

## 集合知と進化のしくみ A Mechanism of Collective Intelligence and Evolution

池上高志  
Takashi Ikegami

東京大学総合文化研究科広域科学専攻、広域システム科学系  
The University of Tokyo, General Systems Studies

A large web data gathers users with similar tendencies, allowing the network to develop a core and peripheral structure, and it is shown that the generation rate of new tags and their usage frequency are high with the higher-order cliques. Novelty is considered not to be simply creating a new pattern randomly, but rather to be generating a pattern by creating a large and structured network. It also shows new development in the adjacent possible theory as a challenge to understand a new kind of collective intelligence.

### 1. はじめに

これまでの集団知というのは、集団になることで生まれる知性ということだった。具体的に知っている例としては、例えば蟻の合理的な判断が群れを作ること増幅されるというような例である。新しく巣を作ろうとする実験において、偽物の場所(decoy)があってもそれを無視して、より最適な巣の場所を選ぶことができるという実験がある。しかしもちろん集団を作ること帰って非合理的になる場合もある。愚かな集団知と賢い集団知を分けているのはなんだろうか。そこでこの集団知と呼ばれる現象をウェブデータの例を元に議論してみたい。

### 2. RoomClip のデータ

RoomClip とは、ユーザーが典型的には部屋のインテリアの写真を、幾つかのタグのセットとともに投稿する、いわゆるソーシャル・ネットワーク・サービスである。各投稿に対しては、投稿したユーザーを follow し、その投稿に"like"をつけることができる。使用されたタグのリストはみんなが見ることができる。サービス開始以来、ユーザーの数は徐々に増え、150 万を超える写真(2016 年 10 月現在)が投稿されている。このウェブサービスの時間発展を解析した。

#### 1) Like の数の時間変化

ユーザーの数は徐々に増えていくが、その成長カーブとは独立にサービス全体の活性度は何度か遷移を繰り返す。だからこの活性度の変化を引き起こすのは、集団サイズの問題だけではない。しかしユーザーの総数がある閾値を越したために、ある時点を境に Like の数が急に増え流ということが起こっているかもしれない。この Like をつけるという行為は、それを観察する他のユーザーに対するフィードバック効果を持っている。すなわち、Like がつくほどより Like がつきやすくなる。

このようなフィードバック効果は、ポアソン過程にフィードバック効果を付け足した、Network

Hawkes 過程として解析することができる。それは簡単に言うと、Like がつくことで Like がつく確率が高くなる、preferential attachment の仕掛けがあるということだ(例えば[Oka 2015]を参照)。

Like の数の時系列に対して、Network Hawkes 過程をフィットさせると、あるところで、このフィードバック項が立ち上がっている傾向を見ることができる[Miyake 2017]。例えば応答関数のカーネルとして指数関数を用いると、この時の Like のつく応答関数は、ラプラス変換後、ちょうどその応答関数の振幅が発散する特異点に向かって進化する傾向を持つことがわかる。昔からの言い方言えば、カオスの淵に向かって進化すると言えるだろう(図1)。

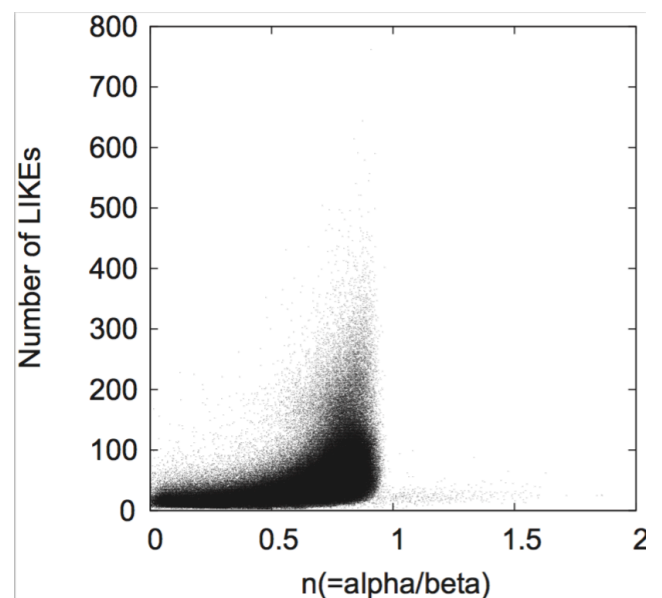


図1 横軸を Hawkes 過程の時定数と振幅の比とし、その値が1に近いところで、Likesの数も発散しているのがわかる。

#### 2) ユーザーのネットワーク

しかし 1)のような構造を作り出すのは、ユーザーがある特徴的な集団を作っているためであるとも推

察される。この時ユーザー同士はどのような集団を作っているだろうか。そこで、ユーザーの類似性を、対称化された KD (Kullback-Leibler) 発散である JS (Jensen-Shannon) 発散を用いて調べてみる [Kobori 2016]。これはユーザーごとに、タグの使用履歴を調べ、どういうタグを使っているかという使用頻度の分布を元に近さを計算し、ネットワークを作るという方法である。ユーザー同士を線でつなげるのは、その結果のエッジの数が完全グラフの 1 パーセントになるように決定された。その結果、ユーザーの類似度ネットワークには、コア (核) とペリフェリー (周辺部) があり、コア部分には 3,4,5 次数のクリークが抽出できる (図 2)。

