

یادگیری تقویتی عمیق

مرزهای هوش مصنوعی

مؤلف

محیط سواک

مترجم

ایوب ترکیان

نیاز دانش

سرشناسه	: سواک، محیط	Sewak, Mohit
عنوان و نام پدیدآور	: یادگیری تقویتی عمیق: مرزهای هوش مصنوعی / مولف محیط سواک؛ مترجم ایوب ترکیان	
مشخصات نشر	: تهران: نیاز دانش، ۱۳۹۸.	
مشخصات ظاهری	: ۲۱۴ص: مصور، جدول، نمودار.	
شابک	: 978-600-8906-58-2	
وضعیت فهرست‌نویسی	: فیپا	
یادداشت	: عنوان اصلی: Deep Reinforcement Learning : Frontiers of Artificial Intelligence, 2019	
عنوان دیگر	: مرزهای هوش مصنوعی	
موضوع	: یادگیری تقویتی	Reinforcement learning
موضوع	: رمزگذاری داده‌ها	Data encryption (Computer science)
شناسه افزوده	: ترکیان، ایوب، ۱۳۳۷ - مترجم.	
رده‌بندی کنگره	: Q۳۲۵/۶	
رده‌بندی دیویی	: ۰۰۶/۳۱	
شماره کتابشناسی	: ۵۸۷۱۳۲۸	



نام کتاب	: یادگیری تقویتی عمیق: مرزهای هوش مصنوعی
مؤلف	: محیط سواک
مترجم	: ایوب ترکیان
مدیر اجرایی - ناظر بر چاپ	: حمیدرضا محمد شیرازی - محمد شمس
ناشر	: نیاز دانش
صفحه‌آرا	: واحد تولید انتشارات نیاز دانش
نوبت چاپ	: اول - ۱۳۹۸
شمارگان	: ۱۰۰ نسخه
قیمت	: ۳۶۰۰۰۰ ریال

ISBN:978-600-8906-58-2

شابک: ۹۷۸-۶۰۰-۸۹۰۶-۵۸-۲

هرگونه چاپ و تکثیر (اعم از زیراکس، بازنویسی، ضبط کامپیوتری و تهیه‌ی CD) از محتویات این اثر بدون اجازه کتبی ناشر ممنوع است، متخلفان به موجب بند ۵ از ماده ۲ قانون حمایت از مؤلفان، مصنفان و هنرمندان تحت پیگرد قانونی قرار می‌گیرند.

کلیه حقوق این اثر برای ناشر محفوظ است.

آدرس انتشارات: تهران، میدان انقلاب، خیابان ۱۲ فروردین، تقاطع وحید نظری، پلاک ۲۵۵، طبقه ۱، واحد ۲

۰۲۱-۶۶۴۷۸۱۰۶-۶۶۴۷۸۱۰۸-۰۹۱۲۷۰۷۳۹۳۵

www.Niaze-Danesh.com

مشاوره جهت نشر: ۰۹۱۲-۲۱۰۶۷۰۹

فهرست مطالب

شماره صفحه	عنوان
۹	فصل ۱ / مقدمه
۹	۱.۱ ارتباط یادگیری تقویتی با AI
۱۰	۲.۱ شناخت طراحی پایه
۱۱	۳.۱ تعیین تابع پاداش
۱۱	۱.۳.۱ پاداش‌های آتی
۱۲	۲.۳.۱ پاداش‌های تصادفی/غیرقطعی
۱۲	۳.۳.۱ تخصیص پاداش
۱۴	۴.۳.۱ تعیین تابع پاداش مناسب
۱۴	۵.۳.۱ انواع پاداش
۱۵	۶.۳.۱ جوانب زمینه
۱۶	۴.۱ حالت در یادگیری تقویتی
۱۷	۱.۴.۱ چهارخونه‌بازی
۱۸	۲.۴.۱ مسئله موازنه پایه چرخ
۲۱	۳.۴.۱ ماریو و شاهزاده
۲۴	۵.۱ عامل در یادگیری تقویتی
۲۴	۱.۵.۱ تابع ارزش
۲۵	۲.۵.۱ اقدام-ارزش/تابع Q
۲۵	۳.۵.۱ کاوش یا بهره‌برداری
۲۶	۴.۵.۱ رویکردهای سیاست
۲۷	۶.۱ خلاصه

فصل ۲ / شناخت ریاضی و الگوریتمی

۲۹	۱.۲ فرایند تصمیم مارکوف
۳۰	۱.۱.۲ نشانه گذاری MDP
۳۱	۲.۱.۲ MDP - هدف ریاضی
۳۱	۲.۲ معادله Bellman
۳۲	۱.۲.۲ تخمین تابع ارزش
۳۳	۲.۲.۲ تخمین تابع Q
۳۴	۳.۲ برنامه سازی پویا و تابع بلمن
۳۴	۱.۳.۲ برنامه سازی پویا
۳۵	۲.۳.۲ بهینه گی حل معادله بلمن
۳۵	۴.۲ روش های استبرگشت ارزش و سیاست
۳۶	۱.۴.۲ تابع بلمن ارزش و سیاست بهینه
۳۶	۲.۴.۲ رفت و برگشت ارزش
۳۷	۳.۴.۲ رفت و برگشت و ارزش و سیاست
۳۷	۵.۲ خلاصه

