



بهینه‌سازی پرتفوی سهام با استفاده از استراتژی یادگیری تقویتی کیو عمیق

مبنتی بر ماتریس حالت - عمل

مهدی اسفندیاری^۱

محمدعلی کرامتی^۲

رضا غلامی جمکرانی^۳

محمد رضا کاشفی نیشابوری^۴

تاریخ دریافت مقاله : ۱۴۰۲/۰۲/۲۵ تاریخ پذیرش مقاله : ۱۴۰۲/۰۲/۲۷

چکیده

هدف این مقاله بهینه‌سازی پرتفوی سهام با استفاده از استراتژی یادگیری تقویتی کیو عمیق مبتنی بر ماتریس حالت-عمل می باشد. بدین منظور، برای بهینه‌سازی و سودآوری پرتفوی متشکل از سهام، عملکرد استراتژی یادگیری تقویتی مبتنی بر الگوریتم کیو عمیق و استراتژی منفعل خرید و نگهداری در دو حالت بازارهای صعودی و نزولی طی دوره زمانی ۱۳۹۶-۱۴۰۰ مورد بررسی قرار گرفت. جامعه آماری ۶۷۲ شرکت پذیرفته شده در بورس اوراق بهادار تهران بوده که از میان آنها تعداد ۷ شرکت (نمونه آماری) مناسب دانسته شد. مقایسه دو استراتژی نشان می‌دهد استراتژی یادگیری تقویتی، در بازارهای صعودی و نزولی در مقایسه با روش معاملاتی خرید و نگهداری که منجر به زیان شده است، در عمل توان بالایی برای سودآوری در بازار بورس اوراق بهادار ایران دارد. براساس نتایج، پیشنهاد می‌شود کارگزاران و شرکتهای بورسی و تحلیلگران از راهبرد یادگیری تقویتی برای سودآوری و بهینه‌سازی پرتفوی سهام استفاده کنند. همچنین، مقایسه نتایج این دو رویکرد روشن می‌کند که کاربرد یادگیری تقویتی برای سرمایه‌گذارهایی که توان ریسک‌پذیری بالای رهیافت خرید و نگهداری را ندارند، مناسب‌تر است.

کلمات کلیدی

بهینه‌سازی پرتفوی، معاملات الگوریتمی، یادگیری تقویتی، الگوریتم کیو عمیق، بورس اوراق بهادار

تهران. طبقه‌بندی JEL: G11, G17, P45.

۱- دانشجوی دکتری، گروه مدیریت صنعتی، واحد قم، دانشگاه آزاد اسلامی، قم، ایران. esfandiar.mehdi@yahoo.com

۲- دانشیار، گروه مدیریت صنعتی، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران. (نویسنده مسئول)
mohammadalikeramati@yahoo.com

۳- دانشیار، گروه حسابداری، واحد قم، دانشگاه آزاد اسلامی، قم، ایران. accountghom@gmail.com

۴- دانشیار، گروه مدیریت مالی، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران. mm.kashefi63@gmail.com

بهینه‌سازی پورتفولیو (پرتفوی)^۱، مولفه اصلی نظام معاملاتی است. مدیریت پرتفوی یک استراتژی سرمایه‌گذاری است که با هدف به حداکثر رساندن بازده مورد انتظار سرمایه و در عین حال به حداقل رساندن ریسک مالی با تخصیص مجدد مستمر دارایی‌های پرتفوی، یعنی با نسبت دادن وزن مناسب به هر ابزار مالی، انجام می‌شود (سلیمانی^۲ و همکاران، ۲۰۲۰). مارکوئیتز^۳ نخستین اقتصاددانی بود که این نظریه را مطرح کرد و این نظریه با عنوان «نظریه پورتفولوی مدرن» (MPT)^۴ قبول عام یافت. مزیت اصلی چنین پورتفولیویی مبتنی بر ترویج تنوع‌بخشی است که «منحنی سرمایه‌سازان سهام»^۵ را هموار می‌سازد؛ یعنی، نسبت به معاملات دارایی انفرادی به بازده بیشتری منجر می‌شود. این امر بدین معناست که ریسک (نوسان) پورتفولیوی طولانی‌مدت همواره کمتر از ریسک دارایی انفرادی است (ژانگ، زهرن و رابرتس،^۶ ۲۰۲۰). به‌باور مارکوئیتز، سرمایه‌گذاران، ریسک و بازده را باهم در نظر می‌گیرند و میزان تخصیص سرمایه به فرصت‌های سرمایه‌گذاری گوناگون را براساس تعامل ریسک و بازده انتخاب می‌کنند (ریوز، مووز و ونما،^۷ ۲۰۱۴).

از طرفی یکی از حوزه‌های جدید در بازارهای مالی، معاملات الگوریتمی مبتنی بر یادگیری تقویتی و توسعه آن در بازار سرمایه‌های سراسر دنیا است و می‌توان گفت بخش قابل توجهی از معاملات، توسط معاملات خودکار صورت می‌گیرد. سه نوع تحلیل شناخته شده در بازارهای مالی مورد استفاده قرار می‌گیرد. تحلیل بنیادی بر پایه عملکرد شرکت‌ها و رشد سودآوری آن‌ها بنا شده است؛ تحلیل مالی رفتاری، حوزه‌ای از دانش مالی است که از نظریه‌های مبتنی بر روانشناسی برای توضیح رفتار بازارهای مالی بهره می‌گیرد و تحلیل تکنیکال سومین روش است که بر پایه سابقه معاملات یک دارایی مالی از طریق نمودار قیمت و فرمول‌های ریاضی که اندیکاتورهای تکنیکال نامیده می‌شوند، بنا شده است. در سال‌های اخیر از هوش مصنوعی نیز برای پیش‌بینی بازار استفاده شده است که ترکیب آن با تحلیل تکنیکال می‌تواند منجر به ایجاد سیستم‌های خودکار معاملاتی و الگوریتمی شود (ها^۸ و همکاران، ۲۰۲۰).

در سیستم مبتنی بر معاملات الگوریتمی، سفارش‌های بزرگ معمولاً به سفارش‌های کوچک شکسته و در چندین مرحله انجام می‌شود. استراتژی معاملاتی در این نوع معاملات شامل نحوه تصمیم‌گیری در خصوص مبلغ سرمایه‌گذاری، مدیریت اجرای سفارش، قیمت سفارش‌گذاری و غیره است. معاملات مبتنی بر الگوریتم، برنامه‌ای شامل تعداد مراحل و قواعد از پیش تعریف شده است که به منظور دستیابی به یک هدف معین طراحی می‌شوند (آگاروال^۹ و همکاران، ۲۰۰۶).

بهینه‌سازی پرتفوی سهام با استفاده.../اسفندیاری، کرامتی، غلامی جمکرانی و کاشفی نیشابوری

همچنین معاملات مبتنی بر یادگیری تقویتی، نوعی از معاملات خودکار بوده که شامل برنامه‌های کامپیوتری برای ارسال سفارشات همراه با الگوریتم‌های تصمیم‌گیری هستند که این الگوریتم‌ها، خود بر اساس پارامترهای منحصر به فرد سفارش، مانند زمان، قیمت و یا مقدار سفارش می‌باشند. در بازارهای مالی الکترونیکی، معاملات مبتنی بر یادگیری تقویتی به معنای استفاده از برنامه‌های کامپیوتری برای ورود سفارش‌های معاملاتی است که سیگنال‌های معاملاتی آن توسط بخشی دیگر از سیستم تولید شده است (سلیمانی و همکاران، ۲۰۲۰).

پژوهش‌ها نشان می‌دهد استفاده از ابزارها و روش‌های پیش‌بینی سنتی، خطای بالایی دارد و در بیشتر موارد، در مقایسه با روش‌های جدیدتر و مدل‌های غیرخطی عملکرد ضعیف‌تری دارند. به عبارتی پیشرفت‌های فن‌آوری، بر معاملات بازار سرمایه نیز همانند بسیاری از عرصه‌های دیگر، اثرات زیادی گذاشته است. حجم داده‌های بازار سرمایه که امروزه تولید می‌شوند، به حدی زیاد است که بدون بکار بردن فن‌آوری‌های مدرن و رایج، نمی‌توان به تصمیم‌گیری مناسب و معامله اقدام نمود. معاملات الگوریتمی، خانواده بزرگی از انواع روش‌ها را در خود جای می‌دهند، که انواع روش‌های یادگیری ماشین و الگوریتم‌های متنوع، در معاملات الگوریتمی کاربرد دارند. نمودار ریخت‌شناسی روش یادگیری تقویتی^{۱۰} به خوبی مبین انواع روش‌های اجرای روش یادگیری تقویتی است که تنها یکی از روش‌های یادگیری ماشین و معاملات الگوریتمی^{۱۱} می‌باشد، و الگوریتم یادگیری تقویتی کیو عمیق^{۱۲}، یکی از روش‌های انجام یادگیری تقویتی است که در این پژوهش مورد توجه قرار دارد و بکار برده شده است. با پیشرفت‌های اخیر در یادگیری ماشینی و یادگیری عمیق، پیش‌بینی رفتارهای مالی پیچیده و خودکارسازی فرآیند تصمیم‌گیری حداقل در زمینه بهینه‌سازی پرتفوی تا حدی ممکن شده است (سلیمانی و همکاران، ۲۰۲۰). از منظر معاملات یادگیری تقویتی، پیدا کردن مدلی برای سفارش‌گذاری بهینه دارای اهمیت است. نهادهای فعال بازار در هر بازه زمانی، استراتژی مشخصی برای انجام معامله و کسب سود دارند. استراتژی‌های معاملاتی بایستی از لحاظ عملکرد و بازده همواره مورد پایش قرار گیرند. ارزیابی استراتژی معاملاتی بویژه در معاملات مبتنی بر یادگیری تقویتی و با بسامد بالا دارای اهمیت است، چراکه ریزساختار بازار با سرعت بالا در حال تغییر است (محمد^{۱۳} و همکاران، ۲۰۲۳).

از طرفی با توجه به این مهم که در یادگیری تقویتی، خروجی الگوریتم‌های مورد استفاده ماتریس حالت-عمل می‌باشد، می‌بایست مفهوم حالت را در قالب حالت بازار دانست که در تحقیق حاضر، از میان سه حالت مطرح برای بازار سرمایه (صعودی، نزولی و خنثی)، دو حالت صعودی و نزولی مطرح شده اند، چرا که در حالت خنثی نوسانات اندک در قیمت سهام اجازه نخواهد داد که معامله‌ای توسط سیگنال

توصیه شده و در نتیجه، به صورت منطقی می‌توان از بررسی حالت خنثی صرف نظر نمود. از آنجاکه هدف موردانتظار در روش یادگیری تقویتی، یادگیری عامل از محیط و پاداش و جریمه‌های اکتسابی در محیط می‌باشد، که در نتیجه این یادگیری از اقدامات، ماتریس کیو مبتنی بر معادله مورد استفاده در آن الگوریتم ایجاد می‌شود، اهمیت و برتری در میان الگوریتم‌ها، به سیاست‌های بهینه‌سازی سبد بستگی پیدا می‌کند. لذا در صورتی که به دنبال منافع آنی از اقدامات عامل در محیط و به دنبال بهره‌برداری باشیم و از اکتشاف چشم پوشی کنیم، روش کیو عمیق، با توجه به سیاست غیرحریصانه‌ای که دارد، الگوریتم مناسب‌تری خواهد بود. لذا الگوریتم یادگیری کیو الگوریتم قدرتمندی است، اما قابلیت تعمیم‌پذیری ندارد و همین مسئله را می‌توان بزرگ‌ترین نقطه‌ضعف آن دانست. اگر الگوریتم یادگیری کیو را به‌روزرسانی اعداد موجود در یک ارائه دو بعدی (شامل: فضای اقدام× فضای وضعیت) در نظر بگیریم، متوجه شباهت آن با برنامه‌نویسی پویا خواهیم شد. این موضوع برای ما روشن می‌سازد که وقتی عامل تصمیم‌گیرنده در الگوریتم یادگیری کیو با وضعیتی کاملاً جدید روبه‌رو شود، هیچ راهی برای شناسایی و انتخاب اقدام مناسب نخواهد داشت. به عبارت دیگر، عامل تصمیم‌گیرنده الگوریتم یادگیری کیو توانایی تخمین ارزش وضعیت‌های ناشناخته را ندارد. برای حل این مشکل، شبکه DQN ارائه دو بعدی را حذف و شبکه عصبی را جایگزین آن می‌کند. (محمد و همکاران، ۲۰۲۳) در پژوهش حاضر برخلاف روش‌های سنتی که در آن، در وهله نخست، معمولاً با بهره‌گیری از مدل‌های اقتصادسنجی، بازده مورد انتظار پیش‌بینی می‌شود، این گام پیش‌بینی‌کننده کنار گذاشته می‌شود تا مستقیماً تخصیص‌های دارای به‌دست آید. بنابراین، با بهینه‌سازی مستقیم «نسبت شارپ»^{۱۴} بازده در واحد ریسک به حداکثر می‌رسد. از این‌رو، هدف این مقاله بررسی امکان و توان کاربرد یادگیری کیو عمیق مبتنی بر ماتریس حالت-عمل برای بهینه‌سازی پورتفوی با توجه به محدودیت‌های تعیین‌شده توسط بازار سهام، مانند نقدینگی و هزینه‌های مبادله است.

