイート件数の取得にはTOPSY[8]も用いる.  
TOPSY：リアルタイム性の検索サービス
- ✓ 取得するデータは**大量で雑多**であり,**精査する**必要がある.

### ◆ 分析対象

#### ✓ 分析ワード:「選挙」

理由:取得期間中,参議院選挙運動が行われていたため.

#### ✓ 取得データ概要

- 取得期間中の1日当たりの平均ツイート数は200161.3件.
- ツイートを投稿した利用者数の1日平均は188818人.
- Twitter利用者の平均フォロワー数は54.4人[7].

### ◆ 取得の背景

- ✓ インターネット選挙運動解禁に係る公職選挙法の一部改正より同年の夏の参議院選挙にてネットの使用が可能になった。
  - 各ネット事業者はCGMと選挙の連動企画を実施した。
  - Twitterでは立候補者の公式アカウントを設けており、選挙前(6/30時点)は153件確認されている。
  
- ✓ ネット選挙とTwitterの親和性[1]
  - 書き込み(ツイート)の件数を定量的に計りやすい。
  - 短文投稿に特化したことによる**即時性**。
  - 情報の受け手が送り手になれる**媒介性**。

## ◆ モデルの選出[6],[9]

### ✓ 感染症の数理モデルを用いる.

- 伝染病における個体の健康状態に従い,集団内の個体を分類するモデル.
- 病気に限らず,情報伝播や流行などの社会現象に適用可能.
- このモデルは状態の組み合わせ次第で,様々な形態を持つ.

### ✓ モデルの運用

- SIRモデル,SEIRモデルを既存モデルとして用いる.
- 2つのモデルの差異から提案モデル構築へのアプローチを模索していく.



## ◆ SIRモデル[2],[3]

## ✓ モデルの概要

- 感染症の数理モデルの基本形とも言えるモデルの1つ.

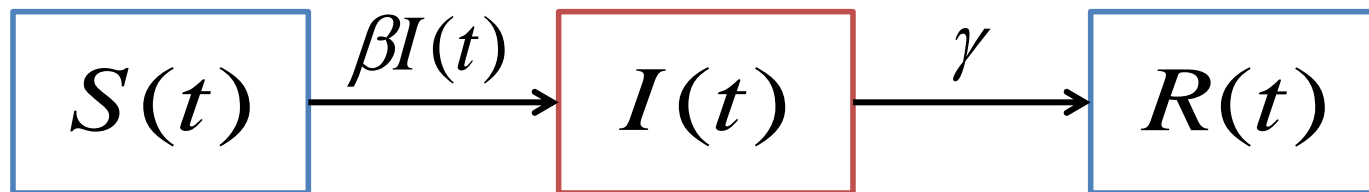
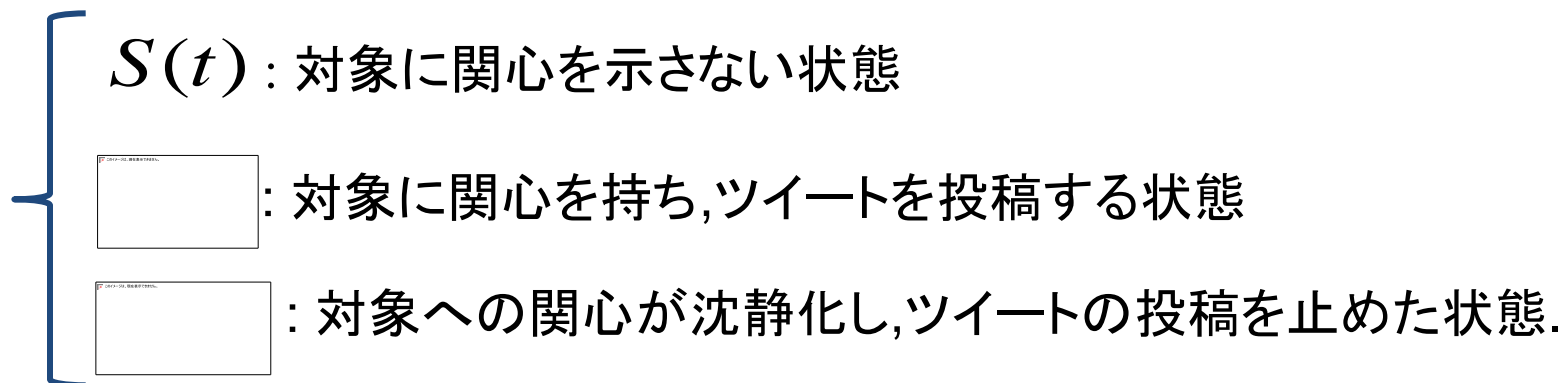


