



پرديس علوم
دانشكده رياضي، آمار و علوم كامپيوتر

به كارگيري روش‌هاي يادگيري ماشين براي معامله‌گري در بازار رمزارزها

نگارنده: علي باجلان

استاد راهنما: دكتور باقر بااعلي

پايان‌نامه براي دريافت درجه كارشناسي
در رشته علوم كامپيوتر

تير ماه ۱۴۰۱

چکیده

پیش‌بینی قیمت در بازارهای مالی همیشه هدف اصلی سرمایه‌گذاران بوده است و فعالین بازار با روش‌های گوناگونی دست به پیش‌بینی قیمت آینده می‌زنند. یکی از روش‌هایی که در سال‌های اخیر بسیار مورد توجه قرار گرفته است بهره‌گیری از تکنیک‌های مختلف هوش مصنوعی برای این مهم می‌باشد. یادگیری تقویتی عمیق زیرمجموعه‌ای از هوش مصنوعی است که با ترکیب دو حوزه یادگیری تقویتی و یادگیری عمیق به دنبال حل دسته گسترده‌ای از مسائل شامل رباتیک، بازی‌های رایانه‌ای و ... می‌باشد. در این پروژه با استفاده از داده‌های قیمت و حجم معاملات گذشته‌ی جفت ارز بیت‌کوین-تتر^۱ یک عامل معامله‌گر بر مبنای یادگیری تقویتی عمیق آموزش داده می‌شود که با تشخیص الگوهای قیمتی-حجمی به خرید و فروش بیت‌کوین با رمزارز پایدار^۲ تتر^۳ پردازد و در عین توجه به سوددهی از مدیریت ریسک نیز غافل نشود.

BTC/USDT^۱

Stable Coin^۲

Tether^۳

فهرست مطالب

۱	مقدمه	۱
۱	۱.۱ قیمت	۱.۱
۲	۱.۱.۱ شمع	۱.۱.۱
۳	۲.۱ تحلیل در بازارهای مالی	۲.۱
۳	۱.۲.۱ تحلیل بنیادی	۱.۲.۱
۴	۲.۲.۱ تحلیل فنی	۲.۲.۱
۴	۳.۲.۱ تحلیل احساسات	۳.۲.۱
۵	۲ یادگیری ماشین	۲
۵	۱.۲ مقدمه	۱.۲
۵	۲.۲ یادگیری عمیق	۲.۲
۶	۱.۲.۲ شبکه های پرسپترون	۱.۲.۲
۷	۲.۲.۲ تابع اتلاف	۲.۲.۲
۷	۳.۲.۲ بازتنظیم پارامترها	۳.۲.۲
۱۰	۳.۲ یادگیری تقویتی	۳.۲
۱۰	۱.۳.۲ فرآیند تصمیم گیری مارکوف	۱.۳.۲
۱۱	۲.۳.۲ یادگیری Q	۲.۳.۲
۱۱	۳.۳.۲ تابع Q	۳.۳.۲
۱۲	۴.۳.۲ یادگیری Q*	۴.۳.۲
۱۳	۴.۲ یادگیری تقویتی عمیق	۴.۲
۱۳	۱.۴.۲ یادگیری عمیق Q	۱.۴.۲
۱۳	۲.۴.۲ تابع اتلاف	۲.۴.۲

۱۳	پس انتشار خطا	۳.۴.۲
۱۴	بازپخش تجربه	۴.۴.۲
۱۴	اکتشاف در برابر ارتفاع	۵.۴.۲
۱۴	اکتشاف ε -حریصانه	۶.۴.۲
۱۵	پیاده سازی و نتایج	۳
۱۵	بیان مسئله	۱.۳
۱۵	محیط مسئله	۲.۳
۱۷	داده های ورودی	۱.۲.۳
۱۸	استخراج ویژگی	۲.۲.۳
۱۹	پاداش	۳.۲.۳
۱۹	ریسک	۴.۲.۳
۲۰	تابع پاداش	۳.۳
۲۰	شبکه عصبی	۴.۳
۲۱	کتابخانه های مورد استفاده	۵.۳
۲۱	نتایج	۶.۳
۲۲	جمع بندی و پیشنهادها	۴
۲۲	جمع بندی	۱.۴
۲۲	پیشنهادها برای ادامه ی کار	۲.۴
۲۵	منابع	

فصل ۱

مقدمه

رمزارز^۱ (ارز دیجیتال) گونه‌ای پول دیجیتال است که در آن تولید واحد پول و تایید اصالت تراکنش با استفاده از الگوریتم‌های رمزنگاری کنترل می‌شود و معمولاً به طور نامتمرکز (بدون وابستگی به بانک‌های مرکزی) فعالیت می‌کنند. با افزایش تعداد و گوناگونی این رمزارزها و بالارفتن مقبولیت آنان در مبادلات، بازارهایی همانند فارکس^۲، که مبادلات جفت ارزهای ملی در آن انجام می‌شود، برای خرید و فروش و مبادله رمزارزها به یک‌دیگر یا ارزهای مرجع مثل دلار شکل گرفت و باعث جذب عده دیگری از سرمایه‌گذاران و معامله‌گران به سمت رمزارزها شد. در دو سال اخیر با جهش قیمت و استقبال بیشتر مردم و موسسات مالی، اندازه این بازار در مقاطعی حتی به بیش از ۸.۲ تریلیون دلار رسید که نشان‌دهنده بزرگی و رشد خیره‌کننده آن می‌باشد.

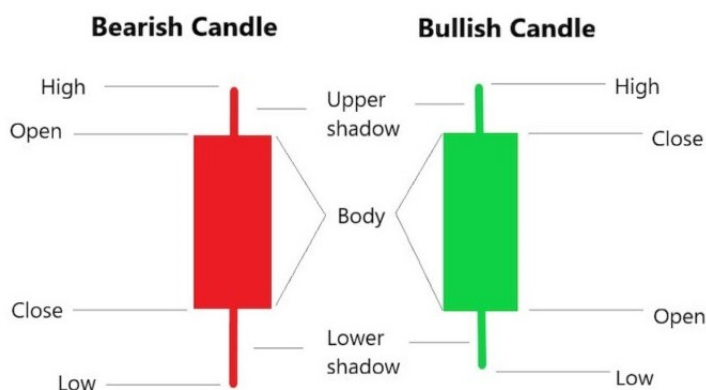
۱.۱ قیمت

در یک بازار آزاد برای هر کالای قابل معامله، نقطه‌ای که نیروهای عرضه و تقاضا به تعادل می‌رسند قیمت تعادلی نام دارد و معاملات با آن قیمت صورت می‌پذیرد. در بازارهای مالی به دلیل نوسان زیاد در نیروهای عرضه و تقاضا، قیمت نیز ثابت نمی‌ماند و نوسان می‌کند. برای بیان دقیق‌تر نوسانات قیمتی، بازه‌های زمانی ثابتی مانند روزانه، ساعتی و دقیقه‌ای را در نظر می‌گیرند و تغییرات قیمت در این بازه زمانی را با شکلی شبیه شمع نمایش می‌دهند.

