

학부생 연구기회 프로그램 (UROP) 공고

◆ 담당교수 : 황승원	◆ 연구실명 :언어및 데이터지능 연구실
◆ UROP 연구 과제명 :언어모델 데이터셋 편향 제거	
◆ 모집대상 : 자연어 관련 연구에 관심있는 3,4학년 학부생	
◆ 모집기간 : ~ 2021년 6월 말	

자연 언어 모델 훈련을 위해 제공되는 데이터셋에 존재하는 편향으로 인해, 언어모델들이 특정 데이터셋에서만 좋은 성능을 보이고 다른 데이터셋에서 잘 작동되지 않는 현상을 보인다(lack of robustness). 이를 막기 위한 편향 제거법을 연구한다.

Task:

- (1) 편향 검출
- (2) 편향 제거를 위한 데이터/모델 생성

Requirements: 자연어 처리 기술, CNN/RNN, Transformer, BERT

연구실 관련연구:

Robustifying Multi-hop QA through Pseudo-Evidentiality Training, ACL 2021

Less is More: Attention Supervision with Counterfactuals for Text Classification, EMNLP 2020

Conditional Response Augmentation for Dialogue using Knowledge Distillation, INTERSPEECH 2020



서울대학교 컴퓨터공학부
Seoul National University
Dept. of Computer Science and Engineering

문의

seungwonh@snu.ac.kr