مهندسی صنایع و مدیریت شریف (تابستان ۱۴۰۳)
دوره‌ی ۴۰، شماره‌ی ۱، ص. ۴۳-۵۴، (پژوهشی)

Original Article

بررسی سودآوری معاملات زوجی با استفاده از الگوریتم مبتنی بر یادگیری‌ماشین و الگوریتم ژنتیک در بورس اوراق بهادار تهران

بهناز رضایی (دانشجوی کارشناسی ارشد)
رضا برادران کاظم‌زاده * (استاد)
محمد علی رستگار (استادیار)
دانشکده‌ی مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس

در این مقاله، به مسئله‌ی چگونگی یافتن زوج های سودآور با اعمال محدودیت خودکار در فضای جستجوی زوج دارایی‌ها با استفاده از تکنیک‌های یادگیری ماشین و ادغام یک الگوریتم یادگیری بدون نظارت، OPTICS، به فرایند شناسایی و انتخاب زوج‌ها در معاملات زوجی پرداخته شده است. همچنین، جهت بهینه‌سازی سبد متشکل از زوجدارایی‌ها 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 به‌روش‌های متداول جستجوی زوج دارایی‌ها توسط سرمایه‌گذاران بهتر بوده و منجر به دستیابی به میانگین نرخ بازگشت سرمایه و نسبت شارپ بالاتری برای سبد در معاملات با استفاده از زوج های منتخب از خوشه‌ها شده است. این معیارهای ارزیابی محاسبه شده برای سبد، بعد از به‌کارگیری الگوریتم بهینه‌سازی ژنتیک دوهدفه ارتقا یافته‌اند. 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 شده است.

واژگان کلیدی: معاملات زوجی، یادگیری ماشین، یادگیری بدون نظارت، الگوریتم بهینه سازی ژنتیک، هزینه‌های معاملاتی.

behnaz_rezaei@modares.ac.ir
rkazem@modares.ac.ir
ma_rastegar@modares.ac.ir

۱. مقدمه

معاملات الگوریتمی استفاده از قدرت کامپیوتر جهت اتخاذ موقعیت‌های معاملاتی است که نحوه‌ی اجرای آن با استفاده از الگوریتم‌ها کنترل می‌شود. به‌عبارت دیگر معاملات الگوریتمی خرید و فروش اوراق بهادار براساس 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. معاملات زوجی یکی از انواع معاملات الگوریتمی شناخته شده در دهه ۱۹۸۰، است که از رایج‌ترین استراتژی‌های آربیتراژ آماری و بازار خنثی بشمار می‌رود. این استراتژی که موضوع اساسی در این تحقیق است به‌عنوان یکی از مهمترین ابزارهای سرمایه‌گذاری، توسط صندوق‌ها و موسسات سرمایه‌گذاری و معامله‌گران استفاده شده است.[1] معاملات زوجی یک استراتژی معاملاتی بر مبنای استفاده از الگوی اسپرد و سودآوری از طریق ناکارایی بازار بر یک زوج سهام است. اسپرد، معمولا برابر با تفاضل یا تفاضل لگاریتم زوجدارایی‌ها 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، به‌عنوان 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 به‌شیوه ای 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 براساس مدل معاملاتی، مورد استفاده می‌شود. فرض بر این است که اگر دو سری قیمت از

* نویسنده مسئول
تاریخ: دریافت ۱۴۰۱/۶/۱۲، اصلاحیه ۱۴۰۲/۲/۱۸، پذیرش ۱۴۰۲/۶/۶.

استناد به این مقاله:
رضایی، بهناز، برادران کاظم‌زاده، رضا، و رستگار، محمدعلی، ۱۴۰۳. بررسی سودآوری معاملات زوجی با استفاده از الگوریتم مبتنی بر یادگیری‌ماشین و الگوریتم ژنتیک در بورس اوراق بهادار تهران. مهندسی صنایع و مدیریت شریف، ۴۰(۱)، صص. ۴۳-۵۴. DOI:10.24200/j65.2023.60603.2308

دارایی‌ها در گذشته و در بلندمدت رفتار مشابهی داشته باشند، در آینده نیز انتظار پایداری این رفتار می‌رود. بنابراین، فرض می‌شود که اسپرد از یک الگوی بازگشت به میانگین پیروی می‌کند. به‌عبارت دیگر در این الگو، روند با هر انحراف از تعادل طولانی مدت تصحیح شده و به تعادل خود باز می‌گردد و یا اینکه روند موجود دچار تغییر شده و یک تعادل جدید برای اسپرد ایجاد می‌شود. به بیان 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 فرصت‌هایی، فاصله‌ی سری قیمت‌های زوج های معاملاتی (اسپرد) به‌طور مداوم نظارت می‌شود و هنگامی‌که 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 به‌ارزش نسبی آنها متکی است. یکی از چالش‌ها در این موضوع، غلبه بر فرایند 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 براساس قیمت‌گذاری نسبی، این مسئله حل می‌شود. همچنین، 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. معمولا سرمایه‌گذاران روی یک زوج معامله نمی‌کنند و سبدی 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 دقت‌عملکرد 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 به‌طور گسترده در زمینه‌های مختلف از جمله مالی استفاده شده است. اکثر کاربردها، مبتنی بر یادگیری با نظارت بوده و در مقابل، یادگیری بدون نظارت به‌ندرت در ادبیات مالی مورد استفاده 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 لزوما از صنعت یکسان) صدر 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 به‌دنبال آن اعمال الگوریتم بدون نظارت خوشه‌بندی OPTICS [1] روی داده‌های مربوط به دارایی‌ها به‌منظور خوشه‌بندی خودکار و ایجاد بستری برای استخراج ترکیب‌های زوجی (دوتایی) سودآورو در مرحله‌ی انجام معاملات سبد زوج های منتخب از خوشه‌ها نیز، با توجه به رویکرد تخصیص سرمایه و وزن یکسان درنظر گرفته شده برای زوج های 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 خوشه بند 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 زوج های 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 زوج های منتخب در روش‌های مختلف خوشه‌بندی (خوشه‌بندی با OPTICS، خوشه‌بندی براساس 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 زوج های منتخب از روش خوشه‌بندی برتر) بدست می‌آید. روش خوشه‌بندی برتر براساس بخش عملکرد معاملاتی و مقایسه سه روش خوشه‌بندی مشخص می‌شود و قابل ذکر است تعداد خوشه‌ها در روش خوشه‌بندی براساس صنعت، برابر با تعداد صنایع استفاده شده در مقاله است. بخش پنجم شامل نتیجه‌گیری و ارائه پیشنهادات آتی برای تحقیق است.