^۱ Cryptocurrency
^۲ Forex

۱.۱.۱ شمع

هر شمع^۳ با استفاده از ۴ قیمت ساخته می‌شود: قیمت آغازین (قیمت باز شدن^۴)، قیمت پایانی (قیمت بسته شدن^۵)، بیشترین قیمت (بالا^۶) و کمترین قیمتی (پایین^۷) که کالا در آن بازه زمانی معامله شده است. شمع‌ها به دو دسته صعودی و نزولی تقسیم می‌شوند که در شمع صعودی قیمت پایانی بیشتر از قیمت آغازین است و اگر قیمت پایانی کمتر از قیمت آغازین باشد به آن شمع نزولی گویند. هر شمع را از لحاظ ظاهری می‌توان با سه مولفه بدنه^۸، سایه و رنگ متمایز کرد. به فاصله بین قیمت آغازین و قیمت پایانی بدنه شمع گفته می‌شود. هر شمع می‌تواند حداکثر دو سایه داشته باشد؛ در شمع صعودی به فاصله بین قیمت پایانی با بیشترین قیمت، سایه بالا^۹ و به فاصله بین قیمت آغازین و کمترین قیمت سایه پایین^{۱۰} گفته می‌شود که برای شمع نزولی بالعکس می‌باشد. اگر این فاصله‌ها وجود نداشته باشند و قیمت‌ها برابر باشند سایه‌ای شکل نمی‌گیرد. همچنین رنگ هر شمع را از روی صعودی یا نزولی بودن آن تعیین می‌کنند که عموماً شمع صعودی را با رنگ سبز و شمع نزولی را با رنگ قرمز نمایش می‌دهند.



شکل ۱.۱: شمع‌های صعودی و نزولی [۱]

- Candle^۳
- Open^۴
- Close^۵
- High^۶
- Low^۷
- Body^۸
- Upper Shadow^۹
- Lower Shadow^{۱۰}

۲.۱ تحلیل در بازارهای مالی

در بازارهای مالی سرمایه‌گذاران و معامله‌گران با تجزیه و تحلیل عملکرد گذشته و بررسی پارامترهای تاثیرگذار بر آینده برآوردی از ریسک و سودآوری آینده کالا یا سهام مربوط بدست می‌آورند و طبق آن تصمیم به معامله یا رها کردن می‌کنند. به طور کلی این تحلیل‌ها رو می‌توان به ۳ بخش تقسیم کرد:

- تحلیل بنیادی^{۱۱}
- تحلیل فنی^{۱۲}
- تحلیل احساسات^{۱۳}

در ادامه به توضیح مختصری از هر کدام از موارد فوق خواهیم پرداخت.

۱.۲.۱ تحلیل بنیادی

در تحلیل بنیادی پیش‌فرض این است که هر دارایی، بنیاد و زیرساختی داشته و این زیرساخت‌ها بر قیمت آن دارایی مؤثرند. این مؤلفه‌های زیرساختی فقط منحصر به خود آن دارایی و شرکتی که آن را عرضه می‌کند نبوده، بلکه در سطح کلان‌تر و از دید بالاتر باید عوامل مؤثر بر دارایی را نگاه کرد. ارزش نهایی یک دارایی را پارامترهای مختلف، از ابعاد کلان تا کوچک‌ترین پارامترها، مشخص کرده و در تحلیل بنیادی همه این عوامل باید بررسی شود تا بتوان قیمت دارایی موردنظر را تشخیص داد یا پیش‌بینی کرد. مثلاً، در قدم اول، روابط سیاسی کشور با جهان بررسی می‌شود، سپس ساختار اقتصادی کشور، در قدم بعد وضعیت کلی آن صنعت را باید بررسی کرد، و در آخرین مرحله شرایط شرکت ارائه‌دهنده (نقشه راه، تیم مدیریتی، برنامه توسعه و ...) مورد ارزیابی قرار می‌گیرد تا ارزش آن دارایی تعیین شود. بعد از اینکه با تحلیل بنیادی ارزش یک دارایی مشخص شد، یکی از این سه گزاره نتیجه می‌شود: ارزش فعلی دارایی بیشتر از چیزی باشد که در آینده ممکن است رخ دهد، یا قیمت فعلی مناسب است، یا قیمت فعلی کمتر از چشم‌انداز آینده است و در آینده قیمت با جهش روبرو می‌شود. هرکدام از این نتایج به دست آمد، سرمایه‌گذار می‌تواند بهتر تصمیم بگیرد و احتمال سوددهی بیشتر خواهد شد. برای بیت‌کوین می‌توان به تورم جهانی و نرخ بهره (به ویژه ایالات متحده آمریکا)، نرخ جهانی طلا به عنوان دارایی ضد تورمی، قوانین بانک‌های مرکزی بر نقل و انتقال رمزارزها،

Fundamental Analysis^{۱۱}

Technical Analysis^{۱۲}

Sentiment Analysis^{۱۳}

پذیرش بیت کوین به عنوان پول قابل مبادله توسط شرکت‌های بزرگ، نرخ هاش استخراج بیت کوین و ... به عنوان مولفه‌های بنیادین توجه داشت تا ارزندگی آن تعیین شود.

۲.۲.۱ تحلیل فنی

در این نوع تحلیل تنها تغییرات قیمت و حجم معاملات یک دارایی بررسی می‌شود تا رفتار قیمت در آینده مشخص شود. این تحلیل از اواخر قرن ۱۸۰۰ میلادی و بر مبنای تئوری داو^{۱۴} شکل گرفت. [۲] پیش‌فرض این تحلیل این است که با نمودارهای قیمتی می‌توان آینده را پیش‌بینی کرد و همه‌ی آنچه که سرمایه‌گذار به آن نیاز دارد، از جمله اطلاعات بنیادین، در نمودار قیمت و حجم خود را نشان می‌دهد. در تحلیل فنی از شاخص‌های فنی^{۱۵} بسیاری برای تفسیر تغییرات قیمتی - حجمی استفاده می‌شود که از معروف‌ترین آن‌ها می‌توان به میانگین متحرک^{۱۶} و شاخص قدرت نسبی^{۱۷} نام برد.

۳.۲.۱ تحلیل احساسات

همواره گفته می‌شود یکی از اصول معامله‌گری و سرمایه‌گذاری دوری از تصمیمات احساسی است و می‌بایست دانش و اطلاعات خود را به طور مداوم افزایش داد و طبق آنان تصمیم گرفت اما چه تعداد از کسانی که در بازارهای مالی در حال فعالیتند، دانش کافی دارند و از ابزارهای درست هم استفاده می‌کنند؟ یا چند درصد از سرمایه‌گذاران بدون دخالت دادن احساسات خرید و فروش انجام می‌دهند؟ با وجود اینکه همه می‌دانیم برای سرمایه‌گذاری موفق باید براساس دانش معامله کرد، ولی واقعیت این است که خیلی از معامله‌گران براساس حس‌شان خرید و فروش انجام می‌دهند. وقتی تعداد این افراد زیاد شود، به احتمال زیاد روند کلی بازار تغییر می‌کند و نتیجه‌ای برخلاف تحلیل‌های اصولی رخ می‌دهد. بنابراین لازم است که به احساسات معامله‌گران و جو روانی بازار نیز توجه کرد. برای مثال می‌توان به شاخص ترس و طمع^{۱۸} و تحلیل احساسی توییت‌ها^{۱۹} به عنوان نمونه‌های پرکاربرد این روزها اشاره کرد.

