

بررسی سودآوری معاملات زوجی با استفاده از الگوریتم مبتنی بر یادگیری ماشین و الگوریتم

ژنتیک در بورس اوراق بهادار تهران

بهناز رضایی^۱، رضا برادران کاظم زاده^{۲*}، محمد علی رستگار^۳

۱- دانشجوی کارشناسی ارشد، دانشکده مهندسی صنایع و سیستم ها، دانشگاه تربیت مدرس، ایمیل: behnaz_rezaei@modares.ac.ir

۲- استاد، دانشکده مهندسی صنایع و سیستم ها، دانشگاه تربیت مدرس، ایمیل: rkazem@modares.ac.ir

۳- استادیار، دانشکده مهندسی صنایع و سیستم ها، دانشگاه تربیت مدرس، ایمیل: ma_rastegar@modares.ac.ir

چکیده

در این مقاله، به مسئله‌ی چگونگی یافتن زوج‌های سودآور با اعمال محدودیت خودکار در فضای جستجوی زوج‌داری‌ها با استفاده از تکنیک‌های یادگیری ماشین و ادغام یک الگوریتم یادگیری بدون نظارت، *OPTICS*، به فرآیند شناسایی و انتخاب زوج‌ها در معاملات زوجی پرداخته شده است. همچنین جهت بهینه‌سازی سبد متشکل از زوج‌داری‌ها و تخصیص سرمایه بهینه به آن‌ها، از الگوریتم مبتنی بر ژنتیک با هدف افزایش نسبت شارپ استفاده شده است. عملکرد تکنیک پیشنهادی برای خوشه‌بندی خودکار، نسبت به روش‌های متداول جستجوی زوج‌داری‌ها توسط سرمایه‌گذاران بهتر بوده و منجر به دستیابی به میانگین نرخ بازگشت سرمایه و نسبت شارپ بالاتری برای سبد در معاملات با استفاده از زوج‌های منتخب از خوشه‌ها شده است. این معیارهای ارزیابی محاسبه شده برای سبد، بعد از به‌کارگیری الگوریتم بهینه‌سازی ژنتیک دوهدفه ارتقا یافتند. این مطالعه با استفاده از داده‌های قیمتی درون‌روزی گروهی از سهام‌های بورس اوراق بهادار تهران بین سال‌های ۲۰۱۵ تا ۲۰۲۰ و در نظر گرفتن هزینه‌های معاملاتی شبیه‌سازی شد.

کلمات کلیدی: معاملات زوجی، یادگیری ماشین، یادگیری بدون نظارت، الگوریتم بهینه‌سازی ژنتیک، هزینه‌های معاملاتی

* نویسنده مسئول: تهران، جلال آل احمد، پل نصر، دانشگاه تربیت مدرس، کد پستی: ۱۴۱۱۵۱۱۱

فرصت‌هایی، فاصله‌ی سری قیمت‌های زوج‌های معاملاتی (اسپرد) به طور مداوم نظارت می‌شود و هنگامی که یک انحراف یا بی‌نظمی آماری در رفتار سری اسپرد شناسایی شود، موضع معاملاتی اخذ می‌شود و این انحراف یک فرصت خرید روی دارایی با روند کاهشی و فرصت فروش روی دارایی با روند افزایشی ایجاد می‌کند و این موقعیت‌ها تا زمان بازگشت اسپرد به تعادل خود حفظ می‌شوند و در نهایت بعد از اصلاح روند نهایی در سری اسپرد، موضع معاملاتی بسته می‌شود. قابل ذکر است این استراتژی صرف نظر از ارزش مطلق دو دارایی به ارزش نسبی آنها متکی است. یکی از چالش‌ها در این موضوع، غلبه بر فرآیند دشوار شناسایی دارایی‌ها از نقطه نظر ارزشگذاری است که یک گام اساسی در تصمیم‌گیری برای فروش دارایی‌هایی با ارزشگذاری بیش از ارزش ذاتی و خرید دارایی‌هایی با ارزشگذاری کمتر از ارزش ذاتی است که با تمرکز بر ایده قیمت‌گذاری نسبی، این مسئله حل می‌شود. علاوه بر این، این استراتژی یک استراتژی بهینه صرف نظر از جهت بازار (افزایشی یا کاهشی) است. هدف اصلی شناخت دو سهم یا روند قیمتی مشابه برای بهره‌گیری از موقعیت‌های فروش و خرید همزمان می‌باشد و با استفاده از این فرصت و استراتژی معاملاتی می‌توان به سودآوری قابل توجهی دست یافت. معمولاً سرمایه‌گذاران روی یک زوج معامله نمی‌کنند و سبدهای از زوج‌ها را جهت معامله در نظر می‌گیرند و از این رو به جستجوی دارایی‌ها و بررسی ترکیبات آن‌ها توسط آزمون‌ها می‌پردازند. اعمال محدودیت‌های متداول برای فضای جستجوی زوج دارایی‌ها توسط سرمایه‌گذاران همانند بررسی ترکیبات ممکن از دارایی‌های موجود در یک صنعت، می‌تواند منجر به شناخت زوج‌های متداول در معاملات شده و بازه سود محدودی را باقی بگذارد و از طرفی عدم اعمال محدودیت برای فضای جستجو و بررسی همه ترکیبات ممکن از دارایی‌های مختلف جهت شناسایی زوج دارایی‌های مالی مناسب، یافتن زوج‌های بهینه و با پتانسیل سودآوری را با توجه به افزایش دسترسی به داده‌ها، دشوارتر می‌کند و نیازمند تکنیک‌هایی با دقت عملکرد بسیار بالا (بدون خطا) است. بنابراین اعمال تقسیم‌بندی‌های غیرمتداول و هوشمند جهت ایجاد فضای جستجوی منسجم و مناسب برای بررسی ترکیبات ممکن و شناسایی زوج دارایی‌های مالی، می‌تواند علاوه بر کاهش زیان‌های احتمالی این استراتژی، منجر به سودآوری بالاتری شود. به‌کارگیری تکنیک‌های یادگیری ماشین در سال‌های اخیر شتاب بالایی گرفته و به طور گسترده در زمینه‌های مختلف از جمله مالی استفاده شده است. اکثر کاربردها مبتنی بر یادگیری با نظارت بوده و در مقابل، یادگیری بدون نظارت به ندرت در ادبیات مالی مورد استفاده

معاملات الگوریتمی استفاده از قدرت کامپیوتر جهت اتخاذ موقعیت‌های معاملاتی است که نحوه‌ی اجرای آن با استفاده از الگوریتم‌ها کنترل می‌شود. به عبارت دیگر معاملات الگوریتمی خرید و فروش اوراق بهادار بر اساس تصمیمات خرید یا فروش اخذ شده توسط الگوریتم‌های کامپیوتری می‌باشند. هدف در این سیستم‌ها شناسایی ناهنجاری‌های زودگذر در قیمت‌های بازار، رسیدن به سود با استفاده از الگوهای آماری در بازارهای مالی، پیاده‌سازی بهینه‌ی سفارش‌های بازار، استتار تمایلات معامله‌گران و یا کشف و بهره‌مندی از استراتژی‌های رقبا است. یکی از انواع معاملات الگوریتمی شناخته شده در دهه ۱۹۸۰، معاملات زوجی است که یکی از رایج‌ترین استراتژی‌های آربیتراژ آماری و بازار خنثی است. این استراتژی به عنوان یکی از مهم‌ترین ابزارهای سرمایه‌گذاری توسط صندوق‌ها و موسسات سرمایه‌گذاری و معامله‌گران استفاده شده است [۱] و موضوع اساسی در این تحقیق است. معاملات زوجی یک استراتژی معاملاتی بر مبنای استفاده از الگوی اسپرد و سودآوری از طریق ناکارایی بازار بر یک زوج سهام است. اسپرد معمولاً برابر با تفاضل یا تفاضل لگاریتم زوج دارایی‌ها در نظر گرفته می‌شود و باید قابل پیش‌بینی برای یک دوره‌ی یک ساله باشد. این استراتژی شامل دو مرحله است. نخست، نیاز به شناسایی دو دارایی است، به عنوان مثال دو سهام که سری قیمت‌های مربوط به آنها در بلندمدت یک رفتار مشابه را نشان می‌دهد و این موضوع نشان‌دهنده‌ی این است که هر دو دارایی در معرض فاکتورهای ریسک یکسان و مرتبط هستند و تمایل دارند به شیوه‌ای مشابه واکنش نشان دهند. دو دارایی که رابطه تعادل بین سری قیمت‌های آنها تایید می‌شود، تشکیل یک زوج معاملاتی را می‌دهند. بعد از شناسایی زوج‌های معاملاتی، سرمایه‌گذار وارد گام دوم استراتژی و اتخاذ مواضع معاملاتی بر اساس مدل معاملاتی مورد استفاده می‌شود. فرض اصلی این است که اگر دو سری قیمت از دارایی‌ها در گذشته و در بلندمدت رفتار مشابهی داشته‌اند، در آینده نیز انتظار پایداری این رفتار می‌رود. بنابراین فرض می‌شود که اسپرد از یک الگوی بازگشت به میانگین پیروی می‌کند. به عبارت دیگر در این الگو، روند با هر انحراف از تعادل طولانی‌مدت تصحیح شده و به تعادل خود بازمی‌گردد و یا اینکه روند موجود دچار تغییر شده و یک تعادل جدید برای اسپرد ایجاد می‌شود. به عبارت دیگر اگر بی‌نظمی و خروج از قاعده در رفتار سری قیمت رخ دهد، انتظار اصلاح روند قیمتی یک فرصت معاملاتی جهت کسب سود را فراهم می‌کند. برای یافتن چنین

قرار گرفته است. الگوریتم بدون نظارت در این تحقیق جهت خوشه‌بندی خودکار و ایجاد فضاهای جستجوی محدود (خوشه‌ها) بر اساس ویژگی‌های دارایی‌ها مورد استفاده قرار می‌گیرد و دارایی‌های مرتبط (نه لزوماً از صنعت یکسان) در خوشه یکسان قرار داده می‌شوند و ضمن داده محور بودن الگوریتم، سوگیری سرمایه‌گذاران نیز تأثیری در نوع خوشه‌بندی ندارد و نیازی به تعیین پارامتر توسط آن‌ها نیست. چارچوب در نظر گرفته شده مبتنی بر تجزیه و تحلیل مؤلفه اصلی (PCA) جهت کاهش ابعاد داده‌ها و استخراج فاکتورهای ریسک مشترک از دارایی‌ها و به دنبال آن اعمال الگوریتم بدون نظارت خوشه‌بندی $OPTICS^2$ روی داده‌های مربوط به دارایی‌ها به منظور خوشه‌بندی خودکار و ایجاد بستری برای استخراج ترکیب‌های زوجی (دوتایی) سودآور در مرحله انجام معاملات سبد زوج‌های منتخب از خوشه‌ها نیز، با توجه به رویکرد تخصیص سرمایه و وزن یکسان در نظر گرفته شده برای زوج‌های سبد، الگوریتمی مبتنی بر ژنتیک جهت بهینه‌سازی سبد مورد نظر پیشنهاد می‌گردد. در ادامه پژوهش، در بخش دوم به ادبیات و پیشینه‌ی تحقیق و در بخش سوم به تعریف مساله و مدل‌سازی آن متشکل از چهار بخش ۱- کاهش ابعاد، ۲- یادگیری بدون نظارت، ۳- معیارهای انتخاب زوج و ۴- شبیه‌سازی پرداخته می‌شود. در بخش یادگیری بدون نظارت به الزامات الگوریتم خوشه‌بند خودکار و مناسب از جمله ۱- عدم نیاز به پیش تعیین تعداد خوشه‌ها، ۲- رفتار بهینه نسبت به نقاط پرت و عدم نیاز به خوشه‌بندی همه‌ی داده‌ها و ۳- مناسب بودن برای خوشه‌های با چگالی متغیر اشاره شده و الگوریتم بدون نظارت مبتنی بر چگالی $OPTICS$ انتخاب می‌شود. در بخش چهارم نتایج در سه بخش ۱- عملکرد انتخاب زوج دارایی (اعلام تعداد خوشه‌های بدست آمده، ترکیبات دوتایی ممکن از خوشه‌ها و زوج‌های منتخب (هم‌انباشته) از بین ترکیبات بدست آمده و مقایسه نتایج حاصل از خوشه‌بند $OPTICS$ با نتایج روش‌های موجود در ادبیات پیشین از جمله ۱- خوشه‌بندی بر اساس صنایع و بررسی دارایی‌ها با صنعت یکسان و ۲- عدم خوشه‌بندی و بررسی همه ترکیبات دوتایی ممکن از دارایی‌های مختلف)، ۲- عملکرد معاملاتی (اجرای مدل معاملاتی مبتنی بر آستانه روی زوج‌های منتخب در روش‌های مختلف خوشه‌بندی (خوشه‌بندی با $OPTICS$ ، خوشه‌بندی بر اساس صنعت، تک خوشه‌ای) با در نظر گرفتن وزن یکسان تخصیص سرمایه به زوج‌ها) و ۳- عملکرد سبد بهینه (اجرای مدل معاملاتی با

تخصیص بهینه سرمایه بین زوج‌های منتخب از روش خوشه‌بندی برتر) بدست می‌آید. روش خوشه‌بندی برتر بر اساس بخش عملکرد معاملاتی و مقایسه سه روش خوشه‌بندی مشخص می‌شود و قابل ذکر است تعداد خوشه‌ها در روش خوشه‌بندی بر اساس صنعت، برابر با تعداد صنایع استفاده‌شده در مقاله است. بخش پنجم شامل نتیجه‌گیری و ارائه پیشنهادات آتی برای تحقیق است.