図2：SIRモデル

- 利用者の状態を以下の通りに示す.



## ✓ 利用者の総和

- 利用者の総数 $N$ は常に一定である.

$$S(t) + I(t) + R(t) = N \quad (1)$$

## ✓ 状態推移の式

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \quad (2) \quad \frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t) \quad (3)$$

$$\frac{dR(t)}{dt} = \gamma I(t) \quad (4)$$

$S(t)I(t)$  : 関心を持った利用者と持たない利用者の接触回数  
 $\beta$  : 関心を持ち、ツイートを投稿する率  
 $\gamma$  : 関心が冷め飽きる率

## ◆ SEIRモデル[2],[3]

## ✓ モデルの概要

- SIRモデルに状態 $E(t)$ を追加したモデル.

$E(t)$  :対象への関心を持ちつつもツイートの投稿をしない状態

- $E(t)$ は潜在的な関心を持った利用者と言える.
- 仮定：潜在的な関心は必ず顕在化する

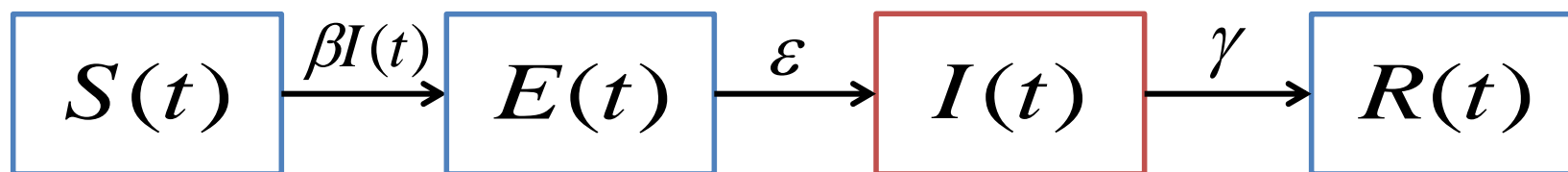


図3：SEIRモデル

## ✓ 利用者の総和

- 式(1)と同様に総数 $N$ は常に一定である.

$$S(t) + E(t) + I(t) + R(t) = N \quad (5)$$

## ✓ 状態推移の式

- SEIRモデルにおいても,利用者が関心を持つのはツイートを投稿するものに限られている.

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \quad (6) \quad \frac{dE(t)}{dt} = \beta S(t)I(t) - \varepsilon E(t) \quad (7)$$

$$\frac{dI(t)}{dt} = \varepsilon E(t) - \gamma I(t) \quad (8) \quad \frac{dR(t)}{dt} = \gamma I(t) \quad (9)$$

$\varepsilon$  : 潜在的な関心が顕在化する率

## ◆ 分析手順

### ✓ パラメータの設定手順

- SIR・SEIRモデル

$\beta$  : 取得した利用者のフォロー数・フォロワー数から推定する.

$\gamma$  : ツイート件数の増減から関心が持続する期間を推定し,その逆数から求める.

$\varepsilon$  :  $\gamma$ と同様

- 提案モデル

小刻みに値を変化させ,最も当て嵌まりの良い値を採用.

