

۲۱۰۰/۱
۹۹,۲,۲



یادگیری تقویتی

نویسندگان: ریچارد ساتن، اندرو بارتو

مترجم: سجاد کردانی مقدم

سرشناسه	: ساتون، ریچارد اس. Sutton, Richard S
عنوان و نام پدیدآور	: یادگیری تقویتی / نویسنده ریچارد ساتن، اندرو بارتو؛ مترجم سجاد کردانی مقدم.
مشخصات نشر	: تهران: گسترش علوم پایه، ۱۳۹۹.
مشخصات ظاهری	: ۲۴۴ ص.: مصور.
شابک	: 978-964-490-790-6
وضعیت فهرست نویسی	: فیبا
یادداشت	: Reinforcement learning : an introduction, 2nd ed, c2012. عنوان اصلی:
موضوع	: یادگیری تقویتی
موضوع	: Reinforcement learning
سنا سه افزوده	: بارتو، اندرو جی.
سنا سه افزوده	: Barto, Andrew G
سنا سه افزوده	: کردانی، سجاد، ۱۳۷۰ مترجم
رده بندی دیره	: Q۱۳۲۵/۶
رده بندی سیب	: ۰۰۶/۳۱
شماره کتابشناسی ملی	: ۶۱۰۹۸۵

نام کتاب: یادگیری تقویتی
 مترجم: سجاد کردانی مقدم
 ناشر: انتشارات گسترش علوم پایه
 مدیر فنی: مهدی زنگنه
 صفحه آرایی: سیما ابویی
 طراحی جلد: غزاله قیاسی
 لیتوگرافی: مهرشاد
 چاپخانه: چاپخانه: راه
 سال نشر: ۱۳۹۹
 نوبت چاپ: اول
 شمارگان: جلد ۱
 قیمت کتاب: ۴۰۰۰۰۰ ریال

ISBN: 978-964-490-790-6

شابک ۹۷۸-۹۶۴-۴۹۰-۷۹۰-۶

حق چاپ و نشر محفوظ و مخصوص ناشر می باشد.

دفتر انتشارات: میدان انقلاب، ابتدای کارگر جنوبی، کوچه مهدیزاده، پلاک ۹، طبقه همکف

تلفکس: ۰۲۱-۶۶۹۰۵۳۱۶

تلفن: ۱۵ ~ ۶۶۹۰۵۳۱۲

نمایندگی تهران: خ انقلاب، نبش ۱۲ فروردین، پ ۱۴۴۴، کتابفروشی الیاس-۶۶۴۰۵۰۸۴

Email: gostaresh_op@yahoo.com

www.gostaresh-pub.com

ضمین: فروشنده و خواننده گرامی: این کتاب دارای برچسب اصالت کالا (هولوگرام) در روی جلد است که با تشکر از همکاری شما

مقدمه مترجم:

کتاب حاضر ترجمه هفت فصل اصلی کتاب یادگیری تقویتی نوشته پروفیسور ساتن میباشد که نسخه اولیه آن در سال ۱۹۹۸ نوشته شده است و در سال ۲۰۱۲ مورد ویرایش و بازنویسی قرار گرفته است. این کتاب به عنوان مهمترین مرجع یادگیری تقویتی شناخته میشود و در اکثر دانشگاه‌ها به عنوان مرجع اصلی معرفی میگردد. یادگیری تقویتی یکی از مهمترین انواع یادگیری در بین موجودات است، این نوع یادگیری به زبان ساده با کمک آزمون و خطا صورت میگیرد، عملی که باعث یک موفقیت و در نتیجه پاداش شود در آینده با احتمال بالاتری باید انتخاب شود و عملی که باعث یک شکست و جریمه شود در آینده با احتمال کمتری باید انتخاب شود. یادگیری تقویتی در سال‌های اخیر در حوزه یادگیری ماشین به یکی از مهمترینباحث تبدیل شده است که یکی از دلایل آن پیشرفت‌های سخت‌افزاری است که امکان استفاده از این نوع یادگیری را برای ما ممکن کرده است. یادگیری تقویتی از سال ۲۰۱۴ که گرگور مایند عامل DQN و یادگیری تقویتی عمیق به کمک شبکه‌های عصبی را ارائه داد وارد فاز جدی شد و به عنوان یکی از داغ‌ترین مباحث یادگیری مورد توجه محققان قرار گرفت. این کتاب در فصل هشتم این کتاب با عنوان یادگیری تقویتی عمیق توسط مترجم تالیف شده است که در این فصل با مروری بر شبکه‌های عصبی، یادگیری تقویتی عمیق و عامل DQN و دوره‌هایی که تا سال ۲۰۲۰ بر روی آن صورت گرفته بررسی شده است.

