نشریــه علمــی دانش مالی تحلیل اوراق بهادار
دوره ۱۷ / شماره ۶۱ / بهار ۱۴۰۳
صفحه ۱۹۶ تا ۲۱۵

طراحی سیستم معاملاتی الگوریتمیک بر پایه یادگیری تقویتی عمیق مورد مطالعاتی: بورس اوراق بهادار تهران

سعید کاظمیان حسین‌آبادی[۱]
سید محمدرضا داودی[*۲]
محمد مشهدی زاده[۳]
پارسا جوزی[۴]

تاریخ دریافت: ۱۴۰۱/۱۱/۰۴ تاریخ پذیرش: ۱۴۰۲/۰۳/۱۳

چکیده

امروزه معاملات الگوریتمی استفاده گسترده‌ای در مدیریت معاملات دارد. سبد گردانی الگوریتمی نوع جدیدی از این سامانه‌هاست که از طریق آن سبد گردان با استفاده از ابزارهای الگوریتمی به بالا بردن کیفیت سود و کاهش ریسک‌های سبد خود کمک می‌کند. هدف از پژوهش حاضر طراحی سیستم معاملاتی الگوریتمیک بر پایه یادگیری تقویتی عمیق به کمک شبکه عصبی است. در این رویکرد عامل یا معامله‌گر در فضای جستجو برای یافتن پاداش بیشتر که همان بازده بیشتر می‌باشد، به جستجو می‌پردازد. عامل معامله‌گر با سیگنال‌های تکنیکی شامل شاخص قدرت نسبی، نوسانگر تصادفی، نشانگر همگرایی-واگرایی و قیمت‌های کمینه، بیشینه، بسته شدن و باز شدن مواجه می‌شود. یادگیری تقویتی عمیق جدول تابع ارزش یا کیفیت Q را با یک شبکه عصبی جایگزین می‌کند. شبکه عصبی مذکور درنهایت با دریافت ورودی حالت، یکی از سه عمل فروش، خرید و نگهداری را پیشنهاد می‌کند. این پیشنهاد به‌صورت سه احتمال با مجموع یک می‌باشد و پیشنهاد با حداکثر احتمال مورد پیاده‌سازی قرار می‌گیرد. نتیجه پیاده‌سازی سیستم معاملاتی یادگیری تقویتی عمیق بر روی شاخص کل بورس اوراق بهادار تهران در بازه ۱۳۹۱ تا ۱۴۰۰ نشان می‌دهد که سیستم پژوهش در میانگین و شاخص همگراییی-واگرایی دارای تفاوت معناداری با سه سیستم دیگر داشت. همچنین نسبت شارپ سیستم پژوهش نسبت به سه مدل دیگر رشد حداقل ۱/۴ برابری را نشان داد.

کلمات کلیدی: شبکه‌های عصبی، یادگیری تقویتی، یادگیری تقویتی عمیق، نوسانگر تکنیکی.

[۱].کارشناسی ارشد مدیریت مالی، واحد دهاقان،دانشگاه آزاد اسلامی، دهاقان، ایران.saeedkazemian1365@yahoo.com
[۲]. استادیار، گروه مدیریت، واحد دهاقان، دانشگاه آزاد اسلامی، دهاقان، ایران(نویسنده مسئول). *smrdavoodi@ut.ac.ir*
[۳]استادیار،گروه مدیریت ، واحد مبارکه ،دانشگاه ازاد اسلامی ، مبارکه ،ایران*m.mashhadi@mau.ac.ir*
[۴]کارشناسی ارشد مدیریت مالی، واحد دهاقان، دانشگاه آزاد اسلامی، دهاقان،ایران. *Parsajozipersonal@gmail.com*

۱. مقدمه

سرمایه‌گذاری و انباشت سرمایه در تحول اقتصادی کشور نقش بسزایی داشته است. اهمیت این عامل و نقش مؤثر آن را می‌توان به‌وضوح در سیستم کشورهایی با نظام سرمایه‌داری مشاهده کرد. بدون شک بورس یکی از مناسب‌ترین جایگاه‌ها جهت جذب سرمایه‌های کوچک و استفاده از آن‌ها جهت رشد یک شرکت، در سطح کلان و نیز رشد شخصی فرد سرمایه‌گذار است (شمس و عطایی،۱۳۹۵). ازآنجایی‌که هدف و تعریف سرمایه‌گذاری، به تعویق انداختن مصرف جهت مصرف بیشتر و بهتر در آینده است؛ افراد با سرمایه‌گذاری انتظار دستیابی به سود مورد انتظار خود را دارند. بنابراین مهم‌ترین امر در این زمینه، خرید یک سهم به قیمت پایین و فروش آن به قیمت بالاتر است که این موضوع؛ به معنی تعیین مناسب موقعیت معاملاتی است. معاملات الگوریتمی در بازارهای مالی، به معنای استفاده از برنامه‌های کامپیوتری برای انتخاب سهام از بین صدها موقعیت موجود، ورود سفارش‌های خریدوفروش به سیستم‌های معاملاتی، کمک به اندازه‌گیری ریسک و بهینه‌سازی پرتفوی و کمک به مدیریت دارایی با بهره‌وری بالاتر می‌باشد. در این روش یک یا چند الگوریتم در انتخاب و اعمال این سفارش‌ها از جنبه‌های مختلف مانند زمان‌بندی، قیمت یا حجم معاملات آن تصمیم می‌گیرند. معاملات الگوریتمی در بسیاری از مواقع، بدون دخالت انسان انجام می‌شود(تیبات و دامین[1]،۲۰۲۰). امروزه معاملات الگوریتمی استفاده گسترده‌ای در مدیریت معاملات بانک‌های سرمایه‌گذاری (تامین سرمایه‌ها)، صندوق‌های بازنشستگی (و شرکت‌های سرمایه‌گذاری) و صندوق‌های سرمایه‌گذاری مشترک دارد. سبد گردانی الگوریتمی نوع جدیدی از این سامانه‌هاست که از طریق آن سبد گردان با استفاده از ابزارهای الگوریتمی به بالا بردن کیفیت سود و کاهش ریسک‌های سبد خود کمک می‌کند. همچنین با استفاده از این ابزار، سبد گردان قادر می‌شود برای تعداد زیادی مشتری سبد گردانی واقعا به‌صورت اختصاصی انجام دهد. هوش مصنوعی و به‌ویژه شبکه‌های عصبی از ابزارهای نوین در پیاده‌سازی معاملات الگوریتمی محسوب می‌شود(ونگ و همکاران[2]،۲۰۱۵). نتایج تحقیقات زیادی نشان داده‌اند که رفتار قیمتی سهام از نوع غیرخطی می‌باشد و سیستم‌های غیرخطی موفقیت بیشتری در استخراج ویژگی‌های بازار و درنهایت تعیین استراتژی‌های معاملاتی سودمند دارند. سیستم‌های غیرخطی دارای انعطاف بیشتری در مدل‌سازی داده‌ها و درک ساختار و روابط پیچیده بین متغیرها هستند. یکی از این مدل‌های پرکاربرد غیرخطی، شبکه‌های عصبی[3] می‌باشد. شبکه‌های عصبی سامانه‌های مدل‌سازی هستند که به‌صورت غیرخطی بر روی‌داده‌های ورودی تغییراتی را ایجاد می‌کنند و آن را به خروجی تبدیل می‌کنند. این سیستم‌ها از بدن انسان الهام گرفته‌اند و برای مدل‌سازی از مفاهیمی همچون نرون و توابع برانگیختگی استفاده می‌کند. یک شبکه عصبی برای کاربرد باید مورد آموزش قرار گیرد تا پارامترهای مجهول ساختاری آن مقداردهی بهینه شوند (کراماتیکی و همکاران[4]،۲۰۲۱). در مبحث یادگیری ماشین و شبکه‌های عصبی مفهوم مهمی بانام یادگیری عمیق قرارداد. یادگیری عمیق[5] شاخه‌ای از بحث یادگیری ماشین و هوش مصنوعی و مجموعه‌ای از الگوریتم‌هایی است که تلاش می‌کنند مفاهیم انتزاعی سطح بالا را با استفاده از یادگیری در سطوح و لایه‌های

[1] *Thibaut and Damien*
[2] *Wang et al*
[3] *Neural network*
[4] *Koratamaddi et al*
[5] *Deep Learning*

