

原発を対象としたTwitter利用者の 関心度移行の推定と検証

◆ 目次

- 研究背景 (p2~4)
- 研究目的 (p5)
- 既存モデル (p6~10)
- データ収集・分析 (p11~12)
- 分析結果 (p13~18)
- 提案モデル (p19~23)
- 最後に (p24~25)

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

◆CGM (consumer generated media) の普及 [2], [3]

* インターネットを通じて、利用者自ら情報を発信し生成するメディアの総称

- 通信インフラの充実と高速化.
- 無料で気軽に閲覧・書き込みができる.
- 日本のSNS利用者は**4289万人 (普及率45%)** に上る.

2014年末には5643万人へ

毎日利用する：64%
ほとんど利用しない：20%

- クチコミサイトの**月間8356万人**利用している。(2012年3月調べ)
※株式会社カカクコム及び連携子会社運営サイト



年々利用者が増加している傾向にある.

◆ Twitterとは[4], [7]

➤ 概要

- 140文字以内の文を投稿するSNSサイト.
- Twitterの利用者数は**1392万人**に及ぶ。(2012年8月調べ)
- Facebookに次ぐ日本で2番目の規模を持つ.

➤ 用語一覧

ツイート： 140文字以内の投稿を指す.

フォロー： 特定のユーザーのツイートを閲覧する仕組み.

フォロワー： フォローしているユーザーのこと.

リツイート： 他のユーザーのツイートを自分のフォロワーにも見せること.

➤ 特徴

- コンセプトは “「今」何をしている？” である。
- 仮定：簡潔に書くことで印象や評価がはっきりする。
- お互いの承認を必要せずリンク可能。
* フォロー, フォロワーを指す
- フォロワー数やリツイート数からオピニオンリーダーの目星がつけやすい。

➤ 採用理由

- 検索サービスが充実している。
- 観測が容易であり、定量的なデータとして扱いやすい。

◆研究目的

- Twitterを用いて“原発”における情報の拡散と**利用者の関心の推移を定量的に明らかにする。**

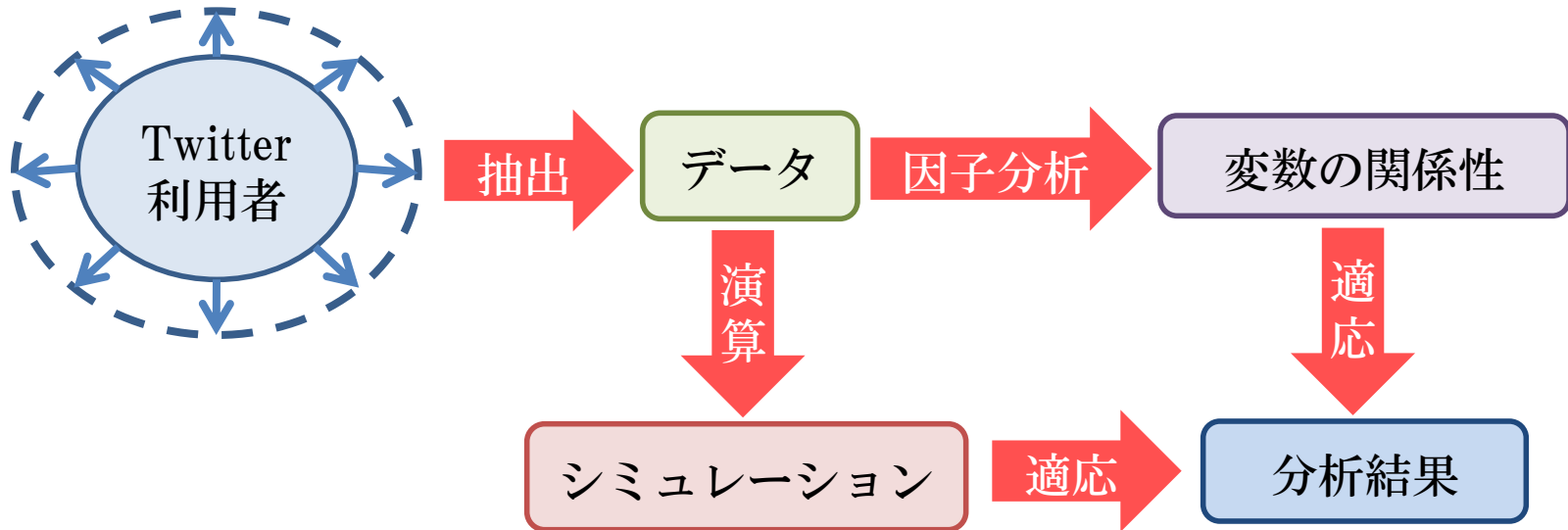


図1:研究概要図

- 抽出データの関係性を明らかにする.
- 実データとシミュレーションを比較.
- モデル(拡張含む)の妥当性の評価を行う.

◆標準モデル[5]

- ブーム時における利用者の状態の推移を表すモデル。
 - * ブーム:関心が高まる状態
- 各状態の利用者の人口に依存している。

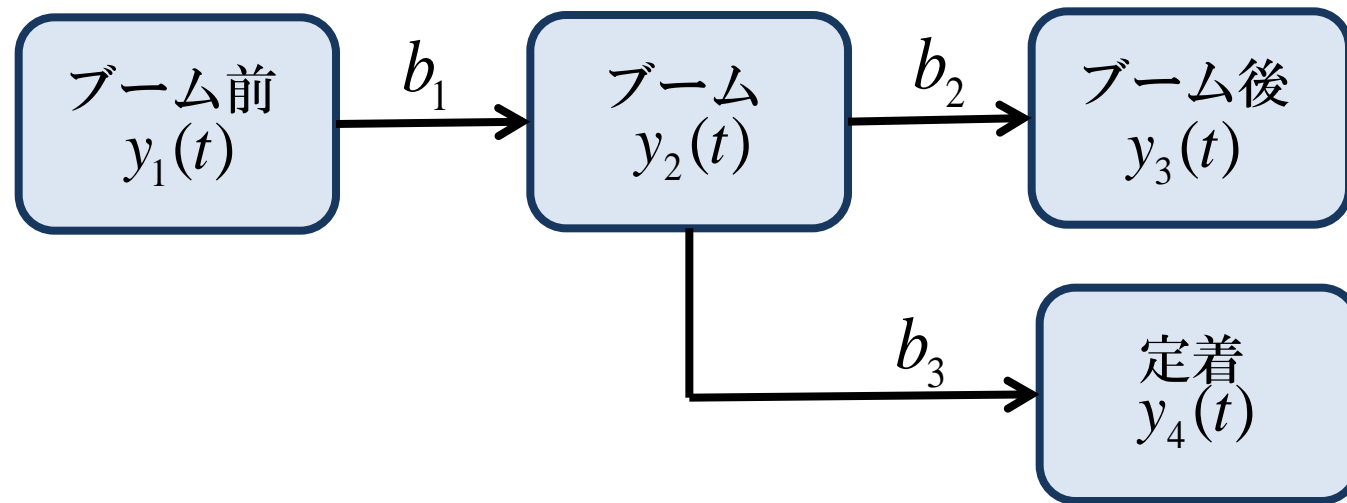


図2:利用者の状態推移 (標準モデル)