۲- ادبیات و پیشینه پژوهش

به طور کلی معاملات زوجی در دو حیطة ی «شناسایی و تشکیل زوجها» و «مدلسازی استراتژی معاملاتی زوج‌ها» مورد بررسی قرار می‌گیرد. مرحله‌ی انتخاب زوج‌ها برای معاملات زوجی شامل یافتن زوج‌های کاندید مناسب و انتخاب امیدوارکننده‌ترین آن‌ها از لحاظ سودآوری است. با شروع جستجو برای کاندیدهای مناسب، سرمایه‌گذار باید ابتدا دارایی‌های مورد نظر (مانند سهام، ETF و...) را انتخاب کرده و ترکیبات احتمالی را جستجو کند. معمولاً دو روش برای این مرحله پیشنهاد می‌شود: انجام یک جستجوی جامع برای همه ترکیب‌های دوتایی ممکن در میان دارایی‌های انتخاب شده [۲، ۳]، یا مرتب کردن آن‌ها در گروه‌های معنادار، معمولاً بر اساس صنعت و محدودسازی ترکیب‌ها به زوج‌هایی که از دارایی‌های موجود در یک صنعت تشکیل می‌شوند [۴، ۵]. در حالی که رویکرد نخست ممکن است زوج‌های غیرمعمول‌تری در معاملات پیدا کند و باعث سودآوری بالاتری شود اما رویکرد دوم با توجه به در نظر گرفتن فضای بررسی محدود و اعمال خطای پایین‌تر، احتمال یافتن روابط جعلی بین دارایی‌ها را کاهش می‌دهد و ایجاد تعادل بین سودآوری ناشی از افزایش دامنه‌ی زوج‌های مورد معامله به زوج‌های غیرمتداول‌تر و میزان خطای شناسایی آن‌ها، امکان توسعه معاملات زوجی با استفاده از روش‌های خودکار و داده‌محور جهت ایجاد فضای جستجوی بهینه و تشکیل سبد معاملاتی با زوج دارایی‌های سودآور ضمن کاهش ریسک سبد را فراهم می‌آورد. نویسنده در [۶] نیز از استراتژی مبتنی بر ETF هر صنعت و PCA برای ایجاد فضای جستجوی سهام‌های آمریکا استفاده می‌کند. سپس سرمایه‌گذار باید تعیین کند که از بین زوج‌های کاندید یافت شده، چگونه یک زوج واجد شرایط برای معامله تلقی می‌شود. رایج‌ترین رویکردها، رویکردهای فاصله، همبستگی و هم‌انباشتگی هستند. رویکرد فاصله، پیشنهادی توسط [۷]، زوج‌هایی را انتخاب می‌کند که مجموع مربعات فواصل بین سری تاریخی قیمت دو دارایی را به حداقل می‌رساند. این روش به طور

گسترده مورد استفاده قرار می‌گیرد، اما طبق نظر [۳] بهینه نیست. برای اثبات این عبارت، اینگونه فرض می‌شود که $P_{j,t}$ و $P_{i,t}$ قیمت های نرمال شده دو دارایی i و j هستند که یک زوج را تشکیل می‌دهند. علاوه بر این، S_{Pi-Pj}^2 واریانس اسپرد طبق (۱) است.

$$S_{Pi-Pj}^2 = \frac{1}{T} \sum_{t=1}^T (P_{i,t} - P_{j,t})^2 - \left(\frac{1}{T} \sum_{t=1}^T (P_{i,t} - P_{j,t}) \right)^2 \quad (1)$$

میانگین مجموع مربعات فواصل، $\overline{SSd} P_{i,Pj}$ در دوره تشکیل زوج‌ها نیز برابر است با:

$$\overline{SSd} P_{i,Pj} = \frac{1}{T} \sum_{t=1}^T (P_{i,t} - P_{j,t})^2 = S_{Pi-Pj}^2 + \left(\frac{1}{T} \sum_{t=1}^T (P_{i,t} - P_{j,t}) \right)^2 \quad (2)$$

یک زوج مطلوب مطابق معیار SSd می‌تواند زوجی باشد که معادله (۲) را به حداقل برساند. اما این بدان معناست که یک زوج با اسپرد صفر در طول دوره تشکیل، بهینه در نظر گرفته می‌شود. از نظر منطقی، این با ایده یک زوج بالقوه پیشنهادی مطابقت ندارد زیرا یک کاندید خوب باید از واریانس اسپرد بالا و ویژگی‌های بازگشت به میانگین برای ارائه فرصت‌های معاملاتی برخوردار باشد. از آنجا که این معیار این نیاز را در نظر نمی‌گیرد، احتمالاً زوج‌هایی با واریانس اسپرد کم و پتانسیل سود محدود ایجاد کرده و فرصت‌های معاملاتی سودآور را فراهم نمی‌کند. استفاده از همبستگی پیرسون به عنوان یک معیار انتخاب توسط [۸]، تحلیل شده است. نویسندگان کاربرد آن را در سری‌های بازگشتی با همان نمونه داده مورد استفاده توسط [۷] بررسی کرده و دریافته‌اند که همبستگی عملکرد بهتری را نسبت به روش حداقل فاصله نشان می‌دهد. با این وجود، این معیار بی‌خطا نیست زیرا ممکن است دو دارایی با بازده‌های همبسته، رابطه تعادلی نداشته باشند و بازگشت واگرایی در روند را نمی‌توان به صورت نظری توضیح داد. در نهایت، رویکرد هم‌انباشتگی مستلزم انتخاب زوج‌هایی است که دو جزء Y_t و X_t با یکدیگر هم‌انباشته هستند، در این صورت طبق تعریف، سری حاصل از ترکیب خطی زیر است.

$$S_t = Y_t - \beta X_t \quad (3)$$

β به عنوان فاکتور هم‌انباشتگی باید مانا باشد و انتظار می‌رود با تعریف سری اسپرد به این روش، رفتار بازگشت به میانگین داشته باشد و هر واگرایی در اسپرد با همگرایی دنبال شود. از این رو، این رویکرد از نظر اقتصادسنجی روابط تعادلی قابل اطمینان‌تری در زوج‌ها پیدا می‌کند. علاوه بر این، [۹]، یک مطالعه مقایسه‌ای بین رویکرد

هم‌انباشتگی و رویکرد فاصله انجام دادند و دریافته‌اند که رویکرد هم‌انباشتگی به طور قابل توجهی از رویکرد فاصله بهتر است. همچنین در تحقیق [۱۰] به این نکته اشاره شده است که عملکرد روش هم‌انباشتگی در شرایط آشفتگی بازار نسبت به سایر روش‌های شناسایی زوج، عملکرد بهتری داشته است. قابل ذکر است که استفاده از ضرایب استاتیک و ثابت به عنوان ضریب هم‌انباشتگی می‌تواند منجر به زیانی بالقوه و نامطلوب در معاملات زوجی شود بنابراین با استفاده از تکنیک‌هایی نظیر اعمال فیلتر کالمن (*Kalman Filter*)، می‌توان از ریسک سیستماتیک آن کاست [۱۱]. تشکیل ترکیبی از دارایی‌های با حرکت و رفتار یکسان یک گام حیاتی در معاملات زوجی است. جهت سودآوری در این زمینه لازم است روش‌هایی مورد بررسی قرار بگیرد که در شرایط ناکارایی بازار سودآورترین ترکیب از دارایی‌ها را انتخاب کند زیرا تعداد قابل توجهی از زوج‌های معاملاتی رفتار و روند ثابتی ندارند و امکان ایجاد روند متفاوت در رفتار اسپرد و عدم ایجاد تعادل مجدد در آن وجود دارد. همچنین ممکن است تاخیر و جلو افتادگی در رفتار هم‌روندی دارایی‌ها مشاهده شود، لذا سرمایه‌گذاران باید با دقت بیشتری به شناسایی و انتخاب زوج‌ها بپردازند [۱۲، ۱۳]. مرحله‌ی بعد از انتخاب زوج‌های مناسب، شامل اخذ مواضع معاملاتی با در نظر گرفتن یک مدل معاملاتی است. طبق نخستین چارچوب پیشنهاد شده توسط [۷]، معیار اتخاذ مواضع معاملاتی بر اساس واگرایی اسپرد و مدل مبتنی بر آستانه استاندارد است. در مدل مذکور، اگر فاصله بین دو سری قیمتی مربوط به یک زوج بیش از دو (به عنوان ضریب) انحراف استاندارد تاریخی آن فاصله باشد، معامله باز شده و مواضع معاملاتی اتخاذ می‌شود. این معامله در صورت مشاهده همگرایی اسپرد به میانگین خود در پایان دوره معاملاتی یا هنگام خروج دارایی از لیست بورس اوراق بهادار بسته می‌شود و یا به عبارت دیگر مواضع معاملاتی عکس جهت سودآوری اتخاذ می‌شود. در این مدل، میانگین و انحراف استاندارد مربوط به اسپرد (S_t) در طول دوره شکل‌گیری زوج‌ها به صورت (μ_S و σ_S) محاسبه شده و بر اساس آن‌ها، آستانه‌های مدل شامل آستانه خرید (α_L ، *long*)، آستانه فروش (α_S ، *short*)، و آستانه خروج (α_{Exit}) تعریف می‌شوند. سپس روند اسپرد در دوره‌ی معاملاتی کنترل می‌شود و در صورت عبور از α_L ، مواضع خرید روی اسپرد با خرید Y و فروش X و در صورت عبور از α_S ، مواضع فروش روی اسپرد با فروش Y و خرید X اتخاذ می‌شود و هنگام عبور از α_{Exit} ، اگر موضعی در حال حفظ باشد، خروج از مواضع انجام می‌گیرد. مدل معرفی شده به طور مکرر در این زمینه استفاده می‌شود اما با این وجود، نقاط ورودی تعریف شده