## ◆ シミュレーション結果

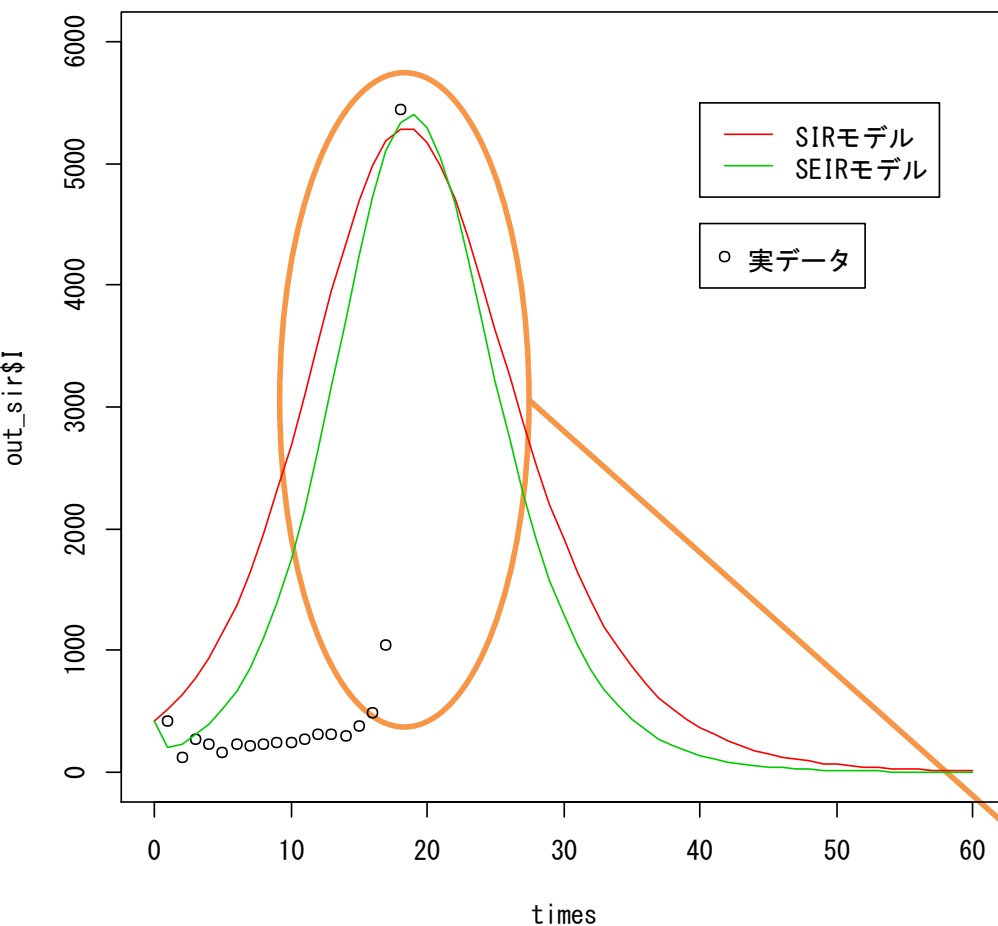


図5より

SEIRモデルによる推定値の方が実データに近い。



投稿件数の急上昇に推定値が追いついていない。

実データの急上昇に推定値が追いついていない。

図4: 既存モデルによるシミュレーション結果

## ✓ 各状態の推移

- 状態Sと状態Rが対照的に推移していることから、「選挙」のキーワードがTwitter全体に与える影響力は小さい。

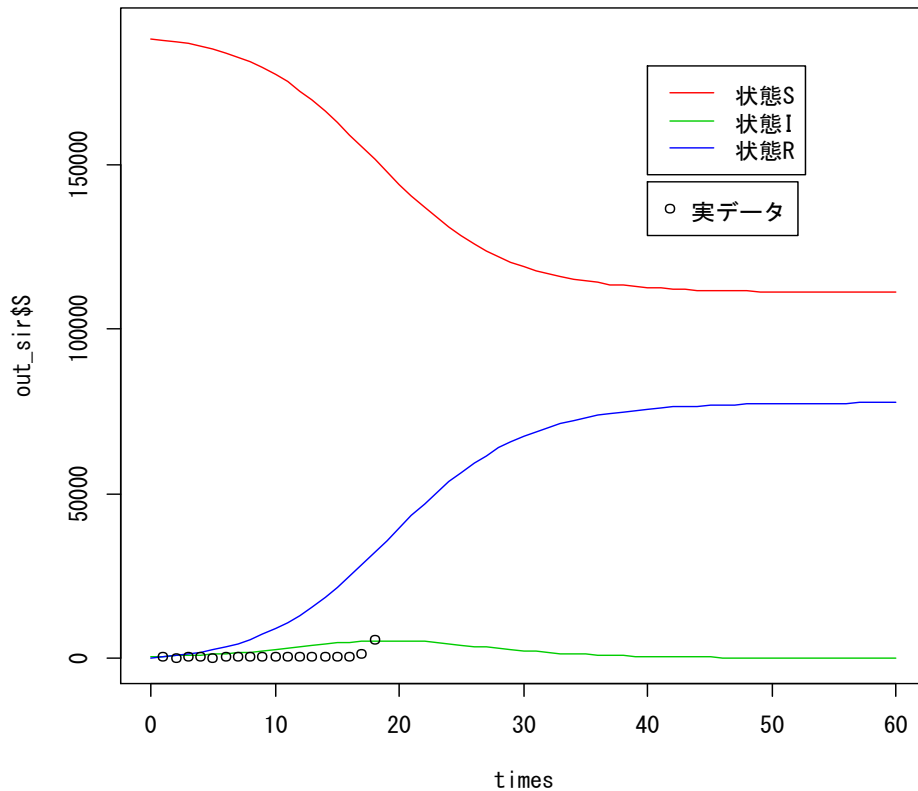


図5：SIRモデルによる各状態の推移

