

ウェブにおける生命性の指標について

Vitality Indexes in the Web Services

岡瑞起 *1
Mizuki Oka池上高志 *2
Takashi Ikegami*1 筑波大学大学院システム情報工学研究科
Department of Computer Science, University of Tsukuba*2 東京大学大学院総合文化研究科
Graduate School of Arts and Sciences, The University of Tokyo

Vitality index is proposed to measure the life-like activity in the web systems. The vitality of the web system is driven by its number of users or the number of submissions, but they cannot simply become the causes of vitality. From the artificial life perspectives, we measure the network vitality by using the Hawkes Process and by revealing a network structure. A sudden rise of network activity and the change of the underlying network structure is identified in this study. Together with the idea of biological adaptation and homeostasis, we argue the possible vitality index of web service and of other human community activities.

1. はじめに

一細胞、一生物個体、社会性昆虫の集団、人間のコミュニティーに至るさまざまなスケールに渡って、生物システムにはその物質的な構成要素では決まらない性質が創発する。例えばそれはシステムの「元気さ」ではないか。元気な細胞は「ホメオスタティックな」意味で状態であり、また適応的でもあるだろう。人間のコミュニティーにも通じるもののように考えられる。しかし、細胞の構成要素である、ミトコンドリアやゴルジ体などにその元気さの起源を見つけるのは難しいだろう。あるいは、コミュニティーのひとりひとりの性格や行動が、そのままコミュニティーの元気さを決定するわけではないだろう。元気さは、もっとダイナミックなものである。

逆に、非生命システムではどうか。たとえばウェブサービスにも「元気なもの」とそうではないものが有るように思われる。実際、元気なウェブサービスは参加者の数も増えるし、投稿数も増すだろう。しかし、ウェブサービスの元気さはそれだけではない。ここでは、時系列解析として Hawkes Process [Hawkes 71] を用いてその「元気さ」を議論し、背後のネットワークのダイナミクスにそのメカニズムを探る。

2. 解析

2.1 データ : RoomClip

以下では、RoomClip *1 というウェブサービスをもとに解析する。RoomClip は写真共有サイトであり、特に部屋やインテリア写真の投稿サイトである。ひとつの投稿写真には、初期にサービス側によって用意されたいくつかのタグがあり、それとユーザー固有のタグをいくつでも投稿に含めることができる。また、ユーザーはどの人の投稿を自分のタイムラインで閲覧できるようにするか、を決定することができる。これをフォロー機能という。また、その投稿が気に入ればユーザーはその投稿にライク (like) することができる。ライクした投稿は、ユーザのタイムラインで共有される。本研究では、2012 年から 2015 年までの約 3 年分の投稿写真と投稿写真に付けられたタグデータを使用し、背後のユーザー間ネットワークの構造とともに、その時系列的な変化を解析する。

連絡先: mizuki@cs.tsukuba.ac.jp

*1 株式会社 Tunnel が運営するインテリア写真共有ソーシャルネットワークサービス。

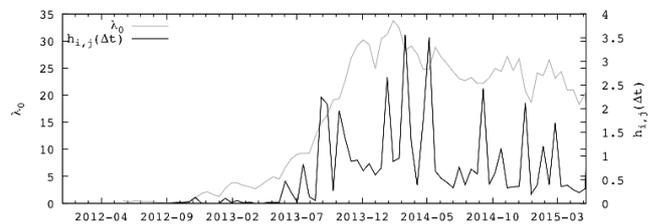


図 1: Hawkes Process による解析結果。点線はデフォルトのポアソン過程による効果を表し、実線は投稿イベント間の相互作用の効果を表す。

2.2 Hawkes Process によるイベント相互作用効果の解析

RoomClip の投稿の時系列をもとに、Hawkes Process [Hawkes 71] で解析した結果を図 1 に示す。Hawkes Process は、デフォルトで投稿イベントが互いに独立なポアソン確率であると仮定し、そこからズレとして投稿イベント間の相互作用の影響を見ていこうとするものである。図では、薄い点線がデフォルトのポアソン過程による効果を、実線が投稿イベント間の相互作用の効果を表している。解析においては、Hawkes Process のパラメータを時系列に対して対数尤度関数を最大化するようにフィッティングし、その値を使って図にプロットしている。

