Open-ended evolution in a web system

Mizuki Oka[1], Yasuhiro Hashimoto[1] and Takashi Ikegami[2]

[1]University of Tsukuba, Japan, [2]The University of Tokyo, Japan
mizuki@cs.tsukuba.ac.jp

Abstract

By using an online photo sharing social network service called RoomClip, new aspects of open-ended evolution (OEE) in a web system are discussed. The reconstruction of a phylogenetic tree with relevant genotype-phenotype mappings is a necessary condition for discussing a Web system as an evolutionary system. By considering a photo as a phenotype and the annotated tags on the photo as genotypes, we found that OEE arises from innovative tags. Herein, we define innovation as the vitality of the service; it is triggered by the emergence of new tags that combine with many other tags. In other words, OEE can emerge as a progressive occurrence of such innovative tags. Based on our findings, we discuss the importance of genotype-phenotype mapping in deriving OEE.

Introduction

Evolutionary biology and its mathematical framework are widely used in various evolutionary analyses of nonbiological systems such as newsgroups, patents, languages, and food recipes. These systems have two properties in common; reconstruction of a phylogenetic tree with relevant genotype-phenotype mappings. For example, in case of US patent evolution (Chalmers et al., 2010), the emergence of a superstar patent (i.e., a patent that produces many other patents) was identified and characterized by measuring its term frequency-inverse document frequency (known as tf-idf).

Studies of nonbiological evolution have aimed to reveal the mechanism of open-ended evolution (OEE), one of the unsolved problems in artificial life studies (Bedau et al., 2000). OEE is defined as the continuous development of innovative technologies and functions found in a phylogeny of evolutionary systems (Chalmers et al., 2010). Thus far, we believe that OEE is found only in the evolution of life on Earth and in the technological evolution of mankind. In studying OEE, it is important to define and identify an *innovation* in an evolutionary system. In the case of US patents, innovative patents were identified by measuring the tf-idf and the number of citations by other patents.

In this study, we use an online social network service (called RoomClip) as an example to study OEE. RoomClip is an online photo sharing service in which registered users post photos with tags. There exist follow-follower relations, and users can either 'like' or 'clip' posted photos (corresponding to retweet and favorite in Twitter, respectively).

We have access to the complete dataset from the launch of the service, and therefore, we can trace back all the relations created along the system's development. A genotype is considered equivalent to a set of tags and a photo, as the phenotype that is annotated by the set of tags. In this study, we define the system as showing OEE if it continues to create new tags, and we define innovation as the vitality (user activities) of a Web service system.

The main finding of this study is that OEE arises from tags that derive many co-occurrences with other tags, as a certain combination of tags invites more postings of photos, thus activating the entire service. As a result, new tags (or vocabularies) are created incrementally. We study the number of likes on photos by using the Hawkes process to look at the users' activation level. We show that the service becomes activated some time after the launch, and the activation level is maintained. We hope that our findings provide useful insights for studying the evolution of biological systems.

Analysis and Results

We used the data obtained from RoomClip (http://roomclip.jp), which was launched in April 2012. Since the launch of the service, the number of users has shown sustainable growth, and the total number of users is around 410,000, as of April 2015. For our analysis, we have gathered all of the posts, tags, likes, and clips as well as information on follow-follower relations for the period between April 2012 and April 2015.

Evidence of OEE: Continuous creation of new tags

IIn Fig. 1, we show the distribution of the distinct number of tags (vocabularies) over the number of annotations (Left) and the average number of tags used per photo over time (Right). It implies that the number of distinct tags is constantly increasing over time along with the number of tags used for each photo. New tags are constantly cre-

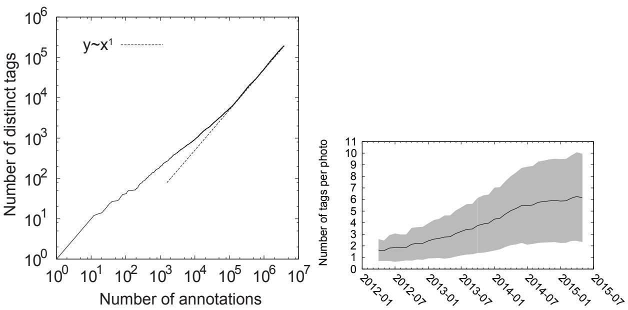

Figure 1: ((Left) Distribution of the distinct number of tags (vocabularies) over the number of annotations of the number of tags. (Right) Average number of tags used per photo over time.

ated (around 5% of tags are newly created on each posted photo), and more tags are used to characterize each photo as the service grows. It has been found that the Yule-Simon process (Simon, 1955) well describes the behaviour of the relation between the vocabulary growth and the number of annotations (Hashimoto et al., 2015).

