Developing a High-Frequency Trading system with Dynamic Portfolio Management using Reinforcement Learning in Iran Stock Market

Mohammad Ali Rastegar
*Corresponding author, Assistant Prof., Faculty of Industrial Engineering, Tarbiat Modares University, Tehran, Iran. E-mail: ma_rastegar@modares.ac.ir

Mohsen Dastpak
M.Sc. in Financial Engineering, Faculty of Financial Science, Kharazmi University, Tehran, Iran. E-mail: moh.dastpak@gmail.com

Abstract

Objective: Presence of the considerable gap between the time of receiving the buy/sell signals and the beginning of the price change trend provides an appropriate situation for implementation of algorithmic trading systems. Tehran stock exchange is one of these markets. A high-frequency trading system has some advantages (exploiting intraday stock market volatility) and disadvantages (high amounts of transaction cost due to the high transaction volume), thus we can augment the advantages and control the disadvantages by designing the system elaborately and modifying the trading regulations.

Methods: In this research, the "Local Traders" approach has been utilized to predict the future trend of stock and Reinforcement Learning has been used for dynamic portfolio management. According to the "Local Traders" approach, there is a local trader (an agent) for each stock that is expert at it. It predicts the future trend of its own stock based on stock's intraday data and their technical indicators by determining how beneficial it is to buy, sell or hold. In this research, 2 models will be proposed based on Local Traders. Based on the first one, trades with fixed lot size were sought by exploiting the local traders' recommendations. In the second model which is an extension of first model, one can dynamically manages the portfolio using reinforcement learning and local traders' recommendations.

Results: Results showed that, the proposed models outperformed the Buy and Hold strategy in Normal and Descending markets. Furthermore, in all kinds of markets, the second model outperformed the first one.

Conclusion: Generally, the Buy and Hold strategy works the best in an Ascending market, hence the proposed algorithms are not expected to outperform this strategy. However, the performance of the proposed approach along with Neural Network method to anticipate the future trend of stocks was considerable in Normal and Descending markets. In addition, the implementation of Reinforcement Learning model to dynamically manage the portfolio has improved the results.

Keywords: Algorithmic trading, Dynamic portfolio management, High-frequency trading, Intra-day data, Reinforcement learning.

Citation: Rastegar, M.A., Dastpak, M. (2018). Developing a High-Frequency Trading system with Dynamic Portfolio Management using Reinforcement Learning in Iran Stock Market. Financial Research Journal, 20(1), 1-16. (*in Persian*)

Financial Research Journal, 2018, Vol. 20, No.1, pp. 1-16
DOI: 10.22059/jfr.2017.230613.1006415
Received: April 5, 2017; Accepted: August 29, 2017

ارائه مدل معاملاتی با فراوانی زیاد، همراه با مدیریت پویای سبد سهام به روش یادگیری تقویتی در بورس اوراق بهادار تهران

محمدعلی رستگار

* نویسنده مسئول، استادیار گروه مهندسی مالی، دانشکده مهندسی صنایع دانشگاه تربیت مدرس، تهران، ایران. رایانامه: ma_rastegar@modares.ac.ir

محسن دستپاک

کارشناس ارشد مهندسی مالی، دانشکده علوم مالی، دانشگاه خوارزمی، تهران، ایران. رایانامه: moh.dastpak@gmail.com

چکیده

هدف: شکاف بین زمان دریافت سیگنال خرید/ فروش و آغاز روند تغییر قیمت در بازارهای نوظهور، بستر مناسبی برای پیاده‌سازی سیستم‌های معاملات الگوریتمی ایجاد می‌کند. ارائه یک سیستم معاملاتی با تکرار زیاد، مزایا (استفاده از نوسان‌های درون‌روزی) و معایبی (هزینه زیاد معاملاتی) دارد که با طراحی درست آن و اصلاح مقررات معامله، می‌توان مزایای آن را افزایش داد و معایبش را کنترل کرد.

روش: در این پژوهش، به ارائه رویکرد استفاده از خودمعامله‌گرها برای پیش‌بینی روند آتی سهم و بهره‌گیری از روش یادگیری تقویتی به منظور مدیریت پویای سبد سهام پرداخته شده و دو مدل بر همین پایه ارائه شده است. مدل نخست با بهره بردن از پیشنهاد خودمعامله‌گرها، به معامله با مقدار ثابت اقدام می‌کند. مدل دوم که به نوعی بسط داده شده مدل نخست است، به کمک روش یادگیری تقویتی، به مدیریت پویای سبد سهام می‌پردازد.

یافته‌ها: نتایج نشان می‌دهد عملکرد هر دو مدل در بازارهای نزولی و نرمال، بهتر از استراتژی خرید ـ و ـ نگهداری است. همچنین بر اساس نتایج، در تمام بازارها مدل دوم در مقایسه با مدل نخست، عملکرد بهتری دارد.

نتیجه‌گیری: به طور کلی در بازار صعودی بهترین استراتژی، خرید ـ و ـ نگهداری دارایی است، در نتیجه نمی‌توان از الگوریتم‌های پیشنهادی عملکردی بهتر از این استراتژی انتظار داشت. از سویی دیگر می‌توان گفت روش شبکه عصبی برای پیش‌بینی روند آتی سهم با رویکرد ارائه شده در این پژوهش، عملکرد بسیار مناسبی در بازارهای نزولی و نرمال داشته است. همچنین پیاده‌سازی روش یادگیری تقویتی به منظور مدیریت پویای سبد سهام توانسته عملکرد مدل را بسیار بهبود بخشد.

کلیدواژه‌ها: معاملات الگوریتمی، معاملات با فراوانی زیاد، مدیریت پویای سبد سهام، داده‌های درون‌روزی، یادگیری تقویتی.

استناد: 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)، ۱- ۱۶.

فصلنامه تحقیقات مالی، ۱۳۹۷، دوره ۲۰، شماره ۱، صص. ۱- ۱۶

DOI: 10.22059/jfr.2017.230613.1006415

دریافت: ۱۳۹۶/۰۱/۱۶، پذیرش: ۱۳۹۶/۰۶/۰۷

مقدمه

در بازارهای پیشرفته، معاملات الگوریتمی با استفاده از داده‌های درون‌روزی، سهم عمده‌ای از معاملات بازار در دهه گذشته را به خود اختصاص داده است. امکان استفاده از معاملات الگوریتمی در بازار بورس اوراق بهادار تهران تا چندی پیش وجود نداشت، اما با مجوزهایی که اخیراً صادر شده است، این گونه معاملات میسر شده و برخی نهادهای مالی برای انجام معاملات، از الگوریتم‌های از پیش تهیه شده یا توسعه یافته خودشان استفاده می‌کنند. به این منظور، روش‌های بسیاری برای پیش‌بینی متغیرهای بازار سرمایه پیاده‌سازی شده است (کندال و اورد، ۱۹۹۷). در بیشتر این پژوهش‌ها، 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 تا کنون در زمینه مدیریت سبد دارایی انجام شده، دارای دو بخش هستند: پیش‌بینی قیمت سهم و مدیریت سبد دارایی. در اغلب این پژوهش‌ها، از «یادگیری نظارت‌شده»[1] برای ایجاد ارتباط میان یکسری داده ورودی و خروجی مطلوب[2] استفاده شده است (دودا، هارد و استورک، ۲۰۰۰). در این میان، از آنجا که روش‌های پیش‌بینی به کمک شبکه عصبی مصنوعی[3] به مدل پارامتری نیازی ندارند، از محبوبیت بسیاری در پیش‌بینی قیمت سهام برخوردارند (ساد و پروخوروف، ۱۹۹۸). فَن و همکارانش از روش «ماشین بردار پشتیبان»[4] برای دسته‌بندی و انتخاب سهم بهره بردند (فن و پالانیسوامی، ۲۰۰۱).