۲. ادبیات و پیشینه پژوهش

به‌طورکلی معاملات زوجی در دو حیطه‌ی «شناسایی و تشکیل زوج‌ها» و «مدل‌سازی استراتژی معاملاتی زوج‌ها» مورد بررسی قرار می‌گیرد. مرحله‌ی انتخاب زوج‌ها برای معاملات زوجی شامل، یافتن زوج های 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،[۳و۲] یا مرتب کردن آنها در گروه‌های معنادار، معمولاً بر اساس صنعت و محدودسازی ترکیب‌ها به زوج هایی که از دارایی‌های موجود در یک صنعت تشکیل می‌شوند.[۴و۵] در شرایطی که رویکرد نخست، ممکن است زوج های غیرمعمول‌تری در معاملات پیدا کند و باعث سودآوری بالاتری شود اما رویکرد دوم، با توجه به در نظرگرفتن 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 زوج های مورد معامله به زوج های 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 به‌حداقل می‌رساند. این روش به‌طور گسترده مورد استفاده قرار می‌گیرد، اما طبق نظر[۳] بهینه نیست. برای اثبات این عبارت، این‌گونه فرض می‌شود که $p_{i,t}$ و $p_{j,t}$ قیمت‌های نرمال شده دو دارایی i و j هستند که یک زوج را تشکیل می‌دهند. علاوه بر این، $S^{۲}_{pi-pj}$ واریانس اسپرد طبق ۱ است.

$$S^{۲}_{Pi-Pj} = \frac{۱}{T}\sum_{t=۱}^{T}(p_{i,t}-p_{j,t})^{۲} - (\frac{۱}{T}\sum_{t=۱}^{T}(p_{i,t}-p_{j,t}))^{۲} \qquad (۱)$$

میانگین مجموع مربعات فواصل، $SSd\ p_i, p_j$ در دوره تشکیل زوج‌ها نیز برابر است با:

$$\overline{ssd}_{Pi,Pj} = \frac{۱}{T}\sum_{t=۱}^{T}(p_{i,t}-p_{j,t})^{۲} = S^{۲}_{Pi-Pj} + (\frac{۱}{T}\sum_{t=۱}^{T}(p_{i,t}-p_{j,t}))^{۲} \qquad (۲)$$

یک زوج مطلوب مطابق معیار SSd می‌تواند زوجی باشد که معادله ۲ را به‌حداقل 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 زوج هایی با واریانس اسپرد کم و پتانسیل سود محدود ایجاد کرده و فرصت‌های معاملاتی سودآور را فراهم نمی‌کند. استفاده از همبستگی پیرسون به‌عنوان یک معیار انتخاب توسط[۸]، تحلیل شده است. نویسندگان کاربرد آن‌را در سری‌های بازگشتی با همان نمونه داده مورد استفاده توسط[۷] بررسی کرده و دریافتند که همبستگی، عملکرد بهتری را نسبت به‌روش 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 به‌صورت نظری توضیح داد. در نهایت، رویکرد هم‌انباشتگی مستلزم انتخاب زوج‌هایی است که دو جزء Y_t و X_t با یکدیگر هم‌انباشته هستند، در این‌صورت طبق تعریف، سری حاصل از ترکیب خطی زیر است.

$$S_t = Y_t - \beta X_t \qquad (۳)$$

β به‌عنوان 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 دریافتند که رویکرد هم‌انباشتگی به‌طور 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 روش های شناسایی زوج، عملکرد بهتری داشته است. قابل ذکر است که استفاده از ضرایب استاتیک و ثابت به‌عنوان 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 دارایی‌ها با حرکت و رفتار یکسان، یک گام حیاتی در معاملات زوجی است. جهت سودآوری در این زمینه لازم است روش‌هایی مورد بررسی قرار گیرد 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.[۱۲و۱۳] مرحله‌ی بعد از انتخاب زوج‌های مناسب، شامل اخذ مواضع معاملاتی با در نظرگرفتن یک مدل معاملاتی است. طبق نخستین چارچوب پیشنهاد شده توسط،[۷] معیار اتخاذ موضع 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 (به‌عنوان ضریب) انحراف استاندارد تاریخی آن فاصله باشد، معامله باز شده و موضع 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 به‌عبارت دیگر موضع معاملاتی عکس جهت سودآوری اتخاذ می‌شود. در این مدل، میانگین و انحراف استاندارد مربوط به اسپرد (S_t) در طول دوره شکل‌گیری زوج‌ها به‌صورت (σ_S, μ_S) محاسبه شده و بر اساس آنها، آستانه‌های مدل شامل آستانه خرید (long)، α_L آستانه فروش $(short)$، α_S و آستانه خروج (α_{Exit}) تعریف می‌شوند. سپس روند اسپرد در دورهی معاملاتی کنترل می‌شود و درصورت عبور از α_L، موضع خرید روی اسپرد با خرید Y و فروش X و در صورت عبور از α_S، موضع 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 موضع انجام می‌گیرد. مدل معرفی شده به‌طور 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 فرایند آماری، برنامه‌ریزی‌ها و مدل‌های ریاضی، الگوریتم‌های ابتکاری و فراابتکاری و یادگیری ماشین وغیره مورد مطالعه قرارگرفته‌اند و نتایج خوبی در این زمینه‌ها گزارش شده است و به‌طور خاص، نتایج به‌دست آمده با رویکردهای یادگیری ماشین، بسیار امیدوارکننده بوده است. عمدتا تکنیک‌های یادگیری ماشین در زمینه‌ی معاملات زوجی برای پیش‌بینی و مدل‌سازی اسپرد و به‌طور کلی برای مدل‌سازی استراتژی معاملاتی استفاده شده است. نویسندگان مقالات،[۱۴-۱۶] کاربرد شبکه‌های عصبی مصنوعی (ANN) را برای پیش‌بینی تغییرات اسپرد بررسی می‌کنند.[۱۷] یک سیستم آربیتراژ آماری تجربی، مبتنی بر مدل‌های ناهمگونی خودبازگشتی مشروط تعمیم یافته شبکه عصبی (GARCH) برای مدل‌سازی مکانیسم تصحیح قیمت‌گذاری نادرست بین قیمت‌های نسبی دارایی‌های یک زوج، پیشنهاد می‌کند. رویکرد مورد استفاده در مقالات[۱۸و۱۹] 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 تقویت شده و جنگل‌های تصادفی در زمینه آربیتراژ آماری با