^{۱۴} Dow Theory

^{۱۵} Technical Indicator

^{۱۶} Moving Average

^{۱۷} Relative Strength Index (RSI)

^{۱۸} Fear and Greed Index

^{۱۹} Tweet Sentiment Analysis

فصل ۲

یادگیری ماشین

۱.۲ مقدمه

یادگیری ماشین شاخه‌ای از هوش مصنوعی است که به برنامه‌های رایانه‌ای قابلیت یادگیری از داده‌ها بدون استفاده از قوانین از پیش تعیین شده و صریح را می‌دهد. الگوریتم‌های یادگیری ماشین با استفاده از داده‌های گذشته، نتایج آینده را پیش‌بینی می‌کنند. یادگیری ماشین کاربردهای وسیعی در زمینه‌های مختلف مانند زبان‌شناسی، پزشکی، اتوماسیون و ... دارد. این حوزه به تمایز رقابتی مهمی برای بسیاری از شرکت‌ها تبدیل شده است.

۲.۲ یادگیری عمیق

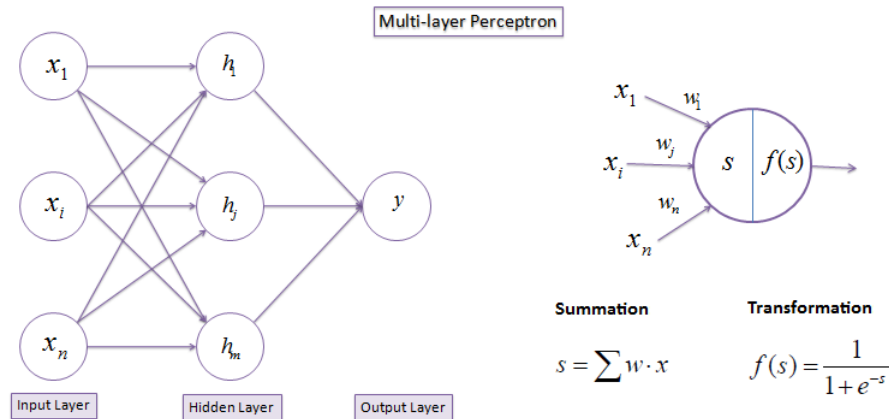
یادگیری عمیق زیرمجموعه‌ای از یادگیری ماشین است که از شبکه‌های عصبی زیستی مغز انسان الهام گرفته است. این نوع یادگیری نیازی به قوانین از پیش طراحی شده نداشته و صرفاً از مقدار زیادی داده برچسب زده شده برای یادگیری استفاده می‌کند. یادگیری عمیق با استفاده از چند لایه شبکه عصبی مصنوعی طراحی شده که هر لایه داده‌های ورودی را به گونه‌ای متفاوت تفسیر می‌کند و در نهایت برای داده ورودی برچسبی را پیش‌بینی می‌کند. به عمل پیش‌بینی برچسب داده‌ها طبقه‌بندی^۱ گفته می‌شود. طبقه‌بندی درست داده‌ها با استفاده از یادگیری عمیق بسیار به نوع و حجم داده‌ها بستگی دارد از این رو مراحل پیش پردازش^۲ ،

Classification^۱
Pre-Processing^۲

استخراج ویژگی^۳ و انتخاب ویژگی^۴ بسیار مهم می‌باشند.

۱.۲.۲ شبکه های پرسپترون

هر شبکه عصبی مصنوعی متشکل از نورون‌های مصنوعی است که از طریق اتصالاتی با نورون‌های دیگر در ارتباط است. هر ارتباط وزن مخصوصی دارد که طی فرایند یادگیری تقویت یا تضعیف می‌شود. پرسپترون چند لایه^۵ یک شبکه عصبی مصنوعی پیش‌خور کننده است زیرا اطلاعات هر نورون از لایه قبلی به لایه بعدی منتقل می‌شود و مسیر اطلاعات حلقه یا دوری ندارد. این شبکه با گرفتن دادگان آموزشی و برچسب آنان تلاش می‌کند با رسم یک خط، صفحه یا ابر صفحه عمل طبقه‌بندی را در فضای دادگان انجام دهد و دسته‌های متفاوت را از یک‌دیگر جدا کند. در یک MLP معمولی همه نورون‌های یک لایه به تک‌تک نورون‌های لایه بعد متصل‌اند. همواره لایه ابتدایی ورودی‌ها را دریافت می‌کند و سپس یک یا چند لایه مخفی وظیفه محاسبات و تفسیر ورودی را دارند تا در انتها به لایه خروجی متصل شوند و آن لایه دسته داده را پیش‌بینی کند. معمولاً تعداد نورون‌های لایه خروجی برابر با تعداد دسته‌های مسئله است.



شکل ۱.۲: شبکه پرسپترونی با یک لایه مخفی [۳]

هر نورون در MLP یک پرسپترون نامیده می‌شود که n مشخصه از داده ورودی دریافت می‌کند و برای هر مشخصه وزنی تخصیص می‌دهد سپس مجموع وزن‌دار مشخصه‌های ورودی همراه با پارامتری به نام بایاس

^۳ Feature Extraction
^۴ Feature Selection
^۵ Multi Layer Perceptron(MLP)

برای تابع فعال‌سازی پرسپترون ارسال می‌شود و آن خروجی را تولید می‌کند.

$$u(x) = bias + \sum_{i=1}^n w_i x_i$$

تابع فعال‌سازی معمولاً پله‌ای می‌باشد تا ورودی‌های مختلف را به خوبی تفکیک کند و تصمیم بگیرد آن نورون فعال شود یا خیر و آن را به لایه بعد ارسال کند.

$$f(x) = \begin{cases} 1 & \text{if } u(x) > 0 \\ 0 & \text{otherwise} \end{cases}$$

اگر از توابع فعال‌سازی غیر خطی استفاده نکنیم وزن‌ها و بایاس فقط یک معادله خطی را شکل می‌دهند که برای حل مسائل پیچیده کمکی به ما نمی‌کند و می‌توان گفت یک شبکه عصبی بدون توابع فعال‌ساز غیر خطی فقط یک مدل رگرسیون خطی ساده است. از پرکاربردترین توابع فعال‌ساز غیرخطی می‌توان به سیگموئید^۶، تانژانت هایپربولیک^۷، واحد یک سو شده‌ی خطی^۸ و بیشینه هموار^۹ اشاره کرد.

