

2ちゃんねるにおけるコミュニケーションの時系列解析

松村 真宏 *1, 柴内康文 *2, 西村博之 *3, 三浦麻子 *4, 西田豊明 *1

*1 東京大学大学院情報理工学系研究科

*2 同志社大学文学部社会学科

*3 東京プラス株式会社

*4 大阪大学大学院人間科学研究科

連絡先: 〒 113-8656 東京都文京区本郷 7-3-1

東京大学情報理工学系研究科 西田・黒橋研究室

matumura@kc.t.u-tokyo.ac.jp

<http://www.kc.t.u-tokyo.ac.jp/~matumura>

1 はじめに

昨年の MYCOM2002 では, 2ちゃんねるが盛り上がるメカニズムを解明するために, メッセージのサイズや投稿数, 返信率, 投稿間隔などの基本的な属性に加え, 2ちゃんねるに特徴的な2ちゃんねる語やアスキーアート(AA)などの定型的な表現技法に注目した[松村 02]. 本稿ではまず2章で, その後の解析によって明らかになった2ちゃんねるが盛り上がるモデル(2ちゃんねるモデル)[松村 03]について述べ, このモデルの応用として取り組んだ2ちゃんねるスレッドの議論傾向の分類についての考察を3章で述べる. 4章で本稿のテーマである2ちゃんねるにおけるコミュニケーションの時系列解析について述べる.

2 2ちゃんねるが盛り上がるモデル

2ちゃんねるモデルを構築するために用いた指標は, C, A, I, S, V, AA, N, ABON の8つである. C (Contents) はスレッドがどれくらい内容のある議論を繰り広げているかを測る指標であり, 1メッセージあたりのサイズ(バイト数)として求まる. A (Activity) はスレッドがどれくらい活気があるかを測る指標であり, 1スレッドあたりのメッセージの数として算出する. I (Interaction) は参加者同士がインタラクションしている程度を測る指標であり, 1メッセージあたりの平均返信数である. S (Speed) はスレッドの盛り上がる早さを測る指標であり, 1日あたりに投稿されるメッセージの数とする. V (Vocabulary) はスレッドの2ちゃんねるらしさを測る指標であり, スレッドのサイズに占める2ちゃんねる語のサイズの割合として求まる. AA (ASCII Art) はスレッドにアスキーアートがどれくらい使われているかを測る指標であり, スレッドのサイズに占めるアスキーアートのサイズの割合とする. N (Nameless) はメッセージが名無しで投稿されている程度を測る指標であり, 1メッセージあたりの名無しで投稿されるメッセージの数とする. ABON は議論が荒れている程度を測る指標であり, 1スレッドあたりの2ちゃんねるの管理人によって削除されたメッセージの数として求まる.

