

1) مفهوم یادگیری ماشین: (یادگیری: تغییر مثبت در رفتار و بهبود عملکرد در نتیجه تجارب و ایجاد درک روشنتر از آینده [1])

علمیست که کامپیوترها بدون اینکه به طور صریح برنامه ریزی شده باشند ، عمل می کنند.(Andreu Ng)[2]

شاخه ای از هوش مصنوعی است که یک برنامه کامپیوتری میتواند یاد بگیرد و با داده های جدید سازگار شود، بدون دخالت انسانی. (Jake Frankenfield)[3]

در ابتدایی ترین حالت خود استفاده از الگوریتم ها برای تجزیه داده ها ، یادگیری از آنها و سپس تعیین یا پیش بینی درباره چیزی در جهان است.(Michael Copeland)[4]

بر اساس الگوریتم هایی است که می تواند بدون اتکا به برنامه نویسی مبتنی بر قوانین ، از داده ها یاد بگیرد (Dorian Pyle and Cristina San José) [5]

الگوریتم های یادگیری ماشین می توانند با تعمیم مثال ها ، نحوه انجام کارهای مهم را دریابند.(Pedro Domingos) [6]

2) یادگیری تقویتی: [7] [8]

یکی از روشهای یادگیری است که بین دو حالت با نظارت و بی نظارت است. اصطلاحاً علم تصمیم گیری است و مبتنی بر پاداش و جریمه میباشد.

یادگیری تقویتی مجموعه‌ای از مسائل یادگیری را تشریح می‌کند که در آن‌ها یک «عامل» (agent) باید «اعمالی» (actions) را در یک «محیط» (environment) به منظور بیشینه‌سازی «تابع پاداش» (reward function) تعریف شده انجام دهد [9].

با نظارت نیست چون:	همه داده ها برچسب ندارند.
بدون نظارت نیست چون:	عامل باید با اقدامات صحیح پاداش دریافت کند .

فرضیه پاداش: هدف ما ماکزیمم کردن پاداش نهایی است. (پس ممکن است که در عمل های میانه پاداش ماکزیمم نباشد)

H: تاریخچه: ترتیبی از مشاهدات، عملها و پاداش‌ها تا زمان الان. ---> تصمیم جدید

در تصمیم گیری قدم جدید، اطلاعات و قدم های قبلی و پاداش هایشان موثر هستند.

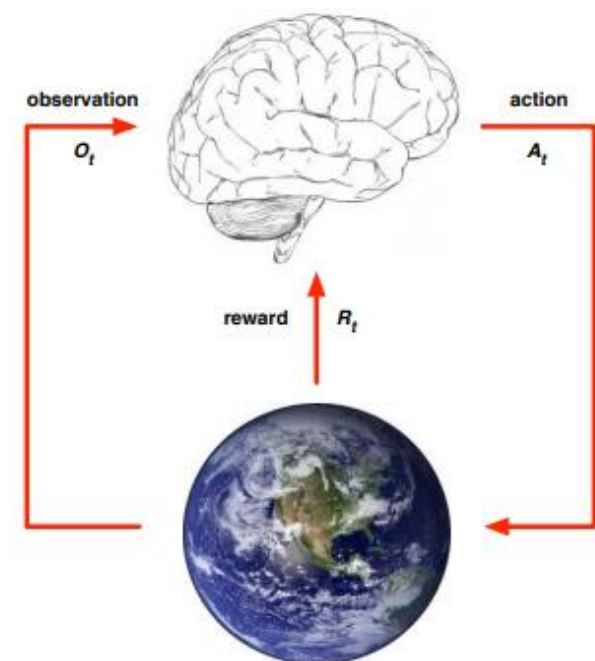


Figure 1: مدل یادگیری تقویتی

State: حالت: تابعی از تاریخچه است. (گام زمانی)

حالت، اطلاعاتی است که برای تعیین قدم بعدی استفاده میشود.

(الف) حالت محیط: نمایشی خصوصی از محیط که ممکن است برای عامل غیر قابل مشاهده باشد و بای انتخاب بعدی استفاده میشود.

(ب) حالت عامل: نمایش داخلی عامل که برای انتخاب بعدی استفاده میشود.

(ج) حالت اطلاعات یا مارکوف: اطلاعات مفیدی از تاریخچه که برای مدلسازی آینده و دور ریختن تاریخچه مفید است.

چیزی که باور داریم در قدم بعدی رخ میدهد، به نمایش عامل از حالت وابسته است.

محیط:

كاملا قابل مشاهده: عامل مستقیماً حالت محیط را میبیند. (فرایند تصمیم ماركوف=MDP)

بخشی قابل مشاهده: عامل، غیر مستقیم محیط را میبیند.

اجزای عامل:

سیاست: رفتار عامل نگاشتی از حالت به عمل است.

تابع ارزش: یک پیشبینی از پاداش آینده با توجه به حالت و عمل

مدل: ارایه مدلی از محیط. پیش بینی هایی از حالت و پاداش بعدی.

الگوریتم های یادگیری تقویتی [10]

از ۴ مورد زیر، دو مورد اول معروف ترین الگوریتم های یادگیری تقویتی محسوب می شوند:

Q-Learning Algorithm

State-action-reward-state-action Algorithm (SARSA)

Deep Q Network Algorithm (DQN)

Deep Deterministic Policy Gradient Algorithm (DDPG)

کاربرد یادگیری تقویتی

بیشتر در حوزه بازی های کامپیوتری (Video Games) و هوش مصنوعی است. موارد دیگر کاربرد یادگیری تقویت شده را به شرح زیر ملاحظه می کنید.

مدیریت منابع در خوشه های رایانه ای

کنترل چراغ راهنمایی

رباتیک

پیکربندی سیستم وب

توصیه های شخصی سازی شده

یادگیری عمیق

References

- [1] <https://mindtoolbox.ir/article/1992/10-%D8%AA%D8%B9%D8%B1%DB%8C%D9%81-%DB%8C%D8%A7%D8%AF%DA%AF%DB%8C%D8%B1%DB%8C/>.
- [2] <https://www.coursera.org/learn/machine-learning>.
- [3] <https://www.investopedia.com/terms/m/machine-learning.asp>.
- [4] <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>.
- [5] <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/an-executives-guide-to-machine-learning>.
- [6] P. Domingos, "A Few Useful Things to Know About Machine Learning," *communications of the acm*, vol. 55, 2012.
- [7] <https://www.youtube.com/watch?v=2pWv7GOvuf0>.
- [8] https://www.davidsilver.uk/wp-content/uploads/2020/03/intro_RL.pdf.

[9] <https://blog.faradars.org/when-reinforcement-learning-not-used/>.

[10] <https://faranesh.com/blog/machine-learning-tutorial>.