동국대학교

dongguk UNIVERSITY

2020년 - 데이터 청년 캠퍼스

데이터사이언스 기반 지능소프트웨어 과정

데이터 사이언스 개론

4) 강화 학습 (Reinforcement Learning)

학습 내용

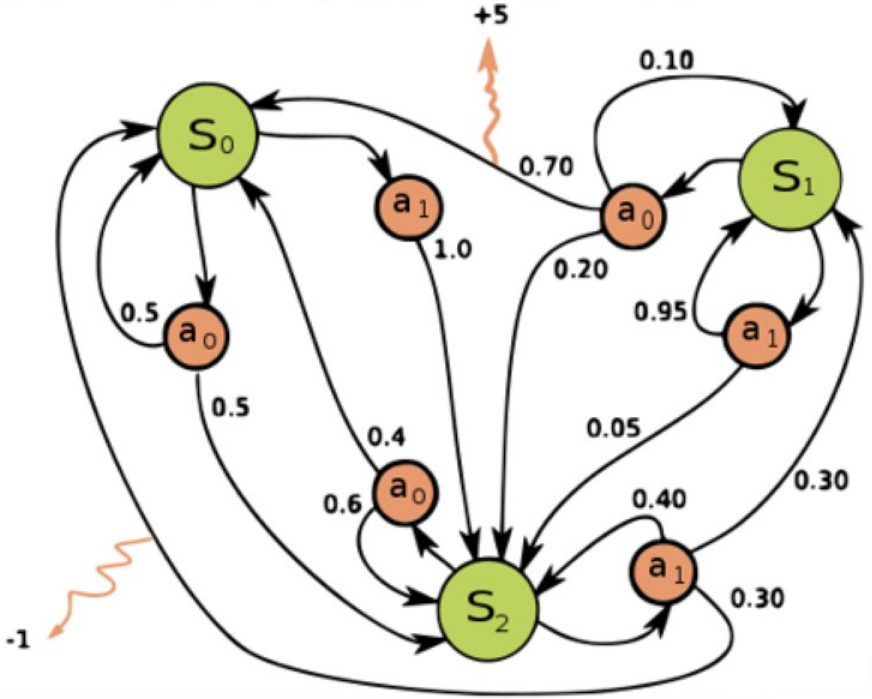
- **비지도 학습 (Unsupervised Learning)**
  - K-평균 군집화 (k-means Clustering)
  - ...
- **지도 학습 (Supervised Learning)**
  - 회귀 분석 (Regression Analysis)
  - ...
- **강화 학습 (Reinforcement Learning) : 1가지**
  - 멀티-암드 밴딧 (Multi-Armed Bandits)

강화 학습 (Reinforcement Learning)

- 데이터 셋 (Dataset) 의 패턴 (Pattern) 을 분석하고 추가되는 데이터로부터 정확도를 개선

- 비지도 / 지도학습은 계산된 결과가 틀리더라도 모델이 변경되지 않지만, 강화학습은 모델이 변경하기도 함

- 마르코프 결정 과정 (Markov Decision Process)
  - 어떤 행동 (a_n) 에 따라 상태 (S_n) 가 변함



강화 학습 (Reinforcement Learning)

■ 다음 광고 중 좀 더 효과적인 문구는 무엇일까?

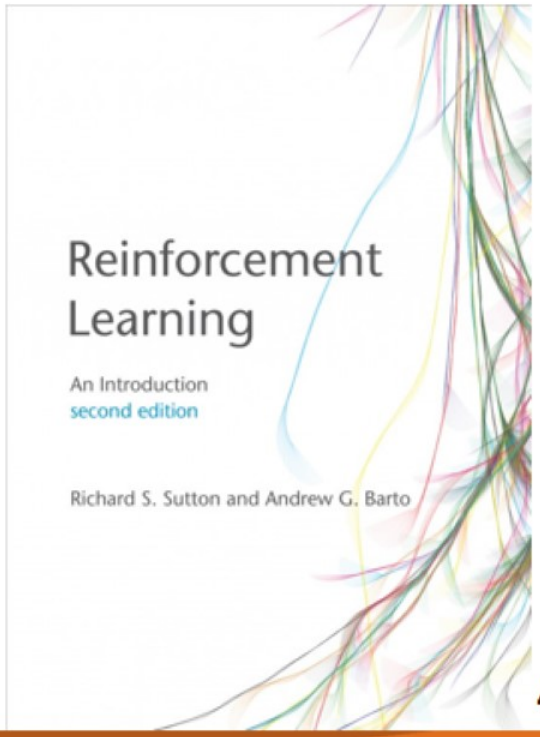
- 최대 50% 할인 / 최대 반값 할인
- 최대 100% 증정 / 최대 2배 증정

- 효과적인 문구를 선택해야 광고의 효율이 증가함 (A / B Test)
- 만약 잘못된 문구를 선택하게 된다면?

