

인공지능을 위한 머신러닝 알고리즘

1. 머신러닝 개요

CONTENTS

1

인공지능과 머신러닝

2

머신러닝 알고리즘의 분류

3

일상 생활 속 머신러닝의 예

4

머신러닝의 구성 요소

학습 목표

- 인공지능을 달성하기 위한 수단으로써 머신러닝을 이해할 수 있다.
- 머신러닝 알고리즘의 종류와 차이점을 이해할 수 있다.
- 일상생활 속 머신러닝이 사용되고 있는 예제들을 파악할 수 있다.



1. 인공지능과 머신러닝

■ 인공지능의 정의

지능

문제를 해결할 수 있는 능력

지능 작업의 예

- 문서의 내용에 따라 항목 분류
- 병 진단
- 바둑

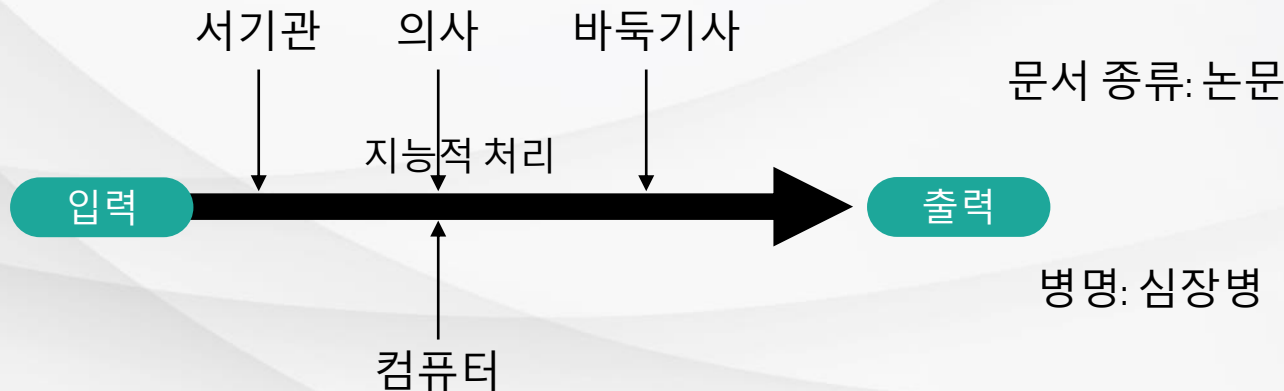
인공지능의 정의

인공지능

지능 작업을 수행할 수 있는 기계의 능력

In this paper, we proposed
the video story QA model
...

환자 I 번 기록
나이: 67
성별: 남자
가슴통증 종류: 무증상
혈압: 160mm Hg
혈중 콜레스테롤: 286mg/dl
혈당: < 120mg/dl
...



■ 인공지능의 역사



Alan Turing(1912~1954)

인공지능의 서막을 열다

낙관의 시대
AI 용어등장('56년)

암흑기
('70년중반~'80년초)

암흑기
('80년후반~'90년초)

인공지능,
바둑 챔피언을 이김
4차 산업혁명 첫 언급

2016

인공지능,
퀴즈 챔피언을 이김

2011

인공지능,
DARPA 그랜드 챌린지에서 우승

인공지능,
체스 챔피언을 이김

2005

1997

전문가 시스템

■ 인공지능 구현 방법: 합리주의자 vs. 경험주의자

◉ 지식의 근원은 어디서부터 왔을까?

◉ 지식공학 (합리주의자/이성주의자)

- 대표적 인물: **Mavin Minsky, Noam Chomsky, Descartes**
- **Top-down**
- 특정 분야의 전문가나 장인들이 학문 연구, 오랜 실무 경험으로 터득한 지식을 사람이 직접 컴퓨터에게 제공
- **1980년대** 전문가 시스템



데카르트

■ 인공지능 구현 방법: 합리주의자 vs. 경험주의자

◉ 머신러닝 (경험주의자)

- 대표적 인물: **Andrew ng, Pedro Domingos, David Hume**
- **Bottom-up**
- 컴퓨터가 데이터로부터 지식을 직접 학습
- 현재의 딥러닝



데이비드 흄

인공지능 구현 방법: 지식공학 vs. 기계학습

물체의 속성:

질량, 온도, 가속도,
부피, 색깔, ...

$f=ma$

Data



Program



Computer



Output

물체가 받는 힘

물체의 속성:

질량, 온도, 가속도,
부피, 색깔, ...