برای این منظور، مقاله در پنج بخش سازمان‌دهی می‌شود: بعد از مقدمه، در بخش دوم، ادبیات پژوهش بررسی می‌شود؛ در بخش سوم، روش پژوهش بیان می‌شود و بخش‌های چهارم و پنجم نیز به یافته‌ها و نتیجه‌گیری و پیشنهادها اختصاص می‌یابد.

مبانی نظری

در این بخش، برای برقراری ارتباط بهتر مفاهیم و نظریات و درنهایت، دست‌یابی به بینشی عمیق نسبت به مسائل اقتصادی، اصطلاحات و مفاهیم اساسی تعریف و تحدید می‌شود.

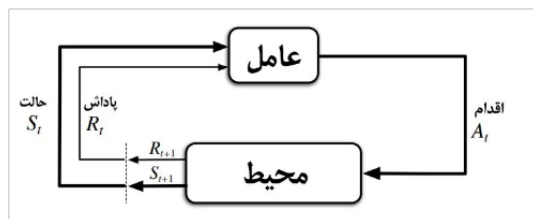
بهینه‌سازی پورتفوی سهام با استفاده.../اسفندیاری، کرامتی، غلامی جمکرانی و کاشفی نیشابوری

بهینه‌سازی پورتفویو

بهینه‌سازی پورتفویو، عبارت است از انتخاب بهترین ترکیب از دارایی‌های مالی به نحوی که باعث شود، تاحد ممکن بازده پورتفویو، بیشینه و ریسک آن، کمینه شود. ایده اساسی نظریه مدرن پورتفویو این است که اگر در دارایی‌هایی که به‌طور کامل هم‌بستگی ندارند، سرمایه‌گذاری شود؛ ریسک آن دارایی‌ها یکدیگر را خنثی کرده و می‌توان یک بازده ثابت با ریسک کمتر به دست آورد (مارکوئیتز، ۱۹۵۲). در بهینه‌سازی پورتفویو، مساله اصلی، انتخاب بهینه دارایی‌ها و اوراق بهاداری است که با مقدار مشخصی سرمایه می‌توان به دست آورد؛ اگرچه، کمینه کردن ریسک و بیشینه کردن بازده سرمایه‌گذاری ساده به نظر می‌رسد؛ روش‌های متعددی برای تشکیل پورتفویوی بهینه به کار رفته است (ترلیون^{۱۵} و همکاران، ۲۰۱۳).

یادگیری تقویتی

یادگیری تقویتی^{۱۶} یکی از زیرشاخه‌های یادگیری ماشین است که در آن یک عامل یادگیرنده در تعامل با محیط سعی می‌کند به یک سیاست بهینه دست یابد. عامل یادگیرنده با مشاهده وضعیت سیستم (S)، اقدام (A) را انتخاب می‌نماید. محیط بازخورد این اقدام را در قالب پاداش (R) و حالت بعدی سیستم به عامل بازمی‌گرداند. عامل مجدداً با مشاهده پاداش و حالت سیستم، اقدام بعدی را انتخاب می‌کند و این فرآیند تا زمان رسیدن به سیاست بهینه ادامه پیدا می‌کند (امیری^{۱۷} و همکاران، ۲۰۱۸).



شکل ۱- نحوه عملکرد سیستم یادگیری تقویتی (منبع: امیری و همکاران، ۲۰۱۸)

انواع الگوریتم‌های یادگیری تقویتی

در یک دسته بندی کلی میتوان الگوریتم‌های یادگیری تقویتی را به دو دسته الگوریتم‌های مبتنی بر سیاست^{۱۸} و الگوریتم‌های مستقل از سیاست^{۱۹} تقسیم بندی نمود (امیری و همکاران، ۲۰۱۸):

الگوریتم‌های یادگیری تقویتی مبتنی بر سیاست که در این الگوریتم‌ها تابع ارزش بر اساس سیاست و اقدام فعلی عامل به روز می‌شود.

الگوریتم‌های یادگیری تقویتی مستقل از سیاست که در این الگوریتم‌ها تابع ارزش مستقل از سیاست

و اقدام فعلی عامل به روز می‌شود.

الگوریتم کیو عمیق یکی از معروفترین الگوریتم های on-policy یادگیری تقویتی است و الگوریتم یادگیری کیو^{۲۰} نیز یکی از معروف ترین الگوریتم های off-policy یادگیری تقویتی است.

معاملات الگوریتمی در ایران

هم‌زمان با ورود معاملات برخط^{۲۱} به بازار ایران، یک‌دریچه جدید برای فعالان بازار سرمایه گشوده شد. با استفاده از این ابزار امکان ورود سیستم‌های الگوریتمی و خودکار به بازار کشور فراهم گشت (پویان فر، ۱۳۹۴). با توجه به این که که معاملات الگوریتمی یکی از سریع‌ترین روش‌ها جهت تبدیل دانش یا ایده معاملاتی به پول و دارایی می‌باشد، در سال‌های اخیر شاهد ورود کارشناسان فنی، مهندسی و ریاضیات به بازار ایران هستیم. علاوه بر این فعالیت برخی شرکت‌ها برای سرمایه‌گذاری در زیرساخت‌های معاملات الگوریتمی از سال ۱۳۹۱ شروع شده است. در این فاصله ریزنی‌هایی با سازمان بازار سرمایه و شرکت بازار سرمایه و فرابازار سرمایه جهت معرفی این سامانه‌ها به بازار انجام گرفت. بدین ترتیب، اولین دستورالعمل معاملات الگوریتمی (معاملات شرطی و خودکار) و مجوز مورد نیاز برای انجام این فعالیت در سال ۱۳۹۴ صادر گردید. علاقه‌مندی نهادهای بزرگ مالی جهت استفاده از این ابزار در سال ۱۳۹۵ شدت گرفت. در حال حاضر در حوزه معاملات الگوریتمی فعالیت‌های قابل اجرا توسط نهادهای مالی شامل بازارگردانی اوراق و سهام توسط نرم افزارهای الگوریتمی تولید شده و سبدگردانی می‌باشد. از سوی دیگر سرمایه‌گذاران قادرند فعالیت‌های پردازش اطلاعات، تحلیل، تصمیم‌گیری خودکار، ارسال سفارش‌ها به صورت دستی یا سیستمی را انجام دهند و یا از ابزارهای معاملات الگوریتمی ارائه شده توسط شرکت‌های معاملات برخط و کارگزاری‌های معتبر (صحرا، تدبیر پرداز) استفاده نمایند. با این حال نیاز به تدوین و تست استراتژی‌های معاملاتی در این حوزه به شدت احساس می‌شود (کیایی، ۱۳۹۷).

پیشینه تحقیق

پژوهش‌های زیادی با موضوع بهینه‌سازی پرتفوی انجام شده است که در آنها از روش‌های مختلفی استفاده شده است. از نظر تاریخی، پژوهش اسکابار و کلوت^{۲۲} (۲۰۰۲) نخستین اثری است که سعی کرده با ترکیب تحلیل تکنیکال و یادگیری تقویتی بهینه‌سیستم معاملاتی خودکاری طراحی کرده و گسترش دهد. پارک، سیم و چویی^{۲۳} (۲۰۲۰) بیان کرده‌اند که عامل هوشمند برای شناسایی اقدام معاملاتی بهینه برپایه یادگیری کیوی عمیق آموزش می‌بیند و نتایج بهتری از راهبردهای استاندارد کسب می‌شود. ژانگ^{۲۴} و همکاران (۲۰۲۰) در تحقیقی بیان کرده‌اند که روش آنها از مدل‌های پایه بهتر عمل کرده

بهینه‌سازی پرتفوی سهام با استفاده... / اسفندیاری، کرامتی، غلامی جمکرانی و کاشفی نیشابوری

و با وجود هزینه‌های سنگین تراکنش، سود مثبتی به دست آمده است.

میزبان و همکاران (۱۳۹۱)، در پژوهش خود از الگوریتم ازدحام ذرات برای بهینه‌یابی سبد دارایی مارکوویتز با توجه به معیارهای متفاوت اندازه‌گیری ریسک یعنی میانگین واریانس، میانگین نیم-واریانس و میانگین قدر مطلق انحرافات و همچنین محدودیتهای موجود در بازار واقعی مانند "اندازه ثابت تعداد سهام" و "محدودیت خرید" استفاده کردند. نتایج به دست آمده از این پژوهش حاکی از عملکرد موفق الگوریتم ازدحام ذرات در محاسبه مرز کارای مارکوویتز در تعاریف مختلف اندازه‌گیری ریسک است.

رستگار و دستپاک (۱۳۹۷) در پژوهش خود رویکرد بهره‌گیری از خودمعامله‌گرها برای پیش‌بینی روند آتی سهم و روش یادگیری تقویتی برای مدیریت پویای سبد سهام را مطرح کرده‌اند. نتایج نشان داد که یادگیری تقویتی عملکرد مدل را بهبود بخشیده است.

فلاح‌پور و حکیمیان (۱۳۹۸) در تحقیقی با به‌کارگیری داده‌های معاملات روزانه (درون‌روزی)^{۲۵} در بورس اوراق بهادار تهران بیان کرده‌اند که معاملات زوجی، معروف‌ترین و قدیمی‌ترین نوع سیستم‌های معاملات الگوریتمی است که کارایی و سودآوری آن در بسیاری از پژوهش‌های بازارهای مالی آشکار شده است. در سال‌های اخیر، تحقیقات مهمی درباره معاملات الگوریتمی براساس یادگیری ماشینی صورت گرفته است. نتایج آزمایش‌روی داده‌های معاملات روزانه زوجی سهام‌های منتخب‌نشان داد که کاربرد روش یادگیری تقویتی در طراحی سیستم معاملات در معامله‌های زوجی نسبت به سایر روش‌ها برتری دارد.

امیری و همکاران (۱۳۹۹)، در تحقیق خود از الگوریتم فراابتکاری به نام جستجوی انطباق تصادفی حریصانه^{۲۶} برای رفع مشکل بهینه‌سازی پرتفوی با محدودیت کاردینالیته^{۲۷} استفاده نمودند و به جهت تطابق بیشتر با دنیای واقعی، دو مجموعه محدودیت شامل محدودیت‌های کف و سقف و محدودیت کاردینالیته را به مدل مارکوویتز اضافه نمودند. بررسی نتایج حاصل از بهینه‌سازی پرتفوی با الگوریتم انطباق تصادفی حریصانه بر روی ۱۹۹ شرکت طی دوره ۵ ساله (۱۳۹۱-۱۳۹۵)، در بورس اوراق بهادار تهران نشان می‌دهد براساس معیار شارپ در هر پرتفوی ۵، ۱۵ و ۳۰ شرکتی الگوریتم انطباق تصادفی حریصانه در بهینه‌سازی پرتفوی کارا تر از مدل مارکوویتز عمل می‌کند.

گل‌ارزی و انصاری (۱۴۰۱) در تحقیقی به مقایسه عملکرد الگوریتم ژنتیک مرتب‌سازی نامغلوب با الگوریتم تکاملی قدرت پارتو بهبودیافته در انتخاب پورتفولیوی بهینه در بورس اوراق بهادار تهران پرداختند. نتایج نشان داد که در مدل، میانگین انحراف معیار عملکرد الگوریتم ژنتیک مرتب‌سازی نامغلوب نسبت به عملکرد الگوریتم تکاملی قدرت پارتو مطلوب‌تر و در مدل میانگین واریانس و میانگین نیم‌واریانس عملکرد الگوریتم تکاملی قدرت پارتو نسبت به عملکرد الگوریتم ژنتیک مرتب‌سازی نامغلوب مطلوب‌تر است.

