فصلنامه مهندسی مالی و مدیریت اوراق بهادار
دوره شانزده، شماره شصت و دو، بهار ۱۴۰۴
نوع مقاله : علمی پژوهشی
صفحات :۱۸۶-۱۶۸

استفاده از خوشه‌بندی برای انتخاب زوج‌های مناسب در راهبرد معامله‌ی زوج‌ها

مجید ابتیاع[۱]
محمدرضا آریافر[۲]
سیدمحمد حسینی[۳]

تاریخ دریافت مقاله: ۱۴۰۲/۰۹/۲۳ تاریخ پذیرش مقاله: ۱۴۰۳/۱۱/۱۴

چکیده

برای سرمایه‌گذاری مناسب و شناسایی موقعیت‌های درست خرید و فروش، وجود یک سیستم و راهبرد مشخص و کارا لازم است. سیستم معامله‌ی زوج‌ها یکی از مشهورترین و قابل‌فهم‌ترین سیستم‌های معاملات الگوریتمی می‌باشد. در این سیستم، از یک زوج سهام که در بلندمدت روند قیمتی یکسان دارند (خاصیت بازگشت به میانگین) و در کوتاه‌مدت دارای نوسانات (اسپرد) است، بهره گرفته می‌شود. مهم‌ترین نکته در این سیستم، شناسایی و پیدا کردن چنین زوج سهم‌هایی است. در این پژوهش از یک رویکرد نوین و متفاوت برای شناسایی زوج سهم‌ها و همچنین یافتن موقعیت‌های درست خرید و فروش در معامله‌ی زوج‌ها استفاده می‌شود. در گام اول، ابتدا یک زوج سهام یا دارایی که یک رابطه آماری بلندمدت دارند انتخاب می‌شوند. برای محاسبه فاصله بین هر زوج سهام و شباهت حرکتی آن‌ها از روش انحراف زمان پویا استفاده می‌شود. سپس از روش خوشه‌بندی سلسله مراتبی برای خوشه‌بندی سهم‌ها بهره گرفته می‌شود و جفت سهمی که میزان شباهت بیشتری دارند انتخاب می‌شوند. در گام دوم از روش ماشین بردار پشتیبان برای شناسایی موقعیت‌های خرید و فروش استفاده می‌شود. به‌منظور بررسی عملکرد و کارایی روش از شاخص S&P که شامل ۵۰۰ سهام برتر بازار بورس نیویورک می‌باشد، استفاده شده است.

کلمات کلیدی

معاملات الگوریتمی، معامله‌ی زوج‌ها، انحراف زمان پویا، خوشه‌بندی، سهام

۱-گروه ریاضی، دانشکده علوم پایه، دانشگاه آیت ا... بروجردی(ره)، بروجرد، ایران. majid.ebtia@gmail.com
۲-گروه اقتصاد منابع و انرژی، دانشکده علوم اداری و اقتصاد، دانشگاه فردوسی مشهد، مشهد، ایران. m.ariafar@mai.ac.ir
۳-گروه ریاضی، دانشکده علوم پایه، دانشگاه آیت ا... بروجردی(ره)، بروجرد، ایران. (نویسنده مسئول) sm.hoseini@abru.ac.ir

مقدمه

امروزه حجم بالای معاملات در بازارهای مالی و فراگیر شدن معاملات آنلاین، موجب پیدایش و پیشرفت معاملات الگوریتمی و سیستم‌های مالی خودکار شده است. هدف از معاملات الگوریتمی، خودکار کردن معاملات و مدیریت فرآیندهای معاملاتی در سرعت بالا توسط الگوریتم‌های رایانه‌ای است، که با کاهش چشم‌گیر دخالت انسان همراه است. معامله‌ی زوج‌ها (اسپردی) یکی از جذاب‌ترین و مهم‌ترین راهبردهاست و در بازارهای غیرقابل‌پیش‌بینی و بی‌ثبات مالی قابل‌استفاده است. این دیدگاه، یک راهبرد معاملاتی خنثی در بازار آربیتراژ آماری است. در این روش‌ها، معاملات دارایی‌ها و سهم‌هایی که معمولاً همبستگی بالایی دارند مورد توجه و بررسی قرار می‌گیرند و هدف، کسب سود از خرید یک دارایی و فروش یک دارایی دیگر به‌صورت هم‌زمان است. معامله‌ی زوج‌ها از قدیمی‌ترین سیستم‌های معاملات الگوریتمی است و معامله‌گران را قادر می‌سازد که در حرکات مختلف بازار ازجمله نوسانات قیمت به‌صورت صعودی یا نزولی کسب سود کنند.

ایده معامله‌ی زوج‌ها اولین‌بار در اواسط دهه ۱۹۸۰ توسط مورگان استنلی معرفی شد و برای موفقیت‌های بزرگ در صنعت مالی مورد استفاده قرار گرفت. یک نکته بسیار مهم برای راهبرد معامله‌ی زوج‌ها، انتخاب یک زوج ابزار مالی مناسب است. به‌عنوان مثال، اگر قیمت اوراق بهادار A افزایش یابد زمانی که قیمت اوراق B افزایش یابد، به‌نظر می‌رسد که این زوج سهام می‌توانند برای معامله‌ی زوج‌ها استفاده شود. با این‌حال، رابطه صریح بین قیمت‌ها ممکن است به‌اندازه کافی برای یک زوج خوب نباشد. زوج‌های خوب باید دارای ویژگی‌های درونی یکسانی باشند. هدف از انجام این پژوهش، ارائه یک رویکرد نوین مبتنی بر الگوریتم ماشین بردار پشتیبان در معامله‌ی زوج‌ها برای بهبود شناسایی بهتر سیگنال‌های معاملاتی است که سبب کاهش زیان و افزایش سودآوری می‌شود. راهبرد معامله‌ی زوج‌ها شامل رویکردهای مختلفی است که می‌توان به رویکرد فاصله‌ای، پیش‌بینی ترکیبی، تفاضل قیمتی تصادفی و هم‌انباشتگی که چهار رویکرد پرکاربرد در این زمینه هستند، اشاره کرد (سریرامولو، ۱۹۹۹).

یکی از موارد بسیار اساسی برای کسب سود در بازار، استفاده از ابزارها و تحلیل تکنیکال است، زیرا که این قواعد منطبق با منطق‌های ریاضی و آماری است و می‌تواند به‌خوبی، نقاط ورود و خروج از سهام‌های مختلف را برای معامله‌گران مشخص سازد. برای انتخاب یک جفت سهم مناسب، جهت پیاده‌سازی راهبرد معامله‌ی زوج‌ها، باید از اندیکاتورهای[1] تحلیل تکنیکال استفاده شود. در واقع، جفت سهم مورد نظر نه‌تنها باید روند قیمتی مشابهی داشته باشند، بلکه باید تعدادی از شاخص‌های تحلیل

تکنیکال آن‌ها نیزمشابه یکدیگرباشند، تا بتوان راهبرد مناسب را برای معامله در نظر گرفت. به عبارتی دیگر، اندیکاتورها ابزارهایی هستندکه با استفاده از اطلاعات قیمت، زمان، حجم و در نظرگرفتن توابع ریاضی می‌توانندداطلاعات‌مفیدی را در جهت انتخاب سهم و کسب‌سود، دراختیارمعامله‌گران قراربدهند.