استفاده از سهام $S\&P\ 500$ تحلیل شده است. مقالات[۲۰و۲۱] 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 رضایت بخشی 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 به‌طور گسترده در معاملات زوجی مورد استفاده قرار نگرفته و نتایج به‌دست آمده 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 زوج های 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 بااستفاده 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 داده محور یادگیری ماشین جهت شناسایی زوج‌های سودآور با رویکرد تقسیم‌بندی فضای جستجوی دارایی‌ها به‌طور خودکار و به‌دور 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، به‌هایپر پارامترهای 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 افرازبندی، DBSCAN مبتنی بر چگالی و Agglomerative مبتنی بر سلسه مراتب در خوشه‌بندی داده‌ها جهت معامله زوجی پرداخته و نسبت شارپ بالاتری درصورت استفاده از الگوریتم خوشه‌بندی تجمعی (سلسله مراتبی) اعلام شده است. سه الگوریتم ذکر شده نیاز به تعیین هایپر پارامترهایی از جانب سرمایه‌گذار همانند «تعداد خوشه‌ها» و غیره 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 دارد. تشکیل یک سبد بازار خنثی، ساخت سبدی از دارایی‌هاست 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 داشته باشند[۲۵و۲۶] و بهینه سازی 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 بنابراین از روش یادگیری آماری برای یافتن زوج مناسب در سبد در هر لحظه معاملاتی مبتنی بر پیش بینی نوسانات با استفاده از روش میانگین متحرک نمایی یا GARCH استفاده شده است. سیستم هوشمند ارائه شده در،[۱۲] به بهینه سازی ترکیبات زوجی با در نظرگرفتن دو هدف متعارض بازده و ریسک به‌ترتیب با مفاهیم «واریانس اسپرد» و «رفتار بازگشت به میانگین» با استفاده از فرمول بندی برنامه‌ریزی عدد صحیح و الگوریتم چند هدفه‌ی ژنتیک پرداخته و مجموعه‌ای از فرصت‌های معاملات زوجی چند متغیره سودآور فراهم شده‌است. 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 به‌منظور بهینه‌سازی بازده دوره مورد نظر با متغیرهای تصمیم تخصیصات معامله و پارامترهای سیگنال به‌کارگرفته شده است. با تکامل و توسعه الگوریتم ژنتیک برای حل مسائل بهینه سازی چند هدفه (مانند الگوریتم ژنتیک مرتب سازی غیر مغلوب (NSGA۲)،[۲] امکان تخصیص سرمایه بهینه بین زوج ها 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 نظرگرفتن 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 به‌عنوان مسئله مقایسه‌های چندگانه یاد می‌شود و احتمالا باعث ارتکاب حداقل یک خطای نوع I (α) در هنگام انجام مقایسات مستقل (با مقدار α به‌عنوان سطح اطمینان) یا نرخ خطای خانوادگی می‌شود که با افزایش دارایی‌ها که به‌طور 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 زوج هایی 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 به‌دنبال این موارد است: پیش تقسیم‌بندی مؤثر و خودکار اجتماع دارایی‌ها که علاوه بر عدم محدودسازی نتایج بررسی ترکیب دارایی‌ها به زوج های 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. به‌عبارت 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 دارایی ها بر اساس مجموعه ای از معیارها است.

۱.۳. 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 به‌عنوان نماینده یک عامل ریسک در نظرگرفت و برای استخراج عوامل ریسک اساسی مشترک،

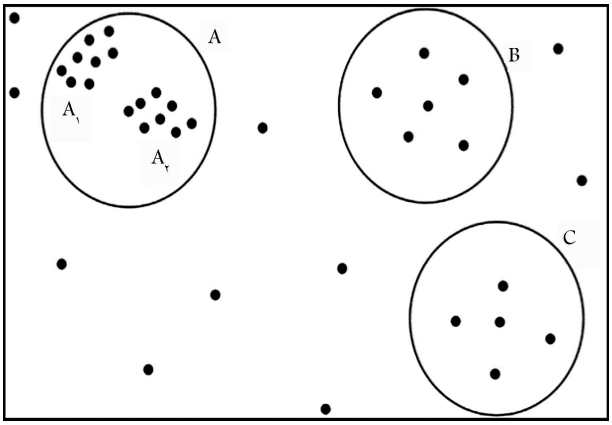
استفاده از PCA را در سری‌های بازده نرمالایز شده پیشنهاد می‌شود که به‌صورت زیر تعریف شده است:

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

که در آن $P_{i,t}$ سری قیمت یک دارایی i است. استفاده از سری قیمت ممکن است به‌دلیل روندهای زمانی زمینه‌ای مناسب نباشد. تعداد مولفه‌های اصلی مورد استفاده، تعداد ویژگی‌ها را برای هر نمایش دارایی مشخص می‌کند. با توجه به‌اینکه 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 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 به‌کار 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 نظرگرفتن این موضوع، کران بالا برای تعداد ابعاد PCA، این مقدار در نظر گرفته شده و به‌صورت تجربی انتخاب می‌شود.

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 گروه بندی 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 گرو بندی همه دارایی‌ها ایجاد نمی‌شود. علاوه بر این، انتساب در الگوریتم موردنظر باید سختگیرانه باشد، در غیر این صورت منجر به افزایش تعداد ترکیب‌های دوتایی ممکن می‌شود که با هدف اولیه در تضاد است. در نهایت، به‌دلیل 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 به‌نظر می‌رسد زیرا به تشکیل خوشه‌هایی با اشکال دلخواه پرداخته و بنابراین، نیاز به اتخاذ فرضیاتی برای نرمال بودن نیست و به‌دلیل عدم گروه‌بندی همه نقاط مجموعه داده، به‌طور 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 [۳] تأثیرگذارترین الگوریتم در این دسته است. به‌طور 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 به‌یکدیگر نزدیک باشند تا به‌عنوان نقاط مجاور تشکیل‌دهنده یک خوشه در نظرگرفته شوند و MinPts حداقل تعداد نقاط برای تشکیل یک خوشه است. با استفاده از این دو پارامتر، درکنار 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، به‌عنوان نقاط پرت طبقه‌بندی می‌شوند. قابل ذکر است الگوریتم DBSCAN با این فرض مناسب است که خوشه‌ها به‌طور



شکل ۱. خوشه ها با چگالی متغیر.

یکنواخت متراکم هستند و نشانگر ضعف این الگوریتم است. به‌عبارت دیگر، اگر نواحی در فضا چگالی‌های متفاوتی داشته باشند، یک ε ثابت ممکن است به‌خوبی 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 $A۱$ و $A۲$ 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 فرایند خوشه‌بندی با پارامتر MinPts 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.

