

데이터 편향성이 예측에 미치는 영향 1

➤ 티쳐블머신을 이용한 이미지 식별

승실대학교 베어드교양대학
서유환 교수
yhsuh@ssu.ac.kr

목차

- ▶ 티쳐블머신을 이용한 지도학습 모델 생성
- ▶ 데이터 편향에 따른 예측 결과 비교
- ▶ 인공지능 편향의 문제점 생각해보기

학습 목표

- ▶ 데이터 편향에 대해 이해한다.
- ▶ 데이터 편향이 인공지능의 예측에 미치는 영향을 설명할 수 있다.
- ▶ 인공지능 편향이 우리 생활에 미칠 수 있는 문제점과 우리가 앞으로 나아가야 할 방향에 대해 생각해 본다.

지난시간 배운 내용

주	주제	온라인	오프라인
1	인공지능의 과거 현재와 미래	1. 강의 및 교과목 소개(공통, 핵심만) 2. 인공지능의 과거와 현재 3. 인공지능의 미래와 다양한 시선 4. 인공지능 개발환경 구축과 사용법(Anaconda/Colab)	1. 강의 및 교과목 소개(분반별, 자세히) 2. 다양한 인공지능 기술 경험하기 (자연어처리, 시각, 음성) 3. 인공지능 챗봇만들기(IBM 왓슨 어시스턴트)
2	공공데이터를 이용한 사회문제 발견과 해결책 모색	1. 빅데이터의 정의와 가치 2. 공공데이터 수집하기 3. 공공데이터로부터 새로운 인사이트 발견하기 - 행정구역별 인구 데이터와 공공의료기관 현황 데이터 분석	1. 서울시 CCTV설치 현황 분석하기 2. 서울시 범죄발생 현황 분석하기
3	인공지능의 개요 및 머신러닝을 이용한 예측	1. 인공지능의 정의와 분류 2. 인공지능 학습방법 이해하기 3. 인공지능 알고리즘 소개	1. 머신러닝을 이용한 이미지 식별(구글 티쳐블 머신) 2. 머신러닝을 이용한 보스톤 집값 예측
4	인공지능과 데이터 윤리	1. 데이터의 불완전성과 결함에 따른 예측 오류와 차별 2. 데이터 편향성이 예측에 미치는 영향 (구글티쳐블머신) 3. 지도학습(SVM)을 이용한 타이타닉호 생존자 예측	1. 타이타닉호 생존자 예측 - 데이터 편향성이 예측에 미치는 영향 - 데이터 왜곡에 따른 예측 결과 비교
5	인공지능과 알고리즘 윤리	1. 알고리즘 기반 의사결정 시스템의 한계 2. 윤리가 적용된 인공지능 알고리즘	1. 알고리즘에 따른 예측 결과 비교 - 보스톤 집값 예측 - 폐암환자 생존 여부 예측
6	인공지능에 대한 다양한 이슈와 우리의 자세 고찰	1. 인공지능의 윤리적/법적 쟁점 (자율주행자동차, AI로봇, 트랜스 휴먼 등) 2. 인공지능시대 사회, 경제적 불평등 문제 3. 인공지능과 프라이버시 4. 인공지능의 윤리적 대응과 규제	1. 자율주행 자동차의 행동학습 시나리오 경험하기 2. 비윤리적 데이터 생성과 수집(웹 크롤링을 이용한 데이터 수집)
7		기말고사	

구글 티처블 머신(Teachable Machine)

▶ <https://teachablemachine.withgoogle.com/>

▶ 이미지, 사운드, 자세를 인식하여 검사

▶ 웹기반으로 기계학습 모델을 생성

▶ 사용법

1. 이미지 모으기

학습시키고자하는 이미지들을 수집
이미지 그룹별로 그룹핑하는 것이 좋음

2. 이미지 데이터 업로드(Image Upload)

3. 학습(Train)

4. 새로운 데이터로 예측 결과 보기

5. 내보내기

▶ 모델 파일을 다운로드하거나 온라인으로 요청하여 데이터 수신

시작하기

Teachable Machine

이미지, 사운드, 자세를 인식하도록 컴퓨터를 학습시키세요.

사이트, 앱 등에 사용할 수 있는 머신러닝 모델을 쉽고 빠르게 만들어 보세요. 전문지식이나 코딩 능력이 필요하지 않습니다.

시작하기

새 프로젝트

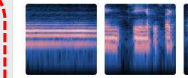
Drive에 있는 기존 프로젝트를 엽니다.

파일에서 기존 프로젝트를 엽니다.



이미지 프로젝트

파일 또는 웹캠에서 가져온 이미지를 기반으로 학습시키세요.



오디오 프로젝트

파일 또는 마이크에서 가져온 1초 분량의 사운드를 기반으로 학습시키세요.



포즈 프로젝트

파일 또는 웹캠에서 가져온 이미지를 기반으로 학습시키세요.

