Archive of SID

بهینه‌سازی سبد سهام با استفاده از الگوریتم ژنتیک و روش K-means بهبود یافته بر پایه الگوریتم ژنتیک

عادل آذر[۱]، احمدرضا یزدانیان[۲] ، مریم قندهاری[۳]

[۱] استاد دانشکده مدیریت دانشگاه تربیت مدرس
AZARA@modares.ac.ir
[۲] استادیار دانشکده ریاضی دانشگاه سمنان،
Ahmadreza.yazdanian@gmail.com
[۳] دانشجوی دکتری دانشکده مدیریت دانشگاه سمنان، mary_ghandehari@yahoo.ca

چکیده: در این مقاله به بهینه سازی سبد سهام با رویکرد کمینه کردن ریسک در معرض خطر (VaR) و بیشینه کردن بازده پرتفوی به صورت همزمان خواهیم پرداخت. الگوریتم مورد استفاده، الگوریتمی پویا براساس الگوریتم ژنتیک و مفاهیم ارزش در معرض خطر می‌باشد. الگوریتم مذکور متفاوت از الگوریتم ژنتیک کلاسیک است. این الگوریتم با نگه داشتن بهترین کروموزوم در هر نسل و استفاده از آن در جواب نهایی باعث بهبود الگوریتم کلاسیک شده است. از الگوریتم مذکور برای به دست آوردن وزن‌های بهینه در سبد سهام و همچنین بهبود الگوریتم K-means برای طبقه‌بندی داده‌ها استفاده شده است. بنابراین دیدگاهی که در این مقاله وجود دارد، شامل طبقه‌بندی اطلاعات خام سهم‌ها با روش k-means بهبود یافته و سپس استفاده از بهینه-ترین طبقه جهت تعیین سبد بهینه سهام می‌باشد. به همین منظور اطلاعات بورس اوراق بهادار تهران طی سال-های ۱۳۹۱، ۱۳۹۲ و ۱۳۹۳ به صورت روزانه مورد استفاده قرار گرفته است. از آنجا که توزیع داده‌ها نرمال نبوده، از آزمون‌های آمار ناپارامتریک جهت آزمون فرضیات استفاده شده است.

نتایج تحقیقات بیانگر آن است که طبقه‌بندی داده‌ها و سپس اجرای الگوریتم ژنتیک روی طبقه بهینه موجب دستیابی به پرتفوی می‌شود که نسبت به پرتفوهای حاصل از اجرای الگوریتم ژنتیک کلاسیک به تنهایی و یا حتی همراه با الگوریتم k-means ساده جهت طبقه‌بندی، دارای ریسک کمتر و بازدهی بیشتر است.

کلمات کلیدی: الگوریتم ژنتیک، اگوریتم طبقه بندی، روش k-means، روش k-means بهبودیافته، ارزش در معرض خطر.

۱ مقدمه

www.SID.ir

Archive of SID

بهینه‌سازی سبد سهام به مفهوم تعیین نسبت سرمایه‌گذاری در دارایی‌هایی است که می‌توان در سبد سهام نگهداری کرد. این نسبت‌ها باید به گونه‌ای انتخاب شود که سبد انتخابی از نظر معیارهایی که به صورت مستقیم یا غیرمستقیم ترکیبی از پارامترهای بازده مورد انتظار سبد، پراکندگی بازده‌ها و ریسک مالی است، بهتر از هر سبد دیگری باشد. بنابراین انتخاب بهینه سبد سهام، انتخاب ترکیبی بهینه از دارایی‌هایی است که می‌تواند در کنار بیشینه‌سازی نرخ بازده مورد انتظار، ریسک نرخ بازده را به طور همزمان کمینه کند (اوریاخی و لوکاس، 2011).

هدف این پژوهش، انتخاب سبد سهامی از میان سبدهای مختلف سرمایه‌گذاری است، به گونه‌ای که بیشترین بازده و کمترین ریسک را به دنبال داشته باشد. این هدف برای پاسخ به سوالی که سرمایه‌گذاران بسیاری در هنگام ورود به بازار سرمایه با آن مواجه می‌شوند در نظر گرفته شده است. جهت دستیابی به هدف این پژوهش با استفاده از الگوریتم طبقه‌بندی K-means بهبود یافته و همچنین به کمک الگوریتم فراابتکاری ژنتیک به انتخاب سبد سهام و تعیین مرز کارا پرداخته می‌شود.

همانطور که بیان شد معیار سنجش ریسک در این پژوهش ارزش در معرض خطر یا همان VaR است. تفاوت روش VaR با روش مارکوویتز در این است که در روش VaR ریسکی کاهش می‌یابد که منجر به افزایش بازده نگردد. تفاوت دیگر این است که در مدل مارکوویتز توزیع نرخ بازده دارایی‌ها حتماً باید نرمال باشد اما در استفاده از VaR هم می‌توان از توزیع‌های نرمال و هم غیر نرمال استفاده کرد.

