پژوهش‌های نوین در تصمیم‌گیری
دوره ۷، شماره ۴، زمستان ۱۴۰۱، صص ۵۱-۸۰
نوع مقاله: پژوهشی

پیش‌بینی رفتار بورس اوراق بهادار با به‌کارگیری اندیکاتورهای تکنیکال، مبتنی بر رویکردهای یادگیری تقویتی عمیق و شبکه‌های کانولوشن

(مطالعه موردی: بورس اوراق بهادار ایران)

آنیتا هادی‌زاده[۱]، محمد جعفر تارخ [۲*]، مجید میرزایی قزآنی[۳]

۱. دانشجوی دکتری مهندسی فناوری اطلاعات، دانشکده مهندسی صنایع، دانشگاه صنعتی خواجه نصیر، تهران، ایران
۲. استاد، گروه مهندسی فناوری اطلاعات، دانشکده مهندسی صنایع، دانشگاه صنعتی خواجه نصیر، تهران، ایران
۳. استادیار گروه مهندسی مالی، دانشکده مهندسی صنایع، دانشگاه صنعتی خواجه نصیر، تهران، ایران

تاریخ دریافت: ۱۴۰۱/۰۳/۲۵ تاریخ پذیرش: ۱۴۰۱/۰۷/۲۷

چکیده

امروزه بورس در کشورهای مختلف نقش مهمی در اقتصاد کشورها ایفا می‌کند. فراوانی داده‌ها در بورس و لزوم پردازش سریع و صحیح داده‌ها و اتخاذ تصمیم مناسب استفاده از کامپیوترها را اجتناب‌ناپذیر نموده است. در این مقاله با استفاده از یادگیری تقویتی عمیق مدلی ارائه شده است تا وظایف یک معامله‌گر در بورس ایران را با توجه به سهم‌های نقد شونده مدل‌سازی کند. در مرحله اول، تاریخچه‌ی قیمت‌های سهام به همراه اندیکاتورهای مبتنی بر آن به‌عنوان ورودی به شبکه عصبی کانولوشن داده می‌شود. در مرحله بعد، به‌منظور محاسبه میزان تطبیق خروجی کانولوشن با خروجی مورد انتظار، از تابع هزینه‌ی مجموع مربعات خطا استفاده می‌شود که به‌نوبه خود، در فرایند بهینه‌سازی کمینه می‌شود. ازآنجاکه داده‌ها در بورس ایران محدود است، استفاده از مدل کانولوشن به‌جای جدول Q در مدل تقویتی عمیق از بیش برازرش مدل جلوگیری می‌کند. به‌منظور ارزیابی مدل،از داده‌های بورس اوراق بهادار تهران در بازه زمانی ۱۳۹۰ تا ۱۴۰۰ استفاده شده و بازده روش پیشنهادی با استراتژی خرید و نگهداری مقایسه شده است. نتایج آزمایش‌ها نشان می‌دهد، در مواردی سود حاصل از روش پیشنهادی و خرید نگهداری به ترتیب معادل ٪۲۱ و ٪۷- حاصل می‌شود.

کلیدواژه‌ها: یادگیری تقویتی عمیق، کانولوشن، بورس، اندیکاتور تکنیکال، تحلیل تکنیکال

---

* نویسنده مسئول مقاله E-mail: mjtarokh@kntu.ac.ir

۱- مقدمه

هدف از بورس، ایجاد بازاری منسجم، قانونمند، پیوسته و شفاف برای تجمیع سرمایه‌های خرد و کلان سرمایه‌گذاران به منظور تأمین مالی شرکت‌ها و شراکت در تقسیم سود آنها است. بورس بازاری پر نوسان است که به‌طور بالقوه می‌تواند سود و زیان‌های قابل توجهی را متوجه سرمایه‌گذاران نماید. افت و خیز این بازار در صعود و نزول رشد اقتصادی کشورها دارای اهمیت است[۱]. بنابراین توجه بسیاری از سرمایه‌گذاران را برای تحلیل بازار و دستیابی به سودهای چشم‌گیر معطوف به خود می‌کند. در گذر زمان، با ورود فناوری‌های نوین در پیش‌بینی و شناسایی رفتارهای بازار، ابزارهای مناسبی در اختیار تحلیل‌گران قرار گرفته است. روش‌های سنتی، برای پیش‌بینی بورس، بر مدل‌های خطی و رگرسیون متمرکز بوده است. در این روش‌ها حافظه جایگاه چندانی نداشت و امکان تعامل با داده‌های زیاد وجود نداشت. با ورود دانش یادگیری ماشین در مباحث مربوط به پیش‌بینی و تصمیم‌گیری، روش‌های SVM[۱] و درخت تصمیم توجه بسیاری از سرمایه‌گذاران را به خود جلب کرد. امروزه به لطف پیدایش کلان‌داده‌ها و تکنیک‌های یادگیری عمیق، امکان پیش‌بینی دقیق‌تری فراهم شده است. تکنیک یادگیری عمیق مبتنی بر ادراک قوی و استخراج خودکار ویژگی است. اما نقطه ضعف آن تغییرپذیری محیط معاملات و ضرورت دسترسی به داده‌های با مقیاس زیاد است. به‌منظور مواجهه با محیط‌های تغییرپذیر، مدل‌های مبتنی بر یادگیری تقویتی امکان تحلیل در محیط‌های تغییرپذیر را فراهم نموده است. به‌طوری‌که سیستم‌های مبتنی بر این روش قادر هستند خود را با تغییرات موجود در بورس تطبیق دهند. روش‌های یادگیری تقویتی بر اساس پاداش و جریمه است[۲]. بنابراین ادغام یادگیری عمیق با یادگیری تقویتی مکمل یکدیگر خواهد بود و رویکردی یکپارچه می‌تواند طرحی را برای ساخت یک سیستم تصمیم‌گیری پویا ارائه دهد. به‌طور سنتی، روش‌های بسیاری برای تحلیل و پیش‌بینی بازار سهام وجود دارد که یکی از این روش‌ها روش تحلیل تکنیکال است. در این روش با استفاده از فرمول‌های ریاضی و سابقه قیمت‌ها، اندیکاتورها به‌دست می‌آید که از آن می‌توان برای پیش‌بینی بورس استفاده کرد[۳]. در این مقاله، هدف این است که با استفاده از داده‌های تاریخچه قیمت‌ها و اندیکاتورهای مذکور، داده‌های مورد نیاز مدل عمیق پیشنهادی فراهم گردد. در نهایت، نتیجه به‌دست آمده با استراتژی خرید و نگهداری مقایسه می‌گردد تا بهینگی روش پیشنهادی نسبت به آن مقایسه گردد. در سال‌های اخیر بورس ایران، طرفداران بسیاری یافته که علاقمند به سرمایه‌گذاری و

---

[۱] Support Vector Machine

دستیابی به سود بیشتر هستند. اما به دلیل شرایط بورس ایران که در ادامه بررسی خواهد شد، بهتر است از سهم‌های نقد شونده استفاده شود تا روند خرید و فروش دچار اختلال نشود. پس هدف این است که در بازاری مانند ایران که شرایطی کمابیش متفاوت با سایر کشورها دارد، با استفاده از یادگیری تقویتی عمیق، پیش‌بینی‌های لازم برای خرید و یا فروش از سهم‌های نقد شونده انجام پذیرد. بنابراین، در این مطالعه، بورس اوراق بهادار ایران به‌عنوان بازار هدف انتخاب گردیده و نمادهای معاملاتی نقد شونده شامل "فولاد"، "فملی"، "شبندر"، "شپنا"، "وپاسار"، "وبملت"، "خودرو"، "بپاس"، "بورس"، "شپدیس "، "وامید"، "کگل"، "خساپا"، "شتران"، "وغدیر"، "خبهمن"، "برکت"، "فارس"، "وتجارت جم"، "فخوز"، "آریان"، "فرابورس"، "پارسان"، "تاپیکو"، "پارس"، "پاکشو"، "ومعادن"، "کچاد" به‌عنوان نمادهای منتخب در بازه زمانی ۱۳۹۰ تا ۱۴۰۰ مورد بررسی می‌باشد.

۲- پیشینه پژوهش

در این بخش مروری بر منابع در دو بخش انجام می‌شود. بخش اول متمرکز بر بررسی منابع بورس ایران می‌باشد. قابل ذکر است پژوهش‌های انجام شده در ایران محدود به حوزه یادگیری عمیق و دیدگاه تکنیکال می‌باشد. در بخش دوم متمرکز بر منابع بورس بین‌الملل با شیوه یادگیری تقویتی عمیق می‌باشد.

۲-۱- بررسی بورس ایران

این پژوهش مبتنی بر بورس ایران است. این بازار ویژگی‌هایی دارد که آن را از سایر بازارهای مالی متمایز می‌نماید. در ادامه ویژگی‌های بورس ایران بر اساس شواهد تجربی بیان می‌شود:

۱. بازار یک‌طرفه- بورس ایران از جمله معدود کشورهایی است که یک‌طرفه است و فعالان این بازار در زمان رشد قیمت‌ها امکان کسب بازده بیشتری از آن بازار دارند و در ریزش قیمت‌ها می‌توانند موقعیت سودآور را شناسایی کنند تا متحمل زیان در اصل سرمایه نشوند.

۲. زمان‌بری دوران رکود بورس- در بازارهای یک‌طرفه مانند بورس ایران، به دلیل فضای تورمی و در نتیجه آن مواجهه فروشِ دارایی با زیان، سهام‌داران ناگزیر به نگهداری سهام خود هستند تا سرانجام با افزایش قیمت سهام از زیان خارج شوند. از طرفی، مادامی‌که بازار نزولی می‌شود، امکان ایجاد معاملات سودآور وجود ندارد و بازار دچار رکود می‌شود[۴].

۳. گذرا بودن سرمایه‌گذاری در بورس–یکی از اهداف بنیادین بورس ترغیب مشارکت همگانی در تولید است. در بورس یک‌طرفه به‌ناچار بسیاری در امواج صعودی بورس وارد این بازار می‌شوند و در فاز رکودی، از بورس خارج شده و به سایر بازارها مانند مسکن مهاجرت می‌کنند[۵].

٤. ریسک نقدشوندگی– سرمایه‌گذاران در دوران رکود بورس، به دلیل عدم وجود خریدار، به‌سختی می‌توانند سهام خود را نقد و تغییر رویکرد سرمایه‌گذاری ایجاد نمایند؛ بنابراین، سرمایه‌گذاران با پول‌های کلان از ورود به بورس بر حذر شوند. از سویی دیگر، به هنگام فروش مقدار زیاد سهام به دلیل عدم وجود خریدار به میزان کافی، در برخی از سهم‌ها (شرکت‌های کوچک) سهام‌داران با مشکل مواجه می‌شوند[۶].

۵. حباب معکوس در دوران کسادی– در دوران نزولی بورس، به دلیل بیم سهام‌داران از بی‌ارزش شدن دارایی خود در اثر حرکت گله‌ای و رقابت در فروش با نرخ پایین‌تر از حد نرمال، ریزش غیرمنطقی ایجاد می‌شود[۷, ۸].

٦. حباب در دوران رونق– در دوران رونق، سرمایه‌گذاری در بورس همواره با بازدهی خوبی همراه است. رشد نقدینگی در سطح جامعه و گسیل آن به سمت بورس، با توجه به عمق اندک این بازار، منجر به رشدهای چند برابری در قیمت برخی از سهم‌های بازار می‌شود. بنابراین، تحلیل نقشی کم‌رنگ‌تر به خود گرفته و سهم‌های گوناگون صرفاً به دلیل ورود نقدینگی و نه ارزش ذاتی رشد قیمتی را تجربه می‌کنند[۸].