この8指標を2ちゃんねるの5748スレッドから算出し, 共分散構造分析[狩野 02]によって解析

2ちゃんねるが盛り上がるモデル
 カイ2乗値(自由度)=13.522 (14) p値=.486
 GFI=.910 CFI=1.000 RMSEA=.000

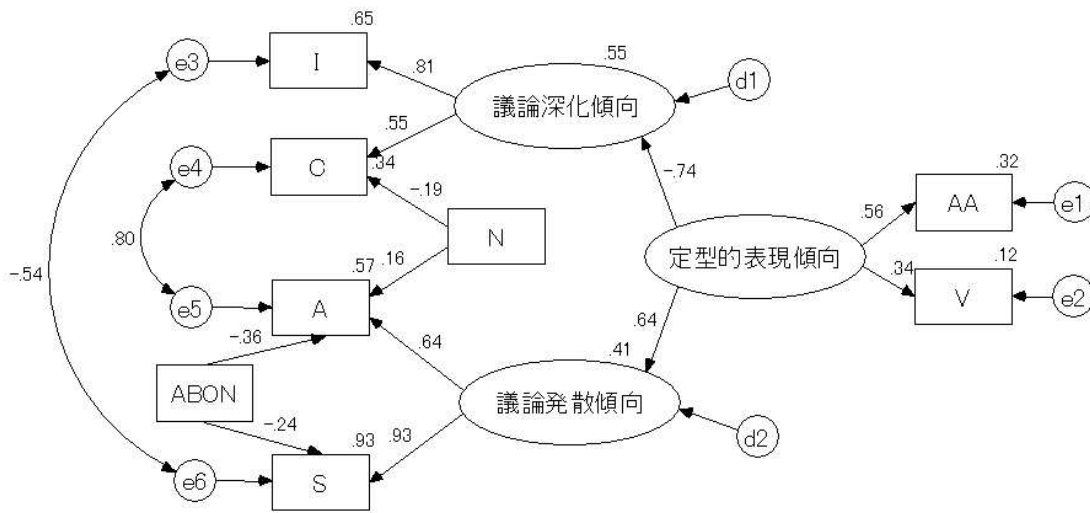


図 1: 2ちゃんねるが盛り上がるモデル

して得られた2ちゃんねるモデルを図1に示す[松村 03]。この2ちゃんねるモデルはさまざまな知見を示唆しており、例えば、2ちゃんねるにおけるコミュニケーションの大きな傾向として、2ちゃんねるの語やアスキーアートを使う定型的表現傾向が強ければ、そこで繰り広げられる議論は発散する傾向になり、逆に定型的表現傾向が弱ければ議論は深まる傾向があることが分かる。また、発散的に議論をしたいときには名無しは有効であるが、議論を深めたいときには名無しが逆効果になっていることも興味深い。これらのことから、2ちゃんねるのユーザは定型的固有表現傾向と名無しをうまく使いわけることで、2ちゃんねるというコミュニティを維持し、多様なコミュニケーションを楽しんでいることが明らかになった。

3 2ちゃんねるスレッドの議論傾向

本章では2ちゃんねるモデルの応用として、2ちゃんねるスレッドを議論深化傾向、議論発散傾向に分類することを考える。まず、2ちゃんねるモデルから得られる議論深化傾向、議論発散傾向が満たすべき条件を整理すると以下ようになる。

議論深化傾向： 議論深化傾向は定型的表現傾向（AA と V）に反比例する。また、I と C（C は N に反比例）は議論深化傾向に比例する。

議論発散傾向： 議論発散傾向は定型的表現傾向に比例する。また、A（A は N に比例し、ABON に反比例）と S（ABON に反比例）にも比例する。

また、指標間の関係の強さは2ちゃんねるモデルに相関係数として表されている。これらの関係から、本稿では議論深化傾向度、議論発散傾向度を式(1)(2)として定式化する。

$$\begin{aligned} \text{議論深化傾向度} &= -0.74 \times (0.56 \times AA + 0.34 \times V) \\ &\quad + 0.91 \times I + 0.55 \times (C - 0.19 \times N) \end{aligned} \quad (1)$$

$$\begin{aligned} \text{議論発散傾向度} &= 0.64 \times (0.56 \times AA + 0.34 \times V) \\ &\quad + 0.64 \times (A + 0.16 \times N - 0.36 \times ABON) \\ &\quad + 0.93 \times (S - 0.24 \times ABON) \end{aligned} \quad (2)$$

式(1)(2)を用いれば、2ちゃんねるのスレッドから議論深化傾向度、議論発散傾向度を数値として求めることができるようになる。たとえば、「AA 長編板 総合雑談スレッド Part2¹」($C: 0.362, A: 0.231, I: 0.402, S: -0.198, V: 1.809, AA: 2.186, N: 0.893, ABON: 0.400$)だと、

$$\text{議論深化傾向度} = -0.888, \text{議論発散傾向度} = 1.04$$

となり、議論深化傾向度より議論発散傾向度の方が大きいので、このスレッドは議論発散傾向だと判断できる。また、「日本の軍国化を憂慮する若者が集うスレ3²」($C: 0.729, A: -0.112, I: 0.732, S: -0.098, V: 0.253, AA: -0.232, N: -0.032, ABON: -0.096$)だと、

$$\text{議論深化傾向度} = 1.102, \text{議論発散傾向度} = -0.150$$

となり、議論深化傾向の強いスレッドであることが分かる。

ここで、2ちゃんねるの5748スレッドの議論発散傾向度、議論深化傾向度を2次元平面にプロットした結果を図2に、2ちゃんねるのカテゴリごとにまとめた結果を図3に示す。図2を見ると、2ちゃんねるのスレッドは全般に議論発散傾向は強いが、議論深化傾向はほぼ均等に散らばっていることがわかる。また、議論発散傾向と議論深化傾向は相反するものではなく、同時に成り立つこともわかる。つまり、2ちゃんねるでは気軽なコミュニケーションが多いが、議論している内容にまで目を向けると、実はしっかりした議論が繰り返されていることも多いのである。

