زن زندگی آزادی

دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر

گزارش پروژه امتیازی
درس یادگیری تعاملی
پاییز ۱۴۰۱

نام و نام خانوادگی
سیاوش رزمی
شماره دانشجویی
۸۱۰۱۰۰۳۵۲

فهرست

خلاصه

موضوع این پروژه پیاده‌سازی معامله‌گری الگوریتمی با استفاده از الگوریتم های یادگیری تقویتی است هدف در این پروژه این است که با استفاده از الگوریتم های مذکور عاملی را آموزش دهیم که به شکل خودکار به معامله‌گری در بازار های مالی بپردازد، برای پیاده‌سازی این عامل از الگوریتم Deep Recurrent Q network استفاده شده است که با توجه به همبستگی زمانی مشاهدات در محیط به نظر مناسب‌تر از مابقی الگوریتم ها می‌باشد، همچنین برای آموزش و تست از محیط Anytrading کتابخانه‌ی Gym استفاده شد که به دلیل وجود برخی محدودیت‌ها قسمت‌هایی از آن تغییر داده شد.

مقدمه

صنعت نو ظهور Fintech امروزه یکی از پر رونق ترین زیر مجموعه های صنایع مربوط به تکنولوژی محسوب می‌گردد، بازیگران این صنعت هدف نسبتاً ساده ایی دارند: چطور از تکنولوژی های روز برای پیشبرد فعالیت های مربوط به مباحث مالی استفاده کنیم؟، پیش بینی میشود که در سال های آتی فینتک بسیاری از مسائل مربوط تصمیم گیری در حوزه مالی مانند معامله‌گری، سرمایه‌گذاری، مدیریت ریسک، مدیریت پورتفوی سرمایه‌گذاری، تشخیص کلاه‌برداری مالی و مشاوره مالی را منقلب کند، مسائل ذکر شده بالا به دلیل ماهیت ترتیبی(Sequential) و تصادفی (Stochastic) و محیط های نیمه مشاهده پذیر (Partially observable) معمولاً بسیار پیچیده و حل آنها دشوار است، یکی از بخش های کلیدی و پر طرفدار حوزه فینتک معامله‌گری الگوریتمی است که سعی دارد با استفاده از قدرت پردازشی رایانه ایی و همچنین قوانین ریاضی به معامله‌گری در بازارهای مالی بپردازد از متد های مرسوم معامله گری استفاده از تحلیل تکنیکال، فاندمنتال و یا متد های کمی (Quantitative) است که با استفاده از داده‌های تاریخی بازار و تحلیل‌های آماری سعی به پیش‌بینی بازار می‌کند، در سال‌های اخیر با گسترش هوش مصنوعی و کلان داده و به وجود آمدن قدرت پردازشی بالا به رویکرد های هوش مصنوعی و یادگیری ماشین برای پیش‌بینی بازار مالی توجه ویژه ایی شده است، مزیت الگوریتم های یادگیری تقویتی نسبت به دیگر روش‌ها امکان یادگیری آن‌ها در محیط های non-stationary است، با تغییر قوانین، شرایط اقتصادی، منتشر شدن اخبار و تمام عوامل مرتبط با بازار، رفتار قیمت ها نیز در طول زمان تغییر می‌کنند و به اصطلاح Regime shift ایجاد می‌شود، بسیاری از روش‌های یادگیری ماشین توانایی همراهی کردن با این تغییرات را ندارند و با تغییر رژیم بازار عمل‌کرد آن‌ها تنزل پیدا کرده و یا به طور کامل بلااستفاده می‌شوند، بنابرین با توجه به این موضوع الگوریتم های یادگیری تقویتی یکی از بهترین گزینه ها برای حل این نوع مسائل می‌باشند، برای پیاده‌سازی این پروژه از محیط gym-anytrading استفاده شد که به دلیل محدودیت‌های آن تغییراتی در آن ایجاد شد، داده استفاده شده نیز داده ۲ ماه گذشته‌ی دو رمز ارز بیتکوین و اتریوم با Timeframe یک دقیقه‌ای است.