$y_1(t)$: ブーム前状態の利用者の数	b_1 : ブームに乗じる利用者の割合
$y_2(t)$: ブーム状態の利用者の数	b_2 : 利用を中止するブーム利用者の割合
$y_3(t)$: ブーム後状態の利用者の数	b_3 : 定着に移行するブーム利用者の割合
$y_4(t)$: 定着状態の利用者の数	

➤ パラメータの範囲

$$b_1 \geq 0 \quad (1) \quad b_2 \geq 0 \quad (2) \quad b_2 + b_3 \geq 0 \quad (3)$$

- (3) は逆流 (定着→ブームに移行) することも考慮している.

➤ 仮定：利用者の総人口は常に一定.

$$y_1(t) + y_2(t) + y_3(t) + y_4(t) = S \quad (4) \quad S : \text{利用者総人口}$$

➤ 状態推移の式

- 図2より(5)~(8)の微分方程式が得られる.

$$y_1'(t) = -b_1 y_1(t) \quad (5)$$

$$y_2'(t) = b_1 y_1(t) - (b_2 + b_3) y_2(t) \quad (6)$$

$$y_3'(t) = b_2 y_2(t) \quad (7)$$

$$y_4'(t) = b_3 y_2(t) \quad (8)$$

➤ 初期条件

$$y_1(t) = (1-k)S \quad (9) \quad y_2(t) = 0 \quad (10)$$

$$y_3(t) = 0 \quad (11) \quad y_4(t) = kS \quad (t \leq T) \quad (12)$$

$k(\geq 0)$: 既に定着状態になっている割合 T : ブームが始まる時刻

➤ 微分方程式の解

- 微分方程式(5)～(8)に初期条件(9)～(12)を与えて解く.

① ブームが始まる前の解 ($t \leq T$)

初期条件(9)～(12)と同様.

② ブームが始まった後の解 ($t > T$)

解法には積分因子を用いる手法を採用.

◆ 数値シミュレーション

- 微分方程式の解のパラメータに任意の値を代入する。

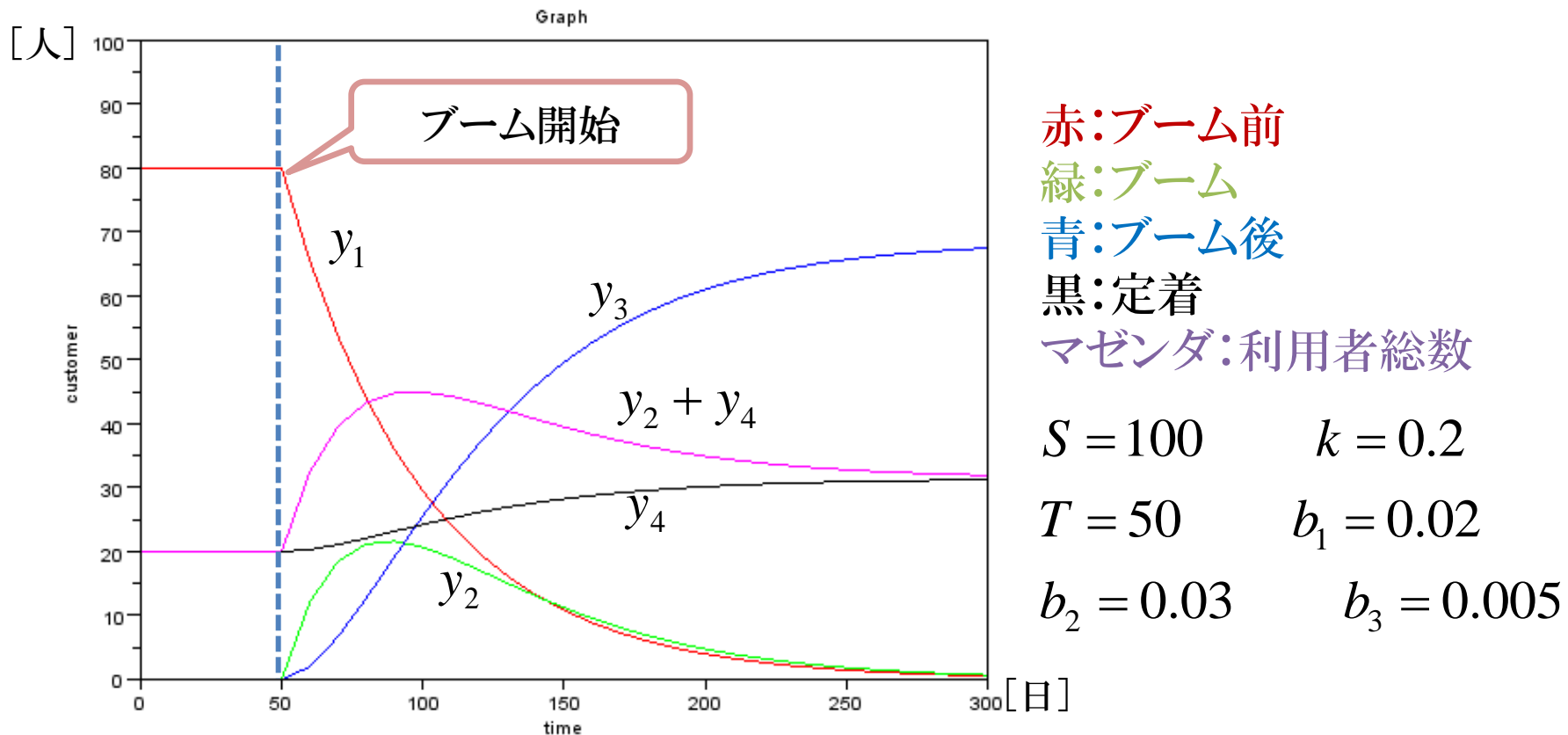


図3: シミュレーション結果(数値計算ソフト: Scilab)