새 이미지 프로젝트

표준 이미지 모델

대부분의 용도에 적합

224 x 224px 정사각 이미지

TensorFlow, TFLite, TF.js로 내보내기

모델 크기: 약 5mb

삼입력 이미지 모델

마이크로 컨트롤러에 적합

96 x 96px 그레이스케일 이미지

마이크로컨트롤러용 TFLite, TFLite, TF.js로 내보내기

모델 크기: 약 500kb

이 모델을 지원하는 하드웨어를 확인하세요.

남자1 과 여자1의 클래스 생성과 학습데이터 업로드

▶ 지도학습 모델 만들기

남자1

① 클래스 이름입력 및 생성 (정답)

이미지 샘플 추가:

웹캠

업로드

② 업로드(드래그 앤 드롭)

여자1

① 클래스 이름입력 및 생성 (정답)

이미지 샘플 추가:

웹캠

업로드

② 업로드(드래그 앤 드롭)

이름

테스트 데이터

학습 데이터 1

학습 데이터 2

상태

✓

✓

✓

이름

남자1

여자1

상태

✓

✓

학습시키기

남자1

6 이미지 샘플

웹캠

업로드

6 이미지 샘플

여자1

파일

파일에서 이미지를 선택하거나 여기로 드래그 앤 드롭하세요.

Google Drive에서 이미지 가져오기

6 이미지 샘플

학습

모델 학습시키기

고급

미리 보기

모델 내보내기

여기에서 모델을 미리 확인하려면 먼저 왼쪽에서 모델을 학습시켜야 합니다.

테스트시킴

9



생각할 문제

10

- ▶ 인공지능 모델은 남자와 여자를 정확하게 구분하였는가?
- ▶ 인공지능 모델이 남자와 여자를 정확하게 구분하지 못하는 이유는 무엇일까?
- ▶ 만약 목욕탕에서 여러분이 만든 인공지능 모델로 남자와 여자를 구분하여 출입을 허용한다면 어떤 일이 일어날까?

테스트 결과

11

남자						
	출력 남자1 100%	출력 남자1 100%	출력 남자1 100%	출력 남자1 99%	출력 남자1 99%	출력 남자1 99%
	출력 여자1	출력 여자1	출력 여자1	출력 여자1	출력 여자1	출력 여자1
여자						
	출력 남자1	출력 남자1 99%	출력 남자1 100%	출력 남자1 100%		
	출력 여자1 100%	출력 여자1 100%	출력 여자1 100%	출력 여자1 100%		

새 프로젝트 생성하기

12



학습데이터2로 학습모델 만들기

13

The interface shows a '학습' (Learning) section with a '모델 학습시키기' (Train Model) button. On the left, there are two image sets: '남자2' (Male 2) and '여자2' (Female 2), each with a '클래스 입력(정답)' (Class Input (Answer)) field. Below these, there are instructions for uploading images: '파일' (File) and 'Google Drive에서 이미지 가져오기' (Get image from Google Drive). A red dashed box highlights the '학습' button and the '모델 학습시키기' button. A red arrow points from the '학습' button to the '모델 학습시키기' button.

테스트시킴

14

The interface shows a '테스트' (Test) section with a '모델 테스트하기' (Test Model) button. On the left, there are two image sets: '남자2' (Male 2) and '여자2' (Female 2), each with a '클래스 입력(정답)' (Class Input (Answer)) field. Below these, there are instructions for uploading images: '파일' (File) and 'Google Drive에서 이미지 가져오기' (Get image from Google Drive). A red dashed box highlights the '테스트' button and the '모델 테스트하기' button. A red arrow points from the '테스트' button to the '모델 테스트하기' button.

