社団法人 情報処理学会 研究報告　　2008－MPS－71 (13)
IPSJ SIG Technical Report　　2008／9／19

ウェブ掲示板における投稿数次数分布の調査と分布関数の導出

成瀬 継太郎 [††]， 久保 正男 [†]， 佐藤 浩 [†]， 松原 隆 [†]

[††] 会津大学　[†] 防衛大学校

本論文の目的は、電子掲示版システム（BBS）におけるコミュニケーションの様相を理解することである．特に本論では次の三点について議論する．(1) 調査：オープン型の BBS のログを分析し、多くの BBS に共通する特性を調べた結果，各々のユーザーによって投稿される一日あたりの記事の数が対数正規分布に従うことを明らかにする．これらの特徴は各ユーザーに明示的にそのように振る舞うように定められたものではなく，自由な相互作用の結果現れた創発現象がもつ性質である．そこで，この現象を理解するために (2) 創発現象を起こす個人特性の提示と (3) 創発現象下のユーザーの振る舞いの定式化を行い，投稿件数分布を導出している．

Understanding Emergent Behaviour of Posting Activities on Web Bulletin Board System

Keitaro Naruse [††]　Masao Kubo [†]　Hiroshi Sato [†]　Takashi Matsubara [†]

[††] University of Aizu　[†] National Defense Academy

The objective of this paper is to understand an aspect of human social interaction in bulletin board systems on internet. When an individual submits an article to a BBS, it is potentially influenced by articles from other users. A submission sometimes starts a long and hot chain of articles, but often does not. This paper tries to answer the question of why and how such a chain of articles emerges. In other words, we attempt to reveal a mechanism linking the individual voluntary activity of article submission and the social phenomenon of a long article chain.

1　はじめに

本論文では，まず，巨大匿名掲示板を対象とし，各掲示板に共通の特徴を調査する．その結果，ほぼすべての掲示板において，一人当たりの投稿数とその頻度が対数正規分布になるという特徴があることを示す．次に，この性質がなぜ発生するのかを明らかにするために，十分現実的であると思われる簡単な仮定から出発し，記事の投稿とその際の個人間の相互作用をエージェント (動機モデルと呼ぶ) として定式化する．次にこれを分析して，ある条件下で動機モデルの結果と現実の対数正規分布特性がよく一致することを示す．これにより返報性がこの創発現象を引き起こす因子の一つであることを示す．この結果に基づき，投稿ダイナミクスを BA モデルの一種として捉え，定式化する．初期条件の違いなど投稿過程に適するよう変更を行い，最終的に導出した投稿数分布と実際の投稿数分布の形状が合致することを示す．

2　掲示板における投稿数分布の調査

2.1　対象掲示板

議論の一貫性を保つために，対象を大量のデータを含む単一の掲示板に絞るものとし，投稿の容易さという観点から２ちゃんねる [1] というオープン型の巨大匿名掲示板群を対象とする．

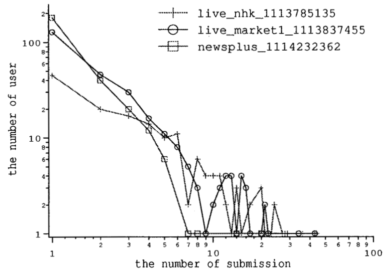

Fig. 1　各投稿者当たりの記事数の分布．

2.2　スレッドに共通性質 (1 ユーザー当たりの投稿件数分布)

以下では掲示板コミュニティの特徴を，記事の投稿者について統計をとった．その結果，投稿件数には次のような規則性がみられた．

Fig1 は，異なる三つのスレッドにおいて，各投稿者がそのスレッドに何件の記事を書き込んだかを示して

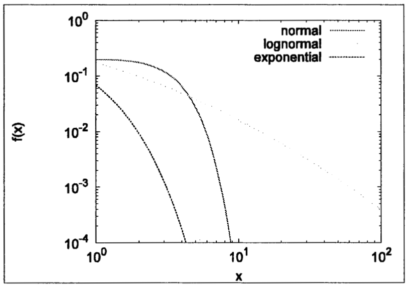

Fig. 2　両対数グラフにおける正規分布，対数正規分布，指数分布

Table 1　AIC の算出結果

Model	Normal	Exponential	Lognormal
# of parameter	2	1	2
Avg. AIC	961.4	723.7	603.7
Min. AIC	103.1	80.1	48.9
Max. AIC	2283.3	1824.7	1316.6

