NJIT

Department of Computer Science

Universal Natural Language Processing with Limited Annotations

Wenpeng Yin
Salesforce Research, Palo Alto, California

Hosted by Jing Li

DATE: Monday, March 1, 2021
TIME: 2:30 PM - 3:30 PM
LOCATION: https://njit.webex.com/njit/j.php?MTID=maf47c56a06a477d9331d7fdb574fe648

https://cs.njit.edu/seminars

Abstract: Much ink has been spilled on a number of fundamental challenges regarding the building of an intelligent machine given task-specific annotated data. This has considerably pushed the frontiers of natural language processing (NLP). However, we are still far away from our dream— universal NLP, typically because: (i) different NLP problems are often modeled independently. This means we have to build new machines (often from scratch) for the new tasks. Human beings, instead, can correlate multiple tasks and leverage knowledge learned from one task to solve another one; (ii) current learning algorithms are overly dependent on large-scale human annotations. We lack a clear solution to make machines work in real-world scenarios where annotations are limited (e.g., zero-shot or few-shot problems).

In this talk, I will introduce our progress towards the goal of universal NLP in three stages. First, we start with unifying various text classification problems where a piece of text can be associated with various open-domain, open-aspect labels such as topics, emotions, and so on while no label has annotated data. Second, we discover and define a more realistic challenge -- "incremental few-shot text classification" where new labels keep coming to the machine while each one is accompanied with only K examples. The machine needs to learn the new concepts with a few examples without re-training using the preceding data. Lastly, we go beyond the text classification problems to unify more distinct NLP problems (e.g., Question Answering, Coreference Resolution, Relation Extraction, etc.) that have a common reasoning process and each has only K annotated examples. Overall, the research of universal NLP pushes us to think more about the shared logic behind various problems and how to make use of indirect supervision to solve the new challenges.

Bio: Dr. Wenpeng Yin is currently a research scientist at Salesforce Research, Palo Alto, California. Before, he was a postdoctoral researcher at the University of Pennsylvania, working with Prof. Dan Roth in textual entailment and information extraction. Wenpeng received a Ph.D. degree in Oct. 2017 from Prof. Hinrich Schütze's group at the University of Munich, Germany. He had an internship at the IBM Watson Research Center in early 2016, and got multiple awards in the past, including WISE2013 "Best Paper", "Baidu Ph.D. Fellowship" in 2014&2015, "Chinese Government Award for Outstanding Self-financed Ph.D. Students Abroad" in 2016, and "Area Chair Favorites" paper award in COLING2018. He was an invited Senior Area Chair for NAACL'21, Area Chairs for NAACL'19 and ACL'19&21.