

Machine Intelligence from Multimodal Data Streams

徐宏民 (Hsu, Winston)

国立台湾大学情报工学科 教授
NVIDIA AI Lab (NTU) Director

【要旨】

Images, videos, audios, and user-contributed logs, etc., are major data types nowadays essential for disruptive opportunities in social media, entertainments, education, healthcare, IoT, etc. However, the developed techniques are far behind the dire needs. For leveraging multimodal data streams with effective deep neural networks, we will present advanced and novel methods for jointly considering spatial and sequential neural networks and their variations, which well approximate the multimodal data streams. We will show the importance and the problems for cross-domain learning – computing data of different types in the same semantic space – and present several solutions. We will demonstrate how to utilize such diverse modalities for improving challenging learning tasks in the end-to-end neural networks. We will showcase some recent works published in the top venues, e.g., CVPR, ICCV, etc., and cognitive applications having been deployed for online services collaborating with international partners (e.g., Microsoft Research, IBM, etc.).