いる．横軸に一人当たりの投稿記事数（何件投稿した人か），縦軸に頻度（何人いたか）を両対数スケールでプロットしている．$(x, y) = (1, 100)$ は，1 件だけ記事を投稿した人が 100 人 いたことを表している．これらは分野も内容も大きく異なる三つのスレッドであるが，投稿回数と該当者の関係は，ほぼ同じように直線的に減少している．この傾向は，他の大部分のスレッドでも同様に観察された．

2.3　赤池の情報量基準 AIC による評価

次に，これらの実際の分布がどのような統計モデル (分布) により良く表現されるかを AIC (赤池の情報量基準) により検討する．統計的分布には様々なものがあるが，本論文では，正規分布，指数分布，対数正規分布を代表的なものとして考察の対象にした (Fig.2 参照) 冪乗則を示す分布では，パレート分布が有名であるが，ここでは対数正規分布でそれらを代表させるものとする．その理由は，対数正規分布は直線だけでなく放物線状も表現できることと，確率変数の定義域が 0 以上の領域であり取り扱いが容易であるためである (より詳細には，確率変数が 0 の時の確率密度を参照するが，パレート分布の定義域は 1 以上であるため，それができないためである)．データの取得は，2005 年 6 月から 8 月の期間に行った (総スレッド数 327,282)．対象としたデータは 2 ちゃんねるのすべてのカテゴリの中から，1 日あたりの記事数が十分多い（400 件以上）スレッド，計 584 件に限定した．

Table.1 には各分布における AIC の平均、最小値、最大値を示した．AIC は小さいものほど良いとする基準である．対数正規分布に対する AIC はどの値についても他より小さいことがわかる．したがって，実際の掲示板における一人当たりの投稿数の分布は，対数正規分布により最も良く表現されるといえる．

以上から，オープン型の BBS における一日という比較してリアルタイムに近い投稿活動を計測した結果，その投稿件数分布が対数正規分布に従うというマクロ特徴があることを明らかにした．この分布は Fig.2 で示したように両対数スケール上で直線状になるという冪分布と同様の性質がある．

3　エージェントモデルによる集団特性の再現

本節では，この分布特性がどのような個々人の意図から発現するのか調査するために、現実的な 4 つの仮定に基づく個人投稿モデルでこの特性を再現する．その結果，マクロ現象を生み出す要因の一つが返報性であることを示す．

3.1　投稿における仮定

以下の四つを仮定して agent モデルを構築した．

(A) 独立性: 記事を投稿する際には個人の意思のみが働くものとし，他者や投稿を強制されることは一切ないものとする．また，他のメディアによる情報交換は一切ないものとする．

(B) 非決定的優先性: 個人は興味などに応じて記事を投稿する動機を持つものとする．そして高い動機を持つと高い頻度で記事を投稿するものとする．

(C) 時間減衰性: 個人の動機は時間とともに減少するものとする．

(D) 返報性: 個人は掲示板に情報収集あるいは議論など何らかのコミュニケーションを求めて記事を投稿するものとする．そのため自分の投稿した記事への返信などの反応があった場合，記事を投稿する頻度が高くなるものとする．

これらの仮定を数学的に表現するため，本論文では動機値という値を導入する．動機値が高いときはより高い確率で記事を投稿（同 B）するものとする．他者からの影響は返報性によるもののみとし，これにより動機値が増加（同 D）するものとする．

3.2　Agent モデル

まず，投稿者 u_i の時刻 t での動機値を $m_i(t)(\geq 0)$ とする．このとき i が記事を投稿するかは，次の確率により決定されるものとする．

$$p_i(t) = \frac{\exp(m_i(t)/\tau)}{\sum_j^N \exp(m_j(t)/\tau)}. \qquad (1)$$

ここで，N はコミュニティにおける個人の数，τ は確率的選択におけるランダム性を決定するパラメータである．この式では記事を投稿する/しないが，他のメン

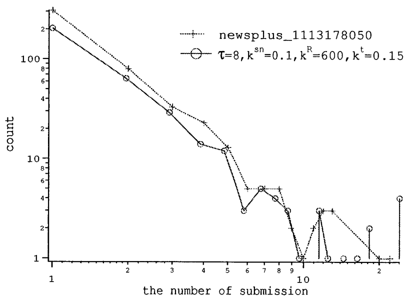

Fig. 3　実際の掲示板における投稿数分布の再現.

