

ソーシャルタギングの形質表現と進化メカニズム

Trait Expression of Social Tagging System and its Evolution Mechanism

西川仁将 *1
Yoshimasa NISHIKAWA岡瑞起 *2
Mizuki OKA橋本康弘 *2
Yasuhiro HASHIMOTO池上高志 *3
Takashi IKEGAMI*1筑波大学情報学群
School of Informatics, University of Tsukuba*2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*3東京大学大学院総合文化研究科
Graduate School of Arts and Sciences, The University of Tokyo

1. はじめに

進化生物学やその数学的枠組みは、言語やソーシャルタギングシステム (Social Tagging System, STS) を採用した Web サービスなどの非生物学的システムの進化的解析において広く用いられ始めている。[Hashimoto 15, 佐藤 15]。STS とはオンラインコンテンツ共有サービスにおいてユーザが任意の文字列 (e.g., タグ) を付与することでコンテンツの管理を行うシステムのことであり、Delicious、Flickr、Twitter、Facebook などがある。STS を使用した Web サービスに対してはこれまで多くの研究がなされている [Cattuto 07]。

一般にサービス開始とともにサービスを利用するユーザーの数は増え、サービスそのものの構造も時に進化する。ユーザー数の増加とともに、どのようにタグがつけられ、どのようなタグが共起するか、どのようなタグが進化するか、という観点から、STS があたかも生物進化のように解析できるという点で非常に興味深い。それは、非生命的な進化を含む形で進化の理論を拡張できるとともに、ダーウィンの進化論を拡張する形で進化生物学にも貢献できるからである。そこにはダーウィン進化では見つからなかった法則があるかもしれない。

しかし、ウェブサービスと生物進化の対応付けは簡単にはいかない。それは Web サービスには、生物システムと同じ意味で「親子関係」が存在しないからである。タグは自律的には複製しない。しかしユーザーの意識を介して、「複製と変異」がタグに起こることは想像できる。それではタグを遺伝子とみなした時、タグ遺伝子はどのような進化ダイナミクスを示すだろうか。

本研究では、進化生物学の数学的枠組みのひとつであるプライス方程式 [Price 70] による形式化を用いて、タグの進化を分析する。プライス方程式は、ある形質に着目しその経時的な変化を扱うが、そこに明示的な親子関係を仮定しない。そのため Web サービスのような親子関係を持たない進化にはうってつけである。ここで問題になるのは、遺伝子の適応度が何によって決まるのか、ということである。次節ではプライス方程式の枠組みを説明し、タグ遺伝子の適応度が何によって与えられているか、それをソーシャルネットワークサービス (i.e., RoomClip *1) のタグの進化を解析の中心に据えて議論する。

連絡先: yoshimasa@websci.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。

2. プライス方程式

本節では実験において使用したプライス方程式について説明する。プライス方程式は、ある形質の進化を適応度と形質の共分散から導いた量的遺伝の方程式である。この形質というのは生物には限定されずさまざまなものに応用することができる。

プライス方程式は形質を z 、形質の適応度を w とすると、次世代での形質の変化 Δz の平均値 $\langle \Delta z \rangle$ は、式 (1) で表される。

$$\langle \Delta z \rangle = \frac{\text{cov}(w, z)}{\langle w \rangle} + \frac{\langle w \Delta z \rangle}{\langle w \rangle}. \quad (1)$$

ここで $\text{cov}(w, z)$ は w と z の共分散を示している。また (1) 式において、第 1 項は形質の淘汰、第 2 項は形質の変異をそれぞれ表している。

次世代での形質の変化を次世代の形質の平均 $\langle z' \rangle$ と $\langle z \rangle$ の平均の差と捉えれば、

$$\langle \Delta z \rangle = \langle z' \rangle - \langle z \rangle \quad (2)$$

とできる。ここである世代の個体 i の形質、適応度、出現頻度をそれぞれ z_i, w_i, q_i とすると、

$$\langle \Delta z \rangle = \frac{1}{w} \langle (w - \bar{w})z \rangle + \frac{1}{w} \langle w \Delta z \rangle \quad (3)$$

となる。ただし w と q 、次世代の出現頻度 q' には以下のような関係式が成り立つ。

$$w_i = \bar{w} \frac{q'_i}{q_i}. \quad (4)$$

さらに w_i が回帰曲線

$$w_i = \beta_1 z + \beta_2 \quad (5)$$

とし、式 (3) に代入すると、第 1 項、第 2 項目はそれぞれ

$$\langle (w - \bar{w})z \rangle = \beta_1 (\langle z^2 \rangle - \langle z \rangle^2) \quad (6)$$

$$\langle w \Delta z \rangle = \beta_1 \langle z \Delta z \rangle + \beta_2 \langle \Delta z \rangle \quad (7)$$