Extraction of a phylogeny tree

We can extract a variety of phylogeny trees using the information available in the system, such as tags and photos. As one trial, we define the phylogeny between photos by the overlap ratio of the tag sets by the Jaccard coefficient (i.e., it is given by $A \cap B / A \cup B$ between a set of tags in photos A and B). The analysis reveals that evolutionary branching is accelerated by certain types of photos or a set of tags.

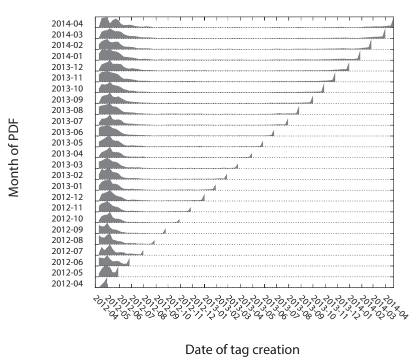

Figure 2: Histogram of used tags mapped to the time when they were created.

Co-occurrence of old and new tags

Fig. 2 shows a histogram of used tags mapped to the time when they were created. Both tags created at the early and most recent stages of the service are used frequently, and most of the others will become obsolete. This implies that the combination of old and new tags are used as a set in newly posted photos.

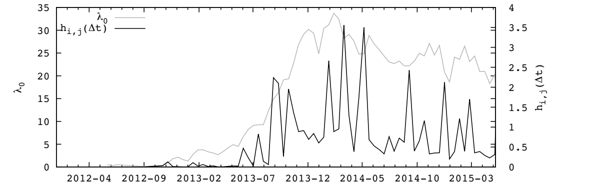

Figure 3: Time evolution of the default activity rate (gray line) and the responses from the inter-user network (black line) calculated by fitting to the Hawkes process.

Measuring the system's vitality

We measured the vitality of the service by using the Hawkes process, the equation of which is given as follows:

$$\lambda^k(t) = \lambda_0^k(t) + \sum_{k'} \int_0^\infty h^{k,k'}(t-\tau)\lambda^{k'}(t-\tau)d\tau,$$

where the first term describes the default activity rate and the second term, the inter-user network activity. In other words, $h^{k,k'}(t-\tau)$ expresses the contribution from user k' to user k with a time delay τ. We fitted the users time series of 'like' events by the Hawkes process by maximizing the likelihood and analysed the users those who posted more than 300 photos.

By using a general response function, we found that the integrated response function occasionally displays sharp bursts, as shown in Fig 3. The responses become bursty when the second part becomes dominant. This bursty nature became apparent after around July 2013.

Discussion

Our hypothesis here is that a Web system maintains its activity by increasing vocabularies in a progressive way. Certain types of tags stimulate users to invent new combinations. At the same time, users take photos to annotate with these tags. This type of tag, which we call innovative tags, can be used to derive the OEE of the service. As this study shows, we claim that genotype-phenotype mapping is a powerful mechanism to maintain and create the OEE, and it is applicable to other nonbiological and, possibly, to biological systems.

Acknowledgements

This work was supported by JSPS Grant-in-Aid for Scientific Research (C) Grant Number 15K00420.

References

Bedau, M. A., McCaskill, J. S., Packard, N. H., Rasmussen, S., Adami, C., Green, D. G., Ikegami, T., Kaneko, K., and Ray, T. S. (2000). Open problems in artificial life. *Artificial Life*, 6:363–376.

Chalmers, D. W., Francis, C. C., Pepper, N. M., and Bedau, M. A. (2010). High-content words in patent records reflect key innovations in the evolution of technology. In *Proceedings of the Twelfth International Conference on the Synthesis and Simulation of Living Systems*, ALIFE 2010, pages 838–845.

Hashimoto, Y., Sato, K., and Oka, M. (2015). Zipf's law and heaps' law in social tagging. Proceedings of the 29th Annual Conference of tje Japan Society of Artificial Intelligence.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3-4):425–440.