

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マルコヴィッチ・オメール*3
Omer Markovitch池上高志*4
Takashi Ikegami

*1筑波大学情報学群

School of Infomatics, 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東京工業大学地球生命研究所

Earth-Life Science Institute, Tokyo Institute of Technology

*4東京大学大学院総合文化研究科

Graduate School 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) 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てきている [Bedau 13][Oka 15]. ソーシャルタギングシステムとはオンラインコンテンツ共有サービスにおいて投稿された画像や動画などのコンテンツにユーザがタグと呼ばれる任意の文字列をアノテーションとして付与していくことによって管理するシステムのことをいう。Twitter, Facebook, Flickrなどがこれにあたり、これまでに多くの研究がなされている [佐藤 15].

一般的に Web サービスはサービス開始とともにユーザ数は増え、ソーシャルタギングシステムにおいては既存のアノテーションが選ばれ、淘汰されたり、全く新しいタグが生み出されていく。また、ウェブサービスそのものも自律的に発展していき構造も進化していく。ユーザ数やコンテンツの増加とともにタグがどのように付けられ、どのようなタグが共起するのか、また、どのようにタグ等のアノテーションが進化していくのか、進化の影響はどのようなものなのかはとても興味深い。本研究ではソーシャルタギングシステムを採用するウェブサービスの一つである RoomClip *1 を対象としたときに、生物学における個体や、形質、適応度といったパラメータがどのように定義できるかを考察し、ダーウィン進化論の数学的枠組みの一つであるプライス方程式を用いて生物学の進化論的アプローチが適用できるかを解析した。

2. プライス方程式

プライス方程式は集団遺伝学での知見から自然選択による進化プロセスを一般化したもので、親世代と子世代の形質の変化量の平均を形質と適応度の共分散 (covariance) から導いた量的遺伝の方程式である [Price 70][Frank 12]. プライス方程式は親世代と子世代の2世代間の形質の差の平均を見て、その世代間である種がどれくらい進化しているかをみていて、(1) 式のような方程式で表す。

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

ここで z は形質であり、形質とは個体の持つ外に現れた特徴

連絡先: yoshimasa@websci.cs.tsukuba.ac.jp

*1 <http://roomclip.jp/>

のことをいう。虫に例えるならば足の長さや、羽の長さがどれくらいかという指標が形質にあたる。また、(1) 式の w は適応度といい、ある形質を持つ個体がどれだけ環境に適しているかを表す尺度である。自然界に適応する場合には個体が残す子孫の数が使われる。プライス方程式の第1項は自然淘汰、第2項が突然変異を表している。自然淘汰とは世代を経るにつれて環境に適さない個体が淘汰されていくことを指していて、突然変異は何らかのエラーにより形質が変化することを意味している。プライス方程式は形質のゆらぎが大きいほど適応度が高くなる、つまり形質のゆらぎと適応度はグラフをとったときに線形になることを意味している。また、プライス方程式はニュートン方程式のように実験的に確かめられる結果ではなく遺伝的な性質とその適応度があれば必ず成り立つ方程式である。

3. 実験

3.1 データ

本研究ではソーシャルネットワークサービス (Social Networking Service, SNS) の一つである RoomClip のデータを使用する。データの期間は 2012 年から 2015 年までの3年間である。RoomClip は Tunnel 社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で、ユーザが室内の写真をウェブに投稿し、その投稿に対して他ユーザがいいねやコメントをすることができる。RoomClip では投稿した写真に対してユーザが任意の複数のタグを付けることができる。写真につけられたタグ群を本研究ではタグセットと呼ぶこととする。データの統計値は表1の通りである。表1におけるライク数とは

表1: RoomClip の基本データ

タグの数	229,213
写真の数	752,071
ユーザ数	410,391
タグセット数	552,860
ライク数	30,823,840

ユーザによってつけられたいいねの数を意味する。

3.2 コード化

RoomClip データをプライス方程式に適用するには、個体や形質、適応度などを定義する必要がある。本研究では個体は

タグセットとし、形質はタグセットの特徴を表す指標としてライク数、適応度はタグセットの出現数とした。これは、適応度はその時点での環境にどれだけ対応しているかを表す指標で、一般的に同様の形質をもつ新しい個体がどれだけ生まれるかが適応度になることが多ことから、本研究ではタグセットのそれに最も近いタグセットの使用回数とした。ある世代 t における形質 $z(t)$ と適応度 $w(t)$ を以下に示す。

$$z(t) = \frac{\text{likes}(t)}{\text{posts}(t)} \quad (2)$$

$$w(t) = \frac{\text{posts}(t+n)}{\text{posts}(t)} \quad (3)$$

$\text{likes}(t)$, $\text{posts}(t)$ はそれぞれ t 世代におけるライク数、投稿数を表している。また、プライス方程式は親世代と子世代の2世代間にどのような進化が起きたかを定量的に示す方程式で、自然界では親子関係が明確な世代があるので計算できるが RoomClip のようなウェブサービスに親子関係は存在しない。そこでデータを1週間毎に区切ることによって擬似的な世代を定義した。