مختلف مدل کنند. مطالعات بالینی نشان می‌دهند، که ساختار مغز پستانداران از معماری شبکه‌های عصبی عمیق بهره می‌برد که در آن، مفاهیم انتزاعی در لایه‌های مختلف، به ترتیب از مفاهیم و ویژگی‌های ساده تا مفاهیم سطح بالا، در نواحی مختلف قشر مغز پردازش می‌شوند. ایده یادگیری عمیق با الهام از ساختار طبیعی مغز انسان و به کمک امکانات و فن‌آوری‌های جدید، توانسته است در بسیاری از حوزه‌های مربوط به هوش مصنوعی و یادگیری ماشین، موفقیت‌های چشم‌گیری را کسب کند .از مزایای یادگیری عمیق می‌توان به یادگیری خودکار ویژگی‌ها، یادگیری چندلایه ویژگی‌ها، دقت بالا در نتایج، قدرت تعمیم بالا و شناسایی داده‌های جدید، پشتیبانی گسترده سخت‌افزاری و نرم‌افزاری، پتانسیل ایجاد قابلیت‌ها و کاربردهای بیشتر در آینده اشاره کرد. یادگیری تقویتی یکی از گرایش‌های یادگیری ماشینی است که از روانشناسی رفتارگرایی الهام می‌گیرد. این روش بر رفتارهایی تمرکز دارد که ماشین باید برای بیشینه کردن پاداشش انجام دهد. این مسئله، با توجه به گستردگی‌اش، در زمینه‌های گوناگونی مانند نظریه بازی‌ها، نظریه کنترل، تحقیق در عملیات، نظریه اطلاعات، سامانه چندعامله، هوش ازدحامی، آمار، الگوریتم ژنتیک، بهینه‌سازی بر مبنای شبیه‌سازی بررسی و استفاده می‌شود(ژانگ و لی[1]،۲۰۲۲). یادگیری تقویتی با یادگیری با نظارت معمول دو تفاوت عمده دارد، نخست اینکه در آن زوج‌های صحیح ورودی و خروجی در کار نیست و رفتارهای ناکارآمد نیز از بیرون اصلاح نمی‌شوند، و دیگر آنکه تمرکز زیادی روی کارایی زنده وجود دارد که نیازمند پیدا کردن یک تعادل مناسب بین اکتشاف چیزهای جدید و بهره‌برداری از دانش اندوخته شده دارد. یادگیری تقویتی یک روش آموزش یادگیری ماشینی است که بر اساس پاداش دادن به رفتارهای دلخواه و/یا تنبیه رفتارهای نامطلوب است. به‌طورکلی، یک عامل یادگیری تقویتی قادر است محیط خود را درک و تفسیر کند، اقداماتی انجام دهد و از طریق آزمون‌وخطا یاد بگیرد. این اهداف بلندمدت به جلوگیری از توقف عامل در اهداف کمتر کمک می‌کند. باگذشت زمان، عامل یاد می‌گیرد که از منفی‌ها دوری کند و به دنبال مثبت باشد. این روش یادگیری در هوش مصنوعی به‌عنوان راهی برای هدایت یادگیری ماشینی بدون نظارت از طریق پاداش و جریمه پذیرفته‌شده است. معاملات الگوریتمی در بازار سرمایه به‌سرعت در حال گسترش می‌باشد. در دنیا درصد قابل‌توجهی از معاملات با این روش انجام می‌شود. امروز بخش عمده‌ای از فعالیت‌های بازار سرمایه از حالت دستی خارج‌شده و الکترونیکی انجام می‌شود و استفاده از فرصت‌های زودگذر و آربیتراژی نیازمند سرعت عمل بالا در شناسایی و اجرا است. طراحی سیستم‌های هوشمند جهت کسب چنین بازده‌هایی ضروری است. پژوهش حاضر نیز با استفاده از ابزارهای هوش مصنوعی نسبت به طراحی یک سیستم معاملات الگوریتمی اقدام می‌کند. تاکنون یادگیری تقویتی عمیق و سیستم معاملاتی مبتنی بر آن در بورس اوراق بهادار تهران مورد پیاده‌سازی و بررسی سودآوری قرار نگرفته است و لذا خلأ پژوهشی در تحقیق حاضر استفاده از این تکنیک به‌منظور پیاده‌سازی یک سیستم معاملاتی خودکار می‌باشد. بنابراین در پژوهش حاضر یک سیستم معاملاتی بر اساس رویکرد یادگیری تقویتی و به کمک شبکه‌های عصبی عمیق طراحی می‌شود.

[1] *Zhang and Lei*

به‌منظور واکاوی مفهوم سیستم معاملاتی الگوریتمیک بر پایه یادگیری تقویتی عمیق سؤالات زیر مطرح می‌شود که در این پژوهش به آن پاسخ داده می‌شود:

۱-چگونه می‌توان به کمک یادگیری تقویتی عمیق یک سیستم معاملاتی طراحی کرد؟

۲-سودآوری سیستم معاملاتی طراحی‌شده چگونه است؟

در مبانی نظری به مطالعه اجمالی در مورد پیشینه پژوهش با تأکید بر سیستم معاملاتی الگوریتمیک بر پایه یادگیری تقویتی عمیق به کمک شبکه عصبی پرداخته می‌شود.

۲. مبانی نظری و مروری بر پیشینه پژوهش

پیشرفت‌های اخیر در حوزه فناوری‌های کامپیوتری، گسترش به‌کارگیری فنّاوری‌های معاملات الگوریتمی و خودکار، توسعه استراتژی‌های معاملاتی و به‌تبع آن سرعت‌بالای معاملات، منجر به افزایش پیچیدگی‌های سرمایه‌گذاری در بازار سهام شده‌اند(ژانگ و همکاران[۱]،۲۰۱۶). معاملات الگوریتمی نوعی از معاملات خودکار بوده که شامل برنامه‌های کامپیوتری برای ارسال سفارشات همراه با الگوریتم‌های تصمیم‌گیری هستند که این الگوریتم‌ها خود بر اساس پارامترهای منحصربه‌فرد سفارش مانند زمان، قیمت و یا مقدار سفارش می‌باشند. در بازارهای مالی الکترونیکی، معاملات الگوریتمی به معنای استفاده از برنامه‌های کامپیوتری برای ورود سفارش‌های معاملاتی است که سیگنال‌های معاملاتی آن توسط بخشی دیگر از سیستم تولیدشده است(رستگار و صداقتی پور،۱۳۹۷).

یادگیری تقویتی می‌تواند در یک موقعیت عمل کند تا زمانی که بتوان پاداش روشنی را اعمال کرد. در مدیریت منابع سازمانی، الگوریتم‌های یادگیری تقویتی می‌توانند منابع محدودی را به وظایف مختلف اختصاص دهند تا زمانی که هدف کلی وجود داشته باشد. هدف در این شرایط صرفه‌جویی در زمان یا حفظ منابع خواهد بود. در رباتیک، یادگیری تقویتی به تست‌های محدود راه پیداکرده است. این نوع یادگیری ماشینی می‌تواند به روبات‌ها توانایی یادگیری وظایفی را که معلم انسانی نمی‌تواند نشان دهد، انطباق مهارت‌های آموخته‌شده با یک کار جدید یا دستیابی به بهینه‌سازی علیرغم کمبود فرمول‌بندی تحلیلی در دسترس، ارائه دهد. یادگیری تقویتی همچنین در تحقیقات عملیات، نظریه اطلاعات، نظریه بازی، نظریه کنترل، بهینه‌سازی مبتنی بر شبیه‌سازی، سیستم‌های چندعاملی، هوش ازدحام، آمار و الگوریتم‌های ژنتیک استفاده می‌شود (یو و یان[۲]،۲۰۲۰).

صیادی نژاد و همکاران(۱۴۰۲) در پژوهش خود بیان کردند با افزایش محبوبیت و فراگیر شدن رمزارزها، ایجاد و توسعه روش‌های پیش‌بینی حرکت‌های قیمتی در این حوزه، توجهات زیادی را به خود جلب کرده است. دراین‌بین مدل‌های یادگیری عمیق (DL) با ساختارهایی مانند حافظه طولانی کوتاه‌مدت $(LSTM)$و شبکه عصبی کانولوشنی (CNN) پیشرفت‌هایی در تحلیل این نوع از داده‌ها ایجاد کرده است. یکی دیگر از رویکردهایی که می‌تواند در تحلیل قیمتی بازار رمزارزها کارا باشد تجزیه سیگنال‌های از طریق الگوریتم‌هایی مانند تجزیه مد تجربی یکپارچه کامل $(CEEMD)$ می‌باشد. با توجه به اهمیت مقوله

[۱] *Zhang and Lei*

[۲] Yu and Yan

پیش‌بینی در بازار رمز ارزها، در این تحقیق با ترکیب مدل‌های یادگیری عمیق و روش تجزیه مد تجربی یکپارچه کامل(*CEEMD*) ، مدل هیبریدی *CEEMD- DL(LSTM)* به‌منظور پیش-بینی بازدهی قیمتی رمز ارز بیت‌کوین (به‌عنوان محبوب-ترین رمز ارز) مورداستفاده قرار گرفته‌است. در این راستا از داده‌های روزانه قیمتی بیت‌کوین در دوره زمانی۲۰۱۳/۰۱/۰۱ - ۲۰۲۲/۰۵/۲۸ استفاده گردید و نتایج به‌دست‌آمده با نتایج مدل‌های رقیب بر اساس معیارهای سنجش کارایی مقایسه شد. بر اساس نتایج به‌دست‌آمده، استفاده از مدل معرفی‌شده *(CEEMD-DL(LSTM))*، کارایی و دقت پیش‌بینی‌های بازدهی رمزارز بیت‌کوین را افزایش داده‌است. بر همین اساس کاربرد این مدل به‌منظور پیش‌بینی در این حوزه پیشنهاد می‌گردد.