مشکل روش‌های مشابه یادگیری نظارت شده در این نوع مسائل، این است که هدف آن‌ها کاهش خطای بین پیش‌بینیِ 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 این رو، رویکرد آموزش با تأخیر زمانی، می‌تواند نتایج واقع‌بینانه‌تری داشته باشد؛ زیرا در این رویکرد، با توجه به زنجیره تصمیم‌ها آموزش داده می‌شود. روش «یادگیری تقویتی»[5] با تأخیر زمانی[6]، مجموع پاداش به دست آمده از تصمیم‌های زنجیره‌ای (پاداش تجمعی)[7] را بر ارزش هر تصمیمِ اخذ شده، اثر می‌دهد (واتکینز، ۱۹۸۹؛ سوتان و بارتو، ۱۹۹۸). این روش به دلیل تأثیر دادن تصمیم‌های زنجیره‌ای، از روش‌های اشاره شده منطقی‌تر به نظر می‌رسد.

با توجه به آنچه گفته شد، اجرای این گونه پژوهش‌ها که به ارائه مدلی برای انجام معاملات الگوریتمی در بازار بورس اوراق بهادر تهران با استفاده از روش «یادگیری تقویتی» برای اخذ تصمیم‌های معاملاتی می‌پردازد، ضروری به نظر می‌رسد. در مدل این پژوهش، به ازای هر سهم موجود در سبد سهام، یک خودمعامله‌گر[8] متناظر که مخصوص آن سهم آموزش دیده است، در نظر گرفته می‌شود. از آنجا که استفاده از سری زمانی قیمت و میانگین متحرک، به تنهایی

1. Supervised Learning
2. Desired Output
3. Neural Network
4. Support Vector Machines (SVMs)
5. Reinforcement Learning (RL)
6. Time Delay
7. Cumulative Reward
8. Local Trader

اطلاعات کافی در اختیار ما قرار نمی‌دهند، در این پژوهش از اندیکاتورهای تحلیل تکنیکال و سیگنال‌های هر یک به عنوان داده ورودی استفاده می‌شود؛ سپس با بهره‌مندی از پیشنهادهای این خودمعامله‌گرها و همچنین وضعیت وزن دارایی‌ها در سبد، مدل تکمیل‌تری برای مدیریت پویای سبد سهام به روش یادگیری تقویتی به منظور بهبود عملکرد مدل نخست ارائه می‌شود.

در ادامه ابتدا با پیشینه تحقیق و مبانی نظری آشنا می‌شویم. پس از آن، در بخش روش‌شناسی پژوهش به بیان روش انتخاب سهام مورد بررسی، معرفی مدل و الگوریتم ارائه‌شده می‌پردازیم. در پایان نیز نتایج به دست آمده از مدل‌ها را بررسی می‌کنیم.

پیشینه پژوهش

نونییر (۱۹۹۸) با الگوگیری از فرایند تصمیم‌گیری مارکوف[1]، مدل ساده شده‌ای از یک بازار مالی را با روش یادگیری تقویتی تحلیل کرد. وی با اضافه کردن تحلیل ریسک به مدل خود، از مدل یادگیری کیو[2] برای حل مسئله بهره برد (الیور میهاچ، ۲۰۰۲). گائو و چان (۲۰۰۰) از یادگیری کیو برای مدیریت پرتفولیو و تصمیم‌گیری بین دو سهم، بهره بردند؛ مدل‌سازی آن‌ها به‌صورت دودویی[3] بود و در هر مرحله تنها یک سهم انتخاب می‌شد. موودی و سفل (۲۰۰۱) کار گائو را ادامه دادند، با این تفاوت که به جای مدل یادگیری کیو، از روش یادگیری تقویتی مستقیم[4] استفاده کردند. از آنجا که مدل‌های یادگیری تقویتی، پیچیدگی بسیاری دارند، باید فرض‌هایی را برای ساده‌سازی و کوچک کردن فضای مسئله در نظر گرفت و نونییر (۱۹۹۸) و الیور میهاچ (۲۰۰۲) در پژوهش‌های خود فرض‌هایی را به منظور ساده‌سازی در نظر گرفتند. در تمام پژوهش‌های اشاره شده، هدف محققان تحلیل سهم به صورت جداگانه بوده است. جانگمین و همکارانش نخستین کسانی بودند که به یکپارچه کردن «تحلیل روند سهام» و «مدیریت سبد دارایی» اقدام کردند. آن‌ها ابتدا در پژوهشی با در نظر گرفتن چهار خودمعامله‌گر (هر معامله‌گر رویکرد خاصی را برای بررسی روند سهم دارد) به عنوان پیشنهاددهنده و انتخاب‌کننده سهام برای خرید یا فروش، به کمک الگوریتم‌های تکاملی، درصدی از دارایی را به هر خودمعامله‌گر اختصاص دادند تا آن عامل، پول در اختیارش را صرف خرید سهام پیشنهاد شده خود کند (جانگمین، لی، ژانگ و لی، ۲۰۰۵). در ادامه، این مدل را با روش یادگیری تقویتی پیاده‌سازی کردند (جانگمین، لی، ژانگ و لی، ۲۰۰۶ و ۲۰۰۴). در پژوهش دیگری، جانگمین و همکارانش از چهار عامل[5] (سیگنال‌دهنده خرید، سیگنال‌دهنده فروش، سفارش‌دهنده خرید و سفارش‌دهنده فروش) بهره بردند و برای هر یک از این عامل‌ها یک مدل یادگیری کیو در نظر گرفتند[6] (لی، پارک، هونگ، جانگمین و لی، ۲۰۰۷). در این پژوهش از تغییرات قیمت درون‌روزی که عامل‌های سفارش‌دهنده خرید و فروش، قیمت‌های بهینه برای سفارش را بر اساس آن‌ها تعیین می‌کند، استفاده شده است. اگرچه

1. Markov Decision Process (MDP)

۲. Q-Learning یکی از روش‌های یادگیری ماشین است.

3. Binary
4. Direct Reinforcement Learning
5. Agent
6. Multiple Q-Learning (MQL)

در رویکرد یادگیری تقویتی با فضای بسیار بزرگی از مسئله مواجه‌ایم و مدل پیچیدگی بسیاری دارد (لی، پارک، جانگمین، لی و هونگ، ۲۰۰۷؛ لی، سونگ دونگ، جانگوو و جیسئوک، ۲۰۰۳؛ لی و ژانگ، ۲۰۰۲)، قدرت و کارایی آن به مراتب بیشتر از سایر مدل‌هاست. بهلولی (۱۳۹۱) سیستم یکپارچه‌ای برای مدیریت پویای سبد دارایی به کمک یادگیری تقویتی ارائه کردند. آن‌ها هر دو بخش پیش‌بینی سهام و مدیریت سبد دارایی را در قالب مدل یادگیری تقویتی روی داده‌های انتهای روز (و نه درون‌روزی) بورس نیویورک پیاده‌سازی کردند. جانگمین، لی، لی و ژانگ (۲۰۰۶) خودمعامله‌گرها را بر اساس ترتیب تقاطع‌ها و صعودی/نزولی بودن سری‌های زمانی میانگین متحرک[1] ۵، ۱۰ و ۲۰ دوره‌ایِ قیمت سهم طراحی کردند.