در این پژوهش، از مدل اسپرد، مدل بازگشت به میانگین اورنشتاین-اولن‌بک[۲] و ماشین بردار پشتیبان برای ایجاد یک راهبرد معاملاتی استفاده می‌شود. ابتدا، جزئیات مدل اسپرد و بازگشت به میانگین اورنشتاین-اولن‌بک ارایه می‌شود. سپس از دو مدل گفته‌شده برای مدل‌سازی شاخص‌های فنی دو اوراق بهادار استفاده شده است. همچنین سیگنال‌های معاملاتی، با پردازش انواع مختلف اسپرد ساخته می‌شود. از این سیگنال‌های معاملاتی به‌عنوان ویژگی‌های ورودی برای دسته‌بندی ماشین بردار پشتیبان استفاده شده است. به‌جای استفاده از روش‌های سنتی پیش‌بینی بر مبنای داده‌های گذشته، از دسته‌بند ماشین بردار پشتیبان برای شناسایی سیگنال‌های معاملاتی استفاده خواهد شد. برای این منظور، منابع اولیه قیمت‌گذاری برای نمونه‌های برچسب‌دار بازسازی شد و از دو روش برای بازسازی نمونه‌های برچسب‌دار استفاده شده است: یکی برای اندازه‌گیری توانایی راهبرد در معرفی فرصت‌های سود و دیگری برای اندازه‌گیری توانایی راهبرد در جهت‌دهی به پیش‌بینی‌ها. در این مطالعه، برای انتخاب راهبرد کسب سود مناسب‌تر در معامله‌ی زوج‌ها از اندیکاتورهای میانگین متحرک ساده[۳]، میانگین متحرک وزنی[۴]، شاخص جریان پول[۵] و شاخص قدرت نسبی[۶] استفاده می‌شود.

مبانی نظری و مروری بر پیشینه پژوهش

امروزه با افزایش معاملات برخط، نیاز است تا داده‌های بازارهای مالی با سرعت بیشتری تجزیه و تحلیل شوند. از گذشته تاکنون پژوهش‌های مختلفی در زمینه معاملات الگوریتمی و همچنین پژوهش‌هایی برای پیشرفت و توسعه راهبرد معامله‌ی زوج‌ها صورت گرفته است که همگی نشان از عملکرد و سودآوری این راهبرد در بازارهای مالی است.

معامله‌ی زوج‌ها معمولاً برای تعیین راهبرد و کسب سود در بازارهای مالی مورد استفاده قرار می‌گیرد. این رویکرد از روش‌های متداول برای اتخاذ تصمیم مناسب جهت خرید و فروش جفت دارایی بهره می‌گیرد. معمولاً روش‌های مختلفی در این زمینه مورد بررسی قرار می‌گیرد، که در ادامه به آن‌ها اشاره خواهد شد. در معامله‌ی زوج‌ها تلاش می‌شود تا زوج‌های سهامی که تا حدودی به‌صورت معکوس حرکت می‌کنند، شناسایی شوند. این روش معمولاً با استفاده از مدل‌های آماری برای تشخیص نوسانات ناهماهنگ در قیمت‌ها انجام می‌شود (طادی، ۱۳۹۷). برای روشن شدن موضوع، ادبیات مورد استفاده در مطالعات مختلف مورد بررسی قرار داده شد. در واقع، هدف از معامله‌ی زوج‌ها شناسایی

موقعیت‌های مناسب جهت ورود به معامله در یک بازار ناکارآمد که منجر به قیمت‌گذاری‌های غیرمنطقی و اشتباه اوراق بهادار شده است، می‌باشد (لیو و وو، ۲۰۱۳).

راهبرد معامله‌ی زوج‌ها توسط گروهی از افراد با زمینه تخصصی در علوم مختلف به‌وجود آمد. این گروه توسط نونزیو تارتاگلیا که تحلیل‌گر کمّی در زمینه‌ی علوم مالی بود، مدیریت و راهبری می‌شد. هدف مهمی که این گروه به‌دنبال آن بودند، این بود که معاملات و راهبرد با استفاده از کامپیوتر انجام گیرد و عامل انسانی که دارای ضریب خطای بالایی هست، حذف شود و کامپیوتر به‌عنوان تصمیم‌گیرنده به انجام معاملات بپردازند. این راهبرد معاملاتی، با اعتماد به تحلیل دقیق داده‌ها و استفاده از الگوریتم‌های پیچیده، به تصمیم‌گیری خودکار و حذف تأثیر احتمالی خطاهای انسانی پرداخت. تحلیل‌های کمّی انجام‌شده توسط نونزیو تارتاگلیا در زمینه‌ی علوم مالی، این گروه را قادر ساخت تا با دقت بالا به تشخیص الگوهای معاملاتی بپردازد و به‌طور خودکار و سریع، تراکنش‌های مالی را اجرا کند. هرچند که این راهبرد تا مدتی موفق بود، اما چالش‌ها و پیچیدگی‌های بازار مالی ممکن است باعث شکست آن گردد. منحل شدن این گروه نشان‌دهنده‌ی ضرورت پیشرفت مداوم در تحلیل داده‌ها و تطابق با متغیرهای متنوع بازارهای مالی است (فلاح پور و همکاران، ۲۰۱۹).

از مهم‌ترین رویکردها در تعیین راهبرد معامله‌ی زوج‌ها، می‌توان به چهار مورد اشاره کرد که عمده تفاوت این موارد در میزان محاسبه اسپرد و شدت بازگشت آن‌ها به میانگین است. این رویکردها عبارت‌اند از رویکرد فاصله‌ای، همگرایی زوجی، رگرسیون زمانی و مقایسه‌ای.

ازرویکردهای مهم درتحلیل و انجام راهبردمعامله‌ی‌زوج‌ها، رویکردفاصله‌ای است که در این پژوهش نیز مورد استفاده قرارخواهدگرفت. در رویکرد فاصله‌ای، نوسانات قیمتی جفت سهم مورد نظر به‌صورت فاصله‌ی قیمتی و یا مجموع مربعات فواصل بین دو سری قیمت نرمال شده، محاسبه می‌شود (مسعود طادی و همکاران ۱۳۹۶). در پژوهشی که سارمنتو و هورتا (۲۰۲۰) ارائه کردند، به بررسی معامله‌ی زوج‌ها با استفاده از رویکرد فاصله‌ای پرداختند. آن‌ها در پژوهش خود، در فرایندی که منجر به تشکیل زوج‌های معاملاتی می‌شود، محدودیتی مانند بزرگی یا کوچکی صنعت را لحاظ نکردند و تنها با شرط حداقل فاصله، زوج‌های سهام را بررسی نمودند. رویکرد فاصله‌ای که در این پژوهش به آن اشاره شده، یکی از رویکردهای مهم در تحلیل و انجام راهبرد معامله‌ی زوج‌ها است. این رویکرد بر اندازه فاصله‌های قیمتی بین دو سهم مختلف تمرکز دارد و از این فاصله‌ها به عنوان یک معیار برای انجام معاملات استفاده می‌کند. محاسبه فاصله‌های قیمتی می‌تواند به‌عنوان یک شاخص مفید برای تشخیص الگوهای بازارو تصمیم‌گیری درمورد زمان‌مناسب برای ورودیاخروج از معاملات، مورداستفاده قرارگیرد.