اسفندیار و همکاران(۱۴۰۱) در پژوهش خود بیان کردند که این مقاله کاربرد معاملات الگوریتمی با تمرکز بر رویکرد یادگیری تقویتی برای بهینه‌سازی پرتفوی سهام‌های منتخب را بررسی می‌کند. این پژوهش از حیث هدف، کاربردی و ازنظر نوع داده، کمّی و از لحاظ روش، توصیفی - اکتشافی و از منظر طرح تحقیق، پس‌رویدادی است. جامعه آماری این پژوهش، ۶۷۲ شرکت بورسی است که از این تعداد، داده‌های پنج شرکت (نمونه آماری) طی دوره زمانی ۱۳۹۶-۱۴۰۰ بررسی شده است. یافته‌های تحقیق در دوره‌های صعودی و نزولی بازار نشان داد که رویکرد یادگیری تقویتی در بازارهای صعودی و نزولی به‌صورت معناداری بر رویکرد خرید و نگهداری برتری دارد و عملکرد بهتری ارائه داده است و نتایج با عملکرد الگوریتم‌ها در بازارهای بورس سازگار است. نتایج آشکار کرد که از دیدگاه سودآوری، رویکرد یادگیری تقویتی نسبت به رهیافت خرید و نگه‌داری، عملکرد بهتر و موثرتری داشته است؛ بنابراین، به‌کارگیری روش یادگیری تقویتی پیشنهاد می‌شود. دسارنج و همکاران(۱۳۹۹) در پژوهشی خود با هدف طراحی سیستم معاملات تکنیکی سهام با استفاده از مدل ترکیبی شبکه عصبی *MLP* و الگوریتم‌های تکاملی نشان دادند که مدل‌های ترکیبی *MLP* و الگوریتم‌های تکاملی عملکرد بهتر و معناداری نسبت به روش خرید و نگهداری و مدل *MLP-BP* داشته است و مدل *MLP_PSO*بازدهی بیش‌تری نسبت به سایر مدل‌ها کسب کرده است. یافتیان و رستگار(۱۳۹۹) در پژوهش خود با هدف طراحی یک سیستم معاملاتی خودکار با استفاده از شبکه عصبی پیچشی نشان دادند که برای بررسی بازدهی و ریسک سیستم ارائه‌شده، یک روش برای خریدوفروش بر اساس نتایج مدل در زمان گذشته معرفی‌شده است که در ۸۰٪ موارد، این روش بازدهی بیشتری نسبت به استراتژی مرسوم خرید و نگهداری کسب کرده است. همچنین همواره ازنظر معیارهای ریسک انحراف معیار و بیشترین افت بهتر عمل می‌کند. همچنین، نتایج نشان‌دهنده‌ی تأثیر زیاد کارمزد معاملات بورس اوراق بهادار تهران بر روی بازدهی مدل است. به‌گونه‌ای که مدل چند برابر سود کسب‌شده را برای پرداخت کارمزد از دست می‌دهد. خنجر پناه و همکاران(۱۳۹۷) در پژوهش خود با هدف کاربرد روش تکنیکال برای پیش‌بینی قیمت سهام: رویکرد مدل‌های احتمال غیرخطی و شبکه‌های عصبی مصنوعی نشان دادند که در آزمون نا پارامتری برابری نسبت‌ها، ازلحاظ آماری مدل‌های ارائه‌شده تفاوت معناداری باهم نداشته-اند، اما معیارهای سنجش خطا بیان می‌کند که مدل پروبیت، خطای کمتری در پیش‌بینی سهام در بازار بورس تهران دارد.

جین[1](۲۰۲۳)در مقاله خود یک الگوریتم نوآورانه را برای حل مسئله سبد بهینه در فعالیت‌های معاملاتی بازار سهام پیشنهاد می‌کند. استراتژی جدید معاملات پرتفوی پژوهش از سه ویژگی برای پیشی گرفتن از سایر استراتژی‌های معیار در یک محیط بازار واقعی استفاده می‌کند. ابتدا، یک مدل بهینه‌سازی پورتفولیو میانگین-واریانس را پیشنهاد می‌کند که راه‌حل آن بر اساس معماری بازیگر- منتقد است. برخلاف ادبیات موجود که انتظار بازده تجمعی را می‌آموزد، ماژول منتقد توزیع بازده تجمعی را با رگرسیون کمی می‌آموزد و ماژول بازیگر وزن بهینه پورتفولیو را با حداکثر کردن تابع هدف مدل بهینه‌سازی خروجی می‌دهد. ثانیه، از یک تابع تبدیل خطی برای تحقق فروش کوتاه‌مدت استفاده می‌کند، سوم، یک روش چند فرآیندی به نام *Ape-x* برای تسریع سرعت آموزش یادگیری تقویتی عمیق استفاده شد. برای اعتبار سنجی رویکرد پیشنهادی خود، این پژوهش برای دو نمونه کارنامه نماینده بک تست انجام می‌دهد و مشاهده می‌کند که مدل پیشنهادی در این کار برتر از استراتژی‌های معیار است.

لی[2] و همکاران(۲۰۲۲) یک مدل یادگیری تقویتی جدید را برای پیاده‌سازی معاملات سهام پیشنهاد کردند که بازار سهام را از طریق داده‌های سهام، شاخص‌های فنی و نمودارهای شمعی تحلیل می‌کند و استراتژی‌های معاملاتی پویا را می‌آموزد. نتایج نشان داد که استراتژی معاملاتی می‌تواند در مقایسه با سایر استراتژی‌های تجاری سود بیشتری به دست آورد. ژانگ و لی[3](۲۰۲۲) از یادگیری تقویتی عمیق به‌منظور پیش‌بینی قیمت سهام استفاده کردند. آن‌ها هدف خود را افزایش دقت و سرعت پیش‌بینی نوسانات قیمت سهام با استفاده از مدل یادگیری تقویتی عمیق روش استراتژی گرادیان یا *PG*[4] قراردادند. آزمایش‌های صورت گرفته نشان می‌دهند که دقت پیش‌بینی الگوریتم جدید و سرعت هم‌گرایی پاداش به‌طور قابل‌توجهی بالاتر از الگوریتم یادگیری تقویتی عمیق سنتی است. درنتیجه، الگوریتم جدید با شرایط نوسان بازار سازگارتر است. کراماتیکی و همکاران[5](۲۰۲۱) یک رویکرد یادگیری تقویتی عمیق جدید را برای آموزش مؤثر یک معامله‌گر خودکار هوشمند پیشنهاد می‌کنند، که نه‌تنها از داده‌های تاریخی قیمت سهام استفاده می‌کند، بلکه احساسات بازار را برای یک سبد سهام متشکل از شرکت‌های داو جونز درک می‌کند. نشان داده می‌شود که رویکرد پژوهش در مقایسه با خطوط پایه موجود در معیارهای استانداردشده مانند نسبت شارپ و بازده سرمایه‌گذاری سالانه قوی‌تر است. تیبات و دامین[6](۲۰۲۰)در پژوهشی یک رویکرد نوآورانه مبتنی بر یادگیری تقویت عمیق (*DRL*) برای حل مشکل تعیین موقعیت بهینه(خرید، فروش و نگهداری) در یک معامله الگوریتمیک در طول فعالیت تجاری در بازارهای سهام را ارائه می‌دهند. استراتژی تجاری جدید *DRL* پیشنهادشده به دنبال بیشینه‌سازی نسبت شارپ در طیف وسیعی از بازارهای سهام می‌باشد. یانگ و ژای[7](۲۰۲۰) برای پیش‌بینی قیمت سهام از یک روش ترکیبی در یادگیری عمیق استفاده کردند. ابتدا به کمک شبکه‌های عصبی پیچشی به استخراج ویژگی از ورودی‌ها شامل یک تنسور سه بعدی از ورودی‌های تکنیکی، قیمت و شاخص‌ها پرداختند و در ادامه به کمک

[1] *Jin*
[2] *Li et al*
[3] *Zhang and Lei*
[4] *policy gradient strategy*
[5] *Koratamaddi et al*
[6] *Thibaut، Damien*
[7] *Yang and Zhai*

