

딥러닝 보조 레이블링 작업 시스템

서울시립대학교
실전문제연구단
인공지능솔루션연구팀

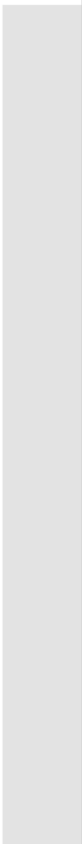
2020 연구성과 공유 한마당



서울시립대학교
UNIVERSITY OF SEOUL



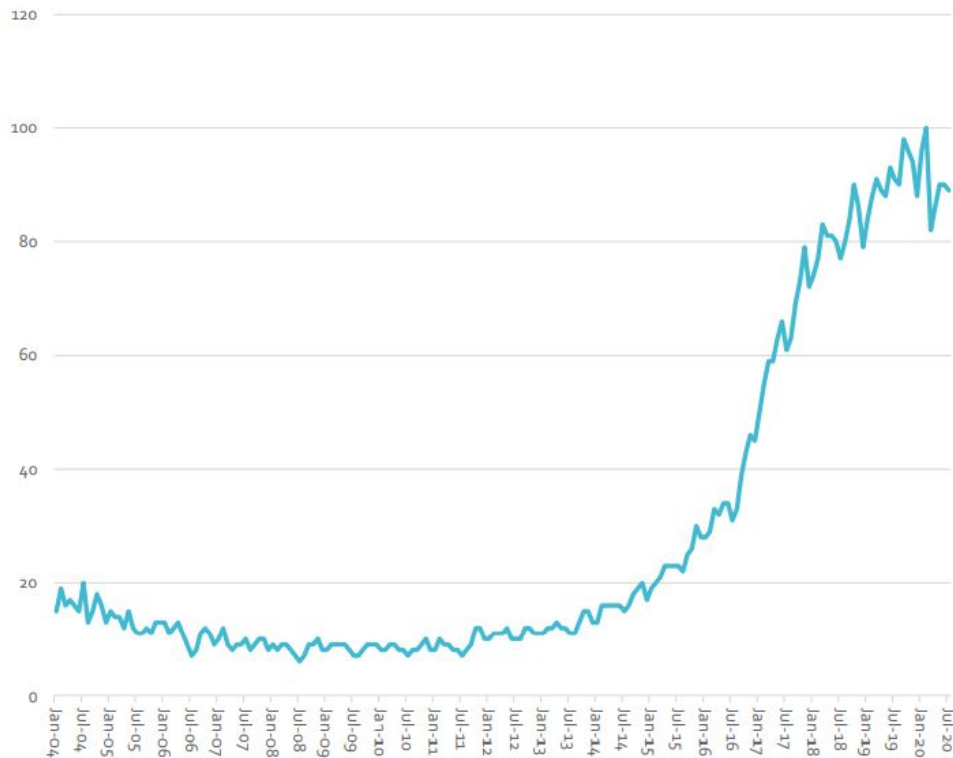
OUTLINE

- 서론
 - 설계 및 구현
 - 결론 및 향후연구
- 

서론

Google Trends, 2014.01. ~ 2020.07.

Machine learning





지도학습 (Supervised Learning)



준지도학습 (Semi-supervised learning)



비지도학습 (Unsupervised learning)

머신러닝의
세 분류

지도학습

- AlexNet

- ILSVRC 2012 에서 최고의 성능을 보여줌
- Convolution Neural Network 를 사용하여 지도학습을 실시
- ImageNet 으로부터 1500만개 이상의 레이블된 사진 데이터를 사용하여 학습을 진행

문제점

- ImageNet

- 14,197,122개의 데이터를
- 21,841개의 카테고리
- 전부 수작업으로 분류



Deep learning Assisted Labeling Task

- 기본개념

- 레이블링 작업이 완료된 데이터를 실시간으로 학습
- 레이블링 작업자에게 가능성 높은 선택지를 우선적으로 보여줌
- 작업자는 분류작업을 하는 것이 아닌, 분류가 제대로 되었는지 검증작업만 하게함으로 생산성을 높이는 것을 목표로 함.

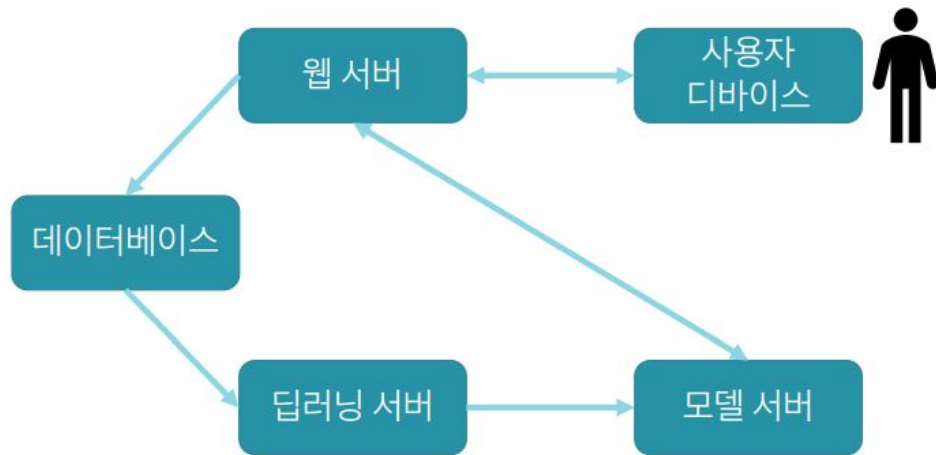


Deep learning Assisted Labeling Task

- 웹 브라우저를 통해 레이블링 작업공간에 접속
- 레이블링 할 원본 데이터를 상단에 표시
- 현재 학습된 모델이 판단하기에 가능성 높은 카테고리를 강조해서 표현

