



## کاوش قوانین پیوستگی کمی در بازار سهام با استفاده از الگوریتم های فراابتکاری چندهدفه مبتنی بر الگوریتم ژنتیک

مصطفی زندیه<sup>۱</sup>  
سیما مردانلو<sup>۲</sup>

تاریخ دریافت: ۹۵/۰۷/۱۵ تاریخ پذیرش: ۹۵/۱۰/۲۶

### چکیده

پیش بینی بازده سهام موضوعی مهم در حوزه مالی است که توجه محققان را برای سالهای بسیاری به خود جلب کرده است. سرمایه گذاران همواره در تلاش برای پیدا کردن راهی برای پیش بینی قیمت سهام و پیدا کردن سهام و زمان مناسب برای خرید و یا فروش هستند. اخیراً، از تکنیک های داده کاوی و تکنیک های هوش مصنوعی در این حوزه استفاده می شود. کشف قوانین پیوستگی یکی از رایج ترین روش های داده کاوی مورد استفاده جهت استخراج دانش از مجموعه داده های بزرگ است. برخی محققین کاوش قوانین پیوستگی را به عنوان یک مسئله چند هدفه بیان کرده اند، که به طور مشترک چند معیار را برای به دست آوردن یک مجموعه با قوانین جالب تر و دقیق تر بهینه سازی می کند. در این پژوهش، یک مدل تکاملی چند هدفه جدید ارائه می دهیم که قابلیت درک، جالب بودن و کارایی را به منظور کاوش مجموعه ای از قوانین پیوستگی کمی از داده های مالی، شامل ۱۰ تا از رایج ترین نشانگرهای تحلیل تکنیکی، حداکثر می کند. برای این منظور، این مدل، دو الگوریتم تکاملی چندهدفه معروف الگوریتم ژنتیک مرتب سازی نامغلوب II و الگوریتم ژنتیک رتبه بندی شده نامغلوب را برای انجام فراگیری تکاملی فواصل ویژگی ها و انتخاب شرایط برای هر قانون گسترش می دهد. علاوه براین، مدل ارائه شده، یک جمعیت خارجی و یک فرایند شروع مجدد را برای مدل تکاملی به منظور ذخیره تمام قوانین نامغلوب یافته شده و بهبود تنوع مجموعه قوانین به دست آمده معرفی می کند. نتایج به دست آمده بر روی داده های سهام در دنیای واقعی، اثربخشی روش ارائه شده را نشان می دهد.

**واژه های کلیدی:** قوانین پیوستگی کمی، داده کاوی، الگوریتم تکاملی چندهدفه، نشانگر های تکنیکی، سیگنال خرید و فروش.

۱- دانشیار گروه مدیریت صنعتی، دانشگاه شهیدبهبشتی، تهران، ایران

۲- کارشناس ارشد مهندسی مالی، دانشگاه رجا، قزوین، ایران (نویسنده مسئول) s\_mardanlu@yahoo.com

## ۱- مقدمه

امروزه سرمایه گذاری در بورس، بخش مهمی از اقتصاد کشور را تشکیل می دهد. به همین دلیل تصمیم گیری صحیح در انتخاب سهام برای سهامداران از اهمیت خاصی برخوردار است تا بتوانند با ریسک متناسبی، بالاترین بازده را از سرمایه گذاری خود کسب کنند. در سالهای اخیر، تحقیقاتی در زمینه استفاده از تکنیک های داده کاوی از جمله قوانین پیوستگی<sup>۱</sup> به عنوان یک روش نوین و ابزار قوی به همراه الگوریتم های فراابتکاری<sup>۲</sup> به منظور ایجاد یک سیستم تولید کننده استراتژی معاملاتی انجام شده است.