ممکن است بهینه نباشند زیرا هیچ اطلاعاتی در مورد جهت اسپرد در زمان بعدی در تصمیم معاملات گنجانده نشده است و تحقیقات مختلفی جهت پیشنهاد مدل‌های معاملاتی قوی‌تر ظاهر شده‌اند. تکنیک‌هایی از زمینه‌های مختلف، مانند تئوری کنترل تصادفی، مدل‌سازی فرآیند آماری، برنامه‌ریزی‌ها و مدل‌های ریاضی، الگوریتم‌های ابتکاری و فراابتکاری و یادگیری ماشین و ... مورد مطالعه قرار گرفته‌اند و نتایج خوبی در این زمینه‌ها گزارش شده است و به طور خاص، نتایج به‌دست‌آمده با رویکردهای یادگیری ماشین بسیار امیدوارکننده بوده است. عمدتاً تکنیک‌های یادگیری ماشین در زمینه‌ی معاملات زوجی برای پیش‌بینی و مدل‌سازی اسپرد و به طور کلی برای مدل‌سازی استراتژی معاملاتی استفاده شده است. نویسندگان مقالات [۱۴، ۱۵، ۱۶]، کاربرد شبکه‌های عصبی مصنوعی (ANN) را برای پیش‌بینی تغییرات اسپرد بررسی می‌کنند. [۱۷] یک سیستم آربیتراژ آماری تجربی مبتنی بر مدل‌های ناهمگونی خودبازگشتی مشروط‌تعمیم‌یافته شبکه عصبی (GARCH) برای مدل‌سازی مکانیسم تصحیح قیمت‌گذاری نادرست بین قیمت‌های نسبی دارایی‌های یک زوج، پیشنهاد می‌کند. رویکرد مورد استفاده در مقالات [۱۸، ۱۹] مبتنی بر پیش‌بینی برای بازه زمانی یک هفته آینده با استفاده از شبکه‌های عصبی بازگشتی (RNN) بوده که از طریق آن بازده پیش‌بینی شده محاسبه می‌شود. [۱۹] اثربخشی شبکه‌های عصبی عمیق، درختان با گرادیان تقویت‌شده و جنگل‌های تصادفی در زمینه آربیتراژ آماری با استفاده از سهام S&P ۵۰۰ تحلیل شده است. مقالات [۲۰، ۲۱] نیز از مدل یادگیری عمیق جهت پیش‌بینی تغییرات اسپرد و شناسایی فرصت‌های معاملاتی بین ETF و سهام‌های هم‌انباشته استفاده می‌کنند. اخیراً، [۲۲] پتانسیل به‌کارگیری یادگیری تقویتی عمیق در تنظیم معاملات زوجی بررسی شده و نتایج رضایت‌بخشی در مقایسه با روش‌های سنتی‌تر گزارش شده است. همچنین در تحقیق [۲۳] از مدل یادگیری تقویتی جهت تشخیص ریسک‌های مهم از جمله شکست‌های ساختاری و ریسک‌های بسته شدن بازار در معاملات زوجی استفاده شده است. با این وجود، تکنیک‌های یادگیری ماشین هنوز به طور گسترده در معاملات زوجی مورد استفاده قرار نگرفته و نتایج به‌دست‌آمده از به‌کارگیری مدل‌های یادگیری ماشین در «مدلسازی استراتژی معاملاتی» نشان‌دهنده جهتی امیدوارکننده برای تحقیقات آینده در زمینه ارتقای سودآوری معاملات زوجی است و با توجه به اهمیت موضوع شناسایی ترکیباتی از زوج‌های مناسب و سودآور در سبد سرمایه‌گذاری (سبد بازار خنثی) در معاملات زوجی، امکان بسط

کاربرد تکنیک‌های یادگیری ماشین در این موضوع نیز جهت افزایش میزان سودآوری این معاملات با استفاده از این تکنیک‌ها وجود دارد. همانطور که اشاره شد، با توجه به دو رویکرد کاهنده سود و اعمال شده توسط سرمایه‌گذاران جهت شناسایی زوج‌ها شامل بررسی همه ترکیبات دوتایی ممکن از «دارایی‌های یک صنعت» و «همه دارایی‌ها»، استفاده از تکنیک‌های داده‌محور یادگیری ماشین جهت شناسایی زوج‌های سودآور با رویکرد تقسیم‌بندی فضای جستجوی دارایی‌ها به طور خودکار و به دور از سوگیری سرمایه‌گذاران احساس می‌شود. الگوریتم‌های خوشه‌بندی بدون نظارت هستند، به این معنا که به مقادیر واقعی (برچسب‌ها) جهت هدف‌گیری دسترسی ندارند و به گروه‌بندی داده‌های ورودی می‌پردازند و برخلاف مدل‌های یادگیری نظارت شده مانند یادگیری عمیق، به هاپرپارامترهای زیادی نیاز ندارند و فضای کمتری برای صید داده (Data Snooping) باقی می‌گذارند. [۲۴] به مقایسه عملکرد سه الگوریتم نظارت نشده K-Means مبتنی بر افزایش‌بندی، DBSCAN مبتنی بر چگالی و Agglomerative مبتنی بر سلسه مراتب در خوشه‌بندی داده‌ها جهت معامله زوجی پرداخته و نسبت شارپ بالاتری در صورت استفاده از الگوریتم خوشه‌بندی تجمعی (سلسله‌مراتبی) اعلام شده است. سه الگوریتم ذکر شده نیاز به تعیین هاپرپارامترهایی از جانب سرمایه‌گذار همانند «تعداد خوشه‌ها» و ... دارند و با توجه به رویکرد تقسیم‌بندی خودکار مورد نیاز این تحقیق برای دارایی‌ها، الگوریتمی با ماهیت عدم نیاز به تعیین پارامتر جهت توسعه تحقیقات نیاز است. تشکیل یک سبد بازار خنثی عبارت است از ساخت سبدهای از دارایی‌ها که همبستگی بین بازده آن و بازده بازار در کمترین حد ممکن باشد و رویکردها بر ساختن سبدهایی متمرکز هستند که بازدهی ثابتی را بدون توجه به شرایط بازار دارند [۲۵، ۲۶] و بهینه‌سازی سبد موردنظر با استفاده از استراتژی‌های مختلف منجر به کسب سود همراه با کاهش ریسک است. استراتژی مورد استفاده در [۲۷]، ترکیب دو موضوع «تنوع بخشی» و «معاملات زوجی» بوده و به این منظور از روش یادگیری آماری برای یافتن زوج مناسب در سبد در هر لحظه معاملاتی مبتنی بر پیش‌بینی نوسانات با استفاده از روش میانگین متحرک نمایی یا GARCH استفاده شده است. سیستم هوشمند ارائه شده در [۱۲]، به بهینه‌سازی ترکیبات زوجی با در نظر گرفتن دو هدف متعارض بازده و ریسک به ترتیب با مفاهیم «واریانس اسپرد» و «رفتار بازگشت به میانگین» با استفاده از فرمول بندی برنامه‌ریزی عدد صحیح و الگوریتم چندهدفه‌ی ژنتیک پرداخته و مجموعه‌ای از فرصت‌های معاملات زوجی چند متغیره سودآور فراهم

شده است. الگوریتم ژنتیک مورد استفاده در [۲۸]، جهت بهینه‌سازی همزمان تخصیص سرمایه بین مجموعه کوچکی از زوج‌های کاندید و سیگنال‌های معاملاتی در یک سیستم معاملاتی با رفتار بازگشت به میانگین بوده و الگوریتم ژنتیک با طرح کدینگ باینری در کروموزوم بر روی داده‌های آموزشی به منظور بهینه‌سازی بازده دوره مورد نظر با متغیرهای تصمیم تخصیص معامله و پارامترهای سیگنال به کار گرفته شده است. با تکامل و توسعه الگوریتم ژنتیک برای حل مسائل بهینه‌سازی چند هدفه (مانند الگوریتم ژنتیک مرتب سازی غیر مغلوب^۳ NSGA2)، امکان تخصیص سرمایه بهینه بین زوج‌ها در سبد جهت دستیابی به حداکثر سود یا بازده همراه با کاهش ریسک فراهم می‌شود. در صورتی که سبد معاملاتی از زوج‌های سودآور تشکیل شده باشد، بهینه‌سازی آن می‌تواند منجر به رشد مضاعف میزان سودآوری سبد شود.

۳- تعریف مساله و مدل‌سازی

هدف مورد بررسی تحقیق چگونگی دستیابی به زوج دارایی‌های سودآور در سبد معاملاتی توسط سرمایه‌گذار بدون قرارگیری در معرض ناملایمات تکنیک‌های متداول جستجوی زوج‌های معاملاتی است. از جمله ناملایمات می‌توان به دو مورد اشاره کرد. نخست، هزینه محاسباتی آزمون احتمال بازگشت به میانگین برای همه ترکیبات ممکن که با در نظر گرفتن دارایی‌های بیشتر به شدت افزایش می‌یابد و دارایی‌ها دائماً برای تشکیل زوج‌های جدید نظارت می‌شوند. مورد دوم زمانی ایجاد می‌شود که آزمون فرضیه‌های متعدد در یک لحظه انجام می‌شود که از آن به عنوان مسئله مقایسه‌های چندگانه یاد می‌شود و احتمالاً باعث ارتکاب حداقل یک خطای نوع α در هنگام انجام مقایسات مستقل (با مقدار α به عنوان سطح اطمینان) یا نرخ خطای خانوادگی می‌شود که با افزایش دارایی‌ها که به طور تصادفی انتخاب شده‌اند، افزایش می‌یابد. از یک طرف، اگر سرمایه‌گذار جستجوی خود را به دارایی‌ها در یک بخش محدود کند، احتمال کمتری وجود دارد که زوج‌هایی را بیابد که هنوز در حجم زیادی معامله نشده‌اند و بنابراین حاشیه کمی برای سود باقی می‌گذارند اما از سوی دیگر، اگر سرمایه‌گذار هیچ محدودیتی بر فضای جستجو اعمال نکند، ممکن است به اجبار ترکیب‌های بسیاری را مورد بررسی قرار داده و احتمالاً روابط جعلی بین دارایی‌ها بیابد. این عدم تعادل، انگیزه جستجو برای روشی را فراهم می‌کند که به دنبال این موارد است: پیش تقسیم‌بندی مؤثر و خودکار اجتماع دارایی‌ها که علاوه

بر عدم محدودسازی نتایج بررسی ترکیب دارایی‌ها به زوج‌های نسبتاً رایج در معاملات زوجی، به اعمال فضای جستجوی گسترده نیز نپردازد. دستیابی به تعادل با به‌کارگیری یک الگوریتم یادگیری بدون نظارت جهت ایجاد خوشه‌های معنادار و بر اساس ویژگی‌های دارایی‌ها توسط آن صورت می‌پذیرد. به عبارت دیگر، گروه‌بندی دارایی‌ها توسط سرمایه‌گذار تعریف نمی‌شود و داده‌ها به صراحت خود را در خوشه‌ها نشان می‌دهند. روش پیشنهادی شامل مراحل کاهش ابعاد، اعمال یک الگوریتم خوشه‌بندی مناسب و انتخاب زوج دارایی‌ها بر اساس مجموعه‌ای از معیارها است.

۳-۱ کاهش ابعاد

اولین گام در جهت خوشه‌بندی شامل کاهش ابعاد و یافتن یک نمایش فشرده برای هر دارایی، از سری قیمت آن است که تجزیه و تحلیل مؤلفه اصلی (PCA) پیشنهاد شده است. PCA یک روش آماری است که از یک تبدیل متعامد برای تبدیل مجموعه‌ای از مشاهدات متغیرهای احتمالاً همبسته به مجموعه‌ای از متغیرهای خطی غیرهمبسته، اجزای اصلی، استفاده می‌کند. هر جزء را می‌توان به عنوان نماینده یک عامل ریسک در نظر گرفت و برای استخراج عوامل ریسک اساسی مشترک، استفاده از PCA را در سری‌های بازده نرمالایز شده پیشنهاد می‌شود که به صورت زیر تعریف شده است:

$$R_{i,t} = \frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}} \quad (۴)$$

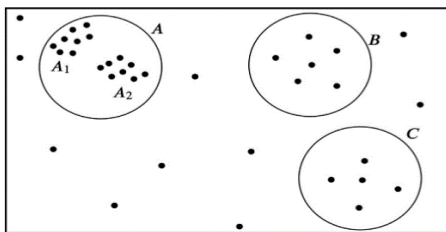
که در آن $P_{i,t}$ سری قیمت یک دارایی i است. استفاده از سری قیمت ممکن است به دلیل روندهای زمانی زمینه‌ای مناسب نباشد. تعداد مولفه‌های اصلی مورد استفاده، تعداد ویژگی‌ها را برای هر نمایش دارایی مشخص می‌کند. با توجه به اینکه یک الگوریتم یادگیری بدون نظارت در یادگیری ماشین روی این داده‌ها اعمال خواهد شد، تعداد ویژگی‌های زیاد و ابعاد گسترده داده‌ها مناسب نیست زیرا با وجود ویژگی‌های بیشتر، احتمال یافتن ویژگی‌های نامربوط را افزایش می‌یابد و همچنین در اثر مشکل نفرین ابعاد (اصطلاح توسط بلمن [۲۹] برای توصیف افزایش تصاعدی حجم همراه با افزودن ابعاد اضافی به فضای اقلیدسی به کار رفته است)، کارایی روش خوشه‌بندی پایین می‌آید. با توجه به [۳۰]، این اثر برای ابعاد بزرگتر از ۱۵ شروع به شدید شدن می‌کند و

با در نظر گرفتن این موضوع، کران بالا برای تعداد ابعاد PCA ، این مقدار در نظر گرفته شده و به صورت تجربی انتخاب می شود.