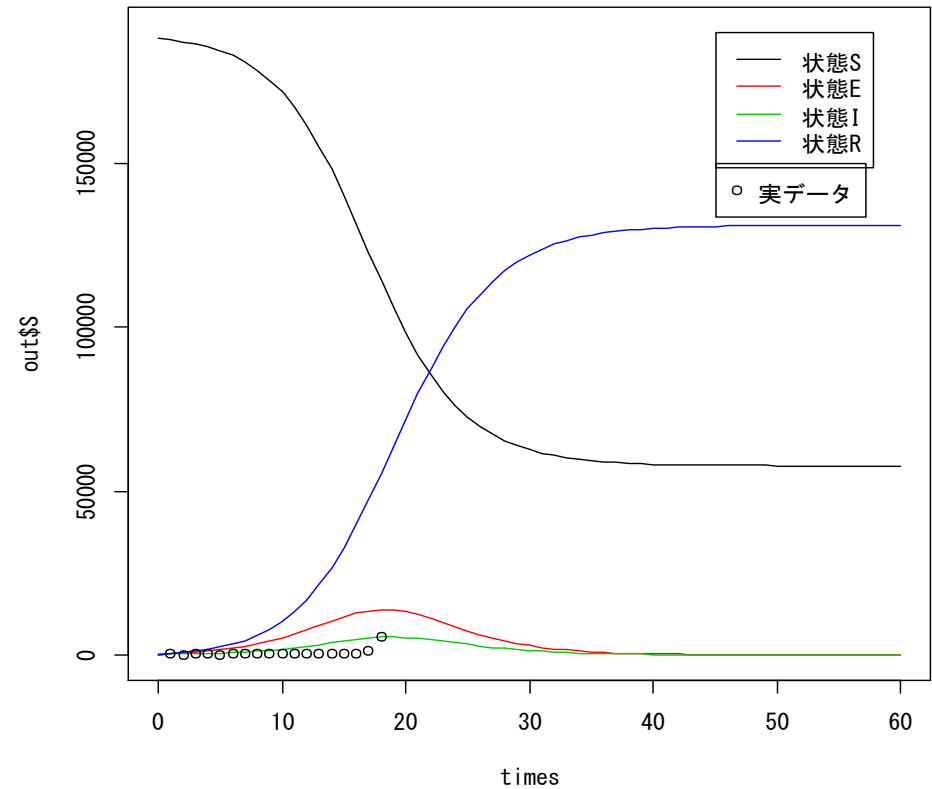


図6：SEIRモデルによる各状態の推移

## ◆ 提案モデル

✓ 図4よりSIRモデルに比べSEIRモデルの方が当て嵌まりが良い。

➡ 状態Eの存在が起因していると仮定する。

➡ 提案モデルはSEIRモデルを基に構築する。

✓ 潜在的な関心の推移に注目したモデルを提案する。

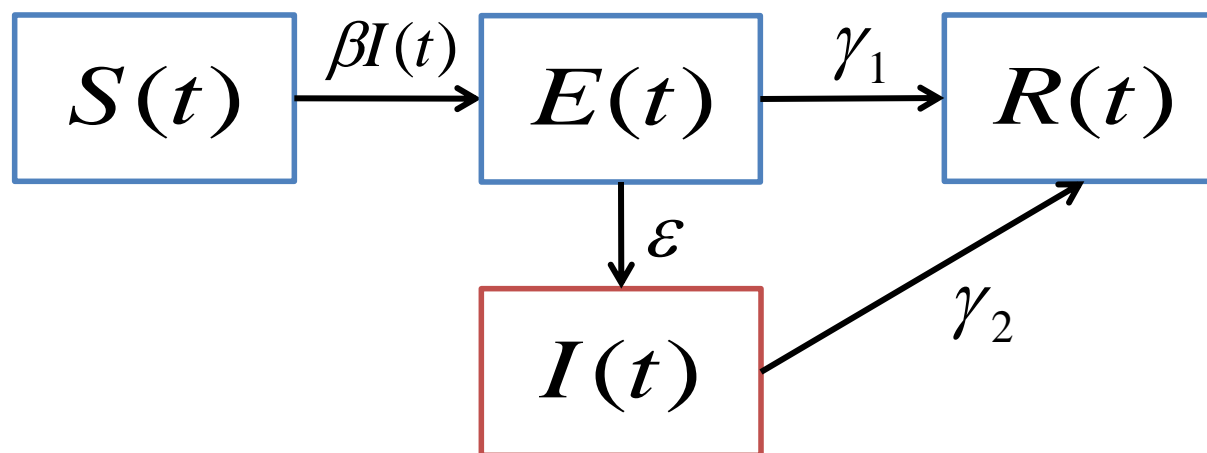


図7：提案モデル



- 式(5)と同様に利用者の総数 $N$ は常に一定である.

$$S(t) + E(t) + I(t) + R(t) = N \quad (10)$$

### ✓ 状態推移の式

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \quad (11) \quad \frac{dE(t)}{dt} = \beta S(t)I(t) - (\varepsilon + \gamma_1)E(t) \quad (12)$$

$$\frac{dI(t)}{dt} = \varepsilon E(t) + \gamma_2 I(t) \quad (13) \quad \frac{dR(t)}{dt} = \gamma_1 E(t) + \gamma_2 I(t) \quad (14)$$

