MODELING STATISTICAL PROPERTIES OF HUMAN BLOGGING

Y.Sano[a], K.Yamada[a], K.Watanabe[a], T.Mizuno[b], H.Takayasu[c] and M.Takayasu[a]

[a]Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology
4259-G3-52 Nagatsuta-cho, Midori-ku, Yokohama 226-8502, Japan
sano@smp.dis.titech.ac.jp

[b]Institute of Economic Research, Hitotsubashi University

[c]SONY Computer Science Laboratories

Collective human behaviors are attracting increasing attention by statistical physicists in recent years such as number of registers on conferences [1]. In our presentation, we focus on word frequencies of blogs as a result of collective human phenomena.

First, we discuss about statistical properties of word frequencies observed in real blogs. It is already known that by grouping words depending on their frequencies, there are two limiting cases: the dilute limit and the dense limit for frequency of words [2]. Especially for frequent words which appear almost everyday, we confirm that their appearance number fluctuations are clearly deviated from Poisson statistics and we can approximate the relation between the average number of word appearance $<n>$ and the standard deviation σ empirically by a linear relation as shown in Fig.1. For fashionable new words, the appearance number tends to grow exponentially, and for seasonal words such as "Christmas", the number time series often shows a sharp peak characterized by power laws of time difference measured from the peak day.

To understand these behaviors we develop a simple threshold based agent model. Each agent i has a variable $I_i(t)$ which represents his/her interest and he/she posts a blog entry when his/her interest exceeds a given threshold. We compare the number of blog entries $n(t)$ at time t with the real data. We also introduce a forgetting factor with intensity level α a feedback factor from him/her-self with intensity level β and a feedback factor from number of the whole blog entries with intensity level γ.

$$I_i(t+1) = \alpha I_i(t) + \beta \langle n_i \rangle + \gamma n(t) + f_i(t)$$

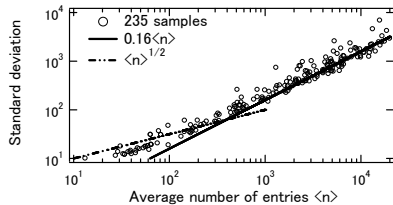where $f_i(t)$ denotes a random noise term. We show numerical results of this model.

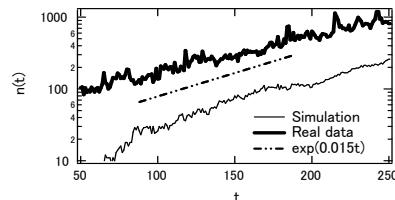
Fig.1 Standard deviation and average number

Fig.2 Simulation results and empirical data

Keywords
human dynamics, blog analysis, time series analysis, agent-based simulation

References
[1] V. Alfi, G. Parisi and L. Pietronero, *Nature Physics*, **3**, 746, 2007.
[2] R. Lambiotte, M. Ausloos and M. Thelwall, Journal of Informetrics, V.1, P. 277-286, 2007.