물체가 받는 힘

Data



Output



Computer



Program

$f=ma$

■ 기계학습과 농작물 재배

씨앗 = 알고리즘

영양분 = 데이터

농부 = 프로그래머

식물 = 프로그램





2. 머신러닝 알고리즘의 분류

■ 학습의 방법

❖ 지도학습 (Supervised Learning)

- ◉ 학습 데이터마다 레이블을 가지고 있음

❖ 비지도학습 (Unsupervised Learning)

- ◉ 학습 데이터가 레이블을 가지고 있지 않음

❖ 준지도학습 (Semi-Supervised Learning)

- ◉ 학습 데이터가 약간의 레이블을 가지고 있음

❖ 강화학습 (Reinforcement Learning)

- ◉ 최종 출력이 바로 주어지지 않고 시간이 지나서 주어지는 경우

■ 지도학습 (Supervised Learning)

◉ 주어진 입력-출력 쌍들을 매핑해주는 함수를 학습

◉ $D=\{X, Y\}$ 로부터 $F(X)=Y$ 를 만족시키는 함수 F 를 학습

- X 의 예> 물체의 속성들. 질량, 온도, 가속도, 부피, 색깔 등.. / 숫자 이미지 파일
- Y 의 예> 물체의 힘/숫자 레이블 (0~9)
- F 의 예> 뉴턴 운동 법칙/숫자 자동 인식기

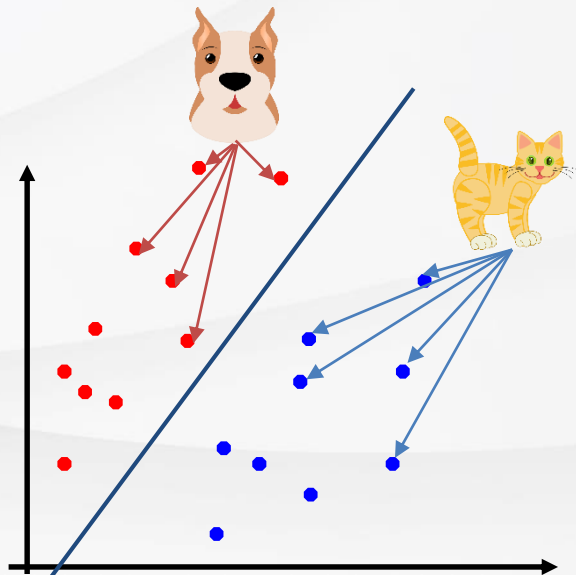
지도학습 (Supervised Learning)

◎ 새로운 데이터 \mathbf{x}' 의 출력을 함수 \mathbf{f} 를 사용하여 예측

- $\mathbf{f}(\mathbf{x})$ 가 이산적 (Discrete)일 때: 분류 (Classification) 문제
- $\mathbf{f}(\mathbf{x})$ 가 연속적 (Continuous)일 때: 회귀 (Regression) 문제
- $\mathbf{f}(\mathbf{x})$ 가 \mathbf{x} 의 확률 $\mathbf{p}(\mathbf{x})$ 일 때: 확률 추정 (Estimation) 문제

◎ 보다 정확한 학습을 할 수 있음

- 딥러닝이 대표적인 예
- 사용할 수 있는 데이터에 한계가 있음
- 데이터를 생성하는데 비용이 많이 듦



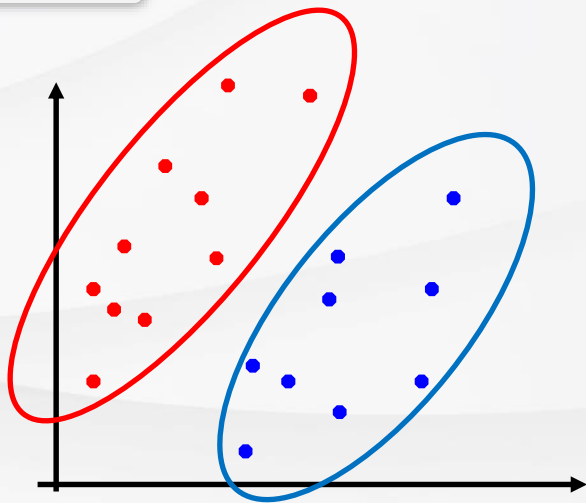
지도 학습의 예: 데이터에 레이블이 있을 때 개와 고양이의 분류 문제

■ 비지도학습 (Unsupervised Learning)