۳.۳. معیارهای انتخاب زوج

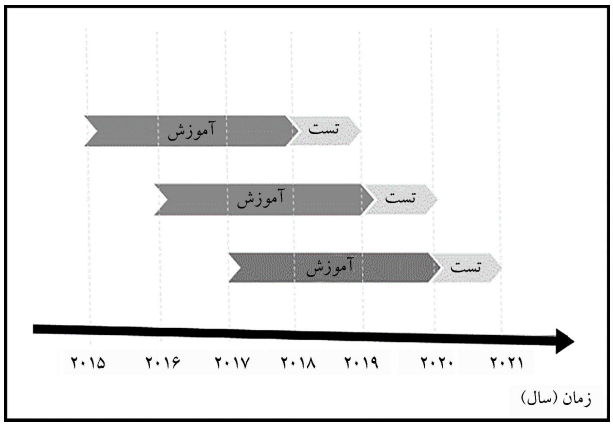
پس از ایجاد خوشه های دارایی‌ها و محاسبه‌ی همه ترکیبات ممکن در خوشه‌ها، لازم است مجموعه‌ای از شرایط برای شناسایی و انتخاب زوج ها و به‌کارگیری آنها در معامله تعریف شود. رویکردهای شناسایی مبتنی بر هم‌انباشتگی، عملکرد خوبی را در انتخاب زوج ها به‌دست می‌آورند زیرا این از نظر اقتصادسنجی، روابط تعادلی قابل اطمینان‌تری در زوج ها 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 به‌منظور آزمون این شرط، استفاده از آزمون انگل ـ گرنجر (Engle-Granger) پیشنهاد می‌شود. آرمسترانگ[۳۱] به‌عنوان یکی از منتقدان به آزمون انگل ـ گرنجر اشاره می‌کند که انتخاب متغیر وابسته ممکن است به‌نتایج متفاوتی منجر شود. برای رفع این موضوع، پیشنهاد می‌شود که آزمون انگل ـ 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 زوج ها 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. به‌این منظور و طبق نمای کلی طرح تحقیق، تکنیک پیشنهادی شامل اعمال خودکار فضاهای جستجوی ترکیبات ممکن با استفاده از الگوریتم خوشه‌بندی OPTICS به‌عنوان منابع استخراج زوج دارایی‌های سودآور، با دو حالت جستجوی ترکیبات با در نظرگرفتن فرض عدم خوشه‌بندی داده‌ها و خوشه‌بندی بر اساس صنعت دارایی‌ها مقایسه می‌شود. به‌عبارت دیگر، زوج‌های منتخب از ترکیبات هر سه فضای جستجو مورد معامله قرارگرفته و نتایج مورد ارزیابی و مقایسه قرار می‌گیرد.

۴.۳. شبیه‌سازی

جهت انجام معاملات به این نکته توجه می‌شود که از آنجایی که مقایسه نتایج حاصل از تکنیک‌های خوشه‌بندی برای انتخاب زوج‌های سودآور نسبت به‌یکدیگر به‌عنوان 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 سهم‌هایی متعلق به صنایع مختلفی از جمله مخابرات، بیمه و صندوق بازنشستگی، ماشین آلات و دستگاه‌های برقی، محصولات شیمیایی، ماشین‌آلات و تجهیزات، مواد و محصولات دارویی، انبوه‌سازی، املاک و مستغلات، حمل‌ونقل، انبارداری و ارتباطات، استخراج نفت گاز و خدمات جنبی جز اکتشاف، 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 غیره 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.

سوالی که در ادامه فرایند پیش می‌آید این است که سرمایه برای هر زوج در سبد

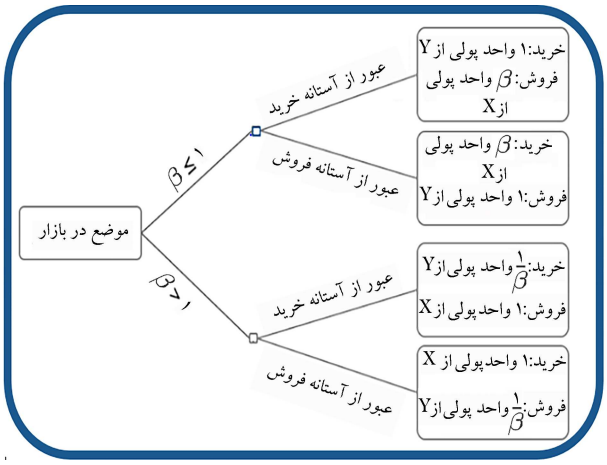


شکل ۲. دوره‌های زمانی آموزش و تست تحقیق.

چگونه تخصیص می‌یابد. در تئوری، اگر ارزش یکسانی در مواضع خرید و فروش دارایی‌های یک زوج در نظرگرفته شود، هزینه ورود به موقعیت خرید توسط سرمایه به‌دست آمده از موقعیت فروش بطورکامل پوشش داده می‌شود 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 نظرگرفته می‌شوند و برای ساده کردن 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 می‌تواند کسری از دارایی‌های معاملاتی را خریداری کند، زیرا خود دارایی‌ها به احتمال زیاد 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 $leg۱$ و $leg۲$ 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.

$$\max(leg۱, leg۲) = ۱ \qquad (۵)$$

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 درنظرگرفته می‌شود. هزینه‌های معاملاتی نیز برای تمام نتایج ارائه شده در این‌کار محاسبه می‌شود که عبارتند از: هزینه‌ی استقراض (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 ـ 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 کارانه‌ی یک دوره‌ای (۱۵ دقیقه) برای ورود به یک موضع در نظر



شکل ۳. چارچوب موضع‌گیری معاملاتی بازار.

جدول ۱. رویکردهای خوشه‌بندی.

	۲۰۱۵ – ۲۰۱۸	۲۰۱۶ – ۲۰۱۹	۲۰۱۷ – ۲۰۲۰
رویکرد تک خوشه‌ای			
تعداد خوشه‌ها	۱	۱	۱
تعداد ترکیبات دوتایی ممکن	۴۹۵۰	۴۹۵۰	۴۹۵۰
زوج های منتخب	۶۵۵	۸۴۷	۲۰۲۳
رویکرد خوشه‌بندی با الگوریتم OPTICS			
تعداد خوشه ها	۱۲	۶	۵
تعداد ترکیبات دوتایی ممکن	۱۸۰	۱۲۸	۱۴۷
زوج های منتخب	۲۳	۲۶	۸۴
رویکرد خوشه‌بندی براساس صنعت دارایی‌ها			
تعداد خوشه ها	۲۴	۲۴	۲۴
تعداد ترکیبات دوتایی ممکن	۲۲۵	۲۲۵	۲۲۵
زوج های منتخب	۴۸	۷۵	۹۷

گرفته می‌شود. سیستم موردنظر فاقد ادراک حد ضرر بوده و یک موضع معاملاتی تنها در صورت مشاهده رفتار همگرایی اسپرد زوج یا پایان دوره معاملاتی بسته می‌شود.