سجاد نردانی مقدم

Sajjad.moghadam@ut.ac.ir

فهرست مطالب

۷	فصل اول: مقدمه
۸	۱-۱ یادگیری تقویتی
۱۱	۲-۱ مثال‌ها
۱۳	۳-۱ عناصر یادگیری تقویتی
۱۶	۴-۱ یک مدل قابل تعمیم: بازی دوز
۲۴	۵-۱ نسخه‌های مختلف
۲۵	۶-۱ تاریخچه یادگیری تقویتی
۳۷	فصل دوم: مسائل دسته‌های بازی
۳۸	۱-۲ مساله‌ی دسته‌های بازی
۴۱	۲-۲ روش‌های ارزش-اقدام
۴۵	۳-۲ انتخاب اقدام بیشینه هموار
۴۶	۴-۲ پیاده‌سازی افزایشی
۴۸	۵-۲ ردیابی مسائل دینامیک
۵۱	۶-۲ مقادیر اولیه‌ی خوش بینانه
۵۳	۷-۲ جست و جوی وابسته (مسائل مبتنی بر زمینه)
۵۵	۸-۲ نتیجه‌گیری‌ها
۵۷	۹-۲ یادداشت‌های کتابشناختی و تاریخی
۵۹	فصل سوم: مسئله یادگیری تقویتی
۵۹	۱-۳ چرخه تعاملی عامل-محیط
۶۴	۲-۳ اهداف و پاداش‌ها
۶۶	۳-۳ پاداش بازگشتی
۶۹	۴-۳ نشانه‌گذاری یکسان برای حالت چندبخشی و مداوم
۷۰	۵-۳ خاصیت مارکوف

۶-۳	فرآیند تصمیم‌گیری مارکوف	۷۵
۷-۳	تابع ارزش	۷۸
۸-۳	تابع ارزش بهینه	۸۶
۹-۳	بهینگی و تقریب	۹۳
۱۰-۳	خلاصه فصل	۹۴
۱۱-۳	اشارات تاریخی و علمی	۹۶

فصل چهارم: برنامه‌نویسی پویا

۱-۴	ارزیابی سیاست	۱۰۰
۲-۴	بهبود سیاست	۱۰۵
۳-۴	تکرار سیاست	۱۰۹
۴-۴	تکرار ارزش	۱۱۳
۵-۴	برنامه‌نویسی پویا غیرزمان	۱۱۷
۶-۴	تکرار سیاست تعمیم یافته	۱۱۸
۷-۴	بهینگی برنامه‌نویسی پویا	۱۲۱
۸-۴	خلاصه فصل	۱۲۲
۹-۴	اشارات تاریخی و علمی	۱۲۴

فصل پنجم: روش‌های مونته کارلو

۱-۵	ارزیابی سیاست مونته کارلو	۱۲۸
۲-۵	تخمین مونته کارلو از مقادیر اقدام	۱۳۴
۳-۵	کنترل مونته کارلو	۱۳۶
۴-۵	کنترل «برسیاستی» مونته کارلو	۱۴۰
۵-۵	ارزیابی یک سیاست حین پیروی از دیگری (ارزیابی سیاست برون‌سیاستی)	۱۴۴
۶-۵	کنترل برون‌سیاستی مونته کارلو	۱۴۶
۷-۵	پیاپیاده‌سازی افزایشی	۱۴۹
۸-۵	خلاصه فصل	۱۵۰
۹-۵	اشارات تاریخی و علمی	۱۵۲

فصل ششم: یادگیری تفاوت زمانی

۱-۶	پیش‌بینی TD	۱۵۶
-----	-------------	-----

۱۶۲	مزیت‌های روش‌های پیش‌بینی TD	۲-۶
۱۶۶	بهینگی $TD(0)$	۳-۶
۱۷۱	سارسا: کنترل TD بر سیاست	۴-۶
۱۷۵	یادگیری-کیو: کنترل TD برون سیاست	۵-۶
۱۷۸	بازی‌ها، پس‌وضعیت‌ها و دیگر موارد خاص	۶-۶
۱۸۰	خلاصه	۷-۶
۱۸۲	اشارات کتاب‌شناسی و تاریخی	۸-۶

فصل هفتم: ردیابی‌های شایستگی

۱۸۶	۱-۷ پیچیدگی TD چندگامی	۱-۷
۱۹۳	۲-۷ نگاه روبه‌جلوی $TD(\lambda)$	۲-۷
۱۹۸	۳-۷ نگاه روبه‌عقب $TD(\lambda)$	۳-۷
۲۰۲	۴-۷ همسان‌سازی راه‌رو، جلو و روبه‌عقب	۴-۷
۲۰۶	۵-۷ سارسا (λ)	۵-۷
۲۱۰	۶-۷ $Q(\lambda)$	۶-۷
۲۱۴	۷-۷ ردیابی‌های جایگزینی	۷-۷
۲۱۷	۸-۷ مشکلات پیاده‌سازی	۸-۷
۲۱۸	۹-۷ λ متغیر	۹-۷
۲۱۹	۱۰-۷ نتیجه‌گیری‌ها	۱۰-۷
۲۲۱	۱۱-۷ اشارات کتاب‌شناسی و تاریخی	۱۱-۷

فصل هشتم: یادگیری تقویتی عمیق

۲۲۴	۱-۸ شبکه‌های عصبی	۱-۸
۲۳۷	۲-۸ روش‌های شبکه‌های عصبی عمیق	۲-۸
۲۴۶	۳-۸ یادگیری تقویتی عمیق	۳-۸
۲۶۲	۴-۸ خلاصه فصل	۴-۸