$S(t)I(t)$  : 関心を持った利用者と持たない利用者の接触回数

$\beta$  : 伝達係数.関心の伝達のし易さ

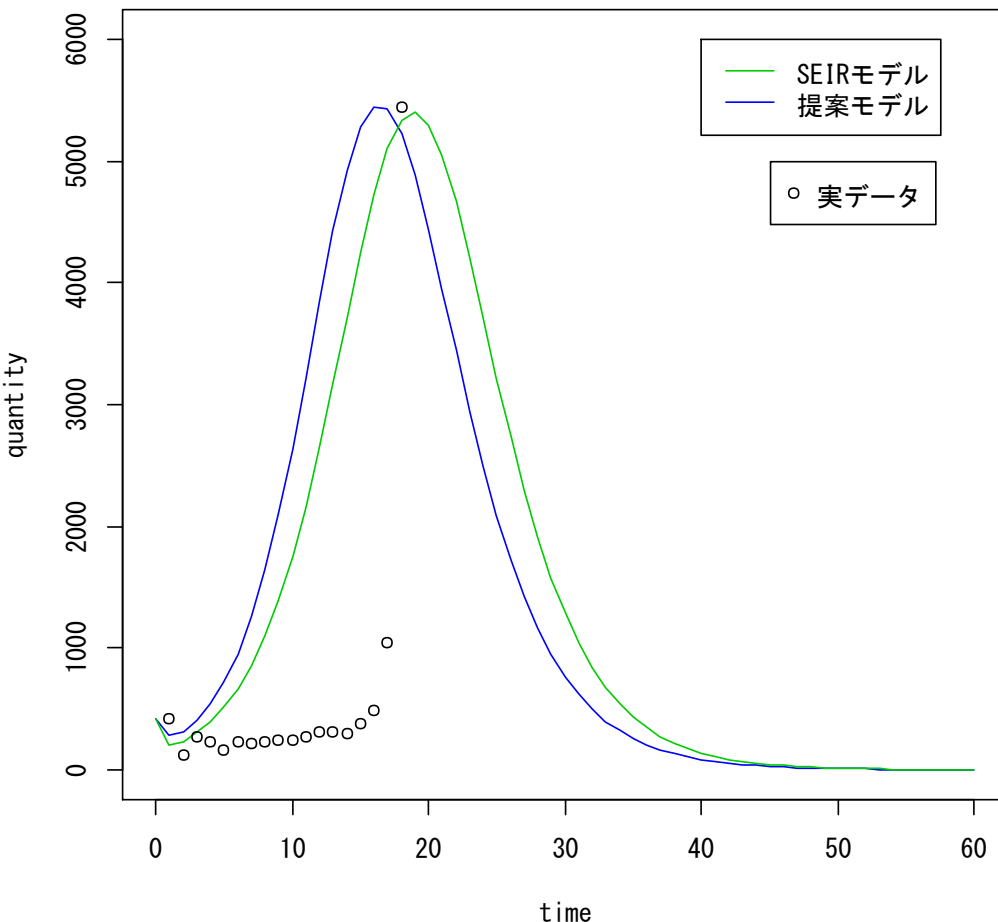
$\gamma_1$  : 関心が顕在化することなく沈静化する率

$\gamma_2$  : 関心が顕在化した後に冷める率

$\varepsilon$  : 潜在的な関心が顕在化する率

## ◆ シミュレーション結果

✓ SEIRモデルの推定結果と大差の無い結果となった。



仮説

潜在的な関心を持ったまま  
関心を失う利用者は少ない



各パラメータが推移に与える  
影響から確認する。

図8: 提案モデルによるシミュレーション結果

## ✓ 提案モデルとパラメータ

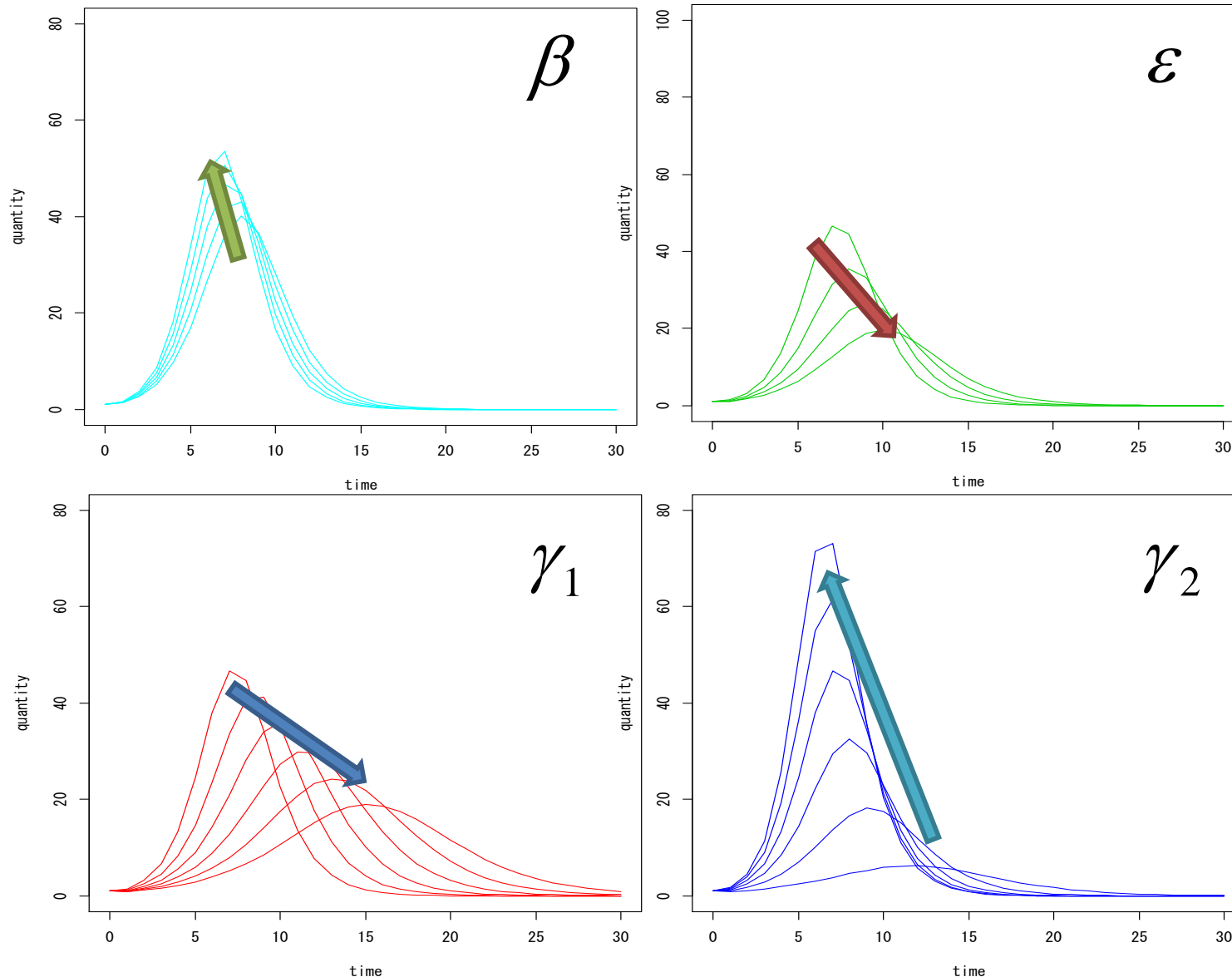


図9: 各パラメータの影響力(値が大きくなる程矢印の向きに推移する)

## ◆ 課題・改善点

### ✓ シミュレーション結果より


- 実データの値の急変は状態Eの推移の改変のみでは再現できない。

### ✓ パラメータの影響力より

- 図9右下の $\gamma_1$ の推移から、潜在的な関心を持った利用者が沈静へ向かう割合は小さくない。

 値の調整次第で推定の精度の向上が期待できる。

- 沈静へ向かう割合( $\gamma_1, \gamma_2$ )の変化が大きい。

 関心を持たせるよりも関心を失う要因を考慮する必要がある

## ◆ まとめ

### ✓ 分析結果より

- ツイート全体の推移から、「選挙」は利用者全体に与えた影響は小さい.
- 選挙運動期間から20日程度で関心は沈静化する.
- 関心の持続日数は0.65～1.4日とWeb上の出来事らしく非常に短い.

### ✓ 提案モデルの運用

- 現行では,SEIRモデルと大差がない.
- ツイートの投稿件数の急変の要因を導出する必要がある

## ◆ 今後の課題

### ✓ 分析対象ワード

#### 分析対象の性質

定められた日時になると大きな変動が生じるケース

Ex) 今回の分析対象の「選挙」

利用者の投稿のみに左右されるケース



性質を考慮した分析を行っていく。

### ✓ モデルの構築・改良

- 提案モデルの課題・改良点の解決.
- 考慮できなかったTwitterの特性の確認及び修正・追加.