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

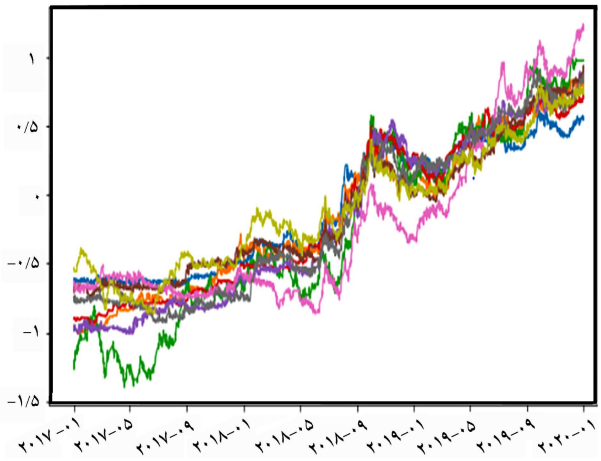
۴. نتایج

۱.۴. عملکرد انتخاب زوج دارایی‌ها

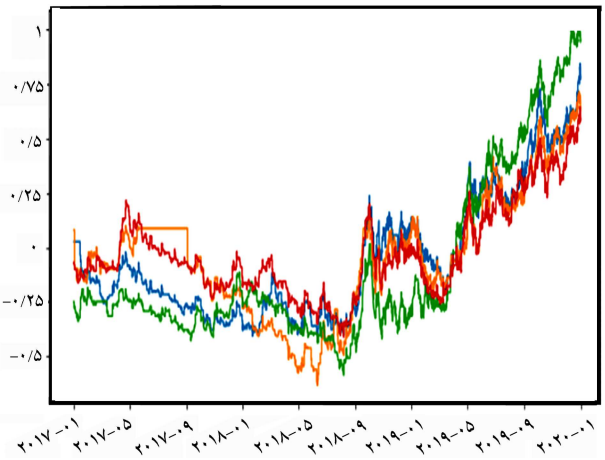
در این بخش به نتایج ارائه شده شامل تعداد زوج های منتخب توسط سیستم تحت سه روش مختلف خوشه‌بندی پرداخته می‌شود. همانطورکه انتظار می‌رود، رویکرد تک‌خوشه‌ای دارای مجموعه بزرگتری از سهم‌ها جهت بررسی ترکیبات ممکن خواهد بود و قاعدتا زوج های بیشتری با استفاده از این تنظیم انتخاب می‌شوند و بررسی میزان سودآوری آنها مسئله‌ی باقیمانده در این مرحله از تحقیق است. با توجه به‌نوع عملکرد مورد انتظار الگوریتم OPTICS، تعداد ترکیبات ممکن استخراج شده از خوشه‌ها می‌تواند به‌طور قابل ملاحظه‌ای کمتر باشد و علاوه برکاهش بار محاسباتی انجام تست‌های آماری متعدد، احتمال یافتن زوج های جعلی نیز کاهش می‌یابد. عملکرد خودکار و اتکای این الگوریتم به داده‌ها، با مشاهده تغییر تعداد خوشه‌ها در دوره‌های مختلف اثبات می‌شود. همانطورکه 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 براساس صنعت دارایی‌ها، کاهش قابل توجهی داشته است. زوج های 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 و ۵ مربوط به آخرین دوره‌ی زمانی آموزش مورد نظر این تحقیق طبق شکل‌های ۴ و 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 به‌یکدیگر 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 چند رشته‌ای صنعتی، عرضه برق، گاز، بخار و آب گرم، فرآورده‌های نفتی کک و سوخت هسته‌ای، فلزات اساسی و استخراج کانه‌های فلزی هستند. چند مورد از سری‌های شناسایی شده در این خوشه مربوط به گروه یکسان و از شرکت‌های چندرشته‌ای 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 همه‌ی این گروه‌ها می‌توانند زیرمجموعه‌ای از گروه بزرگتری مانند وابستگی به فعالیت استخراج از منابع و معادن هستند. این اولین مدرکی است که نشان می‌دهد رویکرد مبتنی بر الگوریتم معرفی شده می‌تواند درگروه‌بندی دارایی‌ها، یک سطح عمیق‌تر پیش رود زیرا قادر است بخش‌های خاص و بزرگتر را نیز شناسایی کند. وضعیت مشابهی نیز در خوش ی شماره ۴ در جدول ۲ مشاهده می‌شود. در این مورد، دارایی‌های شناسایی شده متعلق به گروه‌های محصولات شیمیایی، شرکت‌های چندرشته‌ای صنعتی، فرآورده‌های نفتی کک و سوخت هسته‌ای، فلزات اساسی و استخراج کانه‌های فلزی هستند. طبق شکل ۵، خوشه پنجم از دارایی‌های گروه و صنعت کاملا یکسان تشکیل شده است و دارایی‌های آنها به‌ترتیب متعلق به گروه خودرو و ساخت قطعات هستند. مابقی خوشه‌ها نیز ثابت می‌کنند که قابلیت خوشه‌بندی این الگوریتم فراتر از انتخاب دارایی‌ها از یک صنعت است و یک رابطه قابل مشاهده بین سری‌های قیمتی شناسایی شده وجود دارد. بنابراین، می‌توان تایید کرد که الگوریتم پیشنهادی، به گروه‌بندی دارایی‌ها از یک صنعت می‌پردازد در شرایطی که مانعی برای خوشه‌هایی حاوی دارایی‌ها از بخش‌های مختلف وجود ندارد.

جدول ۲. خوشه‌بندی با الگوریتم OPTICS دوره زمانی ۲۰۱۷ تا ۲۰۲۰.

شماره خوشه	شماره نماد	نماد	گروه
	۳	بالبر	ماشین آلات و دستگاه‌های برقی
	۱۵	ثشاهد	انبوه سازی، املاک و مستغلات
	۱۸	حتوکا	حمل و نقل، انبارداری و ارتباطات
	۳۷	زمگسا	زراعت و خدمات وابسته
	۳۸	سبهان	سیمان، آهک و گچ
	۳۹	ستران	سیمان، آهک و گچ
۱	۴۱	سیدکو	سیمان، آهک و گچ
	۵۱	غبشهر	محصولات غذایی و آشامیدنی به جز قند و شکر
	۵۲	غمارگ	محصولات غذایی و آشامیدنی به جز قند و شکر
	۹۸	قلرست	قند و شکر
	۶۷	لبوتان	ماشین آلات و تجهیزات
	۷۸	وصنا	سرمایه گذاری‌ها
	۸	پکرمان	لاستیک وپلاستیک
	۱	البرز	بیمه و صندوق بازنشستگی
	۲۷	دالبر	مواد و محصولات دارویی
۲	۳۰	دسبحا	مواد و محصولات دارویی
	۳۲	دکوثر	مواد و محصولات دارویی
	۱۷	جم	محصولات شیمیایی
	۴۳	شبهرن	فرآورده های نفتی، کک و سوخت هسته‌ای
	۵۸	فرآور	فلزات اساسی
	۹۵	مبین	عرضه برق، گاز، بخار و آب گرم
۳	۷۱	وامید	شرکت های چند رشته‌ای صنعتی
	۷۹	وصندوق	شرکت های چند رشته‌ای صنعتی
	۸۵	ونیکی	شرکت های چند رشته‌ای صنعتی
	۶۳	کرماشا	محصولات شیمیایی
	۶۴	کروی	استخراج کانه‌های فلزی
	۸۷	تاپیکو	محصولات شیمیایی
	۴۶	شتران	فرآورده های نفتی، کک و سوخت هسته‌ای
	۹۱	شپنا	فرآورده‌های نفتی، کک و سوخت هسته‌ای
۴	۹۲	فخوز	فلزات اساسی
	۵۹	فملی	فلزات اساسی
	۷۲	وبانک	شرکت های چند رشته‌ای صنعتی
	۸۱	ومعادن	استخراج کانه‌های فلزی
	۲۱	خزامیا	خودرو و ساخت قطعات
	۲۲	خساپا	خودرو و ساخت قطعات
۵	۲۵	خودرو	خودرو و ساخت قطعات
	۹۰	خپارس	خودرو و ساخت قطعات



شکل ۴. خوشه‌ی ۳ ـ دوره زمانی ۲۰۱۷ تا ۲۰۲۰.