شبکه‌های حافظه طولانی کوتاه‌مدت به پیش‌بینی پرداختند. نتیجه پژوهش بر روی ۵ شاخص ازجمله $S\&P$ نشان می‌دهد که دقت روش ترکیبی از روش‌های دیگر چون بردار پشتیبان و ترکیب شبکه‌های عصبی با تحلیل مولفه اصلی بیشتر است. نبی پور و همکاران(۲۰۲۰) برای پیش‌بینی قیمت سهام از الگوریتم‌های مختلفی در حوزه هوش مصنوعی استفاده کردند. این الگوریتم‌ها شامل شبکه‌های عصبی،درخت‌های تصمیم، روش‌های جمعی، جنگل تصادفی و یادگیری حافظه طولانی کوتاه‌مدت می‌باشد. نتیجه پژوهش بر روی‌داده‌های تاریخی ۱۰ ساله از بورس اوراق بهادار تهران بر روی ۵ شاخص با ورودی‌های تکنیکی نشان می‌دهد که شبکه عصبی بازگشتی دارای بالاترین دقت می‌باشد. ژو و کسلج[1](۲۰۱۹) برای پیش‌بینی شاخص بازار از الگوریتم یادگیری عمیق شبکه‌های حافظه طولانی کوتاه‌مدت استفاده کردند. ورودی‌های سیستم پیش‌بینی شامل داده‌های تاریخی، نما گرهای تکنیکی و احساسات سهامداران می‌باشد. برای سنجش احساسات سهامداران از توییت منتشرشده در توییتر استفاده‌شده است. نتیجه پژوهش نشان می‌دهد که سیستم پیش‌بینی با در نظر گرفتن احساسات سرمایه‌گذاران دارای دقت بالاتری نسبت به سیستم عمیقی می‌باشد که این ورودی را ندارد. همچنین نتیجه پژوهش نشان می‌دهد که توییت افراد دارای تعداد دنباله کنندگان بالا نقش مهمی در تبیین خروجی دارد.

۳. روش‌شناسی پژوهش

هدف پژوهش حاضر طراحی و ارزیابی عملکرد یک سیستم معاملاتی مبتنی بر یادگیری تقویتی عمیق می‌باشد. روش تحقیق در این پژوهش توصیفی-تحلیلی می‌باشد. این تحقیق بر اساس هدف یک تحقیق کاربردی است. با توجه به هدف، پژوهش حاضر در دسته کلی طراحی سیستم‌های معاملاتی مبتنی بر هوش مصنوعی جای می‌گیرد. ورودی سیستم معاملاتی شامل قیمت‌های باز و بسته شدن، بیشینه و کمینه قیمت به همراه شاخص‌های تکنیکی قدرت نسبی، شاخص تصادفی و همگرایی-واگرایی می‌باشد. خروجی سیستم معاملاتی مبتنی بر یادگیری تقویتی عمیق نیز شامل اتخاذ یکی از سه موقعیت خرید، فروش و نگهداری می‌باشد. این سیستم به‌جای استفاده از جدول q از یک شبکه عصبی عمیق برای رکورد نتایج حاصل از اقدامات استفاده می‌کند. پس از آموزش سیستم معاملاتی بر روی‌داده‌های آموزشی، دقت مدل پیش‌بینی بر روی‌داده‌های تست مورد ارزیابی قرار می‌گیرد. برای تجزیه‌وتحلیل اطلاعات از کد نویسی با زبان پایتون و پکیج تنسورفلو و همچنین زبان برنامه‌نویسی متلب استفاده خواهد شد. مدل پژوهش برگرفته از تحقیق لی و همکاران(۲۰۲۲) می‌باشد. جامعه آماری پژوهش شاخص کل بورس اوراق بهادار تهران به‌عنوان نماینده بازار می‌باشد و نمونه آماری به‌صورت زمانی شامل شاخص کل بورس اوراق بهادار تهران در بازه شروع سال ۱۳۹۱ تا پایان تیر ۱۴۰۱ می‌باشد.

۱،۳. تبیین مدل پژوهش

در این بخش به طراحی سیستم معاملاتی بر پایه یادگیری تقویتی عمیق پرداخته می‌شود. اما در ابتدا به معرفی سه نوسان‌نمای تکنیکی می‌پردازیم که بخشی از ورودی‌های سیستم معاملاتی می‌باشد. یکی از این نوسانماهای معروف و پرکاربرد، قدرت نسبی یا RSI

[1] *Xu and Keselj*

می‌باشد. در تنظیمات این اندیکاتور عموماً از تنظیم ۱۴ روزه استفاده می‌شود. فرمول محاسبه این نوسان‌نما به‌صورت زیر می‌باشد:

(۱)

$$RSI = (100 - \frac{100}{1+RS})$$

$$RS = Average\,Gain \,/\, Average\,Loss$$

صورت کسرRS برابر میانگین تمام افزایش قیمت‌ها و مخرج برابر مجموع قدر مطلق تمام کاهش قیمت‌ها در ۱۴ روز گذشته می‌باشد. ازآنجاکهRS همواره عددی مثبت است به‌راحتی می‌توان دید کهRSI عددی مابین صفر و یک خواهد شد. غالباً و به‌صورت تجربی، دو سطح کلیدی ۳۰ و ۷۰ را به ترتیب به‌عنوان سطوح اشباع فروش(فزونی پیروزی بر شکست) و اشباع خرید(فزونی شکست بر پیروزی) می‌نامند.

نوسان‌نمای استوکستیک یا SO نیز دارای نوسانی در بازه ۰ تا ۱۰۰ می‌باشد. در محاسبه این نوسانگر دو مقدار %Kو%D به‌صورت رابطه (۲) محاسبه می‌شوند.

(۲)

$$\%K = \frac{(Current\ Close - Lowest\ Low)}{(Highest\ High - Lowest\ Low)} \times 100$$

که $\%D$ میانگین متحرک ساده سه‌روزه $\%K$ می‌باشد. در محاسبات بالا کمترین و بیشترین قیمت همواره در یک دوره زمانی مشخص محاسبه می‌شوند که همگرایی ساخته می‌شود که از آن برای تولید سیگنال خریدوفروش استفاده می‌شود. بنابراین درصورتی‌که $\{MA_n(t)\},\{MA_m(t)\}$ دو سری میانگین متحرک به‌طورمعمول این دوره ۱۴ روزه می‌باشد. مفهوم این نوسانگر به این صورت است که همواره بزرگی قیمت نسبت به فاصله بیشترین و کمترین قیمت در دوره زمانی موردبررسی سنجیده می‌شود . بنابراین هر چه %Kبزرگتر شود و نزدیکی این عدد به ۱۰۰ نشان از پتانسیل بیشتر سهم برای ریزش قیمت دارد. از طرفی هر چه%K کمتر شود و نزدیکی این عدد به ۰ نشان از پتانسیل بیشتر سهم برای ریزش قیمت دارد. غالباً برای این نوسانگر نیز دو حد آستانه خرید افراطی و فروش افراطی در نظر گرفته می‌شود که معمولاً این دو خط به ترتیب خطوط ۸۰ و ۲۰ می‌باشند. میانگین متحرک‌ها گروهی از عملگرهای هموار کننده می‌باشند که به‌جای سیگنال اصلی قیمت در هرلحظه، میانگین قیمت‌ها را در یک‌فاصله زمانی مشخص محاسبه می‌کنند و به‌جای آن قرار می‌دهند. میانگین متحرک‌ها با هموارسازی نمودار قیمت، روند حرکت آن را به‌صورت عینی‌تری به نمایش می‌گذارند. درصورتی‌که $\{p(t)\}_{t=1}^{T}$ سری قیمت یک سهام باشد، سری میانگین متحرک ساده به طول n برابر است با رابطه (۳):

$$MA_n(t) = \frac{p(t) + p(t-1) + p(t-n+1)}{n}$$

درصورتی‌که n عدد بزرگی باشد، میانگین متحرک روند تاریخی را بهتر نشان می‌دهد و هرچه n کوچک‌تر باشد حساسیت آن به داده‌های جدیدتر بیشتر می‌شود. با استفاده از ترکیب میانگین متحرک‌ها نوسان‌نماهای جدیدی به نام نوسان‌نمای واگرایی- باشند که $n > m$، آنگاه درصورتی که $MA_n(t) < MA_m(t)$ سیگنالی برای خرید دریافت

می‌شـود و اگر $MA_n(t) > MA_m(t)$ سـیگنالی برای فروش دریافت می‌شود.