۲-۳ یادگیری بدون نظارت

پس از ایجاد یک نمایش فشرده برای هر دارایی، یک تکنیک خوشه‌بندی اعمال می‌شود. برای انتخاب الگوریتم مناسب جهت خوشه‌بندی خودکار، ابتدا برخی از الزامات خاص مسئله مانند عدم نیاز به پیش‌تعیین تعداد خوشه‌ها، عدم نیاز به خوشه‌بندی و گروه‌بندی تمام دارایی‌ها، تخصیص دقیق و شناسایی نقاط پرت و عدم در نظر گرفتن فرضی مبنی بر شکل خوشه‌ها تعریف می‌شوند. با تعیین تعداد خوشه‌های مبتنی بر داده توسط الگوریتم، امکان سوگیری کمتری از جانب سرمایه‌گذار ایجاد می‌شود و با توجه به توانایی الگوریتم در تشخیص مقادیر پرت و عدم تخصیص آن‌ها به خوشه‌ها، الزامی جهت گروه‌بندی همه دارایی‌ها ایجاد نمی‌شود. علاوه بر این، انتساب در الگوریتم مورد نظر باید سختگیرانه باشد، در غیر این صورت منجر به افزایش تعداد ترکیب‌های دوتایی ممکن می‌شود که با هدف اولیه در تضاد است. در نهایت، به دلیل عدم وجود اطلاعات قبلی مبنی بر شکل‌گیری منظم خوشه‌ها، الگوریتم انتخاب شده نباید مبنی بر این فرض عمل کند. با در نظر گرفتن الزامات توصیف شده، یک الگوریتم خوشه‌بندی مبتنی بر چگالی انتخاب مناسبی به نظر می‌رسد زیرا به تشکیل خوشه‌هایی با اشکال دلخواه پرداخته و بنابراین نیاز به اتخاذ فرضیاتی برای نرمال بودن نیست و به دلیل عدم گروه‌بندی همه نقاط مجموعه داده، به طور طبیعی نسبت به موارد پرت قوی است و نیازی به تعیین تعداد خوشه‌ها نیز ندارد. الگوریتم خوشه‌بندی مکانی بر مبنای چگالی در کاربردهای دارای نویز، $DBSCAN$ ، تأثیرگذارترین الگوریتم در این دسته است. به طور خلاصه، $DBSCAN$ خوشه‌هایی از نقاط را بر اساس چگالی آنها تشخیص می‌دهد و جهت اعمال آن روی داده‌ها، دو پارامتر باید تعریف شوند: ϵ ، نشانگر این است که نقاط چه میزان باید به یکدیگر نزدیک باشند تا به عنوان نقاط مجاور تشکیل‌دهنده یک خوشه در نظر گرفته شوند و $MinPts$ ، حداقل تعداد نقاط برای تشکیل یک خوشه است. با استفاده از این دو پارامتر، در کنار برخی از مفاهیمی که در اینجا حذف می‌شود، خوشه‌هایی از نقاط مجاور شکل می‌گیرد و نقاطی که در نواحی با کمتر از $MinPts$ در دایره‌ای به شعاع ϵ قرار می‌گیرند، به عنوان نقاط پرت طبقه‌بندی می‌شوند. قابل ذکر است الگوریتم $DBSCAN$ با این فرض مناسب است که خوشه‌ها به طور

یکنواخت متراکم هستند و نشانگر ضعف این الگوریتم است. به عبارت دیگر، اگر نواحی در فضا چگالی‌های متفاوتی داشته باشند، یک ϵ ثابت ممکن است به خوبی با چگالی یک خوشه منطبق باشد اما برای دیگری غیر واقعی باشد. طبق شکل ۱، بدیهی است که خوشه A ، B و C ممکن است با استفاده از یک ϵ یافت شوند اما تمایزی بین $A1$ و $A2$ قائل نشده است. الگوریتم خوشه‌بندی ترتیب نقاط برای تشخیص ساختار خوشه، $OPTICS$ ، شبیه و بر اساس الگوریتم $DBSCAN$ بوده و با معرفی برخی مفاهیم، به پیاده‌سازی فرآیند خوشه‌بندی با پارامتر ϵ متغیر می‌پردازد و مشکل ایجاد شده در الگوریتم $DBSCAN$ را برطرف می‌کند. در تنظیمات ارتقا داده شده برای این الگوریتم، سرمایه‌گذار فقط ملزم به تعیین پارامتر $MinPts$ است زیرا مناسب‌ترین ϵ برای هر خوشه توسط الگوریتم تشخیص داده می‌شود. بنابراین استفاده از الگوریتم $OPTICS$ ، نه تنها برای محاسبه چگالی خوشه‌های متغیر، بلکه برای تسهیل کار سرمایه‌گذار پیشنهاد می‌شود.

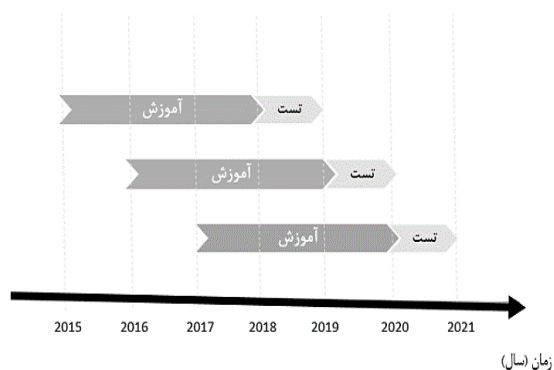


شکل ۱. خوشه‌ها با چگالی متغیر

۳-۳ معیارهای انتخاب زوج

پس از ایجاد خوشه‌های دارایی‌ها و محاسبه‌ی همه ترکیبات ممکن در خوشه‌ها، لازم است مجموعه‌ای از شرایط برای شناسایی و انتخاب زوج‌ها و به کارگیری آنها در معامله تعریف شود. رویکردهای شناسایی مبتنی بر هم‌انباشتگی عملکرد خوبی را در انتخاب زوج‌ها به دست می‌آورند زیرا این از نظر اقتصادسنجی، روابط تعادلی قابل اطمینان‌تری در زوج‌ها در مقایسه با روش‌های مبتنی بر فاصله و همبستگی توسط این رویکردها شناسایی می‌شود. بنابراین یک زوج فقط در صورتی واجد شرایط برای معامله تلقی می‌شود که دو دارایی تشکیل‌دهنده زوج با یکدیگر هم‌انباشته شوند. جهت سادگی و به منظور آزمون این شرط، استفاده از آزمون انگل-گرنجر ($Engle-Granger$) پیشنهاد می‌شود. آرمسترانگ [۳۱] به عنوان یکی از منتقدان به آزمون انگل-گرنجر اشاره می‌کند که انتخاب متغیر وابسته ممکن است به نتایج متفاوتی منجر

خودرو و ساخت قطعات، رایانه و فعالیت‌های وابسته به آن، خدمات فنی و مهندسی، زراعت و خدمات وابسته، لاستیک و پلاستیک، سیمان، آهک و گچ، فرآورده‌های نفتی، کک و سوخت هسته‌ای، محصولات غذایی و آشامیدنی به جز قند و شکر، ساخت محصولات فلزی، فلزات اساسی، سایر محصولات کانی، غیر فلزی، استخراج کانه‌های فلزی، استخراج سایر معادن، سرمایه‌گذاری‌ها، شرکت‌های چند رشته‌ای صنعتی، بانک‌ها و موسسات اعتباری، عرضه برق، گاز، بخار و آب گرم و قند و شکر و ... هستند. داده‌ها باید در دو دوره آموزش برای یافتن سودآورترین زوج‌های کاندید معامله و دوره‌ی تست برای ارزیابی نحوه عملکرد استراتژی تقسیم شوند. شکل ۲ نحوه چیدمان دوره‌ها را نشان می‌دهد.



شکل ۲. دوره‌های زمانی آموزش و تست تحقیق

سوالی که در ادامه فرآیند پیش می‌آید این است که سرمایه برای هر زوج در سبد چگونه تخصیص می‌یابد. در تئوری، اگر ارزش یکسانی در مواضع خرید و فروش دارایی‌های یک زوج در نظر گرفته شود، هزینه ورود به موقعیت خرید توسط سرمایه به‌دست‌آمده از موقعیت فروش به‌طور کامل پوشش داده می‌شود و بنابراین نیازی به تامین مالی اولیه سبد نیست و توسط خود تامین مالی می‌شود. اگرچه این فرض در تئوری معقول است اما در عمل صدق نمی‌کند، زیرا برای استقراض دارایی‌ها همیشه یک وثیقه مورد نیاز است و سرمایه‌گذاری اولیه صفر را غیرممکن می‌کند. با فرض استفاده از ابزار فروش استقراضی (*Short Selling*)، سرمایه حاصل از موقعیت فروش بلافاصله در موقعیت خرید اعمال می‌شود و این نوع اهرم توسط اکثر صندوق‌های پوشش ریسک پذیرفته می‌شود. جهت تخصیص سرمایه به سبد، زوج‌ها با وزن یکسان در سبد در نظر گرفته می‌شوند و برای ساده‌کردن محاسبات و طبق رویکرد مورد استفاده توسط اکثر نویسندگان در این زمینه، یک واحد پولی در هر زوج در هر معامله سرمایه‌گذاری می‌گردد و برای مقاصد مقایسه نیز مناسب‌تر است. این تقریب با این فرض کار می‌کند که سرمایه‌گذار

شود. برای رفع این موضوع، پیشنهاد می‌شود که آزمون انگل-گرنجر برای هر دو ترکیب ممکن از متغیر وابسته اجرا شود و ترکیبی انتخاب شود که کمترین آماره را ایجاد کند. تا این مرحله، چارچوب روش پیشنهادی تحقیق برای انتخاب زوج‌ها با رویکرد ایجاد فضاهای جستجوی ترکیبات به‌طور داده‌محور و گروه‌بندی خودکار دارایی‌های مالی توسط الگوریتم پیشنهادی در قالب سه بخش قبلی به تفصیل شرح داده شد. نحوه پیشرفت در طول سه بخش به این صورت است که نخست سری قیمتی دارایی‌ها استخراج و سپس با کاهش ابعاد داده‌ها، هر دارایی نه تنها با سری قیمتی خود بلکه با نمایشی فشرده توصیف می‌شود که حاصل استفاده از *PCA* در سری بازده است و با اجرای این اقدامات، الگوریتم *OPTICS* قادر به سازماندهی دارایی‌ها در خوشه‌ها است. سپس با جستجو و آزمون ترکیب‌های زوجی (دوتایی) ممکن در خوشه‌ها، زوج‌های مناسب طبق معیار هم‌انباشتگی انتخاب می‌شوند. در ادامه به جنبه‌ی پیاده‌سازی و ارزیابی روش پیشنهادی پرداخته و میزان دستیابی به هدف تحقیق در رابطه با بهینه‌سازی تکنیک جستجوی دارایی‌ها مشخص می‌شود. به این منظور و طبق نمای کلی طرح تحقیق، تکنیک پیشنهادی شامل اعمال خودکار فضاهای جستجوی ترکیبات ممکن با استفاده از الگوریتم خوشه‌بندی *OPTICS* به عنوان منابع استخراج زوج دارایی‌های سودآور، با دو حالت جستجوی ترکیبات با در نظر گرفتن فرض عدم خوشه‌بندی داده‌ها و خوشه‌بندی بر اساس صنعت دارایی‌ها مقایسه می‌شود. به عبارت دیگر، زوج‌های منتخب از ترکیبات هر سه فضای جستجو مورد معامله قرار گرفته و نتایج مورد ارزیابی و مقایسه قرار می‌گیرد.