۷. زمان‌بندی نشر خبر– برخی بازارها، همچون فارکس، ساعت مشخصی برای پخش اخبار مربوطه دارد. در این زمان‌ها، فعالان آن بازار دست از فعالیت کشیده و متمرکز خبر می‌شوند تا با تحلیل مناسب فعالیت خود را از سر گیرند. اما بورس ایران فاقد این ویژگی است و در هر زمانی از شبانه روز اخبار تحول‌آفرین امکان پخش دارد که بسیاری از آنها در حد شایعه هستند. بنابراین عدم تقارن اطلاعاتی میان ذینفعان شکل می‌گیرد[۹].

۸. ماهیت متزلزل– ماهیت بازارهای سهام مبتنی بر سرمایه‌گذاری است واساسا در دراز مدت ماهیتی صعودی دارد. اما در بورس ایران با وجود دوره‌های طولانی رکود، امکان از دست رفتن سرمایه وجود دارد[۷].

۹. ارزش ارز– ارزش لحظه‌ای دلار و طلا می‌تواند بر روند تغییرات بورس ایران تغییرات هیجانی ایجاد کند. از دیگر سو، در بازار جهانی، ارزش طلا و ارز اساساً برخلاف یکدیگر تغییر می‌کند. اما در بورس ایران افزایش یکی، افزایش دیگری را به دنبال دارد[۷].

۱۰.صف- در بورس ایران به دلیل حجم معاملات اندک، صف خرید و صف فروش ایجاد می‌شود. بدین معنا، پس از تحلیل و دستیابی به نتیجه، امکان انجام خرید و فروش ممکن است وجود نداشته باشد[۵].

۱۱.محدودیت نوسان و سودآوری با استفاده از شرایط ناپایدار- بورس ایران محدودیت در نوسان دارد و بهای سهم نمی‌تواند بیش از ۵٪ رشد و یا ریزش نماید. به‌این‌ترتیب، در شرایط یکسانی قیمت باز و بسته در بازه زمانی مشخص که در سایر بازارها ناکارآمد می‌باشد، در ایران می‌تواند موقعیت مناسبی برای خرید یا فروش باشد. از طرفی، در بازار ایران افت و خیزهای جعلی بسیار است که در اثر هماهنگی بین غول‌های بازار به معنی ورود یا خروج کوتاه مدت پول‌های درشت، به دلیل اخبار هیجانی ایجاد می‌شود[۷].

۱۲.خریدار و فروشنده حقیقی و حقوقی- در بسیاری مواقع خریدار و فروشنده حقیقی نیستند. بلکه تبانی‌هایی است که میان صاحبان سهام برای ایجاد تغییر در قیمت‌ها صورت می‌پذیرد که منجر به روند قیمت غیرواقعی می‌شود.

در بررسی بورس ایران و روش‌های یادگیری، پژوهش [۱۰] مدلی ارائه کرده است که پتانسیل آتی سهام با در نظر گرفتن شاخصهای تحلیل تکنیکال با استفاده از شبکه فازی و الگوریتم ژنتیک در بازه زمانی پنج ساله پیش‌بینی و بر مبنای عواملی چون میانگین، واریانس و چولگی انجام داده است. پژوهش [۱۱] بر بررسی بازار سهام تهران متمرکز است که داده‌ها در دو گروه بررسی شده‌اند. در گروه اول داده‌های مربوط به سال ۲۰۰۹ تا ۲۰۱۱ با هدف کمینگی ریسک و بیشینگی بازدهی و دستیابی به مدلی برای مدیریت سبد خرید طراحی شده و گروه دوم داده‌های مربوط به سال ۲۰۱۲ تا ۲۰۱۴ با هدف مدیریت بهینه پورتفولیو بر اساس پیش‌بینی‌های سه ماهه طراحی شده است. مدل‌های ICA[۱]، ARMA[۲] و GARCH[۳] مورد نظر بوده و برای تحلیل از میانگین، قیمت بیشینه، قیمت کمینه، چولگی، کشیدگی[۴]، آزمون نرمال بودن جارک برا[۵] استفاده شده است. پژوهش [۱۲] بر بورس ایران از ۱۹۹۹ تا ۲۰۱۶ متمرکز است. برای پیش‌بینی از مدل‌های زنجیره مارکو، E-GARCH[۶]، ARIMA-GARCH، ARIMA-GARCH[۷] استفاده شده است. در این مقاله رویکرد فازی از دیگر روش‌های GARCH پاسخ بهتری داده است. برای تحلیل داده از ارزش بازار سهام و روند کاهش و یا

---

۱ Imperialist Competitive Algorithm
۲ Autoregressive Moving-Average model
۳ Generalized Autoregressive Conditional Heteroscedasticity
۴ Kourtosis
۵ JARQUE-Bera
۶ Exponential GARCH
۷ Autoregressive Integrated Moving Average (ARIMA)

افزایش شاخص TEDPIX[1] استفاده شده است. برای ارزیابی نیز معیار متوسط مربعات خطا مورد نظر است. پژوهش [۱۳] و [۱۴] بر پیش‌بینی بورس ایران بر گروه‌های نفتی، مواد معدنی غیرفلزی و فلزات اساسی در بورس اوراق بهادار تهران متمرکز است. داده‌ها به‌صورت ده ساله و بر اساس پیش‌بینی‌های ۲، ۵، ۱۰، ۱۵، ۲۰ و ۳۰ روزه با استفاده از الگوریتم‌های درخت تصمیم، جنگل تصادفی، تقویت تطبیقی[2]، تقویت گرادیان[3] و تقویت گرادیان فوق‌العاده[4] و شبکه‌های عصبی مصنوعی[5]، شبکه عصبی بازگشتی (RNN) و LSTM مورد تحلیل قرار گرفته است. ده شاخص فنی شامل میانگین متحرک ساده و وزن‌دهی شده، ممنتوم، استوکستیک، استوکستیک k%، اندیس قدرت نسبی، نوسان‌گر A/D، اندیس ویلیام و اندیس کانال کالا، میانگین متحرک ساده n روزه[6]، میانگین متحرک وزنی ۱۴ روزه[7]، ممنتوم[8]، استوکستیک K%، استوکستیک D%، RSI[9]، ویلیام R%، اسیلاتور A/D[10] و CCI[11] به‌عنوان ورودی هر یک از مدل‌های پیش‌بینی در نظر گرفته شده است. بر اساس نتایج حاصل از آزمایش‌ها، در بین این روش‌ها، LSTM[12] بهترین نتیجه و درخت تصمیم پرخطاترین روش معرفی شده است. همچنین، در مدل‌های مبتنی بر درخت، رقابت تنگاتنگی بین تقویت تطبیقی، تقویت گرادیان و تقویت گرادیان فوق‌العاده وجود دارد. معمولاً مقادیر خطاها با افزایش دوره زمانی مورد پیش‌بینی بیشتر می‌شود. با بررسی نتایج مشخص شده است که پیش‌بینی گروه فلزات نتایج مناسب‌تری دارد.

مطالعات انجام شده نشان می‌دهد، پژوهش در بورس ایران نیازمند بررسی‌های عمیق‌تر و دقیق‌تری با توجه به فنون پیش‌بینی روز دنیا در حوزه هوش مصنوعی و بخصوص یادگیری عمیق است.

۲-۲- بررسی بورس بین‌الملل مبتنی بر یادگیری تقویتی عمیق

در این بخش مطابق با جدول ۱ مروری بر منابع مرتبط با روش یادگیری تقویتی عمیق انجام می‌شود.

[1] TEHRAN PRICE INDEX
[2] Adaboost
[3] Gradient boosting
[4] XGBoost
[5] Artifitial Neural Network (ANN)
[6] Simple n-day moving average
[7] Weighted ۱۴-day moving average
[8] Momentum
[9] Relative Strength Index
[10] Accumulation/Distribution
[11] Commodity channel index
[12] Long short-term memory

جدول ۱. منابع مورد مطالعه در حوزه یادگیری تقویتی عمیق و بورس

مرجع	روش مورد استفاده	داده‌ها و شاخص‌های استفاده شده	نتایج به‌دست آمده	نکات قابل توجه	اثربخشی در پژوهش حاضر
[۱۵]	RL+DL	عدم انتخاب شاخص تکنیکالی و انتخاب ویژگی بصورت خودکار	اثربخشی سیستم یادگیری در خلاصه‌سازی هم‌زمان شرایط بازار و یادگیری اقدام بهینه	یادگیری فازی در مدل DL برای کاهش عدم قطعیت	انتخاب همه شاخص‌های تکنیکال موجب کند شدن در تحلیل و افزونگی در محاسبات می‌شود.
[۱۶]	RL	مقایسه با شاخص داوجونز و استراتژی تخصیص پورتفولیوی با واریانس کمینه در ۳۰ نماد بزرگ بازار مبتنی بر قیمت روزانه	عملکرد مناسب‌تررویکرد یادگیری تقویتی عمیق از نظر نسبت شارپ[۱] و بازده تجمیعی	روش مناسب برای در متعادل کردن ریسک و بازده	با توجه به اثربخشی نسبت شارپ، این نسبت برای پژوهش حاضر مورد استفاده قرار می‌گیرد.
[۱۷]	RL مطالعه ۵۰ نشریه و دسته بندی به سه رویکرد منتقد، بازیگر و منتقد-بازیگر	انتخاب بهترین کارایی و نسبت شارپ توجه به بهبود کارایی سیستم در مواجهه با داده‌های با حجم زیاد با استفاده از تکنیک بارگذاری هنگام نیاز[۲]	در بررسی وضعیت تابع پاداش و فضای عمل عامل، رویکرد منتقد مناسب انتخاب میان N دارایی قابل معامله و یا تصمیمات معاملاتی بسیار خرد، رویکرد بازیگر مناسب برای فضای کنش مداوم و همگرایی سریع‌تر و شفافیت بیشتر، رویکردمنتقد-بازیگر (ترکیب دو روش)	توجه به محدودیت‌های مهم مانند هزینه‌های مبادله، نقدینگی بازار و درجه ریسک‌گریزی سرمایه‌گذار	تاکید بر RL به دلیل امکان ترکیب یکپارچه پیش‌بینی و ساخت نمونه و در نتیجه دستیابی به هماهنگی بیشتر عملکرد یادگیری ماشین با اهداف سرمایه‌گذار
[۱۸]	مقایسه مدل‌های سنتی با سه مدل یادگیری	مقایسه اطلاعات ۳۰ سهام از داده‌های داوجونزی و استفاده از معیارهای دقت،	عملکرد مناسب‌ترDQN در تصمیم‌سازی	اثبات قابلیت اطمینان و در دسترس بودن این مدل	مدل‌های یادگیری تقویتی عمیق بیشترین شباهت را به یادگیری

[۱] Sharp Ratio

[۲] Load-On-Demand Technique

مرجع	روش مورد استفاده	داده‌ها و شاخص‌های استفاده شده	نتایج به‌دست آمده	نکات قابل توجه	اثربخشی در پژوهش حاضر
	تقویتی عمیق شامل DQN[1]، DDQN[2] و DDDQN[3]	کارایی[4]، بهینه‌سازی[5]، محبوبیت[6] برای مقایسه		با داده‌های تجربی	انسانی دارند. اما سرعت همگرایی کُند از جمله چالش‌هایی که است که این روش با آن مواجه است.
[۱۹]	پیش‌بینی شاخص‌های S&P۵۰۰-S، S&P۵۰۰-L در بلند مدت، کوتاه مدت با استفاده از Q-Learning	عوامل قیمتی باز، بسته، پایین، بالا و مقایسه با استراتژی خرید و نگهداری ارزیابی نتیجه با استفاده از دقت و نسبت سورتینو[7]، حداکثر ریزش[8]، پوشش[9] و منحنی حقوق صاحبان سهام[10]	اکثر رویکردهای موفق به‌صورت نظارت شده عمل می‌کنند	تنها استفاده از داده‌های قیمتی پردازش‌شده	استفاده از عوامل قیمت در کنار سایر شاخص‌ها اثرگذار است.
[۲۰]	RL	قیمت باز، بسته، بالا و پایین و نیز حجم معاملات با شاخص‌های میانگین صنعتی داوجونز[11]، S&P۵۰۰[12]، راسل ۲۰۰۰، نزدک[13]، شاخص بورس شانگهای، شانگهای-	کشف وابستگی‌های پنهان و پویایی‌های نهفته در داده‌های سهام		استفاده از حجم در بهبود نتیجه اثرگذار است.

