

یادگیری تقویتی عمیق

مرزهای هوش مصنوعی

مؤلف

محیط سواک

مترجم

ایوب ترکیان

نیاز دانش

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

۲۹	فصل ۲ / شناخت ریاضی و الگوریتمی
۲۹	۱.۲ فرایند تصمیم مارکوف
۳۰	۱.۱.۲ نشانه‌گذاری 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 برنامه‌سازی پویا و تابع بلمن
۳۴	۱.۳.۲ برنامه‌سازی پویا
۳۵	۲.۳.۲ بهینه‌گی حل معادله بلمن
۳۵	۴.۲ روش‌های رفت‌وبرگشت ارزش و سیاست
۳۶	۱.۴.۲ تابع بلمن ارزش و سیاست بهینه
۳۶	۲.۴.۲ رفت‌وبرگشت ارزش
۳۷	۳.۴.۲ رفت‌وبرگشت و ارزیابی سیاست
۳۷	۵.۲ خلاصه

۳۹	فصل ۳ / کدسازی محیط و حل MDP
۳۹	۱.۳ مثال مسئله دنیای شبکه
۳۹	۱.۱.۳ شناخت دنیای شبکه
۴۰	۲.۱.۳ انتقال‌های حالت مجاز
۴۱	۲.۳ ساخت محیط
۴۱	۱.۲.۳ ارث یا ساخت طبقه محیط
۴۳	۲.۲.۳ دستورالعمل 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 رفت‌وبرگشت ارزش
۵۵	۶.۳ کد رویکرد رفت‌وبرگشت سیاست
۵۹	۷.۳ خلاصه

۶۱	فصل ۴ / یادگیری تفاضل زمانی، SARSA، و یادگیری Q
۶۱	۱.۴ چالش‌های برنامه‌سازی 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

۷۷	۱.۵ ساختار پروژه و وابستگی‌ها
۷۹	۲.۵ کد
۷۹	۱.۲.۵ واردات و لاگ کردن (Q_learning.py)
۸۰	۲.۲.۵ کد طبقه سیاست رفتار
۸۲	۳.۲.۵ کد طبقه عامل یادگیری Q
۸۵	۴.۲.۵ کد تست پیاده‌سازی عامل (تابع اصلی)
۸۵	۵.۲.۵ کد استثناءهای سفارشی (rl_exceptions.py)
۸۵	۳.۵ نمودار آمار آموزش

فصل ۶ / مقدمه یادگیری عمیق ۸۷

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

۱۰۱	۱.۷ تنها نیستید
۱۰۳	۲.۷ محیط‌ها و پلاگین‌های آموزش استاندارد

۱۰۳	Retro و OpenAI دنیای
۱۰۳	OpenAI Gym
۱۰۴	DeepMind آزمایشگاه
۱۰۴	DeepMind مجموعه کنترل
۱۰۴	Malmo پروژه
۱۰۴	گاراژ
۱۰۵	کتابخانه‌های توسعه و پیاده‌سازی عامل
۱۰۵	DeepMind TRFL
۱۰۵	OpenAI خطوط مبنای
۱۰۵	RL کراس
۱۰۶	Coach
۱۰۶	RLlib

فصل ۸ / شبکه عمیق Q، دوتایی، و دوئل ۱۰۷

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 DQN دوتایی
۱۱۹	۵.۸ DQN دوئلی
۱۲۱	۶.۸ خلاصه

فصل ۹ / کدسازی DQN دوتایی ۱۲۳

۱۲۳	۱.۹ ساختار پروژه و وابستگی‌ها
۱۲۵	۲.۹ کد عامل DQN (پرونده DoubleDQN.py)
۱۳۳	۱.۲.۹ کد طبقه سیاست رفتار (پرونده behavior_policy.py)
۱۳۷	۲.۲.۹ کد طبقه حافظه بازپخش تجربه (پرونده Experience_Replay.py)
۱۳۹	۳.۲.۹ کد طبقات استثناء سفارشی (پرونده RL_Exceptions.py)
۱۳۹	۳.۹ نمودارهای آمار آموزشی

فصل ۱۰ / رویکردهای سیاست پایه ۱۴۱

۱۴۱	۱.۱۰ مقدمه
۱۴۳	۲.۱۰ تفاوت رویکردهای ارزش پایه و سیاست پایه

۱۴۶	مشکلات محاسبه گرادیان سیاست
۱۴۸	۴.۱.۰ الگوریتم 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 تخفیف‌یافته
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 مفهومی
۱۶۰	۳.۱.۱ معماری برای پیاده‌سازی
۱۶۲	۱.۳.۱.۱ روش فاعل-منتقد و DQN دوگلی
۱۶۴	۲.۳.۱.۱ مزیت معماری مدل فاعل-منتقد
۱۶۵	۴.۱.۱ پیاده‌سازی فاعل-منتقد مزیت غیرسنکرونه (A3C)
۱۶۷	۵.۱.۱ پیاده‌سازی فاعل-منتقد مزیت سنکرونه (A2C)
۱۶۹	۶.۱.۱ خلاصه

فصل ۱۲ / کدسازی A3C

۱۷۱	۱.۱۲ ساختار و وابستگی‌های پروژه
۱۷۵	۲.۱۲ کد (A3C_Master_File:a3c_master.py)
۱۷۹	۱.۲.۱۲ A3C_Worker (File: a3c_worker.py)
۱۸۵	۲.۲.۱۲ مدل فاعل-منتقد (TensorFlow (File: actorcritic_model.py)
۱۸۷	۳.۲.۱۲ SimpleListBasedMemory (File: experience_replay.py)
۱۹۰	۴.۲.۱۲ Custom Exceptions (rl_exceptions.py)
۱۹۰	۳.۱۲ نمودارهای آمار آموزش

فصل ۱۳ / گرادیان سیاست قطعی و DDPG

۱۹۳	۱.۱۳ گرادیان سیاست قطعی (DPG)
۱۹۵	۱.۱.۱۳ مزایای گرادیان سیاست قطعی
۱۹۷	۲.۱.۱۳ نظریه گرادیان سیاست قطعی
۱۹۸	۳.۱.۱۳ فاعل-منتقد گرادیان پایه سیاست قطعی خاموش
۱۹۹	۲.۱۳ گرادیان سیاست قطعی عمیق (DDPG)

۲۰۰	DDPG اصلاحات یادگیری عمیق
۲۰۳	DDPG شبیه‌کد الگوریتم
۲۰۳	۳.۱۳ خلاصه

۲۰۷ فصل ۱۴ / کدسازی DDPG

۲۰۷	۱.۱۴ کتابخانه‌های زورق 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_action.py)
۲۱۴	۵.۱۴ عامل بازیگر محیط MountainCarContinuous-v0