Computing Human Languages Fast and Accurately

Humans think with language, and verbalize their experiences in the real world to convey them to others. Aiming to observe humans and the real world from text and aiding our language activities, our laboratory studies on natural language processing (NLP) that computerizes text efficiently and accurately. The pursuit of efficient but accurate NLP models leads to computational linguistics (CL) that reveals the mechanism of languages and ultimately human intelligence.

Read massive and diverse text faster

Social big data contains massive text, in which the quality and quantity change over time. To go beyond the speed limit of the current NLP technologies, we leverage various approaches studied in computer science to make the fastest NLP. Specifically, we have developed a classifier that becomes faster when redundancy in text increases, and realized a morphological analyzer that can process 1,000,000 sentences per second. Meanwhile, we have established a method that finds the emergence and disappearance of real-world entities from text.

These methods realize the infrastructure for analyzing social big data, which allows us to understand situations in disasters such as the 3.11 earthquake and public opinions during the COVID-19 pandemic.

Speak accurately considering situation

Although the meaning of text depends on the situation, namely, by whom, when, and where the text is written, most of the existing NLP benchmark datasets contain fragments of text and ignore those contexts (incomplete problem settings). Respecting the surrounding contexts around text, we study various applications such as dialogue systems and machine translation. Specifically, we have developed methods for text understanding and text generation, while considering surrounding text, time information, and speakers and listeners.

To push past the limits of static NLP models based on machine learning, we make a neural model seek for knowledge when it lacks confidence in its inference. We also promote automatic “subjective” evaluation for personalized text generation.

We study all aspects of NLP/CL, and expect students to voluntarily set up their research themes. We welcome those who are willing to design novel and important NLP tasks, without being hidebound by the current research method that solves classic tasks on worn-out datasets.

Our laboratory has been collaborating with Toyoda and Goda laboratories in IIS, and sharing research environments (data, computational resources, and student rooms). The members will receive feedback for their research from various perspectives, and can perform out-of-the-box, cross-cutting research.

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