فرمول بندی

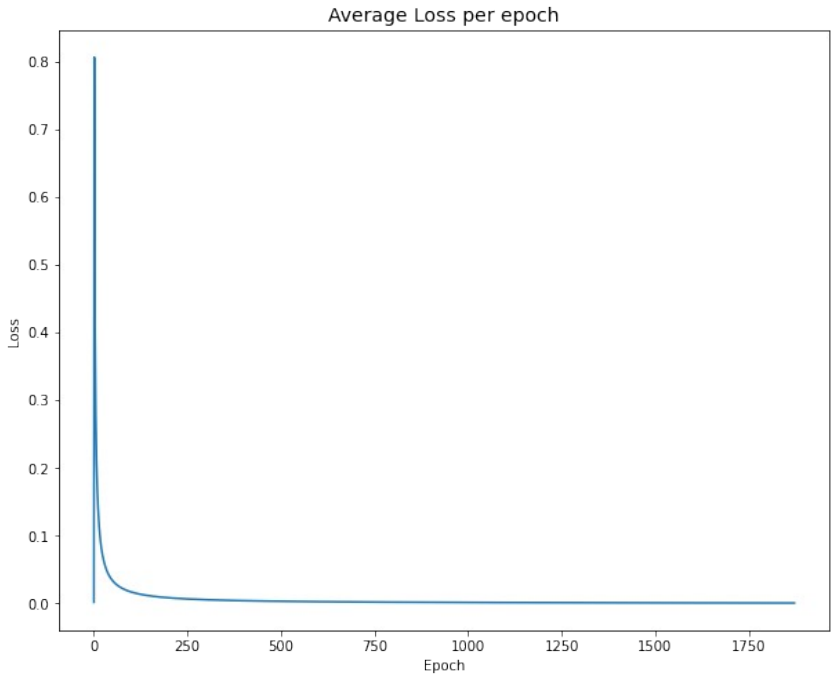
برای این حل مسأله میتوان به اشکال مختلف مسأله را مدل کرد، حدف مسأله ما آموزش عاملی است که بتواند با دریافت داده‌های قیمتی مربوط به دو رمز ارز بیتکوین و اتریوم صرفاً در بازار بیتکوین به معامله‌گری سودده بپردازد، دلیل استفاده از داده‌های بازار اتریوم همبستگی قیمتی بالای این دو رمز ارز با یکدیگر است، با بررسی مقاله های منتشر شده و همچنین محدودیت‌های محیط مورد استفاده در نهایت محیط به شکل یک مسأله حالت پیوسته و اکشن گسسته مدل شد، حالت محیط تابعی از مشاهدات عامل شامل قیمت های هر کندل (داده های هر یک دقیقه از بازار به شکل نمودار شمع ژاپنی)، حجم معامله شده در ۱ دقیقه، حجم خریداری شده، تعداد معامله انجام شده، به همراه دو اندیکاتور معروف MACD و RSI است که برای تحلیل تکنیکال در بازار مالی استفاده می‌شود در step اطلاعات دقیقه قبلی به عامل داده می‌شود، با استفاده از این اطلاعات حالت محیط به دست می‌آید، ۲ اکشن برای این محیط تعریف شده که شامل خرید و فروش سهم می‌باشد و پاداش آن‌ها برابر با تفاوت قیمت در زمان خرید و فروش (میزان سود ما از معامله) است که این مقدار از قیمتClose نرمال شده به دست می‌آید.

متد

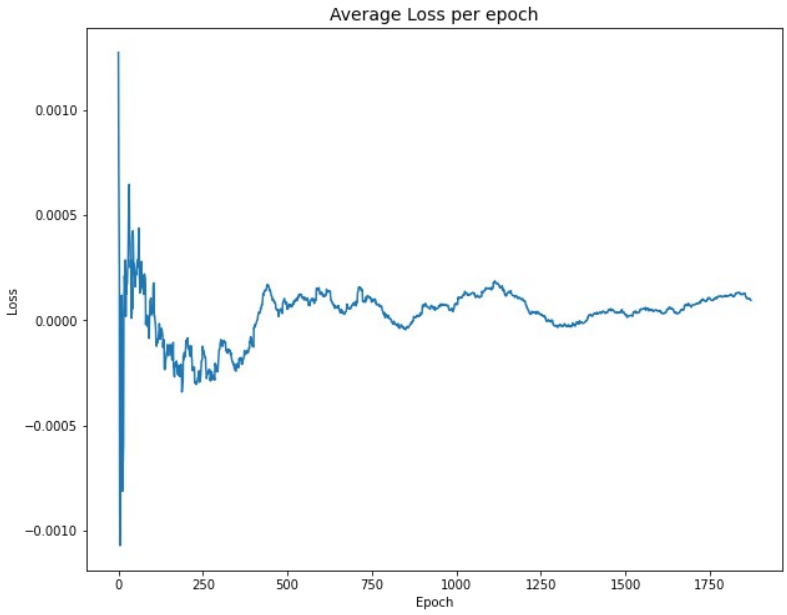
برای حل این مسأله تا به حال از روش‌های متفاوتی استفاده شده است که مرسوم ترین آن‌ها DQN، A2C است، اما در این پروژه تصمیم گرفتیم از روش Deep Recurrent Q network ها استفاده کنیم، دلیل استفاده از این الگوریتم این بود که مشاهدات انجام شده در محیط به شدت با همدیگر همبستگی زمانی دارند بنابرین تنها با در نظر گرفتن یک مشاهده عامل نمیتواند پیشبینی درستی از محیط انجام دهد، راهکاری که اکثر مقالات برای حل این مشکل در نظر گرفته‌اند ایجاد یک پنجره زمانی برای ارسال مشاهدات با یکدیگر است به این شکل که تعدادی از مشاهدات قبلی که در محیط انجام شده به شکل یک‌جا به شبکه داده می‌شود تا همبستگی زمانی در محیط لحاظ شود اما این راه حل در بازار های مالی کارا نیست به این دلیل که در بازار های مالی بسته به شرایط بازار میزان متفاوتی از نوسان قیمتی داریم و به اصطلاح Volatility قیمت در زمان های مختلف متفاوت است و یک پنجره زمانی ثابت باعث می‌شود که در زمان هایی که بازار نوسان کمی دارد تعداد زیادی از مشاهده به محیط بدهیم و در زمان نوسان بالا مشاهدات ما کم باشد، بنابرین روش بهتر این است که عامل با استفاده از یک حافظه درونی خود همبستگی زمانی محیط را در نظر بگیرید. برای پیاده‌سازی این روش از ماژول های LSTM که معمولاً برای سری های زمانی در نظر گرفته می‌شود استفاده شد، شبکه DRQN شامل یک لایه خطی، یک لایه LSTM و یک لایه خطی نهایی است، همچنین از Experience Replay برای بهینه سازی فرآیند یادگیری استفاده کردیم و به دلیل اینکه مشاهده تصادفی در این محیط بی‌معنی است کل هر Trajectory را یک جا به عامل می‌دهیم.