۳-۴ شبیه سازی

جهت انجام معاملات به این نکته توجه می‌شود که از آنجایی که مقایسه نتایج حاصل از تکنیک‌های خوشه‌بندی برای انتخاب زوج‌های سودآور نسبت به یکدیگر به عنوان هدف اصلی این مرحله از تحقیق است، بهینه‌سازی مدل معاملاتی مدنظر نیست و مدل مبتنی بر آستانه استاندارد جهت معامله اعمال می‌شود. در این تحقیق، تمرکز روی مسئله‌ی تک‌متغیره است و در دو طرف اسپرد از یک سهام استفاده شده است. داده‌های مورد استفاده مربوط به سهام‌هایی متعلق به صنایع مختلفی از جمله مخابرات، بیمه و صندوق‌بازنشستگی، ماشین‌آلات و دستگاه‌های برقی، محصولات شیمیایی، ماشین‌آلات و تجهیزات، مواد و محصولات دارویی، انبوه‌سازی، املاک و مستغلات، حمل و نقل، انبارداری و ارتباطات، استخراج نفت و گاز و خدمات جنبی جز اکتشاف،

جهت ارزیابی نتایج سودآوری معاملات سبد زوج دارایی‌ها، معیارهای نرخ بازگشت سرمایه (ROI)، نسبت شارپ (SR) و حداکثر سقوط سبد (MDD) پیشنهاد می‌شود.

۴- نتایج

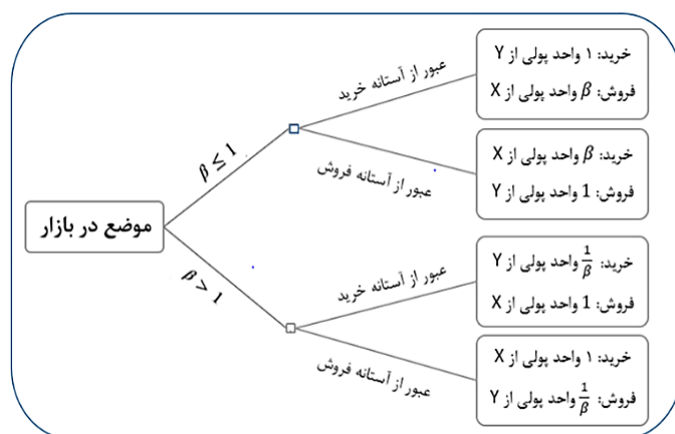
۴-۱- عملکرد انتخاب زوج دارایی‌ها

در این بخش به نتایج ارائه شده شامل تعداد زوج‌های منتخب توسط سیستم تحت سه روش مختلف خوشه‌بندی پرداخته می‌شود. همانطور که انتظار می‌رود، رویکرد تک‌خوشه‌ای دارای مجموعه بزرگ‌تری از سهم‌ها جهت بررسی ترکیبات ممکن خواهد بود و قاعدتاً زوج‌های بیشتری با استفاده از این تنظیم انتخاب می‌شوند و بررسی میزان سودآوری آن‌ها مسئله‌ی باقی‌مانده در این مرحله از تحقیق است. با توجه به نوع عملکرد مورد انتظار الگوریتم $OPTICS$ ، تعداد ترکیبات ممکن استخراج شده از خوشه‌ها می‌تواند به طور قابل ملاحظه‌ای کمتر باشد و علاوه بر کاهش بار محاسباتی انجام تست‌های آماری متعدد، احتمال یافتن زوج‌های جعلی نیز کاهش می‌یابد. عملکرد خودکار و اتکای این الگوریتم به داده‌ها، با مشاهده تغییر تعداد خوشه‌ها در دوره‌های مختلف اثبات می‌شود. همانطور که در جدول ۱ مشخص است، سه رویکرد در دوره‌های مختلف مورد بررسی بوده و تعداد ترکیبات ممکن با اعمال الگوریتم خوشه‌بند مبتنی بر چگالی $OPTICS$ روی داده‌های مربوط به دارایی‌ها و رویکرد خوشه‌بندی بر اساس صنعت دارایی‌ها، کاهش قابل توجهی داشته است. زوج‌های منتخب و واجد شرایط معامله از طریق آزمون هم‌انباشتگی در هر دوره و رویکرد مشخص شده است. برای بررسی ساختار دارایی‌های تشکیل‌دهنده هر خوشه ناشی از به‌کارگیری الگوریتم بدون نظارت پیشنهادی، روند لگاریتمی سری قیمت دارایی‌های مربوط به خوشه‌های شماره ۳، ۵ مربوط به آخرین دوره‌ی زمانی آموزش مورد نظر این تحقیق طبق شکل‌های ۴، ۵ رسم می‌شود و سری قیمت نشان داده شده حاصل از تفاضل میانگین سری قیمت اصلی جهت تسهیل تجسم است. با بررسی اشکال خوشه‌ها، دریافت می‌گردد که روند سری قیمتی مربوط به دارایی‌ها در هر خوشه به یکدیگر بسیار نزدیک بوده و تمایز آن‌ها از یکدیگر دشوار است و نیازی جهت یافتن دلیل اساسی برای توجیه این رفتار احساس می‌شود. شکل ۴ مربوط به خوشه شماره سه بوده و دارایی‌های تشکیل‌دهنده‌ی آن طبق جدول ۲ به گروه‌های محصولات شیمیایی، شرکت‌های چندرشته‌ای صنعتی، عرضه برق، گاز، بخار و آب گرم، فرآورده‌های نفتی

می‌تواند کسری از دارایی‌های معاملاتی را خریداری کند زیرا خود دارایی‌ها به تا حتمال زیاد دقیقاً یک واحد پولی ارزش ندارند. بر این اساس، نیاز به تنظیم چارچوبی است که تضمین کند هر معامله فقط با یک واحد پولی انجام شده و از رابطه ۵ استفاده کند که در آن leg_1 و leg_2 نشان‌دهنده میزان سرمایه‌گذاری انجام شده در هر پایه از زوج هستند. قابل ذکر است نسبت هم‌انباشتگی β بین دو اوراق بهادار توسط این رویکرد رعایت می‌شود و همانطور که در شکل ۳ نمایش داده شده است، جهت اخذ مواضع معاملاتی از نسبت هم‌انباشتگی استفاده می‌شود.

$$\max(leg_1, leg_2) = 1 \quad (5)$$

با ادامه‌ی روند معاملات توسط سرمایه‌گذار، فرضی مبنی بر به‌کارگیری مجدد تمام سرمایه و سود کسب شده توسط یک زوج در طول دوره معاملاتی در معامله بعدی آن زوج در نظر گرفته می‌شود. هزینه‌های معاملاتی نیز برای تمام نتایج ارائه شده در این کار محاسبه می‌شود که عبارتند از: هزینه‌ی استقرای ($Rental$)، هزینه‌ی کمیسیون ($Commission$) و اثر بازار ($Market Impact$). از آنجایی که یک سیستم معاملاتی نمی‌تواند فوراً عمل کند، ممکن است یک انحراف کوچک در قیمت ورودی وجود داشته باشد که به دلیل تاخیر در ورود به یک موضع معاملاتی است. با احتساب این عامل و اطمینان از اینکه استراتژی در عمل قابل اجرا است، یک تاخیر محافظه کارانه یک دوره ای (۱۵ دقیقه) برای ورود به یک موضع در نظر گرفته می‌شود. سیستم مورد نظر فاقد ادراک حد ضرر بوده و یک موضع معاملاتی تنها در صورت مشاهده رفتار همگرایی اسپرد زوج یا پایان دوره معاملاتی بسته می‌شود.



شکل ۳. چارچوب موضع‌گیری معاملاتی بازار

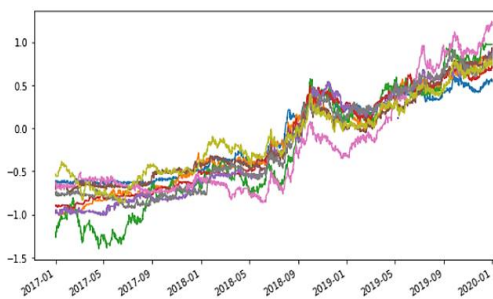
جدول ۱. رویکردهای خوشه بندی

۲۰۱۷-۲۰۲۰	۲۰۱۶-۲۰۱۹	۲۰۱۵-۲۰۱۸	
رویکرد تک خوشه ای			
۱	۱	۱	تعداد خوشه ها
۴۹۵۰	۴۹۵۰	۴۹۵۰	تعداد ترکیبات دوتایی ممکن
۲۰۲۳	۸۴۷	۶۵۵	زوج های منتخب
رویکرد خوشه بندی با الگوریتم OPTICS			
۵	۶	۱۲	تعداد خوشه ها
۱۴۷	۱۲۸	۱۸۰	تعداد ترکیبات دوتایی ممکن
۸۴	۲۶	۲۳	زوج های منتخب
رویکرد خوشه بندی بر اساس صنعت دارایی ها			
۲۴	۲۴	۲۴	تعداد خوشه ها
۲۲۵	۲۲۵	۲۲۵	تعداد ترکیبات دوتایی ممکن
۹۷	۷۵	۴۸	زوج های منتخب

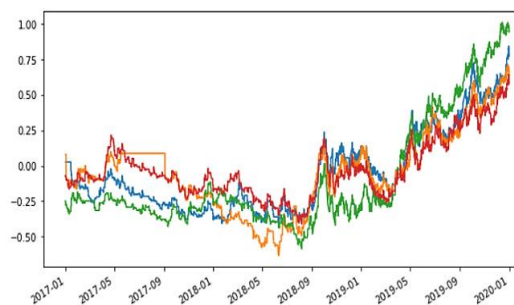
جدول ۲. خوشه بندی با الگوریتم OPTICS - دوره زمانی ۲۰۱۷ تا ۲۰۲۰

شماره خوشه	شماره نماد	نماد	گروه	
۱	۳	بالبر	ماشین آلات و دستگاه های برقی	
	۱۵	نشاءد	انبوه سازی، املاک و مستغلات	
	۱۸	حتوکا	حمل و نقل، انبار داری و ارتباطات	
	۳۷	زمگسا	زراعت و خدمات وابسته	
	۳۸	سپهان	سیمان، آهک و گچ	
	۳۹	ستران	سیمان، آهک و گچ	
	۴۱	سیدکو	سیمان، آهک و گچ	
	۵۱	غشهر	محصولات غذایی و آشامیدنی به جز قند و شکر	
	۵۲	غمارگ	محصولات غذایی و آشامیدنی به جز قند و شکر	
	۹۸	قارست	قند و شکر	
۲	۶۷	لیوتان	ماشین آلات و تجهیزات	
	۷۸	وصنا	سرمایه گذاریها	
	۸	پکرمان	لاستیک و پلاستیک	
	۱	البرز	بیمه و صندوق بازنشستگی	
	۲۷	دالبر	مواد و محصولات دارویی	
	۳۰	دسیجا	مواد و محصولات دارویی	
	۳۲	دکوثر	مواد و محصولات دارویی	
	۱۷	جم	محصولات شیمیایی	
	۴۳	شپهرن	فرآورده های نفتی، کک و سوخت هسته ای	
	۵۸	فراور	فلزات اساسی	
۳	۹۵	مبین	عرضه برق، گاز، بخار و آب گرم	
	۷۱	وامید	شرکت های چند رشته ای صنعتی	
	۷۹	وصندوق	شرکت های چند رشته ای صنعتی	
	۸۵	ونیکو	شرکت های چند رشته ای صنعتی	
	۶۳	کرمانشا	محصولات شیمیایی	
	۶۴	کروی	استخراج کانه های فلزی	
	۴	۸۷	تاپیکو	محصولات شیمیایی
		۴۶	شتران	فرآورده های نفتی، کک و سوخت هسته ای
		۹۱	شپنا	فرآورده های نفتی، کک و سوخت هسته ای
		۹۲	فخوز	فلزات اساسی
۵۹		فملی	فلزات اساسی	
۷۲		ویانک	شرکت های چند رشته ای صنعتی	
۸۱		ومعدان	استخراج کانه های فلزی	
۲۱		خرامیا	خودرو و ساخت قطعات	
۲۲		خساپا	خودرو و ساخت قطعات	
۵		۲۵	خودرو	خودرو و ساخت قطعات
	۹۰	خیارس	خودرو و ساخت قطعات	