پس از معرفی سه نوسانگر به بررسی مدل یادگیری تقویتی عمیق پرداخته می‌شود. یکی از اجزای تشکیل‌دهنده یادگیری تقویتی مجموعه حالات می‌باشد. در سیستم معاملاتی پژوهش با توجه به پیوستگی قیمت و شاخص‌های تکنیکی، متغیرهای حالت پیوسته می‌باشند و لذا یادگیری Q بر اساس جدول قابلیت کاربرد ندارد و به‌جای آن از یادگیری تقویتی عمیق استفاده می‌شود. بردار حالت عبارت است از

$$s_t = \begin{bmatrix} p_t \\ o_t \\ c_t \\ h_t \\ l_t \\ rsi_t \\ so_t \\ macd_t \end{bmatrix} \quad (۴)$$

که در آن

جدول۱: معرفی اجزای بردار حالت در یادگیری تقویتی عمیق

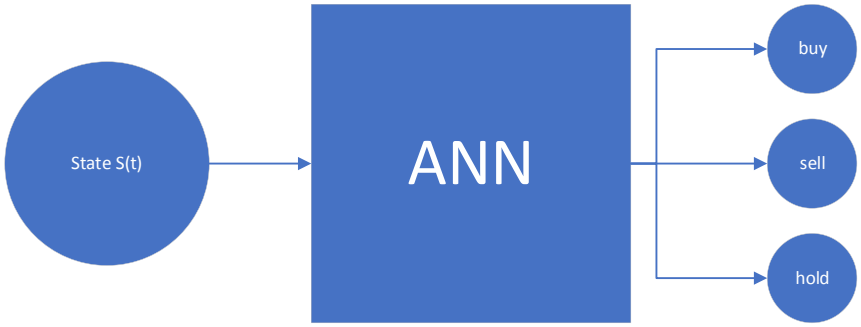
تعریف	متغیر
قیمت روزانه	p_t
قیمت باز شدن	o_t
قیمت بسته شدن	c_t
بیشترین قیمت روزانه	h_t
کمترین قیمت روزانه	l_t
شاخص قدرت نسبی	rsi_t
شاخص تصادفی	so_t
شاخص همگرایی-واگرایی	$macd_t$

پس از معرفی مجموعه حالات در یادگیری تقویتی، به معرفی مجموعه اقدامات پرداخته می‌شود. با توجه به ماهیت سیستم معاملاتی در بورس اوراق بهادار، سه عمل یا اقدام تعریف می‌گردد.

جدول۲: معرفی اجزای بردار اقدام

توصیف	عمل
فروش به‌شرط اینکه دارایی قبلاً خریداری‌شده باشد(فاقد فروش استقراضی).	*sell*
خرید	*buy*
نگهداری	*hold*

یادگیری تقویتی عمیق جدول تابع ارزش یا کیفیت Q را با یک شبکه عصبی جایگزین می‌کند. شبکه عصبی مذکور درنهایت با دریافت ورودی حالت، یکی از سه عمل فروش، خرید و نگهداری را پیشنهاد می‌کند. این پیشنهاد به‌صورت سه احتمال با مجموع یک می‌باشد و پیشنهاد با حداکثر احتمال مورد پیاده‌سازی قرار می‌گیرد.



نمودار۱: شبکه عصبی عمیق مورداستفاده در یادگیری تقویتی

پاداش یک عمل در سیستم معاملاتی پژوهش، بازده ناشی از معامله می‌باشد. این پاداش در زمان فروش دارایی محاسبه خواهد شد و برای اقدامات نگهداری و خرید، پاداشی در نظر گرفته نمی‌شود. برای آموزش شبکه عصبی از معادله بلمن[1] استفاده می‌شود. $Q : S \times A \rightarrow \square$ قبل از شروع یادگیری Q به یک مقدار ثابت احتمالاً دلخواه (انتخاب‌شده توسط برنامه‌نویس) مقداردهی اولیه می‌شود. سپس، در هر بار t، عامل یک عمل را انتخاب می‌کند a_t پاداشی را مشاهده می‌کند r_t وارد حالت جدیدی می‌شود S_{t+1} که ممکن است به حالت قبلی S_t و عملکرد انتخاب‌شده بستگی داشته باشد، به‌روز می‌شود. هسته اصلی الگوریتم معادله بلمن به‌عنوان یک به‌روزرسانی تکرار مقدار ساده است که از میانگین وزنی مقدار قدیمی و اطلاعات جدید استفاده می‌کند:

[1] *Bellman equation*

$$Q^{new}(s_t,a_t) \leftarrow \underbrace{Q(s_t,a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \overbrace{\Big(\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1},a)}_{\text{estimate of optimal future value}} - \underbrace{Q(s_t,a_t)}_{\text{old value}} \Big)}^{\text{temporal difference}} \quad (۵)$$

$$\text{new value (temporal difference target)}$$

جایی که r_t پاداش دریافتی هنگام انتقال از حالت s_t به حالت s_{t+1} و α نرخ یادگیری است که $0 < \alpha \leq 1$.

توجه داشته باشید که $Q^{new}(s_t,a_t)$ مجموع سه عامل است:

$(1-\alpha)Q(s_t,a_t)$: مقدار فعلی وزن‌دار شده توسط نرخ یادگیری. مقادیر نرخ یادگیری نزدیک به ۱ باعث می‌شود تغییرات در Q سریع‌تر شود.

αr_t: پاداش $r_t = r(s_t,a_t)$ برای دریافت اینکه آیا عمل a_t در حالت s_t (وزن دهی شده بر اساس میزان یادگیری) انجام می‌شود یا خیر.

$\alpha\gamma \max_a Q(s_{t+1},a)$: حداکثر پاداشی که می‌توان از حالت s_{t+1} به دست آورد (وزن‌بندی شده بر اساس نرخ یادگیری و ضریب تخفیف)

یک اجرا از الگوریتم زمانی به پایان می رسد که حالت s_{t+1} حالت نهایی یا پایانی باشد. اگر ضریب تخفیف کمتر از ۱ باشد، مقادیر کنش محدود هستند حتی اگر مشکل بتواند حلقه‌های بی‌نهایت داشته باشد. برای همه حالت‌های نهایی s_f، $Q(s_f,a)$ هرگز به‌روزرسانی نمی‌شود، اما روی مقدار پاداش r برای مشاهده s_f، تنظیم می‌شود. در بیشتر موارد، $Q(s_f,a)$ را می‌توان برابر با صفر در نظر گرفت.

ضریب یادگیری نشان می‌دهد که تا چه حد اطلاعات تازه به‌دست‌آمده بر اطلاعات قدیمی غلبه می‌کند. ضریب ۰ باعث می‌شود عامل چیزی یاد نگیرد (به‌طور انحصاری از دانش قبلی استفاده می‌کند)، درحالی‌که ضریب ۱ باعث می‌شود عامل فقط جدیدترین اطلاعات را در نظر بگیرد (نادیده گرفتن دانش قبلی برای کشف احتمالات). در محیط‌های کاملاً قطعی، نرخ یادگیری $\alpha_t = 1$ بهینه است. هنگامی‌که مسئله تصادفی است، الگوریتم تحت برخی شرایط فنی روی نرخ یادگیری همگرا می‌شود که نیاز به کاهش آن به صفر دارد. در عمل، اغلب از یک نرخ یادگیری ثابت مانند $\alpha_t = 0.1$ استفاده می‌شود،

ضریب تخفیف γ اهمیت پاداش‌های آینده را تعیین می‌کند. ضریب ۰ عامل را تنها با در نظر گرفتن پاداش‌های فعلی، یعنی r_t (در قانون به‌روزرسانی بالا) "کوته بین" (یا کوته بین) می‌کند، درحالی‌که ضریب نزدیک به ۱ باعث می‌شود آن را وادار کنید تا برای یک پاداش بلندمدت تلاش کند. اگر ضریب تخفیف برابر یا بیشتر از ۱ باشد، مقادیر عمل ممکن است متفاوت باشد. برای $\gamma = 1$، بدون حالت پایانی، یا اگر عامل هرگز به یک حالت نرسد، تمام تاریخچه‌های محیطی بی‌نهایت طولانی می‌شوند و ابزارهای کاربردی با پاداش‌های اضافی

و تخفیف ناپذیر عموماً بی‌نهایت می‌شوند. حتی با یک ضریب تخفیف فقط کمی کمتر از ۱، یادگیری تابع Q منجر به انتشار خطاها و ناپایداری‌ها می‌شود، زمانی که تابع مقدار با یک شبکه عصبی مصنوعی تقریب زده شود. در این صورت، شروع با ضریب تخفیف کمتر و افزایش آن به سمت ارزش نهایی، یادگیری را تسریع می‌کند. برای طراحی سیسم معاملاتی بر پایه یادگیری تقویتی عمیق از دو شبکه عصبی استفاده می‌شود. شبکه اول برای محاسبه مقادیر Q در سمت راست معادله بلمن استفاده می‌شود. این شبکه عصبی آموزش نمی‌بیند و فقط برای مقصد محاسبه احتمالات استفاده می‌شود. پس از محاسبه سمت راست معادله بلمن برای یک ورودی، از یک شبکه عصبی (که همان سیستم معاملاتی می‌باشد) برای آموزش استفاده می‌شود. بدین‌صورت که ورودی آن حالت و خروجی آن مقدار معادله بلمن می‌باشد که توسط شبکه عصبی اول آموزش‌دیده است. در فواصل زمانی منظم شبکه عصبی دوم، جای شبکه عصبی اول را می‌گیرد و شبکه عصبی دوم از نو آموزش می‌بیند.

