# 학부생 연구기회 프로그램 (UROP) 공고

| | |
|---|---|
| ◆ 담당교수 : 전병곤 | ◆ 연구실명 : 소프트웨어 플랫폼 연구실 |
| ◆ UROP 연구 과제명 : 머신러닝을 통한 빅데이터 처리 시스템 최적화 | |
| ◆ 모집대상 : 머신러닝 이론과 Python과 Java에 익숙한 학부생 | |
| ◆ 모집기간 : ~ 2019년 6월 말 | |

## 연구 배경

- 머신러닝, 데이터마이닝 등 빅데이터 처리 방법이 다양해지고, 데이터센터들이 많이 설립되며 다양한 클러스터 환경들이 등장하게 되었습니다.
- 기존의 Apache Spark, Apache Hadoop 등의 데이터 처리 시스템들은 새로운 클러스터 환경에 적응하지 못하는 문제점이 있습니다.
- Apache Nemo는 이러한 문제점에 착안하여 만들어진 여러 환경에 알맞게 데이터 처리 방법을 유연하게 바꿀 수 있는 기능을 가진 시스템입니다.
- 이 기능을 머신 러닝을 이용하여 최적화하여 사용자의 특별한 조작 없이도 알맞게 데이터 처리 방법을 설정해주는 시스템을 만들고자 합니다.
- Website: http://nemo.apache.org / ATC'19 Paper: https://bit.ly/2KfmYSt

## 연구 내용

- 세계 최대 빅데이터 커뮤니티인 Apache Software Foundation 산하의 오픈소스 소프트웨어 개발에 참여할 수 있는 경험을 쌓는다.
- 머신러닝 지식을 실제 시스템 위에 구현하고 결과를 도출한다.
- Apache Nemo위에서 머신 러닝을 활용하여 데이터 처리 방법을 최적화하는 방법에는 여러가지가 있다. 학생은 시스템 최적화를 위해 관심사와 알맞는 머신러닝 모델을 작성하거나 개선해보고, 그 결과를 관찰한다.
  - 모델의 종류에는 decision trees (random forest, boosted trees), Bayesian optimization, deep neural networks, reinforcement learning 등이 존재한다.
- 시간 및 기간은 같이 이야기 하면서 유연하게 조정 가능합니다.

## 사전 지식 및 요건

- Linux 사용 경험, Python과 Java 코드를 읽고 작성할 수 있어야 함.
- Regression/classification, supervised/unsupervised learning, cost function/regularization 등 기본적 머신 러닝 이론에 대한 지식.
- [선택] TensorFlow 등 딥 러닝 프레임워크 사용 경험.


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

소프트웨어 플랫폼 연구실
Seoul National University
Software Platform Lab

문의 송원욱 (wonook@apache.org)