図 1 をみると 2013 年の 7 月過ぎ頃から、急激に投稿イベントのデフォルト発生件数と、投稿イベント間の相互作用の効果が間欠的に出現しているのを見て取ることが出来る。これはこの時期を境に、投稿イベントの数も増加し、かつ投稿イベント間に相互フィードバックが生まれ、ウェブサービスの実効的な「雰囲気」が変化したのではないかと考えられる。ユーザーの数はこの時期に必ずしも急激に増えたわけではないので、この上昇は、登録したユーザーの数では決まらない。なにか、このサービスに参加するユーザーのネットワークに生まれた雰囲気があるのではないか。この雰囲気を、元気さの指標の候補とし、次節でユーザー間のネットワークを計算する。

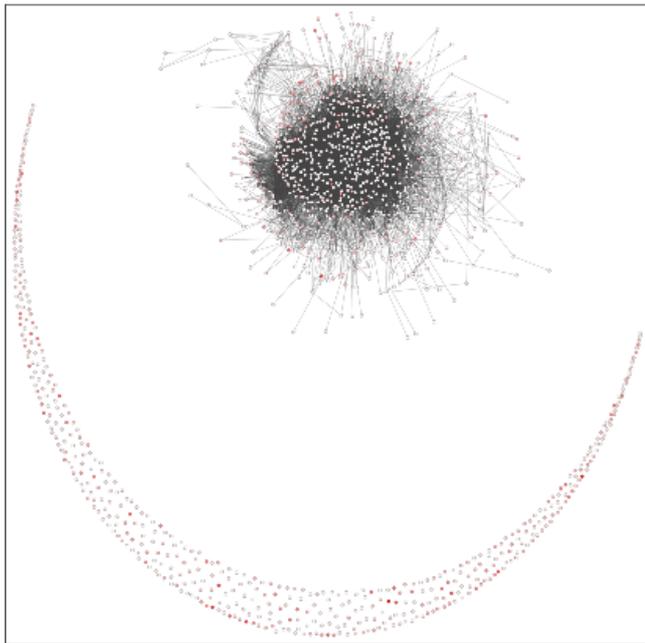


図 2: Jensen Shannon Divergence によるユーザー同士のネットワーク [小堀 16]

2.3 ネットワーク解析

ユーザー同士の距離を Jensen Shannon Divergence (JSD) により計測し、ネットワークを図 2 に示す。JSD とは、ふたつの分布の間の近さ Kullback-Leibler 発散量を比較する 2 つの分布に関して対称になるように一般化し計算するものである。ここではデータセットとして 100 回以上投稿を行っているユーザー 1,493 人それぞれについてタグ使用分布ベクトルを求め、このベクトル同士の距離を計算し、JSD が閾値以下のユーザー同士を類似したユーザーとみなし、ユーザーをノードとしてそのノード間にエッジを張ってネットワークを構築した。ここで JSD の閾値は、エッジ数が完全グラフの場合の 1% である 22,290 に近くなるように調整されている [小堀 16]。

図 2 をみると、ネットワークは典型的にコア（一つのノードから他のノードに強く繋がれたノード群）-ペリフェラル（つなぐ相手が 1,2 個の孤立したノード群）構造を持っていることが分かる。さらにこのネットワークをクリーク・パーコレーション法により部分ネットワーク分割してみると、コアなネットワークほどタグの生成数や新規タグの生成数が多い事がわかった。すなわち類似した相手が一定以上うまると、その結果として新規タグの生成数も増える。

同じようにタグの生成率をユーザー間の like のネットワークを作って解析してみると、やはり同じように大きなクリークほど生成数が多いことが分かった。こうしたことより、ネットワークにはある臨界的なサイズがあり、そこを超えるとある「雰囲気」が立ち上がって、ユーザーのアクティビティが増すのではないと思われる。その結果として新規タグの数も増える。しかしこれはまだ仮説であり、より詳細な解析を行っているところである。そこで次節では、新規タグの生成率を計算し、他のサービスと比較してみた。