الگوریتم K-means

الگوریتم k-means یکی از روشهای مبتنی بر خوشه‌بندی است. این روش در سال ۱۹۶۷ توسط Mc Queen ارائه گردیده است. گام‌های انجام این الگوریتم به شرح زیر می‌باشد.

گام اول: از میان n نقطه که نشانگر کل داده‌ها می‌باشد، k نقطه به عنوان مراکز خوشه‌ها به صورت تصادفی انتخاب می‌شود. مراکز انتخاب شده با عبارت است از $z_1, z_2, \dots, z_k$.

گام دوم: نقاطی که به کلاستر C_i تخصیص میدهیم با استفاده از رابطه زیر مشخص میشوند.

$$\|x_i - z_j\| < \|x_i - z_j\| \quad j = 1,2,\dots,k \quad j \neq i$$

گام سوم: مراکز جدید خوشه ها به صورت زیر محاسبه میشوند.

$$z_i^* = (1/n_i)\sum x_j, \quad i = 1,2,\dots,k \quad x_j \in c_j$$

که n_i تعداد عناصر متعلق به خوشه c_j میباشد.

گام چهارم: اگر به ازای $i = 1,2,\dots,k$ داشته باشیم: $z_i^* = z_i$ آنگاه الگوریتم خاتمه میبابد و خوشه ها مشخص میشود. در غیر این صورت از گام دوم مراحل تکرار می‌شوند.

الگوریتم مطرح شده خلاصه ای از روش خوشه بندی K-Means بود. روش ارائه شده در ادامه مقاله، روش تکامل یافته بر اساس الگوریتم ژنتیک میباشد. [۱۴]

کلاستربندی با استفاده از الگوریتم ژنتیک

در خوشه بندی با استفاده از الگوریتم ژنتیک، معیاری به نام μ به صورت زیر تعریف میشود.

www.SID.ir

Archive of SID

$$\mu(C_1, C_2, \dots, C_k) = \sum\sum \|X_j - Z_i\|$$

وظیفه الگوریتم ژنتیک پیدا کردن مراکز کلاستر مناسب $Z_1, Z_2, \dots, Z_k$ به صورتی است که μ حداقل گردد.
هر رشته دنباله ای از اعداد واقعی است که مراکز خوشه ها را نشان میدهد. در یک فضای N-بعدی طول هر کروموزوم برابر N*K خواهد بود.

$$(N_1, N_2, \dots, N_k)$$

مراکز K کلاستر به صورت تصادفی از داده های موجود انتخاب شده و در یک کروموزوم جای میگیرند. این فرآیند تا جایی ادامه می یابد که p کروموزوم تشکیل دهنده جمعیت ساخته شوند.
به دست آوردن ارزش هر کروموزوم شامل دو مرحله می‌باشد. ابتدا کلاسترها براساس مراکزی که در کروموزوم‌ها قرار گرفته‌اند، شکل می‌گیرند. به این صورت که هر نقطه x_i به یکی از کلاسترهای c_i با مرکز z_i اختصاص داده می‌شود با این شرط که:

$$\|x_i - z_j\| < \|x_i - z_p\|, \quad P = 1,2, \dots, K \quad j \neq p$$

پس از انجام کلاستربندی، مراکز کلاستر موجود در کروموزوم‌ها با میانگین نقاط هر کلاستر جایگزین می‌شود. برای کلاستر c_i مرکز جدید z_i به کمک رابطه زیر به دست می‌آید.

$$z_i^* = 1/n_i \sum x_j, \quad i = 1,2, \dots, k$$

معیار کلاستر بندی (μ) به کمک رابطه زیر محاسبه میشود:

$$\mu = \sum_{i=1}^{k} \mu_i$$

$$\mu_i = \sum \|X_i - Z_i\|$$

تابع *Fitness* هم توسط رابطه زیر بدست می‌آید:

$$F = 1/\mu$$

تابع F زمانی حداکثر میشود که تابع μ حداقل شده باشد. توجه به این نکته ضروری است که حداقل شدن μ هدف الگوریتم است.[۹]
برش (*Crossover*): یک فرآیند احتمالی است که اطلاعاتی را مابین دو کروموزوم والد (*Parent*) برای تولید دو کروموزوم فرزند (*Offspring*) جابجا میکند. در این مقاله از برش تک نقطه ای با نرخ برش ثابت $P_c = 0.8$ استفاده شده است.
جهش (*Mutation*): هر کروموزوم با یک احتمال ثابتی در عملیات جهش شرکت میکند.

www.SID.ir

Archive of SID

۵ فرضیات پژوهش

فرضیه‌های پژوهش به صورت زیر بیان می‌گردد:

فرضیه اول: ریسک بدست آمده با استفاده از الگوریتم طبقه‌بندی و الگوریتم ژنتیک، در پورتفوی سهام حاصل، کمتر از ریسک حاصل از بکارگیری روش الگوریتم ژنتیک به تنهایی است.

