

Fluctuation scalings of the appearance of words in the Japanese blogs — observations and modeling —

Hayafumi Watanabe¹, Yukie Sano², Hideki Takayasu³, Seayor Santi⁴, Koki Uchiyama⁵, and Misako Takayasu⁶

^{1,4,5} Hottolink, Inc., Yonbancho Chiyoda-ku, Tokyo 102-0081, Japan

²Division of Policy and Planning Sciences, University of Tsukuba, 1-1-1, Tennodai, Tsukuba, 305-8573, Japan

³Sony Computer Science Laboratories Inc., 3-14-13 Higashigotanda, Shinagawa-ku, Tokyo 141-0022, Japan

⁶Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8502, Japan

E-mail: ¹h.watanabe@hottolink.co.jp

Abstract

To understand the non-trivial empirical statistical properties of fluctuations of time series of words appearances in blogs, we introduce a solvable model of the system consisted of heterogeneous bloggers who write words on their blogs randomly. Firstly, we analyze this model numerically and analytically, from which we obtain formulas of various types of fluctuation scalings (FS), which are the power law scalings between a mean and a standard deviations. In addition these results imply that FSs are independent of the details of the parameters of individual bloggers. Secondly, we applied the model to time series of appearances of words in 1.8 billion Japanese blog entries. From careful data analysis, we find that the model reproduce the various type of statistical properties of fluctuations of words appearances such as: FSs, distributions and the correlation with the total number of blogs, over 8 orders of magnitude.

Keyword: Complex system, Data analysis, Human dynamics, Fluctuation scaling, Social media data

1. Introduction

Fluctuation scalings (FSs) are observed various complex systems, such as random walk on a complex network, Internet traffic, river flows, animal population, cell numbers, stock market, the social network service and word appearance of Japanese blogs[1,2]. However, there are few rigorous studies about the FS with a non-stationary in spite of the fact that most real time series are non-stationary. Therefore, we conduct the intensive study of the FS of non-stationary time series in the case of appearances of words in Japanese blogs theoretically and empirically.

2. Data

In this study, we focused on time series of appearances of words, that is, time series of a daily total count of a focused word in blogs under observations. The data was collected by

using an internet service called “Kuchikomi@kakaricho” (<http://kakaricho.jp>), from which we can obtain the more than 3.1 billion Japanese blog entries from 2006.

3. Results

We introduce the model of the system consisted of heterogeneous bloggers who behave randomly. The model is related to the Random diffusion model [2]. By applying the model to actual data, we find that the model reproduce the various statistical properties of fluctuations of words appearances of Japanese blogs.

References

- [1] Z. Eisler, I. Bartos, and J. Kertesz, “Fluctuation scaling in complex systems: Taylor’s law and beyond”, *Advances in Physics*, vol.57, 1 (2008).
- [2] Y. Sano et.al, ”Empirical analysis of collective human behavior for extraordinary events in the blogosphere”, *Phys.Rev.E*, vol.87,1(2013).