3.3 結果と考察

前節での定義のもとでプライス方程式を計算した。プライス方程式を計算した結果を図1, 図2, 図3に示す。

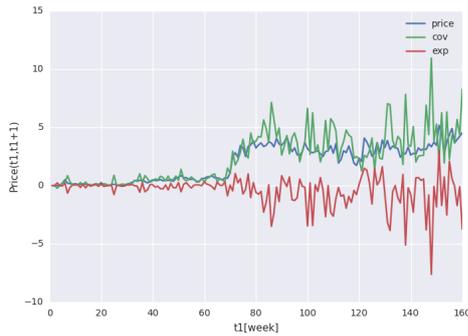


図1: プライス方程式の時間発展

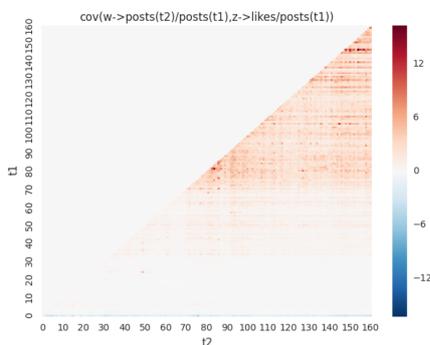


図2: プライス方程式-第1項-

図1はプライス方程式の時間発展を見た図で横軸は世代を表す。縦軸は横軸の世代 ($t1$) を親世代としたときに、次の世代 ($t1+1$) を子世代としてプライス方程式を計算した値であ

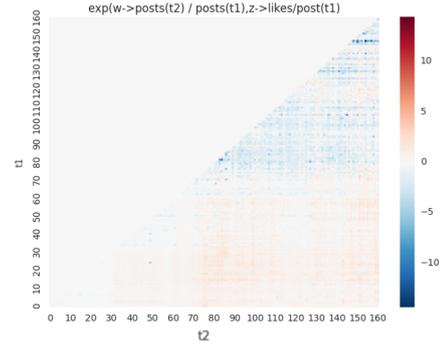


図3: プライス方程式-第2項-

る。緑線は第1項、赤線は第2項を表していて青線はそれらを足し合わせたプライス方程式の値を示している。これを見ると70世代付近で大きな進化が起きていることがわかる。この70世代という時期は RoomClip のインターフェースが大幅にリニューアルされた時期と一致する。また、第1項、すなわち自然淘汰のほうが進化への寄与が大きいことが見て取れる。

図2, 図3は縦軸が親の世代 ($t1$), 横軸が子世代 ($t2$) となっていて、すべての世代の組み合わせでのプライス方程式を計算した結果で図2はプライス方程式の第1項, 図3は第2項を示している。

図2から $t1, t2$ が後半になるに連れて値が大きくなっていることがわかる。つまり後半に出てきたタグセットはその後に出てきたタグセットに影響を与えるということがわかった。第1項は親世代のライク数とタグセット数の比の共分散を表していることから、親世代のライク数が多いほど子世代でのタグセット数が増えることを示している。また、図3から第2項はサービス開始初期に出てきたタグセットが後半のタグセットに影響を与えていることがわかる。第2項はライク数の増加率とタグセット数の相関を表していることから親世代と子世代のライク数の差が大きいほどタグセット数が増えることを示している。

4. まとめ

ウェブサービスに対して生物学の進化論的アプローチが適用できるかをダーウィン進化論の数学的枠組みの一つであるプライス方程式を用いて解析した。その結果、データ全体の進化をみたとき、70世代付近で大きな進化が起きていることがわかった。また、プライス方程式の第1項, 第2項はそれぞれことなる影響を示していることがわかった。

今回の実験では世代を1週間毎に区切り計算したが、本来世代とは親子関係から作られるものであり、各個体ごとに異なるべきである。そこで今後は系統樹等を作成して親子関係を定義し、世代を定義したい。また、形質や適応度を他のデータにしてみても適切な形質や適応度を見つけていく。

参考文献

[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)

[Bedau 13] Bedau, A.M.:Minimal memetics and the Evolution of Patented Technology, Foundations of Science, Vol. 18, Issue 4. pp. 791-807(2013)

[Oka 15] Oka, M., Hashimoto, Y., Ikegami, T:Open-ended Evolution in a Web System, Proceedings of the 13th European Conference on Artificial Life, pp. 159(2015)

[Frank 12] Frank, S.A.:Natural selection.IV. , EUROPEAN SOCIETY FOR EVOLUTIONARY BIOLOGY, pp. 1002-1019(2012)