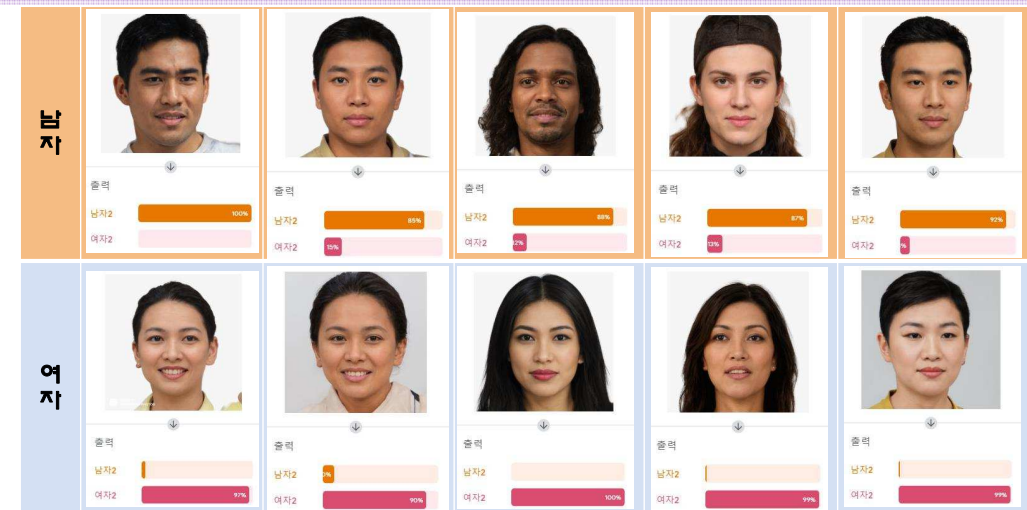
생각할 문제

15

- ▶ 다시 학습시킨 인공지능 모델은 남자와 여자를 정확하게 구분하였는가?
- ▶ 인공지능 모델이 남자와 여자를 정확하게 구분했다면 그 이유는 무엇이고 그렇지 못한 이유는 무엇일까?
- ▶ 나아가 이러한 인공지능 모델이 신용평가, 재범예측, 채용 등에 사용된다면 어떠한 문제가 있을 수 있을까?

테스트 결과

16



스스로 해보기

17

▶ 이외 어떠한 경우 데이터 편향이 발생할 수 있을까?

- ▶ 한쪽 그룹의 데이터만 많이 학습 시키는 경우

“ 인공지능은 지금까지의 전통적인 사회적 편견을 학습하여 사회적 편향을 더욱 강화시킬 수도 있다”

정리

18

▶ 인공지능 편향/ 데이터 편향

- ▶ 인공지능 모델을 만들 때 특정 집단의 특징을 모두 반영하지 못한 데이터를 학습시킴으로써 나타나는 현상

▶ 인공지능 편향으로 나타나는 문제

- ▶ 인공지능 편향은 공정하지 못한 인공지능 모델을 생성
- ▶ 공정하지 못한 인공지능은 인종차별, 남녀차별 등의 문제를 일으킴

▶ 인공지능을 개발과 사용에 있어 윤리가 필요한 까닭

- ▶ 인공지능 개발자는 인공지능 모델을 학습시킬 데이터를 제공하고 제공되는 데이터는 불완전성과 결함을 가질 수 있음
- ▶ 인공지능은 사람의 손으로 만들어지기 때문에 사람의 생각이 반영될 수 밖에 없음
- ▶ 인공지능 사용자는 인공지능에 대한 맹목적 신뢰는 경계할 필요가 있음
- ▶ 공정한 인공지능을 만들기 위한 노력이 필요함

다음시간에 배울 내용

19

주	주제	온라인	오프라인
1	인공지능의 과거 현재와 미래	1. 강의 및 교과목 소개(공통, 핵심만) 2. 인공지능의 과거와 현재 3. 인공지능의 미래와 다양한 시선 4. 인공지능 개발환경 구축과 사용법(Anaconda/Colab)	1. 강의 및 교과목 소개(분반별, 자세히) 2. 다양한 인공지능 기술 경험하기 (자연어처리, 시각, 음성) 3. 인공지능 챗봇만들기(BM 왓슨 어시스턴트)
2	공공데이터를 이용한 사회문제 발견과 해결책 모색	1. 빅데이터의 정의와 가치 2. 공공데이터 수집하기 3. 공공데이터로부터 새로운 인사이트 발견하기 - 행정구역별 인구 데이터와 공공의료기관 현황 데이터 분석	1. 서울시 CCTV설치 현황 분석하기 2. 서울시 범죄발생 현황 분석하기
3	인공지능의 개요 및 머신러닝을 이용한 예측	1. 인공지능의 정의와 분류 2. 인공지능 학습방법 이해하기 3. 인공지능 알고리즘 소개	1. 머신러닝을 이용한 이미지 식별(구글 티쳐블 머신) 2. 머신러닝을 이용한 보스톤 집값 예측
4	인공지능과 데이터 윤리	1. 데이터의 불완전성과 결함에 따른 예측 오류와 차별 2. 데이터 편향성이 예측에 미치는 영향 (구글티쳐블머신) 3. 지도학습(SVM)을 이용한 타이타닉호 생존자 예측	1. 타이타닉호 생존자 예측 - 데이터 편향성이 예측에 미치는 영향 - 데이터 왜곡에 따른 예측 결과 비교
5	인공지능과 알고리즘 윤리	1. 알고리즘 기반 의사결정 시스템의 한계 2. 윤리가 적용된 인공지능 알고리즘	1. 알고리즘에 따른 예측 결과 비교 - 보스톤 집값 예측 - 폐암환자 생존 여부 예측
6	인공지능에 대한 다양한 이슈와 우리의 자세 고찰	1. 인공지능의 윤리적/법적 쟁점 (자율주행자동차, AI로봇, 트랜스 휴먼 등) 2. 인공지능시대 사회, 경제적 불평등 문제 3. 인공지능과 프라이버시 4. 인공지능의 윤리적 대응과 규제	1. 자율주행 자동차의 행동학습 시나리오 경험하기 2. 비윤리적 데이터 생성과 수집(웹 크롤링을 이용한 데이터 수집)
7		기말고사	