となる。

表 1: RoomClip データの基本統計

ユーザ数	タグ数	投稿数
410,440	229,250	873,095

3. 実験

3.1 データ

本研究では STS を採用しているソーシャルネットワークの 1 つである RoomClip のデータを使用する。本研究では、2012 年から 2015 年までの 3 年分の写真につけられたタグデータを使用し、サービスの進化をプライス方程式をよって分析する。データ統計値は表 1 に示す通りである。タグは ID で管理され、それぞれのタグはタグ名を持つ。

3.2 コード化

データをプライス方程式に適用するには、出現頻度 q_i や形質 z_i を定義する必要がある。本稿では、ある世代でタグ i が使われた回数 n_i と、その世代 k でタグが使われた回数 $N(k)$ を用い、出現頻度を式 (8) で与える。

$$q_i = \frac{n_i}{N(k)}. \quad (8)$$

形質を決定する際、どのような形質が適切なのかを調べる必要がある。そこで、まず適応度 (w_i) と形質の相関を測った。本研究では形質が低いタグと高いタグ、すなわち新しいタグと古くから使用されているタグの適応度が比較的高い、「年齢」を形質候補として選択した。つまり若いタグほど使われやすいとして、以下の解析を行なう。

タグが作成されてから使用されるまでの時間を t とすると形質 z_i は、

$$z_i = \frac{t_i}{t_{last}} \quad (9)$$

となる。ここで、 t_{last} は使用するデータの一番最後にタグを使用した日付である。

3.3 結果と考察

適応度 w が式 (5) で表せると仮定し、最小二乗法によって β_1, β_2 を求めた。次に、線形にフィッティングした適応度を用い、式 (7) に代入して計算する。その結果を図 1 に示す。横軸は世代 (1 世代は 30 日間) を表し、縦軸は、形質変化を示す。黒線は、実データでの形質変化、赤線はプライス方程式によるフィッティング結果を示す。黄線は、プライス方程式の第一項、つまり既存タグの形質変化が進化にどれだけ影響を及ぼすかを表す。青線は、プライス方程式の第二項、つまり、新規タグの形質変化が進化にどれだけ影響を及ぼすかを表す。これらの結果から、既存タグが進化に及ぼす影響は、世代を経るにつれて減少していき、進化への影響は第二項の影響、つまり新規タグからの影響、が大きいことが読み取れる。

また、プライス方程式でのフィッティング結果と実データの差は、適応度の線形近似によるのみではなく、タグ間の相互作用やプライス方程式のここで扱う二つの項以外の効果があるからだ。例えば、絶滅や、全く新しいタグの発現など、それが、タグの進化を、特殊なものにしていると考えられる。

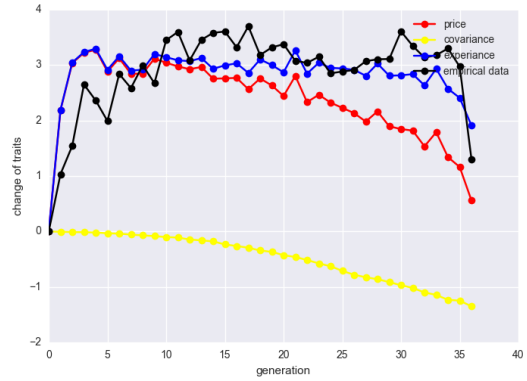


図 1: 実データを用いた形質変化 (黒線) とプライス方程式によるフィッティング結果 (赤線)。黄線はプライス方程式の第一項、青線は第二項を示す。

4. まとめ

タグの進化、すなわちタグの遺伝子の適応度や形質が何によって与えられているかを、進化学の数学的枠組みのひとつであるプライス方程式を用いて調べた。その結果タグの進化はプライス方程式の第二項である新規タグの形質変化に強く依存していることがわかった。更に、実データとプライス方程式によるフィッティングには、ズレが生じていることは、新しい進化のメカニズムがそこに隠れていくことを示唆している。今後は、このズレに注目した解析を進めていきたい。

参考文献

- [Cattuto 07] Cattuto, C., Loreto, V., and Pietronero, L.: Semiotic dynamics and collaborative tagging, *Proceedings of the National Academy of Sciences*, Vol. 104, No. 5, pp. 1461–1464 (2007)
- [Hashimoto 15] Hashimoto, Y.: Growth fluctuation in preferential attachment dynamics, *arXiv:1509.05590* (2015)
- [Price 70] Price, G. R.: Selection and covariance, *Nature*, No. 5257, pp. 520–521 (1970)
- [佐藤 15] 佐藤晃矢, 岡瑞起, 橋本康弘, 加藤和彦? Yule-Simon 過程によるタグ共起ダイナミクスのモデル化と分析, *人工知能学会論文誌*, Vol. 30(5), pp. 667–674 (2015)