فصل ۳ / کد سازی محیط و مدل MDP

۳۹	۱.۳ مثال مسئله دنیای شبکه
۳۹	۱.۱.۳ شناخت دنیای شبکه
۴۰	۲.۱.۳ انتقال های حالت مجاز
۴۱	۲.۳ ساخت محیط
۴۱	۱.۲.۳ ارث یا ساخت طبقه محیط
۴۳	۲.۲.۳ دستور العمل ساخت طبقه محیط سفارشی
۴۵	۳.۳ الزامات پلاتفرم و ساختار پروژه برای کد
۴۷	۴.۳ کد ایجاد محیط دنیای شبکه
۵۲	۵.۳ کد رویکرد رفت و برگشت ارزش
۵۵	۶.۳ کد رویکرد رفت و برگشت سیاست
۵۹	۷.۳ خلاصه

فصل ۴ / یادگیری تفاضل زمانی، SARSA، و یادگیری Q

۶۱	۱.۴ چالش های برنامه سازی پویای کلاسیک
۶۳	۲.۴ رویکردهای مدل پایه و آزاد
۶۴	۳.۴ یادگیری تفاضل زمانی (DP)
۶۵	۱.۳.۴ مسایل تخمین و کنترل
۶۵	۲.۳.۴ TD(0)
۶۶	۳.۳.۴ TD(λ) و ردیابی حائر شرایط
۶۷	۴.۴ SARSA

۶۹	۵.۴ یادگیری Q
۷۱	۶.۴ الگوریتم‌های راهزن
۷۱	۱.۶.۴ ع مقتصدانه
۷۲	۲.۶.۴ الگوریتم‌های سازگار با زمان
۷۳	۳.۶.۴ الگوریتم‌های سازگار با اقدام
۷۳	۴.۶.۴ الگوریتم‌های سازگار با ارزش
۷۴	۵.۶.۴ انتخاب الگوریتم راهزن
۷۴	۷.۴ خلاصه

فصل ۵ / کدسازی یادگیری Q

۷۷	۱.۵ ساختار پروژه و وابستگی‌ها
۷۹	۲.۵ کد
۷۹	۱.۲.۵ وابستگی‌ها و لاگ کردن (Q_learning.py)
۸۰	۲.۲.۵ کد طبقه‌بندی
۸۲	۳.۲.۵ کد طبقه‌بندی کامل یادگیری Q
۸۵	۴.۲.۵ کد تست پیاده‌سازی عامر (تابع اصلی)
۸۵	۵.۲.۵ کد استثناءهای سارشی (rl_exception.py)
۸۵	۳.۵ نمودار آمار آموزش

فصل ۶ / مقدمه یادگیری عمیق

۸۷	۱.۶ نورون‌های مصنوعی
۸۹	۲.۶ شبکه‌های عصبی عمیق پیش‌خور (DNN)
۹۱	۱.۲.۶ مکانیسم پیش‌خور
۹۲	۳.۶ ملاحظات معماری
۹۳	۱.۳.۶ توابع فعال‌سازی
۹۴	۲.۳.۶ توابع اتلاف
۹۵	۳.۳.۶ بهینه‌گرها
۹۶	۴.۶ شبکه‌های عصبی کانولوشن
۹۷	۱.۴.۶ لایه کانولوشن
۹۸	۲.۴.۶ لایه رأی‌گیری
۹۸	۳.۴.۶ لایه‌های مسطح و تمام‌وصل
۹۹	۵.۶ خلاصه

فصل ۷ / منابع پیاده‌سازی

۱۰۱	۱.۷ تنها نیستید
۱۰۳	۲.۷ محیط‌ها و پلافرم‌های آموزش استاندارد

۱۰۳	۱.۲.۷ دنیای OpenAI و Retro
۱۰۳	OpenAI Gym ۲.۷.۲
۱۰۴	آزمایشگاه DeepMind ۳.۲.۷
۱۰۴	مجموعه کنترل DeepMind ۴.۲.۷
۱۰۴	پروژه Malmö ۵.۲.۷
۱۰۴	گاراژ ۶.۲.۷
۱۰۵	۳.۷ کتابخانه‌های توسعه و پیاده‌سازی عامل
۱۰۵	DeepMind TRFL ۱.۳.۷
۱۰۵	خطوط مبنای OpenAI ۲.۳.۷
۱۰۵	RL کراس ۳.۳.۷
۱۰۶	Couch ۴.۳.۷
۱۰۶	K lib ۵.۳.۷