- ◉ 입력만 있고 출력은 없는 상태에서 이뤄지는 학습
 - 데이터에 내재되어 있는 고유의 특징을 탐색하고자 함
- ◉ $D=\{X\}$ 로부터 $F(X)=X$ 를 만족시키는 함수 F 를 학습
 - X 의 예> 유튜브 비디오
 - F 의 예> 비디오 항목 자동 분류기

■ 비지도학습 (Unsupervised Learning)

- 클러스터링이 주로 사용됨: 비슷한 데이터끼리 묶음
- 지도학습에 비해서 학습하기 어려움
- 우리 주변에 있는 대부분의 데이터는 레이블이 없음
 - 대표적인 예> 비디오
- 앞으로 머신러닝이 풀어야 할 중요한 숙제



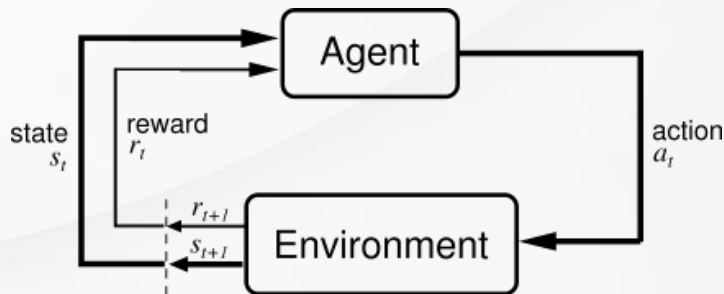
■ 강화 학습 (Reinforcement Learning)

- ◉ 결과(출력)이 바로 주어지지 않고 시간이 지나서 주어지는 경우
 - 예> 바둑: 승/패의 결과(**Final Outcome**)이 바둑 기사(**Agent**) 한 수 두자마자(**Action**) 주어지지 않고 시간이 지나고 나서 주어짐, 바둑 기사는 매 순간 바둑판의 상황(**State**)을 읽고 어떤 수를 두어야 할지 고민함 (**Agent**가 받는 **Reward**를 최대화하는 쪽으로)
 - 게임, 미로 찾기 등...
 - 어떤 **Action**이 최종 출력에 영향을 주었는지 불명확한 문제에 사용됨

■ 강화 학습 (Reinforcement Learning)

● 컴퓨터가 계산해야 할 중요한 이슈

- 매 순간 어떤 **Action**을 선택해야 하는가?
- **State**에 대한 평가



● 우리 주변에 있는 많은 문제들이 강화 학습으로 풀어야 할 문제들

- 예> 대화 - 상점에서 주문, 고객 상담 등..
- 좋은 대학에 입학하기 위한 전략은?
- 축구를 잘 하는 로봇을 만들려면?

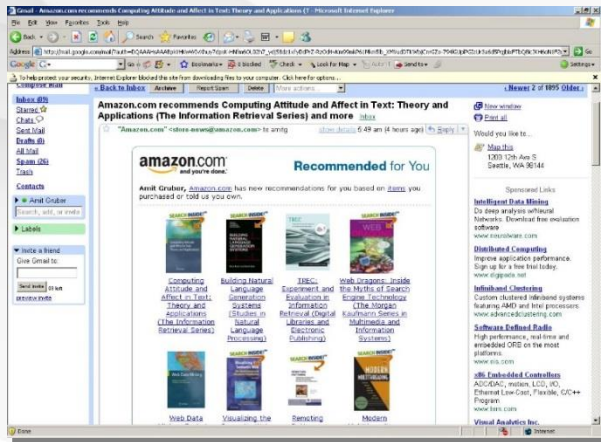


3. 일상 생활 속 머신러닝의 예

Netflix의 추천시스템

❖ 온라인 DVD 대여 업체

- ◉ 10만 개의 영화 보유
- ◉ 고객들이 Netflix가 보유하고 있는 영화를 골고루 볼 수 있도록 관리를 해야 함
- ◉ 고객의 취향에 맞는 비디오 추천
- ◉ 고객의 평점, 구매 내역을 활용