۲.۲.۲ تابع اتلاف

تابع اتلاف^{۱۰} تفاوت بین برجسب واقعی با برجسب پیش‌بینی شده را محاسبه می‌کند و از آن برای تخمین میزان خطای شبکه و میزان تغییر مورد نیاز در وزن‌های شبکه استفاده می‌شود. یکی از ساده‌ترین توابع اتلاف، میانگین مربعات خطا نام دارد که به صورت زیر محاسبه می‌شود:

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2$$

۳.۲.۲ بازتنظیم پارامترها

برای داشتن یک شبکه عصبی خوب نیاز است وزن‌ها را به طور مداوم تغییر دهیم تا میزان اتلاف شبکه کمینه شود پس با یک مسئله بهینه‌سازی روبه‌رو هستیم. پس از محاسبه مقدار زیان در مرحله قبل، این مقدار به

Sigmoid^۶

Hyperbolic tangent (tanh)^۷

Rectified linear unit (ReLU)^۸

Softmax^۹

Loss Function^{۱۰}

صورت معکوس (از لایه خروجی به سمت لایه اول) در شبکه انتشار می‌یابد و با استفاده از مفهوم گرادیان^{۱۱}، وزن‌های شبکه عصبی پرسپترون چند لایه به‌روزرسانی می‌شوند. مراحل یادگیری با استفاده از الگوریتم پس انتشار خطا در شبکه عصبی پرسپترون چند لایه به شرح زیر است:

- با فرض داشتن تعداد J نود لایه نهان و یک نمونه ورودی با تعداد n ویژگی در یک شبکه عصبی پرسپترون چند لایه، خروجی نودهای لایه آخر (لایه خروجی)، از طریق رابطه زیر مشخص می‌شوند. در این رابطه، y_i خروجی نود i هر کدام از لایه‌های نهان را نشان می‌دهد.

$$y_i = f\left(\sum_{j=1}^J w_{ji}y_j\right)$$

- در مرحله بعد، با توجه به مقادیر خروجی مورد انتظار و خروجی پیش‌بینی شده برای هر کدام از نودهای لایه خروجی (i)، مقدار «دلتا» (Delta) برای هر کدام از نودها، توسط رابطه زیر محاسبه می‌شود. در این رابطه، $y_{expected_i}$ خروجی مورد انتظار نود i در لایه خروجی و $y_{predicted_i}$ ، خروجی پیش‌بینی شده توسط نود i در لایه خروجی است.

$$\delta_i = (y_{predicted_i} - y_{expected_i}) \cdot (y_{expected_i}) \cdot (1 - y_{expected_i})$$

- برای هر کدام از نودهای لایه نهان (j)، مقدار دلتا از طریق رابطه زیر محاسبه می‌شود.

$$\delta_j = (O_j) (1 - O_j) \sum_i w_{ji} \delta_i$$

- برای تنظیم کردن وزن‌های تمامی نودها در شبکه عصبی پرسپترون چند لایه، از روش گرادیان کاهشی^{۱۲} استفاده می‌شود. در این روش، وزن‌های تمامی نودها از طریق رابطه عمومی زیر تنظیم می‌شوند.

$$\Delta w_{ij} = \eta \delta_j y_n$$

^{۱۱} Gradient
^{۱۲} Gradient Descent

- به طور خاص، برای تنظیم وزن نود j از لایه نهان (از نود n لایه ورودی به سمت نود j لایه نهان)، از رابطه زیر استفاده می‌شود. در این رابطه، x_n مقدار نود n لایه ورودی را نشان می‌دهد.

$$\Delta w_{nj} = \eta \delta_j x_n$$

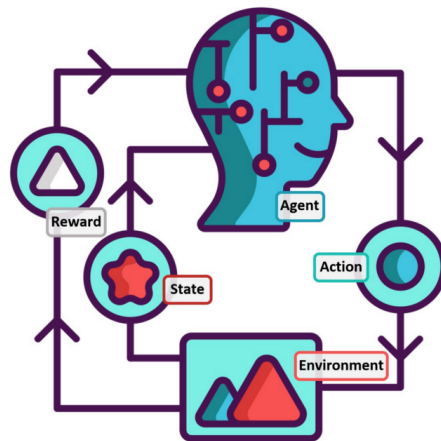
- همچنین، برای تنظیم وزن نود i از لایه خروجی (از نود j لایه نهان به سمت نود i لایه خروجی)، از رابطه زیر استفاده می‌شود. در این رابطه، O_j مقدار خروجی نود j لایه نهان را نشان می‌دهد.

$$\Delta w_{ji} = \eta \delta_i O_j$$

در روابط نمایش داده شده، η نرخ یادگیری شبکه عصبی پرسپترون چند لایه است؛ آبرپارامتری که برای تنظیم نرخ تغییرات وزن‌های شبکه عصبی مورد استفاده قرار می‌گیرد. معمولاً در ابتدای کار شبکه عصبی پرسپترون چند لایه، مقدار کوچکی برای این آبرپارامتر در نظر گرفته می‌شود. با این حال، توصیه می‌شود که در طی فرایند یادگیری و با افزایش تعداد تکرارهای این شبکه، به تدریج مقدار این آبرپارامتر افزایش پیدا کند.

۳.۲ یادگیری تقویتی

یادگیری تقویتی شاخه‌ای از یادگیری ماشین است که طی آن عامل هوشمند با آزمون و خطا نحوه تعامل بهینه با محیط را یاد می‌گیرد تا سود تجمعی کسب شده از محیط را بیشینه کند. در یادگیری تقویتی برخلاف یادگیری نظارت شده نیازی به خروجی برچسب‌دار برای دادگان ورودی نداریم و عامل هوشمند تنها با پاداش‌هایی که از محیط دریافت می‌کند می‌بایست بهترین تصمیم ممکن را تشخیص دهد و همچنین تاثیر انتخاب فعلی خود بر آینده را نیز در نظر گیرد. یادگیری تقویتی الهام گرفته از تعامل نوع بشر با محیط و یادگیری بر اساس تجربیات خود است. تصور کنید که کودکی فصل زمستان در یک اتاق است. کودک شومینه را می‌بیند و به آن نزدیک می‌شود. شومینه گرم است، این امر مثبت تلقی شده و کودک احساس مثبتی دارد (پاداش +۱). در اینجا کودک می‌فهمد که آتش چیز مثبتی است. سپس، کودک سعی می‌کند که آتش را لمس کند. اما آتش دست کودک را می‌سوزاند (پاداش -۲). در اینجا کودک می‌فهمد می‌بایست در فاصله مناسب از آتش قرار بگیرد، زیرا با وجود اینکه گرما تولید می‌کند. اما نزدیک شدن بیش از اندازه موجب سوختن می‌شود.



شکل ۲.۲: شمایی از یادگیری تقویتی [۴]

۱.۳.۲ فرآیند تصمیم‌گیری مارکوف

این فرآیند ابزار اصلی ما برای حل مسائل یادگیری تقویتی است و آن را با یک پنج‌تایی (S, A, R, P, γ) مشخص می‌کنند که:

- S مجموعه‌ای از وضعیت‌های عامل و محیط به نام فضای وضعیت هاست.
- A مجموعه‌ای از اعمال ممکن برای عامل هوشمند در محیط داده شده به نام فضای اعمال است.

- R جایزه داده به عامل توسط محیط برای رفتن از حالت s به s' با عمل a است.
 - P احتمال رفتن از حالت s به s' در زمان t با عمل a را نشان می دهد.
 - γ عددی بین ۰ و ۱ است که نشان دهنده میزان تاثیر پاداش های احتمالی آینده بر پاداش فعلی می باشد که هرچه به صفر نزدیک تر باشد آینده نگری عامل کمتر می شود.
- برای حل مسئله با یادگیری تقویتی نیاز به یافتن استراتژی بهینه برای یک عامل در محیط داریم. یک استراتژی π تابعی از فضای حالت به فضای اعمال است. Π^* یک استراتژی بهینه شناخته می شود اگر:

$$\Pi^* = \operatorname{argmax}_{\Pi} E \left[\sum_{t \geq 0} \gamma^t r_t | \Pi \right]$$

به بیان دیگر Π^* استراتژی ای است که امید ریاضی مجموع پاداش های دریافتی را با در نظر گرفتن نرخ تنزیل γ بیشینه کند.