۱۰۷	فصل ۸ / شبکه عصبی Q دوتایی، و دولی
۱۰۷	۱.۸ هوش مصنوعی عام
۱۰۹	۲.۸ مقدمه DeepMind گوگل و AlphaGo
۱۱۰	۳.۸ الگوریتم DQN
۱۱۲	۱.۳.۸ بازپخش تجربه
۱۱۶	۲.۳.۸ شبکه Q هدف اضافی
۱۱۷	۳.۳.۸ برش پاداش‌ها و جرایم
۱۱۸	۴.۸ دوتایی DQN
۱۱۹	۵.۸ دولی DQN
۱۲۱	۶.۸ خلاصه

۱۲۳	فصل ۹ / کدسازی DQN دوتایی
۱۲۳	۱.۹ ساختار پروژه و وابستگی‌ها
۱۲۵	۲.۹ کد عامل DQN (پرونده DoubleDQN.py)
۱۲۳	۱.۲.۹ کد طبقه سیاست رفتار (پرونده behavior_policy.py)
۱۲۷	۲.۲.۹ کد طبقه حافظه بازپخش تجربه (پرونده Experience_Replay.py)
۱۲۹	۳.۲.۹ کد طبقات استثناء سفارشی (پرونده RL_Exceptions.py)
۱۳۹	۳.۹ نمودارهای آمار آموزشی

۱۴۱	فصل ۱۰ / رویکردهای سیاست پایه
۱۴۱	۱.۱۰ مقدمه
۱۴۳	۲.۱۰ تفاوت رویکردهای ارزش پایه و سیاست پایه

۱۴۶	۳.۱.۰ مشکلات محاسبه گرادیان سیاست
۱۴۸	۴.۱.۰ REINFORCE
۱۵۰	۱.۴.۱.۰ نقایص الگوریتم REINFORCE
۱۵۱	۲.۴.۱.۰ شبه کد الگوریتم REINFORCE
۱۵۱	۵.۱.۰ روش کاهش واریانس
۱۵۱	۱.۵.۱.۰ تخصیص پاداش پایه آتی تجمعی
۱۵۲	۲.۵.۱.۰ پاداش‌های آتی تجمعی تخفیف یافته
۱۵۳	۳.۵.۱.۰ REINFORCE با خط‌مبنا
۱۵۴	۶.۱.۰ انتخاب خط‌مبنا برای الگوریتم REINFORCE
۱۵۵	۷.۱.۰ خلاصه

فصل ۱۱ / مدل‌های فاعل-منتقد

۱۵۷	۱.۱.۱ مقدمه روش‌های فاعل-منتقد
۱۵۹	۲.۱.۱ طراحی مفهوم
۱۶۰	۳.۱.۱ معماری برای پیاده‌سازی
۱۶۲	۱.۳.۱.۱ روش فاعل-منتقد و DQN دوئلی
۱۶۴	۲.۳.۱.۱ مزیت معماری مدل فاعل-منتقد
۱۶۵	۴.۱.۱ پیاده‌سازی فاعل-منتقد مزیت غیر حروبه (A3C)
۱۶۷	۵.۱.۱ پیاده‌سازی فاعل-منتقد مزیت سگروپ (A2C)
۱۶۹	۶.۱.۱ خلاصه

فصل ۱۲ / کدسازی A3C

۱۷۱	۱.۱.۲ ساختار و وابستگی‌های پروژه
۱۷۵	۲.۱.۲ کد (A3C_Master_File: a3c_master.py)
۱۷۹	۱.۲.۱.۲ A3C_Worker (File: a3c_worker.py)
۱۸۵	۲.۲.۱.۲ مدل فاعل-منتقد (File: actorcritic_model.py) Tensorflow
۱۸۷	۳.۲.۱.۲ SimpleListBasedMemory (File: experience_replay.py)
۱۹۰	۴.۲.۱.۲ Custom Exceptions (rl_exceptions.py)
۱۹۰	۳.۱.۲ نمودارهای آمار آموزش

فصل ۱۳ / گرادیان سیاست قطعی و DDPG

۱۹۳	۱.۱.۳ گرادیان سیاست قطعی (DPG)
۱۹۵	۱.۱.۱.۳ مزایای گرادیان سیاست قطعی
۱۹۷	۲.۱.۱.۳ نظریه گرادیان سیاست قطعی
۱۹۸	۳.۱.۱.۳ فاعل-منتقد گرادیان پایه سیاست قطعی سیاست خاموش
۱۹۹	۲.۱.۳ گرادیان سیاست قطعی عمیق (DDPG)

۲۰۰	DDPG اصلاحات یادگیری عمیق
۲۰۳	DDPG شبه کد الگوریتم
۲۰۳	۳.۱۳ خلاصه

فصل ۱۴ / کدسازی DDPG ۲۰۷

۲۰۷	۱.۱۴ کتابخانه‌های ضروری سطح بالا
۲۰۸	۲.۱۴ محیط پیوسته خودروی کوهستان (Gym)
۲۰۸	۳.۱۴ ساختار و وابستگی‌های پروژه
۲۱۱	۴.۱۴ کد (File: ddp_g_continout_actor.py)
۲۱۴	۵.۱۴ عامل بازیگرم محیط MountainCarContinuous-v0

www.ketab.ir