■ Reinforcement Learning : An Introduce

- Richard S. Sutton and Andrew G. Barto
- The MIT Press

2. Multi-armed Bandits



멀티-암드 밴딧 (Multi-Armed Bandits)

- 여러 개의 One-Armed Bandit (OAB) 중,
  전략을 어떻게 구성해야 최선의 지불률을 낼 수 있는가?

- 두 개의 One-Armed Bandit 의 지불률은 다음과 같음

머신	지불률
A	0.5
B	0.4

- 하지만 당연히 이 사실은 모름!!!

멀티-암드 밴딧 (Multi-Armed Bandits)

게임 규칙

- OAB 을 총 2,000번 할 수 있음
- 한 번 맞추면 $1 획득

전체 탐색

- 두 개의 OAB 을 무작위로 사용
- 평균 $900 을 획득

전체 활용

- 내부 OAB 정보를 훔침
- 평균 $1000 을 획득
- 거의 불가능

멀티-암드 밴딧 (Multi-Armed Bandits)

A / B 테스트 (A / B Test) - 1

- 먼저 각각의 OAB 에서 100번씩 게임을 진행
  (탐색, Exploration)
- 그 중 지불율이 높은 OAB 에 나머지 1,800 게임을 진행
  (활용, Exploitation)
- 평균 $976 을 획득
- 경험상 8% 의 확률로 오인할 수 있음

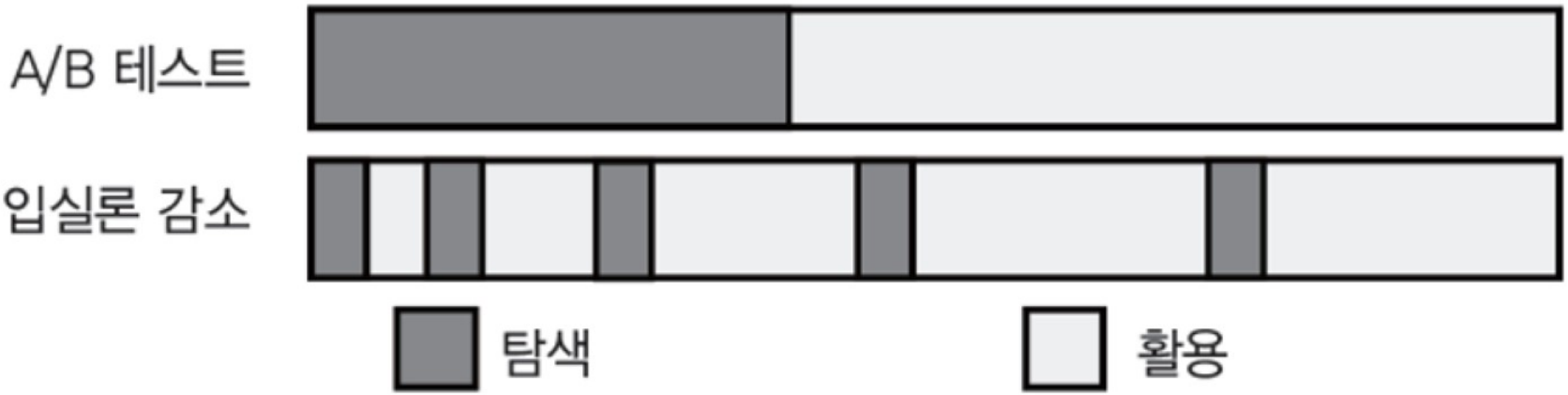
A / B 테스트 (A / B Test) - 2

- 250번씩 총 500 게임을 진행 (탐색)
- 나머지 1,500 게임을 진행 (활용)
- 경험상 1% 의 확률로 오인할 수 있음
- 평균 $963 을 획득

멀티-암드 밴딧 (Multi-Armed Bandits)

■ 입실론 감소 전략 (Epsilon-Decreasing Strategy)

- 탐색과 활용을 적절한 비율로 번갈아 가면서 진행
  (= epsilon + (1 – epsilon))
- 평균 $984 을 획득
- 경험상 4% 의 확률로 오인할 수 있음



멀티-암드 밴딧 (Multi-Armed Bandits)

입실론 감소 전략의 장점

- 최적의 비율 (epsilon) 을 통하여 최고의 수익을 얻을 수 있음

입실론 감소 전략의 단점 (인터넷 광고와 비교)

- OAB 의 지불률은 변하지 않지만, 인터넷 광고는 시간에 따라 클릭률이 감소 할 수 있음
- OAB 의 지불률은 각 게임별로 독립적이지만, 인터넷 광고는 광고 횟수와 관련이 있을 수 있음
- OAB 의 결과는 즉각적으로 발생되지만, 이메일 광고 (Spam mail) 은 결과가 즉각적으로 발생하지 않음


Q & A