図3は図2の示す利用者の移り変わりを示している。

◆突発的なブーム時のモデル[5]

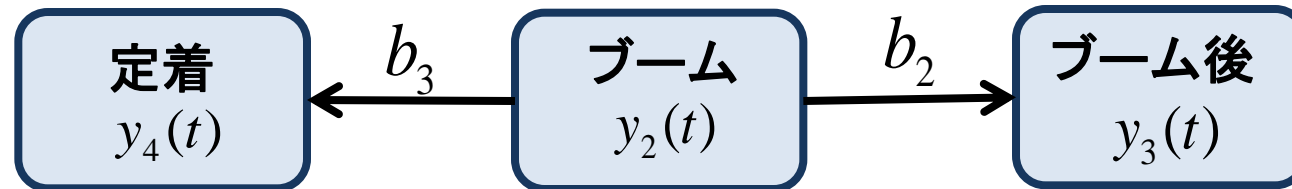


図4:利用者の状態推移(突発的なブーム)

- 図2の標準モデルから派生したモデル.
- 利用者の関心が**一瞬**で高まっている. ($b_1 \rightarrow \infty$)
 - * 図1の“ブーム前”が瞬時に“ブーム”に移行している.
- 対象となる事象の例: 3.11の震災, 狂牛病問題
- 初期条件(突発的なブーム時)

$$y_1(T) = 0 \quad (13) \quad y_2(T) = (1-k)S \quad (14)$$

$$y_3(T) = 0 \quad (15) \quad y_4(T) = kS \quad (16)$$

◆収集方法

- 調査対象： 2011下半期Google急上昇ランキング [1]
- 調査期間： 2010年9月～2012年9月

- 採用理由

原発と同時期に利用者の関心を集めた事柄を把握し、
関心の推移と傾向を知る。

大量のデータを用いての推定を可能にする。

- 使用サイト

- A. GoogleBlog : 検出ワードの検索
- B. Topsy : ツイート検索サイト

- 方法

- ① A. より原発を含めた検出ワードを明示。
- ② B. より、ツイート件数を月毎求める。

◆分析手法

➤ 多変量解析 (因子分析)

- 原発を含めたGoogleの急上昇検索ワードから各データ (変数) の関係性 (背景) を考察する.

- 因子数を2つと想定.
- 因子数の抽出法は最尤法を使用.
- 回転法はバリマックス回転を使用.

➤ 数値シミュレーション

- 構築したモデルによるシミュレーションを行う.
- パラメータの推定にはExcelのソルバーを用いる.
- 単位時間辺りの関心度を $b(y_2 + y_4)$ と仮定する.

* b : 関心の高さを示す係数

◆検出ワード

- 本研究の調査対象である原発と同時期に利用者の関心を集めたワードを以下に示す。

地震, 停電, 放射能, 原発, 東京電力, iPad2, iPhone5
セシウム, Facebook, Google+

- 震災関連, アップルの新製品, SNSサービス, この3つが2011年のトレンドであることが分かる。

◆分析結果

➤ 多変量解析 (因子分析) ①

表1: 因子の寄与量

Factor1	Factor2
3.954218	3.359189

表2: 寄与率

	Factor1	Factor2
SS loadings	3.954218	3.359189
Proportion Var	0.395422	0.335919
Cumulative Var	0.395422	0.731341