جونز (۱۹۹۹) و دمپستر و جونز (۲۰۰۰ و ۲۰۰۲) طی دو پژوهش، سودآوری استفاده از داده‌های درون‌روزی در سیستم‌های معامله‌گر خودکار[2] را نشان دادند. در ادامه توجه خود را به استفاده از اندیکاتورهای تحلیل تکنیکال روی داده‌های درون‌روزی معطوف کردند و سودآوری این مدل را در بازارهای فارکس[3] به نمایش گذاشتند. همچنین، دمپستر، پاین، روماهی و تامپسون (۲۰۰۱) از الگوریتم‌های یادگیرنده برای طراحی سیستم معامله‌گر با استفاده از اندیکاتورهای تحلیل تکنیکال روی داده‌های درون‌روزی بهره گرفتند و نشان دادند که با هزینه معاملاتی صفر، تمام روش‌های اتخاذشده سودآور است، اما با هزینه معاملاتی واقعی هیچ‌یک از این سیستم‌ها سود شایان توجهی ندارد. پژوهش‌های دیگری نیز به تحلیل استفاده از داده‌های درون‌روزی در سیستم‌های معاملات الگوریتمی پرداخته‌اند (دمپستر و روماهی، ۲۰۰۲؛ یاماموتو، ۲۰۱۲؛ نیلی و ولر، ۲۰۰۳؛ تاناکا یاماواکی و توکوکا، ۲۰۰۷). از این بین، تنها پژوهش‌های یاماموتو (۲۰۱۲) و تاناکا- یاماواکی و توکوکا (۲۰۰۷) هم‌زمان از اندیکاتورهای تحلیل تکنیکال و داده‌های درون‌روزی در سیستم معاملات الگوریتمی سهام استفاده کرده‌اند. این دو پژوهش روی فقط یک یا دو دارایی (و نه سبد دارایی سهام) بررسی شده‌اند. در حالی که تاناکا یاماواکی و توکوما (۲۰۰۷) نشان دادند ترکیب بهینه‌ای از اندیکاتورهای تحلیل تکنیکال می‌تواند پیش‌بینی خوبی از قیمت آینده سهم داشته باشد، نتایج پژوهش یاماموتو (۲۰۱۲) نشان می‌دهد هیچ‌یک از استراتژی‌های اتخاذشده در پژوهشش، نمی‌تواند به اندازه استراتژی خرید و نگهداری[4] سودآور باشد. راعی و باجلان (۱۳۸۷) آثار تقویمی در بورس اوراق بهادار تهران را شناسایی کردند و نشان دادند لحاظ کردن آثار تقویمی، موجب افزایش قدرت پیش‌بینی می‌شود.

در بورس اوراق بهادار تهران، تا کنون پژوهش ثبت‌شده‌ای با بهره‌مندی از داده‌های درون‌روزی و استفاده از اندیکاتورهای تحلیل تکنیکال انجام نشده است. پژوهش‌های اخیر اغلب روی داده‌های انتهای روز بوده و ماهیت معامله با فراوانی زیاد را ندارند.

روش‌شناسی پژوهش

این پژوهش روی بورس اوراق بهادار تهران انجام می‌شود و اطلاعات معاملات، اعم از قیمت معامله، بهترین قیمت

1. Moving Average
2. Automated Trading System
3. Foreign Exchange (For-Ex)
4. Buy and Hold

خرید، بهترین قیمت فروش، حجم معاملات انجام شده و ... از طریق سرورهای بورس اوراق بهادار تهران در اختیار مؤسسه‌ها یا افرادی که به این اطلاعات نیاز داشته باشند، قرار می‌گیرد. تمام اندیکاتورها و سری‌های زمانی دیگری که در ادامه توضیح داده خواهد شد نیز، در محیط SQL Server محاسبه و دسته‌بندی می‌شوند. دلیل محاسبه اندیکاتورها و بازده‌ها در محیط یاد شده این است که رویکرد محاسباتی SQL Server به صورت ماتریسی است، در نتیجه در این محیط عملیاتی همچون محاسبه سری زمانی میانگین‌های متحرک، تأخیر دادن به سری زمانی و ... در مقایسه با C# با سرعت بیشتری انجام می‌شود.

خودمعامله‌گر

هر خودمعامله‌گر، عاملی[1] برای پیش‌بینی روند آتی قیمت یک سهم خاص است. در مدل این پژوهش، یک خودمعامله‌گر، یک شبکه عصبی است که بر اساس برداری از اطلاعات (در ادامه به‌تفصیل بیان خواهند شد) نسبت مطلوب آن سهم را برای خرید، فروش یا نگهداری اعلام می‌کند. ایده‌ای که در این پژوهش مدنظر قرار دارد، استفاده از اطلاعاتی است که مجموعه‌ای از اندیکاتورهای تحلیل تکنیکال را در اختیار ما قرار می‌دهند.

سهام مورد بررسی

یکی از الزامات هر سیستم معاملات الگوریتمی[2] و به ویژه معاملات با تکرار زیاد[3]، آن است که سهام مورد بررسی خاصیت معامله‌شوندگی زیادی داشته باشند؛ به این معنا که در هر لحظه، معامله هر مقدار از سهام امکان‌پذیر باشد. از این رو، سهامی انتخاب شده است که ضمن داشتن بیشترین حجم معاملاتی، روند مد نظر را طی کرده باشند. طبق نمودار شاخص کل، سه بازه ۲ ماهه تقریباً «صعودی»، «نزولی» و «نرمال» را در نظر می‌گیریم و در هر فاصله سه سهم را بر اساس بالاترین حجم معاملاتی و دارا بودن روند مرتبط با آن بازه (صعودی، نزولی یا نرمال) انتخاب می‌کنیم. برای مثال، سهم وبشهر در بازه ۱۳۹۲/۰۲/۲۹ تا ۱۳۹۲/۰۴/۲۹ دارای بیشترین حجم معامله بوده و روند صعودی دارد، در حالی که سهم اخابر در بازه ۱۳۹۱/۱۱/۱ تا ۱۳۹۱/۱۲/۲۸ دارای روندی نزولی و بیشترین حجم معاملاتی است.

جدول ۱. فهرست سهام بررسی شده در پژوهش

نوع بازار	بازار نزولی	بازار نرمال	بازار صعودی
بازه	۱۳۹۱/۱۱/۰۱ تا ۹۱/۱۲/۲۸	۱۳۹۱/۱۲/۰۱ تا ۱۳۹۲/۰۱/۲۸	۱۳۹۲/۰۲/۲۹ تا ۱۳۹۲/۰۴/۲۹
سهام منتخب	اخابر	فولاد	بانک
	فملی	پارس	وبشهر
	انصار	صندوق	شپلی

اطلاعات مهم در پیش‌بینی

در این پژوهش، به دنبال یافتن برداری از اطلاعات هستیم تا بتواند برای پیش‌بینی درست و دقیق آینده سهم، اطلاعات

1. Agent
2. Algorithmic Trading
3. High-Frequency Trading

کافی در اختیار ما قرار دهد. بازده سهم یکی از مهم‌ترین عواملِ پیش‌بینی روند آتی یک سهم است. در این پژوهش، بازده یک، دو و سه دوره قبل با تناوب‌های ۱۰، ۳۰ و ۶۰ دقیقه‌ای مد نظر قرار گرفت و از شش اندیکاتور تحلیل تکنیکال پرکاربرد (RSI، MACD، Stochastic Oscillator، Bollinger Band، Alligator و Ichimoku) بهره برده شد (رودریگز- گنزالز، گارسیا،کولومو، اگلسیاس و گومز، ۲۰۱۱؛ مناهوف، هودسان و جبکا، ۲۰۱۴؛ دووینگ، مازا و پتیتجین، ۲۰۱۳؛ دلافونته، گاریدو، لاویادا و گومز، ۲۰۰۶). علاوه بر این‌ها، اطلاعات زمان معامله نیز می‌تواند برای پیش‌بینی آینده سهم، اطلاعات مفیدی در اختیار ما قرار دهد که از جمله آن‌ها می‌توان به زمان انجام معامله در روز (چه ساعتی در روز)، هفته (چه روزی در هفته) و ماه (چه روزی در ماه) اشاره کرد.

یادگیری تقویتی

پایه و اساس یادگیری تقویتی، فرایند تصمیم‌گیری مارکوف است. بر اساس فرایند مارکوف، در هر مرحله از زمان، فرایند در یک حالت[1] از فضای حالت S قرار دارد و تصمیم‌گیرنده ممکن است یک عمل[2] مثل a را که در حالت s در دسترس است، انتخاب کند. این فرایند در مرحله بعدی با حرکتِ تصادفی به حالت جدیدی مثل s' پاسخ داده و به تصمیم‌گیرنده، پاداش[3] مرتبط با آن تصمیم را می‌دهد که به صورت $R_a(s.s')$ است.

احتمال اینکه این فرایند در مرحله بعدی در حالت s' قرار گیرد، به طور خاص توسط تابع تغییر حالت[4] $P_a(s.s')$ بیان می‌شود. با داشتن حالت s و عمل a، حالت مرحله بعدی به طور مشروط از حالت‌ها و عمل‌های قبلی مستقل است؛ به بیان دیگر، انتقال حالت، دارای خاصیت مارکوف[5] است.