سایر رویکردها در تحلیل و انجام راهبرد معامله‌ی زوج‌ها می‌توانند شامل موارد زیر باشند:

۱. رویکرد همگرایی زوجی[۷]: در این رویکرد، تلاش بر این است که زوج‌های سهامی که تا حدودی به‌صورت همگرا حرکت می‌کنند، شناسایی شوند. این همگرایی نشان‌دهنده ارتباط ثابت زمانی بین دو سهم است و ممکن است در انجام راهبردهای معاملاتی مؤثر باشد (هاک، ۲۰۱۵).

۲. رویکرد رگرسیون زمانی[۸]: این رویکرد بر اساس تحلیل رگرسیون زمانی بر روی داده‌های زمانی قیمت‌ها و بازده‌های سهام استوار است. این مدل‌ها ممکن است برای پیش‌بینی حرکت‌های آینده قیمت‌ها و ارتباطات بین سهام مفید باشند (لوو، ۲۰۰۹).

۳. رویکرد مقایسه‌ای[۹]: در این رویکرد، تفاوت‌های نسبی بین عملکرد دو سهم در طول زمان مورد تجزیه و تحلیل قرار می‌گیرد. این مدل به افتراق قدرت نسبی سهام و انتخاب سهم قوی‌تر بر اساس عملکرد تاریخی می‌پردازد (آلبانس، ۲۰۱۳).

هر یک از این رویکردها می‌توانند برای تحلیل زوج‌های معاملاتی و انجام راهبردهای مختلف در بازارهای مالی مورد استفاده قرار گیرند.

چندین پژوهش از سال‌های گذشته وجود دارد که روش‌ها و نتایج جالبی در رابطه با معامله‌ی زوج‌ها به نمایش می‌گذارند. در فرآیند ماشین بردار پشتیبان برای راهبردهای معامله‌ی زوج‌ها، نویسندگان ابتدا از رگرسیون خطی برای محاسبه و تولید سیگنال‌های معاملاتی استفاده می‌کنند و ضریب ۱.۱۴ را از آزمون برگشتی به‌دست می‌آورند و همچنین یک راهبرد معاملاتی را با استفاده از الگوریتم EM و Kalman filter پیشنهاد می‌کنند (شوارتز و همکاران، ۲۰۱۴). همچنین، در فرایند ماشین بردار پشتیبان در آربیتراژ آماری، نویسندگان آربیتراژ شاخص را بررسی می‌کنند. آن‌ها ابتدا از تحلیل مؤلفه اصلی برای انتخاب ۱۲ مؤلفه اول از ۱۰۰ مورد (مؤلفه‌های شاخص FTSE100) و سپس از مدل اورنشتاین-اولن‌بک برای یک فرآیند بازگشت میانگین استفاده می‌کنند؛ که به تولید سیگنال معاملاتی کمک می‌کند.

مورائس‌سارمنتو (۲۰۲۱) در پژوهش خود از یک مدل بازگشت به میانگین تحت فرآیند مکانیزم یادگیری ماشین برای بررسی جفت سهم گوگل و گوگ بهره برده است و دو مبحث را مورد بررسی و مدل‌سازی قرار داده است: اولین مورد نه‌تنها اسپرد[۱۰] قیمتی بلکه چند شاخص فنی را نیز در برگرفته است و دوم، استفاده از دومعیار جدیدبرای اندازه‌گیری راهبرد معاملاتی بجای استفاده از روش آزمایش.

ابراهیمی‌پور و داودی (۱۴۰۰) در پژوهش خود با استفاده از یک مدل حالت-فضای خطی فیلتر

کالمن به بررسی معامله‌ی زوج‌ها در بورس اوراق بهادار پرداخته‌اند. در این مطالعه علاوه بر مدل حالت-فضا، از روش بازگشت به میانگین استفاده‌شده است؛ که درنهایت مدل آن‌ها به نتایج مناسبی دست‌یافت و سودآوری را تحت این نوع راهبرد به‌نمایش گذارد.

در مطالعه‌ای که فلاح‌پور و حکیمیان (۱۳۹۶) در رابطه با بهینه‌سازی راهبرد معامله‌ی زوج‌ها و روش یادگیری تقویتی انجام داده‌اند، معامله‌ی زوج‌ها را از قدیمی‌ترین سیستم‌های معاملاتی معرفی کرده‌اند. همچنین از اصول مهم در راهبرد معامله‌ی زوج‌ها را وجود روابط تعادلی بلندمدت یا همان بازگشت به میانگین دانسته‌اند. این مطالعه با استفاده از رویکرد هم‌انباشتگی نشان داده است که این راهبرد در تمامی شرایط بازار توسط معامله‌گران حقیقی و حقوقی می‌تواند مورد استفاده قرار گیرد، خواه بازار در شرایط رکود[۱۱] یا رونق[۱۲] باشد خواه در شرایط خنثی[۱۳]. معامله‌ی زوج‌ها همراه با آربیتراژ آماری، از راهبردهایی بوده‌اند که از اواخر سال‌های ۱۹۹۰ توسط صندوق‌های تأمین مالی استفاده‌شده است (تیلور، ۲۰۰۴).

بهبود عملکرد الگوریتم معامله‌ی زوج‌های پربسامد، با تلفیق الگوریتم ژنتیک و کنترل فرایند آماری فازی موضوعی بود که دستوری و مرادپور (۱۴۰۰) بر اساس آن پژوهش خود را ارائه کردند. نتیجه این پژوهش بیانگر این موضوع بود که با استفاده از این نوع مدل‌سازی می‌توان افزایش بازدهی را داشت.

دولو و یزدی (۱۴۰۱) نیز در مطالعه خود به بررسی راهبرد معامله‌ی زوج‌ها در بازار آتی سکه پرداختند که نتیجه این پژوهش به اثربخشی این نوع راهبرد تأکید دارد. این نوع راهبردها برای کسب سود از نوسانات در بازارها است و می‌توان به‌صورت مدیریت‌شده از آن‌ها استفاده کرد. معامله زوج‌ها را می‌توان در کوتاه‌مدت و بلندمدت مورداستفاده قرارداد تا بتوان در زمان‌های مختلف سود کسب کرد. راهبرد معامله‌ی زوج‌ها بدون ریسک نیست، در نتیجه اشتباه در محاسبات و تشخیص دو جفت سهم می‌تواند منجر به شکست و ضرر در معاملات شود (بلازکز و همکاران، ۲۰۱۸).

معرفی اندیکاتورها و اسیلاتورها[۱۴] (نوسان‌نماها)

در ادامه، اندیکاتورهای تکنیکالی که در این پژوهش مورد استفاده قرار گرفته‌اند، معرفی می‌شود. اندیکاتورهایی که رفتارهای مشابهی برای جفت سهم مورد نظر ارائه دهند، مناسب هستند. مورائس‌سارمنتو (۲۰۲۱)، در پژوهش خود ۴ اندیکاتور زیر را مورد استفاده قرار داده است، که در این پژوهش نیز از آن‌ها بهره برده خواهد شد:

۱- میانگین متحرک ساده[۱۵]: یکی از متداول‌ترین و پرکاربردترین اسیلاتورهای تکنیکال است. این نوسان‌نما به این دلیل که کاربرد بسیار ساده و قابل‌گسترشی دارد، پایه‌ی اساسی اسیلاتورها است (الیس و پاربری، ۲۰۰۵). میانگین متحرک ساده وزن قیمت را در همه‌ی زمان‌ها یکسان در نظر می‌گیرد، به این معنی که قیمت در زمان‌های دورتر نیز اطلاعات نظیر قیمت‌های زمان‌های نزدیک‌تر را در خود جای می‌دهند. میانگین متحرک ساده به‌صورت زیر محاسبه می‌شود:

$$SMA_t = \sum_{i=t-L+1}^{t} p_i^{close},$$

این اسیلاتور در زمان t برای قیمت بسته‌شدن سهم i را مدنظر قرار می‌دهد و در آن L طول دوره‌ی مورد نظر است.