شکل ۵. خوشه‌ی ۵ ـ دوره زمانی ۲۰۱۷ تا ۲۰۲۰.

۲.۴. 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 رضایت بخش‌تر باشد اما از سوی دیگر، ممکن است این نتیجه حاصل نشود زیرا به تعدیل و حذف زوج‌هایی که سودآور نبوده‌اند، نپرداخته‌ایم. بنابراین، این انگیزه پدید می‌آید که می‌توان به بهینه‌سازی سبد یا میزان سرمایه‌گذاری در هرکدام 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 رضایت بخش، سودآور است و تا این مرحله می‌توان موفقیت فرایند خوشه‌بندی خودکار توسط الگوریتم پیشنهادی در انتخاب زوج‌هایی با سودآوری پایدار در هریک از محیط‌های آزمایش شده تأیید کرد و به‌طور خاص، این ایده تایید می‌شود که روش مطرح شده برای انتخاب زوج دارایی‌ها، روش قدرتمندی است. نکته‌ی قابل توجه در این روش انتخاب زوج دارایی‌ها 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. همانطورکه در جدول پیوست ۱ مشخص است، معامله در چهار سطح آستانه به‌صورت ضریبی از انحراف استاندارد اسپرد ($0/5\sigma, 0/75\sigma, 1\sigma, 2\sigma$) به‌عنوان 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 مختف به‌دست آمده است. مقادیر متوسط نسبت شارپ و نرخ بازگشت سرمایه در حالت استفاده از الگوریتم OPTICS برای شناسایی زوج‌های معاملاتی به‌ترتیب مقادیر ۴/۷۰ و ۱۳۰/۶۶ است که نسبت به‌حالت تک خوشه‌ای با مقادیر ۳/۴۴ 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 نظرگرفته 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 $0/75\sigma$، مقدار نسبت شارپ با در نظرگرفتن 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 به‌دست آمده است اما در بقیه موارد ارتقاء در مقدار معیار نسبت شارپ و نرخ بازگشت سرمایه و به‌طورکلی 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 صنعت دارایی‌ها 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.

۳.۴. عملکرد سبد بهینه

از آنجایی که در بخش ۴.۳. نحوه‌ی تشکیل سبد متشکل از زوج دارایی‌ها مطرح شد، وزن یکسانی برای تخصیص سرمایهی اولیه به زوج‌ها در ابتدای دوره (یک واحد پولی) در نظرگرفته شد و همین مسئله انگیزه‌ی بهینه‌سازی سبد و وزن زوج‌دارایی‌ها را ایجاد نمود. بعد از مشخص شدن رویکرد خوشه‌بندی مناسب 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 قرارگرفته و اطلاعات پایش شده (سری بازده) ناشی از به‌کارگیری این زوج دارایی‌ها در مدل معاملاتی در سال‌های مختلف آموزشی مربوط به تشکیل زوج (دوره‌ی شناسایی زوج‌ها) به‌عنوان ورودی الگوریتم جهت بهینه‌سازی نسبت شارپ سالانه (دو هدف متناقض: بیشترین بازده و کمترین ریسک یا انحراف معیاربازده سالانه) در نظرگرفته می‌شود تا وزن‌های بهینه برای تخصیص سرمایه اولیه به هر زوج در سبد در ابتدای دوره‌ی معاملاتی (یک سال)

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

۰/۷۵σ	۲۰۱۸	
	سبد با وزن‌های	
	یکسان	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.

۰/۷۵σ	۲۰۱۹	
	سبد با وزن‌های	
	یکسان	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.

۰/۷۵σ	۲۰۲۰	
	سبد با وزن‌های	
	یکسان	بهینه
نسبت شارپ (SR)	۷/۷۱۱	۹/۸۱۱
نرخ بازگشت سرمایه٪ (ROI)	۳۱۴/۷۸۰	۳۴۰/۸۵۵
بیشترین سقوط ٪ (MDD)	-۳۷/۶۸٪	-۳۷/۶۹٪

استخراج شود. الگوریتم مورد استفاده برای حل مسئله‌ی دو هدفه، الگوریتم تکاملی و مبتنی بر ژنتیک (NSGA۲) 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 خوشه بندی 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 به‌طور 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 به‌طور مجزا با در نظرگرفتن $0/75\sigma$ به‌عنوان آستانه‌ی معامله طبق جداول ۳ تا ۵ بدست می‌آید که نسبت به‌حالت سبد با تخصیص وزن یکسان برای زوج‌ها، افزیش داشته‌اند که نشانگر موثر بودن رویکرد بهینه‌سازی سبد از طریق الگوریتم بهینه‌سازی دوهدفه مبتنی بر ژنتیک است.

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. به‌عبارت دیگر ایجاد سبدی 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 دارایی های 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 به‌ترتیب زمینه‌ی افزایش خطا در شناسایی زوج‌های مناسب و کاهش سود در معاملات را فراهم می‌کنند. بنابراین الگوریتمی داده محور جهت پیش تقسیم‌بندی یا خوشه‌بندی مؤثر و خودکار اجتماع دارایی‌ها مورد نیاز است که علاوه بر عدم محدودسازی نتایج بررسی ترکیب دارایی‌ها به زوج های نسبتاً رایج در معاملات‌زوجی، به اعمال فضای گسترده برای جستجوی ترکیبات نیز نپردازد. الگوریتم خوشه‌بند موردنظر با توجه به عملکرد مورد انتظار موثر 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 ـ 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 هرخوشه به‌طور خودکار است. نتایج نشان می‌دهد در صورت تمرکز سرمایه‌گذار برکسب بیشترین بازده در هر دو حالت توجه یا عدمتوجه به ریسک متحمل شده، استراتژی مبتنی بر الگوریتم OPTICS عملکرد بهتری نسبت به‌حالت تک خوشه‌ای و خوشه‌بندی بر اساس صنعت دارد و خوشه بندی با استفاده از الگوریتم معرفی شده منجر به شناسایی زوج های امیدوارکننده‌تر از لحاظ سودآوری در معاملات زوجی باتوجه به شاخص 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 زوج های هم‌وزن بوده است، جهت کسب نتایج بهتر برای معیارهای ارزیابی، به بهینه‌سازی سبد متشکل از زوج دارایی‌ها و اختصاص بهینه سرمایه به زوج‌ها با استفاده از الگوریتم تکامل یافته NSGA۲ 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 به‌طور 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 ۱ و ۵ دقیقه) و یا استفاده از ویژگی‌های دیگری از دارایی‌ها نیز وجود دارد و ضمنا 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 دارایی‌های مختلف در بازارهای مختلف و ترکیبات آنها در معاملات زوجی نیز می‌تواند به توسعه این معاملات کمک کند.