۴. یافته‌های پژوهش

مدل تحقیق از دو شبکه عصبی عمیق برای پیاده‌سازی جدول Q استفاده می‌کند. شبکه عصبی اول آموزش نمی‌بیند و فقط برای مقصد محاسبه احتمالات استفاده می‌شود. پس از محاسبه سمت راست معادله بلمن برای یک ورودی، از یک شبکه عصبی (که همان سیستم معاملاتی می‌باشد) برای آموزش استفاده می‌شود. بدین‌صورت که ورودی آن حالت و خروجی آن مقدار معادله بلمن می‌باشد که توسط شبکه عصبی اول آموزش‌دیده است. در فواصل زمانی منظم شبکه عصبی دوم، جای شبکه عصبی اول را می‌گیرد و شبکه عصبی دوم از نو آموزش می‌بیند. پارامترهای مورداستفاده در معادله بلمن در جدول۳ ارائه‌شده است.

جدول۳: تنظیمات مورداستفاده در معادله بلمن

مقدار	پارامتر
۰/۱	ضریب یادگیری α
۰/۹۵	ضریب تحقیق یا تنزیل γ

در فرآیند یادگیری تقویتی، در گام‌های نخست به جستجوی تصادفی اهمیت بیشتری داشته می‌شود و کم‌کم با جمع‌آوری اطلاعات از محیط و یادگیری بیشتر شبکه عصبی عمیق، به بهره‌گیری از اطلاعات کسب‌شده اهمیت داده می‌شود. برای این منظور پارامتر اپسیلون تعریف می‌شود که این ضریب را شبیه‌سازی می‌کند و به‌تدریج باگذشت زمان از مقدار آن کم می‌شود. تنظیمات مورداستفاده در تعادل بین جستجوی تصادفی و بهره‌گیری از اطلاعات موجود در جدول۴ ارائه‌شده است.

مشخصات شبکه‌های عصبی عمیق مورداستفاده در یادگیری تقویتی عمیق در جدول۵ ارائه‌شده است.

جدول۴: تنظیمات جستجو

مقدار	پارامتر
۰/۹۵	ضریب توازن اولیه جستجوی تصادفی و بهره‌گیری ε

مقدار کاهش ضریب توازن در هر اجرا	-۰/۰۰۱

جدول۵: مشخصات شبکه‌های عصبی عمیق مورداستفاده در یادگیری تقویتی عمیق

پارامتر	مقدار
نوع	شبکه عصبی عمیق پرسپترون
تعداد لایه	۳
تابع فعالساز لایه‌های میانی	Relu
تابع فعالساز لایه آخر	سافت مکس
تعداد نرون لایه ورودی	۶
تعداد نرون لایه خروجی	۳
الگوریتم بهینه ساز	آدام[1]

در هر ۱۰ تکرار، جای شبکه آموزش‌دیده جای شبکه پیش‌بینی را می‌گیرد. تعداد داده‌های بازده روزانه شاخص کل در بازه پژوهش ۲۶۵۴ داده می‌باشد که از ۱۶۵۴ داده برای آموزش مدل پژوهش استفاده گردید و ۱۰۰۰ داده نیز برای آزمون و بررسی عملکرد مدل پژوهش در داده‌های برون نمونه‌ای استفاده گردید. برای آموزش مدل پژوهش از ۱۰۰۰ تکرار استفاده گردید.

پس از آموزش سیستم معاملاتی پژوهش بر روی ۱۶۵۴ داده آموزشی روزانه، این سیستم بر روی ۱۰۰۰داده تست مورد ارزیابی قرار گرفت. بدین‌صورت زمانی که خروجی شبکه عصبی برابر خرید می‌باشد، شاخص کل خریداری و یک موقعیت معاملاتی باز می‌شود و در ادامه زمانی که خروجی شبکه عصبی عمیق برابر فروش می‌باشد، شاخص خریداری‌شده، فروخته و موقعیت معاملاتی بسته می‌شود و بازده حاصل‌شده به‌عنوان یک نمونه از عملکرد سودآوری سیستم معاملاتی پژوهش ذخیره می‌گردد. مشخصات آماری موقعیت‌های معاملاتی حاصل‌شده در جدول۶ ارائه‌شده است.

جدول۶: مشخصات آماری معادل بازده روزانه سیستم معاملاتی پژوهش

سیستم معاملاتی	مقدار
تعداد موقعیت معاملاتی	۱۳

[1] Adam

۰/۰۱۰۴	بیشینه بازده موقعیت
-۰/۰۰۱۵	کمینه بازده موقعیت
۲۱/۵۳۸۵	متوسط طول زمانی موقعیت
۰/۰۰۱۸	متوسط معادل بازده روزانه
۰/۰۰۳۷	انحراف معیار معادل بازده روزانه
۰/۴۷۹۳	نسبت شارپ

بر اساس اطلاعات حاصل‌شده، ۱۳ موقعیت معاملاتی توسط سیستم پژوهش شناسایی گردید که متوسط بازده روزانه‌ای برابر ۰/۰۰۱۸ را تولید می‌کند که این مقدار برابر متوسط بازده سالیانه‌ای برابر ۰,۹۱ می‌باشد. متوسط زمان هر معامله تقریباً برابر ۲۲ روز می‌باشد. نسبت شارپ سیستم پژوهش که از تقسیم بازده بر انحراف معیار محاسبه می‌شود برابر ۰/۴۷۹۳ می‌باشد و نشان می‌دهد که در هر واحد ریسک، بازده ای برابر ۰/۴۷۹۳تولید می‌شود. شاخص نسبت شارپ بازده در مقیاس ریسک را محاسبه می‌کند و به آن بازده تعدیل‌شده نیز گفته می‌شود و بیشتر بودن این شاخص، مطلوب سرمایه‌گذاران می‌باشد.

در ادامه سودآوری سیستم‌های معاملاتی انفرادی شکل‌گرفته بر اساس شاخص‌های تکنیکی نیز مورد ارزیابی قرار می‌گیرد. در ابتدا سیستم معاملاتی شکل‌گرفته بر اساس شاخص قدرت نسبی با خرید در ۳۰ و فروش در ۷۰ بر روی‌داده‌های آزمون موردمحاسبه و شبیه‌سازی قرار گرفت که نتیجه در جدول۷ ارائه‌شده است.

جدول۷: مشخصات آماری معادل بازده روزانه سیستم معاملاتی شاخص قدرت نسبی

مقدار	مشخصات سیستم قدرت نسبی
۱۴	تعداد موقعیت معاملاتی
۰/۰۰۷۸	بیشینه بازده موقعیت
-۰/۰۰۳۳	کمینه بازده موقعیت
۲۵/۷۷۲۷	متوسط طول زمانی موقعیت
۰/۰۰۰۴	متوسط معادل بازده روزانه
۰/۰۰۲۷	انحراف معیار معادل بازده روزانه
۰/۱۴۹۳	نسبت شارپ

بر اساس اطلاعات حاصل‌شده، ۱۴ موقعیت معاملاتی توسط سیستم پژوهش شناسایی گردید که متوسط بازده روزانه‌ای برابر ۰/۰۰۰۴را تولید می‌کند که این مقدار برابر متوسط بازده سالیانه‌ای برابر ۰/۱۶می‌باشد. متوسط زمان

هر معامله تقریباً برابر ۲۵ روز می‌باشد. نسبت شارپ سیستم پژوهش که از تقسیم بازده بر انحراف معیار محاسبه می‌شود برابر ۰/۱۴۹۳ می‌باشد و نشان می‌دهد که در هر واحد ریسک، بازده ای برابر ۰/۱۴۹۳ تولید می‌شود.

در ادامه سیستم معاملاتی شکل‌گرفته بر اساس شاخص تصادفی با خرید در ۳۰ و فروش در ۷۰ بر روی‌داده‌های آزمون موردمحاسبه و شبیه‌سازی قرار گرفت که نتیجه در جدول۸ ارائه‌شده است.