۲.۳.۲ یادگیری Q

برای بدست آوردن استراتژی بهینه در فرآیند تصمیم گیری مارکوف می توان از این روش بهره برد. ابتدا تابعی از فضای حالت به اعداد حقیقی تعریف می کنیم و آن را تابع ارزش می نامیم که نمایانگر میزان مطلوبیت حضور در حالت s هنگام استفاده از استراتژی π می باشد.

$$V^{\Pi}(s) = E \left[\sum_{t \geq 0} \gamma^t r_t | s_0 = s, \Pi \right]$$

به بیان دیگر این تابع ارزش نشان دهنده امید ریاضی مجموع پاداش های تنزیل یافته در صورت شروع از وضعیت s تا وضعیت انتهایی است پس وضعیتی ارزشمندتر است که با قرارگرفتن در آن بتوانیم مجموع پاداش های آینده خود را بیشینه کنیم.

۳.۳.۲ تابع Q

این تابع نگاشتی از فضای حالات و اعمال به اعداد حقیقی است که نشان دهنده میزان مطلوبیت عمل a در وضعیت s تحت استراتژی π است.

$$Q^{\Pi}(s, a) = E \left[\sum_{t \geq 0} \gamma^t r_t | s_0 = s, a_0 = a, \Pi \right]$$

Q^* را تعریف می‌کنیم به طوری که:

$$Q^\Pi(s, a) = \max_{\Pi} E \left[\sum_{t \geq 0} \gamma^t r_t | s_0 = s, a_0 = a, \Pi \right]$$

پس می‌توان گفت Q^* تابع تحت Q استراتژی بهینه است. در صورت محاسبه Q^* محاسبه Π^* بسیار ساده خواهد بود زیرا کافی است به ازای هر وضعیت s عملی که $Q(s, a)$ را بیشینه می‌کند انتخاب شود.

$$\Pi^*(s) = \operatorname{argmax}_a Q^*(s, a)$$

۴.۳.۲ یادگیری Q^*

می‌دانیم Q^* در رابطه بلمن صدق می‌کند [۵]

$$Q^*(s, a) = E_{s' \sim \mathcal{E}} \left[r + \gamma \max_a Q^*(s', a) | s, a \right]$$

یادگیری Q^* با کمک معادله‌ی بلمن زیر امکان‌پذیر است.

$$Q_{i+1}(s, a) = E_{s' \sim \mathcal{E}} \left[r + \gamma \max_a Q_i(s', a) | s, a \right]$$

پس اگر $i \rightarrow \infty$ آن‌گاه $Q_i \rightarrow Q^*$ و با تکرار معادله بالا می‌توانیم Q^* را به دست آوریم. هرچند این روش به نتیجه می‌رسد ولی باید تمام $Q(s, a)$ ها محاسبه شوند که برای مسائلی با فضای حالت بسیار بزرگ جواب‌گو نخواهد بود.

۴.۲ یادگیری تقویتی عمیق

همان‌طور که دیدیم یادگیری تقویتی برای مسائلی با فضای حالات بسیار بزرگ کارا نیست پس می‌توانیم از شبکه‌های عصبی عمیق برای تخمین مطلوبیت هر حالت بهره ببریم.

۱.۴.۲ یادگیری عمیق Q

شبکه‌های عصبی عمیق تخمین زندگان قدرتمندی برای توابع پیچیده هستند بنابراین می‌توانیم آنان را برای تخمین Q^* به کارگیریم.

$$Q(s, a; \Theta) \approx Q^*(s, a)$$

که Θ پارامترهای شبکه است.

۲.۴.۲ تابع اتلاف

این تابع طی مراحل یادگیری وظیفه محاسبه خطای معادله بلمن را برعهده دارد.

$$L_i(\Theta_i) = E_{s,a \sim \rho(\cdot)} [(y_i - Q(s, a; \Theta_i))^2] \quad y_i = E_{s' \sim \varepsilon} [r + \gamma \max_a Q(s', a; \Theta_{i-1}) | s, a]$$

که y_i مقدار هدف،

$\rho(\cdot)$ توزیع احتمالی به نام توزیع رفتار و

ε شبیه‌ساز محیط است.

۳.۴.۲ پس‌انتشار خطا

برای محاسبه پس‌انتشار خطا کافی است مشتق تابع اتلاف را نسبت به پارامترهای شبکه بگیریم.

$$\nabla_{\Theta_i} L_i(\Theta_i) = E_{s,a \sim \rho(\cdot); s' \sim \varepsilon} [(r + \gamma \max_a Q(s', a; \Theta_{i-1}) - Q(s, a; \Theta_i)) \nabla_{\Theta_i} Q(s, a; \Theta_i)]$$

البته این رابطه بار محاسباتی زیادی دارد و معمولاً بجای آن از گرادیان کاهش تصادفی استفاده می‌شود.

۴.۴.۲ بازپخش تجربه

یادگیری با الگوریتم گرادیان کاهشی تصادفی نیازمند مستقل بودن توزیع نمونه‌هاست که در اکثر محیط‌ها رفتار عامل هوشمند بر محیط اثر گذاشته و شرط استقلال را نقض خواهد کرد. بدین منظور تجربه‌های عامل هوشمند را در انباری ذخیره می‌کنیم و هر بار به طور تصادفی تعدادی از آنان را برای یادگیری مورد استفاده قرار می‌دهیم تا تجارب پشت سر هم، که احتمالا به یک‌دیگر وابسته‌اند انتخاب نشوند و بتوانیم نمونه‌ها را مستقل فرض کنیم که به این روش بازپخش تجربه می‌گویند.

۵.۴.۲ اکتشاف در برابر انتفاع

یکی از مهم‌ترین مسائل در الگوریتم‌های مکاشفه‌ای که می‌توان بررسی اعمال گوناگون در محیط یادگیری تقویتی را نیز در آن دسته قرار داد ایجاد توازن بین اکتشاف و انتفاع است. در هر مرحله فعل و انفعال بین عامل هوشمند و محیط، عامل هوشمند می‌تواند بر اساس استراتژی آموخته خود از تجربیات پیشین، در جهت انتفاع عمل کند یا اینکه با انتخاب عملی تصادفی به اکتشاف فضای حالات و پاداش‌های دیده نشده دست بزند که ممکن است باعث بهبود استراتژی فعلی شود.

۶.۴.۲ اکتشاف ϵ - حریصانه

در این روش هر مرحله با احتمال ϵ به اکتشاف می‌پردازیم و با احتمال $1 - \epsilon$ بر اساس استراتژی فعلی عمل می‌کنیم. معمولا مقدار ϵ در ابتدا زیاد است و با گذر زمان و یادگیری بهتر استراتژی کاهش می‌یابد.