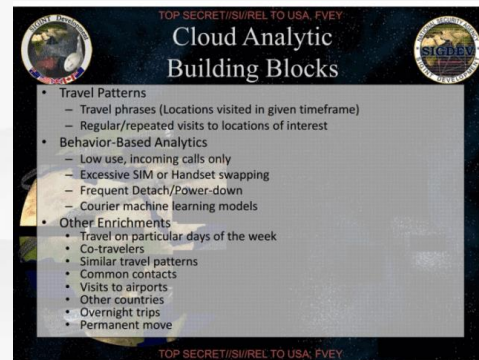
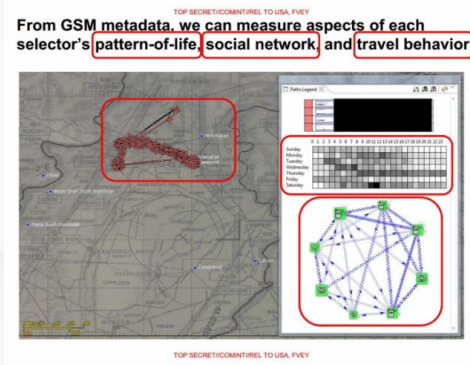
4만 명의
고객

10만 개 영화

1	3	4			??
1	3	5			5
		4	5		5
			3		
			3		
2				2	2
					5
	2	1			1
	3			3	
1					

■ 미국 국가안보국의 SKYNET

- 파키스탄의 테러리스트를 식별하여 사살하기 위한 프로그램
- 2004년부터 2,500명에서 4,000명 가량의 ‘식별된’ 사람들이 사망
- 5,500만 명의 파키스탄 사람들의 휴대전화 기록을 활용
 - 전화 시간, 길이, 수신자/발신자 정보, **GPS**, 이동 기록 등..
 - 전화기를 끄거나 **SIM** 카드를 바꾸면 별도 관리 대상이 됨
- 잘못된 알고리즘 설계로 무고한 많은 사람들이 희생되어 비판 받음



■ 다양한 적용 사례들

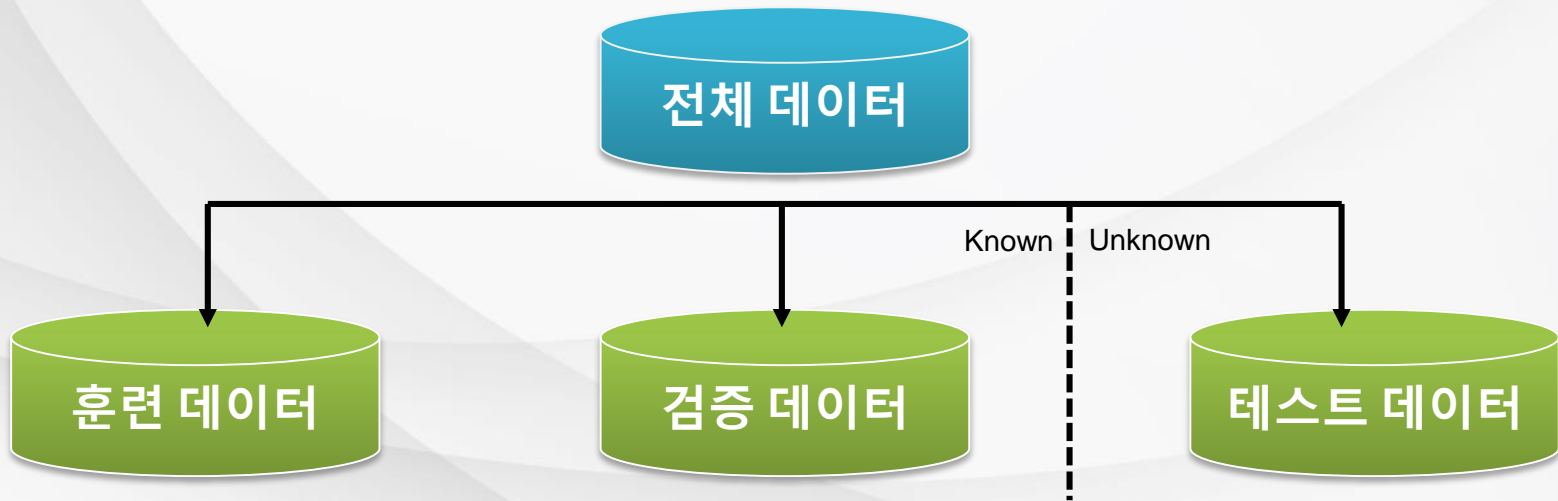
응용 분야	적용 사례
인터넷 정보검색	텍스트 마이닝, 웹로그 분석, 스팸필터, 문서 분류, 여과, 추출, 요약, 추천
컴퓨터 시각	문자 인식, 패턴 인식, 물체 인식, 얼굴 인식, 장면전환 검출, 화상 복구
음성인식/언어처리	음성 인식, 단어 모호성 제거, 번역 단어 선택, 문법 학습, 대화 패턴 분석
모바일 HCI	동작 인식, 제스처 인식, 휴대기기의 각종 센서 정보 인식, 떨림 방지
생물정보	유전자 인식, 단백질 분류, 유전자 조절망 분석, DNA 칩 분석, 질병 진단
바이오 메트릭스	홍채 인식, 심장 박동수 측정, 혈압 측정, 당뇨치 측정, 지문 인식
컴퓨터 그래픽	데이터기반 애니메이션, 캐릭터 동작 제어, 역운동학, 행동 진화, 가상현실
로보틱스	장애물 인식, 물체 분류, 지도 작성, 무인자동차 운전, 경로 계획, 모터 제어
서비스업	고객 분석, 시장 클러스터 분석, 고객 관리(CRM), 마케팅, 상품 추천
제조업	이상 탐지, 에너지 소모 예측, 공정 분석 계획, 오류 예측 및 분류



