

The Master Algorithm

<https://product.kyobobook.co.kr/detail/S000001868277>

"내가 창조하지 못하는 것은 이해하지 못한다." 리처드 파인만

알고리즘 설계 > 자바나 파이썬과 같은 컴퓨터 언어로 변경 > 디버깅 작업

알고리즘은 알고리즘과 결합하여 이용하고 새로운 결과를 추출

알고리즘의 문제

- : 공간 복잡성 - 컴퓨터 저장 장소에 보관할 정보량의 크기
- : 시간 복잡성 - 알고리즘을 수행하는데 걸리는 시간

머신러닝

- : 데이터를 많이 얻을수록 더 많이 배울 수 있음
- : 복잡성 괴물을 베어 내는 칼
- : 패턴 인식, 통계 보형, 데이터 사이언스, 데이터 마이닝, 적응형 시스템, 자기 조직 시스템 등으로 지칭
- : 중개인과 같은 역할 실행

머신러닝은 멋진 신기술이다. 하지만 멋진 신기술이라서 채택하는 것이 아니라 채택하지 않으면 기업이 생존 할 수 없기에 채택한다.

세상의 모든 지식은 단 하나의 보편적 학습 알고리즘으로 데이터에서 얻어 낼 수 있다. ==
마스터 알고리즘

지나친 단순화 경향은 수학의 한계가 아니라 인간 지적 능력의 한계

머신러닝은 수학의 효율성이 데이터의 효율성과 만날 때 얻을 수 있음

베이즈 정리

- : 마스터 알고리즘이거나 마스터 알고리즘을 구동하는 엔진
- : 빅데이터와 빅컴퓨팅이 만나고 광대한 가설의 세계에서 자리를 찾은 것

머신러닝 / 지식공학

- : 학습과 지식은 놀라울 정도로 뒤얽혀 있음

데이터는 인간의 직관을 대체 할 수 없다

== 인간의 직관은 데이터를 대체 할 수 없다. but 증거가 있다고 거부할 이유는 없다.

마스터 알고리즘의 불안

- : 방어선은 좋은 사람이 먼저 마스터 알고리즘을 얻도록 확실한 길을 마련하는 것
- : 머신러닝이 아무리 좋다고 하더라도 사용할 데이터가 좋은 만큼 좋아짐. 데이터를 지배하는 자가 머신러닝을 지배함

기술 사용 위해 훌륭한 개념 모형 확보 필수

마스터 알고리즘 분야에서 사람이 더 잘 할 수 있을까?

- : 세상을 완전히 알아내기에 충분한 데이터 확보 어려움. == 불확정성 원리
- : 어느 시점에 세상에 관한 완전한 지식을 갖는다 하더라도 물리 법칙은 세상의 과거와 미래를 알려주지 않음

기호주의자 / 연결주의자 / 진화주의자 / 베이즈주의자 / 유추주의자

합리주의자 vs 경험주의자

지식 없는 학습은 없다. == 데이터만으로 충분하지 않음.

머신러닝은 지식펌프로 마중물(데이터) 필수

좋은 머신 러닝은 무지와 환각 사이 좁은 길을 영원히 걸어야함

과적합 문제 == 잡음에 의해 심각

단순한 가설에 대한 선호 == 오컴의 면도날

역연역법을 사용하여 새로운 규칙 유도를 반복 == 지식의 선순환.

: 과적합 문제와 계산 비용에 의한 제한 유일

연역연역법 한계

: 계산량이 많아 대용량 데이터 처리 불리 > 기호주의자가 의사결정트리 이용한 귀납법 사용

헵의 규칙

볼츠만 기계

: 신뢰할당문제 이론으로 풀 수 있음

: 학습 느리고 어려움 많아 응용 분야 적용에 비실용적

S자 곡선

: 신뢰 할당 문제에 새로운 해결책

퍼셉트론 알고리즘

: 오류신호는 있음 혹은 없음 - 많은 신경세포로 구성된 신경망일때 비효과적

살아있는 세포는 비선형 시스템의 정수

딥러닝

- : 겹겹히 쌓아 올린 자동부호기
- : 볼츠만 기계 기반
- : 대뇌 피질 시각 영역에 대한 모형을 기반으로 하는 합성곱 신경망

유전 알고리즘의 핵심 입력 = 적합성 함수

역전파보다 국소 최적값에 머무는 경우가 덜하고 원리적으로는 새로운 것을 잘 도출해냄

머신러닝은 폭발적인 조합의 양을 처리시 유리

유전 알고리즘의 유요한 점

- : 각 비트 열이 기하급수로 늘어날 가능성을 내재한 스키마라는 구성요소 보유

머신러닝과 삶에서 가장 중요한 문제는 탐험과 개발 사이의 딜레마

진화주의자 / 연결주의자

- : 자연에서 영감받아 학습 알고리즘 설계
- : 학습구조 집중 / 간단한 구조 선호, 가중치 학습이 모든 일 하도록 요구

볼드윈 효과

설명적 이론 / 규범적 이론

"자연을 움직이는 모든 힘과 사물이 만들어 낸 각각의 상황으로 모두 이해하는 지성이 충분히 거대하여 이 모든 데이터를 분석할 수 있다면 우주의 가장 커다란 천체들과 가장 가벼운 원자들의 운동을 같은 공식으로 파악할 것이다. 그러한 지성에게는 불확실한 것이 없으며 과거와 마찬가지로 미래도 명확히 드러날 것이다." 라플라스

무차별성의 원리 == 이유 불충분의 원리

구글의 페이지랭크

: 마르코프 연쇄

베이즈 네트워크 == 생성 모형

베이즈 네트워크의 문제 해법

: 네트워크 구조에 폐회로 없다고 가정하고 각 지점의 확률이 수렴될 때까지 계속 앞 뒤로 움직이는 것 == 신뢰전파

머신러닝에서 가장 중요한 착상이 등장하는 자리 = 유추

게으른 알고리즘 (최근접 알고리즘)

: 의사결정트리와 같은 일반 모형을 만드는 일이 단지 한 번에 하나씩 특정한 지점들이 어디에 있는지만 알아내는 것보다 훨씬 어렵기 때문

: 지름길 선택

최근접 알고리즘의 문제

: 차원의 저주

: 속성 대부분이 연관성 없음 > 속성의 가중치 구하기로 보완 / 조건부 최적화

사례 기반 추론법

EM알고리즘 = Expectation Maximization

k평균 알고리즘

PCA = 주요 성분 분석

비선형 차원 축소하는 알고리즘 = 이소맵

효과의 법칙 - 에드워드 손다이크

강화학습

청킹

메타학습 알고리즘

== 스택킹 / 배킹 / 랜덤포레스트 / 부스팅 /

alchemy.cs.washington.edu

필터 버블

4종류의 데이터

- : 모든 사람과 공유하는 데이터
- : 친구와 동료하고만 공유하는 데이터
- : 여러 회사와 공유하는 데이터 (일부러 하는 것과 모르고 하는 것)
- : 공유하지 않은 데이터

인공지능이 우려되는 일

- : 인공지능이 악당의 수중에 들어가는 경우
- : 인류가 자진하여 통제권을 내놓는 경우
- : 기계가 우리가 원하는 것 대신 우리가 요구하는 것을 곧이곧대로 줄거라는 점

사람들은 컴퓨터가 너무 똑똑해져서 세상을 지배할 거라고 걱정하지만 실제로 나타난 문제는 컴퓨터가 너무 명청하고 그런 컴퓨터가 세상을 지배하고 있다는 점