گفتنی است پژوهشگران با مدل‌سازی‌های پیچیده ریاضی و مطرح کردن فرضیه‌های بسیار در بازار، به پیش‌بینی تغییرات قیمت سهام پرداخته‌اند، اما به دلیل پیچیدگی‌های زیاد بازار سرمایه، اغلب آنها موفقیت چندانی به دست نیاورده‌اند. اغلب پژوهش‌هایی که تاکنون در زمینه مدیریت سبد دارایی انجام شده، دارای دو بخش هستند: پیش‌بینی قیمت سهم و مدیریت سبد دارایی (رستگار و دستپاک، ۱۳۹۷). در اغلب این پژوهش‌ها، یادگیری نظارت شده برای ایجاد ارتباط میان یک سری داده ورودی و خروجی مطلوب استفاده شده است. در این میان، از آنجایی که روش‌های پیش‌بینی به کمک شبکه عصبی مصنوعی به مدل پارامتری نیازی ندارند، از محبوبیت بسیاری در پیش‌بینی قیمت سهام برخوردارند. مشکل روش‌های مشابه یادگیری نظارت شده در این نوع مسائل، این است که هدف آنها کاهش خطای بین پیش‌بینی حاصل از ورودی‌ها و خروجی مطلوب است، در حالی که در مدیریت سبد دارایی، هدف اصلی افزایش سود است و تنها پیش‌بینی قیمت سهم، دلیل بر سودآوری بالا نیست. از این‌رو، رویکرد آموزش با تأخیر زمانی، می‌تواند نتایج واقع‌بینانه‌تری داشته باشد؛ زیرا در این رویکرد، با توجه به زنجیره تصمیم‌ها، آموزش داده می‌شود که از اهداف پژوهش حاضر می‌باشد.

روش پژوهش

هدف این مقاله بهینه‌سازی پرتفوی سهام با استفاده از استراتژی یادگیری کیو عمیق مبتنی بر ماتریس حالت-عمل است. این پژوهش از حیث هدف، کاربردی و از نظر نوع داده، کمی و از لحاظ روش، توصیفی - اکتشافی و از منظر طرح تحقیق، پس‌رویدادی است. جامعه آماری این پژوهش، ۶۷۲ شرکت بورسی در اسفندماه سال ۱۴۰۰ بوده که از میان آنها، هفت شرکت^{۲۸} (نمونه آماری) انتخاب شده‌اند. روش انتخاب نمونه به صورت خوشه‌ای یک‌مرحله‌ای و سپس، انتخاب هدفمند یک سهم از داخل هر خوشه است؛ به این صورت که انتخاب خوشه‌ها و نیز انتخاب یک سهم از میان سهم‌های (شرکت‌های) هر خوشه به صورت هدفمند انجام شده است. این سهام‌ها (شرکت و نماد اختصاصی) در جدول (۱) عرضه شده است.

جدول ۱. سهام‌های منتخب			
ردیف	نام شرکت	نماد	صنعت
۱	نفت سپاهان	شسپا	نفت و پتروشیمی
۲	همکاران سیستم	سیستم	خدمات
۳	مخابرات ایران	اخابر	مخابرات
۴	کیمیدارو	دکیمی	دارویی
۵	سیمان شرق	سشرق	سیمان

بهبودسازی پرتفوی سهام با استفاده.../اسفندیاری، کرامتی، غلامی جمکرانی و کاشفی نیشابوری

۶	فولاد مبارکه اصفهان	فولاد	فلزات اساسی
۷	پالایش نفت تهران	شتران	فراورده های نفتی

منبع: تارنمای بورس اوراق بهادار تهران^{۲۹} (اسفندماه ۱۴۰۰)

دوره زمانی مورد بررسی سال های ۱۳۹۶-۱۴۰۰ است و در این دوره زمانی، دو دوره ۴۰ روزه معاملاتی به عنوان بازارهای صعودی و نزولی انتخاب شده اند که به شرح زیر می باشد.

معرفی بازارهای صعودی و نزولی بکار برده شده

در پژوهش حاضر، به جهت بررسی عملکرد دو استراتژی خرید و نگهداری و یادگیری تقویتی، از دو حالت بازار صعودی و نزولی استفاده شده است که مشخصات این دوره های صعودی و نزولی بر اساس بررسی نمودار روند بازار هر هفت سهم (در طی پنج سال) بوده که با انتخاب یک دوره مشترک صعودی و نزولی برای هر هفت سهم، یک دوره شامل ۴۰ روز معاملاتی انتخاب شده است. دلیل استفاده از چنین رویکردی، به دلیل تحلیل نمودارها و مشخص نمودن وضعیت سودآوری هر کدام از سبدها در دوره زمانی یکسان در بازار سرمایه اوراق بهادار بوده است، که بدین وسیله امکان تفسیر نتایج سودآوری سبدها و مقایسه سبدها نیز فراهم شود. علاوه بر این، چنین روندی در بازار با روند شاخص کل نیز تا حد زیادی مطابقت دارد. لذا، با توجه به چنین وضعیتی در نمودارها، با اطمینان از تشابه در روند هر هفت سهم مورد مطالعه در این پژوهش، بازارهای صعودی و نزولی به صورت مشترک برای هر هفت سهم انتخاب شده است.

جدول ۲- مشخصات دوره های صعودی و نزولی مورد استفاده

روند	تاریخ شروع	تاریخ پایان	طول دوره
صعودی	۱۳۹۹/۰۱/۰۶	۱۳۹۹/۰۳/۱۰	۴۰
نزولی	۱۴۰۰/۱۰/۰۵	۱۴۰۰/۱۲/۰۲	۴۰

منبع: یافته های پژوهش

از آنجا که هدف تحقیق، بررسی عملکرد یادگیری تقویتی و مقایسه با استراتژی خرید و نگهداری در بازارهای صعودی و نزولی است، حالت بازار خنثی، از مجموع حالت های مورد استفاده در یادگیری کیو حذف می شود. گفتنی است الگوریتم مورد استفاده و همچنین نرخ کامزد نسبتاً بالا که ۱,۵ درصد است، بخشی از نوسانات بازار را برای الگوریتم یادگیری کیو و معاملات الگوریتمی، غیرقابل استفاده می کند، به این معنی که نوسانات اندک در محدوده ۱,۵ درصد، در عمل موجب می شوند که سیگنال ایجاد شده، نگهداری باشد، چرا که تغییرات قیمت سهام در دامنه مثبت و منفی ۱,۵ درصد در مقایسه با قیمت فعلی

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / دوره ۱۵ / شماره ۵۹ / تابستان ۱۴۰۳

سهام، در مقایسه با یک و پنج دوره آتی، قادر نخواهد بود که هزینه های کارمزد معاملاتی خرید و فروش سهام را جبران نماید و در چنین وضعیتی، سهام خریدنی و یا فروختنی نخواهد بود، چرا که معامله سهام قادر نخواهد بود، هزینه های معاملاتی را پوشش بدهد و موجب می شود که ارزش سبد سهام کاهش پیدا کند که منجر به زیان می شود.

از طرفی مفهوم سیگنال نگهداری، با استراتژی خرید و نگهداری، یکسان بوده و در عمل به محقق اجازه مقایسه و بررسی بین دو استراتژی را در حالتی متمایز را نمی دهد. به همین دلیل جهت حذف چالش ایجاد شده از ناحیه نوسانات در قیمت روزانه سهام، حالت بازار خنثی که نوسانات بازار قابل توجه نمی باشند و در نتیجه، سیگنال ایجاد شده برای خودمعامله گر، نگهداری خواهد بود، از فرض مدل تحقیق حذف شده است، چرا که در حالت بازار خنثی، نتایج ایجاد شده با نتایج استراتژی خرید و نگهداری بسیار شباهت می یابد و دلیل این اتفاق، میزان قابل توجه کارمزدهای معاملاتی می باشد.

آماده سازی و پردازش داده ها

مراحل انجام شده جهت آماده سازی فایل اکسل جهت کاربرد در نرم افزار متلب به شرح زیر می باشد:

۱- دریافت فایل اکسل داده های معاملاتی شرکت انتخابی که به صورت اتفاقی ساده در هر گروه صنعتی انتخاب می شود.

۲- بررسی داده های معاملاتی موجود در بازه مورد بررسی (ابتدای سال ۱۳۹۶ الی اسفند ۱۴۰۰)

۳- در صورت موجود بودن داده ها، حضور شرکت در نمونه تحقیق تایید می شود، و در غیر این صورت، به صورت اتفاقی، شرکت بازار سرمایه دیگری در همان گروه صنعتی انتخاب و مورد بررسی قرار می گیرد، تا در نهایت، یک شرکت عضو بازار سرمایه در گروه صنعتی مورد نظر انتخاب شود.

۴- این اقدامات تکرار خواهد شد، تا در نهایت ۷ شرکت بازار سرمایه مورد نظر انتخاب شود.

۵- علاوه بر شرکت های عضو بازار سرمایه، بررسی شاخص کل بازار سرمایه نیز در تحقیق حاضر اهمیت دارد و داده های سالیانه شاخص کل بازار سرمایه در سال های ۱۳۹۶ الی ۱۴۰۰ نیز از تارنمای بازار سرمایه اوراق بهادار دریافت می شود.

۶- داده های تهیه شده شرکت های بازار سرمایه، در دوره های زمانی متفاوتی قرار دارند که می بایست موارد اضافی از این فایل ها حذف شود.

۶- داده های تهیه شده، روزانه می باشد و چیدمان داده های نیز از جدید به قدیم می باشد که می بایست چیدمان به حالت قدیم به جدید تغییر داده شوند که یک نیازمند صرف وقت طولانی جهت انجام این

بهبودسازی بر تقوی سهام با استفاده... / اسفندیاری، کرامتی، غلامی جمکرانی و کاشفی نیشابوری

اقدام است و در این مرحله داده‌های معاملاتی اضافه حذف و چیدمان داده‌ها نیز از قدیم به جدید اصلاح شده است.

۷- پس از آماده‌سازی اولیه، لازم است نسخه دیگری از داده‌های معاملاتی بر اساس دوره زمانی ماهیانه از داده‌های روزانه تهیه شود، به این صورت که قیمت سهام در انتهای هر ماه، به عنوان قیمت ماهیانه سهام منظور و فایل‌های جدیدی تهیه می‌شود، که این فایل مبنای محاسبات ریسک و بازده در طول دوره ۵ ساله مورد بررسی می‌باشد.

۸- پس از ترسیم نمودار عملکرد بازار سهام و شاخص کل بر اساس داده‌های ماهیانه و بازده سالیانه، نسبت به تحلیل روند بازار و انتخاب دو دوره صعودی و نزولی اقدام شده است.

۹- داده‌های معاملاتی روزانه، در دوره‌های صعودی و نزولی در فایل جداگانه تهیه می‌شود، تا از این فایل جهت فراخوانی در نرم‌افزار متلب استفاده شود.

۱۰- با استفاده از فایل‌هایی که در خصوص دوره‌های صعودی و نزولی تهیه شده است، و شامل داده‌های معاملاتی برای ۴۵ روز معاملاتی متوالی می‌باشد، محاسبات بازده رویکرد خرید و نگهداری برای دوره ۴۰ روزه انجام شده است (استفاده از ۵ روز اضافی جهت پیش بینی قیمت آتی سهام و ایجاد سیگنال معاملاتی کاربرد دارد).

همان‌طور که ملاحظه می‌شود به دلیل عدم امکان استفاده مستقیم از داده‌های تهیه شده از وب‌سایت بورس اوراق بهادار تهران، لازم است تا داده‌های اولیه با دقت مورد آماده‌سازی پردازش قرار گرفته و بخشی از محاسبات در داخل نرم‌افزار اکسل کدنویسی شده است و بخش دیگر که مربوط به شبیه‌سازی سبد و اجرای رویکرد یادگیری تقویتی می‌باشد در نرم‌افزار متلب کدنویسی شده است.