فصل ۳

پیاده‌سازی و نتایج

۱.۳ بیان مسئله

ساخت یک ربات معامله‌گر سال‌هاست موضوع بحث تحلیل‌گران بازارهای مالی و برنامه‌نویسان است و افراد و موسسات مالی از راه‌های گوناگون برای رسیدن به یک معامله‌گر خودکار سودده تلاش کرده‌اند و تخمین زده می‌شود ۹۲٪ حجم معاملات فارکس را معاملات الگوریتمی تشکیل دهد. [۶] در این پروژه تلاش می‌کنیم رباتی معامله‌گر برای جفت ارز BTC/USDT در پنجره زمانی^۱ ۱ ساعته طراحی کنیم که علاوه بر سودده بودن ریسک‌گریز هم باشد.

۲.۳ محیط مسئله

با استفاده از داده‌های قیمتی - حجمی صرافی بایننس^۲ به عنوان بزرگ‌ترین صرافی ارزهای دیجیتال یک محیط ایزوله برای خرید و فروش رمزارز بیت‌کوین در نظر گرفته شده که به دلیل دارایی اندک عامل معامله‌گر از تاثیر معاملات ما بر روند بازار چشم‌پوشی شده است. برای واقعی‌تر شدن شبیه‌سازی، کارمزد معاملات نیز همچون حالت پایه صرافی بایننس در نظر گرفته شده است. عامل معامله‌گر ما در ابتدا 1000 USD و 0.05 BTC در حساب خود دارد و در هر مرحله می‌تواند یکی از ۵ تصمیم زیر را اتخاذ کند:

۱. ۷۵٪ دارایی تتر خود را به بیت‌کوین تبدیل کند.

^۱ Time Frame
^۲ Binance

۲. ۲۵% دارایی تتر خود را به بیت کوین تبدیل کند.

۳. معامله‌ای انجام ندهد.

۴. ۲۵% دارایی بیت کوین خود را به تتر تبدیل کند.

۵. ۷۵% دارایی بیت کوین خود را به تتر تبدیل کند.

برخلاف اکثر کارهای مشابه که تنها ۳ حالت خرید کامل، نگهداری و فروش کامل را پشتیبانی می‌کنند این مجموعه اعمال انتخاب شد تا آزادی عمل ربات معامله‌گر افزایش یابد و به واقعیت نزدیک‌تر باشد از طرفی این اجازه را به ربات می‌دهد تا با نگهداری همزمان هر دو رمزارز بتواند بین سود و ریسک تعادل ایجاد کند. در ابتدا فضای اعمال بزرگ‌تر در نظر گرفته شده بود اما به دلیل سختی آموزش مدل با دسته‌های متنوع‌تر، از ۱۱ حالت به ۵ حالت کاهش یافت.

همچنین فرض شد پس از تکمیل یک شمع یک ساعته و دریافت آن توسط ربات، سفارش انتخاب شده به صورت محدود^۳ و با قیمت آغازین شمع بعدی انجام شود که می‌دانیم چنان سرعت عملی دست نیافتنی است زیرا بلافاصله بعد از اتمام یک شمع، شمع بعدی آغاز می‌شود و تا تصمیم ربات بخواهد به صرافی برسد به احتمال زیاد قیمت آن لحظه با قیمت آغازین شمع متفاوت است و ممکن است دیگر بازار آن قیمت را در شمع فعلی نبیند و یا حجم سفارش ما به طور کامل انجام نشود و سفارش ما به درستی انجام نشود برای حل این مشکل می‌بایست از داده‌های دفترچه ثبت معاملات استفاده می‌شد که البته باز هم تضمینی مبنی بر درستی عملکرد در دنیای واقعی وجود نداشت چون به هر حال هر سفارش بر بازار اثر می‌گذارد. به دلیل تاثیر اندک این مشکلات و بار محاسباتی افزودن دفترچه ثبت معاملات به ربات از آن‌ها صرف نظر شد و می‌دانیم در هر صورت می‌توان با سفارش‌گذاری بازار^۴ در اکثر مواقع با اختلاف اندک (کمتر از 0.1%) قیمتی سفارش خود را تکمیل کرد.

Limit Order^۳
Market Order^۴

۱.۲.۳ داده‌های ورودی

همان‌طور که پیش‌تر گفته شد تاریخچه قیمتی-حجمی جفت ارز BTC/USDT از صرافی بایننس دریافت شد. این تاریخچه از ۱۷/۸/۲۰۱۷ تا ۲۶/۶/۲۰۲۲ را شامل می‌شود که در مجموع حاوی ۴۲۵۵۰ شمع یک ساعته این جفت ارز می‌باشد و روندهای مختلف بیت‌کوین را به خوبی نمایندگی می‌کند. در این بازه زمانی قیمت بیت‌کوین بین ۳۰۰۰ دلار تا ۶۹۰۰۰ دلار در نوسان بوده است.



شکل ۱.۳: نمودار شمعی BTC/USDT

برای هر ساعت از معاملات BTC/USDT یک شمع و دو داده حجمی، حجم معاملات به بیت‌کوین و حجم معاملات به تتر، موجود است. شمع ابتدایی برای مرحله یادگیری مورد استفاده قرار خواهند گرفت و ۳۰۰۰ شمع انتهایی (۱۳۰ روز) برای تست نهایی استراتژی آموخته شده به کار گرفته خواهند شد.

۲.۲.۳ استخراج ویژگی

از آن جایی که در تحلیل فنی شکل شمع‌ها (اندازه بدنه و سایه‌ها) اهمیت فراوانی دارند و روندها و اختلافات درصدی قیمت‌ها در بسیاری مواقع از عدد دقیق قیمت‌ها مهم‌ترند تصمیم بر این شد که بدنه و سایه بالا و سایه پایین هر شمع به درصد محاسبه شود. برای مثال اگر قیمت آغازین شمع ۱۰۰۰۰ تتر و قیمت پایانی ۱۰۱۰۰ تتر بوده است بدنه شمع مثبت ۰.۱۰ می‌باشد.

بیش‌ترین (کم‌ترین) قیمت دیده شده در یک بازه زمانی خاص طبق دیدگاه تحلیل فنی سطح مقاومت (حمایت) نامیده می‌شود و نزدیک شدن یا عبور از آنان می‌تواند حاکی از یک روند جدید باشد. بدین منظور بیش‌ترین و کم‌ترین قیمت دیده شده را در ۳ بازه زمانی (۲۴ ساعته، ۷ روزه و ۱ ماهه) برای هر سطر از داده (اطلاعات قیمتی - حجمی یک ساعت) محاسبه کردیم و مانند بخش قبل اختلاف درصدی آن را نسبت به قیمت پایانی شمع فعلی بدست آوردیم. حجم معاملات نیز ارزش تحلیلی بالایی دارد پس میانگین حجم معاملات، به بیت‌کوین و تتر، در همان ۳ بازه زمانی قبلی را برای هر سطر از داده‌ها محاسبه کردیم و برای نرمال سازی نسبت حجم هر داده به میانگین حجمی در بازه‌های زمانی را بدست آوردیم.