کک و سوخت هسته ای، فلزات اساسی و استخراج کانه های فلزی هستند. چند مورد از سری های شناسایی شده در این خوشه مربوط به گروه یکسان و از شرکت های چندرشته ای صنعتی بوده و همچنین دو سری قیمتی از آن متعلق به گروه محصولات شیمیایی هستند. با بررسی کلی در این خوشه قابل استنباط است که همه ی این گروه ها می توانند زیرمجموعه ای از گروه بزرگتری مانند وابستگی به فعالیت استخراج از منابع و معادن هستند. این اولین مدرکی است که نشان می دهد رویکرد مبتنی بر الگوریتم معرفی شده می تواند در گروه بندی دارایی ها یک سطح عمیق تر پیش رود زیرا قادر است بخش های خاص و بزرگتر را نیز شناسایی کند. وضعیت مشابهی در خوشه ی شماره ۴ نیز در جدول ۲ مشاهده میشود. در این مورد، دارایی های شناسایی شده نیز متعلق به گروه های محصولات شیمیایی، شرکت های چندرشته ای صنعتی، فرآورده های نفتی کک و سوخت هسته ای، فلزات اساسی و استخراج کانه های فلزی هستند. طبق شکل ۵، خوشه پنجم از دارایی های گروه و صنعت کاملاً یکسان تشکیل شده است و دارایی های آن ها به ترتیب متعلق به گروه خودرو و ساخت قطعات هستند. مابقی خوشه ها نیز ثابت می کنند که قابلیت خوشه بندی این الگوریتم فراتر از انتخاب دارایی ها از یک صنعت است و یک رابطه قابل مشاهده بین سری های قیمتی شناسایی شده وجود دارد. بنابراین می توان تایید کرد که الگوریتم پیشنهادی، به گروه بندی دارایی ها از یک صنعت می پردازد در حالی که مانعی برای خوشه هایی حاوی دارایی ها از بخش های مختلف وجود ندارد.



شکل ۴. خوشه ی ۳ - دوره زمانی ۲۰۱۷ تا ۲۰۲۰



شکل ۵. خوشه ی ۵ - دوره زمانی ۲۰۱۷ تا ۲۰۲۰

۴-۲ عملکرد معاملاتی

در این بخش، نحوه عملکرد زوج‌های انتخاب شده تحت شرایط خوشه‌بندی مجزا از نقطه نظر معاملاتی تحلیل و مشخص می‌شود که کدام تکنیک خوشه‌بندی، امیدوارکننده‌ترین نمونه را از لحاظ سودآوری ایجاد می‌کند. مدل معاملاتی مبتنی بر آستانه جهت حصول و مقایسه نتایج پیشنهاد شده است. در صورت حصول اطمینان بالاتر از وجود زوج‌های هم‌انباشته در سبد معاملاتی، انتظار می‌رود که نتایج معامله رضایت‌بخش‌تر باشد اما از سوی دیگر، ممکن است این نتیجه حاصل نشود زیرا به تعدیل و حذف زوج‌هایی که سودآور نبوده‌اند، نپرداخته ایم. بنابراین این انگیزه پدید می‌آید که می‌توان به بهینه‌سازی سبد یا میزان سرمایه‌گذاری در هر کدام از زوج‌ها در سبد مورد نظر پرداخت تا نتایج بهبود داده شوند.

طبق نتایج، یکی از ایده‌های کلیدی این است که استراتژی در هر سناریویی با حاشیه سود رضایت‌بخش، سودآور است و تا این مرحله می‌توان موفقیت‌فرآیند خوشه‌بندی خودکار توسط الگوریتم پیشنهادی در انتخاب زوج‌هایی با سودآوری پایدار در هر یک از محیط‌های آزمایش شده تأیید کرد و به طور خاص، این ایده تأیید می‌شود که روش مطرح شده برای انتخاب زوج‌داری‌ها، روش قدرتمندی است. نکته‌ی قابل توجه در این روش انتخاب زوج‌داری‌ها این است که اگر سرمایه‌گذار نگران ریسکی باشد که سبد خود در معرض آن قرار دارد، می‌تواند با استفاده از الگوریتم خوشه‌بند خودکار معرفی شده در این مقاله، به تعادل بهتری بین بازده و ریسک دست یابد. همانطور که در جدول پیوست ۱ مشخص است، معامله در چهار سطح آستانه به صورت ضربی از انحراف استاندارد اسپرد ($0.5\sigma, 0.75\sigma, 1\sigma, 2\sigma$) به عنوان سیگنال‌های اخذ مواضع خرید و فروش با استفاده از سه دوره‌ی تست مجزا (سال‌های ۲۰۲۰، ۲۰۱۹، ۲۰۱۸) انجام شده و نتایج در قالب میانگین نتایج حاصل شده در سطوح آستانه‌ای در دوره‌های مختلف بدست آمده است. مقادیر متوسط نسبت شارپ و نرخ بازگشت سرمایه در حالت استفاده از الگوریتم *OPTICS* برای شناسایی زوج‌های معاملاتی به ترتیب مقادیر ۳.۴۴ و ۴.۷۰ است که نسبت به حالت تک خوشه‌ای با مقادیر ۳.۲۸ و ۱۰.۸۲ بزرگتر شده و نشان‌دهنده‌ی عملکرد خوب رویکرد معرفی شده برای انتخاب زوج‌های معاملاتی است. اگر به جزئیات نتایج هر دوره توجه شود، مشاهده می‌شود که در اولین دوره‌ی تست و در حالتی که آستانه یک برابر انحراف استاندارد در نظر گرفته شده است، معامله‌ی انجام شده در اولین دوره‌ی تست مدنظر با در نظر گرفتن رویکرد استفاده

از الگوریتم *OPTICS* برای خوشه‌بندی داده‌ها، نسبت شارپی معادل ۴.۷۱ را اتخاذ کرده که در مقایسه با نسبت شارپ در رویکرد تک خوشه‌ای، ۵.۱۲، کمتر است. همچنین در دوره‌ی تست دوم مسئله و با آستانه‌ی ۰.۷۵ σ ، مقدار نسبت شارپ با در نظر گرفتن رویکرد استفاده از الگوریتم *OPTICS* برای خوشه‌بندی داده‌ها، مقدار ۱.۶۸ است که در مقایسه با نسبت شارپ معامله در حالت استفاده از رویکرد تک خوشه‌ای، ۱.۸۸، مقدار کمتری بدست آمده است اما در بقیه موارد ارتقاء در مقدار معیار نسبت شارپ و نرخ بازگشت سرمایه و به طور کلی در میانگین سه ساله این معیارها قابل مشاهده است. از لحاظ معیار بیشترین سقوط، رویکرد معرفی شده نسبت به رویکرد تک خوشه‌ای خوب عمل نکرده و باعث شده است که متوسط معیار بیشترین سقوط عدد ۰.۱۲ را در مقابل ۰.۱۰، اتخاذ نماید. همچنین قابل ذکر است که درصد زوج‌های سودآور و درصد سودآوری هر زوج در معامله‌ی انجام گرفته با استفاده از الگوریتم معرفی شده جهت خوشه‌بندی داده‌ها افزایش داشته است. همچنین رویکرد معرفی شده می‌تواند عدد بزرگتری از لحاظ نسبت معاملات سودآور به کل معاملات انجام گرفته به ارمغان آورد. رویکرد خوشه‌بندی بر اساس صنعت‌داری‌ها نیز با توجه به نتایج و متوسط معیارهای ارزیابی نسبت به دو رویکرد قبلی ضعیف‌تر عمل کرده و فقط در متوسط معیار حداکثر سقوط نسبت به رویکرد خوشه‌بندی با الگوریتم *OPTICS* مقداری بهتر عمل کرده است.

۴-۳ عملکرد سبد بهینه

از آنجایی که در بخش ۳-۴ نحوه‌ی تشکیل سبد متشکل از زوج‌داری‌ها مطرح شد، وزن یکسانی برای تخصیص سرمایه‌ی اولیه به زوج‌ها در ابتدای دوره (یک واحد پولی) در نظر گرفته شد و همین مسئله انگیزه‌ی بهینه‌سازی سبد و وزن زوج‌داری‌ها را ایجاد نمود. بعد از مشخص شدن رویکرد خوشه‌بندی مناسب بنابراین زوج‌های منتخب از رویکرد خوشه‌بندی برتر برای انتخاب زوج در طرح تحقیق که عملکرد بهتری در مقایسه با رویکردهای دیگر از نظر معیارهای ارزیابی در مدل معاملاتی کسب کرده‌اند، مورد استفاده‌ی الگوریتم بهینه‌سازی سبد قرار گرفته و اطلاعات پایش شده (سری بازده) ناشی از به‌کارگیری این زوج‌داری‌ها در مدل معاملاتی در سال‌های مختلف آموزشی مربوط به تشکیل زوج (دوره‌ی شناسایی زوج‌ها) به عنوان ورودی الگوریتم جهت بهینه‌سازی نسبت شارپ سالانه (دو هدف متناقض: بیشترین بازده و کمترین ریسک یا انحراف معیاربازده سالانه) در نظر گرفته می‌شود تا وزن‌های بهینه برای تخصیص سرمایه اولیه به هر زوج در سبد در ابتدای

۵- نتیجه گیری و پیشنهادات آتی

هدف مهم دنبال شده در این تحقیق، بررسی عملکرد الگوریتم بدون نظارت *OPTICS* در ایجاد تعادل بین سودآوری و خطای ناشی از شناسایی زوج دارایی‌ها در معاملات زوجی است. به عبارت دیگر ایجاد سیدی از زوج دارایی‌ها با سودآوری بالا، نیازمند تقسیم‌بندی فضای جستجوی زوج دارایی‌ها جهت بررسی ترکیبات دوتایی از هر بخش توسط آزمون‌های آماری شناسایی زوج‌ها است که علاوه بر ایجاد بستر شناسایی زوج‌های غیرمتداول و دارای پتانسیل سودآوری در معاملات، باعث افزایش میزان خطا و شناسایی روابط جعلی بین دارایی‌ها ناشی از تعدد بررسی ترکیبات نشود. نحوه جستجوی ترکیبات در تحقیقات گذشته مبتنی بر دو رویکرد غیر خودکار ۱- یک جستجوی جامع برای همه ترکیب‌های دوتایی ممکن در میان دارایی‌های مختلف و ۲- جستجو و محدود کردن ترکیب‌ها به دارایی‌های موجود در یک صنعت (به عنوان معیار متداول سرمایه‌گذاران جهت تقسیم‌بندی فضای جستجوی زوج دارایی‌ها) بوده است که به ترتیب زمینه‌ی افزایش خطا در شناسایی زوج‌های مناسب و کاهش سود در معاملات را فراهم می‌کنند. بنابراین الگوریتمی داده‌محور جهت پیش‌تقسیم‌بندی یا خوشه‌بندی مؤثر و خودکار اجتماع دارایی‌ها مورد نیاز است که علاوه بر عدم محدودسازی نتایج بررسی ترکیب دارایی‌ها به زوج‌های نسبتاً رایج در معاملات زوجی، به اعمال فضای گسترده برای جستجوی ترکیبات نیز نپردازد. الگوریتم خوشه‌بند مورد نظر با توجه به عملکرد مورد انتظار مؤثر و خودکار خود نباید نیازی به تعیین پارامترها توسط سرمایه‌گذار از جمله تعداد خوشه‌ها داشته باشد و از طرفی به خوشه‌بندی همه دارایی‌ها نپردازد و رفتار بهینه نسبت به نقاط پرت داشته باشد. از این رو با توجه به اهداف تعیین شده برای عملکرد الگوریتم مورد نظر، الگوریتم‌های مبتنی بر چگالی در میان الگوریتم‌های بدون نظارت مطلوب به نظر می‌رسند. حاصل بررسی الگوریتم‌های مبتنی بر چگالی، به‌کارگیری الگوریتم *OPTICS* با توجه به مطلوبیت برای خوشه‌هایی با چگالی متغیر و توانایی در تشخیص پارامتر مناسب برای هر خوشه به طور خودکار است. نتایج نشان می‌دهد در صورت تمرکز سرمایه‌گذار بر کسب بیشترین بازده در هر دو حالت توجه یا عدم توجه به ریسک متحمل شده، استراتژی مبتنی بر الگوریتم *OPTICS* عملکرد بهتری نسبت به حالت تک خوشه‌ای و خوشه‌بندی بر اساس صنعت دارد و خوشه‌بندی با استفاده از الگوریتم معرفی شده منجر به شناسایی زوج‌های امیدوارکننده‌تر از لحاظ سودآوری در معاملات زوجی باتوجه به شاخص