پانوشت‌ها

1. Ordering Points to Identify the Clustering Structure
2. Non-Dominated Sorting Genetic Algorithm
3. Density-Based Spatial Clustering of Applications With Noise

منابع (References)

1. Cavalcante, R., Brasileiro, R., Souza, V., Nobrega, J. and 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. and 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. and 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), pp.689-702. https://doi.org/https://doi.org/10.1016/j.ejor.2016.10.031.
4. Do, B. and Faff, R., 2010. Does simple pairs trading still work? *Financial Analysts Journal*, *66*(4), pp.83-95. https://doi.org/10.2469/faj.v66.n4.1.
5. Dunis, C., Giorgioni, G., Laws, J. and 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. and Lee, J.H., 2010. Statistical arbitrage in the US equities market. *Quantitative Finance*, *10*(7), pp.761-782. https://doi.org/10.1080/14697680903124632.
7. Gatev, E., Goetzmann, W.N. and Rouwenhorst, K.G., 2006. Pairs trading: Performance of a relative-value arbitrage rule. *The Review of Financial Studies*, *19*(3), pp.797-827. https://doi.org/10.1093/rfs/hhj020.
8. Chen, H., Chen, J., Chen, Z. and 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. and Afawubo, K., 2015. Pairs trading and selection methods: Is cointegration superior? *Applied Economics*, *47*(6), pp.599-613. https://doi.org/10.1080/00036846.2014.975417.

10. Rad, H., Low, R.K.Y. and Faff, R., 2016. The profitability of pairs trading strategies: Distance, cointegration and copula methods. *Quantitative Finance*, *16*(10), pp.1541-1558. https://doi.org/10.1080/14697688.2016.1164337.

11. Li, Y. and 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. and Dehghanimohammadabadi, M., 2019. Evolutionary multi-objective optimization for multivariate pairs trading. *Expert Systems with Applications*, *135*, pp.113-128. https://doi.org/https://doi.org/10.1016/j.eswa.2019.05.046.

13. Gupta, K. and Chatterjee, N., 2020. Selecting stock pairs for pairs trading while incorporating lead-lag relationship. *Physica A: Statistical Mechanics and its Applications*, *551*, p.124103. https://doi.org/https://doi.org/10.1016/j.physa.2019.124103.

14. Dunis, C.L., Laws, J. and 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. and 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. and 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), pp.352-375. https://doi.org/10.1080/1351847X.2013.830140.

17. Thomaidis, N., Kondakis, N. and 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), pp. 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. and Horta, N., 2020. Enhancing a Pairs Trading strategy with the application of machine learning. *Expert Systems with Applications*, *158*, p.113490. https://doi.org/https://doi.org/10.1016/j.eswa.2020.113490.

21. Flori, A. and Regoli, D., 2021. Revealing Pairs-trading opportunities with long short-term memory networks. *European Journal of Operational Research*, *295*(2), pp.772-791. https://doi.org/https://doi.org/10.1016/j.ejor.2021.03.009.

22. Kim, T. and 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. and Dai, T.S., 2022. Structural break-aware pairs trading strategy using deep reinforcement learning. *The Journal of Supercomputing*, *78*(3), pp.3843-3882. https://doi.org/10.1007/s11227-021-04013-x.

24. Han, C., He, Z. and Toh, A.J.W., 2023. Pairs Trading via unsupervised learning. *European Journal of Operational Research*, *307*(2), pp.929-947. https://doi.org/https://doi.org/10.1016/j.ejor.2022.09.041.

25. Valle, C.A., Meade, N. and Beasley, J.E., 2014. Market neutral portfolios. *Optimization Letters*, *8*(7), pp.1961-1984. https://doi.org/10.1007/s11590-013-0714-6.

26. Valle, C.A., Meade, N. and Beasley, J.E., 2014. Absolute return portfolios. *Omega*, *45*, pp.20-41. https://doi.org/https://doi.org/10.1016/j.omega.2013.12.003.

27. Lin, T.Y., Chen, C.W.S. and Syu, F.Y., 2021. Multi-asset pair-trading strategy: A statistical learning approach. *The North American Journal of Economics and Finance*, *55*, p.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. and 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

پیوست

جدول پیوست ۱. عملکرد معاملاتی در رویکردهای خوشه‌بندی داده‌ها.