➤ 多変量解析 (因子分析) ②

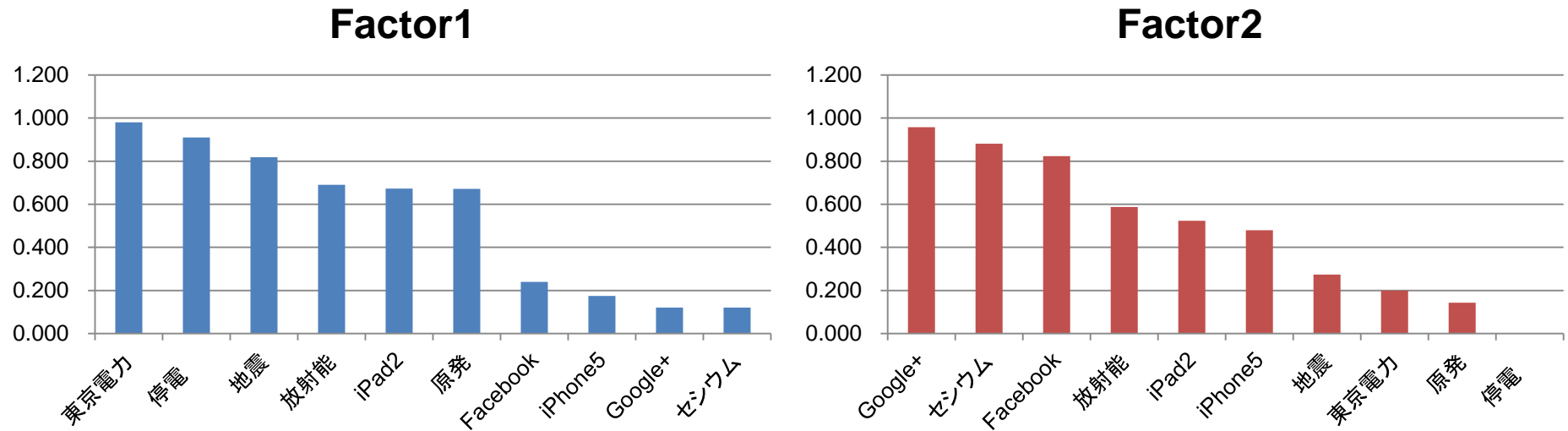


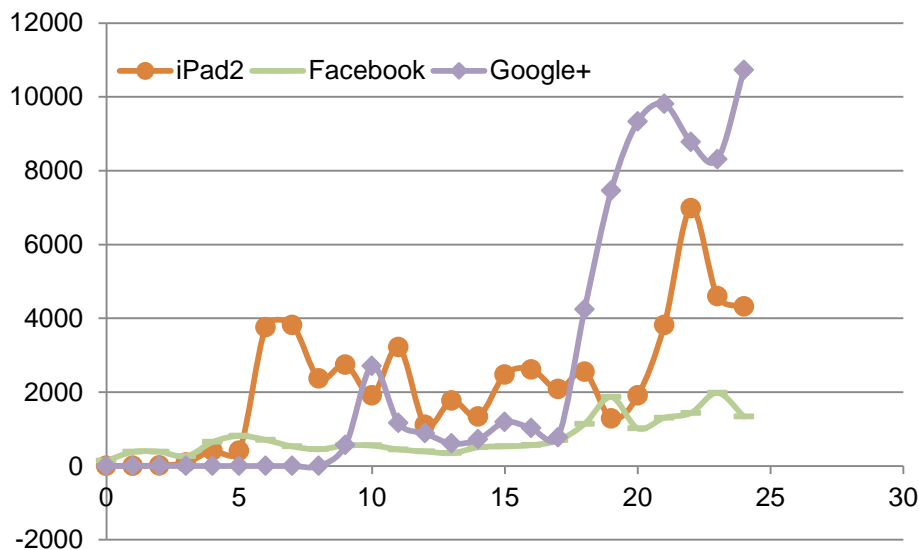
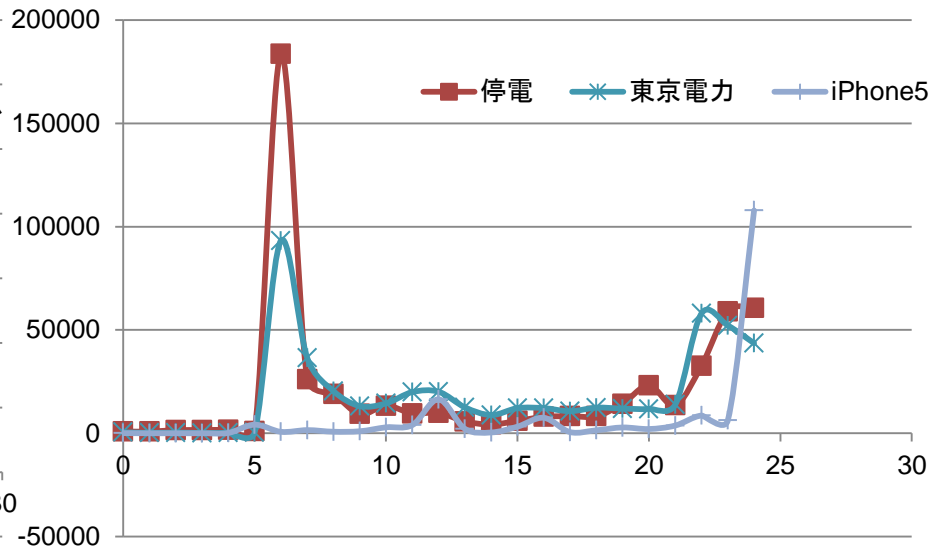
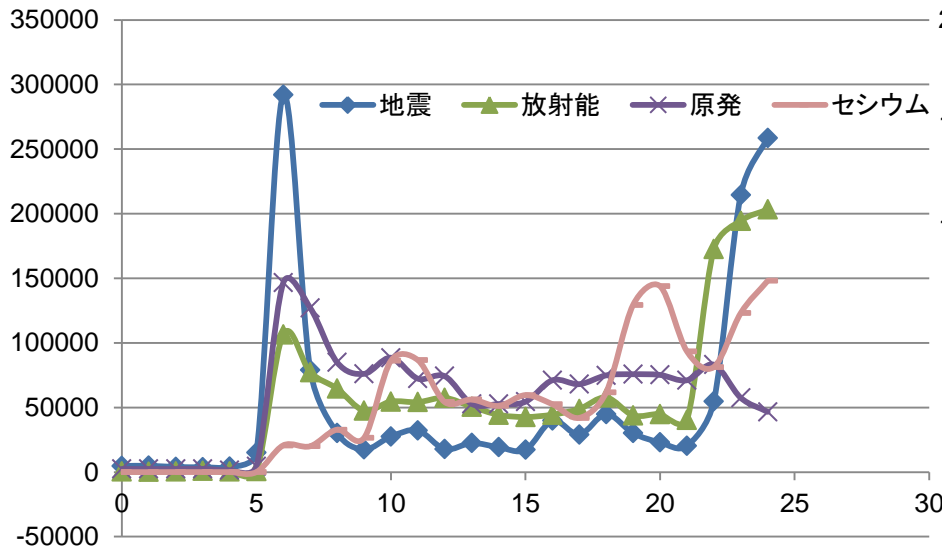
図5:各因子に対する因子負荷量

- 図5に示した因子負荷量のみでは各因子が何を表しているのか不透明である。



- ワードの名称やトレンドによる分類ではなく、ツイート件数の推移, 因子得点の散布図から判別を試みる。

➤ 多変量解析 (因子分析) ③



ツイートの推移の形から

地震, 原発, 停電, 東京電力
放射能, iPad2

セシウム, iPhone5,
Facebook, Google+

と区分けできる。

図6:検出ワードのツイート件数の推移 (縦軸:ツイート件数,横軸:経過月数)