تشریح عملکرد خودمعامله‌گر مورد استفاده در تحقیق

سیستم خودمعامله‌گر در نرم‌افزار متلب کدنویسی شده است و براساس داده‌های بورسی پردازش شده در نرم‌افزار اکسل، محیط معاملاتی را شبیه‌سازی و درخصوص سیگنال‌دهی برای انجام معاملات اقدام می‌کند. روش اجرایی خودمعامله‌گر بدین شکل است که فایل اکسل مربوط به دوره معاملاتی را فرامی‌خواند و داده‌های قیمت سهام را وارد حافظه خودمعامله‌گر می‌کند که یک ماتریس به صورت 1×45 (در نظر داشتن پنج دوره مازاد بر دوره موردنظر) است. ماتریس قیمت‌های سهام، در نقش داده‌های قیمت سهام در بازار واقعی عمل می‌کند و خودمعامله‌گر، قادر است با کمک تابع مطلوبیتی که برای خودمعامله‌گر طراحی شده است، قیمت‌های روز جاری و قیمت‌های آتی را که توسط الگوریتم مورد

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / دوره ۱۵ / شماره ۵۹ / تابستان ۱۴۰۳

بررسی قرار می‌گیرد، درک نموده به تصمیم‌گیری برای سیگنال‌دهی بپردازد. در نهایت، داده‌های معاملاتی خودمعموله‌گر در قالب فایل اکسل ثبت می‌شود تا بتوان با تحلیل خروجی آن، تفسیر بهتری از روند معاملات و تصمیم‌گیری الگوریتم به‌دست آورد. محاسبات تغییرات رشد یا بازده سهام، از طریق فرمول زیر به‌دست می‌آیند.

$$R_n = (P_{t+n} - P_t) / P_t \quad \text{رابطه (۱)}$$

معرفی الگوریتم سیگنال‌دهی

الگوریتم سیگنال‌دهی، براساس مقایسه نرخ بازده آتی محاسبه و ایجاد می‌شود. در این پژوهش، نرخ بازده آتی، حاصل بررسی و مقایسه یک دوره آتی و پنج دوره آتی است. پس از عرضه سیگنال اولیه، الگوریتم یادگیری تقویتی، بر مبنای معادله بلمن، سودآوری تصمیم اخذشده را با سایر گزینه‌های محتمل قابل انتخاب، سنجیده و مقدار محاسبه‌شده جدید را در ماتریس کیو، اصلاح می‌نماید. در مدل طراحی‌شده، روند تغییرات رشد روزانه محاسبه‌شده است که به محاسبه میانگین تغییرات رشد منجر می‌شود که همان بازده سهام طی دوره مورد بررسی است.

الگوریتم سیگنال‌دهی، پنج سیگنال خرید قوی، خرید ضعیف، فروش قوی، فروش ضعیف و عدم‌معامله را ایجاد می‌کند که در حالت قوی، ۱۰۰ درصد و در حالت ضعیف، ۵۰ درصد سرمایه نقدی و یا سهام موجود در سبد برای معاملات خرید و یا فروش استفاده می‌کنند. البته می‌توان دقت پیشنهادات را در حوزه خرید و فروش، افزایش داد، و از درصدهای مختلفی برای خرید و فروش استفاده نمود. که در این تحقیق، به جهت سادگی منطقی، سیگنال فروش قوی، به معنای فروش ۱۰۰ درصد سهام، و سیگنال ضعیف فروش، به معنای فروش ۵۰ درصد سهام، استفاده شده است. در سیگنال خرید قوی، پیشنهاد خرید با استفاده از تمام موجودی پول نقد، و در پیشنهاد خرید ضعیف، خرید با ۵۰ درصد پول نقد موجود، مد نظر می‌باشد. در حالی که سیگنال، عدم خرید - فروش، تنها به یک سیگنال محدود می‌شود. به عبارتی در یک مدل ساده‌تر، می‌توان سه سیگنال متمایز به صورت، خرید، فروش، نگهداری را انتظار داشت، و با افزایش دقت سیگنال‌ها، می‌توان سطوح اقدام متناظری با درصد مورد نظر برای خرید و فروش را شاهد بود.

الگوریتم یادگیری کیو و معادله بلمن

یادگیری تقویتی براساس معادله بلمن^{۳۰} بنا شده است،

رابطه (۲)

بهبودسازی پرتفوی سهام با استفاده.../اسفندیاری، کرامتی، غلامی جمکرانی و کاشفی نیشابوری

در رابطه (۲) متغیر نرخ یادگیری α ، تعیین می‌کند که تا چه میزان اطلاعات جدید بر اطلاعات قدیمی ترجیح داده شود. مقدار صفر باعث می‌شود که عامل، چیزی یاد نگیرد و مقدار یک باعث می‌شود که عامل فقط اطلاعات جدید را ملاک قرار دهد. همچنین، متغیر نرخ تنزیل (گاما) γ اهمیت پاداش‌های آینده را تعیین می‌کند. مقدار صفر باعث می‌شود که عامل، ماهیت فرصت طلبانه گرفته و فقط پاداش‌های فعلی را مدنظر قرار دهد؛ درحالی که مقدار یک عامل را ترغیب می‌کند، برای یک دوره زمانی طولانی برای پاداش تقلا کند.

مدیریت پویای سبد سهام به روش یادگیری تقویتی

از آنجا که با افزایش میزان تنوع دارایی‌های سبد، تعداد ترکیب‌های احتمالی سبد و اقدامات به شدت افزایش می‌یابد، و ترکیب دارایی‌های مختلف در سبد، رو به ازدیاد می‌گذارد که موجب پیچیده‌تر شدن عملیات می‌شود، که در نتیجه به زمان بیشتر برای محاسبات نیاز می‌باشد، یا می‌بایست از سیستم‌های کامپیوتری قوی‌تری استفاده نمود، تا عملیات محاسبات در زمان مناسب به اتمام برسد.

در این تحقیق از آنجا که هدف بررسی عملکرد منطقی مدل می‌باشد، به سبب از دارایی‌ها، مشتمل بر پول نقد و یک نوع سهام، اکتفا می‌شود، زیرا افزایش تعداد دارایی، تاثیری بر منطق عملیاتی مدل و عملکرد قابل انتظار در مدل ندارد، و موجب می‌شود که زمان اجرای عملیات افزایش پیدا کند و تحلیل و تفسیر نتایج نیز فرآیندی سخت و زمان‌بر و محتمل به بروز خطا بشود. آنچه اهمیت دارد، رویکردی است که مدل در شرایط مختلف محیط اتخاذ می‌کند و در نتیجه، اثراتی بر عملکرد سبد ایجاد می‌شود که در این تحقیق، باهدف سنجش عملکرد معاملات الگوریتمی که به روش یادگیری تقویتی انجام می‌شوند، می‌باشد و هدف این نیست که سبدهایی با دارایی‌های متنوع، موجب کندی روند اجرای تحقیق و پیاده سازی مدل بشویم. در نتیجه به حداقل‌های منطقی که تغییری در نتایج ارزیابی نمی‌گذارند، اکتفا می‌نماییم. حالت شروع، در الگوریتم به صورت پیش فرض، با خرید سهم به میزان تمام دارایی نقدی موجود در سبد تنظیم شده است، و در هر دو بازار صعودی و نزولی در ابتدای اجرای مدل، سرمایه نقدی به سهام تبدیل می‌شود.

در محاسبات خرید، پارامترهای قیمت سهام، کارمزد یک درصدی برای خرید، و خرید تعداد سهمی که به صورت عددی صحیح می‌باشند، لحاظ شده است، به این معنی که در محاسبات ریاضی، ارقامی که کم‌تر از ۱۰۱ درصد از قیمت خرید یک سهم در زمان معامله می‌باشند، به عنوان مانده حساب نقدی، که قابلیت خرید یک سهم به صورت کامل و پرداخت کارمزد یک درصدی را ندارند، در حساب باقی می‌ماند.

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / دوره ۱۵ / شماره ۵۹ / تابستان ۱۴۰۳

چنین تصمیمی در ابتدای اجرای الگوریتم، موجب می‌شود که هر دو استراتژی (خرید و نگهداری، یادگیری تقویتی) در اولین قدم، شروعی مشترک را داشته باشند و بتوان در بررسی‌ها، اختلاف‌ها در عملکرد سبد دارایی (سهام + پول نقد) را ناشی از اقداماتی دانست که در ادامه مسیر توسط استراتژی یادگیری تقویتی اجرا شده است. چنین تصمیمی برای شروع، موجب می‌شود که مقایسات تا حد زیادی به یک مقایسه واقعی و شرایط منطقی نزدیک‌تر شده و از خطاهای انسانی در شروع معاملات در مدل مورد بررسی فاصله بگیریم.

ماتریس حالت-عمل

دو رکن سازنده ماتریس حالت-عمل که از آن به جدول کیو^{۳۱} نیز یاد می‌شود، عبارتند از:

- حالت، ردیف را در ماتریس تعیین می‌کند (بازار صعود، بازار نزولی)
- عمل، ستون‌های ماتریس را تعیین می‌کند (پنج سیگنال معاملاتی خرید قوی، خرید ضعیف، نگهداری، فروش قوی، فروش ضعیف)

حالت‌ها

حالت‌های مطرح برای بازار سهام، سه حالت می‌باشد که از میان این حالت‌ها که صعودی، نزولی و خنثی می‌باشند (در مدل تحقیق حاضر، حالت خنثی حذف شده است)

اقدامات

سیستم خودمعامله‌گر، یا الگوریتم معاملاتی، جهت انجام هر معامله، نیازمند تصمیم‌گیری، و سپس ایجاد سیگنال مربوطه و در ادامه اجرای آن اقدام یا سیگنال می‌باشد. تعداد تصمیم‌های مجاز در این مدل، به پنج تصمیم یا سیگنال به شرح زیر محدود شده است.

جدول ۳- معرفی سیگنال‌های معاملاتی^{۳۲}

سیگنال	رقم سیگنال	توضیحات
خرید (قوی)	۱	خرید با استفاده از ۱۰۰ درصد نقدینگی موجود
خرید (ضعیف)	۰,۵	خرید با استفاده از ۵۰ درصد نقدینگی موجود
فروش (قوی)	-۱	فروش ۱۰۰ درصد سهام موجود در سبد
فروش (ضعیف)	-۰,۵	فروش ۵۰ درصد سهام موجود در سبد
نگهداری	۰	هیچ معامله خرید یا فروشی انجام نمی‌شود

با توجه به پنج سیگنال مورد توجه در این تحقیق و استفاده از دو حالت بازار صعودی و بازار نزولی، می‌توان ماتریس کیو یا ماتریس حالت عمل را به صورت زیر طراحی نمود.

بهبودسازی پرتفوی سهام با استفاده.../اسفندیاری، کرامتی، غلامی جمکرانی و کاشفی نیشابوری

جدول ۴- معرفی ماتریس اولیه حالت-عمل (کیو)

عمل حالت	خرید - قوی	خرید - ضعیف	فروش - ضعیف	فروش - قوی	نگهداری
صعودی
نزولی
خنثی

ماتریس حالت-عمل ارائه شده در بالا، دارای سه حالت می‌باشد، که در حالت در سطرهای ماتریس قرار گرفته است، که طبق آنچه توضیح داده‌ایم، حالت خنثی را حذف می‌کنیم و به یک ماتریس دو در پنج به صورت زیر می‌رسیم که همان ماتریسی است که در محاسبات مورد استفاده قرار گرفته است و در خصوص دلایل حذف حالت خنثی، در بخش قبلی، توضیح کامل ارائه شده است.

تفسیر کلی ارقام ماتریس حالت-عمل

جدول ۵- تفسیر کلی ماتریس حالت-عمل

رقم	تفسیر
صفر	معنای رقم ۰، عدم وقوع معامله در این حالت-عمل خاص، تا این زمان است. و این به معنای این است که عامل در خصوص چنین ترکیبی از حالت-عمل، هیچ تجربه‌ای کسب نکرده است و شرایط کاملاً جدید می‌باشد.
مثبت	ارقام مثبت به معنای این است که عامل در این شرایط تصمیمات قبلی را که اخذ نموده است، منتهی به پاداش بوده است و عملکرد رضایت بخش بوده است.
منفی	ارقام منفی، مبین این است که اقدامات قبلی عامل در چنین ترکیبی از حالت-عمل، منجر به پاداش منفی (جریمه) شده است و عامل نمی‌بایست، چنین عملی را اجرا کند، مگر این که اقدام جدید بتواند تغییر در جهت مثبت را در رقم فعلی موجب شود.