また、図3を見ると、カテゴリごとに特徴的な盛り上がりの傾向を示していることがわかる。2ちゃんねるには、カテゴリごとに議論傾向を誘導するような明文化されたルールは存在しない。それにも関わらず、カテゴリごとに盛り上がりの傾向がこれだけはっきりと現れていることは、議論傾向を支配する秩序が自然発生的に生まれていることを示唆しており、大変興味深い。

4 時系列解析へのアプローチ

3章までは、スレッドごとに算出した8指標に基づいた2ちゃんねるが盛り上がるメカニズムの解明と、その応用として2ちゃんねるスレッドの議論傾向の分類について述べてきた。

ここでさらに2ちゃんねるモデルを利用すれば、スレッドの盛り上がりをコントロールすることへの知見も得られる。例えば、2ちゃんねるモデルよりABONが盛り上がりにも負の要因として働くことが明らかになったので、議論発散傾向の強いスレッドを鎮めたいときには、下手に議論を正そうとすると盛り上がり、逆にもっと荒らせば沈静化することがわかる。また、定型的固有表現傾向がほとんど見られなければ、議論深化傾向に傾きやすくなっているため、他の参加者とインタラクションしながら内容のあるメッセージをハンドルネームなど情報発信者が識別できる状態で投稿すれば、まじめな議論になることがわかる。