۲- میانگین وزنی متحرک[۱۶]: تفاوت این میانگین با میانگین متحرک ساده در وزن‌دهی به قیمت‌ها است، به این معنا که قیمت‌های نزدیک‌تر وزن بیشتری را نسبت به قیمت‌های دورتر در خود جای می‌دهند (عطاران و صالحی، ۱۴۰۰). میانگین متحرک وزنی عبارت است از:

$$WMA_t = \sum_{i=t-L+1}^{t} w_i\, P_i^{close},$$

به طوری‌که: $w_i = \frac{i-(t-L)}{\sum_{j=t-L+1}^{t} j-(t-L)}.$

۳- شاخص جریان نقدینگی[۱۷]: این شاخص نمایانگر میزان ورود و خروج نقدینگی در بازار سرمایه می‌باشد (لو، ۲۰۱۸). نحوه‌ی محاسبه‌ی آن به‌صورت زیر است:

$$MFI_t = 100 - \left(\frac{100}{1 + Money\ Flow\ Ratio_t}\right),$$

نسبت جریان نقدینگی در زمان t را می‌توان با استفاده از معادله‌ی زیر نشان داد:

$$\frac{L - period\ Positive\ Money\ Flow_t}{L - period\ Negative\ MoneyFlow_t},$$

که در آن $Positive\ Money\ Flow_t$ یا نقدینگی مثبت، جمع تمام روزهایی است که میانگین سه قیمت بالاتر از روز قبل خود است، و $Negative\ MoneyFlow_t$ یا جریان نقدینگی منفی، حاصل جمع تمام روزهایی است که میانگین سه قیمت پائین‌تر از روز قبل خود است. جریان نقدینگی در زمان t نیز با استفاده از فرمول زیر محاسبه می‌شود:

$$Money\ Flow\ Ratio_t = volume * \frac{P_t^{high} + P_t^{low} + P_t^{close}}{3}.$$

زمانی جریان نقدینگی مثبت است که $\frac{P_{t-1}^{high}+P_{t-1}^{low}+P_{t-1}^{close}}{3} < \frac{P_t^{high}+P_t^{low}+P_t^{close}}{3}$.

۴- شاخص قدرت نسبی[۱۸]: این شاخص از اندیکاتورهای مهم در تحلیل تکنیکال می‌باشد. این اندیکاتور، میزان اشباع خرید یا فروش را با استفاده از محاسبات ریاضی به‌نمایش می‌گذارد و معامله‌گران بسیاری از این اندیکاتور جهت تحلیل سهم‌ها استفاده می‌کنند (آدریان،۲۰۱۱). با استفاده از معادله زیر می‌توان به‌روشنی اندیکاتور شاخص قدرت نسبی را مورد بررسی قرار داد:

$$RSI_t = 100 - \left(\frac{100}{1+RSt}\right),$$

$$RS = \frac{average\ of\ lenght - period\ up\ closes}{average\ of\ lenght - period\ down\ closes'}$$

که در آن RS از تقسیم میانگین قیمت رو به بالا به میانگین قیمت رو به پایین به‌دست می‌آید.

درنهایت باید گفت L طول دوره‌ای است که برای تحلیل تنظیم می‌شود. در این پژوهش، این پارامتر ۵ در نظر گرفته شده است، زیرا هدف، پیش‌بینی در آینده‌ی نزدیک برای تحلیل است.

اورنشتاین-اولن‌بک

اورنشتاین-اولن‌بک یک فرآیند تصادفی است و در مدل‌سازی سهم‌هایی که همیشه به سمت میانگین بلندمدت خود در حرکت هستند، استفاده می‌شود. این فرایند با استفاده از مدل ریاضی زیر بیان می‌شود:

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t,$$

که در آن پارامترهای θ، μ و σ با استفاده از رگرسیون، تخمین زده خواهند شد. فرایند اورنشتاین-اولن‌بک درواقع نوعی از فرایند مارکوف در حالت فضا-حالت و زمان پیوسته است که ویژگی‌هایی از حرکت براونی را در بردارد (بذرائی و همکاران، ۱۴۰۰).

مدل‌سازی به روش اورنشتاین-اولن‌بک یک روش تصادفی است، و مسائل مدل‌سازی شده در این فرآیند به سمت میانگین بلندمدت خود بازمی‌گردند (مورائس‌سارمنتو، ۲۰۲۱). مدآورام (۲۰۱۳) در پژوهش خود با استفاده از مدل بازگشت به میانگین یا اورنشتاین-اولن‌بک و به‌کارگیری فرایند ماشین بردار پشتیبان به مبحث آربیتراژ شاخص برای پیش‌بینی آن پرداخته است. این پژوهش نشان می‌دهد که فرآیند ماشین بردار پشتیبان عملکرد بهتر اما نه در خور توجه از مدل بازگشت به میانگین دارد.

خوشه‌بندی

خوشه‌بندی[۱۹] یکی از روش‌های یادگیری ماشین بدون‌نظارت است و برای گروه‌بندی و افراز داده‌ها به کار می‌رود، در اینجا هدف از خوشه‌بندی سهام این است که سهم‌های درون یک خوشه بیشترین

شباهت را به یکدیگر داشته باشند و همچنین میزان شباهت بین خوشه‌ها کمترین مقدار باشد (طباطبایی و همکاران، ۱۴۰۱). انواع خوشه‌بندی عبارت‌اند از: تفکیکی، سلسله مراتبی، بر مبنای چگالی، بر مبنای مدل. در این پژوهش از خوشه‌بندی سلسله مراتبی استفاده می‌شود.

در روش تفکیکی، بر اساس n مشاهده و k خوشه، خوشه‌بندی انجام می‌شود. به‌این ترتیب، تعداد خوشه‌ها یا گروه‌ها بایستی از قبل مشخص شود. با طی مراحل خوشه‌بندی، هر سهم فقط و فقط به یک خوشه تعلق خواهد داشت و هیچ خوشه‌ای بدون عضو باقی نمی‌ماند. برخلاف خوشه‌بندی تفکیکی که سهم‌ها را در گروه‌های مجزا افراز می‌کند، خوشه‌بندی سلسله مراتبی[۲۰]، در هر سطح از فاصله، نتیجه خوشه‌بندی را نشان می‌دهد. این سطوح به‌صورت سلسله مراتبی هستند.

روش‌های خوشه‌بندی تفکیکی قادر به تشخیص خوشه‌هایی دایره شکل هستند. به این معنی که برای تشخیص خوشه‌ها از مجموعه داده‌هایی به شکل‌های کوژ[۲۱] یا محدب خوب عمل می‌کنند. در عوض برای تشخیص خوشه‌ها برای مجموعه داده‌های کاو[۲۲] یا مقعر دچار خطا می‌شوند. در روش خوشه‌بندی برمبنای چگالی، مکان‌های با تراکم زیاد شناسایی و داده‌های هر یک از آن‌ها در یک خوشه قرار می‌گیرند.

در روش خوشه‌بندی برمبنای مدل، یک توزیع آماری برای داده‌ها فرض می‌شود. هدف در اجرای خوشه‌بندی برمبنای مدل، برآورد پارامترهای توزیع آماری به همراه متغیر پنهانی است که به‌عنوان برچسب خوشه‌ها در مدل معرفی می‌شود (هان و دیویس، ۲۰۰۶).