دوره‌ی معاملاتی (یک سال) استخراج شود. الگوریتم مورد استفاده برای حل مسئله‌ی دو هدفه، الگوریتم تکاملی و مبتنی بر ژنتیک (*NSGA2*) است. دوره‌ی بهینه‌سازی مورد نظر در این کار یک ساله است و هدف ایجاد تعادل بین دو هدف متناقض بازده و ریسک (واریانس) سالانه است. رویکرد بهینه برای تقسیم‌بندی فضای جستجوی زوج‌ها و خوشه‌بندی دارایی‌ها، رویکرد پیشنهادی مبنی بر استفاده از الگوریتم *OPTICS* است و اوزان بهینه برای زوج‌های منتخب ناشی از به‌کارگیری این الگوریتم در دوره‌های آموزشی سه ساله به طور مجزا و با استفاده از الگوریتم مبتنی بر ژنتیک پیشنهادی استخراج می‌شود. سرمایه اولیه با توجه به اوزان بهینه به زوج‌ها جهت اجرای معامله اختصاص داده می‌شود و نسبت شارپ و نرخ بازگشت سرمایه حاصل از اجرای معامله در دوره‌های تست به طور مجزا با در نظر گرفتن 0.75σ به عنوان آستانه‌ی معامله طبق جداول ۳ تا ۵ بدست می‌آید که نسبت به حالت سبد با تخصیص وزن یکسان برای زوج‌ها، افزایش داشته‌اند که نشانگر مؤثر بودن رویکرد بهینه‌سازی سبد از طریق الگوریتم بهینه‌سازی دوهدفه مبتنی بر ژنتیک است.

جدول ۳. مقایسه عملکرد معاملاتی سبد زوجی با وزن یکسان و وزن بهینه سال ۲۰۱۸

۲۰۱۸		۰.۷۵σ
سبد با وزن های یکسان	سبد با وزن های بهینه	
نسبت شارپ (SR)	۴.۸۶۱	۷.۴۳۶
نرخ بازگشت سرمایه % (ROI)	۱۰۰.۰۹۵	۱۴۱.۱۸۴
بیشترین سقوط % (MDD)	-۴.۸۹%	-۵.۸۱%

جدول ۴. مقایسه عملکرد معاملاتی سبد زوجی با وزن یکسان و وزن بهینه - سال ۲۰۱۹

۲۰۱۹		۰.۷۵σ
سبد با وزن های یکسان	سبد با وزن های بهینه	
نسبت شارپ (SR)	۱.۶۸۸	۳.۱۲۰
نرخ بازگشت سرمایه % (ROI)	۷۲.۵۶۵	۸۵.۸۷۱
بیشترین سقوط % (MDD)	-۱۸.۱۳%	-۱۸.۸۷%

جدول ۵. مقایسه عملکرد معاملاتی سبد زوجی با وزن یکسان و وزن بهینه - سال ۲۰۲۰

۲۰۲۰		۰.۷۵σ
سبد با وزن های یکسان	سبد با وزن های بهینه	
نسبت شارپ (SR)	۷.۷۱۱	۹.۸۱۱
نرخ بازگشت سرمایه % (ROI)	۳۱۴.۷۸۰	۳۴۰.۸۵۵
بیشترین سقوط % (MDD)	-۳۷.۶۸%	-۳۷.۶۹%

<https://doi.org/https://doi.org/10.1016/j.ejor.2016.10.031>

4. Do, B., & Faff, R. (2010). Does Simple Pairs Trading Still Work? *Financial Analysts Journal*, 66(4), 83-95. <https://doi.org/10.2469/faj.v66.n4.1>
5. Dunis, C., Giorgioni, G., Laws, J., & Rudy, J. (2010). Statistical Arbitrage and High-Frequency Data with an Application to Eurostoxx 50 Equities. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2272605>
6. Avellaneda, M., & Lee, J.-H. (2010). Statistical arbitrage in the US equities market. *Quantitative Finance*, 10(7), 761-782. <https://doi.org/10.1080/14697680903124632>
7. Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs Trading: Performance of a Relative-Value Arbitrage Rule. *The Review of Financial Studies*, 19(3), 797-827. <https://doi.org/10.1093/rfs/hhj020>
8. Chen, H., Chen, J., Chen, Z., & Li, F. (2017). Empirical Investigation of an Equity Pairs Trading Strategy. *Management Science*, 65. <https://doi.org/10.1287/mnsc.2017.2825>
9. Huck, N., & Afawubo, K. (2015). Pairs trading and selection methods: is cointegration superior? *Applied Economics*, 47(6), 599-613. <https://doi.org/10.1080/00036846.2014.975417>
10. Rad, H., Low, R. K. Y., & Faff, R. (2016). The profitability of pairs trading strategies: distance, cointegration and copula methods. *Quantitative Finance*, 16(10), 1541-1558. <https://doi.org/10.1080/14697688.2016.1164337>
11. Li, Y., & Law, K. K. F. (2021). Systematic risk in pairs trading and dynamic parameterization. *Economics Letters*, 202, 109842. <https://doi.org/https://doi.org/10.1016/j.econlet.2021.109842>
12. Goldkamp, J., & Dehghanimohammadabadi, M. (2019). Evolutionary multi-objective optimization for multivariate pairs trading. *Expert Systems with Applications*, 135, 113-128. <https://doi.org/https://doi.org/10.1016/j.eswa.2019.05.046>
13. Gupta, K., & Chatterjee, N. (2020). Selecting stock pairs for pairs trading while incorporating lead-lag relationship. *Physica A: Statistical Mechanics and its Applications*, 551, 124103. <https://doi.org/https://doi.org/10.1016/j.physa.2019.124103>

های مورد ارزیابی تحقیق از جمله میانگین نرخ بازگشت سرمایه و نسبت شارپ سبد در سه سال است. همانطور که محاسبات مبتنی بر فرض معامله‌ی سبد با زوج‌های هم‌وزن بوده است، جهت کسب نتایج بهتر برای معیارهای ارزیابی، به بهینه‌سازی سبد متشکل از زوج‌داری‌ها و اختصاص بهینه سرمایه به زوج‌ها با استفاده از الگوریتم تکامل یافته *NSGA2* با هدف افزایش بازده همراه با کاهش ریسک پرداخته شد و ارتقا در نتایج قبلی سبد از جمله نسبت شارپ و نرخ بازگشت سرمایه در سه سال به طور مجزا حاصل شد.

در مدل معاملاتی زوجی، پارامترهای زیادی از جمله مدت دوره‌های زمانی آموزش و تست، سطح اطمینان در آزمون (هم انباشتگی) دیکی فولر، دامنه‌ی انحرافات استاندارد برای آستانه وجود دارد و با توجه به اینکه تحت مقادیر مختلف برای آستانه‌ها، نتایج مختلفی حاصل شده است، بنابراین جهت کسب نتایج بهتر می‌توان به بهینه‌سازی پارامترها و مقادیر آستانه با توجه به ویژگی‌های سری‌های زمانی در هر خوشه با به‌کارگیری روش‌ها و تکنیک‌های مختلف پرداخت. همچنین می‌توان معاملات را با تنظیم سطح توقف ضرر یا اعمال فیلترکالمن روی سری‌های زمانی انجام داد. امکان تغییر فرکانس داده‌های درون‌روزی (مثلاً به فرکانس ۵ دقیقه) و یا استفاده از ویژگی‌های دیگری از داری‌ها نیز وجود دارد و ضمناً می‌توان به بررسی شرایط بازاری که در آن استراتژی پیشنهادی بازده بیشتری ایجاد می‌کند پرداخت و مشخص کرد که آیا سرمایه‌گذار می‌تواند از پیش‌بینی چنین سناریوهایی به سودآوری برسد. استفاده از داری‌های مختلف در بازارهای مختلف و ترکیبات آنها در معاملات زوجی نیز می‌تواند به توسعه این معاملات کمک کند.

فهرست منابع

1. Cavalcante, R., Brasileiro, R., Souza, V., Nobrega, J., & Oliveira, A. (2016). Computational Intelligence and Financial Markets: A Survey and Future Directions. *Expert Systems with Applications*, 55. <https://doi.org/10.1016/j.eswa.2016.02.006>
2. Caldeira, J., & Moura, G. (2013). Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy. *Revista Brasileira de Finanças*, 11. <https://doi.org/10.2139/ssrn.2196391>
3. Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689-702.

- Supercomputing, 78(3), 3843-3882. <https://doi.org/10.1007/s11227-021-04013-x>
24. Han, C., He, Z., & Toh, A. J. W. (2023). Pairs trading via unsupervised learning. *European Journal of Operational Research*, 307(2), 929-947. <https://doi.org/https://doi.org/10.1016/j.ejor.2022.09.041>
 25. Valle, C. A., Meade, N., & Beasley, J. E. (2014). Market neutral portfolios. *Optimization Letters*, 8(7), 1961-1984. <https://doi.org/10.1007/s11590-013-0714-6>
 26. Valle, C. A., Meade, N., & Beasley, J. E. (2014). Absolute return portfolios. *Omega*, 45, 20-41. <https://doi.org/https://doi.org/10.1016/j.omega.2013.12.003>
 27. Lin, T.-Y., Chen, C. W. S., & Syu, F.-Y. (2021). Multi-asset pair-trading strategy: A statistical learning approach. *The North American Journal of Economics and Finance*, 55, 101295. <https://doi.org/https://doi.org/10.1016/j.najef.2020.101295>
 28. Huang, C.-F., Hsu, C.-J., Chen, C.-C., Chang, B. R., & Li, C.-A. (2015). An Intelligent Model for Pairs Trading Using Genetic Algorithms. *Computational Intelligence and Neuroscience*, 2015, 939606. <https://doi.org/10.1155/2015/939606>
 29. Bellman, R. E. (2010). *Dynamic Programming*. Princeton University Press. <https://doi.org/doi:10.1515/9781400835386>
 30. Berkhin, P. (2006). A Survey of Clustering Data Mining Techniques. In J. Kogan, C. Nicholas, & M. Teboulle (Eds.), *Grouping Multidimensional Data: Recent Advances in Clustering* (pp. 25-71). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-28349-8_2
 31. Armstrong, J. (2001). Combining Forecasts. In (Vol. 30). https://doi.org/10.1007/978-0-306-47630-3_19
 14. Dunis, C. L., Laws, J., & Evans, B. (2016). Modelling and trading the gasoline crack spread: A non-linear story. In (pp. 140-160). <https://doi.org/10.1057/9781137554178>
 15. Dunis, C., Laws, J., Evans, B., & John, L. (2006). Modelling and Trading the Soybean-Oil Crush Spread with Recurrent and Higher Order Networks: A Comparative Analysis. *Neural Network World*, 16. <https://doi.org/10.4018/978-1-59904-897-0.ch016>
 16. Dunis, C. L., Laws, J., Middleton, P. W., & Karathanasopoulos, A. (2015). Trading and hedging the corn/ethanol crush spread using time-varying leverage and nonlinear models. *The European Journal of Finance*, 21(4), 352-375. <https://doi.org/10.1080/1351847X.2013.830140>
 17. Thomaidis, N., Kondakis, N., & Dounias, G. (2006). An Intelligent Statistical Arbitrage Trading System. https://doi.org/10.1007/11752912_77
 18. Huck, N. (2009). Pairs selection and outranking: An application to the S&P 100 index. *European Journal of Operational Research*, 196(2), 819-825. <https://doi.org/https://doi.org/10.1016/j.ejor.2008.03.025>
 19. Huck, N. (2010). Pairs trading and outranking: The multi-step-ahead forecasting case. *European Journal of Operational Research*, 207(3), 1702-1716. <https://doi.org/https://doi.org/10.1016/j.ejor.2010.06.043>
 20. Sarmento, S. M., & Horta, N. (2020). Enhancing a Pairs Trading strategy with the application of Machine Learning. *Expert Systems with Applications*, 158, 113490. <https://doi.org/https://doi.org/10.1016/j.eswa.2020.113490>
 21. Flori, A., & Regoli, D. (2021). Revealing Pairs-trading opportunities with long short-term memory networks. *European Journal of Operational Research*, 295(2), 772-791. <https://doi.org/https://doi.org/10.1016/j.ejor.2021.03.009>
 22. Kim, T., & Kim, H. Y. (2019). Optimizing the Pairs-Trading Strategy Using Deep Reinforcement Learning with Trading and Stop-Loss Boundaries. *Complexity*, 2019, 3582516. <https://doi.org/10.1155/2019/3582516>
 23. Lu, J.-Y., Lai, H.-C., Shih, W.-Y., Chen, Y.-F., Huang, S.-H., Chang, H.-H., . . . Dai, T.-S. (2022). Structural break-aware pairs trading strategy using deep reinforcement learning. *The Journal of*