فرضیه دوم: بازده بدست آمده توسط الگوریتم طبقه بندی و الگوریتم ژنتیک، در پورتفوی سهام حاصل، بیشتر از بازده حاصل از بکارگیری الگوریتم ژنتیک به تنهایی است.

جامعه آماری این تحقیق کلیه شرکت‌های پذیرفته شده در سازمان بورس و اوراق بهادار تهران می‌باشند. با توجه به اینکه در جامعه آماری مورد نظر شرکت‌ها و صنعت‌های مختلف جای گرفته‌اند، می‌توان جامعه را به طبقه-های مختلف تقسیم کرد به نحوی که هر طبقه دارای صنایع همگنی باشد. نمونه با اعمال یک سری محدویت ها از میان جامعه انتخاب شد:

-کلیه شرکت‌هایی که تا پایان سال ۱۳۹۰ در بورس اوراق بهادار تهران پذیرفته شده و تا پایان سال ۱۳۹۳ همچنان در فهرست بورس، باقی مانده باشند.

-اطلاعات مالی 3 ساله آنها موجود باشد و دسترسی به آن امکان پذیر باشد.

با توجه به این معیارها و همچنین توجه به شرکت هایی که در محدوده این سه سال فعال‌تر بوده اند و نیز توجه به کلیه صنایع، 100 شرکت از میان جامعه مورد بررسی، به عنوان نمونه برگزیده شد.

برای آگاهی از چگونگی توزیع داده ها ابتدا چولگی و کشیدگی داده‌ها آزمون می‌شود. چولگی معیاری از تقارن یا عدم تقارن تابع توزیع می‌باشد. نتایج حاصل از این آزمون برای داده‌های این پژوهش به صورت جدول زیر می باشد:

جدول۱: نتایج حاصل از آزمون چولگی و کشیدگی

شرح	تعداد	میانگین	چولگی	کشیدگی
میانگین بازده	۱۲۴	۰,۰۰۰۵۷	۱,۲	۴,۵
ارزش در معرض خطر	۱۲۴	۴,۳۷	۵,۷	۳۰

با توجه به اطلاعات مندرج در جدول فوق، چولگی متغیر بازده در محدوده نرمال بوده و توزیع آن متقارن است. در صورتی که متغیر ریسک نرمال نبوده و دارای چولگی مثبت است. از لحاظ کشیدگی، متغیر ریسک و بازده، هر دونرمال نبوده و دارای کشیدگی مثبت قابل توجهی است.

با توجه به جدول فوق و مقادیر سطح معناداری برای متغیرها می‌توان گفت که توزیع متغیرها نرمال نمی‌باشد. بنابراین با رسم نمودارهای جعبه‌ای و نمودار شاخه و برگ می‌توان به طور جزئی به نحوه توزیع داده‌ها پی برده و علت نرمال نبودن داده‌ها را مشاهده نمود. با استفاده از روش‌های مذکور و یافتن شرکت‌هایی که دارای داده-های پرت بودند و حذف آنها از نمونه آماری به نمونه نهایی دست یافتیم.

متغیرهای اصلی تحقیق بازده و ارزش در معرض خطر می‌باشد که محاسبات برای هرشرکت عضو نمونه، به طریق زیر صورت گرفته است:

بازده روزانه شرکت‌ها، بر اساس تغییرات قیمت روزانه سهام شرکت‌ها، از طریق فرمول زیر محاسبه شده است.

www.SID.ir

Archive of SID

$$\mathrm{R_{t+1}} = Ln\frac{P_{t+1}}{P_t}$$

R_{t+1}: بازده لگاریتمی امروز، P_{t+1}: قیمت امروز سهم و P_t: قیمت سهم در روز گذشته

به این ترتیب برای هر شرکت بازده روزانه سه سال آن محاسبه شده است و نهایتاً میانگین بازده هرشرکت برآورد شده است.

برای محاسبه ریسک شرکت از معیار ارزش در معرض خطر بهره گرفته می شود. از فرمول زیر برای محاسبه آن استفاده شده است.

$$VaR = N.Z_{\alpha}.\delta\sqrt{T}$$

N: ارزش بازار سهام، Z_{α}: سطح اطمینان، δ: انحراف معیار بازده، T: تعداد روزها.