با خواندن رقم موجود در ماتریس کیو، سه احتمال اصلی در خصوص رقم موجود در این خانه از ماتریس وجود دارد، که در جدول بالا معرفی شده است.

- خود معامله‌گر، رقم اولیه را می‌خواند و به حافظه منتقل می‌کند،
- معامله جدید را طبق همان تصمیم گرفته شده محاسبه می‌کند، و پاداش جدید را بدست می‌آورد.
- محاسبات ماتریس حالت-عمل را از طریق معادله بلمن انجام می‌دهد، و مقدار جدید را برای ماتریس حالت-عمل در محل مربوطه در ماتریس جایگزین می‌نماید.
- نتیجه یادگیری تقویتی، و محاسبات معادله بلمن، ایجاد یک ماتریس حالت-عمل مناسب و

کارا می‌باشد که عامل می‌تواند بر اساس پاداش‌های گذشته، از آن برای تصمیم‌گیری‌ها در خصوص عملی که باید در حالت خاصی انجام بدهد، از مقادیر آن استفاده و تصمیم‌گیری نماید.

الگوریتم کیو عمیق

اگر الگوریتم یادگیری کیو را به‌روزرسانی اعداد موجود در یک ارائه دو بعدی (شامل: فضای اقدام×فضای وضعیت) در نظر بگیرید، متوجه شباهت آن با برنامه‌نویسی پویا خواهید شد. این موضوع برای ما روشن می‌سازد که وقتی عامل تصمیم‌گیرنده در الگوریتم یادگیری کیو با وضعیتی کاملاً جدید روبه‌رو شود، هیچ راهی برای شناسایی و انتخاب اقدام مناسب نخواهد داشت. به عبارت دیگر، عامل تصمیم‌گیرنده الگوریتم یادگیری کیو توانایی تخمین ارزش وضعیت‌های ناشناخته را ندارد.

برای حل این مشکل، شبکه DQN ارائه دو بعدی را حذف و شبکه عصبی را جایگزین آن می‌کند. شبکه DQN به کمک یک شبکه عصبی، تابع Q-value را تخمین می‌زند. وضعیت فعلی به عنوان ورودی به این شبکه داده می‌شود،

سپس مقدار Q-value متناظر با هر اقدام به‌عنوان خروجی از شبکه دریافت خواهد شد. (هوشیو، ۱۴۰۰) لذا، در روش کیو عمیق که به آن شبکه کیو نیز گفته می‌شود، ماتریس کیو حذف و بجای آن خروجی‌های یک شبکه عصبی مصنوعی جایگزین می‌شود. همان‌طور که در معادله مربوط به دو الگوریتم ملاحظه می‌شود، تنها تفاوت در بخش محاسبه وضعیت ناشی از سیاست در معادله می‌باشد. اقدام، اقدامی است که در وضعیت بعدی یعنی تحت سیاست فعلی انجام خواهد گرفت.

در این رابطه، متغییر نرخ یادگیری α ، تعیین می‌کند که تا چه میزان اطلاعات به دست آمده جدید بر اطلاعات قدیمی ترجیح داده شود. مقدار صفر باعث می‌شود که عامل چیزی یاد نگیرد و مقدار یک باعث می‌شود که عامل فقط اطلاعات جدید را ملاک قرار دهد. همچنین متغییر نرخ تنزیل γ ، اهمیت پاداش‌های آینده را تعیین می‌کند. مقدار صفر باعث می‌شود که عامل ماهیت فرصت‌طلبانه گرفته و فقط پاداش‌های فعلی را مدنظر قرار دهد. در حالی که مقدار یک عامل را ترغیب می‌کند، برای یک دوره زمانی طولانی برای پاداش تقلا کند.

یادگیری کیو و کیو عمیق، دو الگوریتم محبوب و مستقل از مدل برای یادگیری تقویتی هستند. تمایز این الگوریتم‌ها با یکدیگر در استراتژی‌های جست‌وجوی آن‌ها محسوب می‌شود در حالی که استراتژی‌های استخراج آن‌ها مشابه است. در حالی که یادگیری کیو، یک روش مستقل از سیاست است که در آن عامل ارزش‌ها را براساس عمل (a) که از سیاست دیگری مشتق شده می‌آموزد، کیو عمیق،

بهبودسازی پرتفوی سهام با استفاده.../اسفندیاری، کرامتی، غلامی جمکرانی و کاشفی نیشابوری

یک روش مبتنی بر سیاست محسوب می‌شود که در آن ارزش‌ها را براساس عمل کنونی (a) که از سیاست کنونی آن مشتق شده می‌آموزد. پیاده‌سازی این دو روش آسان است اما فاقد تعمیم‌پذیری هستند زیرا دارای توانایی تخمین ارزش‌ها برای حالت‌های مشاهده نشده نیستند.

یافته‌های پژوهش

در این بخش بازده روزانه هفت سهم مورد استفاده در پژوهش در بازار صعودی و نزولی بطور مجزا مورد بررسی قرار گرفته است و سه پارامتر عمده که در شناخت سهم و سبد سهام اهمیت دارند، به ترتیب واریانس، ریسک (انحراف معیار پرتفوی) و بازده سهام، در جداول ۴ و ۵، برای هر کدام از هفت سهم و شاخص کل بازار سرمایه محاسبه شده است. تحلیل داده‌های بازار سرمایه روزانه که در این بخش انجام شده است، درک عمیق از وضعیت بازار و بازدهی استراتژی‌های خرید و نگهداری و معاملات الگوریتمی به روش یادگیری تقویتی و رویکرد کیو، را مقدور می‌کند.

بررسی بازده روزانه سهام در بازار صعودی

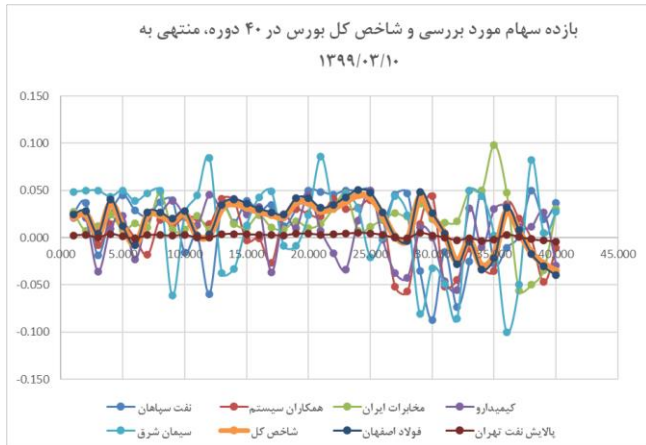
در این بخش، هر هفت سهم و شاخص کل بازار سرمایه در ۴۰ دوره منتهی به تاریخ ۱۳۹۹/۰۳/۱۰، مورد بررسی و تحلیل قرار گرفته است. چنانچه نمودار ۴ نشان می‌دهد گرچه دوره انتخاب شده، یک دوره صعودی می‌باشد اما در بازارهای صعودی نیز تلاطم در تغییرات بازده به خوبی ملاحظه است.

جدول ۶- داده‌های سهام مورد بررسی در ۴۰ دوره منتهی به ۱۳۹۹/۰۳/۱۰

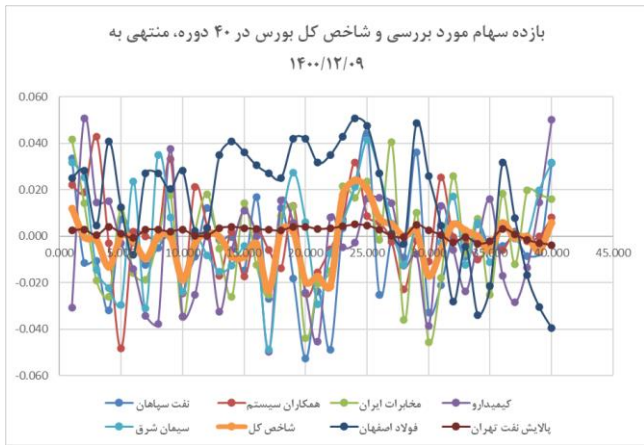
	نام شرکت	نماد	واریانس	ریسک	بازده
۱	نفت سپاهان	شسپا	۰,۰۰۱۲۵۳	۰,۰۳۵۴۰۰	۰,۰۱۶۲۳۵
۲	همکاران سیستم	سیستم	۰,۰۰۰۸۴۹	۰,۰۲۹۱۳۳	۰,۰۰۵۷۲۸
۳	مخابرات ایران	اخبر	۰,۰۰۰۶۵۰	۰,۰۲۵۴۸۸	۰,۰۱۸۴۵۹
۴	کیمیدارو	دکیمی	۰,۰۰۰۸۱۰	۰,۰۲۸۴۶۳	۰,۰۰۷۷۴۹
۵	سیمان شرق	سشرق	۰,۰۰۲۱۸۶	۰,۰۴۶۷۵۸	۰,۰۱۴۲۳۷
۶	شاخص کل	-	۰,۰۰۰۴۶۸	۰,۰۲۱۶۳۹	۰,۰۱۴۷۰۶
۷	فولاد اصفهان	فولاد	۰,۰۰۱۲۵۳	۰,۰۵۵۴۰۰	۰,۰۱۸۲۳۵
۸	پالایش نفت تهران	شتران	۰,۰۰۰۸۴۹	۰,۰۴۹۱۳۳	۰,۰۰۶۷۲۸

(منبع: یافته‌های محقق)

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / دوره ۱۵ / شماره ۵۹ / تابستان ۱۴۰۳



نمودار ۲: نمودار بازده روزانه سهام‌های مورد بررسی در بازار
صعودی (منبع: محقق)



نمودار ۱: نمودار بازده روزانه سهام مورد بررسی در بازار نزولی
(منبع: محقق)

بررسی بازده روزانه سهام در بازار نزولی

در بخش قبلی، تحلیل بهینه‌سازی پرتفوی سهام، دوره انتخابی، یک دوره با روند صعودی یا به عبارتی یک بازار صعودی بوده است. به جهت بررسی کارایی استراتژی‌های معاملاتی مورد بررسی، در این بخش دوره‌ای انتخاب شده است که بازار نزولی باشد. در این بخش، هر هفت سهم و شاخص کل بازار سرمایه در ۴۰ دوره منتهی به تاریخ ۱۴۰۰/۱۲/۰۲، مورد بررسی و تحلیل قرار گرفته است.

جدول ۷- داده‌های سهام مورد بررسی در ۴۰ دوره منتهی به ۱۴۰۰/۱۲/۰۲

بازده	ریسک	واریانس	نماد	نام شرکت	
-۰,۰۰۴۷۳۰	۰,۰۲۰۵۵۳	۰,۰۰۰۴۲۲	شسپا	نفت سپاهان	۱
۰,۰۰۰۲۹۸	۰,۰۱۷۱۶۸	۰,۰۰۰۲۹۵	سیستم	همکاران سیستم	۲
-۰,۰۰۱۳۱۵	۰,۰۲۲۱۴۱	۰,۰۰۰۴۹۰	اخبر	مخابرات ایران	۳
-۰,۰۰۴۹۳۹	۰,۰۲۴۰۸۳	۰,۰۰۰۵۸۰	دکیمی	کیمیدارو	۴
۰,۰۰۰۸۵۲	۰,۰۲۰۳۸۰	۰,۰۰۰۴۱۵	سشرق	سیمان شرق	۵
-۰,۰۰۱۸۶۴	۰,۰۱۰۵۲۹	۰,۰۰۰۱۱۱	-	شاخص کل	۶
۰,۰۱۸۲۳۵	۰,۰۵۵۴۰۰	۰,۰۰۱۲۵۳	فولاد	فولاد اصفهان	۷
۰,۰۰۶۷۲۸	۰,۰۴۹۱۳۳	۰,۰۰۰۸۴۹	شتران	پالایش نفت تهران	۸

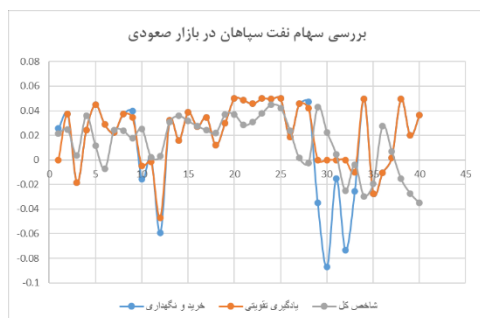
(منبع: یافته‌ها محقق)