	۰.۵σ			۰.۷۵σ			۱σ			۲σ			متوسط(Avg)
	۲۰۱۸	۲۰۱۹	۲۰۲۰	۲۰۱۸	۲۰۱۹	۲۰۲۰	۲۰۱۸	۲۰۱۹	۲۰۲۰	۲۰۱۸	۲۰۱۹	۲۰۲۰	
رویکرد تک خوشه ای													
نسبت شارپ (SR)	۳.۷۷۰	۰.۱۳۷	۱.۲۷۶	۴.۲۱۹	۱.۸۸۸	۶.۷۵۹	۵.۱۲۳	۴.۰۳۵	۵.۲۲۱	۲۶۱۷	-۰.۶۹۸	۷.۰۳۴	۳.۴۴۸
نرخ بازگشت سرمایه٪ (ROI)	۷۹.۵۹۵	۵۷.۲۵۶	۲۶۹.۷۵۰۴	۷۱.۱۰۱	۶۵.۳۱۱	۲۷۴	۶۸.۲۰۴	۶۹.۷۶۹	۲۵۷.۷۵۷	۲۷.۶۸۰	۱۶.۳۳۱	۴۲.۶۳۳	۱۰۸.۲۸
بیشترین سقوط٪ (MDD)	-۵.۹۴٪	-۲۵.۶۹٪	-۲۱.۰۳٪	-۴.۵۱٪	-۲۱.۱۳٪	-۱۳.۶۷٪	-۲.۷۵٪	-۱۷.۳۸٪	-۸.۹۷٪	-۱.۶۸٪	-۲.۵۵٪	-۰.۸۰٪	-۰.۱۰۵
کل جفت ها	۶۵۵	۸۴۷	۲۰۲۳	۶۵۵	۸۴۷	۲۰۲۳	۶۵۵	۸۴۷	۲۰۲۳	۶۵۵	۸۴۷	۲۰۲۳	۱۱۷۵
درصد جفت های سودآور	۸۸.۰۹۱	۸۱.۸۱۸	۸۷.۴۴۴	۹۰.۰۷۶	۸۹.۰۲۰	۸۸.۷۷۹	۹۰.۰۷۶	۹۳.۹۷۸	۹۰.۶۰۸	۸۵.۸۰۱	۶۵.۲۸۹	۸۴.۳۷۹	۸۶.۲۸۰
درصد معاملات سودآور	۷۶۶۶۲	۷۱.۰۰۲	۷۵.۷۳۸	۷۷.۶۱۶	۷۱.۴۹۴	۷۷.۸۲۵	۷۸.۰۸۹	۷۱.۴۴۳	۷۹.۹۰۱	۸۰.۵۷۷	۶۸.۸۴۸	۸۵.۸۳۵	۷۶.۲۵۲
متوسط سودآوری هر جفت ٪	۷۱.۵۲۳	۶۵.۳۷۳۹	۷۵.۷۷۲	۷۲.۴۷۴	۶۴.۹۱۶	۷۸.۲۵۸	۷۳.۰۹۷	۶۵.۳۰۰	۸۰.۶۰۱	۷۸.۱۰۶	۶۲.۲۱۴	۸۶.۲۸۱	۷۲.۸۲۶
رویکرد خوشه بندی با الگوریتم OPTICS													
نسبت شارپ (SR)	۴.۵۹۰	۱.۱۲۸۷	۱.۹۳۶	۴.۸۶۱	۱.۶۸۸	۷.۷۱۱۶	۴.۷۱۹	۴.۴۰۹	۱۴.۵۴۴	۲.۹۵۴	۰.۷۶۴	۷.۱۳۱	۴.۷۰۳
نرخ بازگشت سرمایه٪ (ROI)	۱۱۱.۱۷۱	۶۴.۹۹۷	۳۰۸.۹۲۳	۱۰۰.۰۹۵	۷۲.۵۶۵	۳۱۴.۷۸۰	۸۷.۵۰۵	۸۶.۳۲۸	۳۰۲.۳۹۰	۳۹.۷۳۶	۲۱.۶۴۹	۵۷.۸۱۷	۱۳۰.۶۶۳
بیشترین سقوط٪ (MDD)	-۵.۳۴٪	-۲۲.۶۷٪	-۲۳.۳۶٪	-۴.۸۹٪	-۱۸.۱۳٪	-۳۷.۶۸٪	-۴.۲۳٪	-۱۲.۴۷٪	-۱۰.۵۵٪	-۱.۷۲٪	-۱.۹۶٪	-۱.۹۳٪	-۰.۱۲۰
کل جفت ها	۲۳	۲۶	۸۴	۲۳	۲۶	۸۴	۲۳	۲۶	۸۴	۲۳	۲۶	۸۴	۴۴.۳۳۳
درصد جفت های سودآور	۸۶.۹۵۶	۹۲.۳۰۷	۸۸.۰۹۵	۸۶.۹۵۶	۹۲.۳۰۷	۹۰.۴۷۶	۹۵.۶۵۲	۹۶.۱۵۳	۹۲.۸۵۷	۸۶.۹۵۶	۸۰.۹۵۲	۸۹.۲۸۵	۸۹.۹۱۳
درصد معاملات سودآور	۸۰.۱۹۸	۶۶۶۶۶	۷۵	۸۱.۹۵۴	۷۰.۱۷۵	۷۶.۷۱۵	۸۰	۷۰.۸۳۳	۸۱.۴۷۰	۸۸.۶۳۶	۸۰	۹۰.۶۲۵	۷۸.۵۲۲
متوسط سودآوری هر جفت ٪	۷۷.۳۸۹	۶۴.۱۷۷	۷۵.۲۴۷	۷۵.۹۲۲	۶۳.۵۲۸	۷۶.۳۳۳	۷۶.۷۵۸	۶۴.۸۲۴	۸۱.۲۶۶	۸۴.۴۲۰	۸۲.۲۸۰	۸۹.۸۷۳	۷۶.۰۰۱۸
رویکرد گروه بندی بر اساس صنعت (تعداد خوشه ها= تعداد صنایع مورد استفاده در مقاله)													
نسبت شارپ (SR)	۳.۶۲۶	۰.۳۴۳	۱.۲۱۹	۳.۶۳۴	۱.۱۳۷	۲.۰۵۵	۴.۱۹۵	۱.۰۲۰۱	۸.۴۳۹	۲.۴۰۱	-۱.۸۷۱	۳.۸۳۵	۲.۵۰۳
نرخ بازگشت سرمایه٪ (ROI)	۸۰.۷۱۱	۴۰.۵۹۹	۲۱۵.۴۰۲	۶۸.۱۴۵	۴۷.۲۴۴	۲۲۲.۵۷۸	۶۳.۶۳۰	۵۴.۴۲۲	۲۱۱.۲۷۸	۳۰.۳۷۸	۱۱.۴۲۵	۳۹.۳۵۷	۹۰.۴۳۱
بیشترین سقوط٪ (MDD)	-۳.۱۷٪	-۳۲.۵۶٪	-۱۷.۶۲٪	-۳.۴۰٪	-۲۶.۹۱٪	-۱۴.۰۶٪	-۳.۰۱٪	-۲۲.۴۲٪	-۷.۹۳٪	-۱.۴۰٪	-۴.۹۶٪	-۱.۱۲٪	-۰.۱۱۵
کل جفت ها	۴۸	۷۵	۹۷	۴۸	۷۵	۹۷	۴۸	۷۵	۹۷	۴۸	۷۵	۹۷	۷۳.۳۳۳
درصد جفت های سودآور	۹۷.۹۱۶	۷۶	۸۵.۵۶۷	۹۷.۹۱۶	۸۴.۰۰	۸۹.۶۹۰	۱۰۰	۹۴.۶۶۶	۹۱.۷۵۲	۹۱.۶۶۶	۵۲	۷۷.۳۱۹	۷۹.۶۱۱
درصد معاملات سودآور	۷۷.۲۲۳	۶۸.۶۶۹	۷۱.۸۷۵	۷۶.۷۶۷	۶۵.۸۳۸	۷۵	۷۹.۸۲۰	۶۵.۸۹۱	۷۸.۲۸۵	۹۰.۳۲۲	۵۹.۶۱۵	۸۵.۷۱۴	۷۴.۵۸۵
متوسط سودآوری هر جفت ٪	۷۴.۴۵۷	۶۴.۳۳۵	۷۲.۲۷۲	۷۴.۰۲۴	۶۲.۱۲۱	۷۵.۳۷۳	۷۷.۶۷۸	۶۱.۸۴۰	۷۸.۱۱۷	۸۹.۶۴۵	۴۹.۰۶۱	۸۳.۴۲۹	۷۱.۸۶۲