[1] Deep Q-Network
[2] Dual Deep Q-Network
[3] Duelling Dual Deep Q-Network
[4] Performance
[5] Optimization
[6] Popularity
[7] Sortino ratio
[8] Maximum drawdown
[9] Coverage
[10] The equity curve
[11] Dow Jones Industrial Average (DJI)
[12] Standard and Poor's ۵۰۰ Index
[13] NASDAQ

مرجع	روش مورد استفاده	داده‌ها و شاخص‌های استفاده شده	نتایج به‌دست آمده	نکات قابل توجه	اثربخشی در پژوهش حاضر
		شنژن و شاخص کیفیت و رشد CSI ۵۰۰ و انتخاب تصادفی شاخص‌های بازار سهام و ارزیابی با استفاده از معیار Accuracy			
[۲۱]	استراتژی معاملاتی برمبنای TDQN[۱]	شاخص عملکرد نسبت شارپ در طیف گسترده‌ای از بازارهای سهام	استراتژی DRL می‌تواند روندهای اصلی را تشخیص دهد و از آنها بهره‌مند شود، اما در تغییرات رفتاری بازار حین افزایش نوسانات مردد است.	آموزش عامل یادگیری تقویتی به‌طور کامل بر اساس مجموعه محدودی از داده‌های تاریخی بازار سهام	عامل معاملاتی از روند بازار عقب است، و الگوریتم TDQN بیشتر واکنش‌پذیر است تا فعال. اما RL می‌تواند موقعیت معاملاتی خود را پیش از وارونگی روند با توجه به افزایش ناگهانی نوسانات تطبیق دهد.
[۲۲]	LSTM و RL و مقایسه با مدلهای سنتی	استفاده از مجموعه داده‌های شش ماهه، شامل ۱۰ سهام، از مجموعه داده‌های شش ماهه، شامل ۱۰ سهام از بورس اوراق بهادار شانگهای در چین	عملکرد بهتر مدل پیشنهادی حداقل ۱۰.۶۳ درصد بهتر از مدل‌های	LSTM برای استخراج الگوهای زمانی و یادگیری روابط حالت-عمل از طریق RL	استفاده از مدل یادگیری تقویتی نتایج بهتری نسبت به مدل‌های سنتی به همراه دارد
[۲۳]	مقاله مروری	تاکید بر DRL[۲] بعنوان مدل مطلوب و معرفی طیف وسیعی از معیارهای ارزیابی شامل شامل دقت (بیشترین فراوانی در طبقه‌بندی			بیشترین پژوهش‌ها مبتنی بر سوابق قیمت

۱ Trading Deep Q-Network algorithm

۲ Deep Reinforcement Learning

ردیف	روش مورد استفاده	داده‌ها و شاخص‌های استفاده شده	نتایج به‌دست آمده	نکات قابل توجه	اثربخشی در پژوهش حاضر
		داده‌ها)، فراخوانی[1]، صحت[2]، حساسیت[3]، ویژگی[4]، F1[5]، MAFS[6]، MCC[7]، AUC[8]، ضریب U تی تیل[9]، نسبت ضربه[10]، ARV[11]، میانگین خطای مطلق[12]، خطای میانگین مربعات نرمال[13] و ریشه[14] و نسبی[15]، میانگین درصد مطلق خطا [16]، خطای نسبی ریشه میانگین مربعات[17]، اطلاعات دوجانبه[18]، نسبت شارپ[19]، تست دیبولد-ماریانو[20]، تست کروسکال والیس[21]، حداکثر برداشت[22]، نوسانات سالانه			سهام هستند و در رتبه دوم در حدود ۲۵٪ از پژوهش‌ها متمرکز بر سوابق قیمت‌ها همراه با تحلیل تکنیکال می‌باشند.

از مطالعات انجام شده این نتیجه حاصل می‌شود که علیرغم منابع محدود درباره DRL، این روش برای پیش‌بینی بازار سهام پیشنهاد می‌گردد. از طرفی به‌کارگیری داده‌های قیمت و اندیکاتورها و نسبت شارپ می‌تواند در دستیابی به نتیاج دقیق‌تر موثر واقع گردد.

۳- مبانی نظری

در این بخش، مبانی نظر مورد نیاز مورد توجه قرار گرفته است. به‌عبارت دیگر، مفاهیمی چون نقد شوندگی در بورس ایران، فاکتورهای موثر در پیش‌بینی روند، اندیکاتورهای مورد استفاده، شاخص‌های کل و هم‌وزن بورس ایران و چگونگی استفاده از معیارها تشریح شده است.

۳-۱- سهام نقد شونده

پیش از سرمایه‌گذاری در سهام، اولین نکته در انتخاب سهم این است که سهمی انتخاب شود که در صورت نیاز به سرمایه، سریع‌تر بتوان به پول نقد تبدیل نمود و در صف خرید یا فروش آن باقی نماند و به عبارتی، ریسک نقدشوندگی کمتری داشته باشد. اما امکان نقدشوندگی همه سهم‌ها به یک میزان نیست. عوامل بسیاری بر نقدشوندگی سهام از جمله حجم معاملات بالا (باتوجه به میزان عرضه و تقاضا)، حجم مبنا (تعداد لازم از یک نوع اوراق بهادار برای معامله

[1] Recall
[2] Precision
[3] Sensitivity
[4] Specificity
[5] Score(F1)
[6] Macro-average F-score
[7] Matthews Correlation Coefficient
[8] Average AUC Score
[9] Theil's U Coefficient
[10] Hit Ratio
[11] Average Relative Variance
[12] Mean Absolute Error
[13] Normalized MSE
[14] Root Mean Absolute Error
[15] Relative RMSE
[16] Mean Absolute Percentage Error
[17] Root Mean Squared Relative Error (RMSRE)
[18] Mutual Information (MUL)
[19] Sharpe Ratio،
[20] Diebold-Mariano Test
[21] Kruskal-Wallis Test
[22] Maximum Drawdown

روزانه با هدف مشخص شدن قیمت روز آینده) و ارزش معاملات اثرگذار است. معیار عدم نقدشوندگی امیهود[۱] رایج‌ترین شاخص مورد استفاده برای درک سهم‌های عدم نقدشونده در ادبیات مالی است. بر این اساس در این پژوهش ۲۹ سهام نقد شونده مورد بررسی قرار می‌گیرد که با توجه به رابطه (۱) محاسبه می‌شود که در این رابطه T تعداد روز، V ارزش کل معاملات برای هر نماد در روز t و r_t بازگشت سرمایه در روز t می‌باشد[۲۴]:

$$ILLIQ = \frac{۱}{N}\sum_{t=۱}^{T}\frac{|r_t|}{\$V_t} \qquad (۱)$$

۳-۲- آلفا و اندیکاتورهای مورد استفاده

مهندسی ویژگی‌ها یکی از عناصر مهم برای پیش‌بینی موفقیت‌آمیز روند تغییرات بازار بورس است. پرسش مطرح این است که کدام ویژگی روند پیش‌بینی را به نتیجه سودمندتری منجر می‌کند. از این رو موضوع استخراج فاکتورهای موثر در پیش‌بینی روند اهمیت می‌یابد. آلفا مقدار بازده مازاد را نسبت به شاخص بازار نشان می‌دهد. به طور کلی آلفا می‌تواند عددی مثبت، منفی یا صفر باشد. هرچه مقدار آلفا بیشتر باشد، یعنی صندوق عملکرد بهتری داشته است. فاکتورهای آلفا به داده‌هایی گفته می شود که هدف آنها پیش‌بینی نوسانات قیمت است. هر فاکتور می‌تواند از ترکیب چند ورودی حاصل گردد. تصمیمات تجاری می‌تواند بر اساس این فاکتورها حاصل گردد. فاکتورهای آلفا در نتیجه‌ی تبدیل داده‌های قیمتی، بنیادی یا جانبی از طریق محاسبات ساده مانند تغییرات مطلق یا نسبی یک متغیر در طول زمان، یا محاسبه در یک پنجره زمانی، مانند نمودار میانگین متحرک، ساده یا نمایی، یا تجزیه و تحلیل قیمت/حجم حاصل می‌شود[۲۵]. تحقیقات در مورد فاکتورهای آلفا که بازده بازار را پیش‌بینی می‌کنند، منجر به خلق مدل‌های چند عاملی شده‌اند که ورودی‌های مفیدی برای مدل‌های یادگیری ماشین هستند. در ادامه اندیکاتورهای مورد استفاده در این مقاله تشریح شده است.

۳-۳- ویژگی حرکتی RSI

یکی از ویژگی‌ها برای شناسایی روند قیمت شاخص قدرت نسبی[۲] است[۲۶]. RSI میزان تغییرات مثبت و منفی قیمت سهام را مقایسه می‌کند تا اشباع خرید یا فروش را تشخیص دهد. میانگین تغییر قیمت را برای تعداد معینی از روزهای معاملاتی پیشین (اغلب ۱۴) به ترتیب با افزایش Δ_P^{UP} و کاهش قیمت Δ_P^{down} به‌صورت رابطه (۲) محاسبه می‌کند. این منحنی در بازه

---

[۱] Amihud

[۲] Relative Strength Index (RSI)

۰ تا ۱۰۰ نوسان می‌کند. اگر RSI از منطقه اشباع فروش خارج شده و افزایش یابد، زمان مناسبی برای خرید و اگر RSI از اشباع خرید سهم خارج شده و کاهش یابد، می‌تواند زمان مناسبی برای فروش سهم باشد[۲۵].

$$RSI=100-\frac{100}{1+\frac{\Delta_P^{UP}}{\Delta_P^{down}}} \quad (2)$$

۳-۴- اندیکاتور ATR

اندیکاتور ATR[1] نوسانات بازار را نشان می‌دهد. هدف آن پیش‌بینی تغییرات روند در بازه زمانی مشخص است. هر چه مقدار ATR بیشتر باشد، احتمال تغییر روند افزایش می‌یابد. اندیکاتورATR جهت روند را نشان نمی‌دهد. بلکه نشان دهنده میزان نوسان بازار است. بنابراین لازم است در کنار سایر اندیکاتورها به‌کار گرفته شود. مطابق با رابطه (۳) ATR در دوره زمانی N محاسبه شده و حداکثر قدر مطلق نوسان را به‌صورت مقدار مطلق بزرگ‌ترین محدوده معاملاتی اخیر اندازه‌گیری می‌کند.

$$\mathrm{TR}_i = \max\left[(\mathrm{High}-\mathrm{Low}).\left|\mathrm{High}-\mathrm{Close}_{\mathrm{prev}}\right|.\left|\mathrm{Low}-\mathrm{Close}_{\mathrm{prev}}\right|\right] \quad (3)$$

محاسبه ATR شامل سه عامل مقدار بالای قیمتی روزانه منهای پایین‌ترین حد قیمت روزانه، مقدار بالای قیمتی روزانه منهای مقدار قیمت بسته شدن روز قبلی و - مقدار بسته شدن روز قبلی منهای پایین‌ترین حد قیمت روزانه می‌باشد. سپس، مطابق با رابطه (۴) برای به دست آوردن اندیکاتور ATR برای دوره مورد نظر کافی است میزان محدوده واقعی بر تعداد دوره زمانی (n) تقسیم َشود[۲۵].