¹<http://aa.2ch.net/aastory/kako/1018/10185/1018507451.html>

²<http://academy.2ch.net/kokusai/kako/1022/10220/1022086171.html>

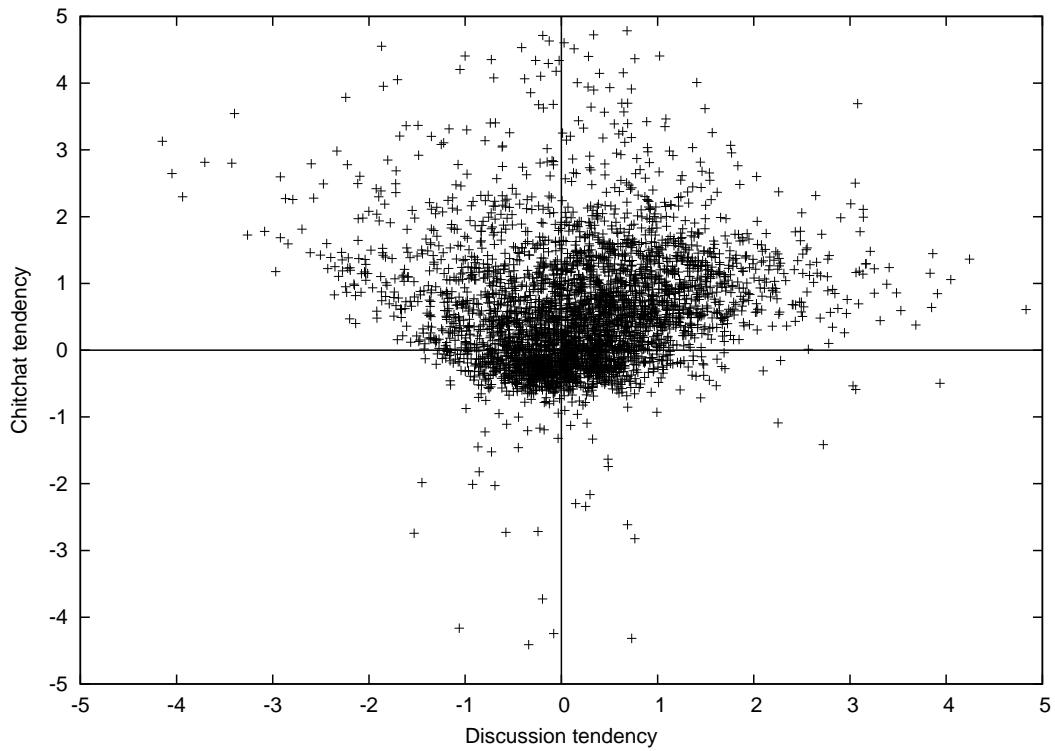


図 2: 2ちゃんねるのスレッドごとの議論深化傾向と議論発散傾向の分布

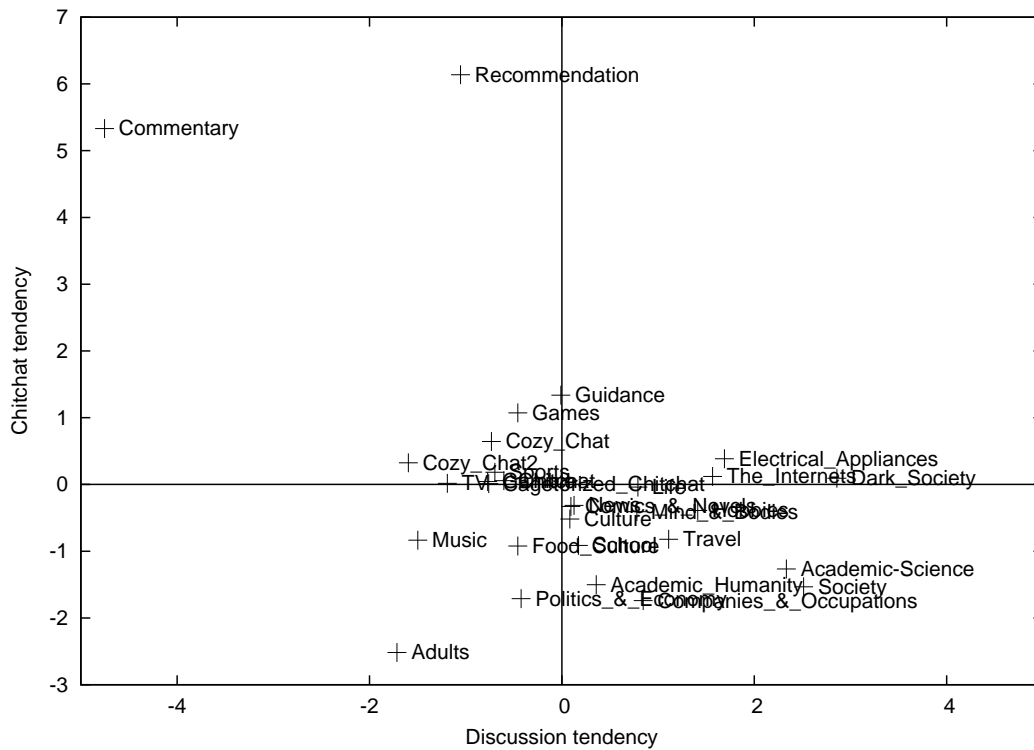


図 3: 2ちゃんねるのカテゴリごとの議論深化傾向と議論発散傾向の分布

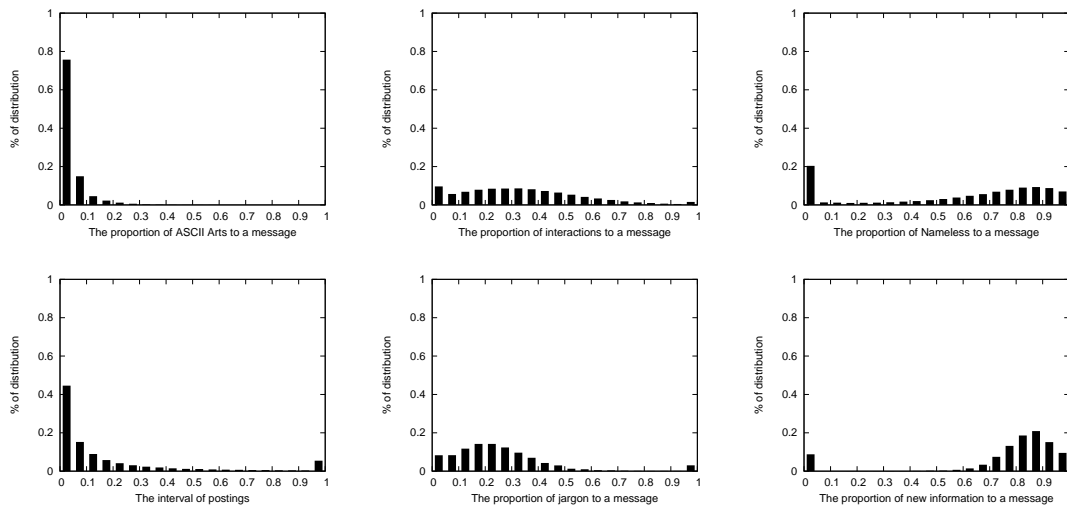


図 4: 3840 スレッドにおける 6 指標の分布

このように、外部から何らかのアクションを与えることでスレッドの盛り上がりをコントロールできる可能性がある。しかし、スレッドの盛り上がりは他の様々な要因との相互作用によってダイナミックに変化しているため、アクションは与えるタイミングも重要になるのであるが、従来の 2 ちゃんねるモデルからはそのタイミングまでは分からない。そこで本章では、このような盛り上がりの時系列変化を支配するメカニズムを捉えるために、2 章で用いた指標を見直し、新たな指標も加えた 6 指標を考える。それぞれの指標の意味と算出方法は以下の通りである。

投稿間隔率 (Time)：投稿が活発なのかゆっくりしているかは、盛り上がりの勢いを表す重要な要因である。これは直前の投稿からの経過時間を見れば簡単にわかるので、その時間を 1 日で正規化した値を投稿間隔率とする。