بهبودسازی پرتفوی سهام با استفاده.../اسفندیاری، کرامتی، غلامی جمکرانی و کاشفی نیشابوری

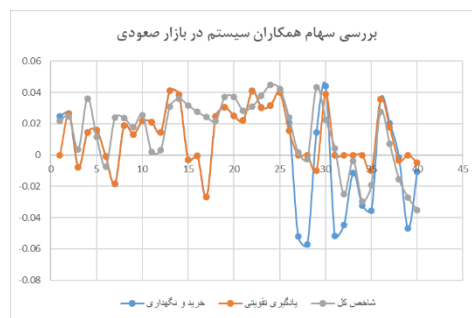
بررسی عملکرد استراتژی یادگیری تقویتی در مقایسه با خرید و نگهداری

بررسی عملکرد سهام در بازار سعودی

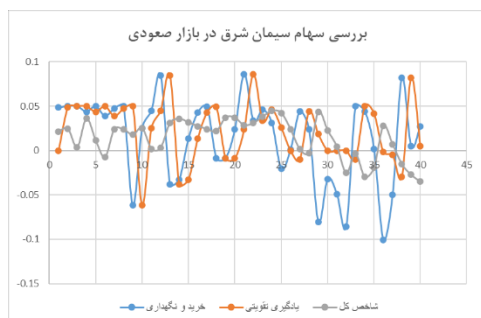
در این بخش عملکرد هفت سهم مورد مطالعه در پژوهش در بازار سعودی، بر اساس استراتژی های خرید و نگهداری و یادگیری تقویتی، مورد بررسی قرار گرفته است، که نتایج این بررسی ها و همچنین عملکرد شاخص کل برای هر کدام از سهم های مورد بررسی در نموداری مجزا ارائه شده است. بر اساس نمودارهای شماره (۳) الی (۷)، همان طور که ملاحظه می شود، روش یادگیری تقویتی، توانسته است، در بیشتر موارد ارزش خود را در محدوده بدون زیان و حتی سودآوری حفظ نماید و عملکردی بهتر از استراتژی خرید و نگهداری ایجاد کند. در نمودارها به خوبی مشخص است که حتی در بازارهای سعودی نیز نوسانات قیمت موجب می شوند که نرخ رشد تغییرات قیمت موجب ایجاد رقم منفی در روند رشد شود که روش یادگیری تقویتی با استفاده از سیستم سیگنال دهی توانسته است در کاهش این وضعیت موفقیت نسبی بدست آورد و این به معنی بهبود در سودآوری سبد می باشد.



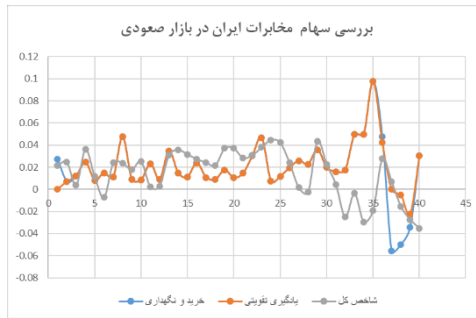
نمودار ۴: مقایسه عملکرد سهام نفت سپاهان در بازار سعودی (منبع: محقق)



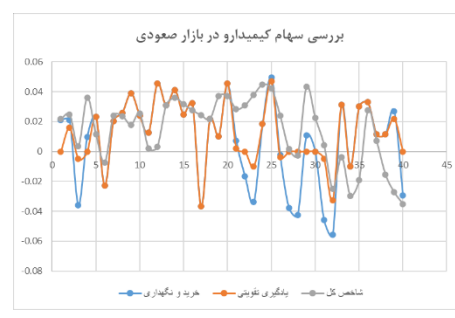
نمودار ۳: مقایسه عملکرد سهام همکاران سیستم در بازار سعودی (منبع: محقق)



نمودار ۷: مقایسه عملکرد سهام سیمان شرق در بازار سعودی (منبع: محقق)



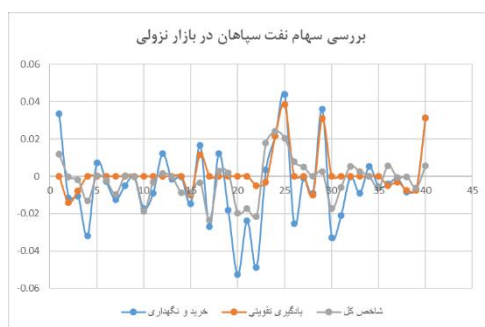
نمودار ۶: مقایسه عملکرد مخابرات ایران در بازار سعودی (منبع: محقق)



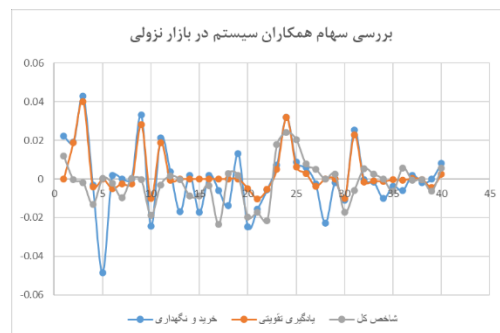
نمودار ۵: مقایسه عملکرد سهام کیمیدارو در بازار سعودی (منبع: محقق)

بررسی عملکرد سهام در بازار نزولی

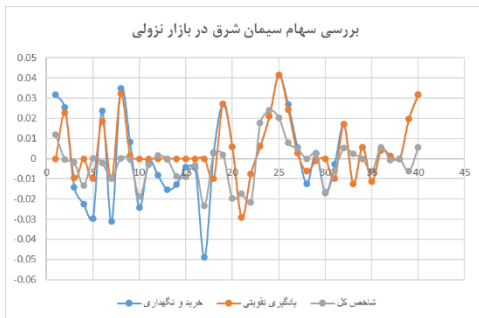
در بررسی‌های انتهای دوره مورد بررسی، عملکرد استراتژی‌های خرید و نگهداری و یادگیری تقویتی در بازار نزولی، همان‌طور که ملاحظه می‌شود، روش یادگیری تقویتی، توانسته است در بیشتر موارد ارزش خود را در محدوده بدون زیان و حتی سودآوری حفظ نماید. از آنجا که نمودارهای بررسی عملکرد بر اساس تغییرات قیمت ترسیم شده‌اند، همان‌طور که در نمودارها مشخص است، در هر دو حالت بازار صعودی و نزولی، نوسانات قیمت دیده می‌شوند و وجود نوسانات در قیمت مهم‌ترین دلیل برای موفقیت‌های معاملات الگوریتمی می‌باشد، چرا که فرصت معامله و کسب سود را فراهم می‌کند.



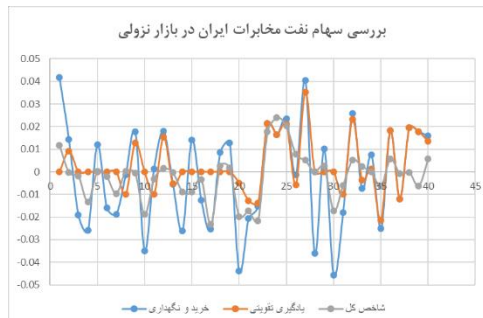
نمودار ۹: مقایسه عملکرد سهام نفت سپاهان در بازار نزولی (منبع: محقق)



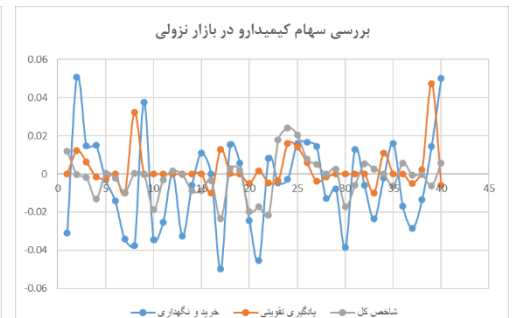
نمودار ۸: مقایسه عملکرد سهام همکاران سیستم در بازار نزولی (منبع: محقق)



نمودار ۱۲: مقایسه عملکرد سهام سیمان شرق در بازار نزولی (منبع: محقق)



نمودار ۱۱: مقایسه عملکرد سهام مخابرات ایران در بازار نزولی (منبع: محقق)



نمودار ۱۰: مقایسه عملکرد سهام کیمیدارو در بازار نزولی (منبع: محقق)

ماتریس حالت-عمل

در این بخش، ارقام محاسبه شده برای هر یک از سهام مورد بررسی (بر اساس قیمت روزانه هر سهم

بهبوده سازی پرتفوی سهام با استفاده.../اسفندیاری، کرامتی، غلامی جمکرانی و کاشفی نیشابوری

و سپس تشکیل سبد سهام) در قالب ماتریس های حال-عمل (ماتریس کیو)، برای دو حالت بازار صعودی و نزولی و پنج عمل ممکن در هر حالت ارائه شده است. نتایج ارائه شده در ماتریس های حال-عمل در این بخش، با استفاده از معادلات روش کیو عمیق محاسبه شده است.

جدول ۸- ماتریس حالت-عمل کیو (الگوریتم کیو عمیق) - شرکت نفت سپاهان(منبع: محقق)

نگهداری	فروش- قوی	فروش-ضعیف	خرید - ضعیف	خرید - قوی	عمل حالت
۰,۰۰۰۰۰۰۰۰	۰,۰۵۲۴۳۲۱۶۳	۰,۰۴۳۸۵۶۴۲۳	۰,۰۷۹۶۱۴۲۱	۰,۰۳۷۰۳۹۰۴۸	صعودی
۰,۰۱۴۵۵۰۲	۰,۰۱۵۵۷۱۱۱۴	۰,۰۲۹۳۹۷۸۶	۰,۰۰۸۰۰۳۲۶	۰,۰۳۱۵۰۰۴۱۴	نزولی

جدول ۹- ماتریس حالت-عمل کیو (الگوریتم کیو عمیق) - شرکت همکاران سیستم(منبع: محقق)

نگهداری	فروش- قوی	فروش-ضعیف	خرید - ضعیف	خرید - قوی	عمل حالت
۰,۰۱۵۶۸۵۱۷	۰,۰۲۱۸۴۳۸۷	۰,۰۱۰۰۴۲۸۵	۰,۰۱۴۷۱۲۵۵	۰,۰۳۰۷۰۴۳۸	صعودی
۰,۰۰۳۳۷۳۶۹	۰,۰۱۰۲۱۴۷۰	۰,۰۰۴۲۹۱۵۵	۰,۰۰۳۰۹۹۴۴	۰,۰۰۳۰۸۴۱۴	نزولی

جدول ۱۰- ماتریس حالت-عمل کیو (الگوریتم کیو عمیق) - شرکت مخابرات ایران(منبع: محقق)

نگهداری	فروش- قوی	فروش-ضعیف	خرید - ضعیف	خرید - قوی	عمل حالت
۰,۰۰۰۰۰۰۰۰	۰,۰۰۰۰۰۰۰۰	۰,۰۲۹۳۵۸۶۱	۰,۰۳۴۸۱۹۱۸	۰,۰۴۳۵۰۰۰۹	صعودی
۰,۰۰۲۳۳۹۶۴	۰,۰۱۹۲۰۰۲۷	۰,۰۱۰۱۹۹۸۴	۰,۰۱۱۷۷۰۳۹	۰,۰۱۷۴۶۹۶۵	نزولی

جدول ۱۱- ماتریس حالت-عمل کیو (الگوریتم کیو عمیق) - شرکت کیمیدارو(منبع: محقق)

نگهداری	فروش- قوی	فروش-ضعیف	خرید - ضعیف	خرید - قوی	عمل حالت
۰,۰۵۳۶۲۴۳۲	۰,۰۳۴۷۲۷۱۷	۰,۰۳۶۹۰۶۵۸	۰,۰۶۲۱۲۵۹۰	۰,۰۴۳۵۳۴۲۱	صعودی
۰,۰۰۵۳۶۷۳۵	۰,۰۴۳۶۹۸۶۳	۰,۰۱۰۷۶۹۴۴	-۰,۰۰۳۶۹۰۲۴	۰,۰۰۶۰۳۶۰۶	نزولی

جدول ۱۲- ماتریس حالت-عمل کیو (الگوریتم کیو عمیق) - شرکت سیمان شرق (منبع: محقق)