یکی از مهم‌ترین الگوریتم‌هایی که از فرایند تصمیم‌گیری مارکوف به منظور مدل‌سازی استفاده می‌کند، الگوریتم یادگیری تقویتی است. مزیت یادگیری تقویتی نسبت به سایر روش‌های تصمیم‌گیری، بر اساس فرایند تصمیم‌گیری مارکوف آن است که در این روش به دانستن تابع تغییر حالت نیازی نیست. روش یادگیری تقویتی دارای سه جزء اصلی زیر است:

۱. محیط[6]: شامل مجموعه حالت‌ها[7] است.
۲. عمل‌ها[8]: تصمیمی است که تصمیم‌گیرنده اتخاذ می‌کند.
۳. پاداش[9]: پاداش یا جریمه‌ای است که تصمیم‌گیرنده بابت اتخاذ تصمیمی که گرفته دریافت می‌کند.

محیط

محیط شامل تمام حالت‌های ممکنی است که می‌توان در آن قرار گرفت. این حالت‌ها با احتمال‌های مختلف و اخذ

1. State
2. Action
3. Reward
4. State Transition Function
5. Markov Property
6. Environment
7. States
8. Actions
9. Reward

تصمیم‌های گوناگون با هم ارتباط دارند. در بسیاری از پژوهش‌ها حوزه یادگیری تقویتی آمده است که «پیاده‌سازی یادگیری تقویتی، هنرِ تعریف درست و مناسب حالت‌ها و پاداش است». در این پژوهش دو عامل کلی را به عنوان حالت در نظر می‌گیریم: ۱. نسبت وزنی سهام در سبد دارایی و ۲. پیشنهادهای خودمعامله‌گرها.

فرض کنید یک سبد دارایی با n سهم و دارایی پول نقد داریم، بنابراین در قسمت «نسبت وزنی سهام در سبد دارایی» n + ۱ پارامتر و در قسمت «پیشنهاد خودمعامله‌گرها» n پارامتر خواهیم داشت. در نتیجه، حالت دارای ۲n+۱ پارامتر خواهد بود که هر یک می‌تواند مقادیر مختلفی داشته باشد.

- نسبت وزنی سهام در سبد دارایی: همان‌طور که گفته شد، این بخش دارای n + ۱ پارامتر است که مقادیر قابل قبول برای هر پارامتر، از ۰، ۰/۰۱ ... تا ۰/۹۹ و ۱ خواهد بود که در مجموع ۱۰۱ حالت مختلف است. از آنجا که باید مجموع وزنی هر n+۱ دارایی (سهام + پول نقد) برابر با ۱ شود، رابطه ۱ برقرار می‌شود.

رابطه ۱) $$\text{کل تعداد حالات ممکن} = \binom{n+100}{n}$$

- پیشنهاد خودمعامله‌گرها: خروجی هر شبکه از خودمعامله‌گرها، از اعداد ۱+، ۰/۵+، ۰، ۰/۵− و ۱− خواهد بود.

عمل‌ها

در این پژوهش، عمل را نسبت وزنی سهام در سبد دارایی برای دوره بعدی در نظر می‌گیریم. همان‌طور که در قسمت قبلی محاسبه شد، تعداد کل حالات نسبت وزنی سهام در سبد دارایی از رابطه ۱ به‌دست می‌آید، به طور مثال برای ۳ دارایی داریم: (۰/۱, ۰/۱, ۰/۸)

رابطه ۲) $$s^A: \{\widetilde{RC}|\widetilde{RP}\} \xrightarrow{\tilde{a}} s^B: \begin{cases} \{\widetilde{RC}'|\tilde{a}\} \\ \{\widetilde{RC}''|\tilde{a}\} \\ \{\widetilde{RC}'''|\tilde{a}\} \\ \vdots \end{cases}$$

در روابط بالا، $\widetilde{RC}$ بردار پیشنهاد خودمعامله‌گرها و $\widetilde{RP}$ بردار نسبت وزنی سهام در سبد دارایی است.

پاداش

اصلی‌ترین جزء یادگیری تقویتی، پاداش‌دهی به تصمیم‌های گرفته شده است. در این پژوهش، سه مورد را به عنوان فاکتورهای اصلی مؤثر در ارزش هر تصمیم معرفی می‌کنیم:

۱. بازده (IR)[1]: بازده‌ای است که پس از ۵ دوره از اخذ تصمیم به دست می‌آید.
۲. نرخ هزینه معاملاتی (RTC)[2]: نسبتِ هزینه معاملاتیِ ناشی از انجام عمل به کل ارزش سبد دارایی است.
۳. ریسک سبد دارایی (PR)[3]: میزان ریسک سبد دارایی است که با انجام آن تصمیم به وجود آمده است.

1. Immediate Return
2. Rate of Transaction Cost
3 Risk of Portfolio

تابع مطلوبیت

یکی از رویکردهای این پژوهش، دخیل کردن میزان ریسک‌گریزی سرمایه‌گذار در مدل است. به همین منظور، ضریبی که بیان‌کننده میزان ریسک‌گریزی سرمایه‌گذار بوده و به تابع مطلوبیت سرمایه‌گذار وابسته است را در مدل وارد می‌کنیم. به دلیل بالا بودن هزینه معاملاتی در بورس ایران و همچنین ماهیت این‌گونه سیستم‌های معاملاتی (بالا بودن حجم معاملات)، لازم است که نرخ هزینه معاملاتی را با ضریبی در تابع مطلوبیت تأثیر دهیم.

تابع مطلوبیت نهایی به‌صورت زیر خواهد شد:

$$U(IR.FR.RTC.PR) = IR - RTC * TA - PR * RA$$ رابطه ۳)

در رابطه ۳، RA میزان ریسک‌گریزی و TA معامله‌گریزی سرمایه‌گذار را نشان می‌دهد. IR بازده حاصل از این تصمیم در پنج دوره آتی است. PR ریسک سبد دارایی و RTC نسبت هزینه معاملاتی پرداخت شده در دوره جاری به ارزش پورتفوی است.

برای تعیین پاداش تصمیم و پیاده‌سازی آن در مدل یادگیری تقویتی، از تکنیکی به نام یادگیری کیو بهره می‌گیریم. این تکنیک روشی را برای ارزش‌دهی به هر تصمیم، ارائه می‌دهد.

یادگیری کیو

یادگیری کیو، تکنیک یادگیری تقویتی است که با یادگیری یک تابع عمل/ ارزش، سیاست مشخصی را برای انجام عمل‌های مختلف در حالت‌های مختلف دنبال می‌کند. یکی از قوت‌های این روش، توانایی یادگیری تابع یاد شده بدون داشتن مدل معینی از محیط است. هر بار که به تصمیم‌گیرنده پاداش داده می‌شود، مقادیر جدیدی برای هر ترکیب حالت/ عمل محاسبه می‌شود. هسته الگوریتم از یک به روزرسانی تکراری ساده تشکیل شده است؛ به این ترتیب که بر اساس اطلاعات جدید، مقادیر قبلی اصلاح می‌شود.