در دنیای تحلیل فنی از شاخص‌های بسیاری استفاده می‌شود که همه آنان طبق قیمت و حجم یک یا چند شمع بدست می‌آیند و برای مقدار و تغییرات آنان تفسیرهایی عنوان شده است. در این پروژه تنها از دو شاخص میانگین متحرک ساده و قدرت نسبی استفاده شده است که بین معامله‌گران مقبولیت بیشتری دارند. شاخص میانگین متحرک ساده نمایان‌گر میانگین قیمت در n شمع گذشته است و شاخص قدرت نسبی بر اساس میانگین قدرت شمع‌های صعودی و شمع‌های نزولی عددی در بازه ۰ تا ۱۰۰ نتیجه می‌دهد که هرچه عدد بیشتر باشد روند طی شده از قدرت صعودی بیشتری برخوردار بوده است و هرچه کمتر باشد نشان دهنده قدرت نزولی روند طی شده است.

$$RSI_{14} = 100 - \left[\frac{100}{1 + RS_{14}} \right], RS_{14} = \frac{\text{UP closes candle's } 14 \text{ Avg}}{\text{DOWN closes candle's } 14 \text{ Avg}}$$

۴ میانگین متحرک ساده با پنجره‌های زمانی ۹، ۱۸، ۴۰، ۱۰۰ شمع که نشان‌دهنده میانگین قیمت در دوره‌های بسیار کوتاه مدت ۹ ساعته تا میان مدت ۱۰۰ ساعته هستند محاسبه شده و اختلاف درصدی آن با قیمت پایانی هر شمع به ویژگی‌های داده افزوده شده است. همچنین یک شاخص قدرت نسبی با پنجره زمانی ۱۴ شمعی هم محاسبه و اضافه شده است.

۳.۲.۳ پاداش

می‌توان گفت سیستم پاداش‌دهی و داده‌های وضعیت محیط، مهم‌ترین بخش‌های هر پروژه یادگیری تقویتی‌ای هستند زیرا یادگیری تنها در صورت شناخت صحیح محیط و پاداش‌دهی مناسب میسر خواهد بود. بسیاری از کارهای مشابه میزان سود عامل معامله‌گر را پاداش آن در نظر گرفتند اما رویکرد کار دو ایراد اساسی دارد: از ریسک معاملات چشم‌پوشی می‌شود و ممکن است صرفاً بخاطر روند کلی بازار نتیجه خوب یا بد باشد و اینکه نمی‌توان زمان دقیقی برای به ثمر رسیدن یک معامله در نظر گرفت پس هر معامله ما پاداشش در آینده نامعلوم مشخص می‌شود و اگر صرفاً سود کلی ربات را در نظر بگیریم به درستی مشخص نخواهد شد کدام اعمال چه تاثیری گذاشته‌اند که در نهایت سود کلی این‌چنین شده است از طرفی اگر اعمال ربات معامله‌گر را به هم پیوسته در نظر بگیریم و پاداش‌های احتمالی آینده را برای هر حالت محاسبه کنیم با فضای حالات فوق‌العاده بزرگ‌تری روبه‌رو خواهیم بود که یادگیری را دشوارتر خواهد کرد. همچنین معامله‌گران حرفه‌ای می‌گویند نیاز نیست همواره در حال خرید و فروش باشید؛ بیشتر اوقات نظاره‌گر باشید و تنها در نقاط خاصی وارد معامله شوید. با این تفاسیر تصمیم‌گرفتم پاداش هر عمل را جدای از اعمال آینده در نظر بگیرم و آن را طبق روند قیمتی چند شمع آینده مشخص کنم تا حالت‌های مختلف بازار بهتر شناسایی شوند و بدون در نظر گرفتن شرایط فعلی (دارایی عامل) و پاداش‌های آینده آن، تنها بهترین عمل در شرایط فعلی بازار آموخته شود.

۴.۲.۳ ریسک

فرض کنید سهامی را در قیمت ۱۰۰ دلار خریداری کرده‌اید سپس قیمت آن به ۹۰ دلار می‌رسد و پس از مدتی به ۱۰۵ دلار افزایش می‌یابد و با آن قیمت سهم را می‌فروشید. در نگاه اول در این معامله ۵٪ سود کرده‌اید اما برای این سود چه ریسکی را متحمل شده‌اید؟ بله ۱۰٪. آیا ارزشش را داشت؟ بسیاری بر این باورند که خیر. از زوایای دیگر اگر شما آن افت قیمت را پیش‌بینی می‌کردید می‌توانستید بجای ۱۰۰ دلار در قیمت ۹۵ دلار سهم را بخرید و در پایان سود بیشتری با ریسک کمتری کسب کنید. شاخص‌های زیادی برای میزان ریسک یک استراتژی معاملاتی ارائه شده است که از معروف‌ترین آنان می‌توان به شاخص کاهش^۵ اشاره کرد. این شاخص بیش‌ترین اختلاف یک قله تا یک قعر پس از آن را به عنوان میزان ریسک استراتژی معاملاتی در نظر می‌گیرد. به بیان دیگر استراتژی معاملاتی شما در بدترین بازه عملکردی آن چند درصد متحمل ضرر شده است.

^۵ Drawdown

۳.۳ تابع پاداش

با توضیحاتی که در قسمت‌های قبلی داده شد این تابع از ۳ مولفه تشکیل شده است:

۱. اختلاف میانگین وزن دار قیمت در ۱۲ شمع پیش‌رو با قیمت پایانی شمع فعلی: میانگینی از قیمت پایانی شمع‌های آینده که هرچه به جلو می‌رویم وزن هر شمع افزایش می‌یابد (دنباله‌ای حسابی) بدست می‌آوریم سپس اختلاف درصدی آن را با قیمت پایانی فعلی محاسبه می‌کنیم

۲. اختلاف بیش‌ترین کاهش در ۱۲ شمع پیش‌رو با قیمت پایانی شمع فعلی: کم‌ترین قیمتی که در شمع‌های پیش‌رو مشاهده خواهد شد را پیدا می‌کنیم و اختلاف درصدی آن با قیمت پایانی فعلی را محاسبه می‌کنیم.

۳. هزینه معامله: ورود به هر معامله علاوه بر کارمزد باعث پیچیده شدن تحلیل استراتژی و دشوار شدن مسئله طبقه‌بندی وضعیت‌ها می‌شود برای همین می‌بایست هزینه‌ای در نظر گرفته شود که تنها وقتی عامل سود خوبی را برای آینده پیش‌بینی کرد وارد معامله شود.

با جمع این ۳ مولفه پاداش هر عمل مشخص می‌شود.

۴.۳ شبکه عصبی

برای تخمین تابع Q از یک شبکه پرسپترون با دو لایه مخفی استفاده شده است. لایه ورودی شامل ۲۵۰ نورون است که هر کدام از آنان یک ویژگی از حالت فعلی محیط را دریافت می‌کنند و سپس هر نورون لایه ورودی به تمام ۱۰۰ نورون لایه مخفی اول متصل می‌شوند و آنان نیز به ۳۵ نورون لایه مخفی دوم و در انتها نورون‌های لایه مخفی دوم به ۵ نورون خروجی که هر کدام برای یک عمل خاص تعریف شده‌اند متصل می‌شوند. پس تعداد نورون‌های شبکه به صورت (5,35,100,250) می‌باشد.