4. 머신러닝의 구성 요소

■ 데이터 준비

❖ 훈련/검증/테스트 데이터 분리



- 모델 학습을 위해 사용하는 데이터

- 모델 검증을 위한 데이터
- 검증 데이터에 대한 성능에 따라 모델을 바꿔보고 통상적으로 성능이 제일 높은 모델을 선택

- 현재 우리가 갖고 있지 않지만 실제 맞닥뜨릴 데이터
- 학습된 모델의 최종 성능 검사를 위해 따로 구별해 놓은 데이터

■ 모델 표현의 방법

❖ 의사 결정 트리

- ◉ 기호주의자 (**Symbolists**)
- ◉ 귀납적 추론, 철학과 심리학, 논리학에서 아이디어를 얻음

❖ 신경망 기반

- ◉ 연결주의자 (**Connectionists**)
- ◉ 두뇌를 분석하고 모방하며 신경과학과 물리학에서 영감을 얻음

❖ KNN, 서포트 벡터 머신

- ◉ 유추주의자 (**Analogizers**)
- ◉ 유사성 판단을 근거로 추정하면서 배우며 심리학과 수학적 최적화의 영향을 받음

■ 모델 표현의 방법

❖ 베이지안 모델

- ◉ 베이즈주의자 (**Bayesians**)
- ◉ 학습이 확률 추론의 한 형태라고 믿으며 통계학에 뿌리를 두고 있음

❖ 유전 알고리즘

- ◉ 진화주의자 (**Evolutionaries**)
- ◉ 컴퓨터에서 진화를 모의시험하며 유전학과 진화생물학에 의존

❖ 모델 앙상블

■ 모델 평가 방법

❖ 에러의 제곱 (Squared Error)

$$\sum\{y - f(x)\}^2$$

❖ 정확도 (Accuracy)

- ◉ 맞힌 테스트 데이터 개수 / 전체 테스트 데이터 개수

❖ 우도 (Likelihood)

$$\log_2(P(y))$$

❖ 정밀도와 재현률 (Precision and Recall)

- ◉ 정보 검색에서 주로 사용

❖ 엔트로피 (Entropy)

❖ 사후 확률 (Posterior Probability)



학습정리

지금까지 [머신러닝 개요]에 대해서 살펴보았습니다.

인공지능과 머신러닝

머신러닝은 인공지능을 구현하기 위한 하나의 방법
지식을 구현하기 위한 두 가지 대립적 방법
지식공학 (합리주의자/이성주의자): 지식을 사람이 직접 컴퓨터에게
제공

머신러닝 (경험주의자): 컴퓨터가 데이터로부터 지식을 직접 학습

머신러닝 알고리즘의 분류

지도학습: 학습 데이터마다 레이블을 가지고 있음
비지도학습: 학습 데이터마다 레이블을 가지고 있지 않음
준지도학습: 학습 데이터가 약간의 레이블을 가지고 있음
강화학습: 최종 출력이 바로 주어지지 않고 시간이 지나서 주어지는
경우

머신러닝의 구성 요소

데이터 준비: 훈련/검증/테스트 데이터 집합으로 분리
모델 표현: 기호/신경망/유추/베이지안/유전 기반 알고리즘 존재
모델 평가 방법: 정확도/에러 제곱/우도 등 각 상황에 맞는 평가 방법 선택