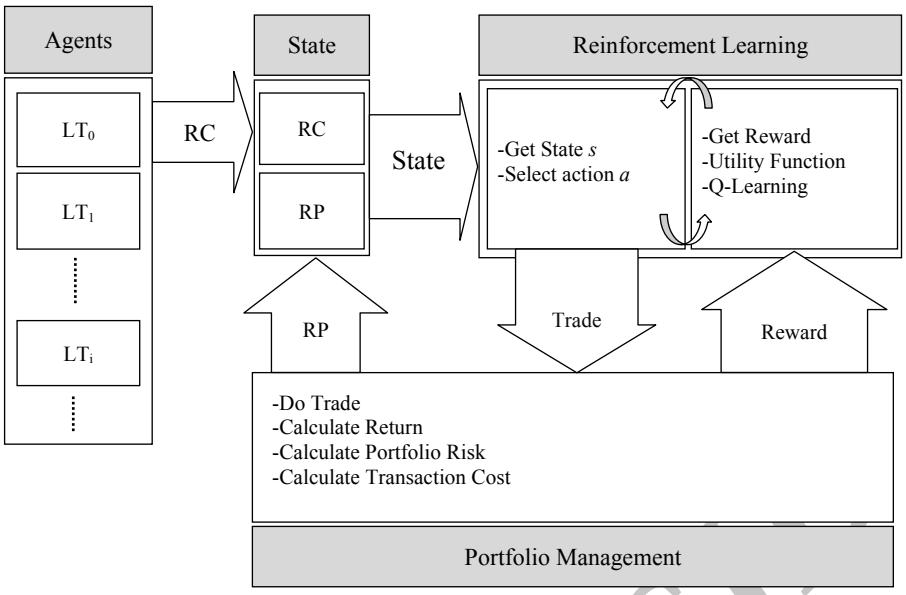
$$Q(s_t.a_t) \leftarrow \underbrace{Q(s_t.a_t)}_{\text{مقدار قبلی}} + \underbrace{\alpha}_{\text{نرخ یادگیری}} \times \left[\overbrace{\underbrace{R(s_t)}_{\text{پاداش}} + \underbrace{\gamma}_{\text{فاکتور تقلیل}} \times \underbrace{\max_{a_{t+1}} Q(s_{t+1}.a_{t+1})}_{\text{بیشترین مقدار آینده}}}^{\text{مقدار یادگرفته شده}} - \underbrace{Q(s_t.a_t)}_{\text{مقدار قبلی}} \right]$$ رابطه ۴)

در مدل مورد بررسی در این پژوهش، از تابع مطلوبیتی که پیش‌تر بیان شد (رابطه ۳) به جای $R(s_t)$ استفاده می‌کنیم و فرمول‌بندی این پاداش را به صورت زیر تغییر می‌دهیم:

$$Q(s_t.a_t) \leftarrow Q(s_t.a_t)(1-\alpha) + \alpha \times \left[U(IR.RTC.PR) + \gamma \times \max_{a_{t+1}} Q(s_{t+1}.a_{t+1}) \right]$$ رابطه ۵)

الگوریتم پیشنهادی

اگر بخواهیم مدل را به صورت یکپارچه در قالب یک سیستم نمایش دهیم، می‌توانیم آن را به صورت زیر بیان کنیم:



شکل ۱. شمای کلی از مدل پیشنهادی

همان‌طور که در شکل ۱ مشاهده می‌شود، الگوریتم پیشنهاد شده، سیستمی است که چهار بخش اصلی را دربرمی‌گیرد: عامل‌ها (Agents)، حالت (State)، یادگیری تقویتی (Reinforcement Learning) و مدیریت سبد (Portfolio Management). مرجع تصمیم‌گیری و یادگیری این الگوریتم، بخش حالت است. الگوریتم، پیشنهادهای معاملاتی را از عامل‌ها و نسبت سهام در سبد را از بخش مدیریت سبد دریافت می‌کند. با این اطلاعات، بخش یادگیری تقویتی عملی (Action) را به کمک تابع توزیع یکنواخت از بین مجموعه عمل‌های ممکن (همانند مثال بیان شده در بخش عمل‌ها) برای حالت موجود انتخاب می‌کند و بر اساس عمل انتخاب شده، معاملات موردنیاز را در بخش مدیریت سبد سهام انجام می‌دهد. پس از هر معامله، بازده متناظرِ آن همراه با هزینه‌های معاملاتی و ریسک مورد انتظار به دست می‌آید و این کار تا انتهای زنجیره انجام می‌شود. در انتهای هر زنجیره، تابع مطلوبیت محاسبه می‌شود و ارزش عمل‌های انتخاب‌شده در هر مرحله با وزن خاصی به روش یادگیری کیو تعدیل می‌شود. این فرایند چندین بار برای کل زنجیره تکرار می‌شود تا الگوریتم بتواند تمام حالت‌های ممکن را تجربه کرده و ارزش مناسبی را برای عمل‌های متناظر آن حالت مقداردهی کند.

یافته‌های پژوهش

در این پژوهش دو مدل (مدل معاملاتی با مقدار ثابت و مدل مدیریت پویای سبد سهام) ارائه شده است. مدل نخست که تنها بر اساس نتایج به دست آمده از خودمعامله‌گرهاست، به مقدار ثابتی از دارایی، با توجه به پیشنهاد معامله‌گر آن دارایی معامله می‌کند؛ سپس با رویکرد مدیریت پویای سبد سهام به صورت چند دوره‌ای و با بهره‌گیری از الگوریتم‌های مبتنی بر فرایند مارکوف، مدل مدیریت پویای سبد سهام به کمک روش یادگیری تقویتی ارائه شده است. در ادامه به نحوه پیاده‌سازی، شرایط و نتایج حاصل، همراه با مقایسه مدل‌های مدیریت پویای سبد سهام، خرید به مقدار ثابت و خرید و نگهداری آورده شده است.

خودمعامله‌گر

ایده‌ای که در این پژوهش مد نظر قرار گرفته است، استفاده از بازده آتی سهم در یک تا پنج دوره بعد است. از آنجا که هزینه معاملاتی برای انجام معامله خرید و فروش در بازار بورس اوراق بهادار تهران ۱/۵ درصد (خرید ۱ درصد و فروش ۰/۵ درصد) است، اگر سهمی بیشتر از این مقدار سود یا ضرر داشته باشد، به ترتیب خریدنی و فروختنی در نظر گرفته می‌شود؛ زیرا آن معامله در بدترین حالت بدون ضرر خواهد بود. از طرفی، در معاملات با فراوانی زیاد، بهینه‌ترین معامله در خرید آن است که در نزدیک‌ترین حالت به سود بیش از ۱/۵ درصد برسد. در معامله فروش نیز، فروختنی‌ترین سهم آن است که در سریع‌ترین حالت انتظار، ضرر بیش از ۱/۵ درصد داشته باشیم. پس نتیجه می‌گیریم که انتظار سود یا ضرر بیشتر و مساوی ۱/۵ درصد در یک و دو دوره آتی، باید نسبت به سه، چهار و پنج دوره بعدی سیگنال قوی‌تری داشته باشد. پس بررسی و مقداردهی خروجی دلخواه در هر نقطه زمانی طبق الگوریتم زیر انجام می‌شود:

اگر r_n را بازده n دوره‌ای سهم در نظر بگیریم، این مقدار به صورت رابطه ۶ محاسبه می‌شود.

$$r_n = \frac{p_{t+n} - p_t}{p_t} \quad \text{رابطه ۶)}$$

$$IF\ (r_2 \geq 0.015)$$
$$Return\ +1.0$$
$$ELSEIF\ (r_5 \geq 0.015)$$
$$Return\ +0.5$$
$$ELSEIF\ (r_2 \leq -0.015)$$
$$Return\ -1.0$$
$$ELSEIF\ (r_5 \leq -0.015)$$
$$Return\ -0.5$$
$$ELSE$$
$$Return\ 0.0$$

در پیاده‌سازی شبکه عصبی، از تابع Bipolar Sigmoid با مقدار پارامتر[1] استاندارد ۲ استفاده شده است. کل اطلاعات به دو گروه اطلاعات برای آموزش و اطلاعات برای آزمایش دسته‌بندی می‌شوند، نسبت این دسته‌بندی ۹۰درصد است؛ به این معنا که ۹۰ درصد اطلاعات برای آموزش و ۱۰ درصد باقی‌مانده برای آزمودن در نظر گرفته شده است. در مدل شبکه عصبی از الگوریتم لونبرگ ـ مارکارد[2] استفاده می‌شود. الگوریتم یاد شده روشی برای یافتن کمینه یک تابع غیرخطی چند متغیره است که روش استانداردی برای حل مسئله کمینه مربعات برای توابع غیرخطی محسوب می‌شود. خروجی، یک شبکه است که با دریافت یک بردار از اطلاعات، عددی بین ۱ و ۱- را به‌عنوان پیش‌بینی نتیجه می‌دهد.