注目すべきは、このクリーク構造と新規タグの生成率の関係である。高次クリークで作られたタグほど使用頻度が高いようにも思えるが、新しいタグの生成率は必ずしもそうではない。したがって大体の傾向としては、新規タグの生成率も使用頻度も、高次クリークの方が大きいことがわかる。この辺は以後の検証が必要である。

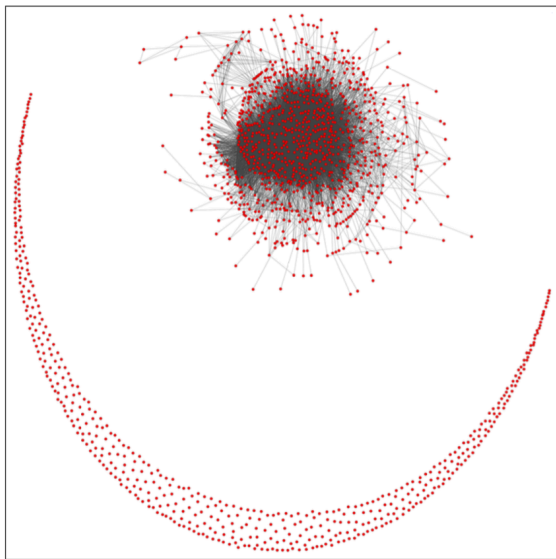


図 2：ユーザーのタグの使用パターンを類似度を図り、それを使ってユーザー間のネットワークを構築。高次クリークほど、新しいタグの生成率が高くなっている傾向がみられる。

新規タグの生成率や使用頻度は、ほぼ核部分で高そうに見える。これは新規タグの生成が集団のネットワークによって加速される。少なくとも核部分で作られたタグの使用頻度が高くなるというのは示唆的である。それは集団知の問題を、集団としての合理的な意思決定を行うようになる、という意味ではなく、新しいものを作り出せる集団の組織化という意味として、集団知を考えることができるということではないか。

### 3. 議論: 集団知と近接可能性理論

ここで見たウェブデータの解析からは、集団知を「新規性の生成する集団」として定義できる。知性には、記憶や知覚、学習と並んで新規性の生成という大事な要素があり、そこを強調する。ここで定義する「新規性」とは、これまでにない新しいパターンや構造を作り出すことである。ウェブサービスを用いた例では、新規性とは投稿される新しいタグのことである。RoomClipの統計的なデータ解析では、新規タグの生成率はある簡単な関数に従っているように見える。この生成率そのものは時間とともに減少するのだが、それぞれの時点での生成率を見ると、サービスが始まってからある時点で、その割合が急激に立ち上がることがわかる。生物進化におけるカンブリア爆発を思わせるこのイベント、新規の爆発的な増加は、一つにはユーザーのネットワーク構造の変化にあると考えられる。ユーザーのネットワークにはコア (核) とペリフェリー (周辺) 構造があるように見える。核は密な相互作用で作られたネットワークで、周辺部は弱いネットワークで繋がったものである。それぞれの領域での新規の生成率を調べると、たくさんと繋がった高次クリークで、新しいタグの生成率が高いことがわかった。

なぜ核の部分で新しいタグの生成率が高いか、たとえば、それはここでの解析の解釈からだけ言えば、ユーザーの似たようなネットワークが作られると、そこには集団的に色々なアイデアを生む作用があるということらしい。これは近接可能性理論 [Kauffman 2003]に関係しているかもしれない。つまり似たようなユーザーが参入することで、ユーザー間ネットワークの構造が核と辺縁部を持ち、新規タグがいくつも生まれる、あるいは Like の数が波動的に増減する幾つかのイノベーションの波がやってくる。そういうストーリーが考えられる。

謝辞：共同研究者の岡瑞起、橋本康弘、小堀晋太郎、および SNS データを提供いただいた Tunnel 株式会社 に感謝いたします

#### 参考文献

- [Kobori 2016] 小堀晋太郎「ソーシャルタギングにおける語彙生成に関する研究」筑波大学・卒業研究 (2016)
- [Miyake 2017] 三宅 雅矩、池上高志「Hawkes Processと最適区間幅推定を用いたWeb Dataの解析」, JSAI 2017.
- [Oka 2015] Mizuki Oka, Yasuhiro Hashimoto and Takashi Ikegami, "An open-ended evolution in a web system", Proceeding of the 13th ECAL (2015).
- [Kauffman 2003] Kauffman, S., In: "The Adjacent Possible: A Talk with Stuart Kauffman" at edge.org, March 11 (2003) .