インタラクション率 (Interaction)：参加者間で起こっているインタラクションは、メッセージの盛り上がりに関与する要因であると考えられる。2 ちゃんねるでは、あるメッセージへの返信を表す記号として引用符“>>”が用いられるので、その割合を調べることでインタラクション率を求めることができる。

名無し率 (Nameless)：メッセージを投稿する際にハンドルネームを用いるか否かはメッセージの識別性に関わり、ひいては盛り上がりの傾向に影響を及ぼす重要な要因の一つである。名無し率は「名無しさんリスト³」に登録されているか否かを調べれば簡単に求まる。

2 ちゃんねる語率 (Jargon)：2 ちゃんねるの参加者にとっては、2 ちゃんねる語はコミュニケーションを促進させる重要な要因である。そこで、1 メッセージあたりに含まれる 2 ちゃんねる語の割合を 2 ちゃんねる語率として求める。本稿では、2 ちゃんねる語の辞書 [2 典 02] を組み込んだ形態素解析システム JUMAN [黒橋 98] を用いて抽出されるちゃんねる語と、辞書に登録されていないカタカナ語を併せて 2 ちゃんねる語とする。

アスキーアート率 (ASCII Art)：円滑なコミュニケーションを行うためには、ユーモア溢れるアスキーアートもまた重要な要因であると考えられる。アスキーアートには、通常のメッセージでは使われない‘多’, ‘又’, ‘川’などの文字がよく使われるので、それらの文字が 1 メッセー