$$ATR=\frac{1}{n}*\sum_i^n TR_i \quad (4)$$

۳-۵- میانگین متحرک[2]

از میان اندیکاتورهای بسیار میانگین متحرک، در این مقاله از دو رابطه میانگین متحرک ساده[3] و نمایی[4] استفاده شده است. مطابق با رابطه (۵)، در سری قیمت P_t با پنجره‌ای به طول N، میانگین متحرک ساده در قیمت‌ها استفاده شده و در زمان t، به‌گونه‌ای محاسبه می‌شود که هر نقطه داده در پنجره به میزان مساوی وزن دارد. میانگین متحرک نمایی نیز در حجم استفاده

[1] Average True Range (ATR)
[2] Moving average
[3] Simple moving average (SMA)
[4] Exponential moving average

شده و در سری قیمت P_t با پنجره‌ای به طول N، در زمان t، به‌صورت بازگشتی و میانگین وزنی قیمت کنونی و آخرین قیمت پیشین به‌گونه‌ای محاسبه می‌شود که هر نقطه داده در پنجره به میزان مساوی وزن دارد[۲۵]:

$$= \frac{۱}{N}\sum_{i=۱}^{N} P_{t-N+۱} SMA = \frac{P_{t-N+۱}+P_{t-N+۲}+P_{t-N+۳}+P_t}{N} \quad (۵)$$

$$, \alpha = \frac{۲}{N+۱} EMA(N)_t = \alpha P_t + (۱-\alpha) EMA(N)_{t-۱}$$

۳-۶- اندیکاتور مکدی

اندیکاتور مکدی[۱] «واگرایی» و «همگرایی» میانگین متحرک قیمت را نمایش می‌دهد که شامل دو میانگین متحرک نمایی به نام‌های " خط مکدی" و "خط سیگنال" می‌باشد. در محاسبه میانگین مکدی، به قیمت‌های نزدیک‌تر به روزهای آخر معاملاتی وزن بیشتری داده شده است. خط مکدی ویژگی تند و پرنوسانی دارد. ولی خط سیگنال بسیار کند و کم تلاطم است که دلیل این تمایز، متفاوت بودن میانگین‌گیری در خط مکدی و خط سیگنال می‌باشد. جزء سوم مکدی هیستوگرام است که در قالب خطوطی عمودی اختلاف خط مکدی و خط سیگنال را نمایش می‌دهد. هنگامی‌که این دو منحنی با یکدیگر برخورد می‌کنند، سیگنالی برای خرید یا فروش آن دارایی می‌باشد. به این صورت که اگر منحنی مکدی ، منحنی سیگنال را از پایین به بالا قطع کند، زمان مناسبی برای خرید و اگر از بالا به پایین قطع کند، زمان مناسبی برای فروش آن دارایی می‌تواند باشد. خط مکدی تفاضل میانگین نمایی ۱۲ دوره‌ای قیمت از میانگین نمایی ۲۶ دوره‌ای قیمت می‌باشد و خط سیگنال، میانگین نمایی ۹ روزه خط MACD است.

۳-۷- استوکستیک

اندیکاتور استوکستیک[۲] رابطه بین قیمت بسته سهم و بازه قیمتی آن را در دوره زمانی مشخص می‌سنجد. رفتار بازار که نمایش تدریجی دارد، اغلب در تحلیل موج قیمتی قبلی هر نمودار کاملاً قابل مشاهده است. اندیکاتور استوکستیک بر اساس تنظیمات دوره‌ای پیش‌فرض ۵ و ۱۴ تدوین شده است. خط K%، درصدی از تغییرات می‌باشد که نسبت آن بین سطوح ۰ تا ۱۰۰ نوسان می‌کند. این خط اساساً یک مقیاس جهت بررسی اندازه نیروی حرکتی در نوسانگر می‌باشد. محاسبه K% از طریق روابط (۶)، (۷) و (۸) قابل محاسبه است[۲۵]:

$$K\% = \frac{(\text{Current Close} - \text{Lowest Low})}{(\text{Highest High} - \text{Lowest Low})} * ۱۰۰ \quad (۶)$$

---

[۱] Moving Average Convergence Divergence

[۲] Stochastic

$$K^{Fast}(T_K) = \frac{P_T - P_{T_K}^L}{P_{T_K}^H - P_{T_K}^L} * ۱۰۰ \quad (۷)$$

$$K^{Slow}(T_{SlowK}) = MA\big((T_{SlowK})\big)[K^{Fast}] \quad (۸)$$

۳-۷-۱- اندیکاتور UO

اندیکاتور UO[۱] رفتار قیمتی کوتاه، متوسط و بلندمدت را با ابزاری تحلیلی ترکیب می‌کند که معیاری برای برآورد فشار و قدرت خرید تعبیر می‌شود [۲۵]. در ابتدا فشار خرید، سپس مقدار True Range به‌صورت روابط (۹) و (۱۰) محاسبه می‌گردد:

$$BP_t = P_t^{Close} - min(P_{t-۱}^{Close}.P_t^{Low}) \quad (۹)$$

$$TR_t = max\big(P_{t-۱}^{Close}.P_t^{High}\big) - min(P_{t-۱}^{Close}.P_t^{Low}) \quad (۱۰)$$

فشار خرید کل در هفت روز گذشته به‌صورت کسری از محدوده کل در همان دوره بیان می‌شود. منظور از BP_t فشار کل امروز است و به همین ترتیب فشار کل روزهای گذشته مطابق با رابطه (۱۱) محاسبه می‌گردد و در نهایت پس از محاسبه $Avg_۷$ و $Avg_{۱۴}$ و $Avg_{۲۸}$، Ultosc به‌صورت رابطه (۱۲) محاسبه می‌گردد:

$$\text{Avg}_t(T) = \frac{\sum_{i=0}^{T-1} BP_{t-i}}{\sum_{i=0}^{T-1} TR_{t-i}} \quad (۱۱)$$

$$ULTOSC_t = ۱۰۰ * \frac{۴\text{Avg}_t(۷) + ۲\text{Avg}_t(۱۴) + \text{Avg}_t(۲۸)}{۴+۲+۱} \quad (۱۲)$$

۳-۸- شاخص کل و هموزن

شاخص کل قیمت[۲]، نشان‌دهنده تغییرات قیمت‌ها در کل بازار است. است و میانگین افزایش یا کاهش قیمت سهام در بازار را بیان می‌کند. این تغییرات نسبت به سال پایه که در سال ۱۳۶۹ است، بیان می‌شود. منظور از سال پایه سالی است که نوسانات قیمت در آن کمتر است و به‌عنوان مبدأ تغییرات انتخاب می‌شود. این شاخص، همان شاخصی است که در خبرها و گزارش‌ها اطلاع‌رسانی می‌شود و نشانه‌ای از رشد یا افت بازار عنوان می‌شود. مقدار شاخص کل از نسبت ارزش جاری بازار سهام در زمان محاسبه به ارزش جاری بازار سهام در تاریخ مبدأ به دست می‌آید. ارزش جاری هر شرکت نیز برابر با حاصل‌ضرب تعداد سهام آن در قیمت هر سهم شرکت مورد نظر است. از مجموع ارزش جاری شرکت‌های پذیرفته شده در بورس،

[۱] Ultimate Oscillator

[۲] Tehran Exchange Price Index

ارزش جاری بازار به دست می‌آید. در کنار شاخص کل، شاخص دیگری به نام شاخص هم‌وزن وجود دارد که اندازه شرکت‌ها در محاسبه آن تأثیری ندارد و وزن همه شرکت‌ها یکسان در نظر گرفته می‌شود. شاخص کل تمایز میان سهم‌های بزرگ و کوچک قائل می‌شود ولی شاخص هم‌وزن، وزن یکسان برای همه شرکت‌ها در نظر می‌گیرد. از دید فعالان بازار سرمایه، شاخص کل دید شفاف‌تری نسبت به روند بازار می‌دهد و امتیاز بیشتر سهم‌های شاخص‌ساز در محاسبات شاخص کل را از آنان می‌گیرد. به همین دلیل بیشتر نیز مورد استفاده قرار می‌گیرد. وقتی شاخص هم‌وزن منفی است، یعنی بازدهی بیش از نصف شرکت‌های بازار کاهش‌یافته است و زمانی که شاخص هم‌وزن مثبت باشد، یعنی بازدهی بیش از نصف شرکت‌های بازار رشد کرده است. و مفهومی از کل روند بازار را در اختیار می‌گذارد.

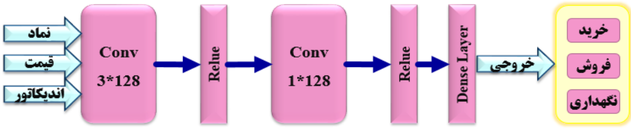
۳-۹- چگونگی استفاده از معیارها

برای طراحی مدل در حدود ۱۹ ویژگی برای یک روز شامل بازده سرمایه ۲، ۵، ۱۰، ۲۱ و ۶۰ روزه، RSI، MACD، ATR، Stochastic، Ultosc، حجم، قیمت باز، قیمت پایانی (جایگزین قیمت بسته)، قیمت بالا و پایین مورد استفاده قرار می‌گیرد. در محاسبه نسبت بازدهی‌های زمانی، بازه‌های زمانی کوتاه مدت ۵ روزه، میان‌مدت ۲۰ روزه و بلند مدت ۶۰ روزه لحاظ می‌گردد. همچنین، حجم معاملات انجام شده در بازه زمانی ۵ ، ۲۰ و ۶۰ روز اخیر محاسبه می‌گردد و نسبت کوتاه مدت به میان‌مدت و نسبت میان‌مدت به بلند مدت برای قیمت و حجم محاسبه می‌گردد. بازه زمانی کوتاه مدت حرکت جدید صعودی یا نزولی بازار را محاسبه می‌کند و سرعت بالاتری دارد. درحالی‌که محاسبه میان‌مدت به بلند مدت واکنش‌های کندتر را نمایش می‌دهد و به محض رشد بازار وارد نمی‌شود. از طرفی به‌رغم سایر بازارهای بورس در دنیا، در ایران دولت نیز در بازار سهام مالکیت دارد. به همین منظور از سهم‌های بزرگ و نقد شونده استفاده می‌شود.

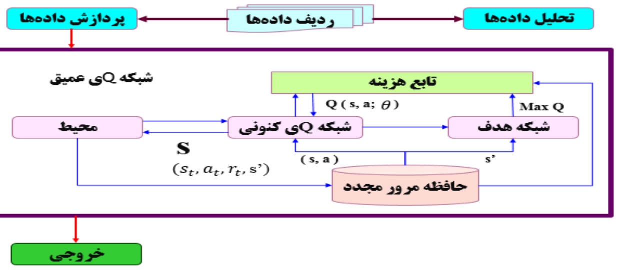
۴- مدل پیشنهادی

همان‌طور که در شکل ۱ نمایش داده شده است، قیمت سهام به همراه اندیکاتورهای مبتنی بر آن به‌عنوان ورودی به شبکه عصبی کانولوشن وارد می‌شوند. مزیت شبکه عصبی کانولوشن این است که ساختار ورودی را تغییر نمی‌دهد و به ارتباط میان همسایگی‌ها اهمیت می‌دهد. همچنین با استفاده از تکنیک به اشتراک‌گذاری وزن‌ها از بیش‌برازش مدل جلوگیری می‌شود. شایان ذکر است اندیکاتورهای محاسبه شده، بر اساس تاریخ با داده‌های قیمت تطبیق داده می‌شود. به منظور محاسبه میزان تطبیق خروجی محاسبه شده با خروجی مورد انتظار، از

تابع هزینه‌ی مجموع مربعات خطا استفاده می‌شود که در فرایند بهینه‌سازی کمینه می‌شود. در شکل ۱ و شکل ۲ نمودار بلوکی نحوه عملکرد معماری پیشنهادی نمایش داده شده است.