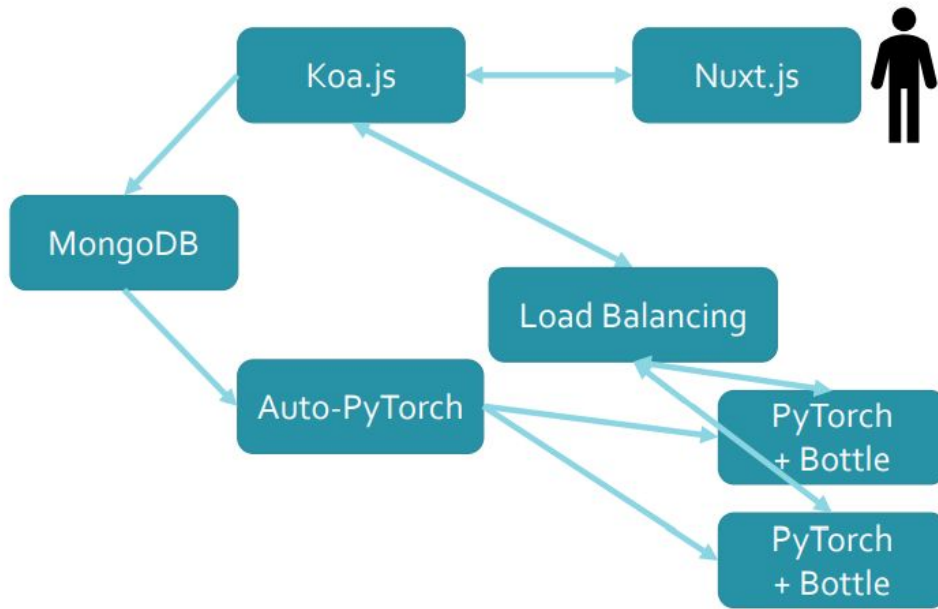


설계 및 구현



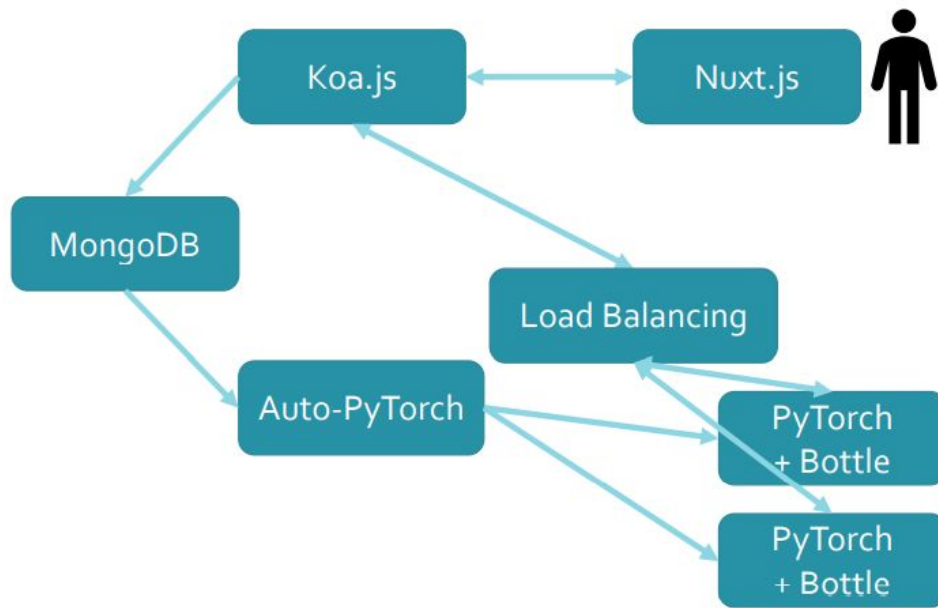
- 사용자의 레이블링 작업결과는 실시간으로 데이터베이스에 기록
- 딥러닝 서버는 매번 데이터베이스로부터 배치를 불러와 학습을 진행
- 딥러닝 서버에서 갱신된 모델은 모델서버에서 REST API 형태로 서비스 제공
- 웹 서버는 모델 서버로부터 원본 데이터의 추론 정보를 제공받음

설계 및 구현



- Docker를 통해 컨테이너 형태로 위 서비스들을 구동

설계 및 구현



- 신속한 학습을 위해서 Auto Encoder 기법을 사용하여 미리 원본 데이터의 압축된 특징데이터를 추출, 특징데이터만 사용하여 학습을 진행함
- Auto-PyTorch를 통해 적절한 파라미터를 찾는 과정을 자동화함

결론 및 향후연구

- 요약
 - 실시간 학습을 통하여, 레이블링 작업을 보조하는 시스템을 설계하고 개발하였다.
- 한계점
 - 실제 레이블링 작업환경에서 작업자의 생산성 향상에 도움을 주는지 아직 검증되지 않았다.
- 향후 연구
 - 레이블링 작업 생산성 향상에 본 시스템이 얼마나 도움을 주었는지 평가.

감사합니다

