

ADL—Model Based

2019年4月23日 上午 11:53

希望可以學環境會做什麼事情，先想想看對手會怎麼回在決定要怎麼下比較好

state做一個action會進到下一個state的機率是多少，model怎麼去學

下一個state會進入哪一個state，model環境的資訊

學到環境怎麼transition，會去偷看會得到啥效果

左某一個動作可能會進入甚麼狀態，類似planning，要知道怎麼去做

model，推測之後會發生的事情，去model環境的好壞

model based很難做的很好

AC本來就已經有去學怎樣環境好壞了，所以model base不一訂好

- 缺點:
 - transition只要一小步model不好就會爛掉，trajectory越長越容易出錯，可能會導致錯誤reward
 - 其他方法已經implicit去model怎麼互動了
 - 疊了十層就已經包含去看下十步了 (lookahead)
- 用12-layer CNN去預測expert要去做動作，和直接去做planning跟105步的動作差不多
- 所以model要多好多步才可以知道效果

RL issue

- learning & planning
 - RL可以假設環境不知道怎麼做的，經過許多次亂玩的結果知道遊戲怎麼玩
 - actor直接知道該怎樣走
 - planning是稍微知道環境狀態怎麼樣的
 - 利用predict後發生的是情，利用當成feed back 去做
- Atari 亂玩就會漸漸學到，不用管遊戲設計
- 預測想好做某個動作會發生啥事情，如果在這個state做action會得到啥分數該做啥action
- Exploration vs Exploitation
 - EL 就是trial and error

=====Unsupervised learning用RL來做MUSE

原本word embedding只能model單一與一，希望可以自動學到sense level embedding

看到一句話都要去做sense selection，decision making，利用學出來的sense embedding

reward基於embedding，去算collocation likelihood當成reward

用Q function 看到這些context選這個sense的value

給定context是這個sense的機率，去看對應的sense之colocation likelihood是高或是低

data不用標記，random sample，兩個字各自選module，哪個sense的機率高，sense extraction

RL+DL 可以變成AGI: artificial general Intelligence