نتایج حاصل از اجرای الگوریتم

قبل از مشاهده نتایج توضیح مختصری در مورد نوع داده‌های استفاده شده بیان می‌گردد. همان طور که قبلاً بیان شد هدف بررسی یک روش بهبود یافته خوشه‌بندی در بهبود نتایج حاصل از بهینه‌سازی سبد سهام با استفاده از الگوریتم ژنتیک است. لازم به توضیح است که الگوریتم ژنتیک طراحی شده با استفاده از نرم افزار مطلب به گونه ای طراحی شده است که بتواند در هر نسل بهترین جواب را حفظ کند. جواب نهایی، انتخاب بهترین جواب از بین جواب‌های نگهداری شده در هر مرحله است. در این مقاله الگوریتم ژنتیک طراحی شده با استفاده از *K-means Clustering* و *GA-Clustering* اجرا شده است و نتایج با هم مقایسه شده است. نتایج نشان میدهد که الگوریتم خوشه بندی بهبود یافته با استفاده از الگوریتم ژنتیک نتایج بهتری را نسبت به روش *K-means* ارائه میدهد.

نتایج ارائه شده در جدول زیر معیار کلاستر بندی یا μ میباشد که هر چقدر کمتر باشد نتیجه مطلوبتری به همراه خواهد داشت.

GA-Clustering	*K-Means Clustering*
29.6173	*32.1988*

نتیجه‌گیری

در این مقاله به بررسی روش تلفیقی الگوریتم ژنتیک همراه با *k-means* بهبود یافته پرداختیم. روش *k-means* یکی از روش‌های خوشه‌بندی است که باعث می‌شود داده‌های ورودی الگوریتم ژنتیک از کیفیت بهتری برخوردار باشد. نتایج تحقیق نشان می‌دهد که بهینه‌سازی سبد سهام با استفاده از روش‌های خوشه‌بندی نتایج بهتری را نسبت به بهینه‌سازی سبد سهام در حالت عادی دارد. روش خوشه‌بندی هر چه بهتر بتواند به دسته

www.SID.ir

Archive of SID

بندی اطلاعات بپردازد نتیجه بهینه سازی بهتر خواهد بود. نهایتاً این نتیجه حاصل شد که با توجه به پیچیده بودن عمل تصمیم‌گیری در انتخاب سبد سرمایه گذاری، تلفیق یک سری از الگوریتم ها ممکن است نتایج بهتری را حاصل کند، همانطور که در این پژوهش این نتیجه حاصل شد که استفاده از این روش ما را به به هدف خود که افزایش بازده و کاهش ریسک است نزدیک‌تر می‌سازد. بنا به یافته‌ها این روش موجب بهبود اجرای الگوریتم ژنتیک می‌شود.

مراجع

۱. البرزی محمود، الگوریتم ژنتیک، چاپ اول، تهران مؤسسه انتشارات علمی دانشگاه صنعتی شریف؛ *1388* .

۲. پورزرندی ابراهیم، کیخا مینا. بهینه سازی سبد سهام با استفاده از روش ***k-means*** و الگوریتم ژنتیک. مجله مهندسی مالی و مدیریت اوراق بهادار. شماره نوزدهم. تابستان ۱۳۹۳. ص ۱۳۱-۱۵۲.

۳. فروزان محمد رضا، نیرومند محمدرضا. روش‌های نوین بهینه‌سازی. چاپ اول. اصفهان: جهاد دانشگاهی.*1388* .

۴. محمد استخری نازنین. انتخاب سبد سهام از میان سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از مدل بهینه‌سازی الگوریتم ژنتیک. مجله توسعه و سرمایه ۱۳۸۶. شماره *1*صفحات ۷۱-۹۲.

۵. مقصودی روح الله، فصیحی نیما، تشنه لب محمد. ارائه مدلی جهت بهبود الگوریتم خوشه بندی *K-means* برپایه الگوریتم های ژنتیک.

6. Das Nadita. Hedge fund classification using k-means clustering method. 9th International conference on computing in Economics and Finance. 2003. Washington. July 11-13.
7. Hachloufi EL. M. Stocks Portfolio Optimization Using Classification and Genetic Algorithm. Mathematical Siences. 2012. Vol 6: 4673-4684.
8. Nanda S R, Mahaty B & Tiwari M K. Clustering Indian Stock Market data for portfolio management. Expert systems with Applications. 2010. 37.
9. Roudier Felix. portfolio optimization and Genetic Algorithms. zurich. 2007.
10. Sefiane Slimane, Benbouziane Mohamed. portfolio selection using genetic algorithm. 2012. Journal of Applied Finance & Banking. vol 2. no4: 143-154.
11. White Richard. Methods for classification. 1996 Agu 16

www.SID.ir