انحراف زمان پویا[۲۳] (DTW) روشی معروف است که برای پیدا کردن یک هم‌ترازی بهینه بین دو دنباله داده‌ی وابسته به زمان استفاده می‌شود. در این روش، دنباله‌ها به‌صورت غیرخطی منبسط یا منقبض می‌شوند تا با یکدیگر مطابقت داشته باشند. در زمینه‌هایی مانند داده‌کاوی و بازیابی اطلاعات، DTW با موفقیت برای مقابله خودکار با تغییر شکل‌های زمانی و سرعت‌های مختلف مرتبط با داده‌های وابسته به زمان استفاده‌شده است (مولر، ۲۰۰۷).

ماشین بردار پشتیبان[۲۴] یکی از روش‌های معروف یادگیری ماشین است که برای هر دو مسئله دسته‌بندی و رگرسیون کاربرد دارد. ایده‌ی این روش ایجاد یک ابر صفحه یا مرز جداکننده بین داده‌ها است به‌نحوی‌که بیشترین تمایز ممکن بین داده‌های کلاس‌های مختلف به‌وجود آید. در این روش، هدف حل مسئله بهینه‌سازی زیر است (لای و همکاران، ۲۰۲۲):

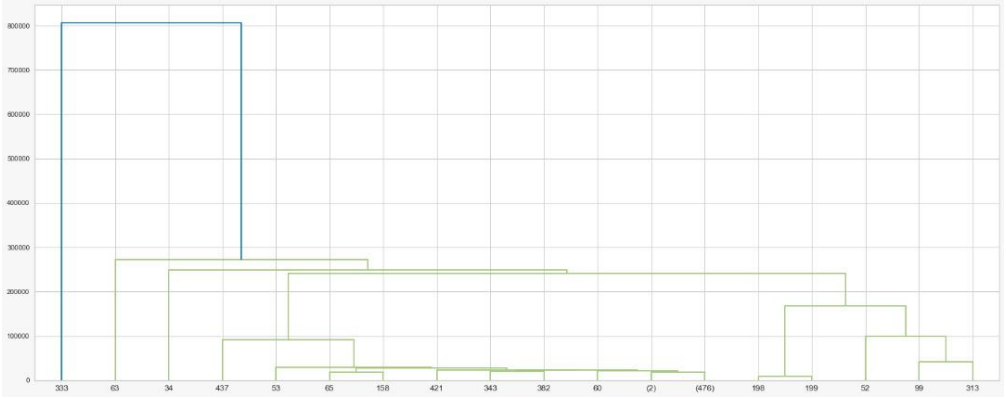
$$min_{\gamma,w,b}\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m}\xi_i,$$

$$s.t. \quad y^{(i)}\left(w^T x^{(i)} + b\right) \geq 1 - \xi_i, \qquad i = 1, \dots, m,$$
$$\xi_i \geq 0, \qquad i = 1, \dots, m.$$

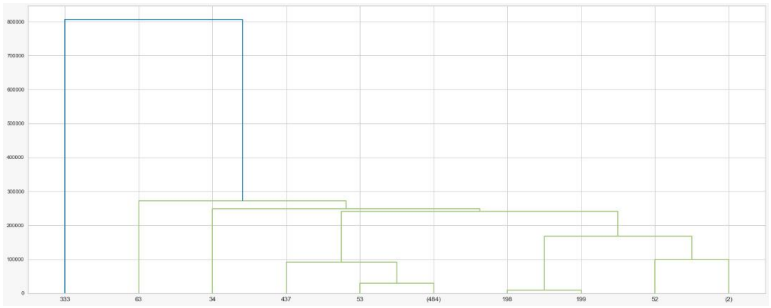
روش‌شناسی و شناخت داده‌ها

این پژوهش در زمره پژوهش‌های کاربردی قرار دارد و هدف این پژوهش ساخت مدلی است که بتواند به‌خوبی موقعیت‌ها و سیگنال‌های معاملاتی را تشخیص دهد که این امر سبب دست‌یابی به سود بیشتر با استفاده از راهبرد معامله‌ی زوج‌ها خواهد شد. داده‌های به‌کار رفته در این پژوهش شامل قیمت نهایی (بسته شدن) ۵۰۰ سهام برتر بازار بورس نیویورک در بازه ۲۰۱۹/۱/۱ تا ۲۰۲۲/۵/۱۶ می‌باشد. مدل پیشنهادی در این پژوهش به پنج گام زیر تقسیم می‌شود.

گام اول: شناسایی دو زوج سهام مناسب است که رابطه آماری بلندمدتی دارند. پژوهش‌های متفاوتی در این زمینه صورت گرفته است که در آن‌ها از رویکردهای فاصله، هم انباشتگی و ... استفاده‌شده است (هاک، ۲۰۱۵؛ مولر، ۲۰۰۷؛ مورائس‌سارمنتو، ۲۰۲۱). در این پژوهش از رویکرد خوشه‌بندی که یک رویکرد نوین به شمار می‌رود بهره گرفته می‌شود. ابتدا به‌کمک روش خوشه‌بندی سلسله‌مراتبی، قیمت سهام‌ها خوشه‌بندی می‌شوند. معیارهای فاصله‌ی بی‌شماری برای خوشه‌بندی وجود دارد که از معیار فاصله انحراف زمان پویا که یک معیار مبتنی بر روش‌های بهینه‌سازی است، در این پژوهش استفاده‌شده است. در این گام بر اساس شکل‌های ۱ و ۲ که نمودار دندروگرام را نمایش می‌دهند، زوج سهامی که بیشترین شباهت را دارند، انتخاب می‌شوند.



شکل ۱: یکی از خوشه‌های مربوط به بازار بورس نیویورک با سطح فاصله کمتر. (منبع: یافته‌های پژوهشگر)



شکل ۲. خوشه‌ی با سطح فاصله کمتر در بازار بورس نیویورک. (منبع: یافته‌های پژوهشگر)

در شکل ۱ و ۲ خوشه‌های منتخب را مشاهده می‌کنید. پس از خوشه‌بندی بر اساس نمودارهای دندروگرام دو سهام گوگل و گوگ بیشترین شباهت و نزدیک‌ترین فاصله را داشتند.

گام دوم: در این گام به تولید ویژگی دو سهم منتخب یعنی گوگل و گوگ پرداخته می‌شود. در این پژوهش برای تولید ویژگی از اندیکاتورهای میانگین متحرک ساده، میانگین وزنی متحرک، شاخص جریان نقدینگی و شاخص قدرت نسبی استفاده می‌شود. سپس درصد تغییرات برای هریک از ویژگی‌های جدید که از گام قبل به‌دست‌آمده‌اند، محاسبه می‌شوند.