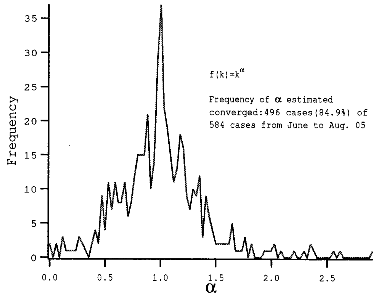

Fig. 4　返報性の確認

バの動機値を観測しながら競合的に行われおり，前述の独立性の仮定に矛盾するように見えるかもしれない．しかし，これは単位時間における投稿数が最大１件という制約からくるものである．実際の掲示板システムにおいても，サーバの動作原理上，ある瞬間の書き込みは１件のみである．

動機値は次の式にように変化するものとする．

$$m_i(t) = k_i^t m_i(t-1) + event_i(t) \tag{2}$$

$$event_i(t) = \begin{cases} k_i^S & \text{when it posts} \\ k_i^R & \text{when it receives a reply} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

ここで，$k_i^t(0 \leq k_i^t \leq 1)$ は動機値の自然減衰係数，$k_i^S(\leq 0)$ は記事を投稿したときの動機値の変化量，$k_i^R(\geq 0)$ は自分の記事に返信がついたときの動機値の変化量である．この動機モデルの特性を理解するために前実験を行った．すべてのパラメータは投稿行動のランダムさの制御に影響を及ぼす．ランダム性が強い場合には、このモデルの投稿回数分布は両対数スケール上で放物線状になる．一方、ランダム性が弱い場合には直線形状に近づいた．さらに、これは実際のデータによく一致した．例として Fig3 を示す．これはうまくフィットするように手作業でパラメータを調整した例である．

本章での議論から，この投稿件数が対数正規分布に従うという性質を発現する要因の一つが，このシミュレーションで用いた A から D までの条件に強い関わりがあることがわかった．特に，返報性は掲示板では以前に投稿した記事は常にだれでも閲覧できるので,「沢山投稿したものほど新たにより多くの投稿をする」という性質を生み出すことが予想される．紙面の制限のため割愛するが，事実このシミュレーションをある条件の元で近似すると，投稿件数が対数正規分布特性を持つ事を示す事ができる．この特性は，バラバシらが提案した優先的選択,preferential attachment(以下 PA)[2] に類似している．事実，投稿件数分布のと PA がもたらす冪分布は値域によっては非常に似た形状になる．そこで，次章では，この優先的選択特性について改めて実証実験を行った後，この BA モデル [2] を元にこの投稿件数に関する集団現象の特性を定式化する．

4　返報性に基づく投稿件数にみられる創発的特徴の理解

まず，優先的選択過程の有無を検証する．ここでは投稿回数変化に着目して検証を行う [3]．

バラバシの優先的選択の考え方を掲示板に応用すると，ある利用者の投稿数の増加率 $\Pi(k)$ はそれまでの投稿数 k に比例して増加することになる．この手法はこれを一般化し

$$\Pi(k) = \frac{k^\alpha}{\Sigma_{k'} k'^\alpha} \tag{4}$$

と仮定し，実際のデータから α を求める．これによって優先的選択の有無が明らかになる．

Fig4 に，この推定手法を前述の 584 スレッドに適用した結果を示す．α が 1 の時には BA モデル [2] でのスケールフリーに一致する．84.9%(496 件) のスレッドについて収束し α を推定することができた．そのときの α の分布を示している．$\alpha = 1$ に高いピークが見られ，返報性の仮定が妥当なものと判断できる．以上から，投稿過程が返報性に基づくもので，またそれを BA モデル [2] で表現してもよいと判断した．

4.1　投稿過程モデルの提案

バラバシらの方法とここで扱う掲示板では次のような違いがある．まず，バラバシらの方法ではノードを追加する度にリンクを増やしてゆくが，(1) 初期ノード数が極めて少ない場合を想定，(2) ノード数が生成中に増加する，という特徴がある．これに対し，ここではノードをコミュニティの各メンバと捉えるので，(1) 初期ノード数は無視できない，(2) 一日という短い時間間隔なのでこの間に利用者の総数にほとんど変化

がないものとする．これを踏まえ，本論では次式の投稿モデルを提案する．

一単位時間に一件の投稿があるものと仮定し、時刻 t での総投稿数を t とする．メンバ i の投稿数を $n_i(t)$ とする．また，コミュニティの総数を N とする．この時、投稿者 i が時刻 t に掲示板に投稿する確率 $\Pi(i)$ を次式で与える

$$\Pi(i) = \frac{wn_i + 1}{\sum_{j=1}^{N}(wn_j + 1)} \tag{5}$$

w は一回の投稿が投稿確率に与える重みであり $w \geq 0$ とする．投稿数が多い投稿者は高い確率で再投稿し，その確率は投稿数 n_i に比例する．また一度も投稿したことがないメンバ (n_i=0) も，低い確率であるが投稿する可能性がある．w=0 の時は各メンバの投稿確率は等しくなり，投稿はランダムに行われることを意味する．