➤ 多変量解析 (因子分析) ④

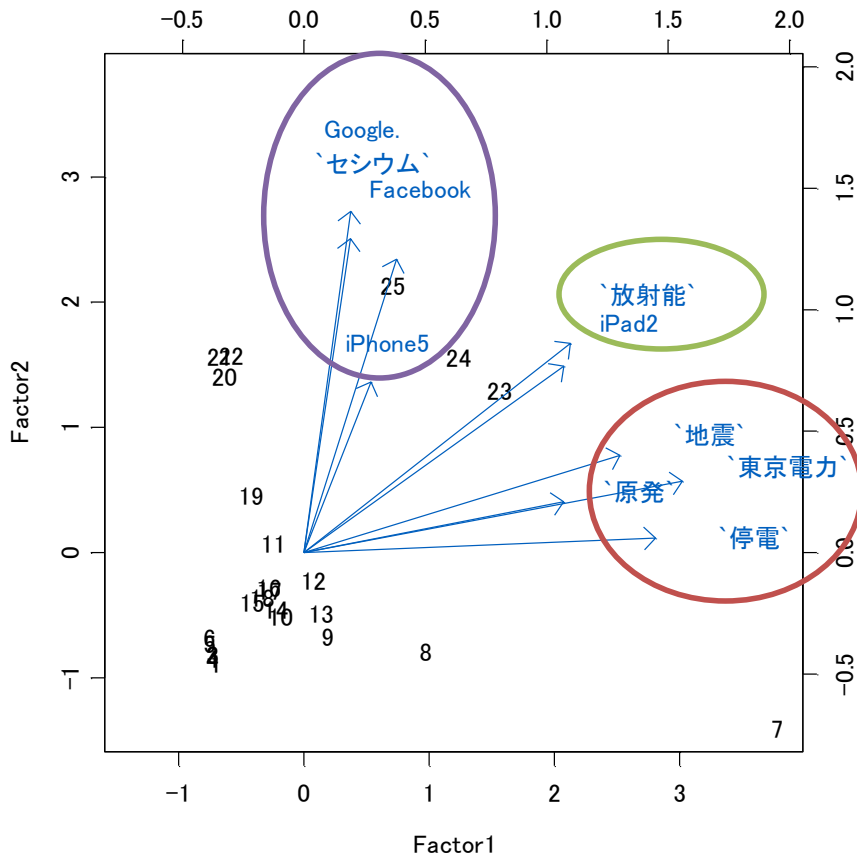


図7:因子得点の散布図

図6, 図7より

Factor1は3.11の震災

Factor2はツイートの持続性

* ツイートの持続性とは、
観測中ツイート件数が落ち込みを
あまり見せず上昇することを指
を説明する因子と定義する。

以上から

原発は震災の影響を大きく受けた
ものの落ち着きを見せつつある。

と仮定できる。* 仮定を(α)とする。

- 原発を含めた各ワードの関係性が明らかになった。
- 上記の仮定をツイート件数の推移から確認を行う。

▶ 数値シミュレーション

- ツイート件数の推定を行い、モデルの妥当性と情報が拡散する様相を検証する。

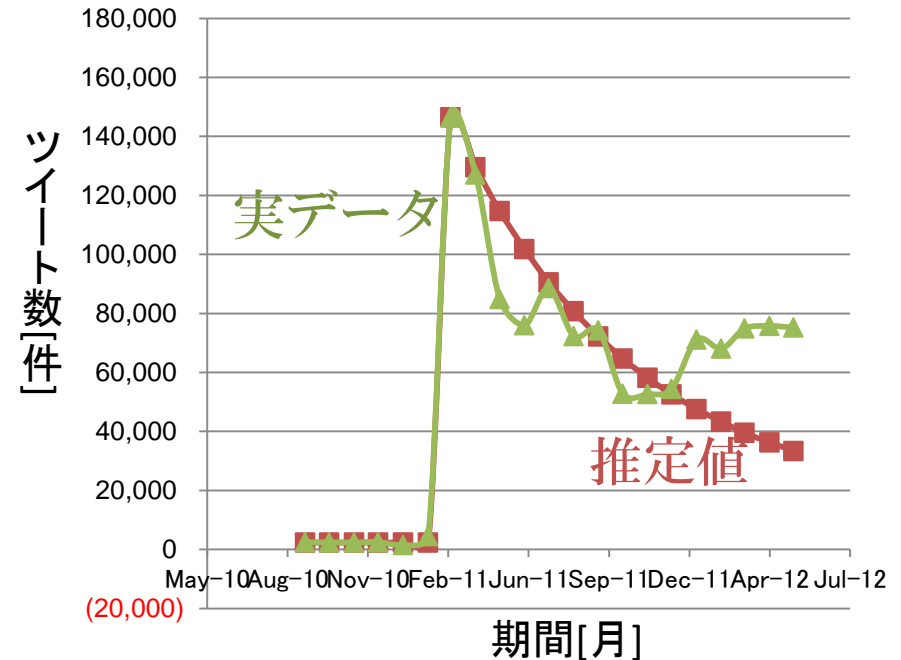
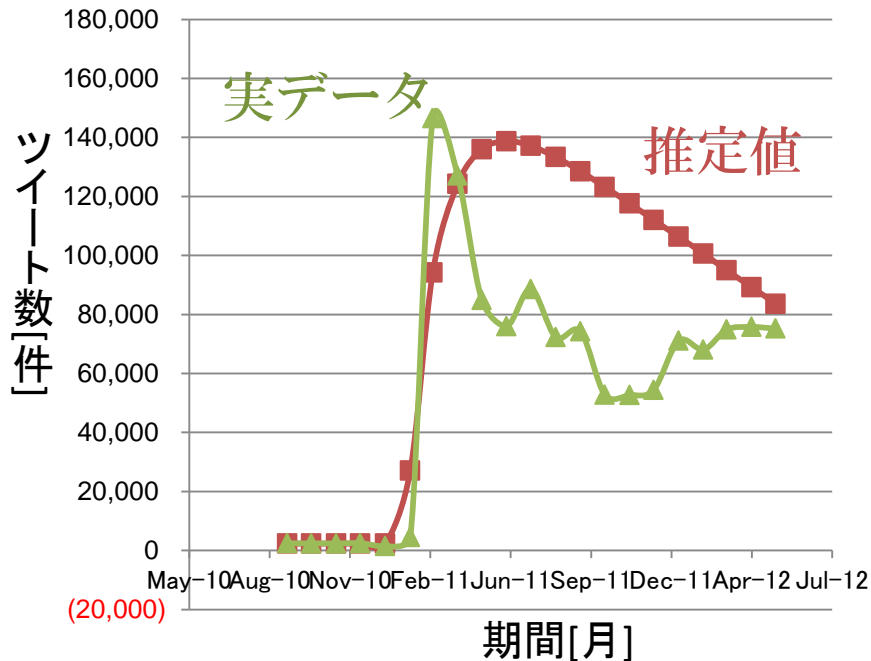


図8: ツイート件数の推移(左: 標準モデル 右: 突発的なモデル)

- 図8より左は右と比べ、実データに沿った推移を取っていることが分かる。
- 以上より、原発は突発的なブームに属していると言える。

◆モデルの改善点

➤ 既存モデルの問題

- Twitterの性質の反映

既存のモデルでは、ツイートをする利用者のみが関心を持つことになっており、関心を持っているがツイートはしない利用者も存在が無視されている。

- 観測時の利用者総人口

利用者の人口は時間変化するものであり、利用者総人口が常に一定のままではモデルの精度に関わる。

➤ シミュレーション結果から分かること

- 図6より、ブーム前後で関心の大きさに大きな差があり、ブーム前の推移は一先ず無視できるのではないか。



本研究向けにモデルの再構築を行う

◆提案モデル (Twitterから利用者の関心を予測するモデル)