شکل ۱. معماری کانولوشن پیشنهادی در پیش‌بینی بورس ایران



شکل ۲. مدل پیشنهادی DRL برای پیش‌بینی بورس ایران

در بخش دوم مدل پیشنهادی از DQN استفاده شده که جایگزینی برای وظایف معامله‌گر است. DQN وظایف یک معامله‌گر را مدل‌سازی می‌کند. اما ایجاد محیطی واقع‌بینانه، چالش‌های فراوانی ایجاد می‌کند. به‌عبارت دیگر، یادگیری تقویتی عمیق که در حوزه‌های دیگر به پیشرفت‌های چشم‌گیری دست یافته، ممکن است با توجه به ماهیت نویزی داده‌های مالی با موانع بزرگ‌تری، از جمله یادگیری تابع مقدار[1] بر مبنای پاداش‌های تأخیری[2]، روبرو شود. روش یادگیری مبتنی بر Q[3] نسخه‌ای از یادگیری تقویتی[4] است که با استفاده از برنامه‌نویسی پویا، بدون دانستن مدل، ماتریس‌های انتقال و پاداش را محاسبه می‌کند. این روش به‌طور مستقیم تابع مقدار-عمل q را به مقدار تقریبی *q بهینه می‌کند. خطمشی**ε**-حریصانه خطمشی است که بررسی تست عمل جدید را در هر وضعیتی تضمین می‌کند و درعین‌حال از تجربیات یاد گرفته شده در گذشته نیز بهره می‌برد. در این مدل استفاده از **ε**-حریصانه پیشنهاد می‌گردد. خطمشی ε-حریصانه عملی را به‌صورت تصادفی و با احتمال ε و بهترین عمل را با توجه به تابع مقدار انتخاب می‌کند. در ابتدا لازم است سعی و خطای بسیاری انجام شود تا فرایند یادگیری بهتر انجام شود. به‌تدریج حضور این عامل تصادفی کاسته می‌شود و بر مبنای

---

[1] value function
[2] delayed rewards
[3] Q-learning
[4] Reinforcment learning

دانش یاد گرفته شده پیش می‌رود. برای همین می‌توان پیش‌بینی کرد درابتدا معامله‌گر مبتنی بر این روش به‌صورت ضرر ده عمل کند و به مرور پاسخ دقیق‌تری ارائه نماید. در اینجا ε دارای مقدار اولیه ۰.۹۹ می‌باشد و در ادامه کوچک‌تر می‌شود تا به صفر برسد. DQN نوعی یادگیری تقویتی است که در آن، برخلاف یادگیری نظارت‌شده، لازم نیست داده‌های آموزشی به‌صورت دستی برچسب زده شوند. بلکه با محیط تعامل و نتیجه را مشاهده می‌کند. این فرایند چندین بار تکرار می‌شود و نمونه‌هایی از تجربه‌های مثبت و منفی به دست می‌آید که به‌عنوان داده‌های آموزشی عمل می‌کنند. به‌این‌ترتیب، با آزمایش و تجربه یادگیری انجام می‌شود. در نتیجه، الگوریتم نیازی به انتخاب اقدامات بر اساس خط‌مشی القا شده ندارد. در این مدل تابع Q با یادگیری نگاشت پیوسته و پارامتری با به‌کارگیری شبکه‌های عصبی عمیق، در اینجا کانولوشن، تقریب زده می‌شود. امروزه شبکه‌های عمیق برای تقریب توابع ارزش، کاربرد بسیاری پیدا کرده‌اند. اما به دلیل تعامل داده‌ها با محیط در مدل، چالش‌های متعددی وجود دارد:

- در توالی‌ها، ممکن است عامل بسیاری از وضعیت‌ها را تجربه ننماید و لازم است تعمیم یابد.
- در حالی‌که در یادگیری نظارت شده فرض بر این است که داده‌های آموزشی از یک توزیع مشخص به‌صورت مستقل نمونه‌برداری شده‌اند که تعمیم‌پذیری نماینده آن به‌درستی برچسب‌گذاری شده است. اما در RL تنها یک نمونه در هر مرحله زمانی وجود دارد، بنابراین لازم است یادگیری به‌صورت برخط انجام شود.
- هنگامی‌که وضعیت‌های متوالی مشابهند و داده‌ها به‌شدت هم‌بسته هستند و رفتار در وضعیت‌ها، توزیع رفتار در حالت‌ها و اقدامات ایستا نیستند، تنها با انجام یادگیری پیوسته تغییرات صورت گرفته پوشش داده می‌شود.

۴-۱- الگوریتم یادگیری عمیق Q

در الگوریتم DQN[۱]، ارزش هر عمل[۲] برای هر حالتِ معین با استفاده از شبکه عصبی عمیق تخمین زده می‌شود. در ابتدا این روش در بازی آتاری معرفی شد که در آن عامل با استفاده از پیکسل‌های ورودی، بازی را می‌آموخت. الگوریتم DQN تابع ارزش-عمل q را با یادگیری مجموعه‌ای از وزن‌های θ تقریب می‌زند. θی حاصل از شبکه عصبی مرتبط با DQN، حالت‌ها

---

[۱] Deep Q-learning

[۲] Action Value

را به اقدامات نگاشت می‌کند تا رابطه q(s,a,θ)≈ q*(a,θ) برقرار گردد. این الگوریتم کاهش گرادیان را بر اساس تابع ضرری اعمال می‌کند که مجذور تفاوت میان DQN برآورد شده و مقدار واقعی را مطابق با رابطه (۱۳) و (۱۴) محاسبه می‌کند[۲۵]:

$$y_i = \mathbb{E}\left[r + \gamma \max_{a'} Q\left(s'.a';\ \theta_{i-1} | s.a\right)\right] \quad (۱۳)$$

$$L_i(\theta_i) = \left(y_i - Q\left(s.a;\ \theta\right)\right)^{۲} \quad (۱۴)$$

که در آن y_i، Qی هدف است و $Q(s.a;\ \theta)$ پیش‌بینی کنونی است و تفاضل $y_i - Q\ (s.a;\ \theta)$ خطای تفاوت زمانی[۱] می‌باشد. وزن‌های DQN، در بهبود نتیجه، نقش تعیین کننده دارد. الگوریتم یادگیری Q به‌جای محاسبه گرادیان کامل، ازSGD[۲] استفاده می‌کند و وزن‌های θ_i را پس از هر مرحله زمانی i به‌روزرسانی می‌کند. عامل تریدر، به منظور کشف فضای حالت-عمل، با احتمال ϵ خطمشی حریصانه را دنبال می‌کند و با احتمال ۱ − ϵ عملی را با بالاترین مقدار Qی پیش‌بینی شده انتخاب می‌کند. معماری پایه DQN به‌گونه‌ای اصلاح شده که فرآیند یادگیری را کارآمدتر کند و نتیجه نهایی را بهبود بخشد. در ادامه به این اصلاحات اشاره شده است.

❖ **تکرار تجربه**

در اینجا از تجربیاتی که در گذشته انجام شده استفاده می‌کند. به این صورت که اشتباهاتی که قبلاً انجام شده را در جدولی نگهداری و آنها را دسته‌بندی می‌کند و با توجه به این اشتباهات، خود را دوباره Train می‌کند و به‌این‌ترتیب، از رخداد دوباره اشتباهات گذشته جلوگیری می‌شود. در تکرار تجارب مربوط به عامل تریدر سابقه‌ای از حالت، پاداش و حالت بعدی ذخیره می‌شود. تکرار تجربه، کارایی نمونه را افزایش می‌دهد و همبستگی خودکار نمونه‌های جمع‌آوری‌شده در طول یادگیری آنلاین را کاهش می‌دهد و بازخورد را محدود می‌کند[۲۵].

❖ **شبکه هدف**

برای تضعیف حلقه‌ی بازخورد ناشی از پارامترهای شبکه در به‌روزرسانی وزن‌های NN، الگوریتم DQN از DeepMind، برای کنترل سطح انسانی، استفاده می‌کند تا شبکه‌ی هدفی را به کار بندد که در طول زمان به آهستگی و به‌آرامی تغییر یابد. معماری شبکه‌ی هدف مشابه معماری شبکه‌ی Q است، اما وزن‌های آن یعنی θ^- پس از گام‌های τ تنها زمانی به‌صورت دوره‌ای به‌روزرسانی می‌شوند که از شبکه‌ی Q کپی شده باشند، در غیر این صورت بی‌تغییر

---
[۱] Temporal Difference: TD
[۲] stochastic gradient descent (SGD)

باقی می‌مانند. شبکه‌ی هدف، پیش‌بینی‌های هدف TD را موجب می‌شود، یعنی مطابق با رابطه ۱۵ جای شبکه‌ی Q را می‌گیرد[۲۷]:

$$y_i = \mathbb{E}[r + \max_{a'} Q(s'.a'; \theta^- | s.a)] \quad (۱۵)$$

❖ **یادگیری Q عمیق مضاعف**

DDQN از دو شبکه عصبی تشکیل شده است. یکی شبکه آنلاین که زودبه‌زود به‌روز می‌شود و دیگری، یعنی شبکه هدف[۱] برای جلوگیری از بیش برازش، دیربه‌دیر به‌روز می‌شود. پس از هر چند بار به‌روز شدن، وزن‌های آموخته شده در شبکه هدف کپی می‌شود. در یکی عمل انتخاب می‌شود. درحالی‌که در دیگری آن عمل برآورد و تخمین زده می‌شود تا عمل[۲] بهتر مشخص گردد. پس دو شبکه عصبی تعریف می‌شود که یکی θ و دیگری θ' است. یکی از این دو بهترین است. هنگامی‌که وضعیت بعدی به دست می‌آید، شبکه مناسب‌تر مشخص می‌گردد و پاداش بر اساس θ' و عمل‌ها بر اساس θ انجام می‌شود. یادگیری Q مقادیر عمل را بیش‌ازحد برآورد می‌کند، زیرا از مقادیر بیشینه‌ی عمل برآورد شده نمونه‌گیری می‌کند. اگر این سوگیری به‌صورت یکسان اعمال نشود و اولویت‌های عمل را تغییر دهد، می‌تواند بر فرآیند یادگیری و خط‌مشی اثر منفی بگذارد[۲۸]. مطابق با رابط (۱۵) برای جداسازی برآورد مقادیر عمل از انتخاب عمل‌ها، الگوریتم DDQN[۳] برای انتخاب بهترین عمل از وزن‌های θی یک شبکه با توجه به حالت بعدی استفاده می‌کند و وزن‌های θ' شبکه‌ی دیگر را برای ارائه‌ی برآورد مقدار عمل مربوطه به کار می‌برد:

$$y_i = \mathbb{E}[r + \gamma Q\left(S'.arg\max_{a'} Q(S_{t+۱}.a; \theta_t)\right); \theta'_t) \quad (۱۵)$$

۴-۲- معیارهای اندازه‌گیری عملکرد

برای ارزیابی و مقایسه استراتژی‌های مختلف و یا بهبود استراتژی موجود، معیارهایی مورد نیاز است تا عملکرد را با توجه به اهداف مورد نظر بیان نماید. در سرمایه‌گذاری و تجارت، رایج‌ترین اهداف بازدهی و ریسک سرمایه‌گذاری است. این معیارها با معیارهای بیانگر فرصت‌های سرمایه‌گذاری، مانند نرخ بهره بانکی، مقایسه می‌شوند. معیارهای بسیاری برای این منظور تعریف شده است. در این بخش به معیارهایی که در این مقاله به کار می‌رود، اشاره می‌شود. این معیارها هنگامی مفید است که رویکردهای مختلف برای بهینه‌سازی عملکرد مورد نظر است:

---
[۱] target network
[۲] Action
[۳] Double DQN

مطابق با رابطه (۱۶)، R سری زمانی حاصل از بهره در یک دوره زمانی از تاریخ ا تا T می‌باشد و مطابق با رابطه (۱۷)، R_f سری زمانی نرخ‌های بدون ریسک می‌باشد که در این پژوهش نرخ بهره بانکی است و در نهایت، مطابق با رابطه (۱۸)، R_e سود اضافی است که از تفریق R و R_f حاصل می‌شود:

$$R= (r_1, ..., r_T) \quad (۱۶)$$

$$R_f= (r_{f1}, ..., r_{fT}) \quad (۱۷)$$

$$R_e = (r_1 - r_{f1}, ..., r_T- r_{fT}) \quad (۱۸)$$

سود و ریسک همواره در تبادل‌اند. ریسک بیشتر می‌تواند سود بیشتری را به همراه داشته باشد، اما در شرایط زیان‌آور احتمال زیان ناشی از آن نیز بیشتر است. برای بیان اینکه چگونه استراتژی‌های مختلف این تبادل را رقم می‌زند، تناسب‌هایی برای نمایش نسبت اندازه‌گیری سود در واحد تعریف می‌شود. دو نسبت مهم نسبت شارپ و نسبت اطلاعات هستند. از این‌رو در ابتدا این دو نسبت توصیف می‌شوند و در ادامه ارزیابی داده‌ها به‌صورت مختصر بیان خواهد شد.

۴-۲-۱- نسبت شارپ

مطابق با روابط (۱۹)، (۲۰)، (۲۱) و (۲۲) نسبت شارپ که با SR نمایش داده می‌شود، به منظور کمک به سرمایه‌گذاران برای درک بهتر سود یک سرمایه‌گذاری نسبت به ریسک آن می‌باشد. این نسبت میانگین بازده به‌دست‌آمده مازاد بر نرخ سود بدون ریسک، به ازای هر واحد از نوسان‌پذیری یا ریسک کل است. با کسر نرخ بدون ریسک از میانگین بازده، عملکرد فعالیت‌های مربوط به قبل از ریسک را می‌توان جدا نمود، یعنی سود مازاد مورد انتظار را با نوسانات آن مقایسه می‌کند که با انحراف استاندارد آن اندازه‌گیری می‌شود. به‌این‌ترتیب، میزان جبران را به‌عنوان میانگین سود اضافی به ازای هر واحد ریسک انجام شده اندازه‌گیری می‌کند. هرقدر میزان نسبت شارپ بالاتر باشد، نشان دهنده آن است که بازدهی به‌دست‌آمده با ریسک کمتری رخ داده است. منفی بودن نسبت شارپ نیز بیانگر آن است که سرمایه‌گذاری در چنین حالتی توجیه‌پذیر نیست و توصیه نمی‌شود. μ متوسط میزان رشد و بازدهی در بازه زمانی مشخص شده است. R_f سود بانکی یا همان سود بدون ریسک است. مقدار نسبت شارپ معمولاً به‌صورت نسبت منفی به‌معنای سرمایه‌گذاری توجیه‌ناپذیر، نسبت ۱ و کمتر از ۱ به‌معنای ضعیف و غیرقابل‌قبول، نسبت بیشتر از ۱ به‌معنای قابل‌قبول، نسبت بالاتر از ۲ به‌معنای بسیار خوب و نسبت ۳ و بالاتر به‌معنای بسیار عالی تحلیل می‌شود[۲۵].

$$\mu = E(R_t) \quad (۱۹)$$

$$\sigma^{۲}{}_{R^e} = Var\ (R_t - R^f) \quad (۲۰)$$

$$SR= \frac{\mu - R^f}{\sigma_{R^e}} \quad (۲۱)$$

$$SR= \frac{\mu - R_f}{\sigma_{R^e}} \quad (۲۲)$$

۴-۲-۲- معیار ارزیابی

در اینجا به منظور ارزیابی خروجی مدل پیشنهادی از مقایسه بازدهی سبدی که از ۲۹ سهم نقد شونده بازار به‌صورت خرید و نگهداری به‌دست آمده و بازدهی حاصل از خرید و فروش این سهم‌ها توسط عامل ترید به‌دست آمده، استفاده می‌شود.

۴-۳- الگوریتم بهینه‌سازی وفقی Adam

در الگوریتم Adam کلمه Adam از عبارت "Adaptive moments" مشتق شده است. این الگوریتم را می‌توان ترکیبی از دو الگوریتم گشتاور و RMSProp در نظر گرفت. اول اینکه در این الگوریتم به‌طور مستقیم از ممان مرتب اول گرادیان، با وزن‌دهی نمایی استفاده شده است. مناسب‌ترین راه برای اضافه کردن گشتاور به RMSProp، اعمال گشتاور به گرادیان مقیاس شده است. دوم اینکه در اینجا تصحیح اریبی بر تخمین‌های ممان مرتبه اول و دوم اعمال می‌شود. در RMSProp از تخمین ممان مرتبه دوم، بدون عامل تصحیح کننده استفاده می‌شود که موجب بالا رفتن اریب در اوایل فرآیند یادگیری می‌شود. الگوریتم Adam در الگوریتم ۱-۰ به‌عنوان الگوریتمی شناخته می‌شود که نسبت به انتخاب فراپارامترها حساسیت زیادی ندارد.

الگوریتم ۱-۰: الگوریتم Adam

ورود نرخ یادگیری:	$(\epsilon = ۰.۰۰۱)$	۱
ورود نرخ میرایی نمایی تخمین ممان‌ها در بازه:	$(\rho_۱ = ۰.۹. \rho_۲ = ۰.۹۹۹)$	۲
تنظیم ثابت کوچک δ جهت پایداری عددی:	$(\delta = ۱۰^{-۸})$	۳
مقداردهی اولیه بردارهای تخمین ممان‌ها:	$(r = ۰. s = ۰)$	٤
مقداردهی اولیه پارامترهای مدل:	θ	٥
while) عدم فرارسیدن معیار توقف(		٦
نمونه‌برداری m نمونه از مجموعه داده‌ها $\{x^{(۱)}. \dots . x^{(m)}\}$ به همراه برچسب‌های مربوطه $\{y^{(۱)}. \dots . y^{(m)}\}$		٧
محاسبه گرادیان:	$g \leftarrow \frac{۱}{m} \nabla_\theta \sum_i L(f(x^{(i)}. \theta), y^{(i)})$	٨

به‌روزرسانی تخمین ممان مرتبه اول: $s \leftarrow \rho_1 s + (1 - \rho_1) g$

به‌روزرسانی تخمین ممان مرتبه دوم: $r \leftarrow \rho_2 r + (1 - \rho_2) g \odot g$

اصلاح اریب ممان مرتبه اول: $\hat{s} \leftarrow \frac{s}{1 - \rho_1}$

اصلاح اریب ممان مرتبه دوم: $\hat{r} \leftarrow \frac{r}{1 - \rho_2}$

محاسبه نرخ یادگیری تطبیقی به‌صورت عنصر به عنصر: $\Delta\theta = -\frac{\hat{s}}{\delta + \sqrt{\hat{r}}} \epsilon$

انجام به‌روزرسانی: $\theta \leftarrow \theta + \Delta\theta$

۵- پیاده‌سازی مدل

برای طراحی از برنامه پایتون و کتابخانه TA-Lib استفاده شده است. TA-Lib شامل بیش از ۱۵۰ اندیکاتور تکنیکال است که توسعه‌دهندگان نرم‌افزارهای تجاری برای تجزیه‌وتحلیل‌های تکنیکال داده‌های بازار مالی استفاده می‌کنند. این ویژگی با استفاده از برنامه پایتون قابل اجراست. برای طراحی بخش یادگیری Q از یک مدل دولایه کانولوشن بهره گرفته شده است. شبکه عصبی کانولوشن، شبکه‌ای است که بر اساس اعمال کرنل‌ها بر ورودی عمل می‌کند. به‌طوری‌که کرنل‌ها برای کل داده‌های ورودی به اشتراک گذاشته می‌شوند. همچنین در شبکه‌های کانولوشن برای کاهش تعداد ویژگی‌ها از لایه‌های پولینگ استفاده می‌شود. در شبکه معرفی شده، پس از دو لایه کانولوشن، لایه سوم به‌صورت شبکه عصبی معمولی اتصال کامل در نظر گرفته شده است که به منظور فراهم کردن خروجی استفاده شده است. در لایه اول داده‌های قیمتی به همراه ۱۴ اندیکاتور منتخب که در بخش پیش بیان شد، وارد سیستم کانولوشن می‌شود. در هر مرحله یک هدف وجود دارد و برای جلوگیری از بیش برازش داده‌های تحلیل شده به‌روز می‌شود و ۱۲۸ خروجی ایجاد می‌کند. این کانولوشن یک بعدی است و روزهای کاری را بررسی می‌کند. به‌این‌ترتیب، در ابتدا ۷۴۰۰ پارامتر قابل مشاهده است. در ادامه نرون‌های ضعیف حذف و نرون‌های قوی باقی می‌مانند. مشخصات و ساختار شبکه عصبی دو لایه کانولوشن و یک لایه خروجی در شکل ۳ نمایش داده شده است.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
Conv_0 (Conv1D)              (None, 3, 128)            5888
_________________________________________________________________
Conv_1 (Conv1D)              (None, 1, 128)            49280
_________________________________________________________________
Output (Dense)               (None, 1, 2)              258
=================================================================
Total params: 55,426
Trainable params: 55,426
Non-trainable params: 0
```

شکل ۳. ساختار شبکه عصبی کانولوشن

در اینجا، کانولوشن‌ها به‌صورت پنجره سه تایی انجام می‌شوند. در هر یک از دو لایه ورودی ۱۲۸ کانولوشن و در لایه خروجی سه نوع خروجی "خرید"، "نگهداری" و "فروش" و در مجموع ۲۵۸ پارامتر وجود دارد. روند آموزش در شکل ۴ نمایش داده شده است. در ابتدا پاداش‌ها مقادیر اندکی دارند و به مرور با پیش‌روی فرایند یادگیری، نتیجه پیش‌بینی به بازار نزدیک و سپس از آن پیشی می‌گیرد. در اینجا منظور از بازار استراتژی خرید و نگهداری است.