式 5 を展開すると、投稿数の単位時間あたりの変化は次の式のように表せる．

$$\frac{\partial n_i}{\partial t} = \Pi(i) = \frac{wn_i + 1}{wt + N} = \frac{n_i + 1/w}{t + N/w} \tag{6}$$

この式を解くと，最終的に

$$n_i(t) = (1 + \frac{1}{w})\frac{t + \frac{N}{w}}{t_i + \frac{N}{w}} - \frac{1}{w} \tag{7}$$

となる．ここで t_i はメンバ i が初めて投稿する時刻である．次に，メンバ i が時刻 t_i で初めて投稿する確率 P_{first} を求める．これは時刻 1 から $t_i - 1$ まで投稿しないで，時刻 t_i で初めて投稿する確率である．従って，

$$P_{first}(t_i) = (\prod_{t=0}^{t_i-1}(1 - \frac{1}{wt + N}))\frac{1}{wt_i + N} \tag{8}$$

となる．
さて、投稿回数の分布関数 $P(n_i(t) < k)$ は式 7 より次のように書ける．

$$P(n_i(t) < k) = 1 - Prob(t_i \leq ((t + \frac{N}{w})\frac{1 + \frac{1}{w}}{k + \frac{1}{w}} - \frac{N}{w})) \tag{9}$$

この式の右辺第二項は t_i がどのような分布をしているかに関する確率密度関数であり，あるメンバ数 N とある時刻 t のもとでその時刻での投稿がそのメンバにとって初めての投稿である確率である．したがって，$P_{first}(t_i)$ にほかならない．そこで，これを上の式に代入して

$$= 1 - \int_0^{(t+\frac{N}{w})\frac{1+\frac{1}{w}}{k+\frac{1}{w}} - \frac{N}{w}} P_{first}(t)dt \tag{10}$$

を得る．その結果，投稿件数分布 $P(k)$ は

$$= \frac{\Gamma(\frac{(1+w)(N+tw)}{w(1+kw)} - \frac{1}{w})}{(1+kw)\Gamma(\frac{(1+w)(N+tw)}{w(1+kw)})} \tag{11}$$

となる．

参考までにいくつかの w=2,3,4 に対して確率分布 $P(k)$ を両対数でプロットした結果を Fig5 に示す．横軸に投稿回数，縦軸に該当者の頻度を示している．N = 1000, t = 1000 とした．対数正規分布と同様ほぼ直線であり定性的に矛盾しない．

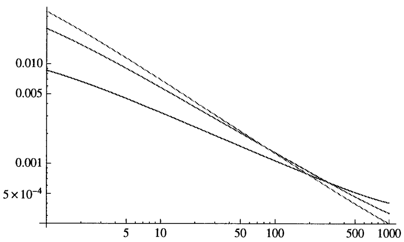

Fig. 5　提案モデルの投稿件数分布例 (式 11)

以上，投稿活動における集団的特性について明らかにした．

5　まとめ

本論文の目的は掲示板コミュニティにおいて，集団の特徴はどこに現れるのか，またこれをもたらす個人の間の相互作用を明らかにすることであった．

この目的に対して，以下の 3 項目を行った．(1) 実際の掲示板のデータを解析し，各投稿者当たりの記事投稿数の頻度は対数正規分布状になることを明らかにした．(2) このよう集団特性が現れる仕組みを明らかにする為に返報性に基づく簡単なモデルを構築した．その結果，実際と同様の特性をもつことがわかった．(3) 返報性は優先的選択ダイナミクスをもたらすことを予想し，改めてユーザーの投稿順序に着目した実証実験をおこなって，実際の掲示板でも「過去に多くの投稿をおこなったものほど投稿しやすい」傾向があることを示した．最後に，初期条件の違いなど投稿過程に適するように投稿数分布を BA モデルから導出し，実際の投稿数分布の形状が定性的に合致することを示した．

参考文献

1) 2 ちゃんねる監修, 2 ちゃんねる公式ガイド 2004, pp. 20–24, コアマガジン (2004)

2) Barabási, A.-L., Albert, R., and Jeong, H., Mean-field theory for scale-free random networks, Physica A, vol. 272, pp. 173–187 (1999)

3) Massen C.P. and Doye J. P.K. :A self-consistent approach to measure preferential attachment in networks and its application to an inherent structure network,Physica A: Statistical Mechanics and its Applications, Volume 377, Issue 1, 1,pp351-362(2007).