نتایج

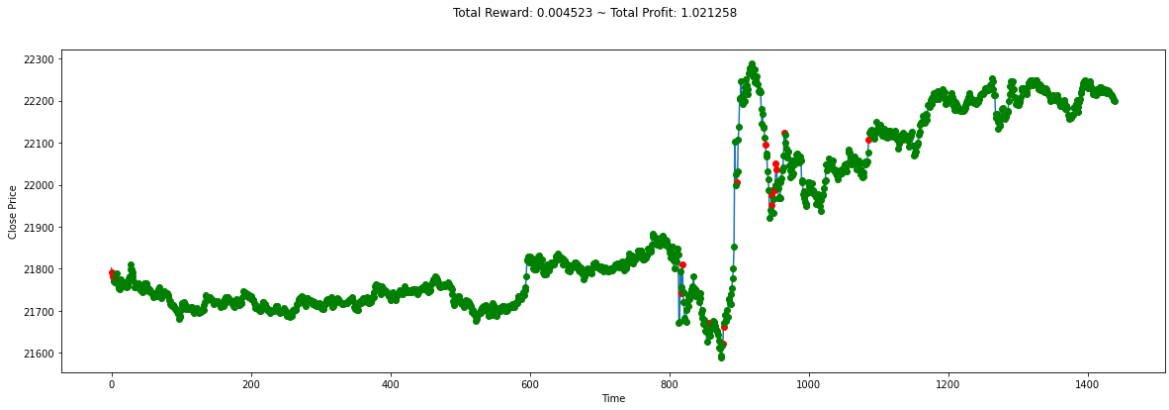
مدل برای ۲۰۰۰ ایپاک آموزش داده شد که در نهایت به نتایج زیر رسید:



شکل ۱: نمودار خطای مدل در هر ایپاک (نمودارها به شکل میانگین تجمعی کشیده شده است)



شکل ۲: نمودار میانگین پاداش در هر ایپاک



شکل ۳: عمل‌کرد مدل در ایپزود تست

با توجه به نتایج به نظر می‌رسد که خطای مدل به شکل مطلوبی کاهش پیدا کرده است و میزان پاداش آن نسبتاً مثبت است، همچنین به نظر می‌رسد که عامل تمایل بیشتری به معامله Long در محیط دارد.

خلاصه

الگوریتم های یادگیری تقویتی یکی از بهترین گزینه ها برای معامله‌گری الگوریتمی هستند که توانایی این را دارند با تغییرات بازار خود را همسو سازند، در این پروژه سعی شد که با پیاده‌سازی الگوریتم های متفاوت نسبت به کارهای انجام شده قبلی و همچنین بهبود محیط و استفاده از ویژگی‌های بهتر میزان عمل‌کرد عامل را بهبود بخشیم، اما این مدل نیز همچنان نیاز به بهبود داشته تا عمل‌کردی مناسب جهت پیاده‌سازی واقعی و معامله‌گری در شرایط واقعی داشته باشد، از جمله کارهایی که در ادامه می‌توان انجام داد بهبود تابع پاداش، بهبود شبکه DRQN و همچنین استفاده از ویژگی‌های مناسب‌تر جهت مدلسازی محیط اشاره کرد.

منابع

[۱] M. Hausknecht and P. Stone, “Deep Recurrent Q-learning for partially observable MDPs,” *arXiv [cs.LG]*, 2015.

[۲] E. Ponomarev, I. Oseledets, and A. Cichocki, “Using reinforcement learning in the algorithmic trading problem,” *arXiv [q-fin.TR]*, 2020.

[۳] T. Théate and D. Ernst, “An application of deep reinforcement learning to algorithmic trading,” *arXiv [q-fin.TR]*, 2020.

[۴] A. Saxena, *Deep-Recurrent-Q-Networks: Implementation of Deep Recurrent Q-Networks for Partially Observable environment setting in Tensorflow*.