학부생 연구기회 프로그램 (UROP) 공고

◆ 담당교수 : 강유	◆ 연구실명 : 데이터 마이닝 연구실
◆ UROP 연구 과제명 : 모바일 디바이스를 위한 RNN 모델 압축 기술 연구	
◆ 모집대상 : 데이터 마이닝 및 기계 학습에 흥미 있는 3-4학년 학부생	
◆ 모집기간 : ~ 2020년 12월 말	

RNN Compression

▶ Model Compression

- Deep Learning models show a powerful performance, but some of them are too big and slow to be used for the mobile device.
- We need to compress a large RNN model to be fast and small enough for the mobile device.

▶ Goal

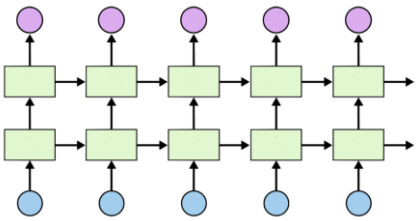
- Compress a large RNN model without sacrificing accuracy.

▶ Requirements

- Basic knowledge about the deep learning frameworks (e.g. PyTorch, TensorFlow)

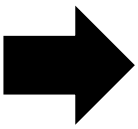
Example of RNN Compression

Original Model

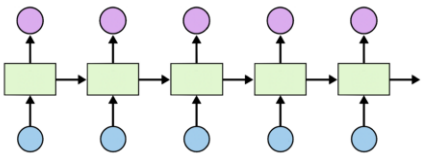


- Size: 1 GB
- Inference time: 10 sec
- Accuracy: 93%

Compress



Compressed Model



- Size: 100 MB
- Inference time: 1 sec
- Accuracy: 93%

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

문의 데이터 마이닝 연구실 박승철 (ant6si@snu.ac.kr)