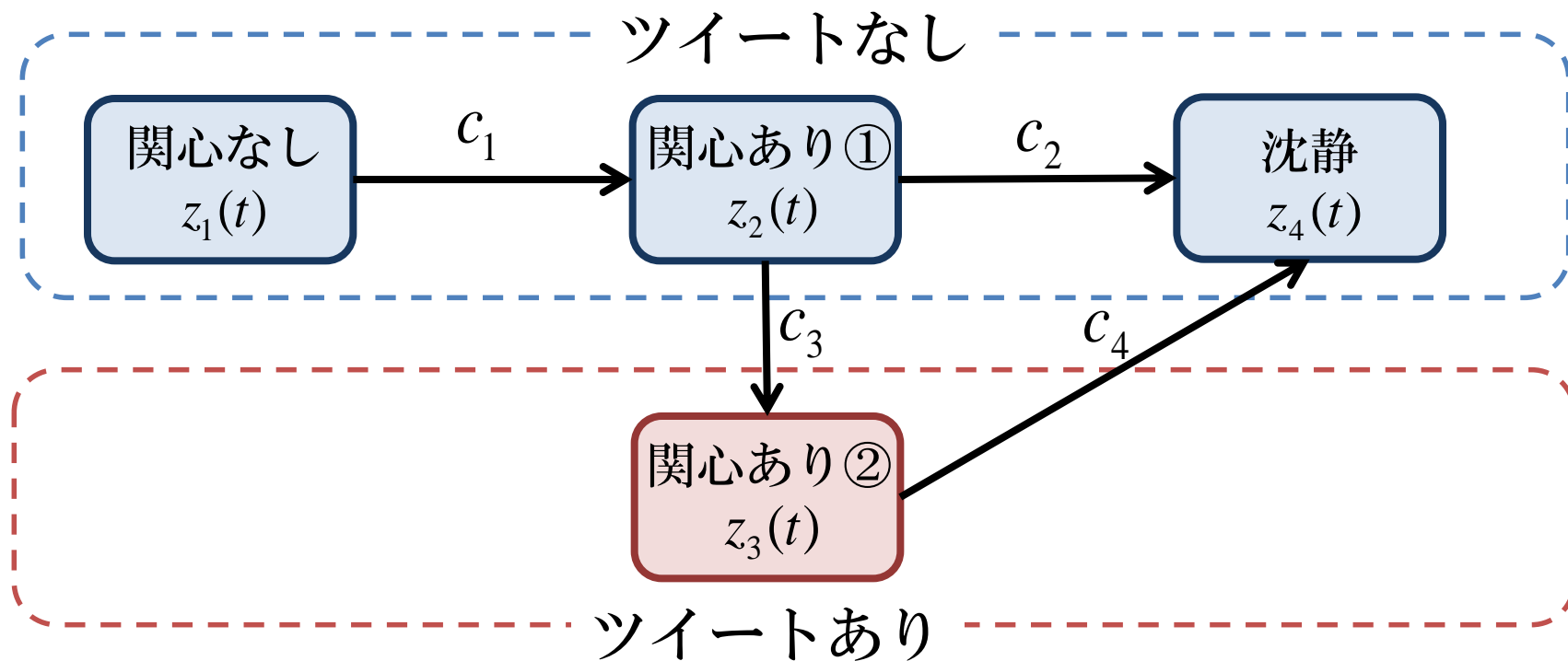


図9:利用者の状態推移 (提案モデル)

- | | |
|----------------------------|------------------------|
| $z_1(t)$: 関心のない利用者の割合 | c_1 : 関心を持ち始める利用者の割合 |
| $z_2(t)$: 関心を持つ利用者の割合 | c_2 : 関心が静まる利用者の割合 |
| $z_3(t)$: 関心を持ち書き込む利用者の割合 | c_3 : ツイートを行う利用者の割合 |
| $z_4(t)$: 関心が落ち着く利用者の割合 | c_4 : ツイートをやめる利用者の割合 |

➤ 状態推移の式

- 図9より(17)~(20)の微分方程式が得られる。

$$z_1'(t) = -c_1 z_1(t) \quad (17)$$

$$z_2'(t) = c_1 z_1(t) - (c_2 + c_3) z_2(t) \quad (18)$$

$$z_3'(t) = c_3 z_2(t) - c_4 z_3(t) \quad (19)$$

$$z_4'(t) = c_2 z_2(t) + c_4 z_3(t) \quad (20)$$

➤ 初期条件

$$z_1(T) = (1-l-m) \quad (21) \quad z_2(T) = l \quad (22)$$

$$z_3(T) = m \quad (23) \quad z_4(T) = 0 \quad (24)$$

$\left[\begin{array}{l} l(\geq 0) : \text{既に関心あり①状態になっている割合} \\ m(\geq 0) : \text{既に関心あり②状態になっている割合} \end{array} \right.$