جدول پیوست ۱. عملکرد معاملاتی در رویکردهای خوشه بندی داده ها

متوسط (Avg)	۲۵			۱۵			۰.۷۵۵			۰.۵۵			
	۲۰۲۰	۲۰۱۹	۲۰۱۸	۲۰۲۰	۲۰۱۹	۲۰۱۸	۲۰۲۰	۲۰۱۹	۲۰۱۸	۲۰۲۰	۲۰۱۹	۲۰۱۸	
رویکرد تک خوشه ای													
۳.۴۴۸	۷.۰۳۴	-۰.۶۹۸	۲.۶۱۷	۵.۲۲۱	۴.۰۳۵	۵.۱۲۳	۶.۷۵۹	۱.۸۸۸	۴.۲۱۹	۱.۲۷۶	۰.۱۳۷	۳.۷۷۰	نسبت شارپ (SR)
۱۰۸.۲۸	۴۲.۶۳۳	۱۶.۳۳۱	۲۷.۶۸۰	۲۵۷.۷۵۷	۶۹.۷۶۹	۶۸.۲۰۴	۲۷۴	۶۵.۳۱۱	۷۱.۱۰۱	۲۶۹.۷۵۰۴	۵۷.۲۵۶	۷۹.۵۹۵	نرخ بازگشت سرمایه % (ROI)
-۰.۱۰۵	-۰.۸۰%	-۲.۵۵%	-۱.۶۸%	-۸.۹۷%	-۱۷.۳۸%	-۳.۷۵%	-۱۳.۶۷%	-۲۱.۱۳%	-۴.۵۱%	-۲۱.۰۳%	-۲۵.۶۹%	-۵.۹۴%	بیشترین سقوط % (MDD)
۱۱۷۵	۲۰۲۳	۸۴۷	۶۵۵	۲۰۲۳	۸۴۷	۶۵۵	۲۰۲۳	۸۴۷	۶۵۵	۲۰۲۳	۸۴۷	۶۵۵	کل جفت ها
۸۶.۲۸۰	۸۴.۳۷۹	۶۵.۲۸۹	۸۵.۸۰۱	۹۰.۶۰۸	۹۳.۹۷۸	۹۰.۰۷۶	۸۸.۷۷۹	۸۹.۰۲۰	۹۰.۰۷۶	۸۷.۴۴۴	۸۱.۸۱۸	۸۸.۰۹۱	درصد جفت های سودآور
۷۶.۲۵۲	۸۵.۸۳۵	۶۸.۸۴۸	۸۰.۵۷۷	۷۹.۹۰۱	۷۱.۴۴۳	۷۸.۰۸۹	۷۷.۸۲۵	۷۱.۴۹۴	۷۷.۶۱۶	۷۵.۷۳۸	۷۱.۰۰۲	۷۶.۶۶۲	درصد معاملات سودآور
۷۲.۸۲۶	۸۶.۲۸۱	۶۲.۲۱۴	۷۸.۱۰۶	۸۰.۶۰۱	۶۵.۳۰۰	۷۲.۰۹۷	۷۸.۲۵۸	۶۴.۹۱۶	۷۲.۴۷۴	۷۵.۷۷۲	۶۵.۳۷۳۹	۷۱.۵۲۳	متوسط سودآوری هر جفت %
رویکرد خوشه بندی با الگوریتم OPTICS													
۴.۷۰۳	۷.۱۳۱	-۰.۷۶۴	۲.۹۵۴	۱۴.۵۴۴	۴.۴۰۹	۴.۷۱۹	۷.۷۱۱۶	۱.۶۸۸	۴.۸۶۱	۱.۹۳۶	۱.۱۲۸۷	۴.۵۹۰	نسبت شارپ (SR)
۱۳۰.۶۶۳	۵۷.۸۱۷	۲۱.۶۴۹	۳۹.۷۳۶	۳۰۲.۳۹۰	۸۶.۳۲۸	۸۷.۵۰۵	۳۱۴.۷۸۰	۷۲.۵۶۵	۱۰۰.۰۹۵	۳۰۸.۹۲۳	۶۴.۹۹۷	۱۱۱.۱۷۱	نرخ بازگشت سرمایه % (ROI)
-۰.۱۲۰	-۱.۹۳%	-۱.۹۶%	-۱.۷۲%	-۱۰.۵۵%	-۱۲.۴۷%	-۴.۲۳%	-۳۷.۶۸%	-۱۸.۱۳%	-۴.۸۹%	-۲۳.۳۶%	-۲۲.۶۷%	-۵.۳۴%	بیشترین سقوط % (MDD)
۴۴.۳۳۳	۸۴	۲۶	۲۳	۸۴	۲۶	۲۳	۸۴	۲۶	۲۳	۸۴	۲۶	۲۳	کل جفت ها
۸۹.۹۱۳	۸۹.۲۸۵	۸۰.۹۵۲	۸۶.۹۵۶	۹۲.۸۵۷	۹۶.۱۵۳	۹۵.۶۵۲	۹۰.۴۷۶	۹۲.۳۰۷	۸۶.۹۵۶	۸۸.۰۹۵	۹۲.۳۰۷	۸۶.۹۵۶	درصد جفت های سودآور
۷۸.۵۲۲	۹۰.۶۳۵	۸۰	۸۸.۶۳۶	۸۱.۴۷۰	۷۰.۸۳۳	۸۰	۷۶.۷۱۵	۷۰.۱۷۵	۸۱.۹۵۴	۷۵	۶۶.۶۶۶	۸۰.۱۹۸	درصد معاملات سودآور
۷۶.۰۰۱۸	۸۹.۸۷۳	۸۲.۲۸۰	۸۴.۴۲۰	۸۱.۲۶۶	۶۴.۸۲۴	۷۶.۷۵۸	۷۶.۳۳۳	۶۳.۵۲۸	۷۵.۹۲۲	۷۵.۲۴۷	۶۴.۱۷۷	۷۷.۳۸۹	متوسط سودآوری هر جفت %
رویکرد گروه بندی بر اساس صنعت (تعداد خوشه ها = تعداد صنایع مورد استفاده در مقاله)													
۲.۵۰۳	۳.۸۳۵	-۱.۸۷۱	۲.۴۰۱	۸.۴۳۹	۱.۰۲۰۱	۴.۱۹۵	۲.۰۵۵	۱.۱۳۷	۳.۶۳۴	۱.۲۱۹	-۰.۳۴۳	۳.۶۲۶	نسبت شارپ (SR)
۹۰.۴۳۱	۳۹.۳۵۷	۱۱.۴۲۵	۳۰.۳۷۸	۲۱۱.۲۷۸	۵۴.۴۲۲	۶۳.۶۳۰	۲۲۲.۵۷۸	۴۷.۲۴۴	۶۸.۱۴۵	۲۱۵.۴۰۲	۴۰.۵۹۹	۸۰.۷۱۱	نرخ بازگشت سرمایه % (ROI)
-۰.۱۱۵	-۱.۱۲%	-۴.۹۶%	-۱.۴۰%	-۷.۹۳%	-۲۲.۴۲%	-۳.۰۱%	-۱۴.۰۶%	-۲۶.۹۱%	-۳.۴۰%	-۱۷.۶۲%	-۳۲.۵۶%	-۳.۱۷%	بیشترین سقوط % (MDD)
۷۳.۳۳۳	۹۷	۷۵	۴۸	۹۷	۷۵	۴۸	۹۷	۷۵	۴۸	۹۷	۷۵	۴۸	کل جفت ها
۷۹.۶۱۱	۷۷.۳۱۹	۵۲	۹۱.۶۶۶	۹۱.۷۵۲	۹۴.۶۶۶	۱۰۰	۸۹.۶۹۰	۸۴.۰۰	۹۷.۹۱۶	۸۵.۵۶۷	۷۶	۹۷.۹۱۶	درصد جفت های سودآور
۷۴.۵۸۵	۸۵.۷۱۴	۵۹.۶۱۵	۹۰.۳۲۲	۷۸.۲۸۵	۶۵.۸۹۱	۷۹.۸۲۰	۷۵	۶۵.۸۳۸	۷۶.۷۶۷	۷۱.۸۷۵	۶۸.۶۶۹	۷۷.۲۲۳	درصد معاملات سودآور
۷۱.۸۶۲	۸۳.۴۲۹	۴۹.۰۶۱	۸۹.۶۴۵	۷۸.۱۱۷	۶۱.۸۴۰	۷۷.۶۷۸	۷۵.۳۷۳	۶۲.۱۲۱	۷۴.۰۲۴	۷۲.۲۷۲	۶۴.۳۳۵	۷۴.۴۵۷	متوسط سودآوری هر جفت %

Investigating the Profitability of Pairs Trading using Machine Learning and Genetic based Algorithms in Tehran Stock Exchange

Behnaz Rezaei¹, Reza Baradaran Kazemzadeh², Mohammad Ali Rastegar³

1 -Master student, industrial and systems engineering department, TMU, email: behnaz_rezaei@modares.ac.ir

2 - Professor, industrial and systems engineering department, TMU, email: rkazem@modares.ac.ir

3 -Assistant Professor, industrial and systems engineering department, TMU, email: ma_rastegar@modares.ac.ir

Abstract

In this paper, the problem of finding profitable pairs by automatically limiting search space of pairs using machine learning techniques and integrating an unsupervised learning algorithm, OPTICS, to pair identification and selection in pair trading is discussed. In addition, in order to optimize the portfolio consisting of pairs of assets and allocate optimal capital to them, a genetic-based algorithm with the aim of increasing the sharpe ratio is used. The proposed technique for automatic clustering is better than the conventional methods of searching for pairs of assets used by investors and leads to higher average rate of return on investment and a higher sharpe ratio for portfolio in trading using selected pairs of clusters. These calculated evaluation criteria for portfolio were improved after using bi-objective optimization genetic algorithm. This study was simulated using intraday price data of a group of stocks in Tehran Stock Exchange between the years 2015 up to 2020 and taking into account the transaction costs.

Keywords: Pairs Trading, Neutral Market, Machine Learning, Unsupervised Learning, Optimization Genetic Algorithm, Transaction Costs.