نرخ یادگیری 0.001 در نظر گرفته شده است و تجربیات عامل در دسته‌های ۶۵ تایی برای یادگیری فراخوانده می‌شوند.

۵.۳ کتابخانه‌های مورد استفاده

برای مرحله پیش‌پردازش و استخراج ویژگی از کتابخانه pandas که مخصوص کار با داده‌هاست استفاده شد و محاسبه شاخص‌های فنی با کتابخانه talib به انجام رسید. محیط عامل^۶، عامل هوشمند و انباره تجربیات آن با کمک کتابخانه TensorFlowAgent طراحی شد و شبکه عصبی نیز با TensorFlow ساخته شد.

۶.۳ نتایج

پس از ۱۰ بار گذراندن داده‌های آموزشی با اعمال تصادفی، عامل معامله‌گر به استراتژی‌ای رسید که در بازار به‌شدت نزولی ماه‌های اخیر (اواسط فوریه ۲۰۲۲ تا اواخر ژوئن ۲۰۲۲) که بیت‌کوین از ۴۴۰۰۰ دلار تا ۲۰۰۰۰ دلار سقوط کرد به سود ۱۲٪ رسید و بیشینه کاهش آن ۴٪ بود در حالی که حساب سرمایه‌گذاری (بدون معامله) ۳۴٪ دارایی خود را از دست داد.

```
Initial: 3111.7 $, Final: 3488.3 $, Available BTC: 0.0027, MaxDrawDown: 0.04
Holder: 2071.6 $, Last BTC price: 21432.69 $
Actions:{0: 31, 1: 5, 2: 3004, 3: 0, 4: 95}
```

شکل ۲.۳: خروجی ارزیابی

همانطور که از دیکشنری اعمال پیداست عامل معامله‌گر در ۹۵٪ اوقات فقط ناظر بوده و وارد معامله‌ای نشده است که طبق توضیحات گذشته رفتار بسیار حرفه‌ای می‌باشد و به دلیل نزولی بودن بازار بیشتر فروشنده بوده است تا خریدار. طبق این موارد و اعداد بدست آمده می‌توان یادگیری عامل را موفقیت‌آمیز قلمداد کرد.

فصل ۴

جمع‌بندی و پیشنهادات

۱.۴ جمع‌بندی

همان‌طور که دیدیم استفاده از هر چهار قیمت یک شمع و در نظر گرفتن بالاترین و پایین‌ترین قیمت‌های اخیر، داده‌های ارزشمندی را برای یادگیری مدل فراهم کرد و به کارگیری یادگیری تقویتی عمیق نیز به قدرت الگوریتم معامله‌گر ما افزود تا در نهایت برخلاف نمونه‌های مشابه که تنها با قیمت پایانی آموزش می‌بینند و عمل‌کردی بهتر از یک مدل رگرسیونی ندارند، در بازار شدیداً نزولی به سوددهی رسید.

۲.۴ پیشنهادات برای ادامه‌ی کار

در این پروژه به دلیل محدودیت‌های پردازشی تنها یک جفت ارز مورد بررسی قرار گرفت اما برای افزایش جامعیت استراتژی نیاز است عامل، با داده‌های چند جفت ارز مطرح دیگر بازار رمزارزها آموزش ببیند و ارزیابی‌هایی نیز محدود به یک جفت ارز و بازه زمانی خاصی نباشد. هم‌چنین با تست پنجره‌های زمانی مختلف و تغییر تعداد شمع‌های ورودی و افق زمانی پاداش می‌توان انتظار داشت نتایج بهتری محقق شود. بخش دیگری که جای پیشرفت و آزمون زیادی دارد شبکه عصبی مورد استفاده عامل است؛ در ابتدای پروژه از یک شبکه عصبی بازگشتی^۱ استفاده شد اما به دلیل بار محاسباتی به یک شبکه پرسپترون ساده بسنده شد. تغییر نوع شبکه و تعداد لایه‌های مخفی و شمار نورون‌های آنان ممکن است قدرت یادگیری عامل را با یک جهش روبه‌رو کند.

^۱ Recurrent Neural Network

واژه‌نامه فارسی به انگلیسی

Cryptocurrency	رمزارز
Stable Coin	رمزارز پایدار
Tether	تتر
Forex	فارکس
Candle	شمع
Open price	قیمت آغازین
Close price	قیمت پایانی
Candle's body	بدنه شمع
Candle's shadow	سایه شمع
Fundamental Analysis	تحلیل بنیادی
Technical Analysis	تحلیل فنی
Sentiment Analysis	تحلیل احساسات
Technical Indicator	شاخص‌های فنی
Moving Average	میانگین متحرک
Relative Strength Index	شاخص قدرت نسبی
Fear and Greed Index	شاخص ترس و طمع
Tweet Sentiment Analysis	تحلیل احساسی توییت‌ها
Classification	طبقه‌بندی
Pre-Processing	پیش پردازش
Feature Extraction	استخراج ویژگی
Feature Selection	انتخاب ویژگی
Multi Layer Perceptron	پرسپترون چند لایه
Sigmoid	سیگموید

Hyperbolic tangent	تانزانٹ هایپربولیک
Rectified linear unit	واحد یک سو شده ی خطی
Softmax	بیشینه هموار
Loss Function	تابع اتلاف
Gradient Descent	گرادیان کاهش
Time Frame	پنجره زمانی
Limit Order	سفارش محدود
Drawdown	کاهش
Environment	محیط
Recurrent Neural Network	شبکه عصبی بازگشتی

منابع

- [١] <https://learn.bybit.com/candlestick/best-candlestick-patterns/>
- [٢] Dow Theory https://en.wikipedia.org/wiki/Dow_theory
- [٣] https://www.saedsayad.com/artificial_neural_network_bkp.htm
- [٤] <https://towardsdatascience.com/reinforcement-learning-fda8ff535bb6?gi=8df1e8000479>
- [٥] F.-F. Li, J. Johnson, and S. Yeung, “Lecture notes in deep reinforcement learning,” May,2017
- [٦] A study in 2019 showed that around 92% of trading in the Forex market was performed by trading algorithms https://en.wikipedia.org/wiki/Algorithmic_trading

Abstract

One major goal that investors consistently try to achieve is to predict the price of the market in the future and as a result, they adopt various approaches to do so. One of these approaches which has recently come to attention, uses different artificial intelligence (AI) techniques for predicting the price. Deep reinforcement learning -which is a combination of reinforcement learning and deep learning- is a subset of AI that is designed to provide an optimal solution to the issues arisen in many fields such as robotic, game development, etc. In this report, the author discusses a trading agent designed on the basis of deep reinforcement leaning which receives the history of price and volume of BTC/USDT transactions as input. This agent learns to recognize price-volume patterns in order to decide whether to sell or to buy BTC/USDT, while the element of money management (profitability and risk prevention) is defined to be of significant importance in the process of decision-making.



College of Science
School of Mathematics, Statistics, and Computer Science

Using machine learning methods to trading in the cryptocurrency market

Ali Bajelan

Supervisor: Bagher BabaAli

A thesis submitted in partial fulfillment of the requirements for
the degree of B.Sc. in Computer Science

July 2022