مدل معاملاتی با مقدار ثابت

یکی از اهدافی که از ابتدا در این پژوهش مد نظر بوده، بررسی این فرضیه است که آیا رویکرد طراحی و استفاده از خودمعامله‌گرها در معاملات با تکرار بالا تأثیرگذار است یا خیر. به همین منظور نیاز است که عملکرد این خودمعامله‌گرها

1. Sigmoid Alpha value
2. Levenberg–Marquardt algorithm

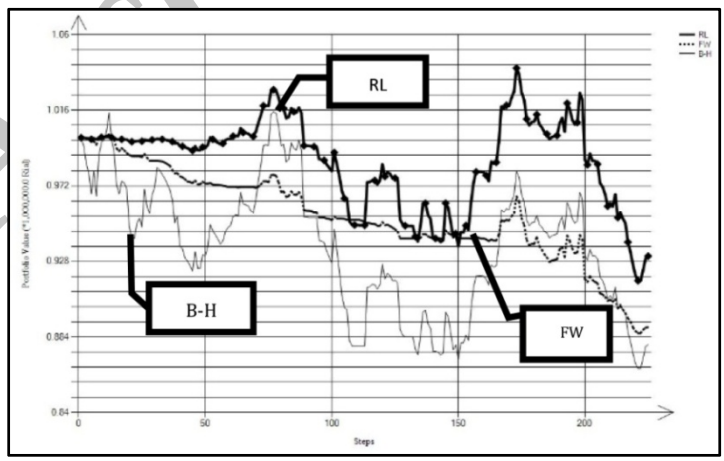
را در قالب یک مدل ساده بررسی کنیم. مدلی که در اینجا استفاده شده، مدلی است که تنها بـر اسـاس میـزان پیشـنهاد خرید یا فروش (عددی بین ۱ و ۱−)، نسبتی از مقدار ثابت دارایی (برای مثال ۱,۰۰۰,۰۰۰ ریال) را می‌خرد یا می‌فروشـد. برای مثال، ۱ به معنای خرید به میزان ۱۰۰ درصد مقدار ثابت از آن دارایی و ۰/۵− به معنـای فـروش ۵۰ درصـد مقـدار ثابت از آن دارایی است.

مدیریت پویای سبد سهام به روش یادگیری تقویتی

همان‌طور که در معرفی الگوریتم یادگیری تقویتی نیز بیـان شـد، محـیط شـامل دو بخـش «نسـبت سـهام در سـبد» و «پیشنهاد خودمعامله‌گر» است. با در نظر گرفتن دو دارایی (پول نقد و یک سهم) تعداد حالات ممکن بـرای محـیط ۵۰۵ (۵۰۵ = ۵ × ۱۰۱) است و در هر یک از حالات محیط، ۱۰۱ عمل مختلف را باید امتیازدهی کند. با اجرای مدل، به میزان مشخصی تلاش می‌کنیم تمام تصمیم‌های ممکن در هر حالت ـ کـه بـا آن روبـه‌رو شـده‌ایـم ـ را ارزش گـذاری کنـیم. متأسفانه با بیشتر شدن تعداد دارایی‌ها، تعداد حالت‌ها در محیط به طور فزاینده‌ای زیاد می‌شود؛ از این رو حجم محاسبات نیز افزایش می‌یابد، به طوری برای اجرای عملیات شبیه‌سازی به مدت زمان بیشتر و سیستم‌های قوی‌تر نیاز پیدا می‌کند. برای نمایش عملکرد مناسب مدل، به نمایش نتایج مدل روی دو دارایی، یک سهم و پول نقد بسـنده مـی‌کنـیم؛ چراکـه تعداد دارایی تناقضی با رویکرد مدل ندارد.

نمایش نتایج طبق الگوی نموداری زیر خواهد بود، بر این اساس B-H معـرف اسـتراتژی خریـد و نگهـداری، FW مدل خرید با مقدار ثابت و RL مدل مدیریت پویای سبد سهام به روش یادگیری تقویتی است.

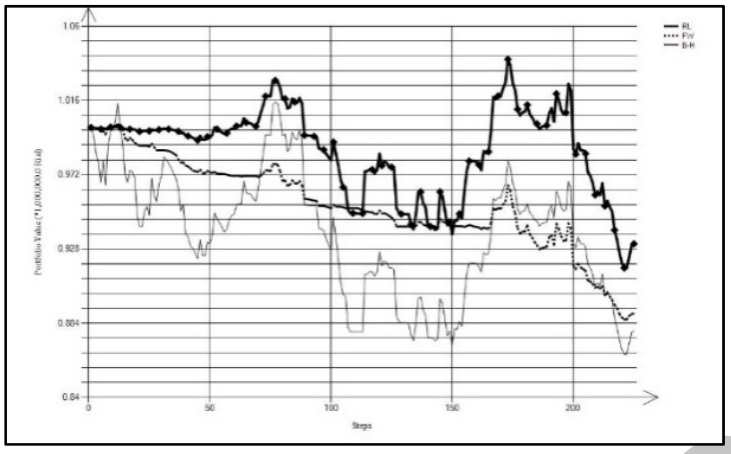
گفتنی است تمام نتایجی که در ادامه آورده شده است (بجز مواردی که قیدشده باشد)، همه با احتساب هزینه کامل معاملاتی بوده و هیچ‌گونه تخفیفی در هزینه معاملاتی لحاظ نشده است.



شکل ۲. نمونه‌ای از نمایش نتایج

نتایج به دست آمده در بازار نرمال

نمودار زیر (شکل ۳) نتایج سبد متشکل از پول نقد و یک سهم (فولاد) را نشان می‌دهـد. بـرای سـایر سـبدها نتـایج بـه صورت جدول و بدون نمودار ارائه شده است.



شکل۳. نمودار نتایج مدل برای سبدی متشکل از پول نقد + سهم (فولاد)

جدول ۲. نتایج مدل ارائه شده برای بازار نرمال

سبد	معیار
سبد متشکل از: پول نقد + سهم (فولاد)	٪ ۵/۳
سبد متشکل از: پول نقد + سهم (حفاری)	٪ ۲/۴
سبد متشکل از: پول نقد + سهم (وپارس)	٪ ۴/۷
سبد متشکل از: پول نقد + سهم (وصندوق)	٪ ۱/۶

خلاصه نتایج مدل در بازار نرمال: بر اساس نتایج، هر دو مدل پیشنهاد شده در بازار نرمال، همیشه جواب بهتری از استراتژی خرید و نگهداری دارند. علاوه بر این، همان‌طور که مشاهده می‌شود، مدل مدیریت پویای سبد سهام، همیشه جواب بهتری از مدل با خرید به مقدار ثابت داشته است و به طور واضح، مدل ارائه شده با رویکرد مدیریت پویای سبد سهام، در مدل ساده خرید به مقدار ثابت بر اساس پیشنهاد خودمعامله‌گرها بهبود ایجاد کرده است. در بالا نمودار مربوط به نتایج بازار در حالت نرمال برای یک سبد آمده است و برای سایر بازارها نتایج فقط توضیح داده شده است.

خلاصه نتایج در بازار صعودی: نکته‌ای که در بازار صعودی باید به آن توجه کرد این است که در این نوع بازار، بهترین استراتژی، خرید و نگهداری است، زیرا بازار همیشه در حال روند صعودی است و بهترین استراتژی تبدیل کل پول نقد در سبد به سهام است. مطابق با انتظارات، هیچ‌یک از مدل‌های ارائه شده نمی‌توانند به خوبیِ استراتژی خرید و نگهداری عمل کنند. اما مدل مدیریت پویای سبد سهام، به صورت مناسب و گاهی چشمگیر، مدل خرید با مقدار ثابت را بهبود داده است.

