

'수학의 쓸모' 정리

작성일: 20.06.22

작성자: 도원진

1. 추천엔진의 핵심

- 개인화
- 조건부확률: P(부분 요소 | 전제)

조건부확률을 전쟁에서 활용한 인물: 에이브러햄 왈드(Abraham Wald)

- 헝가리계 미국 통계학자
- 2차 세계대전의 비행전투에서 조건부확률을 통해 해결

주어진 과제

- 살아돌아오지 못한 비행기로 인해 총상 부위에 대한 데이터를 얻을 수 없었음

해결책

- 여러 조건들을 고려하여 실제 일어날 법한 모형(모델)을 가정함.
- 이를 통해 부족한 데이터 즉, 격추당한 비행기에 대한 데이터를 확보

$P(\text{동체 손상} \mid \text{귀환}) = 33\%$

$P(\text{귀환} \mid \text{동체 손상}) = 105 / (105 + 8) = 93\%$

2. 패턴과 예측

2.1. 패턴 발견의 계기. 안드로메다 은하

1924년까지 태양계가 속한 '우리은하'(=은하수, 밀키웨이)만 유일하다고 굳게 믿음

1. 9세기(800~900)경

- '안드로메다 은하'발견 (당시에 단순한 성운으로 오해함)

2. 1600년대

- 망원경 발명
- 다른 은하 추가발견
- 나선형위주
- 성운으로 굳게 간주함(과학자의 오해)
- 참고)

- 성운: 성간 물질이 모여 있어 마치 구름처럼 보이는 천체(성간물질 무리)

3. 1900년대

- 기존의 이론을 의심하기 시작
- 고성능 망원경 발명
- 신성이 생기는 수가 우리은하에서 발견되는 수보다 압도적으로 큼

4. '안드로메다 성운'의 이동속도 측정

- 도플러 효과를 이용한 분광계를 통해 300km/s 로 이동하는 것을 발견
- '우리은하'의 그 어떤 것 보다 20배 빠른 속도
- 대다수의 다른 은하는 1000km/s로 이동한다는 사실을 추가로 발견
- 그럼에도 논쟁은 1910년 ~ 1920년 동안 끊임없이 이어짐

5. '안드로메다'와 거리를 재려는 시도

- 측정방법
 - 시각차를 이용한 물체와 관찰자간의 거리재기
 - 시각차이로 인해 물체의 관측위치가 이동됨 (=시차)
 - '시차'와 '관측거리'는 반비례관계임
- 한계
 - 우리은하의 너비만 약 10만 [광년](#)
 - 300광년이상일 경우 시차가 3cm밖에 안됨
 - 눈으로 관측하기 어려운 시차임

6. 헨리에타 레빗

- 맥동변광성 조사
 - 뜻 : 주기를 갖고 밝기가 변하는 별
- 1912년
 - '세페이드 변광성' 별집단(성단)발견(25개의 별로 구성됨)
 - 성단(星團, star cluster)은 [중력](#)으로 뭉쳐 있는 [별](#)들의 무리이다(별무리)
 - 참고)
 - 성운 : 성간 물질이 모여 있어 마치 구름처럼 보이는 천체(성간물질 무리)
 - 모두 관측자의 위치에서 **같은 거리**에 있다고 판단
 - 같은거리에 있는 25개의 별들은 밝기가 제각각이며 주기도 제각각이다.
 - 이를 통해 그래프를 그림(X축: 주기, Y축: 밝기, 그래프 선위의 점: 해당 별)
 - 주기가 길수록 밝다(비례관계)
 - 할로새플리
 - 은하수의 폭 거리 측정
- 우리은하(은하수)의 존재하는 각 별들의 절대밝기를 측정
 - 이후, 별과 별 사이의 거리 측정
- 이 거리들을 종합하여 은하수의 폭이 적어도 10만 Light year라는 것을 발견
- 1923년
 - 허블
 - 1919년 부터 나선형 성운들의 맥동변광성을 꾸준히 측정
 - 레빗의 규칙을 이용해 지구~ 안드로메다 사이의 거리 측정
 - 100만Light year

2.3 인공지능 시대가 늦은 이유

1. 개념은 이미 1805년에 나옴

2. 하드웨어 기술력의 한계

- 패턴을 기술하는 방정식에 많은 파라미터가 요구됨.
- 컴퓨터가 연산하기에 무리가 있음. 하드웨어적인 기술력이 뒷받침되지 않았던 현실
- 예) 100만화소
 - RGB 3가지 색상으로 300만개의 숫자가 관련됨.

