Multilingual Multimodal Language Processing Using Neural Networks

Instructors: Mitesh M Khapra and Sarath Chandar

Abstract:

We live in an increasingly multilingual multimodal world where it is common to find multiple views of the same entity across modalities and languages. For example, news articles which get published in multiple languages are essentially different views of the same entity. Similarly, video, audio and multilingual subtitles are multiple views of the same movie clip. Given the proliferation of such multilingual multimodal content it is no longer sufficient to process a single modality or language at a time. Specifically, there is an increasing demand for allowing transfer, conversion and access across such multiple views of the data. For example, users want to translate/convert news articles to their native language, automatically caption their travel photos and even ask natural language questions over videos and images. This has led to a lot of excitement around this interdisciplinary research area which requires ideas from Machine Learning, Natural Language Processing, Speech and Computer Vision among other fields.

In this tutorial we focus on neural network based models for addressing various problems in this space. We will first introduce the participants to some of the basic concepts and building blocks that such approaches rely on. We will then describe some of these approaches in detail. There are two important parts to the tutorial. In the first part, we will talk about approaches which aim to learn a common representation for entities across languages and modalities thereby enabling cross lingual and cross modal access and transfer. In the second part we will talk about multilingual multimodal generation. For example, we will discuss neural network based approaches which aim at (i) generating translations in multiple languages, (ii) generating images given a natural language description and (iii) generating captions in multiple languages.

Outline:

1. Introduction and Motivation [20 mins]
2. Basics [40 mins]
    1. Learning distributed representations using Neural Networks
    2. Convolutional Neural Networks
    3. Recursive Neural Networks and its variants
3. Multilingual/Multimodal Representation Learning [40 mins]
    1. Using parallel data with and without word alignments
    2. Using pivot view in the absence of parallel data
4. Multilingual/Multimodal Generation [80 mins]
    1. Neural Machine Translation systems
    2. Generating captions from images
    3. Answering natural language questions over images

    4. Describing videos
    5. Generating images from a given natural language description
5. Summary

About the Instructors:

Mitesh M Khapra: IBM Research India, mikhapra@in.ibm.com, http://researcher.watson.ibm.com/researcher/view.php?person=in-mikhapra

Mitesh Khapra obtained his Ph.D. from the Indian Institute of Technology, Bombay in the area of Natural Language Processing with a focus on reusing resources for multilingual computation. His areas of interest include Statistical Machine Translation, Text Analytics, Crowdsourcing, Argument Mining and Deep Learning. He is currently working as a Researcher at IBM Research India where he is focusing on mining arguments from large unstructured text. He is also interested in learning common representations across languages and modalities with the view of enabling cross language and cross modal access. He has co-authored papers in top NLP and ML conferences such as ACL, NAACL, EMNLP, AAAI and NIPS.

Sarath Chandar: University of Montreal, apsarathchandar@gmail.com, http://sarathchandar.in/

Sarath Chandar is currently a PhD student in University of Montreal where he works with Yoshua Bengio and Hugo Larochelle on Deep Learning for complex NLP tasks like question answering and dialog systems. His research interests includes Machine Learning, Natural Language Processing, Deep Learning, and Reinforcement Learning. Before joining University of Montreal, he was a Blue Scholar in IBM Research India for a year.