³<http://logo.s3.xrea.com/>

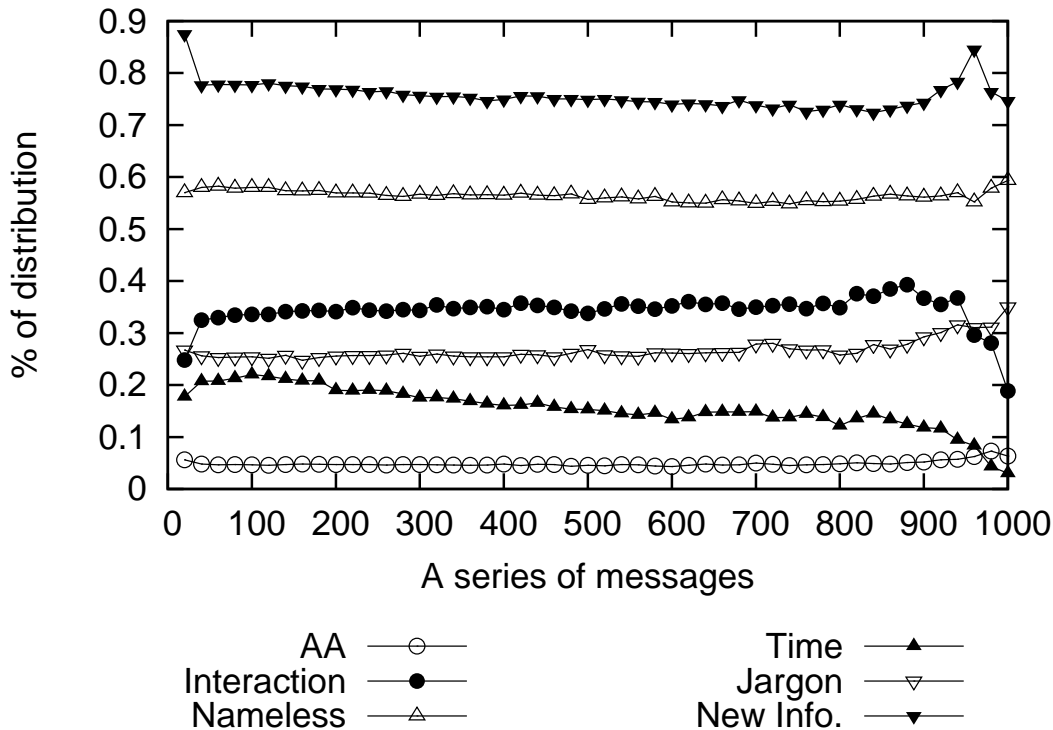


図 5: 3840 スレッドにおける 6 指標の時系列変化

ジあたりに含まれる AA 文字の割合を調べることでアスキーアート率は求まる。なお本稿では、膨大なアスキーアートを公開している Web サイト⁴ を利用してアスキーアートによく使われる文字 (AA 文字) を 77 文字取り出した。

新情報率 (New Info.): 2ちゃんねるにおける 1 メッセージの平均文字数 (アスキーアートは除く) は約 73 文字 (146.3 バイト) である⁵。しかし、コミュニケーションにおいて問題なのは文字数ではなくそのメッセージがもたらす内容であろう。そこで、新しく出てきた名詞が含まれる割合を新情報率として求める。

2ちゃんねるの 3840 スレッド⁶から 20 メッセージごとに算出した 6 指標の分布は図 4 のようになる。ここでは、それぞれの指標の値を 0.5 刻みで 20 等分しており、例えば AA_1 は AA 値が 0~0.05, AA_{20} は AA 値が 0.95~1 であることを意味している。図 4 を見れば明らかなように、各指標はそれぞれ特徴的な分布をしている。この分布が他のオンラインコミュニティとどのように異なるのかも興味深いのが、我々が注目したいのは時系列における指標の変化である。

そこで次に、スレッドの盛り上がりの時系列変化を捉えるために、2ちゃんねるの 3840 スレッドの 20 投稿ごとに算出した 6 指標の時系列変化を図 5 に示す。図 5 を見ると、投稿が長く続いているスレッドには AA は 0.05 辺り、Interaction は 0.3~0.4、Nameless は 0.55~0.6、Jargon は 0.2~0.3、New Info. は 0.7~0.8 といった、ほぼ安定した指標を維持しており、これらが投稿が長く続く心地よいリズムを生み出しているのであろう。なお、Time は 0.2~0.05 と投稿が続くにつれて徐々に投稿のペースが速くなっているが、これは投稿が長く続いているときには投稿者が徐々にのめり込んでいるからだと考えられる。また、それまで安定していた指数が投稿数が 900 を越えた