گام سوم: رگرسیون خطی روی ویژگی‌های تولیدشده‌ی زوج سهم منتخب پیاده‌سازی و با کمک رگرسیون پارامترهای مدل اسپرد معامله‌ی زوج محاسبه می‌شود که فرم این معادله به شکل زیر می‌باشد:

$$\frac{dA_t}{A_t} = \alpha dt + \beta \frac{dB_t}{B_t} + dX_t,$$

که در آن $\frac{dA_t}{A_t}$ و $\frac{dB_t}{B_t}$ به ترتیب بازده زوج سهم‌های انتخاب‌شده را نمایش می‌دهند. X_t بازده اسپرد در زمان t ، α جمله رانش، β نشان‌دهنده نسبت پوششی است که برای عادی‌سازی نوسان بین سهم‌ها عمل می‌کند. همچنین عبارت باقیمانده dX_t محاسبه می‌شود. با مجموع عبارات باقیمانده‌ها فرمول زیر ساخته می‌شود:

$$X_t = \sum_{i=0}^{t} dX_t.$$

گام چهارم: خود رگرسیون مرتبه یک یا وقفه‌ی یک روی شرایط انتشار برای رسیدن به پارامترها با توجه به مدل اورنشتاین-اولن‌بک اجرا می‌شود. سپس برای هر ویژگی به‌طور جداگانه بر اساس فرمول زیر امتیازی محاسبه می‌شود:

$$T_{feature} = \frac{X_t^{feature} - \mu_{feature}}{\sigma_{feature}}.$$

گام پنجم: در گام آخر برچسب‌ها ساخته و به مدل ماشین بردار پشتیبان ارائه می‌شود. برای این

کار ابتدا یک آستانه در نظر گرفته می‌شود. مقدار این آستانه متناسب است با میزان سودی که قرار است کسب شود، یعنی اگر مقدار آستانه بزرگ در نظر گرفته شود مقدار سوددهی بیشتر خواهد شد، اما در مقابل فرصت‌های سوددهی کمتر می‌شود و برعکس. در زمان اولیه و گام شروع میزان سود برابر صفر می‌باشد، اگر تفاضل مقدار باقیمانده‌ی چند گام آینده از زمان حال بزرگ‌تر از مقدار آستانه باشد نشان‌دهنده یک موقعیت معاملاتی است و اسپرد به‌وجود می‌آید و می‌توان کسب سود کرد. این موقعیت با برچسب ۱ (موقعیت سودآور) برچسب‌گذاری می‌شود. در غیر این صورت با برچسب ۰ (موقعیت بدون سود) برچسب‌گذاری می‌گردد. سپس ویژگی‌ها و ویژگی برچسب (هدف) به مدل ماشین بردار پشتیبان که به‌اختصار طبقه‌بند یا دسته‌بند گفته می‌شود، ارائه می‌شود.

برای آموزش ماشین بردار پشتیبان بدین شکل عمل می‌شود. ابتدا به روش اعتبارسنجی، پارامترهای مدل تنظیم و بهترین آن‌ها انتخاب می‌شود. برای محاسبه‌ی پارامترهای بهینه‌ی الگوریتم، یکی از زیرمجموعه‌های داده‌های بخش اعتبارسنجی انتخاب می‌شود. ابتدا یک شبکه از مقادیر ابر پارامترها انتخاب و به‌ازای هر نقطه از این شبکه، ابر پارامترهای روش، مقداردهی می‌شوند. سپس از روش اعتبارسنجی متقابل برای ارزیابی عملکرد ابر پارامتر در نقطه‌ی مذکور بهره برده خواهد شد. در روش اعتبارسنجی متقابل، داده‌های آموزشی به k قسمت مساوی تقسیم می‌شوند. از زیرنمونه‌ها، یکی به‌عنوان داده‌های اعتبارسنجی برای آزمایش مدل حفظ می‌شود و بقیه‌ی زیرنمونه‌ها، مدل را می‌سازند. بعد از ساخت مدل، زیرنمونه‌ی آزمایشی برای میزان اعتبار مدل ساخته‌شده به‌کار گرفته می‌شود. به‌این‌ترتیب به ازای هر نقطه‌ی شبکه، فرآیند اعتبارسنجی متقابل، k بار تکرار می‌شود و در هر بار معیار عملکرد آن نقطه سنجیده می‌شود. در انتها از میانگین معیارهای به‌دست‌آمده برای سنجش میزان اعتبار آن نقطه بهره برده می‌شود. در میان نقاط مختلف شبکه‌ی مذکور، نقطه‌ای که بهترین اعتبار را تولید کرده است، به‌عنوان ابر پارامترهای مدل استفاده می‌شود. بعد از انتخاب بهترین پارامترها، الگوریتم ماشین بردار پشتیبان روی مجموعه‌ی داده اعمال و مدل را می‌سازد.

برآورد مدل و نتایج

برای برآورد عملکرد و محاسبه‌ی کارایی مدل، معیارهای زیر استفاده می‌شوند. این معیارها بر اساس ماتریس درهم‌ریختگی محاسبه شده‌اند. تعداد نمونه‌های کلاس موقعیت‌های سودآور که به‌درستی سودآور تشخیص داده شده‌اند را با TP، تعداد نمونه‌های کلاس بدون سود که به‌درستی بدون سود تشخیص داده شده‌اند را با TN، تعداد نمونه‌های کلاس سودآور که به‌اشتباه بدون سود تشخیص داده شده‌اند را با FP و تعداد نمونه‌های کلاس بدون سود که به‌اشتباه سودآور تشخیص داده

شده‌اند را با FN نمایش دهید. در این‌صورت پنج معیار متداول زیر برای ارزیابی مدل ساخته شده توسط روش ماشین بردار پشتیبان استفاده شده است:

$$TNR = \frac{TN}{TN + FP}, \quad (۱)$$

$$TPR = \frac{TP}{TP + FN}, \quad (۲)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \quad (۳)$$

$$F_1 = \frac{2\,TP}{2\,TP + FP + FN}. \quad (۴)$$

منحنی مشخه عملکرد[۲۵] (ROC) یکی از معیارهای ارزیابی است که برای نمایش توانایی ارزیابی یک دسته‌بند دو کلاسه استفاده می‌شود و مقدار آن متغیر است. هر یک از معیارهای ارزیابی فوق، هرچه به یک نزدیک‌تر باشند مدل از عملکرد بهتری برخوردار است. از آنجا که هر یک از معیارهای ارزیابی از اهمیت خاصی برخوردار هستند، پس لازم است در ارزیابی عملکرد مدل از هر پنج معیار در کنار یکدیگر استفاده شود و مورد بررسی قرار گیرند.

جدول ۱: مقادیر معیارهای ارزیابی مدل حاصل از اعمال روش ماشین بردار پشتیبان.

(منبع: یافته‌های پژوهشگر)

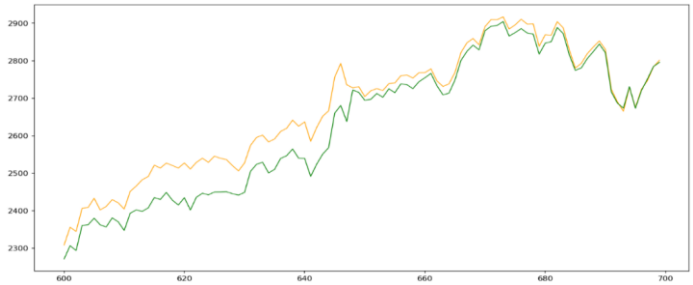
معیار	ROC	F_1	Acc	TPR	TNR
روش ماشین بردار پشتیبان	۷۳،۲۸	۰.۶۶۸	۰.۶۷۰۵	۰.۶۶۷۸	۰.۶۶۸۳

بر اساس جدول ۱، مقدار TNR که نرخ تشخیص داده‌های با برچسب بدون سود را نمایش می‌دهد دارای مقدار ۰.۶۶۸۳ است و قابل‌قبول است. از طرف دیگر TPR نرخ تشخیص داده‌های با برچسب سودآور است و مقدار این معیار نیز ۶۶.۷۸ درصد می‌باشد. معیار ارزیابی Acc دقت کلی مدل را برای هر دو کلاس نشان می‌دهد و مقدار این معیار ۶۷.۰۵ درصد می‌باشد. مقدار F_1 یا دقت مدل برای داده‌های نامتوازن را ۰.۶۶۸ نشان می‌دهد. مقدار ROC مدل ۷۳.۲۸ درصد است و کارایی مناسب مدل را نشان می‌دهد. با توجه به مقادیر به‌دست آمده و مقایسه آن‌ها در کنار یکدیگر، عملکرد مدل مناسب است و به خوبی توانسته موقعیت‌های معاملاتی را شناسایی کند.