➤ 仮定：各状態の総和

$$z_1(t) + z_2(t) + z_3(t) + z_4(t) = 1 \quad (25)$$

- 各状態の利用者の推移を人口ではなく割合として捉える。
- よって利用者総人口に依らない推定が可能になる。

➤ 微分方程式の解

- 既存モデルと同様に求める。

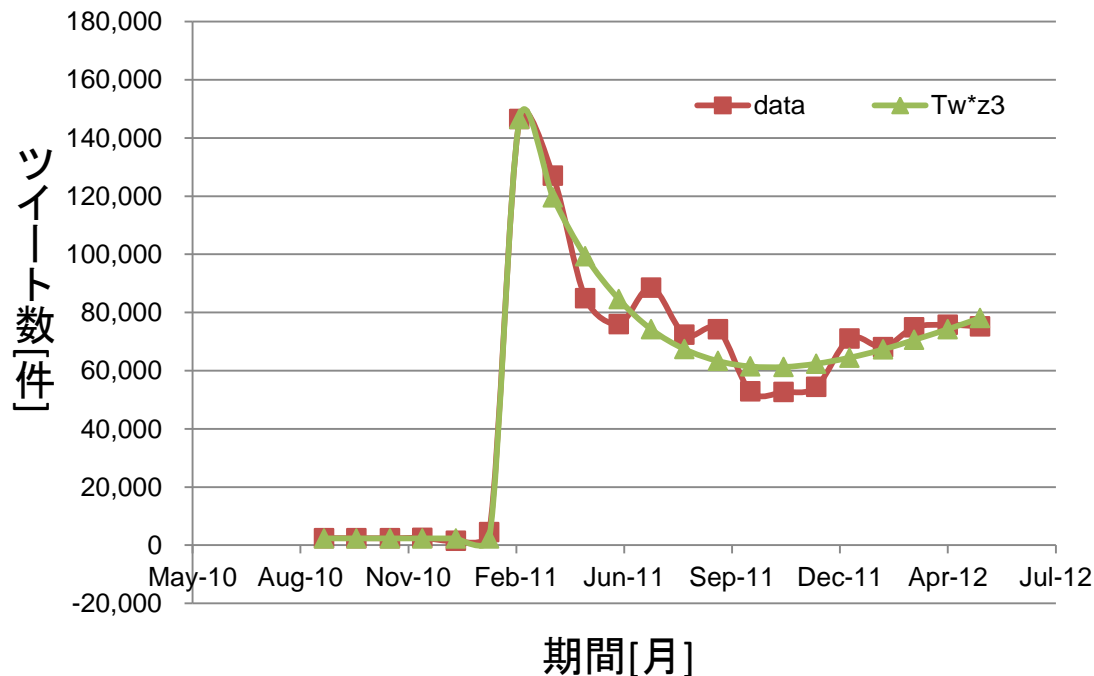
➤ モデルの特徴

- 図9にあるようにツイートの有無で2つの層に分別される。
- ツイート件数から、潜在的な利用者の関心を推定する。
* 関心あり①の状態

◆数値シミュレーション

- 既存モデルと同様の手順で推定し, 比較を行う。
- 推定値を $Tw \times z_3(t)$ とする。
* Tw :1ヶ月当たりの総ツイート数. 今回は $Tw=12$ 億[件]と仮定する. [6]

➤ シミュレーション結果



- 推定値の当て嵌まりの良さは図7の結果以上と言える。
- ツイート件数が再び盛り上がる現象を表現できている。

運用可能なモデルである。

図10:ツイート件数の推移(提案モデル)

➤ 各状態の利用者の推移

- 数値シミュレーションにて導出されたパラメータより求める。

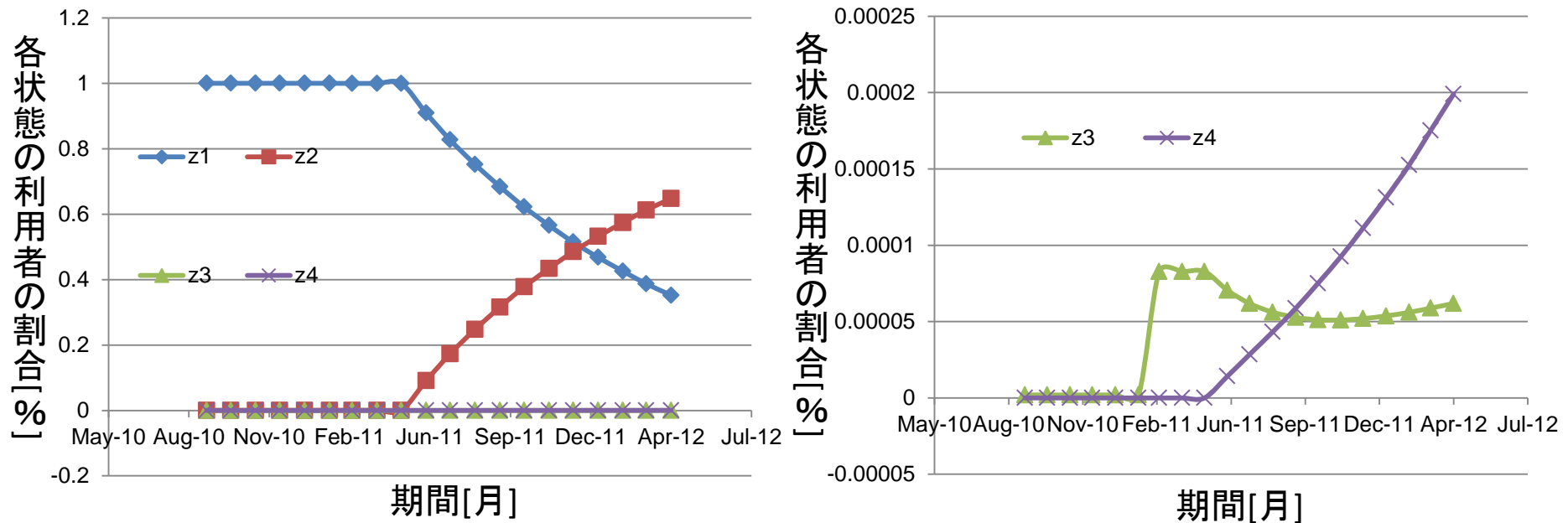


図11:各状態の利用者の推移

- ブーム発生後の“**関心なし**”と“**関心あり①**”の変化は著しく、原発が我々に大きな衝撃を与えたことをよく反映している。
- “**関心あり②**”と“**沈静**”は上記の2つと比べ微小であり、潜在的な原発に対する関心の大きさが改めて伺える。

◆考察

分析結果, 提案モデルより

➤ 仮定(α)は図11より以下のように説明できる.

原発に対する利用者関心はツイート件数では落ち着いている。
しかし、潜在的な関心は高いままである。
状況によっては再びツイート件数に反映され盛り上がりを見せる。

➤ 仮定: 因子分析による解析からツイート件数の推移の形を予測できる。



疑問

- 因子数を増やすことで、予測の精度の向上が見込める?
- ツイート件数のみならず利用者関心の推移も分類できる?