نگهداری	فروش- قوی	فروش-ضعیف	خرید - ضعیف	خرید - قوی	عمل حالت
۰,۰۰۰۰۰۰۰۰	۰,۰۴۴۳۶۳۹۱	۰,۰۱۸۹۰۶۷۳	۰,۰۶۳۵۱۷۱۴	۰,۰۱۵۴۹۹۴۶	صعودی
۰,۰۱۲۳۹۸۳۶	۰,۰۲۰۳۳۳۹۱	۰,۰۰۳۷۸۶۵۸	۰,۰۱۱۵۲۶۹۰	۰,۰۳۷۷۸۳۷۵	نزولی

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / دوره ۱۵ / شماره ۵۹ / تابستان ۱۴۰۳

جدول ۱۳- ماتریس حالت-عمل کیو (الگوریتم کیو عمیق) - شرکت پالایش نفت تهران (منبع: محقق)

عمل حالت	خرید - قوی	خرید - ضعیف	فروش - ضعیف	فروش - قوی	نگهداری
صعودی	۰,۰۳۹۳۰۹۲۴	۰,۰۲۷۴۱۷۰۲	۰,۰۲۰۸۰۳۱۲	۰,۰۰۰۰۰۰	۰,۰۰۰۰۰۰
نزولی	۰,۰۱۹۳۵۸۳۷	۰,۰۰۱۳۲۸۳۹	۰,۰۲۲۶۱۵۳۳	۰,۰۰۸۸۷۱۶۸	۰,۰۰۷۳۵۰۶۷

جدول ۱۴- ماتریس حالت-عمل کیو (الگوریتم کیو عمیق) - شرکت فولاد اصفهان (منبع: محقق)

عمل حالت	خرید - قوی	خرید - ضعیف	فروش - ضعیف	فروش - قوی	نگهداری
صعودی	۰,۵۰۹۸۲۶۰۴	۰,۵۰۸۷۲۶۳۸	۰,۵۰۷۸۲۲۴۳	۰,۵۰۰۰۰۰۰۰	۰,۵۰۰۰۰۰۰۰
نزولی	۰,۵۰۶۰۵۷۹۰	۰,۵۰۱۵۳۹۱۹	۰,۵۰۱۵۳۹۱۹	۰,۵۰۲۹۹۴۶۲	۰,۵۰۲۷۹۸۲۸

در خصوص سهام شرکت نفت سپاهان، همانطور که در جدول (۸) قابل ملاحظه است، در روش کیو عمیق، در بازارهای صعودی، بزرگ‌ترین پاداش کسب شده، از عمل خرید-ضعیف می‌باشد که رقم ۰,۰۷۹ می‌باشد. و اقدام فروش-قوی در رتبه دوم می‌باشد. در حالی که نگهداری، منجر به هیچ پاداشی نبوده است و با بررسی جزئیات سیستم سیگنال‌دهی مشخص می‌شود که هیچ سیگنالی برای نگهداری، در بازارهای صعودی ایجاد نشده است.

در حالت بازار نزولی، خرید-قوی، بالاترین پاداش را در میان سایر اقدامات به خود اختصاص داده است، و فروش-ضعیف با اختلاف بسیار اندکی در رتبه دوم قرار می‌گیرد.

در بازار نزولی، در خصوص سهام شرکت نفت سپاهان، ضعیف‌ترین سیگنال و عملکرد مربوط به خرید-ضعیف می‌باشد.

مفهوم این ارقام در روش یادگیری تقویتی، این است که سهم مورد نظر در طول دوره مورد بررسی، در حالت بازار نزولی، توانسته است با اقدام خرید-قوی، پاداش‌های بیشتری کسب کند. البته در بازار نزولی، سودآوری در معاملات به دلیل نزولی بودن قیمت‌ها، چالش برانگیز است، ولی به دلیل وجود نوسانات در بازار، در بازار نزولی نیز امکان سودآوری وجود دارد و در بازار سرمایه‌های مختلف جهان، سرمایه‌گذاران در بازارهای نزولی نیز امکان سودآوری را دارند.

در حالت بازار نزولی، در مورد سهام شرکت نفت سپاهان، اقدام نگهداری نیز دارای پاداش محاسبه شده می‌باشد، که به معنای این است که اقدام به نگهداری سهام و عدم خرید سهام، در برخی از موارد، توانسته است منجر به پاداش شود.

بهبودسازی پرتفوی سهام با استفاده.../اسفندیاری، کرامتی، غلامی جمکرانی و کاشفی نیشابوری

لذا، خودمعامله گرمی آموزد که در بازارهای صعودی، برای این سهم بخصوص خرید-ضعیف، مناسب ترین اقدام می باشد، و نگهداری نیز، منجر به پاداش نخواهد بود، و البته موجب جریمه نیز نبوده است.

در تحلیل حالت بازار صعودی، در بین هفت سهم مورد بررسی، سیگنال نگهداری، در اکثر موارد، هیچ پاداشی به ارمغان نیاورده است که به معنای این است که طبق سیاست های انتخابی، در بازارهای صعودی، نگهداری سهم در روش یادگیری تقویتی و با تنظیمات اعمال شده بر روی خودمعامله گر، عامل موفق به کسب سود نبوده است، و زیانی نیز به وقوع نپیوسته است. از آنجا که استراتژی معاملاتی در روش یادگیری تقویتی، بهره بردن از فرصت های معاملاتی خرید و فروش می باشد، لذا، سیستم خودمعامله گر با بهره بردن از سیگنال های ایجاد شده، با به وجود آمدن فرصت معاملاتی، نسبت به معامله سهم اقدام می کند و به هیچ عنوان هدف این نیست که بر خرید و نگهداری در بازارهای صعودی تأکید بشود، و یا در بازارهای نزولی، فقط به فروش سهم توجه بشود و از فرصت های معاملاتی خرید غفلت بشود.

لذا، با هدف انجام معاملات بیشتر در اولین فرصت فراهم شده، در واقع از سیاست خرید و نگهداری تا حد امکان دور می شویم، در حالی که در مواردی که شرایط بازار ایجاب کند، سیگنال معاملاتی نگهداری، ایجاد خواهد شد و این اتفاق در مواردی واقع شده است که سیستم سیگنال دهی، اقدام را نگهداری تشخیص داده است، که جزئیات معاملاتی هر سهم در محاسبات ماتریس کدنویسی شده در اکسل و نیز خروجی ماتریس سیگنال معاملاتی قابل ملاحظه می باشد.

همان طور که انتظار می رود، هر کدام از سهم های مورد بررسی ارقام خاص خود را در ماتریس ها، به نمایش گذاشته اند، و تفسیر این ارقام نیز برای هر سهم تا حدی مخصوص به خود آن سهم می باشد، و دلیل وجود این تفاوت ها، با وجود این نکته که تا حد امکان شرایط معاملاتی برای همه هفت سهم مورد بررسی یکسان سازی شده است، نتایج محاسبات ماتریس کیو، متفاوت هستند، به رفتار متفاوت هر سهم در بازار سرمایه باز می گردد. در واقع خودمعامله گر با استفاده از روش یادگیری تقویتی می آموزد که بهترین اقدام در یک حالت بازار مشخص صعودی و یا نزولی چیست و می تواند از این تجربه در انجام معاملات آتی بهره برداری کند.

نکته بسیار مهم، توجه به تابع مطلوبت معاملات می باشد، و در صورتی که مطلوبیت هر معامله در حد قابل قبولی باشد، نسبت به انجام آن اقدام خواهد شد و سیستم ایجاد سیگنال های معاملاتی در واقع بخش اصلی از مطلوبیت معاملات خرید و فروش را ارزیابی می کند. همان طور که ملاحظه می شود، مباحث و تفسیرهای فراوانی را می توان در خصوص هر سهم و رفتار این سهم در بازار مورد بررسی قرار داد، و ارقام ماتریس کیو، حاصل تجربه عامل در محیطی است که عامل معاملات را تجربه کرده است.

ارقام ماتریس کیو، را می‌توان به مانند تفسیر عملکرد معاملاتی خود معامله‌گر در حالت‌های بازار برای یک سهم مشخص در قالب ارقام استاندارد شده و قابل درک برای عامل انسانی و عامل غیرانسانی که همان الگوریتم‌های کامپیوتری می‌باشند دانست. این ماتریس از پیچیدگی‌های درک عملکرد معاملاتی می‌کاهد و در قالب ارقامی معنادار، این عملکرد را تشریح و قابل درک می‌کند. به هر میزان که طول دوره بررسی عملکرد طولانی‌تر بشود، ارقام محاسبه شده، قابلیت اتکای بالاتری پیدا می‌کنند و از کیفیت بهتری در جهت درک رفتار سهم مشخص در حالت بازار ارائه می‌دهند.

نتیجه‌گیری و پیشنهادها

در پژوهش حاضر، برای رسیدن به بینشی عمیق‌تر از بازار بورس و عملکرد آن، داده‌های روزانه هفت شرکت بورسی با استفاده از روش یادگیری کیو عمیق طی دوره زمانی ۱۳۹۶ - ۱۴۰۰ بررسی شد. جهت بررسی دو استراتژی یادگیری تقویتی و استراتژی خرید و نگهداری، از میان سه حالت بازار که صعودی، نزولی، و خنثی می‌باشد، دو حالت اصلی (بازارهای صعودی و نزولی)، مورد توجه قرار گرفته‌اند. چرا که روند بازار در این دو حالت امکان بررسی و مقایسه بهتری در بین دو استراتژی معاملاتی خرید و نگهداری را با روش معاملات الگوریتمی می‌دهد.

نتایج بررسی‌ها، به خوبی برتری معاملات الگوریتمی را در هر دو بازار صعودی و نزولی به اثبات رسانیده است. نتایج کسب شده در بازارهای نزولی که روش معاملاتی خرید و نگهداری، منجر به زیان می‌شود، موفق به سود آوری مناسبی بوده است. مقایسه نتایج این دو رویکرد روشن می‌کند که کاربرد یادگیری تقویتی برای سرمایه‌گذارهایی که توان ریسک‌پذیری بالای رهیافت خرید و نگهداری را ندارند، مناسب‌تر است.

در نهایت، پس از اتمام یادگیری، ماتریس کیو ایجاد شده، می‌تواند در انتخاب عمل مورد نظر در هر کدام از حالت‌های بازار، با در نظر داشتن این نکته که سهم مورد نظر با توجه به کارمزدهای معاملاتی خریدنی و یا فروختنی محسوب می‌شود، می‌توان پس از ایجاد سیگنال معاملاتی، به بررسی عملکرد متناظر با آن سیگنال در ماتریس حالت-عمل، اقدام نمود و با ملاحظه رقم محاسبه شده برای این اقدام، در صورتی که رقم محاسبه شده مثبت می‌باشد، نسبت به اجرای آن سیگنال، اقدام نمود. ماتریس کیو، منعکس کننده دستاوردهای عامل (خود معامله‌گر) از انجام یک عمل در حالت مشخصی از بازار است، لذا، ارقام ایجاد شده در ماتریس کیو می‌تواند در اطمینان بخشی از آن اقدام در آن حالت بازار، کاربرد داشته باشد. کاربرد دیگری که از ماتریس کیو و یادگیری تقویتی می‌توان به عمل آورد، این است که با توجه به ارقام ماتریس کیو، در مواردی که ارقام منفی و یا نزدیک به صفر می‌باشد، می‌توان با بررسی در سوابق

بهینه‌سازی پرتفوی سهام با استفاده.../اسفندیاری، کرامتی، غلامی جمکرانی و کاشفی نیشابوری

معاملاتی، وضعیت را تحلیل نمود و ایرادات احتمالی در سیستم معاملاتی را تشخیص و با برطرف نمودن آن، نسبت به آموزش خود معامله‌گر اقدام نمود و پس از طی مرحله آموزشی، از ارقام ماتریس کیو جهت تأیید سیگنال ایجاد شده توسط سیستم سیگنال‌دهی، استفاده نمود.