با توجه به شکل‌های ۳ و ۴ بین نقاط ۶۰۰ و ۷۰۰ بیشترین اسپرد به وجود آمده و این فاصله موقعیت‌های معاملاتی را نمایش می‌دهد. زمانی که اسپرد از میانگین بلندمدت خود دور بشود، موقعیت های معاملاتی گشوده می‌شود و آنگاه که به میانگین خود نزدیک شود، موقعیت معاملاتی

بسته می‌شود. برای مثال نقاط ۶۰۵ و ۶۱۰ دو موقعیت معاملاتی است همچنین نقاط ۶۱۱ و ۶۲۰ نیز نقاط معاملاتی هستند. در نقطه ۶۲۰ معامله باز می‌شود و در نقطه ۶۳۰ معامله بسته می‌شود. این دو نقطه به عنوان برچسب سودآور شناخته شده و با عدد ۱، TP و بقیه نقاط که اسپرد مناسبی تشکیل نشده مانند ۶۸۰ و ۷۰۰ برچسب بدون سود شناخته شده و با عدد ۰، TN نمایش داده می‌شوند.

شکل ۳: سری‌های زمانی دو سهم گوگل و گوگ (منبع: یافته‌های پژوهشگر)



شکل ۴: اسپرد بین دو سهم گوگل و گوگ، موقعیت‌های معاملاتی. (منبع: یافته‌های پژوهشگر)

نتیجه‌گیری و ارائه پیشنهادات

در این پژوهش، یک مدل نوین برای شناسایی زوج سهم‌ها و موقعیت‌های مناسب خرید و فروش در راهبرد معاملاتی زوج‌ها ارائه شد. مدل پیشنهادی با استفاده از ترکیب روش‌های خوشه‌بندی سلسله‌مراتبی و ماشین بردار پشتیبان طراحی شد و کارایی آن با داده‌های مربوط به شاخص $S\&P\ 500$ مورد ارزیابی قرار گرفت. نتایج حاصل از تحلیل‌ها نشان داد که رویکرد خوشه‌بندی سلسله‌مراتبی همراه با استفاده از معیار فاصله انحراف زمان پویا، قادر است زوج سهم‌هایی با شباهت حرکتی بالا را شناسایی کند. سپس مدل ماشین بردار پشتیبان بر روی زوج سهم با بیشترین شباهت حرکتی اعمال شد که توانست به‌طور مؤثر موقعیت‌های معاملاتی سودآور را شناسایی و پیش‌بینی کند. عملکرد مدل از طریق معیارهای ارزیابی نظیر نرخ تشخیص درست موقعیت‌های سودآور، نرخ تشخیص

موقعیت‌های بدون سود، دقت کلی و شاخص $F1$ مورد بررسی قرار گرفت. مدل توانست با دقت ۶۷.۰۵ درصد و شاخص ROC برابر با ۷۳.۲۸ درصد، عملکرد قابل قبولی ارائه دهد. این نتایج نشان می‌دهد که مدل پیشنهادی می‌تواند به‌عنوان ابزاری مفید برای سرمایه‌گذاران و فعالان بازارهای مالی مورد استفاده قرار گیرد، به‌ویژه در شرایطی که تصمیم‌گیری سریع و مبتنی بر داده ضروری است.

پیشنهادات برای تحقیقات آتی عبارتند از:

۱ .استفاده از سایر مدل‌های یادگیری ماشین: در این پژوهش از ماشین بردار پشتیبان به دلیل ویژگی‌های تفسیری و کارایی آن در داده‌های با ابعاد کوچک استفاده شد. با این حال، پیشنهاد می‌شود در پژوهش‌های آینده، از مدل‌های یادگیری عمیق نظیر شبکه‌های عصبی بازگشتی(RNN) ، شبکه‌های پیچشی (CNN) و مدل‌های ترکیبی استفاده و نتایج آنها با مدل ارایه شده در این پژوهش مقایسه شود.

۲ .بررسی تاثیر سایر شاخص‌ها بر عملکرد دیدگاه ارایه شده: در این پژوهش از تعدادی شاخص مالی مانند میانگین متحرک ساده و شاخص قدرت نسبی استفاده شد. پیشنهاد می‌شود در پژوهش‌های آتی، از ویژگی‌های بیشتری مانند تحلیل احساسات بازار یا داده‌های خبری استفاده شود.

۳ .بررسی بازارهای مالی مختلف: این پژوهش بر اساس داده‌های شاخص $S\&P\ 500$ انجام شد. برای تعمیم‌پذیری بیشتر، پیشنهاد می‌شود که مدل در بازارهای مالی دیگر مانند بورس‌های آسیایی یا بازار ارزهای دیجیتال نیز ارزیابی شود. این کار می‌تواند به شناسایی نقاط قوت و ضعف مدل در محیط‌های مختلف کمک کند.

۴ .استفاده از معیارهای بهینه‌سازی پیشرفته: در فرآیند شناسایی زوج سهم‌ها و موقعیت‌های معاملاتی، استفاده از معیارهای دیگری مانند الگوریتم‌های تکاملی یا بهینه‌سازی ازدحام ذرات (PSO) و مقایسه نتایج آن با پژوهش حاضر پیشنهاد می‌شود.

۵ .ایجاد سامانه معاملات خودکار: پیشنهاد می‌شود از نتایج این پژوهش برای توسعه سامانه‌های معاملات خودکار استفاده شود. این سامانه‌ها می‌توانند با ادغام مدل پیشنهادی و پلتفرم‌های معاملاتی، فرآیند تصمیم‌گیری را برای سرمایه‌گذاران تسهیل کنند و بهره‌وری معاملات را افزایش دهند.

با توجه به نتایج به‌دست‌آمده و پیشنهادات ارائه‌شده، این پژوهش گامی مؤثر در جهت به‌کارگیری یادگیری ماشین برای بهبود راهبردهای معامله‌ی زوج‌ها بوده و می‌تواند به‌عنوان مبنایی برای تحقیقات آتی مورد استفاده قرار گیرد.

منابع

۱) ابراهیمی‌پور، م.، و داودی، س. (۱۴۰۰). بررسی سود آوری راهبرد معامله زوجی بر پایه سیستم حالت-فضای خطی و فیلتر کالمن در بورس اوراق بهادار. دانش سرمایه‌گذاری, ۱۰(۳۷), ۵۷-۷۵.

۲) دستوری، م.، و مرادپور، م. (۱۴۰۰). بهینه‌سازی الگوریتم معاملات زوجی پربسامد با استفاده از تلفیق الگوریتم ژنتیک و کنترل فرایند آماری فازی. دانش سرمایه‌گذاری, ۱۰(۴۰), ۴۷۱-۴۸۴.

