Cognitive biases and bounded rationality

RL v.s IRL

Goal	Key tasks	Optimality?	Sub-fields	Fields
Solve practical decision problems	1. Define appropriate utility function and decision problem. 2. Solve optimization problem	If it's tractable	RL, Game and Decision Theory, Experimental Design	ML/Statistics, Operations Research, Economics (normative)
Learn the preferences and beliefs of humans	1. Collect data by observation/experiment. 2. Infer parameters and predict future behavior	If it fits human data	IRL, Econometrics (Structural Estimation), Inverse Planning	ML, Economics (positive), Psychology, Neuroscience

Table 1: Two uses for formal models of sequential decision making. The heading "Optimality" means "Are optimal models of decision making used?".

Limitations of optimality

- 최적화에 근거한 (PO)MDP 문제 해결
- 하지만, 사람은 항상 최적으로 살지 않는다 (deviate from optimality)
- 그렇다면, 최적화에 근거해서는 사람의 행동 유형을 항상 정확히 예측할 수는 없다.
  - 그러나 initial belief를 다소 이상하게 가진다면, 특이한 행동(irrational behavior) 도 어느 정도 설명 가능하다

The smoker example

- 백해무익한 것을 알지만(true belief) 담배를 끊지 못하는 사람.. irrational behavior
- false belief 로 설명 못한다. 진실은 알고 있으므로..
- softmax action noise 로도 설명 못한다.
  - 담배 예제는 단순한 random error 행위가 아니라, 시스템적으로 비최적화된 행동을 한다.
    - 최적화된 행동에서 벗어나기는 해도 예측 가능하다(즉 담배를 필 것)
    - 한 domain에서의 비최적화된 경향이 모든 domain에서 그러한 것은 아니다.

example of systematic deviations from optimal action

- (체계적으로) 담배를 매주 끊을려고 하지만, 매주 실패하는 것
- 숙제를 항상 마감일 임박해서 끝내는 것, 실상은 체계적으로 일찍 끝내려고 계획은 하면서도.
- 랜덤 문자열(숫자열)을 까먹는 것(ex. password, ID)
- long division과 같은 산술 문제를 체계적으로(?) 실수하는 것

이러한 체계적인 비합리로의 이탈을 크게 다음 두 분류로 설명

- cognitive bias
- cognitive bound

Human deviations from optimal action: Cognitive Bounds

- basic computational constraints 로 인해 비합리적(sub-optimal)으로 행동할 수도 있다.
- bounded rationality, bounded optimality, (Gershman et al., 2015).
- constraint on memory
  - ex) forgetting : 중국 레스토랑 문제에서 (PO)MDP agent는 한번 경험한 것을 안 잊어버림. 하지만 사람은 모두 잊어버림!!]
- constraint on time
  - 가장 단순한 POMDP 예제인 bandit 문제조차도.. intractable
- 인간은 기계가 풀 수 있는(tractable)한 POMDP 문제조차도 실수(systematic error)
  - Zhang and Angela,2013, Doshi-Velez and Ghahramani, 2011

Human deviations from optimal action: Cognitive Biases

- 시공간의 인지 제약으로 인한 sub-optimal은 인간이나 기계에 공통된 한계
- 반면 cognitive bias는 오직 인간에게만 적용되는 이야기
- ex)
  - loss aversion
  - time inconsistency

Learning preferences from bounded and biased agents

- 인간은 보았듯이 인지 제약, 인지 편견이 있다.
- 그러므로 softmax-optimal agent 은 인간의 행동을 예측하는 generative model 로써 한계가 있다.
- 이러한 일탈(deviations)을 고려한 방법들에 대해서 앞으로 배워보자
  - time-inconsistent agents via hyperbolic discounting
  - myopic planning

Time inconsistency I

- Time inconsistency 란?
  - 현재 자아의 미래 선호랑 실제 미래 자아의 선호가 다른 경우
    - inconsistency between what you prefer your future self to do and what your future self prefers to do.
  - ex) 잠잘때는 일찍 일어나서 운동하고 싶은 마음. 막상 아침에는 침대에서 더 자고 싶은 상황
- related to laziness, addiction, procrastination, impulsive behavior,
- pre-commitment behavior
  - 미래 자아가 어쩔 수 없이 involve되도록 미리 무언가 장치를 해놓는 것, ex) 알람종
  - introduced by a Nobel-prize winning economist named Thomas Schelling
- 30일 지나고 100달러 받을래? 31일 지나고 110달러 받을래?
  - 대부분은 110달러 선호. 실제로는 100달러를 받는 경우가 많다.

Time inconsistency due to hyperbolic discounting

- 사람은 동일한 가치의 보상이라면 미래보다는 가급적 빨리 받는 것을 선호한다.
  - discounting function as hyperbolic