```
 10 | 00:00:23 | Agent:  34.9% ( 34.9%) | Market:  97.9% ( 97.9%) | Wins: 20.0% | eps:  0.986
 20 | 00:00:48 | Agent:  37.3% ( 39.7%) | Market: 128.3% (158.8%) | Wins: 25.0% | eps:  0.972
 30 | 00:01:13 | Agent:  30.5% ( 16.9%) | Market:  94.7% ( 27.5%) | Wins: 36.7% | eps:  0.958
 40 | 00:01:41 | Agent:  28.5% ( 22.6%) | Market:  85.9% ( 59.3%) | Wins: 37.5% | eps:  0.943
 50 | 00:02:09 | Agent:  24.8% ( 10.2%) | Market:  71.5% ( 14.0%) | Wins: 40.0% | eps:  0.929
 60 | 00:02:39 | Agent:  24.6% ( 23.2%) | Market:  73.3% ( 82.6%) | Wins: 40.0% | eps:  0.915
 70 | 00:03:10 | Agent:  24.5% ( 23.7%) | Market:  82.1% (134.4%) | Wins: 34.3% | eps:  0.901
 80 | 00:03:41 | Agent:  25.5% ( 32.8%) | Market:  80.8% ( 71.8%) | Wins: 33.8% | eps:  0.887
 90 | 00:04:14 | Agent:  25.3% ( 23.3%) | Market:  82.4% ( 95.7%) | Wins: 34.4% | eps:  0.873
100 | 00:04:47 | Agent:  25.2% ( 24.2%) | Market:  81.8% ( 76.0%) | Wins: 34.0% | eps:  0.859
110 | 00:05:22 | Agent:  23.3% ( 16.6%) | Market:  78.3% ( 63.4%) | Wins: 36.0% | eps:  0.844
120 | 00:05:59 | Agent:  22.7% ( 33.6%) | Market:  78.5% (160.8%) | Wins: 36.0% | eps:  0.830
130 | 00:06:35 | Agent:  21.8% (  7.7%) | Market:  79.7% ( 38.6%) | Wins: 33.0% | eps:  0.816
140 | 00:07:13 | Agent:  22.6% ( 30.3%) | Market:  81.3% ( 75.3%) | Wins: 30.0% | eps:  0.802
150 | 00:07:53 | Agent:  23.4% ( 18.7%) | Market:  90.7% (108.3%) | Wins: 27.0% | eps:  0.788
160 | 00:08:34 | Agent:  26.2% ( 51.4%) | Market:  93.1% (106.6%) | Wins: 24.0% | eps:  0.774
170 | 00:09:14 | Agent:  27.1% ( 31.9%) | Market:  84.8% ( 51.1%) | Wins: 27.0% | eps:  0.760
180 | 00:09:58 | Agent:  25.8% ( 20.8%) | Market:  88.2% (105.9%) | Wins: 26.0% | eps:  0.745
190 | 00:10:43 | Agent:  31.0% ( 74.3%) | Market: 101.7% (230.9%) | Wins: 23.0% | eps:  0.731
```

شکل ۴. روند آموزش در شبکه عصبی کانولوشن

شبیه‌سازی از سه بخش تشکیل شده است. در بخش اول آماده‌سازی داده انجام می‌شود. در بخش دوم داده‌های آماده شده جمع‌آوری می‌شود. سپس شبیه‌سازی انجام می‌گردد. محیط ترید براساس GYM ENV انجام شده است که روند آن در سه گام تعیین میزان هزینه معاملات، تعریف سه عمل اصلی شامل خرید، نگهداری و فروش و تعریف محیط مشاهده چند بعدی شامل ارزش کمینه تا بیشینه می‌باشد. در یادگیری عمیق عامل تریدر نقش تصمیم‌گیری را در انجام معاملات بر عهده دارد که در اینجا شبکه کانولوشن است. DDQN محیط را مشاهده می‌کند و بر اساس آن تصمیم می‌گیرد. در DDQN مشاهدات کانولوشن به ماشین یادگیری سپرده می‌شود و در خروجی شبکه عمیق نتیجه مناسب شامل خرید، فروش یا نگهداری اعلام می‌گردد. در DDQN هر سطر بیانگر حالت‌ها و ستون‌ها بیان‌گر عملیات است. حالت موجود در بورس مطالعه می‌شود و در سطر مناسب جایگزین می‌گردد. هر پیش‌بینی

که منجر به سود شود، به همان نسبت پاداش دریافت می‌کند و پیش‌بینی که منجر به ضرر شود به همان نسبت جریمه دریافت می‌شود. ستونی که بالاترین پاداش را دریافت می‌کند، عمل مورد انتخاب برای خرید یا فروش یا نگهداری است. به‌این‌ترتیب، نتیجه خرید یا فروش برای نتیجه‌گیری روز بعد ذخیره می‌شود. در شبیه‌سازی سود و زیان حاصل از هر عمل (معامله) محاسبه شده و نتیجه را اعلام می‌گردد. قابل ذکر است عامل ترید داده‌های یک سال را ترید می‌کند. اما این‌که سال از کجا شروع می‌شود، تصادفی در نظر گرفته می‌شود. آلفا یک‌دهم در نظر گرفته می‌شود. γ ضریبی برای تعیین میزان دقت پاداش است. دستیابی به پاسخ مورد نظر پس از هر مرحله زمانی همراه با صرف زمان و تأخیر است و این در γ اعمال می‌گردد. γ عامل افت ارزش و عددی ثابت است. در نهایت TD TARGET پاداشی است که دریافت می‌شود.

۵-۱- نتایج تجربی

در سیستم یادگیری تقویتی جدول حالت-عمل[۱] محاسبه می‌شود و بیان‌گر ارزش هر عمل در هر حالت است. هدف این است که برای هر وضعیت بهترین عمل انتخاب شود. این جدول (شبکه عصبی کانولوشن) مدام در حال بهینه شدن است. در تحلیل بورس تعداد وضعیت‌ها محدود نیست و نمودار قیمت حالت‌های بی‌شماری دارد. بنابراین حالات بورس به یادگیری عمیق سپرده می‌شود تا تصمیم‌گیری شود. نحوه یادگیری در یادگیری تقویتی بر اساس پاداش و تنبیه است. در ابتدا عامل ترید به‌صورت تصادفی خرید و فروش می‌کند و سود و زیان حاصل از آن را به‌عنوان دانش ثبت می‌کند. پس از مدتی به‌جای روند تصادفی از دانش نیز استفاده می‌کند و به مرور تصمیمات خود را بر اساس دانش بهینه می‌کند. در DDQN دو شبکه موجود است یکی با هر داده‌ای به‌صورت آنلاین یاد می‌گیرد و دیگری به‌صورت بلند مدت می‌آموزد و بر هدف مورد نظر متمرکز است. شبکه آنلاین آموخته‌های خود را به شبکه هدف منتقل می‌کند و شبکه هدف تصمیمات را اتخاذ می‌کند. به‌این‌ترتیب، دو شبکه وجود دارد. هر بار داده‌ها در ۲۱ روز اخیر بررسی می‌شود. منظور از ۲۱ روز یک ماه کاری است. در این بررسی داده‌های پانزده سال اخیر تحلیل می‌گردد. گام‌های زمانی به‌صورت بازه‌های زمانی ۵ روزه انتخاب می‌شود[۲] و میزان سود به‌دست آمده پس از ۲۱ روز را مشخص می‌کند. کانولوشن حجم خروجی را به مرور کوچک می‌کند. پس در ابتدا اطلاعات ورودی شامل ۵ روز است که هر روز ۱۹ ویژگی دارد و کانولوشن از تمام این ۱۹ ویژگی خروجی می‌دهد. اما در هر مرحله به

[۱] State Action

[۲] Time step

دلیل کانولوشن سه تایی، دو حذف می‌شود. در کل ۱۲۸ کرنل مختلف بر ویژگی‌های ورودی اعمال می‌شود. کرنل‌ها که مسئول استخراج ویژگی هستند، ویژگی‌های مؤثر در پیش‌بینی بازار را استخراج می‌کند. در لایه خروجی، در سه حالت نتیجه نهایی خود را نسبت به هر سهم نقد شونده به‌صورت "خرید"، "نگهداری" و یا "فروش" اعلام می‌کند. در شکل مشاهده می‌شود که در ابتدا بازار وضعیت بهتری دارد اما به مرور معامله‌گر RL از بازار پیشی می‌گیرد. منظور از بازار، میزان سودی است که سرمایه‌گذار در صورت سرمایه‌گذاری در سهم‌های نقد شونده در بازه زمانی یک ساله به‌دست می‌آورد. در نهایت، در تکرار هزارم فرایند یادگیری از یادگیری به بهره‌مندی از دانش شیفت می‌شود. که خروجی بهینه می‌شود. در جدول ۲ میزان سود و زیان در دو حالت خرید و نگهداری و یا تصمیم‌گیری با عامل تریدر نمایش داده شده است. در ستون آخر میزان تفاضل این دو نمایش داده شده است. مشاهده می‌شود که به‌عنوان مثال در سهام "فرابورس"، درصورتی‌که عامل ترید تصمیم‌گیری به معامله نماید، سودی معادل ۲۰.۹۹ حاصل می‌شود. درحالی‌که استراتژی خرید و نگهداری سبب ضرری معادل ۷.۹۶ حاصل می‌شود.

جدول ۲. مقایسه میزان بهره‌وری حاصل از عامل تریدر و بازار

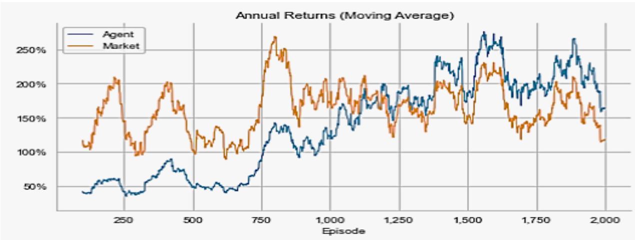
نماد	سود حاصل از استراتژی نگهداری و خرید	سود حاصل از ماشین	تفاضل دو استراتژی
فولاد	۴۹.۱۳	۳۳.۶۱	-۱۵.۵۱
فملی	۳۰.۳۷	۳۵.۷۳	۵.۳۶
شبندر	۹۹.۹۲	۷۲.۰۴	۲۷.۸۷
شپنا	۲۸.۷۰	۵۵.۴۳	۲۶.۷۳
وپاسار	۳۲.۱۰	۳۸.۵۴	۶.۴۴
وبملت	-۲۱.۲۴	۳.۸۴	۲۵.۰۸
خودرو	-۱۷.۹۴	-۰.۷۱	۱۷.۲۲
بپاس	۲۰.۴۹	-۰.۶۷	-۲۱.۱۶
بورس	-۲۷.۷۴	۰.۹۸	۲۸.۷۲
شپدیس	۸۲.۵۳	۲۸.۰۴	-۵۴.۴۹
وامید	۱۸.۷۰	۱۸.۰۴	-۰.۶۶
کگل	۱۰۷.۶۶	۱۱۱.۳۸	۳.۷۲
خساپا	-۱۷.۹۵	۲.۹۷	۲.۹۲
شتران	۱۱.۰۲	-۱۲.۳۰	-۲۳.۳۲
وغدیر	۴۷.۰۴	۷۴.۳۹	۲۷.۳۶
خبهمن	۱۸.۰۰	۴۰.۶۶	۲۲.۶۶

نماد	سود حاصل از استراتژی نگهداری و خرید	سود حاصل از ماشین	تفاضل دو استراتژی
برکت	۳.۴۰	۱۵.۷۸	۱۲.۳۸
فارس	۱۶.۱۱	۲۶.۱۳	۱۰.۰۱
وتجارت	-۳۳.۰۷	-۲۶.۵۹	۶.۴۸
جم	۶۵.۷۰	۴۵.۹۰	-۱۹.۸۰
فخوز	۱۵.۸۶	۱۹.۹۵	۴.۰۹
آریان	-۸.۸۳	-۹.۵۳	-۰.۶۹
فرابورس	-۷.۹۶	۲۰.۹۹	۲۸.۹۴
پارسان	۶۱.۴۵	۶۱.۴۵	۸.۷۳
تاپیکو	۴۴.۰۵	۴۴.۰۵	-۱۴.۳۲
پارس	۳۷.۴۲	۵۳.۹۹	۱۶.۵۷
پاکشو	-۵.۵۴	-۱.۳۸	۴.۱۶
ومعادن	۴۶.۴۹	۲۷.۶۷	-۱۸.۸۲
کچاد	۸۵.۲۵	۷۲.۶۱	-۱۲.۶۴

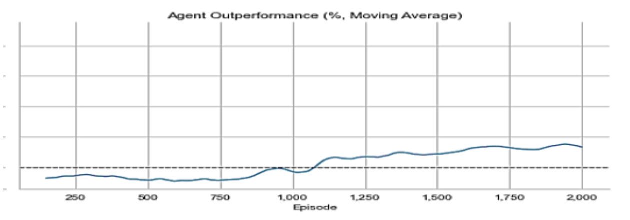
تست داده‌ها از ۹۹/۹/۱۱ تا ۱۴۰۰/۱۱/۴ انجام شده است. درحالی‌که مطابق با شکل ۵، در این دوره، شاخص بازار بازدهی ۱۱٪- را تجربه کرده است. شاخص‌ها نشان می‌دهند درحالی‌که در این آزمایش تمرکز بر ۲۹ سهم نقد شونده بازار بوده است، سود حاصل از عامل ترید برابر با ۲۹٪ می‌باشد. در شکل ۶ و شکل بازگشت سرمایه سالانه[۱] و ماهانه نمایش داده شده است. عامل با رنگ آبی و بازار با رنگ قرمز نمایش داده شده است. منظور از بازار میزان سودآوری حاصل از سرمایه‌گذاری در سهم‌های نقد شونده در بازه زمانی یک ساله است. در این شکل اختلاف خروجی شبکه عصبی با بازار قابل مشاهده است. در ابتدا سرمایه‌گذاری در بازار با سود بیشتری مواجه است. اما به مرور پس از ۱۰۰۰ اپیزود، شبکه طراحی شده پیشی می‌گیرد و عامل ترید بر بازار غلبه می‌کند.