جدول۸: مشخصات آماری معادل بازده روزانه سیستم معاملاتی شاخص تصادفی

مقدار	مشخصات سیستم معاملاتی شاخص تصادفی
۱۱	تعداد موقعیت معاملاتی
۰/۰۰۶۷	بیشینه بازده موقعیت
-۰/۰۰۲۳	کمینه بازده موقعیت
۲۰/۲۷۲۷	متوسط طول زمانی موقعیت
۰/۰۰۰۵۷	متوسط معادل بازده روزانه
۰/۰۰۳۰	انحراف معیار معادل بازده روزانه
۰/۱۹۱۲	نسبت شارپ

بر اساس اطلاعات حاصل‌شده، ۱۱ موقعیت معاملاتی توسط سیستم پژوهش شناسایی گردید که متوسط بازده روزانه‌ای برابر ۰/۰۰۰۵۷ را تولید می‌کند که این مقدار برابر متوسط بازده سالیانه‌ای برابر۰/۲۳ می‌باشد. متوسط زمان هر معامله تقریباً برابر ۲۰ روز می‌باشد. نسبت شارپ سیستم پژوهش که از تقسیم بازده بر انحراف معیار محاسبه می‌شود برابر ۰/۱۹۱۲ می‌باشد و نشان می‌دهد که در هر واحد ریسک، بازده ای برابر ۰/۱۹۱۲ تولید می‌شود. در ادامه به عملکرد سیستم معاملاتی شکل‌گرفته بر اساس شاخص همگرایی-واگرایی پرداخته می‌شود که نتیجه عملکرد آن در جدول۹ ارائه‌شده است.

جدول۹: مشخصات آماری معادل بازده روزانه سیستم معاملاتی شاخص همگرایی-واگرایی

مقدار	مشخصات سیستم معاملاتی شاخص همگرایی-واگرایی
۱۴	تعداد موقعیت معاملاتی
۰/۰۰۹۷۲۹	بیشینه بازده موقعیت
-۰/۰۱۶۵۳	کمینه بازده موقعیت
۳۲/۵۱۳۵۱	متوسط طول زمانی موقعیت
-۰/۰۰۰۴۴	متوسط معادل بازده روزانه
۰/۰۰۴۷۹۳	انحراف معیار معادل بازده روزانه
-۰/۰۹۱۰۳	نسبت شارپ

بر اساس اطلاعات حاصل‌شده، ۱۴ موقعیت معاملاتی توسط سیستم پژوهش شناسایی گردید که متوسط بازده روزانه‌ای برابر ۰/۰۰۰۴۴- را تولید می‌کند. ازاین‌رو سیستم شکل‌گرفته بر اساس شاخص همگرایی-واگرایی زیان ده می‌باشد.

عملکرد چهار سیستم معاملاتی شامل یادگیری تقویتی عمیق، مبتنی بر شاخص قدرت نسبی، مبتنی بر شاخص تصادفی و مبتنی بر شاخص همگرایی و واگرایی مورد ارزیابی قرار گرفت. جدول۱۰ نتایج چهار استراتژی معاملاتی را به‌صورت یکجا نشان می‌دهد.

جدول۱۰: مشخصات آماری چهار استراتژی مورداستفاده در پژوهش

سیستم معاملاتی شاخص عملکرد	یادگیری تقویتی عمیق	قدرت نسبی	تصادفی	همگرایی- واگرایی
تعداد موقعیت معاملاتی	۱۳	۱۴	۱۱	۱۴
بیشینه بازده موقعیت	۰/۰۱۰۴	۰/۰۰۷۸	۰/۰۰۶۷	۰/۰۰۹۷۲۹
کمینه بازده موقعیت	-۰/۰۰۱۵	-۰/۰۰۳۳	-۰/۰۰۲۳	-۰/۰۱۶۵۳
متوسط طول زمانی موقعیت	۲۱/۵۳۸۵	۲۵/۷۷۲۷	۲۰/۲۷۲۷	۳۲/۵۱۳۵۱
متوسط معادل بازده روزانه	۰/۰۰۱۸	۰/۰۰۰۴	۰/۰۰۰۵۷	-۰/۰۰۰۴۴
انحراف معیار معادل بازده روزانه	۰/۰۰۳۷	۰/۰۰۲۷	۰/۰۰۳۰	۰/۰۰۴۷۹۳
نسبت شارپ	۰/۴۷۹۳	۰/۱۴۹۳	۰/۱۹۱۲	-۰/۰۹۱۰۳

جدول۱۱: مقدار احتمال آزمون برابری میانگین برای چهار سیستم معاملاتی

نوسانگر	سیستم پژوهش	شاخص قدرت نسبی	شاخص تصادفی	شاخص همگرایی- واگرایی
سیستم پژوهش		۰/۰۰۶۹	۰/۰۰۹۱۲	۰/۰۰۰
شاخص قدرت نسبی			۰/۸۷۳	۰/۰۰۰
شاخص تصادفی				۰/۰۰۰
شاخص همگرایی-واگرایی				

در ادامه عملکرد سیستم‌های معاملاتی چهارگانه پژوهش از بابت تفاوت معناداری آماری بین میانگین‌های بازده روزانه کسب‌شده مورد به کمک آزمون مقایسات زوجی بررسی قرار گرفت.

بر اساس اطلاعات جدول۱۰ و ۱۱ سیستم پژوهش در میانگین دارای تفاوت معناداری با سه سیستم دیگر می‌باشد. سیستم شکل‌گرفته بر اساس شاخص قدرت نسبی و تصادفی دارای تفاوت معناداری نمی‌باشند و سیستم شکل‌گرفته بر اساس شاخص همگراییی-واگرایی دارای تفاوت معنادار با سه سیستم دیگر می‌باشد. همچنین نسبت شارپ سیستم پژوهش نسبت به سه مدل دیگر رشد حداقل ۱/۴ برابری را نشان می‌دهد.

۵. بحث و نتیجه‌گیری

در پژوهش حاضر یک سیستم معاملاتی بر اساس رویکرد یادگیری تقویتی و به کمک شبکه‌های عصبی عمیق طراحی گردید. در این رویکرد عامل یا معامله‌گر در فضای جستجو برای یافتن پاداش بیشتر که همان بازده بالاتر می‌باشد، به جستجو می‌پردازد. عامل معامله‌گر با سیگنال‌های تکنیکی شامل شاخص قدرت نسبی، نوسانگر تصادفی، شاخص همگرایی-واگرایی و قیمت‌های کمینه، بیشینه، بسته و باز شدن مواجه می‌شود. عامل این اطلاعات را در گام‌های زمانی متوالی دریافت می‌کند و بر اساس آن‌ها یک اقدام انجام می‌دهد که در اینجا شامل خرید، نگهداری و یا فروش دارایی می‌باشد. یادگیری تقویتی به دنبال طراحی یک استراتژی الگوریتمیک یا خودکار می‌باشد که درنتیجه آن میزان پاداش یعنی بازده خریدوفروش بیشینه شود که یکی از الگوریتم‌های مشهور در این زمینه q-یادگیری می‌باشد. غالباً برای پیاده‌سازی یادگیری تقویتی از جدول Q استفاده می‌شود که در آن کیفیت اقدام‌های مختلف در وضعیت‌های مختلف در جدول ضبط و رکورد می‌شود. در مدل موردبررسی پژوهش با توجه به ماهیت پیوسته متغیرهای حالت، جدول Q قابل‌استفاده نیست و به‌جای آن از یک شبکه عصبی عمیق استفاده می‌شود تا استراتژی معاملاتی یادگیری تقویتی مدنظر پژوهش را آموزش ببیند. درنهایت شبکه عصبی آموزش‌دیده در مرحله آزمایش می‌تواند با دریافت اطلاعات تکنکیی و قیمتی مذکور، یکی از سه اقدام خرید، نگهداری و فروش را پیشنهاد دهد. در راستای سؤالات پژوهش به دو سؤال پاسخ داده شد.

۱- چگونه می‌توان به کمک یادگیری تقویتی عمیق یک سیستم معاملاتی طراحی کرد؟

برای طراحی سیسم معاملاتی بر پایه یادگیری تقویتی عمیق از دو شبکه عصبی استفاده می‌شود. شبکه اول برای محاسبه مقادیر Q در سمت راست معادله بلمن استفاده می‌شود. این شبکه عصبی آموزش نمی‌بیند و فقط برای مقصد محاسبه احتمالات استفاده می‌شود. پس از محاسبه سمت راست معادله بلمن برای یک ورودی، از یک شبکه عصبی (که همان سیستم معاملاتی می‌باشد) برای آموزش استفاده می‌شود. بدین‌صورت که ورودی آن حالت و خروجی آن مقدار معادله بلمن می‌باشد که توسط شبکه عصبی اول آموزش‌دیده است. در فواصل زمانی منظم شبکه عصبی دوم، جای شبکه عصبی اول را می‌گیرد و شبکه عصبی دوم از نو آموزش می‌بیند.

