Dynamics of words popularity observed from large scale social data

K. Yamada[1], K. Sato[2], Y. Sano[3], H. Takayasu[4], M. Takayasu[2]

[1]National Institute of Informatics, Japan

[2]Tokyo Institute of Technology, Japan

[3]Tsukuba University, Japan

[4]Sony CSL, Japan

space written by humans such as blog and tweet is especially interesting database because the data directly reflects trends and topics in human society. For example, the number of blog entries including earthquake have a clear peak when a large earthquake occurs. By using a huge Japanese blog data base with the author ID, we can observe not only the number of entries per day for any words, but also personal dynamics of blog entries. In this presentation, we report statistical properties and modeling for four major categories of words.

The first is ordinary words which is used in our daily life, for example soon. The number of entries of soon has a steady fluctuation. The second is Trending words, for example Twitter. The number of entries of Twitter was increasing exponential from Oct. 2008 to Jun 2010. The third is News words, for example Michael Jackson. We can observe clear jump and power law decaying in the number of entries of Michael Jackson after the news of which Michel Jackson died[3].

The fourth is "event words" which has growth and relaxation characterized by a power function around the peak day such as national holidays. We reproduced these dynamics by an agent-based model based on the SIR (Susceptible-Infected-Recovered) model which is well known in mathematical epidemiology to clarify the origin of these dynamics from the view point of bloggers interactions.

In order to reproduce not only an exponential but also a power law growth and relaxation behaviors observed in trending words, we developed the base model by adding some effects to our model, for example an external shock effect, a deadline effect and an amorphous effect. The amorphous effect, inspired by solid physics studies, gives bloggers individual characteristics, in other words individual duration of interest for the specific word. As a result of adding these essential effects, our model reasonably reproduces the dynamics observed from our data. In addition we give master equation of our stochastic agent-based model and introduce the equation of lifecycle of popularity as seen in the trending words.

Our model can be applied to a prediction and control for spreading false rumors as seen in after the huge 2011 earthquake in Japan[4].

[l] D.J Watts, Nature **445**, 489 (2007).

[2] D. Lazer, et al., Science **323**, 721 (2009).

[3] Y. Sano, K. Yamada, H. Watanab, Phys. Rev. E **87**, 012805 (2013).

[4] M. Takayasu, K. Sato, Y. Sano,, PLoS ONE **10**, e0121443 (2015).