3. 2014년에 구글이 딥러닝을 통해 해결

- 신경망
- 38만8736개 매개변수 도입
- 15억번 연산진행
- 현재 Nvidia그래픽카드는 15억 * 1000 번 연산을 1초에 진행함.

3. 데이터의 홍수에서 살아남기

3.1. 로봇이 가장 먼저 인지해야 할 질문

"나는 어디에 있는가"

- 자신이 어디에 있는지 안다는 것
 - 1. 낯선 주위 환경에 대해 지도를 머릿속으로 생각할 수 있다는 것
 - 2. 그 지도에서 나의 위치를 추론할 수 있다는 것
- 베이지 규칙을 통해 해결함.

3.2. 베이지 규칙

3.2.1. 사용배경

- 1968년)
 - 미국, USS 핵잠수함 침몰
 - 이를 수색하기 위해 베이지언 탐색 진행

3.2.2. 베이지언 검색 과정

1. 사전믿음
 - 지도의 칸을 나눈 구간에 타겟이 발견될 확률을 적는다.
 - 가장 높은 발견확률을 가진 칸을 수색한다.
2. 데이터 추가
 - 수색후, 발견이 되지 않으면 주변 확률을 높여 업데이트를 한다.
3. 수정된 믿음

3.3. 베이지 규칙 으로 움직이는 자율주행차

1. GPS로 자율주행하면 되지 않을까?

- 5미터 안으로만 정확하다.
- 터널이나 높은 건물근처에서는 오차가 30~40미터로 늘어난다

2. 추측항법을 이용하는 자율주행

- 내관 및 외삽 항법 이라고도 불림
- 내관
 - 슬력, 바퀴각도, 가속도 등 내부상태 정보 수집
- 외삽
 - 내관 정보를 통해 물리법칙을 적용한 후 다음 짧은 시간동안 자동차의 움직임을 예측

3. 베이지 규칙 순서

- 내관 정보를 통해 사전 믿음 (확률)을 만든다.
- 외관 정보를 통해 데이터 (확률)를 얻는다.
- 사전 믿음 과 데이터 가 갖는 각각의 확률을 종합한다.

3.4. 통상적인 사전확률과 조건부확률(사후확률)은 엄연히 다르다

• 1000명의 유방조영술 실험자

- 실제 10명이 암에 걸림
 - 유방조영술로 오진 : 2건
 - 유방조영술로 진단: 8건
- 실제 990명은 건강
 - 유방조영술로 오진 : 100건
 - 유방조영술로 진단: 890건

• 사전확률 VS 사후확률

- $P(\text{유방암} \mid \text{유방조영술 양성판정}) = 8/108 = 7.4\%$ (사후확률)
- 유방조영술의 정확도 80% (사전확률)
- 사전확률 "유방조영술 정확도 80%"의 의미
- 실제 암에 걸린 경우 80%를 발견함

• 결론

- 올바른 사후확률은 사전확률과 데이터를 결합하여 얻는다.
- 사전확률 : 80% 진단율
- 데이터 : 1000명의 실험자 및 10명의 유방암 환자 (1%의 유방암 발생율)
- 사후확률 : 7.4%

4. 디지털 비서와 대화하는 법

4.1. 프로그래밍언어(그레이스 호퍼)

• 규칙

1. 문법과 어휘 한정
2. 최대한 많은 번역규칙을 기계에 프로그래밍한다.

- 별명(주소)를 통한 명령
 1. 설명(기술)보다는 서브루틴에 번호를 매긴다.
 2. 일일이 부호로 적지 않는다.

4.2. 자연언어 혁명

- 프로그래밍 전략은 자연언어에서 통하지 않는다.
- 언어 인지도가 37%밖에 되지 않는다.
- 이유
 1. 너무 많은 규칙
 2. 견고성 부족
 3. 언어의 모호성

4.2.1. 단어벡터 도입

- 모든 단어를 수치로 기술해주는 모형
- 단어가 대체로 어떤 문맥에서 사용되는지 그 정도를 파악하는 도구

대상 \ 질문	동물	기분이 좋다	말을한다	런던에 거주	곰이다
스크루지	1	0	1	1	0

- 단어들의 조합으로 갖춰진 문장 300개를 기계에게 학습시킨다.
- 덧셈, 뺄셈으로 언어를 표현하기

예) 각국 수도 : 런던 - 영국 + 이탈리아 = 로마

단어 시제 : captured - capture + go = went