⁴ 「AA 大辞典 (仮)」 <http://maruheso.tripod.co.jp/aadic/>

⁵ 2002 年 6 月にダウンロードした 2ちゃんねるの 5746 スレッドから算出した。

⁶ いずれも投稿数が 100 以上のスレッドであり、1 スレッドあたりの平均投稿数は 290.6 であった。

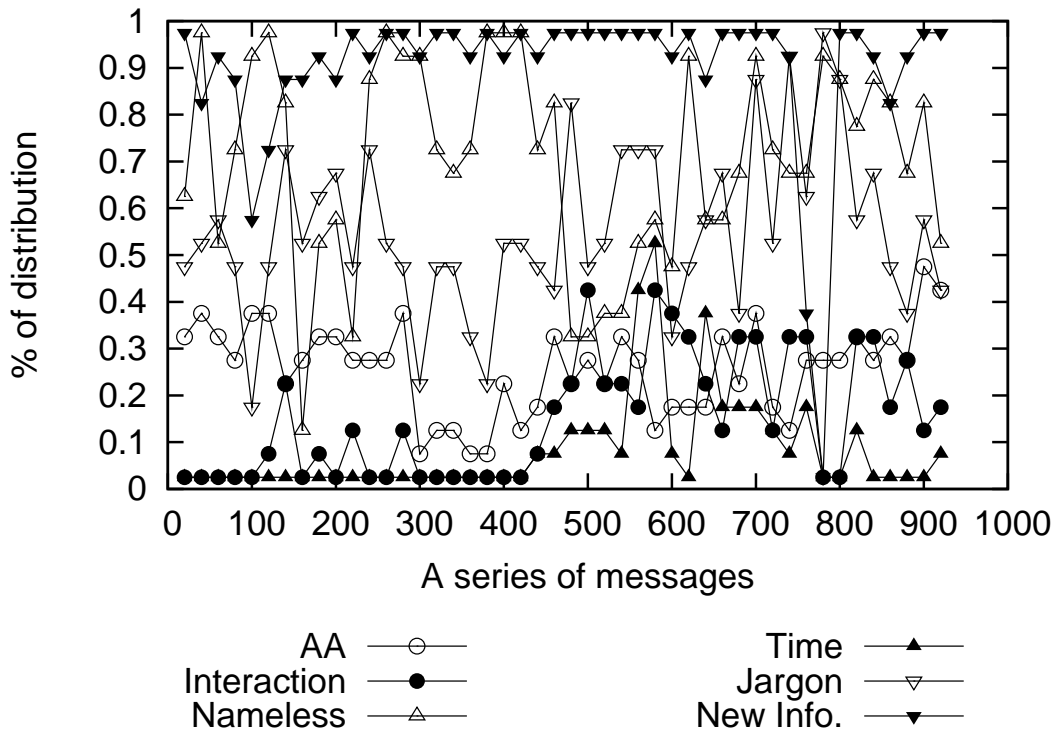


図 6: あるスレッドにおける 6 指標の時系列変化

辺りから急に变化し、インタラクション率は下がり投稿のペースが上がっている。これは 1 スレッドにつき最大 1000 投稿という 2 ちゃんねるの制約のためであり、投稿数が 1000 に近づくと新しいスレッドに移行する話題や、誰が 1000 番目の投稿をするかといった話題で盛り上がっているからだと考えられる。

5 まとめと今後の展望

本稿ではまず昨年 MYCOM2002 で発表した研究のその後の成果について紹介し、その応用として 2 ちゃんねるの議論傾向の分類をする手法を提案した。その結果、2 ちゃんねるでは全体的には議論発散傾向が強いが、そういうスレッドの中にも実は議論深化傾向が強いスレッドも数多く混ざっていることが明らかになった。ここで提案している議論傾向を指標化するアイデアは、オンラインコミュニティにおける盛り上がりのモニタリングなどへの応用が考えられる。

また、時系列におけるスレッドの盛り上がりの傾向を分析した結果、投稿が長く続くスレッドでは指標がほぼ決まった範囲内で安定していることが明らかになった。参加者の居心地がよくないとスレッドから去ってしまうことを考えると、3840 スレッドから明らかになった 6 指標の値は、心地よく盛り上がるためのある種のリズムを表しているのかもしれない。

今回の時系列解析から得られた知見は、2 ちゃんねるの 3840 スレッド全体の傾向から得られたマクロなものである。しかし、例えばあるスレッドにおける 6 指標は図 6 に示すようにもっとダイナミックに変化している。そこで今後は、個々のスレッドのダイナミックな変化の中からその変化を支配するメカニズムを解明することを考えている。

参考文献

- [松村 02] 松村真宏, 中村洋, 大澤幸生, 石塚満: なぜ2ちゃんねるは盛り上がるのか?, 第3回人工知能学会 MYCOM, 2002.
- [松村 03] 松村真宏, 三浦麻子, 柴内康文, 大澤幸生: 2ちゃんねるが盛り上がるメカニズムの解明, 第51回人工知能基礎論研究会 SIGFAI51
- [2典 02] 2典プロジェクト, 2典~2ちゃんねる辞典~, バーチャルクラスター (2002).
(<http://freezone.kakiko.com/jiten/> と同じ内容である)
- [狩野 02] 狩野裕, 三浦麻子: グラフィカル多変量解析, 現代数学社, 2002.
- [黒橋 98] 日本語形態素解析システム JUMAN 使用説明書, 1998.