## 参考文献

- [1] 陳玉霞,「ソーシャルメディアの集合知効果及び企業利用についての一考察」,『大阪産業大学経営論集』, 241-266,2012.
- [2] 板倉豊, 朝日弓未, 山口俊和,「口蹄疫流行に対する感染モデルの提案」,『オペレーションズ・リサーチ』, 56(12), 728-734, 2011.
- [3] Linda J.S. Allen, 生物数学入門,共立出版, 348-358,2011.
- [4] 中桐裕子,栗田治,「社会的なブームの微分方程式モデル」,『日本オペレーション・リサーチ学会和文論文誌』, 83-105,2004.
- [5] 日経MJ,SNS の消費1.7 兆円に, 掲載日2012/12/3.
- [6] 白井嵩士,榊剛史,鳥海不二夫,篠田孝祐,風間一洋,野田五十樹,沼尾正行,栗原聡,「Twitterにおけるデマツイートの拡散モデルの構築とデマ拡散防止モデルの推定」,『第26回人工知能学会全国大会』,2012.

- [7] 総務省, インターネット選挙運動の解禁に関する情報,  
([http://www.soumu.go.jp/senkyo/senkyo\\_s/naruhodo/naruhodo10.html](http://www.soumu.go.jp/senkyo/senkyo_s/naruhodo/naruhodo10.html)),  
最終閲覧日2013/6/20.
- [8] TOPSY, (<http://topsy.com/>), 最終閲覧日2012/9/30.
- [9] 上田浩, 大竹洋平, 河内一樹, 関元秀, 吉田拓, 藁科克英, 「社会における  
伝播現象の数理」, 『MODELing Seminar: 言説・情報・文化の伝播チー  
ム』, 2006.
- [10] 湯浅壘道, 「アメリカにおけるインターネット選挙運動の規制」, 九州国際  
大学法学論集, 17(1), 71-115, 2010.



# Appendix

## ◆ 取得データ詳細

表A1:取得ツイート一覧

DATE	PUBLIC	選挙	選挙_USER
704	193104	423	418
705	67479	120	119
706	230581	266	263
707	248402	234	231
708	173999	162	161
709	198606	223	219
710	185823	213	211
711	209520	228	226
713	207773	249	247
714	213971	268	267
715	222636	313	311
717	193706	302	301
719	209754	488	479
720	211569	1047	1035
721	235496	5448	5408

## ◆ 基本再生産数 $R_0$

### ✓ 基本再生産数とは

- 1人の感染者から感染して発症する二次感染者数の平均値.
- 本研究では,1人当たりの利用者が影響を与える利用者数の平均値をして扱う.

 フォロワー数を指す.

- 導出は式(A1)に示す通りになる.

$$R_0 = \frac{\beta S(0)}{\gamma} \quad (\text{A1})$$

## ◆ S言語プログラム

```
#パラメータの設定
N <- 188818 #利用者の総ツイート数
period <- 50 #期間
#SIRモデルの場合:beta,gamma
# SEIRモデルの場合:beta,epsilon,gamma
#提案モデルの場合:beta,epsilon,gamma1,gamma2
```

```
#各状態の初期値
date <- 1 ; S <- 188400
E <- 0 ; I <- 418 ; R <- 0
```

```
result <- matrix(1:(4*period),nrow=50)
```

```
#SIRモデル
begin <- c(date,S,I,R)
result[1,] <- begin
for (n in 1:(period-1)){
  date <- n+1
  R <- gamma*I
  I <- beta*S*I-gamma*I
  S <- -beta*S*I
  SIR <- c(date,S,I,R)
  result[n+1,] <- SIR
}
```

```
#SEIRモデル
begin <- c(date,S,E,I,R)
result[1,] <- begin
for (n in 1:(period-1)){
  date <- n+1
  R <- gamma*I
  I <- epsilon*E-gamma*I
  E <- beta*S*I-epsilon*E
  S <- -beta*S*I
  SEIR <- c(date,S,E,I,R)
  result[n+1,] <- SEIR
}
```

```
#提案モデル
begin <- c(date,S,E,I,R)
result[1,] <- begin
for (n in 1:(period-1)){
  date <- n+1
  R <- gamma1*E+gamma2*I
  I <- epsilon*E-gamma2*I
  E <- beta*S*I-(epsilon+gamma1)*E
  S <- -beta*S*I
  SEIR2 <- c(date,S,E,I,R)
  result[n+1,] <- SEIR2
}
```