۳) دولو, ی.، یزدی، آ.، و اردوان، م. (۱۴۰۱). معاملات جفتی؛ مقایسه رویکردهای فاصله‌ای و کاپیولا. تحقیقات مالی, ۲۴(۱), ۱۰۴-۱۳۳.

۴) طباطبایی، م.، صالح‌پور جم, م.، امین، س.، مصفایی، س.، و جمال، آ. (۱۴۰۲). شبیه‌سازی رسوب معلق با استفاده از الگوریتم‌های یادگیری ماشینی و داده‌های بارش ماهواره CHIRPS با تاکید بر خوشه‌بندی داده‌ها و آزمون گاما، مطالعه موردی: حوزه آبخیز رامیان، استان گلستان. مهندسی و مدیریت آبخیز، ۱۵(۳)، ۳۲۸-۳۵۰.

۵) عطاران، م.، و صالحی، م. (۱۴۰۰). بررسی عملکرد روش‌های تکنیکال مبتنی بر بازگشت به میانگین، در معاملات مکرر. مطالعات مدیریت و کار آفرینی، ۳۵(۷)، ۴۲۱-۴۲۷.

۶) فلاح‌پور، آ.، و حکیمیان، آ. (۱۳۹۶). بررسی عملکرد سیستم معاملات زوجی در بورس اوراق بهادار تهران: رویکرد هم انباشتگی و بررسی نسبت سورتینو. مهندسی مالی و مدیریت اوراق بهادار، ۳۰(۸)، ۱-۱۷.

۷) نوراحمدی، آ.، و نوراحمدی، م. (۱۴۰۲). کاربرد فیلتر کالمن برای تخمین نسبت پوشش ریسک پویا در راهبرد معاملات زوجی (مطالعه موردی: صنعت خودرو). تحقیقات مالی، ۲۵(۱)، ۶۳-۸۷.

8) Adrian، R. M. (2011). The relative strength index revisited. African Journal of Business Management, 5(14), 5855-5862.

9) Albanese, C., Li, D., Lobachevskiy, E., & Meissner, G. (2013). A comparative analysis of correlation approaches in finance. The Journal of Derivatives, 21(2), 42-66.

10) Carrasco Blázquez, M., & Prado Román, C. (2018). Pairs trading techniques: An empirical contrast. European Research on Management and Business Economics (ERMBE), 24(3), 160-167.

11) Ellis, C. A., & Parbery, S. A. (2005). Is smarter better? A comparison of adaptive, and simple moving average trading strategies. Research in International Business and Finance, 19(3), 399-411.

12) Han, B., & Davis, L. S. (2006). Semi-parametric model-based clustering for DNA microarray data. In 18th International Conference on Pattern Recognition (ICPR'06) (Vol. 3, pp. 324-327). IEEE.

13) Huck, N., & Afawubo, K. (2015). Pairs trading and selection methods: is cointegration superior? Applied Economics, 47(6), 599-613.

14) Lai, X., Deng, C., Tang, X., Gao, F., Han, X., & Zheng, Y. (2022). Soft clustering of retired lithium-ion batteries for the secondary utilization using Gaussian mixture model based on electrochemical impedance spectroscopy. Journal of Cleaner Production, 339, 130786.

15) Liew, R. Q., & Wu, Y. (2013). Pairs trading: A copula approach. Journal of Derivatives & Hedge Funds, 19, 12-30.

16) Lu, C. J., Lee, T. S., & Chiu, C. C. (2009). Financial time series forecasting using independent component analysis and support vector regression. Decision support systems, 47(2), 115-125.

17) Lu, Y. (2018). Construction of financial warning model for the listed companies-based on cash Flow index. Management & Engineering, 33, 12-22.

18) Madhavaram, G. R., Dodds, J. C., & Saint Mary's University (Halifax, N.S.). (2013). Statistical arbitrage using pairs trading with support vector machine learning. Saint Mary's University.

19) Moraes Sarmento, S., & Horta, N. (2020). Enhancing a pairs trading strategy with the application of machine learning. Expert Systems with Applications, 158, 113490.

20) Moraes Sarmento, S., Horta, N., (2021). A Machine Learning based Pairs Trading Investment Strategy. SpringerBriefs in Applied Sciences and Technology, Springer International Publishing.

21) Müller, M. (2007). Dynamic time warping. Information Retrieval for Music and Motion, 14, 69-84.

22) Schwartz, B., Gannot, S., & Habets, E. A. (2014). Online speech dereverberation using Kalman filter and EM algorithm. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23(2), 394-406.

23) Sriramulu, S., Jarvi, T. D., & Stuve, E. M. (1999). Reaction mechanism and dynamics of methanol electrooxidation on platinum (111). Journal of Electroanalytical Chemistry, 467(1-2), 132-142.

24) Taylor, N. (2004). Trading intensity, volatility, and arbitrage activity. Journal of Banking & Finance, 28(5), 1137-1162.

یادداشت‌ها:

1 Indicators
2 Ornstein Uhlenbeck
3 Simple moving average (SMA)
4 Weighted moving average (WMA)
5 Money flow index (MFI)
6 Relative strength index (RSI)
7 Pair integration approach
8 Time Series Regression
9 Comparative approach
۱۰اسپرد به معنای تفاوت بین قیمت خریدار و قیمت فروشنده است.
11 Bear Market
12 Bull Market
13Range market
۱۴اسیلاتور ابزاری است که به معامله‌گر امکان معامله و کسب سود را در دوره‌های تثبیت بازار می‌دهد.
15 SMA (simple moving average)
16 Weighted moving average (WMA)
17 Money flow index (MFI)
18 Relative strength index (RSI)
19 Clustering
20 Hierarchical Clustering
21 Convex
22 Concave
23 Dynamic time warping
24 Support vector machines
25 Receiver Operating Characteristic

Selection in Pairs Trading Strategy via a Clustering Method

Majid Ebtia[1]
Mohammad Reza Aryafar[2]
Sayyed Mohammad Hoseini[3]

Receipt: 14/12/2023 **Acceptance: 02/02/2025**

Abstract

In order to invest properly and identify the right short and long positions, the existence of an efficient strategy is necessary. The pairs trading system is one of the most famous algorithmic trading systems. In this system, a pair of stocks that have same trend in the long time (reversion to the mean) and have fluctuations (spread) in the short time is used. The most important point is to identify such a pair of stock. In this research, an approach is used to identify pairs of stocks and also to find the right short and long positions in trading pairs. In the first step, a stock pair that has a long time statistical relationship is selected. The DTW method is used to calculate the distance between each stock pair and their movement similarity. Then, the hierarchical clustering method is used to cluster stocks, and then pairs of stocks that have a greater degree of similarity are selected. In the second step, the SVM method is used to identify buying and selling positions. In order to check the performance of the method, the S&P 500 index, which includes the top 500 stocks of the New York Stock Exchange, has been used.

Keywords

Algorithmic trading, pairs trading, dynamic time deviation, clustering, stocks

1-Department of Mathematics, Faculty of Basic Sciences, Ayatollah Boroujerdi University, Boroujerd, Iran. majid.ebtia@gmail.com
2-Department of Resources and Energy Economics, Faculty of Administrative Sciences and Economics, Ferdowsi University of Mashhad, Mashhad, Iran. m.ariafar@mai.ac.ir
3-Department of Mathematics, Faculty of Basic Sciences, Ayatollah Boroujerdi University, Boroujerd, Iran. (Corresponding Author) sm.hoseini@abru.ac.ir