از طرفی نتایج یادگیری تقویتی از طریق الگوریتم یادگیری کیو عمیق برای عامل خود معامله‌گر، در دو حالت بازار صعودی و نزولی مشخص کرد که عامل از تجربیات قبلی خود می‌آموزد که سهامی را که تحت اجرا دارد، در روند بازاری که تحت آموزش به روش یادگیری تقویتی به روش کی، چه پاداش‌هایی برای عامل و عملی که در آن محیط انجام می‌دهد، به ارمغان می‌آورد. به عبارتی، نتایج اطلاعاتی خوبی را در خصوص رفتار یک سهم بخصوص در یک حالت مشخص به عامل ارائه می‌کند و چنین جدولی که یک ماتریس حالت-عمل می‌باشد، به عامل در انتخاب بهترین عمل ممکن کمک می‌نماید و نتیجه استفاده از روش یادگیری تقویتی، با استفاده از ضرایب مناسب، این است که عامل از محیط (بازار) می‌آموزد، و پس از این‌که، دوره آموزش مناسبی را در خصوص یک سهم مشخص و حالت‌های بازار، طی دوره زمانی به طول مناسب طی شود، می‌توان، به آموزش و تجربه ای که عامل از حالت-عمل کسب کرده است، تا حد زیادی اعتماد نمود و در تصمیم‌گیری در معاملات از این قابلیت یادگیری تقویتی بهره‌مناسبی برد و سودآوری را افزایش داد. در نتیجه، می‌توان با توجه به بررسی‌ها و تفسیرهای انجام شده، یادگیری تقویتی را روشی مؤثر و مبتنی بر یادگیری عامل از محیط دانست که در بهینه‌سازی پرتفوی و افزایش سودآوری از معاملات مؤثر می‌باشد. لذا با توجه به ماهیت بازار سرمایه که ادامه روند گذشته به صورتی پویا می‌باشد، معاملات الگوریتمی به روش یادگیری تقویتی می‌تواند با انتخاب بهترین عمل با استناد به پاداش - جریمه حاصل از هر اقدام، نسبت به سایر اقدامات، موجب بهینه شدن پرتفوی بشود.

به واسطه تنوع در شرکت‌های بورس اوراق بهادار تهران، یکی از عمده‌ترین محدودیت‌ها و چالش‌های تحقیق حاضر، بررسی و انتخاب شرکت‌هایی بوده است که در دستیابی به نتایج قابل قبول پس از بررسی‌ها، ایجاد تردید ننمایند، به عبارتی، در چنین تحقیقاتی، یکی از محدودیت‌های عمده، انتخاب نمونه‌های مناسب می‌باشد. به عبارتی در بررسی تعداد محدودی از شرکت‌های حاضر در بازار سرمایه، در گام اول، توجه به تنوع این شرکت‌ها در جهت دستیابی به نتایج قابل اعتنا از بررسی اهمیت فراوانی دارد.

منابع

- (۱) امیری، میثم، ابراهیمی سروعلیا، محمدحسن و هاشمی، هما. (۱۳۹۹). بررسی عملکرد الگوریتم GRASP درانتخاب پرتفوی بهینه (با لحاظ محدودیت کاردینالیته). اقتصادمالی، ۱۴(۵۱)، ۱۴۷-۱۷۲.
- (۲) رستگار، محمدعلی، دستپاک، محسن (۱۳۹۷). ارائه مدل معاملاتی با فراوانی زیاد همراه با مدیریت پویای سبد سهام به روش یادگیری تقویتی در بورس اوراق بهادار تهران. فصلنامه تحقیقات مالی، ۲۰(۱): ۱۶-۱.
- (۳) فلاح پور، سعید، حکیمیان، حسن (۱۳۹۸). بهینه‌سازی استراتژی معاملات زوجی با استفاده از روش یادگیری تقویتی، با به‌کارگیری دیتاهای درون‌روزی در بورس اوراق بهادار تهران، فصلنامه تحقیقات مالی، ۲۱(۱): ۳۴-۱۹.
- (۴) گل‌ارضی، غلامحسین، انصاری، حمیدرضا (۱۴۰۱). مقایسه عملکرد الگوریتم‌های تکاملی NSGAI و SPEA2 در انتخاب پرتفولیوی بهینه در بورس اوراق بهادار تهران. فصلنامه تحقیقات مالی، ۲۴(۳): ۴۱۰-۴۳۰.
- (۵) میزبان، هدیه سادات، افچنگی، زهرا، احراری، مهدی، آروین، فرشاد و سوری، علی (۱۳۹۱). بهینه‌سازی سبد سهام با استفاده از الگوریتم ازدحام ذرات در تعاریف مختلف اندازه گیری ریسک. اقتصاد مالی، ۱۹(۶)، ۲۰۵-۲۲۷.
- 6) Agarwal, A. , Hazan, E. , Kale, S. , & Schapire, R. E. (2006). Algorithms for portfolio management based on the newton method. In Proceedings of the 23rd international conference on machine learning (pp. 9–16). ACM .
- 7) Amiri, R., Mehrpouyan, H., Fridman, L., Mallik, R. K., Nallanathan, A., & Matolak, D. (2018). A Machine Learning Approach for Power Allocation in HetNets Considering QoS. In 2018 IEEE International Conference on Communications (ICC). 2018 IEEE International Conference on Communications (ICC 2018). IEEE. <https://doi.org/10.1109/icc.2018.8422864>
- 8) Ha, Y., & Zhang, H. (2020). Algorithmic trading for online portfolio selection under limited market liquidity. In European Journal of Operational Research (Vol. 286, Issue 3, pp. 1033–1051). Elsevier BV. <https://doi.org/10.1016/j.ejor.2020.03.050>
- 9) Markowitz H.M. (1952) . Portfolio Selection. Journal of Finance, 7 (1): 77-91.
- 10) Mohammed, M. A., Lkhan, A., Abdulkareem, K. H., & Garcia-Zapirain, B. (2023). A hybrid cancer prediction based on multi-omics data and reinforcement learning state action reward state action (DEEP Q). In Computers in Biology and

بهبودسازی پرتفوی سهام با استفاده.../اسفندیاری، کرامتی، غلامی جمکرانی و کاشفی نیشابوری

- Medicine (Vol. 154, p. 106617). Elsevier BV. <https://doi.org/10.1016/j.combiomed.2023.106617>
- 11) Park, H., Sim, M. K., & Choi, D. G. (2020). An intelligent financial portfolio trading strategy using deep Q-learning. *Expert Systems with Applications*, 158.
- 12) Reeves, M., Moose, S., & Venema, T. (2014). *The growth share matrix*. BCG—The Boston Consulting Group.
- 13) Skabar, A., & Cloete, I. (2002). Neural networks, financial trading and the efficient markets hypothesis. In *ACSC*: 241-249
- 14) Soleymani, F., & Paquet, E. (2020). Financial portfolio optimization with online deep reinforcement learning and restricted stacked autoencoder—DeepBreath. In *Expert Systems with Applications* (Vol. 156, p. 113456). Elsevier BV. <https://doi.org/10.1016/j.eswa.2020.113456>
- 15) Treleaven, P., Galas, M. & Lalchand, V. (2013). Algorithmic trading review. *Communications of the ACM*, 56(11): 76-85.
- 16) Zhang, Z., Zohren, S., & Roberts, S. (2020). Deep reinforcement learning for trading. *The Journal of Financial Data Science*, 2(2): 25-40.

یادداشت‌ها:

۱ Portfolio

۲ Soleymani

۳ Markowitz

۴ وی به بسط نظریه پورتفولیوی مدرن (MPT) Modern Portfolio Theory (MPT) مبادرت ورزید. نظریه وی با عنوان نظریه انتخاب پورتفولیو (Theory of Portfolio Choice) مشهور شده است.

۵ The Equity Curve

۶ Zhang, Zohren & Roberts

۷ Reeves, Moose & Venema

۸ Ha

۹ Agarwal

۱۰ Reinforcement Learning

۱۱ Algorithm Trading

۱۲ Q-Learning

۱۳ Mohammed

۱۴ Sharpe Ratio

۱۵ Treleaven

۱۶ Reinforcement Learning

۱۷ Amiri

۱۸ On-policy

۱۹ Off-policy

۲۰ Q-learning

۲۱ Online

۲۲ Skabar & Cloete

۲۳ Park, Sim & Choi

۲۴ Zhang

۲۵ معاملات درون‌روزی (Intraday Trading) که به‌عنوان معاملات روزانه نیز شناخته می‌شود به خرید و فروش سهام در همان روز اشاره دارد. بازار سهام در معرض نوساناتی است که به تغییر قیمت سهام در طول روز منجر می‌شود.

۲۶ GRASP

۲۷ CCPO

۲۸ هدف از انتخاب هفت شرکت در صنایع مختلف، تصمیم‌پذیر نمودن نتایج تحقیق به صورت عمومی جهت کاربرد رویکرد یادگیری تقویتی در بهینه نمودن پرتفوی سهام می‌باشد در حالی که حجم محاسبات نیز می‌بایست قابلیت اجرایی داشته باشند، لذا با توجه به ملاحظات اجرایی و محدودیت‌های اجرای تحقیق، و همچنین خوانایی بیشتر جداول و نمودارها، تعداد هفت شرکت بازار سرمایه در پنج گروه صنعتی انتخاب شده‌اند. در خصوص انتخاب صنایع نیز با هدف تنوع‌دهی به صنایع انتخابی تلاش بر انتخاب صنایعی است که متمایز بوده و در نهایت صنایعی که انتخاب می‌شوند، طبق نظریه پرتفوی مدرن کم‌ترین همبستگی را با هم داشته باشند و در نتیجه عملکرد کلی صنایع انتخابی قابل تصمیم‌دهی به سایر صنایع بازار سرمایه متشابه نیز باشد. انتخاب تعداد صنایع با توجه به ملاحظات اجرایی و محدودیت‌های اجرای تحقیق، پس از بررسی اثرات افزایش و کاهش تعداد صنایع بر نتایج تحقیق، توسط محقق تصمیم‌گیری شده است.

۲۹ <http://irbours.com>

۳۰ Bellman Equation

۳۱ Q-table

۳۲ در خصوص سیگنال‌های قابل ایجاد، می‌بایست این نکته را یادآوری نمود که در مدل فعلی، به جهت ساده سازی مدل و تمرکز بر اجرای مدل و تحلیل نتایج، از پیچیده نمودن مدل و ایجاد سیگنال‌های میانی که منجر به پیچیده شدن ساختاری مدل به واسطه افزایش تعداد گزینه‌های محتمل در تصمیم‌گیری و معامله و در نتیجه کاهش دقت در بررسی‌ها و افزایش زمان ایجاد و اجرای مدل می‌شود خودداری شده است.

Stock portfolio optimization using Deep Q Reinforcement Learning strategy based on State-Action matrix

Mehdi Esfandiari¹

Receipt: 15/05/2023 Acceptance: 17/05/2023

Mohammad Ali Karamati²

Reza Gholami Jamkarani³

Mohammad Reza Kashfi Neishabouri⁴

Abstract

The purpose of this paper is to optimize the portfolio consisting of stocks using DEEP Q's reinforcement learning strategy based on the state-action matrix. For this purpose, in order to optimize and make profitable the portfolio consisting of stocks, the performance of the reinforcement learning strategy based on the DEEP Q algorithm and the passive strategy of Buying and Holding in two states of Bullish and Bearish markets during the time period of 2017-2021 were investigated. The statistical population was 672 companies admitted to the Tehran Stock Exchange, of which 7 companies (statistical sample) were considered suitable. The comparison of two strategies shows that the Reinforcement Learning strategy, in the Bullish and Bearish markets, compared to the trading method of buying and holding, which has led to losses, has a high potential for profitability in the Iranian stock market. Based on the results, it is suggested that brokers and stock exchange companies and analysts use the Reinforcement Learning strategy for profitability and stock portfolio optimization. Also, the comparison of the results of these two approaches makes it clear that the application of Reinforcement Learning is more suitable for investors who do not have the high risk-taking ability of the Buy-and-Hold approach.

keywords

Portfolio Optimization, Algorithmic Trading, Reinforcement Learning, DEEP Q Algorithm, Tehran Stock Exchange **JEL:** G11, G17, P45

1-PhD student, Department of Industrial Management, Qom Branch, Islamic Azad University, Qom, Iran. esfandiar.mehdi@yahoo.com

2-Associate Professor, Department of Industrial Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran. (Corresponding Author) mohammadalikeramati@yahoo.com

3- Associate Professor, Department of Accounting, Qom Branch, Islamic Azad University, Qom, Iran accountghom@gmail.com

4-Associate Professor, Department of Financial Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran mm.kashefi63@gmail.com