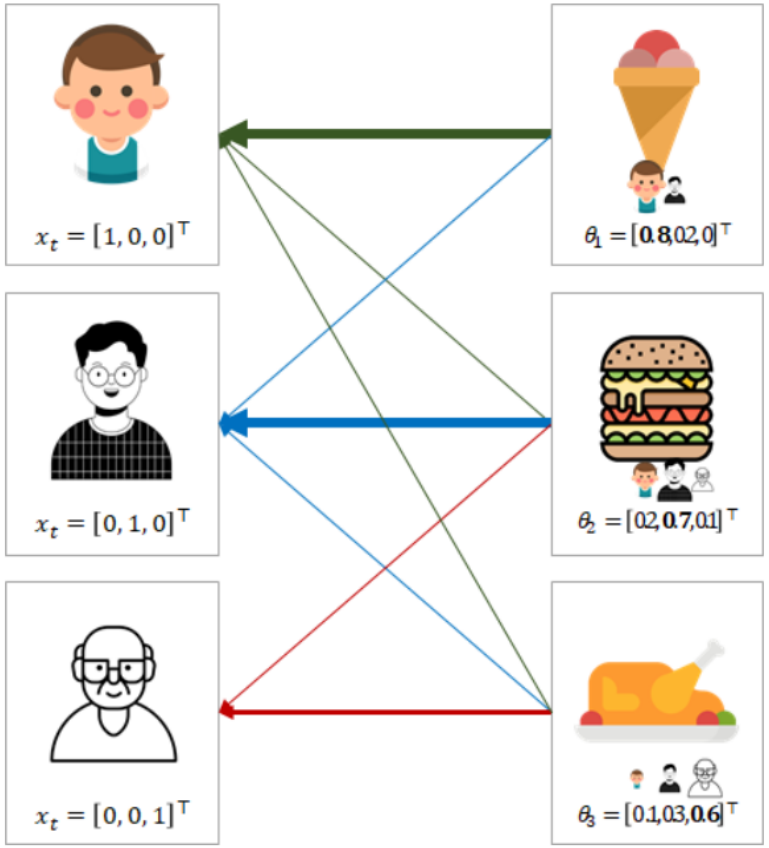
A Contextual-Bandit Approach to Personalized News Article Recommendation

웹 서비스들은 콘텐츠와 고객의 정보를 모두 사용한 서비스를 제공하고 싶어함.

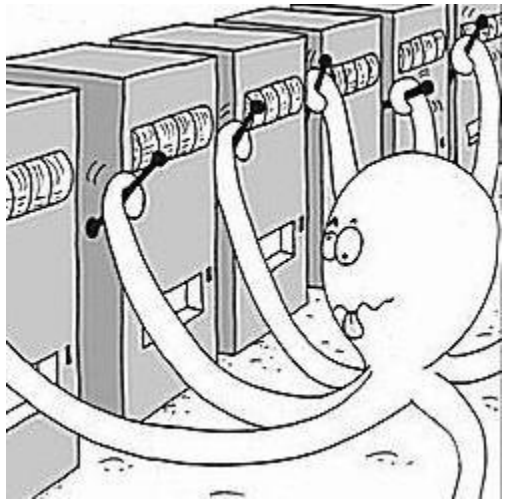
문제점
1. 콘텐츠가 동적으로 변화하기 때문에 기존의 협업 필터링과 같은 방법을 사용할 수 없음.
2. 학습과 연산 모두 빠른 방법을 요구한다.

-> Contextual bandit
유저와 기사에 대한 상황 정보를 기반으로 제공할 기사를 순차적으로 선택
동시에 유저 클릭 피드백을 기반으로 기사 선택 전략을 조정하여 총 사용자 클릭 수를 최대화 한다.

Contextual Bandits 이란?
매 trial의 의사결정 전에, 관측 가능한 정보를 활용하여 최적의 의사결정을 하는 bandits



Multi-Armed Bandits?



N개의 슬롯머신이 존재할 때, 각 슬롯머신의 수익률은 다르며 고객은 각 슬롯머신의 수익률을 모른다.
이때, 고객은 돈을 어느 슬롯머신(bandit)에 걸고 손잡이(arm)를 당겨야 할까?
카지노에는 여러 슬롯머신이 존재하기 때문에 Multi-Armed Bandits 문제이다.

Exploration and Exploitation(탐색과 활용 or 탐험과 탐사)

Exploration: 정보 수집을 위해 무작위로 arm을 선택하는 것.
Exploitation: 충분한 정보 수집 후, 최적의 arm을 선택하는 것.

전략1. greedy
한 번씩 플레이 해보고, 가장 점수가 좋은 슬롯에 올인.
-> 탐험이 충분히 이루어지지 않음.

전략2. e-greedy
탐험이 부족한 greedy 전략을 보완한 전략.
동전을 던졌을 때, 앞면이 나오면 점수가 좋은 슬롯을 선택(활용), 뒷면이 나오면 랜덤하게 선택(탐색).
동전의 앞면이 나올 50%의 확률이 입실론 이라는 하이퍼파라미터.

전략3. UCB(Upper-Confidence-Bound)
좋은 수익률을 보이며 최적의 선택이 될 가능성이 있는 슬롯머신 선택.
탐험을 할 때 랜덤이 아닌, 최적이 될 수 있을만한 가능성을 수치로 계산하여 가장 가능성이 있는 슬롯 머신을 선택.

https://brunch.co.kr/@chris-song/62
http://sanghyukchun.github.io/96/
https://yjjo.tistory.com/38?category=982145

UCB 알고리즘

e-greedy 알고리즘은 과거에 관측된 reward를 고려하지 않아 비효율적인 탐색방법.
UCB 알고리즘은 현재까지 테스트 해본 arm의 평균값과 시행횟수를 이용해 모평균이 어느 범위에 있을지 추정하고, 그 범위의 상한이 가장 높은 arm을 선택하는 알고리즘. 시행횟수가 적은 arm은 범위가 넓어지고, 시행횟수가 많은 arm은 범위가 좁아지도록 하여 탐색과 이용을 적절히 분배 가능.

$$A_t \doteq \arg\max_a \left[Q_t(a) + c\sqrt{\frac{\log t}{N_t(a)}} \right]$$

← 해당 슬롯머신이 최적의 슬롯머신이 될 수도 있는 가능성

t: 모든 슬롯머신을 선택한 횟수의 합(시점)
$Q_t(a)$: t시점까지 슬롯머신 a의 누적 보상
c: 탐색의 정도를 결정하는 하이퍼 파라미터(크면 탐색을 많이, 작으면 활용을 많이 수행)
$N_t(a)$: t시점 전까진 해당 슬롯머신을 선택했던 횟수

랜덤하게 탐색을 하는 것이 아니라 시간이 흘렀음에도 적절한 탐색이 이루어지지 않았다면 가중치를 준다.

https://soobarkbar.tistory.com/135
https://yjjo.tistory.com/42#recentComments