2.4 新規タグの生成率

RoomClip のように投稿とともにいろいろなタグをつけることがユーザーのアクションになる機能は多くのウェブサービス

によって採用されている (e.g., Flickr, Delicious, Instagram)。しかしサービスによらず、そこにはある普遍性が知られている。たとえば異なるタグの数 (=ボキャブラリー数) は、投稿数の増加とともにべき的に増加することが知られている。これを Heaps 則という。このボキャブラリーの増加速度が、新しいタグの生成率である。ユーザーがどのくらいの頻度で新しいタグをつけるのか、ひとつの投稿とともにつけられる新規タグはどう変化するか、そうしたものを横断的にいくつかのウェブサービスの元気さの指標として計算した結果を図 3 に示す。

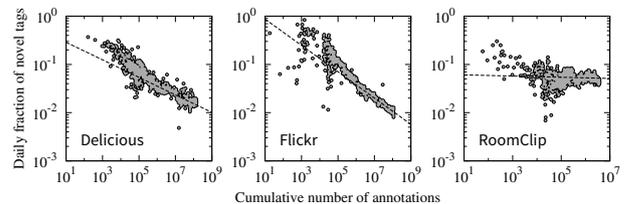


図 3: 異なるサービス間 (Delicious, Flickr, RoomClip) の新規タグ生成率 [Hashimoto]

図 3 にみるように、同じように Heaps 則に従っていても、その新規タグの生成率はサービスによって異なっている [Hashimoto]。この違いは、ひとつにはサービスの性質によって決定すると思われる [Sato]。つまりタグをする目的が、その投稿した写真の種類をよりきちんと分類するものなのか、それともその投稿の注目度をあげようとするものかによって異なっている。Roomclip は後者の性質が多いと考えられる。後者の場合には、新規タグの生成はその時のユーザーの「雰囲気」により影響を受けやすいと考えられる。実際、図 3 では RoomClip の新規タグの生成率はほぼ一定に保たれていることが分かる。

新規タグの成長速度を元気さの指標とするのであれば、RoomClip がこの中では一番元気だといえる。写真ごとにつけられるタグの数に関しても、初期に比べて大きく増加することが分かっている。こうしたことからウェブサービスの進化を生物進化の類推で捉えれば、新規タグの生成率は進化速度であろう [西川 16]。いつまでも進化を続けられること (これを Open-Ended Evolution という [Oka 15]) は、システムの元気さのひとつの大きな指標だといえる。

3. 考察

本研究は、元気さという、漠然としているが生命システムを特徴つける上では、とても大事だと思われるものについて、いくつかの解析例を示してきた。特に、i) Hawkes Process にみるイベント間相互作用が上昇する、ii) 背後のネットワークに大きなクリーク構造が出現、iii) 新規タグの生成率が衰えない、などが具体的な指標として見えてきた。本講演は、人工生命化する社会というオーガナイズドセッションに寄稿するものである。講演ではさらにウェブのサービスが媒介することで社会に元気さを与えていく、それがサービスに跳ね返るような循環構造に関しても議論したい。

参考文献

[Hashimoto] Hashimoto, Y., Sato, K., and Ikegami, M. O. T.: Large fluctuation in the growth

of word occurrences in social tagging dynamics, *to be submitted*.

[Hawkes 71] Hawkes, A. G.: Spectra of Some Self-Exciting and Mutually Exciting Point Processes, *Biometrika*, Vol. 58(1), pp. 83–90 (1971)

[Oka 15] Oka, M., Hashimoto, Y., and Ikegami, T.: An open-ended evolution in a web system, In Late Breaking Papers at the European Conference on Artificial Life (2015), <https://www.cs.york.ac.uk/nature/ecal2015/paper-159.html>

[Sato] Sato, K., Oka, M., Hashimoto, Y., Kato, K., and Ikegami, T.: Tagging Motivation affects the micro structure of Vocabulary creation, *to be submitted*.

[小堀 16] 小堀晋太郎?F ソーシャルタギングにおける語彙生成に関する研究, 卒業研究論文 (2016)

[西川 16] 西川仁將, 岡瑞起, 橋本康弘, 池上高志?F ソーシャルタギングの形質表現と進化メカニズム, 2016 年度人工知能学会全国大会, p. *, 人工知能学会 (2016)