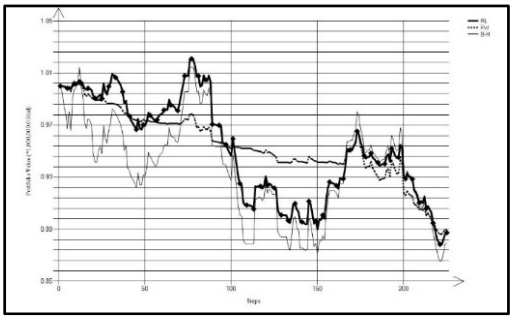
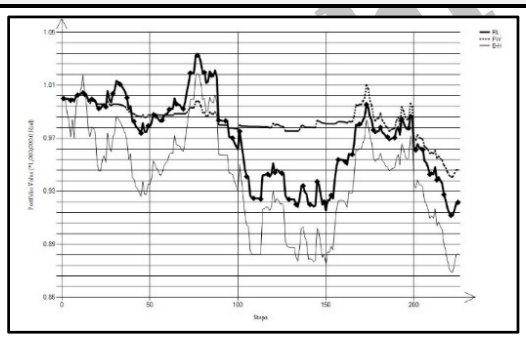
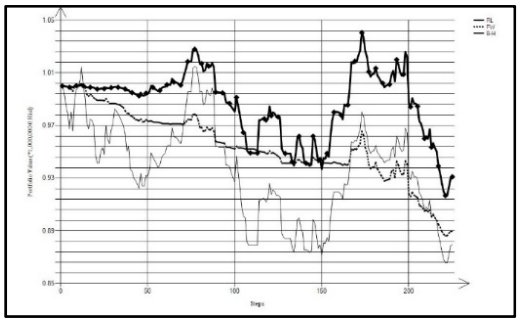
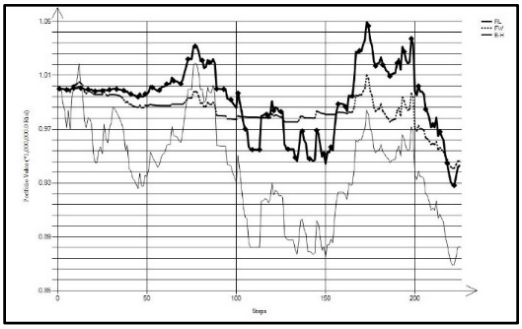
خلاصه نتایج در بازار نزولی: برعکس بازار صعودی، در بازار نزولی انتظار ما از یک مدل خوب افزایش نسبت پول نقد در سبد و استفاده از معدود روندهای صعودی پیش‌آمده طی دوره است. بر اساس نتایج این بازار، هر دو مدل به خوبی انتظارات ما را برآورده کرده‌اند و عملکرد بسیار بهتری از استراتژی خرید و نگهداری از خود بر جای گذاشته‌اند. علاوه بر آن، مدل مدیریت پویای سبد سهام، در دو مورد عملکردی بسیار بهتری از مدل خرید به مقدار ثابت داشته است.

تحلیل حساسیت روی هزینه معاملاتی

همان‌طور که پیش‌تر نیز بیان شد، به دلیل تعدد معاملات در مدل‌های معاملاتی با فراوانی زیاد، سیستم متحمل هزینه سنگینی بابت معاملات خود تحت عنوان هزینه معاملاتی خواهد شد. در مدل مدیریت پویای سبد سهام به روش یادگیری تقویتی، با تأثیر دادن نسبت هزینه‌های معاملاتی به ارزش سبد در تابع مطلوبیت، تلاش شده است تا مدل از انجام معاملات کم‌سود و با هزینه معاملاتی زیاد جلوگیری کند. به منظور بررسی میزان تأثیر هزینه معاملاتی روی مدل‌های ارائه شده، مدل مدیریت پویای سبد سهام را با ضرایب مختلفی از «معامله‌گریزی» (پیش‌تر با TA معرفی شد) طراحی کردیم. همان‌طور که مشاهده می‌شود، در ضرایب پایین معامله‌گریزی (۱۵)، تخفیف در هزینه معاملاتی موجب تغییر شایان توجهی در نتیجه هر دو مدل (مدیریت پویای سبد سهام و خرید به مقدار ثابت) شده است، اما با افزایش ضریب معامله‌گریزی (۳۵)، تخفیف در هزینه معاملاتی اثر چندانی بر مدل مدیریت پویای سبد سهام نداشته و تنها بر مدل خرید به مقدار ثابت تأثیرگذار بوده است. این موضوع گویای آن است که مدل ارائه شده به خوبی توانسته است تأثیر منفی هزینه معاملاتی را تقلیل دهد و این کاهش تأثیرپذیری به دلیل کاهش تعدد معاملات است که می‌تواند از انجام معاملات کوچک با بازده کم جلوگیری کند.

با اینکه مدل ارائه شده به گونه‌ای طراحی شده است که بتواند از انجام معاملات با سود پایین جلوگیری کند، انتظار می‌رود با تخفیف در هزینه معاملاتی برای سیستم‌های معاملاتی با فراوانی زیاد، بدون بالا بردن ضریب معامله‌گریزی بتوان بازده مناسبی از سیستم گرفت.

جدول ۳. نتایج مدل‌ها در بازار نرمال با توجه به هزینه معاملاتی و ضریب معامله گریزی

TA	هزینه کامل معاملاتی	با نصف هزینه معاملاتی
۱۵		
۳۵		

نتیجه‌گیری و پیشنهادها

طبق جدول ۴ که خلاصه نتایج عملکرد مدل‌ها را نشان می‌دهد، در بازارهای نرمال و نزولی هر دو مدل ارائه شده بهتر از خرید و نگهداری عمل کردند، ولی در بازار صعودی هیچ‌یک از مدل‌ها نتوانستند عملکرد بهتری از خرید و نگهداری داشته باشند. مدل مدیریت پویای سبد سهام که با رویکرد بهبود عملکرد مدل ساده خرید به مقدار ثابت بر اساس پیشنهاد خودمعامله‌گرها ارائه شده بود در همه بازارها از مدل نخست عملکرد بهتری دارد.

جدول ٤. خلاصه و جمع‌بندی نتایج مدل‌های ارائه شده

بازار	FW	B-H	RL	B-H	FW	RL
نرمال	>		>		<	
صعودی	>		<		<	
نزولی	>		>		<	

این پژوهش جزء نخستین پژوهش‌های ایران است که در زمینه سیستم‌های معاملاتی روی داده‌های درون‌روزی کار کرده است. از این لحاظ نتایج این پژوهش را تنها می‌توان با هم مقایسه کرد. در هر صورت، نتایج مدل درستی نتایج پژوهش محققانی همچون محمدی (۱۳۸۳)، ستایش، تیمورزاده، پورموسی و ابوذری (۱۳۸۸)، صمدی، ایزدی‌نیا و داورزاده (۱۳۸۹) و رزمی، جولای و امامی (۱۳۸۶) را در مورد کارایی استفاده از اندیکاتورهای تحلیل تکنیکال در بورس اوراق بهادار تهران تأیید می‌کند، اما از آنجا که مدل ارائه شده در این پژوهش معاملات درون‌روزی را مد نظر قرار داده و تمام پژوهش‌های اشاره شده بر اساس داده‌های انتهای روز بوده‌اند، نمی‌توان به صورت مشخص بین این نتایج مقایسه‌ای انجام داد.

در مدل پژوهش حاضر، قابلیت تأثیر ریسک سبد سهام لحاظ شده است، پیشنهاد می‌شود با بهبود سرعت پردازش اطلاعات در شبیه‌سازی مدل، شرایط برای افزایش تعداد دارایی‌ها در سبد فراهم شود. وجود چند سهم با همبستگی‌های متفاوت در یک سبد، موجب بهبود عملکرد سیستم خواهد شد. از سوی دیگر، پیاده‌سازی چنین سیستم‌هایی در سایر بازارها، مانند بازار معاملات ارز خارجی (فارکس) نیز به دلیل تفاوت در هزینه‌های معاملاتی این بازارها با بازار اوراق بهادار می‌تواند نتایج مطلوبی داشته باشد.

منابع

بهلولی خدادادی، محمد (۱۳۹۱). *مدیریت پویای سبد سهام با استفاده از یادگیری تقویتی*. پایان‌نامه کارشناسی ارشد، تهران: دانشکده علوم اقتصادی.

راعی، رضا؛ باجلان، سعید (۱۳۸۷). شناسایی و مدل‌سازی اثرات تقویمی بورس اوراق بهادار تهران با استفاده از مدل های ARCH و GARCH. *فصلنامه پژوهش‌های اقتصادی (رشد و توسعه پایدار)*، ۸(۴)، ۲۱-۴۷.