◆まとめ

- 利用者関心をツイート件数と可視化することで、定量的に観測及び推定することができたと言える。
 - 事象に合わせた拡張を行うことで精度の向上が見込める。
 - S-PLUSの因子分析により各データのツイート件数の推移に影響を与える要因が明らかになった。

◆課題

- 用いるデータ(種類・単位時間の設定)を増やした上で再度検証を行う必要がある。
- ツイート件数と利用者関心の推移の関係を明示する。
- 今回得た考察の更なる検証とモデルへの反映の手立て。

参考文献

- [1] GoogleJapanBlog,Google2012年上半期話題となった検索キーワードランキング,
(<http://googlejapan.blogspot.jp/2012/06/google-2012.html>),最終閲覧日2012/9/30.
- [2] ICT総研, SNS利用動向, 広告活用状況に関する調査,
(http://www.ictr.co.jp/topics_20111227.html), 最終履歴2012/6/27.
- [3] 株式会社カカコム2012年3月期決算説明資料,
(<http://corporate.kakaku.com/ir/>), 最終履歴2012/6/27.
- [4] mixi, Twitter, Facebook, Google+, LinkedIn 2012年7月最新ニールセン調査,
(<http://media.loops.net/sekine/2012/08/23/neilsen-netview-201207/>),
最終閲覧2012/8/31.
- [5] 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.
- [6] NECビッグロブ,5月のTwitter利用者の動向,
(<http://www.biglobe.co.jp/pressroom/info/2012/06/120608-1>),最終閲覧日2012/9/30.
- [7] ツイナビ~Twitter使い方ガイド~,
(<http://twinavi.jp/guide/section/twitter/glossary>), 最終閲覧2012/8/31.

Appendix

◆ p6より②ブームが始まった後の解

$$y_1(t) = (1-k)S \cdot e^{-b_1(t-T)} \quad (26)$$

$$y_2(t) = \frac{b_1(1-k)S}{b_1 - b_2 - b_3} \left\{ -e^{b_1(t-T)} + e^{-(b_2+b_3)(t-T)} \right\} \quad (27)$$

$$y_3(t) = \frac{b_2(1-k)S}{b_2 + b_3} + b_2 E \quad (28)$$

$$y_4(t) = \frac{(b_2 k + b_3)S}{b_2 + b_3} + b_3 E \quad (29)$$

$$E = \frac{(1-k)S}{(b_2 + b_3)(b_1 - b_2 - b_3)} \left\{ (b_2 + b_3)e^{b_1(t-T)} - b_1 e^{-(b_2+b_3)(t-T)} \right\}$$

◆ 突発的なブーム時の解

- 図1より(5)~(8)の微分方程式が得られる.

$$y_2(t) = (1-k)S e^{-(b_2+b_3)(t-T)} \quad (30)$$

$$y_3(t) = -\frac{b_2(1-k)S}{b_2+b_3} e^{-(b_2+b_3)(t-T)} + \frac{b_2(1-k)S}{b_2+b_3} \quad (31)$$

$$y_3(t) = -\frac{b_2(1-k)S}{b_2+b_3} e^{-(b_2+b_3)(t-T)} + \frac{(b_3+kb_2)S}{b_2+b_3} \quad (32)$$

◆総利用者数

- 既存モデルのパラメータ推定にて用いる.

$$Y(t) = \int_0^t b \{y_2(t) + y_4(t)\} dt \quad (33)$$

- 標準モデル

$$Y(t) = \frac{\{(b_2 t + b_3 T)k + b_3(t - T)\}bS}{b_2 + b_3} + \frac{(1-k)bS}{b_1 - b_2 - b_3} \left\{ \frac{b_3 - b_1}{b_1} (1 - e^{-b_1(t-T)}) + \frac{b_1 b_2}{(b_2 + b_3)^2} (1 - e^{-(b_2 + b_3)(t-T)}) \right\} \quad (34)$$

- 突発的なブーム

$$Y(t) = \frac{\{(b_2 k + b_3)t - b_3(1-k)T\}bS}{b_2 + b_3} + \frac{b_2(1-k)bS}{(b_2 + b_3)^2} (1 - e^{-(b_2 + b_3)(t-T)}) \quad (35)$$

◆ 提案モデルの解

$$z_1(t) = (1-l-m)e^{-c_1(t-T)} \quad (36)$$

$$z_2(t) = \frac{c_1(1-l-m)}{c_1-c_2-c_3} \left(-e^{-c_1(t-T)} + e^{-(c_2+c_3)(t-T)} \right) + le^{-(c_2+c_3)(t-T)}$$

$$z_3(t) = \frac{c_2}{c_4-c_3-c_2} \left\{ \frac{c_1(1-l-m)}{c_1-c_2-c_3} \left(-e^{-c_1(t-T)} + e^{-(c_2+c_3)(t-T)} \right) + l \left(e^{-(c_2+c_3)(t-T)} - e^{-c_4(t-T)} \right) \right\} \quad (37)$$

$$+ \frac{c_2}{c_4-c_3-c_2} \cdot \frac{c_1(1-l-m)}{c_1-c_4} \left(e^{-c_1(t-T)} - e^{-c_4(t-T)} \right) + me^{-c_4(t-T)}$$

(38)

$$z_4(t) = \frac{c_3}{c_2+c_3} \left\{ \frac{(1-l-m)}{c_1-c_2-c_3} \left((c_2+c_3)e^{-c_1(t-T)} - c_1e^{-(c_2+c_3)(t-T)} \right) - le^{-(c_2+c_3)(t-T)} \right\}$$

$$+ \frac{c_2}{(c_4-c_3-c_2)(c_2+c_3)} \left\{ \frac{c_4(1-l-m)}{c_1-c_2-c_3} \left((c_2+c_3)e^{-c_1(t-T)} - c_1e^{-(c_2+c_3)(t-T)} \right) + l \left(c_4e^{-(c_2+c_3)(t-T)} + (c_2+c_3)e^{-c_4(t-T)} \right) \right\}$$

$$+ \frac{c_2(1-l-m)}{(c_4-c_3-c_2)(c_1-c_4)} \left(-c_4e^{-c_1(t-T)} + c_1e^{-c_4(t-T)} \right) + me^{-c_4(t-T)} \quad (39)$$

$$+ \frac{c_2(1-l-m)}{(c_4-c_3-c_2)} \left(1 + \frac{c_4}{c_2+c_3} \right) + \frac{1}{c_2+c_3} \left(c_3(1-l-m) + l(c_2-c_3) \right) + m$$

◆ 多変量解析 (因子分析) のプログラム

```
factanal(x = ~ ., factors = 2, method = "mle", data = `論文用データ`, scores  
        = TRUE, type = "regression", rotation = "varimax", na.action =  
        na.omit, control = list(iter.max = 20, unique.tol = 0.0001))
```