۲-سودآوری سیستم معاملاتی طراحی‌شده چگونه است؟

سیستم معاملاتی پژوهش پس از آموزش بر روی ۱۶۵۴ داده، بر روی ۱۰۰۰ داده تست مورد ارزیابی سودآوری قرار گرفت. نتایج نشان داد که سیستم پژوهش در میانگین دارای تفاوت معناداری با سه سیستم دیگر می‌باشد. سیستم شکل‌گرفته بر اساس شاخص قدرت نسبی و تصادفی دارای تفاوت معناداری نمی‌باشند و سیستم شکل‌گرفته بر اساس شاخص همگراییی-واگرایی دارای تفاوت معنادار با سه سیستم دیگر می‌باشد. همچنین نسبت شارپ سیستم پژوهش نسبت به سه مدل دیگر رشد حداقل ۱/۴ برابری را نشان می‌دهد. با توجه به نتایج حاصل‌شده از سیستم معاملاتی پژوهش پیشنهادهای زیر ارائه می‌گیرد.

۱-با توجه به نسبت شارپ بالای سیستم پژوهش نسبت به سیستم‌های انفرادی شکل‌گرفته بر اساس ابزارهای تکنیکی به سرمایه‌گذاران ریسک گریز پیشنهاد می‌شود تا از سیستم معاملاتی پژوهش برای ترکیب سیگنال‌های تکنیکی و اتخاذ موقعیت‌های معاملاتی استفاده کنند.

۲- توجه شود که عملکرد سودآوری سیستم معاملاتی پژوهش باید با توجه به سهام موردنظر مورد بر روی‌داده‌های آزمایشی مورد ارزیابی قرار گیرد. همچنین پیشنهاد می‌شود تا با تغییر ورودی‌های مدل و استفاده از طیف وسعی از متغیرهای تکنیکی و بنیادی دقت سیستم معاملاتی ارزیابی و بهترین مدل انتخاب گردد.

۶. منابع

اسفندیار، مهدی، کرامتی، محمد علی، غلامی جمکرانی، رضا، کاشفی نیشابوری، محمدرضا (۱۴۰۱). بهینه‌سازی پرتفوی سهام در بورس اوراق بهادار تهران (کاربرد رهیافت یادگیری تقویتی). فصلنامه علمی مدلسازی اقتصادی, ۱۶(۵۸), ۵۱-۶۶.

خنجرپناه، حسین، دوروش، داود، شوال پور، سعید، جبارزاده، آرمین. (۱۳۹۷). کاربرد روش تکنیکال برای پیش‌بینی قیمت سهام: رویکرد مدل‌های احتمال غیرخطی و شبکه‌های عصبی مصنوعی. *راهبرد مدیریت مالی*, ۶(۳), ۵۹-۷۹.

رستگار، محمد علی، صداقتی‌پور، امین. (۱۳۹۷). ارایه سیستم معاملات الگوریتمی برای قرارداد آتی سکه طلا مبتنی بر داده‌های درون-روزی. *دانش سرمایه‌گذاری*، ۷(۲۸), ۶۸-۴۹.

سارنج، علیرضا، قاسمی، احمدرضا، ارم، اصغر، تهرانی، رضا. (۱۳۹۹). طراحی سیستم معاملات تکنیکی سهام با استفاده از مدل ترکیبی شبکه عصبی *MLP* و الگوریتم‌های تکاملی. *دانش مالی تحلیل اوراق بهادار*، ۱۳(۴۵), ۶۴-۴۷.

شمس، شهاب الدین؛ عطایی، بهروز. (۱۳۹۵). شناسایی دستکاری قیمت سهام از طریق مدل ترکیبی الگوریتم ژنتیک – شبکه عصبی مصنوعی و مدل *SQDF*. *راهبرد مدیریت مالی*، ۴(۳)، ۱۴۹-۱۷۱.

صیادی نژاد, اسماعیل زاده, & رستمی. (۲۰۲۳). ارائه مدل پیش‌بینی بازدهی بیت‌کوین با استفاده از روش هیبریدی یادگیری عمیق-الگوریتم تجزیه سیگنال (CEEMD-DL). اقتصاد مالی, ۱۷(۶۲), ۲۱۷-۲۳۸.

صیادی نژاد، سکینه، اسماعیل زاده، علی،رستمی، محمدرضا (۱۴۰۲). ارائه مدل پیش‌بینی بازدهی بیت‌کوین با استفاده از روش هیبریدی یادگیری عمیق-الگوریتم تجزیه سیگنال (CEEMD-DL). اقتصاد مالی, ۱۷(۶۲), ۲۱۷-۲۳۸.

یافتیان، امیرحسین، رستگار، محمدعلی. (۱۳۹۹). طراحی یک سیستم معاملاتی خودکار با استفاده از شبکه عصبی پیچشی. *چشم انداز مدیریت مالی*، ۱۰(۳۱), ۱۵۳-۱۸۴.

Jin, B. (2023). A Mean-VaR Based Deep Reinforcement Learning Framework for Practical Algorithmic Trading. IEEE Access, 11, 28920-28933.

Koratamatti, P., Wadhwani, K., Gupta, M., & Sanjeevi, S. G. (2021). Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation. Engineering Science and Technology, an International Journal, 24(4), 848-859.

Li, Y., Liu, P., & Wang, Z. (2022). Stock Trading Strategies Based on Deep Reinforcement Learning. Scientific Programming, 2022.

*Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A and Salwana, E. and S., Shahab. (2020). Deep Learning for Stock Market Prediction. Entropy **2020**, 22(8), 840.*

Thibaut, T., Damien, E (2020). An Application of Deep Reinforcement Learning to Algorithmic Trading, ArXiv2021.

Wang,L.,Wang.Z and Zhao,SH. (2015). Stock market trend prediction using dynamical Bayesian factor graph, Expert Systems with Applications,45,6263-6270.

Xiao-dan Zhang, Ang Li, Ran Pan, (2016),Stock trend prediction based on a new status box method and AdaBoost probabilistic support vector machine. Applied Soft Computing 49 , 385–398.

Xu,Y., and V. Keselj. (2019). Stock Prediction using Deep Learning and Sentiment Analysis, 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 5573-5580.

Yang, C., Zhai, J., Tao, G. (2020). Deep Learning for Price Movement Prediction Using Convolutional Neural Network and Long Short-Term Memory. Mathematical Problems in Engineering, 20,1-13.

Yu, P & Yan, X. (2020). Stock price prediction based on deep neural networks. Neural Computing and Applications. 32. 10.

Zhang, J., & Lei, Y. (2022). Deep Reinforcement Learning for Stock Prediction. Scientific Programming, 2022.

Journal of Financial Knowledge of Securities Analysis
Vol. 17 / No. 61, Spring 2024

Designing algorithmic trading strategy based on deep reinforcement learning
Case study: Tehran Stock Exchange

Saeed Kazemian hossinabadi[1]
Sayyed Mohammad Reza Davoodi[2*]
Mohammad Mashhadizadeh[3]
Parsa Jozzi[4]

Abstract

Today, algorithmic trading is widely used in trading management. Algorithmic portfolio management is a new type of these systems through which the portfolio manager helps to increase the quality of profit and reduce the risks of his portfolio using algorithmic tools. The purpose of this research is to design an algorithmic trading system based on deep reinforcement learning with the help of neural network. In this approach, the agent or trader searches the search space to find more reward, which is the same as more return. The trader is faced with technical signals including relative strength index, stochastic oscillator, convergence-divergence indicator and minimum, maximum, closing and opening prices. Deep reinforcement learning replaces the Q value or quality function table with a neural network. Finally, upon receiving the state word, the mentioned neural network suggests one of the three actions of selling, buying and holding. This proposal is in the form of three possibilities with a total of one, and the proposal with the maximum probability is implemented. The result of the implementation of the deep reinforcement learning trading system on the total index of the Tehran Stock Exchange in the period of 2013 to 2014 shows that the research system had a significant difference with the other three systems. The system formed based on the relative and random power index has no significant difference, and the system formed based on the convergence-divergence index has a significant difference with the other three systems. Also, the Sharpe ratio of the research system compared to the other three models showed a growth of at least 1.4 times.

Keywords: *neural networks, reinforcement learning, deep reinforcement learning, technical oscillator.*

[1]*Master of Financial Management, Dehaghan Branch, Islamic Azad University, Dehaghan, Iran. Saeedkazemian1365@yahoo.com*
[2] *Assistant Professor, Dehaghan Branch, Islamic Azad University, Dehaghan, Iran. (Corresponding author). Smrdavoodi@ut.ac.ir*
[3]*Assistant Professor, Mobarakeh Branch, Islamic Azad University, Mobarakeh, Iran. m.mashhadi@mau.ac.ir*
[4]*Master of Financial Management,Dehaghan Branch, Islamic Azad University, Dehaghan,Iran. Parsajozipersonal@gmail.com*