Exponential discounting for optimal agents

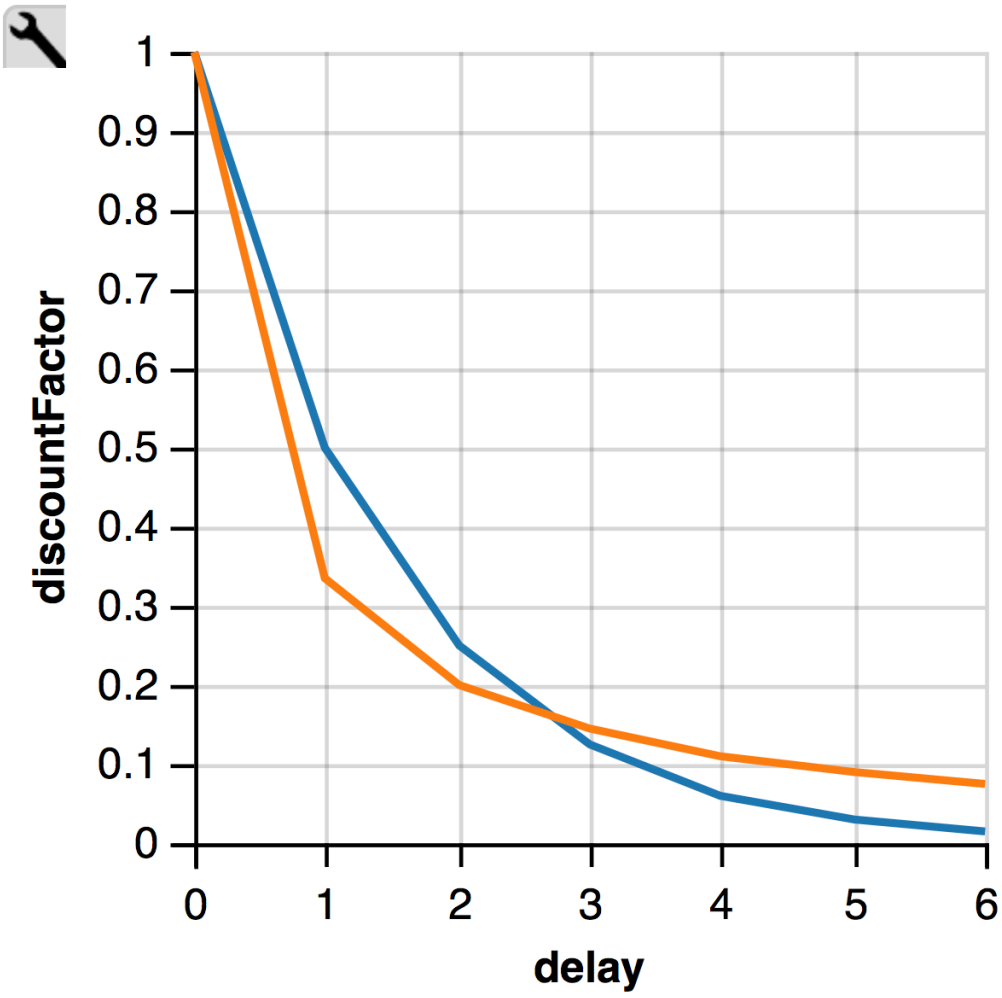
- 이전까지의 예제는 known, finite time horizon 을 기반으로 expected reward 계산
- unbounded or infinite time horizon 에서는 reward 를 그냥 더하면 발산
- 해결책은 discounted expected utility, 특히 exponential discounting

exponential discounting의 효과(부작용)은?

- 현재의 알려진 작은 보상에 만족하는 경향,
- 반대로 미래의 알려지지 않은 보다 큰 보상을 찾는 모험을 꺼리게 됨. 왜냐하면 이 모험을 하는데도 비용이 초래되므로
- 여름 휴가지 결정 예제
  - 알려진 sub-optimal 휴가지에 만족하지, 더 큰 만족을 얻을 휴가지를 탐험하지 않음
  - See code
  - optimal에 비해 상대적으로 exponential trajectory는 explore를 하지 않는다.

Discounting and time inconsistency

- Exponential discounting 은 relative time prefence를 나타낸다.
  - ex) 0 day에서의 보상과 30 day 에서의 보상비는 30 day에서의 보상과 60 day에서의 보상 비율과 같다.
  - 동위원소의 반감기를 생각하면 됨
  - Exponential discounting is time consistent
- expoential 보다 더 smooth한 것, 예를 들어 hyperbolic discounting은 time-inconsistent하다.
  - 초반에는 더 가파르다가 나중에는 덜 가팔라짐
    - 반감기가 점차 길어지는 것
  - preferences that reverse over time (Strotz, 1955)
    - 30일 지나고 100달러 받을래? 31일 지나고 110달러 받을래 문제에서 31일 옵션을 선택할 수도 있음.



Time inconsistency and sequential decision problems

- hyperbolic discount를 사용한다면, 100달러냐 110 달러 선호냐 하는 것은 이제 어떤 시점에 decision making을 하느냐에 따라 달라진다.
- 시간에 따라서 선호 경향이 달라질 수 있으므로, planning이 더 복잡해졌다.
- discount를 아예 안하거나 하더라도 exponential만 사용하는 (PO)MDP 문제에서는 이러한 복잡함이 없다.

대처 방안 :

- Naive agent
  - 현재 자아의 선호를 미래 자아도 따르겠거니 하는 막연한(naive)한 가정을 하는 agent
- Sophisticated agent
  - 현재 자아의 선호가 미래 자아의 선호랑 다를 수 있음을 알고,
  - pre-commitment 를 통해서 미래 자아가 동일한 선호를 가질 수 있도록 correct
- Example
  - Naive 는 채식 선호기는 하지만 북쪽 도넛 가게를 지나가다가 유혹에 빠져 도넛 가게로 들어감
    - 미래의 선호가 바뀔 것을 대비하지 못했음
  - sophisticate 는 도넛 가게 근처로 가면 미래 선호가 바뀔 것을 예측하고, 미리 그쪽 주변을 안지나가도록 우회해서(pre-commitment) 채식 식당으로 간다.

In [ ]: