Understanding human and society using fast and accurate language technologies

Humans think with languages and verbalize experiences in the real world to convey them to others. Our laboratory studies on natural language processing that process languages efficiently and accurately with computers. The pursuit of such technologies leads to computational linguistics that reveals the mechanism of languages and intellectual ability.

Recently, the growth of social media and mobile devices allows us to accumulate our experiences and opinions as social big data (Fig. 1). We deeply analyze massive text in the social big data with computers (Fig. 2) to read society trend (Fig. 3). We also develop technologies that promote language communication and aim at understanding our thinking from what we write. We welcome students who are willing to design novel and important NLP tasks rather than solving classic tasks on worn-out datasets, and those who want to study languages from scientific perspectives.

http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/index.en.html

Reading massive text extremely efficiently

Social big data contains massive text, in which the quality (content) and quantity (volume of flow) changes dramatically over time. We have exploited redundancy in text information to accelerate classifiers for deep text analysis orders of magnitude faster. We have also invented a classifier that adaptively speeds up when the amount of text increases, a methodology for compiling real-word knowledge acquired from time-series text, and a real-time method that discovers emerging entities from microblog streams.

Promoting language communication

To further expand valuable social big data, we develop dialogue systems and cross-lingual language technologies such as machine translation that promote language communication. We have leveraged massive dialogue data on social media to realize an emotion-aware human chatbot and a situation-aware chatbot that exploits non-textual information. We are now studying on multilingual models and language technologies that compliment machine translation.

Understanding humans via languages

Being aware of humans that manipulate languages, we study on methods that read our mind from our writings. We have so far devised a method that induces values from social media. We are now attempting to clarify interpersonal variations in word meanings from review text.

For inquiry, feel free to contact Naoki Yoshinaga (ynaga@iis.u-tokyo.ac.jp).

Fig. 1: Social Big Data.

Fig. 2: Efficient and adaptive fundamental text analysis.

Fig. 3: Tracking topics and events in social media.

Fig. 4: Promoting communication and understanding humans.