رزمی، جعفر؛ جولای، فریبرز؛ امامی، امیرعباس (۱۳۸۶). یک رویکرد «بوت استرپ» برای مقایسه سودآوری شاخص‌های تحلیل تکنیکی ـ بورس اوراق بهادار تهران. تحقیقات اقتصادی، ۴۲(۴)، ۱-۲۶.

ستایش، محمدرضا؛ تقی‌زاده شیاده، تیمور؛ پورموسی، علی اکبر؛ ابوذری لطف، علی. (۱۳۸۸). امکان‌سنجی به کارگیری شاخص‌های تحلیل تکنیکی ـ فنی ـ در پیش‌بینی روند قیمت سهام در بورس اوراق بهادار تهران. *فصلنامه بصیرت*، ۴۲ (۱)، ۱۵۵-۱۷۷.

صمدی، سعید؛ ایزدی نیا، ناصر؛ داورزاده، مهتاب (۱۳۸۹). کاربرد بهره‌گیری از تحلیل تکنیکی در بورس اوراق بهادار تهران (رویکردی بر میانگین متحرک). *پیشرفت‌های حسابداری*، ۲(۱)، ۱۲۱-۱۵۴.

محمدی، شاپور (۱۳۸۳). تحلیل تکنیکی در بورس اوراق بهادارتهران. *فصلنامه تحقیقات مالی*، ۶(۱)، ۹۷- ۱۲۹.

References

Bohluli Khodadadi, M. (2010). *Dynamic portfolio management using reinforcement learning*. Master's Thesis. University of Economic Sciences, Tehran. (*in Persian*)

De la Fuente, D., Garrido, A., Laviada, J., & Gómez, A. (2006). Genetic algorithms to optimise the time to make stock market investment. *In Proceedings of the 8th annual conference on Genetic and evolutionary computation,* 1857-1858. ACM.

Dempster M.A.H. & Romahi Y. (2002). Intraday FX Trading: An Evolutionary Reinforcement Learning Approach. *In: Yin H., Allinson N., Freeman R., Keane J., Hubbard S. (eds) Intelligent Data Engineering and Automated Learning — IDEAL 2002. IDEAL 2002. Lecture Notes in Computer Science, vol 2412.* Springer, Berlin, Heidelberg

Dempster, M. A. H. & Jones, C. M. (2002). Can channel pattern trading be successfully automated? The *European Journal of Finance*, 8 (3), 275-301.

Dempster, M. A. H., & Jones, C. M. (2000). *The profitability of intra-day FX trading using technical indicators.* Judge Institute of Management, University of Cambridge.

Dempster, M. A. H., Payne, T. W., Romahi, Y., & Thompson, G. W. P. (2001). Computational learning techniques for intraday FX trading using popular technical indicators. *IEEE Transactions on Neural Networks,* 12(4), 744-754.

Duda, R. O., Hard, P. E. & Stork, D. G. (2000). *Pattern Classification*. New York, Wiley-Interscience.

Duvinage, M., Mazza, P., & Petitjean, M. (2013). The intra-day performance of market timing strategies and trading systems based on Japanese candlesticks. *Quantitative Finance,* 13(7), 1059-1070.

Fan, A. & Palaniswami, M. (2001). Stock selection using support vector machines. *Proceedings of the International Joint Conference on Neural Networks,* 3, 1793-1798.

Gao, X. & Chan, L. (2000). An algorithm for trading and Portfolio Optimization using Q-Learning and Sharp Ration Maximization. *Proceedings of the international conference on neural information processing,* 832-837.

Jangmin, O., Lee, J. W., Lee, J., & Zhang, B. T. (2004). Dynamic asset allocation exploiting predictors in reinforcement learning framework. *In European Conference on Machine Learning,* Springer Berlin Heidelberg, 298-309.

Jangmin, O., Lee, J., Lee, J. W. & Zhang, B. (2005). Dynamic Asset Allocation for Stock Trading Optimized by Evolutionary Computation. *IEICE Transactions on Information and Systems*, 88 (6), 1217-1223.

Jangmin, O., Lee, J., Lee, J. W. & Zhang, B. T. (2006). Adaptive stock trading with dynamic asset allocation using reinforcement learning. *Information Sciences*, 176, 2121-2147.

Jones, C. M. (1999). *Automated technical foreign exchange trading with high frequency data*. Doctoral dissertation, University of Cambridge.

Kendall, S. M., & Ord, K. (1997). *Time Series*. New York, Oxford.

Lee, J. W., & Zhang, B. T. (2002). Stock trading system using reinforcement learning with cooperative agents. *In Proceedings of the Nineteenth International Conference on Machine Learning,* Morgan Kaufmann Publishers Inc, 451-458.

Lee, J. W., Park, J., Jangmin, O., Lee, J., & Hong, E. (2007). A multiagent approach to Q-learning for daily stock trading. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 37(6), 864-877.

Lee, J. W., Sung-Dong, K. I. M., Jongwoo, L. E. E., & Jinseok, C. H. A. E. (2003). An intelligent stock trading system based on reinforcement learning. *IEICE Transactions on Information and Systems*, 86(2), 296-305.

Manahov, V., Hudson, R., & Gebka, B. (2014). Does high frequency trading affect technical analysis and market efficiency? And if so, how? *Journal of International Financial Markets, Institutions and Money*, 28, 131-157.

Mohamadi, Sh. (2004). Technical analysis in Tehran Stock Exchange. *Financial Research Journal*, 6(1), 97-129. (*in Persian*)

Moody, J. & Saffell, M. (2001). Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks*, 12(4), 875, 889.

Neely, C. J., & Weller, P. A. (2003). Intraday technical trading in the foreign exchange market. *Journal of International Money and Finance*, 22(2), 223-237.

Neuneier, R. (1998). Enhancing Q-learning for optimal asset allocation. *Advances in Neural Information Processing Systems*, 10, 936-942.

Oliver Mihatsch, R. N. (2002). Risk-Sensitive Reinforcement Learning. *Machine Learning*, 49, 267-290.

Raei, R. & Bajelan, S. (2007). Detecting and modeling of calendar effects in Tehran Stcok Exchange. *Quarterly Journal of The Economic Research,* 8 (4), 21-47. (*in Persian*)

Razmi, J., Julay, F., & Emami, A. (2007). A Bootstrap approach for comparing the profitability of technical analysis indicators – Tehran Stock Exchange. *Journal of Economic Researchs*, 85, 85-110. (*in Persian*)

Rodríguez-González, A., García-Crespo, Á., Colomo-Palacios, R., Iglesias, F.G. and Gómez-Berbís, J.M. (2011). CAST: Using neural networks to improve trading systems based on technical analysis by means of the RSI financial indicator. *Expert systems with applications,* 38(9), 11489-11500.

Saad, E. W., Prokhorov, D. V. & Wunsch, D. C. (1998). Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. *IEEE Transactions on Neural Networks,* 9(6), 1456-1470.

Samadi, S., Izadinia, N., & Davarzadeh, M. (2010). The application of exploiting technical analysis in Tehran Stock Exchange (an approach to moving average). *Journal of Accounting Advances*, 2(1), 121-154. (*in Persian*)

Setayesh, M., Taghizadeh, T., Poormoosa, A., & Abuzari, A. (2008). Feasibility of exploiting technical analysis indicators in predicting the price trend of stocks in Tehran Stock Exchange. *Quarterly Basirat,* 7, 155-177. (*in Persian*)

Sutton, R. S. & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MIT Press.

Tanaka-Yamawaki, M., & Tokuoka, S. (2007). Adaptive use of technical indicators for the prediction of intra-day stock prices. *Physica A: Statistical Mechanics and its Applications*, 383(1), 125-133.

Watkins, C. (1989). *Learning from delayed rewards*, Ph.D, Cambridge University.

Yamamoto, R. (2012). Intraday technical analysis of individual stocks on the Tokyo Stock Exchange. *Journal of Banking & Finance*, 36(11), 3033-3047.