[۱] Annual Return

شکل ۵. شاخص بازار در بازه زمانی آذر تا بهمن ۱۴۰۰[۲۹]



شکل ۶. بازگشت سرمایه سالانه



شکل ۷. بازگشت سرمایه ماهانه

۵-۲- چالش‌ها

این مقاله در پایتون پیاده‌سازی و اجرا شده است. در محیطی به نام gym در پایتون، صدها محیط آزمایشگاهی از پیش پیاده‌سازی شده تعبیه شده است. اما برای بورس چنین محیطی وجود ندارد. به‌ناچار از gym غیر بورسی استفاده شد و همه امکانات ویژه بورس و بخصوص بورس ایران در آن پیاده‌سازی شد. از طرفی شرایط بورس ایران متفاوت از دنیا است. به‌عنوان مثال در بورس ایران، سازمان بورس به ازای هر خرید و فروش ۱٪ دریافت می‌کند. درصورتی‌که در سایر نقاط دنیا اغلب این میزان ۰.۱٪ است. تفاوت دیگر در کارگزاری است و اگر مبلغی باقی بماند، در سایر نقاط دنیا مالیات کسر می‌شود، اما در ایران مبلغی دریافت

نمی‌شود. از سویی دیگر یک‌طرفه بودن بورس در ایران شرایط ویژه‌ای ایجاد می‌کند که به‌راحتی قابل پیاده‌سازی نیست و باید تمهیداتی اندیشیده شود.

۶- واکاوی

۶-۱- مباحثه و تفسیر نتایج

در این مقاله با چالش‌های متعددی مواجه شدیم که با به‌کارگیری تمهیدات ویژه اثر آنها تا حدود زیادی خنثی شده است. در گام اول با به‌کارگیری مدل‌های RL و توانایی آنها در مواجهه با محیط‌های غیر ایستا توانستیم با توجه به نمودار شکل ۳-۵ برتری قابل توجهی نسبت به استراتژی خرید و نگهداری به دست آوریم. همچنین با به‌کارگیری مدل‌های کانولوشن به‌جای جداول Q از بیش برارزش مدل به دلیل وجود داده‌های کم برای آموزش مدل جلوگیری به‌عمل آمده است. از طرفی با استفاده از اطلاعات موجود در حجم معاملات از این سیگنال به‌عنوان نقشی مکمل در پیش‌بینی روند آینده سهمها بهره گرفته شده است. در نهایت شبیه‌سازی را نشان می‌دهد مدل پیشنهادی توانسته به‌طور مؤثری هزینه نسبت بالای انجام معاملات در بورس ایران را پوشش دهد.

۶-۲- نتیجه‌گیری

بهینگی استراتژی‌های تصمیم‌گیری از موضوعات جذاب محققین و سرمایه‌گذاران است. در این مقاله براساس یادگیری تقویتی عمیق، مدلی ارائه شد که با استفاده از شبکه عصبی کانولوشن داده‌های قیمتی به همراه اندیکاتورها که توصیف کننده روند قیمت و نوسانات هستند، به‌صورت وضعیت سیستم در نظر گرفته شده‌اند و بهره‌گیری از سیستم تنبیه و پاداش و یادگیری روابط عمل-حالت در DDQN تصمیم‌گیری ترید انجام می‌گیرد. این بررسی در بازه زمانی پانزده ساله شامل ۲۹ سهم نقد شونده در بورس ایران انجام شده است. از جمله نوآوری‌های این مقاله مطالعه بر سهم‌های نقد شونده بورس ایران و نه تمرکز بر گروهی خاص می‌باشد تا بدین ترتیب بتوان از آن بتوان در مدیریت صندوق‌های سرمایه‌گذاری استفاده نمود. در انتها لازم به ذکر است، به منظور دستیابی به نتایج دقیق‌تر، می‌توان ویژگی‌های خرد و کلان اقتصادی را نیز به مجموعه ویژگی‌ها و اندیکاتورهای استفاده شده اضافه نمود. همچنین، می‌توان شاخص‌های بنیادی تعیین کننده در عملکرد اثربخش شرکت‌ها را نیز بررسی نموده و در فرایند تصمیم‌گیری مورد توجه قرار داد.

۷- منابع

[۱] Sharma, A., D. Bhuriya, and U. Singh. Survey of stock market prediction using machine learning approach. in ۲۰۱۷ International conference of electronics, communication and aerospace technology (ICECA). ۲۰۱۷. IEEE.

[۲] Sutton, R.S. and A.G. Barto, Reinforcement learning: An introduction. Robotica, ۱۹۹۹. ۱۷(۲): p. ۲۲۹-۲۳۵.
[۳] Barroso, B., R. Cardoso, and M. Melo, Performance analysis of the integration between Portfolio Optimization and Technical Analysis strategies in the Brazilian stock market. Expert Systems with Applications, ۲۰۲۱. ۱۸٦: p. ۱۱٥٦۸۷.
[٤] DEHGHAN, A. and M. Kamyabi, On Tehran stock market in cyclical condition. ۲۰۱۹.
[٥] Hejazi, R., et al., The Relationship between Stock Prices Crash and Dedicated Institutional Investors and Transient Institutional Investors. ۲۰۲۰.
[٦] Ali Asghar, A.R. and B. Lari Semnani, Assessing the Relationships of Bank Deposits and Governmental Industrial Development Bonds Investments with the Attracitveness of Investing (Lequidity and Capitalization) in Tehran Stock Exchange (TSE). Management Research in Iran, ۲۰۰۷. ۱۱(۲۰): p. ۱-۲۹.
[۷] Fairchild, R.J. and J. Kinsella, An Emotional Finance Framework for Examining Bubbles and Crashes. Available at SSRN ۳۹۹۹۳۲۳, ۲۰۲۲.
[۸] Hatefi Madjumerd, M., G. Zamanian, and M.N. Shahiki Tash, Evaluation of Multiple Bubbles in the Stock Market of Tehran. Quarterly Journal of Quantitative Economics, ۲۰۱۷. ۱٤(۲): p. ۸٥-۱۱۰.
[۹] Bagherzadeh, S., The Initial Public Offerings Underpricing and Its Determinants in Tehran Stock Exchange. Human Sciences Modares, ۲۰۱۱. ۱٥(۱): p. ۷۷-۱۰۷.
[۱۰] Afsar, A. and F. Helyel, A Hybrid Approach to Portfolio Optimization Using Technical Analysis and Data Mining. Modern Research in Decision Making, ۲۰۱۷. ۲(۲): p. ۱-۲۲.
[۱۱] Katani, S., F. Samadi, and Z. Hajiha, Optimization of multi-objective portfolio using imperialist competitive algorithm in Tehran Stock Exchange. ۲۰۲۱.
[۱۲] BELAGHİ, R.A., M. AMİNNEJAD, and Ö.G. ALMA, Stock Market Prediction Using Nonparametric Fuzzy and Parametric GARCH Methods. Turkish Journal of Forecasting, ۲۰۱۸. ۲(۱): p. ۱-۸.
[۱۳] Nabipour, M., et al., Deep learning for stock market prediction. Entropy, ۲۰۲۰. ۲۲(۸): p. ۸٤۰.
[۱٤] Nabipour, M., et al., Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. IEEE Access, ۲۰۲۰. ۸: p. ۱٥۰۱۹۹-۱٥۰۲۱۲.
[۱٥] Deng, Y., et al., Deep direct reinforcement learning for financial signal representation and trading. IEEE transactions on neural networks and learning systems, ۲۰۱٦. ۲۸(۳): p. ٦٥۳-٦٦٤.
[۱٦] Xiong, Z., et al., Practical deep reinforcement learning approach for stock trading. arXiv preprint arXiv:۱۸۱۱,۰۷٥۲۲, ۲۰۱۸.
[۱۷] Meng, T.L. and M. Khushi, Reinforcement learning in financial markets. Data, ۲۰۱۹. ٤(۳): p. ۱۱۰.
[۱۸] Li, Y., P. Ni, and V. Chang, Application of deep reinforcement learning in stock trading strategies and stock forecasting. Computing, ۲۰۲۰. ۱۰۲(٦): p. ۱۳۰٥-۱۳۲۲.

[۱۹] Carta, S., et al., Multi-DQN: An ensemble of Deep Q-learning agents for stock market forecasting. Expert systems with applications, ۲۰۲۱. ۱۶٤: p. ۱۱۳۸۲۰.
[۲۰] Shi, Y., et al., Stock trading rule discovery with double deep Q-network. Applied Soft Computing, ۲۰۲۱. ۱۰۷: p. ۱۰۷۳۲۰.
[۲۱] Théate, T. and D. Ernst, An application of deep reinforcement learning to algorithmic trading. Expert Systems with Applications, ۲۰۲۱. ۱۷۳: p. ۱۱٤٦۳۲.
[۲۲] Sun, T., D. Huang, and J. Yu, Market Making Strategy Optimization via Deep Reinforcement Learning. IEEE Access, ۲۰۲۲.
[۲۳] Jiang, W., Applications of deep learning in stock market prediction: recent progress. Expert Systems with Applications, ۲۰۲۱. ۱۸٤: p. ۱۱٥٥۳۷.
[۲٤] Amihud, Y., Illiquidity and stock returns: cross-section and time-series effects. Journal of financial markets, ۲۰۰۲. ٥(۱): p. ۳۱-٥٦.
[۲٥] Jansen, S., Machine Learning for Algorithmic Trading: Predictive Models to Extract Signals from Market and Alternative Data for Systematic Trading Strategies with Python. ۲۰۲۰: Packt Publishing Limited.
[۲٦] Mehrabanpour, M., A. Azar, and M. Shahrami Babkan, Stock price forecasting by presenting a hybrid model using principal component analysis and rough set theory. Modern Research in Decision Making, ۲۰۲۲. ۷(۲): p. ۱۳۷-۱٦۷.
[۲۷] Minh, D.L., et al., Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network. Ieee Access, ۲۰۱۸. ٦: p. ٥٥۳۹۲-٥٥٤۰٤.
[۲۸] Hessel, M., et al. Rainbow: Combining improvements in deep reinforcement learning. in Thirty-second AAAI conference on artificial intelligence. ۲۰۱۸.
[۲۹] Index. ۱٤۰۰; Available from: https://rahavard۳٦٥.com/.