کشف قوانین پیوستگی یکی از رایج ترین روش های داده کاوی مورد استفاده جهت استخراج دانش از مجموعه داده های بزرگ است. قوانین پیوستگی، همبستگی بین اقلام موجود در یک مجموعه داده را شناسایی می کند و به عنوان بیانی از نوع  $X \rightarrow Y$  که در آن  $X$  و  $Y$  مجموعه اقلام و  $X \cap Y = \emptyset$  می باشد، تعریف می شود.<sup>[۱]</sup> هنگامی که بازه پیوسته است، قوانین پیوستگی به عنوان قوانین پیوستگی کمی شناخته می شوند که در آن هر قلم جفت فاصله ویژگی است.<sup>[۱۷]</sup>

در سال های اخیر، از الگوریتم های تکاملی، به خصوص الگوریتم های ژنتیک، توسط بسیاری از محققان به منظور کاوش قوانین پیوستگی کمی از مجموعه داده ها با مقادیر کمی استفاده شده است.<sup>[۳]</sup> به تازگی، برخی از محققین کاوش قوانین پیوستگی را به عنوان یک مسئله چند هدفه (به جای تک هدفه) با حذف برخی از محدودیت های روش های کنونی ارائه کرده اند. چندین هدف در فرآیند کاوش قوانین پیوستگی در نظر گرفته شده اند که مجموعه ای با قوانین جالب تر و دقیق تر بدست می آورند. به این ترتیب، ما می توانیم به طور مشترک معیارهایی چون تایید<sup>۲</sup>، اطمینان<sup>۴</sup> و غیره را بهینه کنیم که می توانند درجات مختلفی از تعامل را بسته به مجموعه داده استفاده شده و نوع اطلاعاتی که می تواند از آن بدست آید ارائه دهند. از آنجا که این روش طبیعی چند هدفه دارد، استفاده از الگوریتم های فراابتکاری تکاملی چند هدفه برای بدست آوردن مجموعه ای از جواب ها با معیار های مختلف می تواند یک راه کار خوب در این زمینه ارائه دهد.<sup>[۸،۷]</sup>

در این پژوهش، ما یک مدل تکاملی چند هدفه جدید برای کاوش مجموعه ای از قوانین پیوستگی کمی ارائه می دهیم که سه هدف را حداکثر می کند: قابلیت درک<sup>۵</sup> (C)، جالب بودن<sup>۶</sup> (I) و کارایی<sup>۷</sup> (P). برای انجام این کار، این مدل (با نام QAR-CIP) دو الگوریتم تکاملی چندهدفه معروف الگوریتم ژنتیک مرتب سازی نامغلوب<sup>۸</sup> (NSGA II) و الگوریتم ژنتیک رتبه بندی شده نامغلوب<sup>۹</sup> (NRGA) را به منظور انجام فراگیری تکاملی فواصل ویژگیها و انتخاب شرایط برای هر قانون گسترش می دهد. علاوه بر این، مدل ارائه شده، یک جمعیت خارجی<sup>۱۰</sup> (EP) و یک فرایند شروع مجدد<sup>۱۱</sup> به منظور ذخیره تمام قوانین نامغلوب یافته شده و بهبود تنوع مجموعه قوانین به دست آمده معرفی می کند.

هدف از انجام این پژوهش ایجاد یک سیستم سیگنال دهی خودکار به منظور پشتیبانی از تصمیمات سرمایه گذاران با استفاده از الگوریتم های فرا ابتکاری می باشد. الگوریتم ارائه شده با بررسی نشانگرهای

تکنیکی<sup>۱۲</sup> و کاوش قوانین پیوستگی کمی بین داده ها، سیگنال های معاملاتی تولید می کند تا سرمایه گذار بتواند با استفاده از این سیستم برای تصمیم گیری، سود حاصل از سرمایه گذاری خود را بیشینه کند. این پژوهش به ترتیب به شرح زیر است. بخش ۲ مطالعه مختصری بر مبنای نظری و پیشینه پژوهش دارد. بخش ۳ جزئیات اجرای مدل تکاملی ارائه شده برای کاوش قوانین پیوستگی را شرح می دهد. بخش ۴ نتایج به دست آمده از اجرای مدل بر روی داده های مالی را نشان داده و مورد بحث قرار می دهد. در نهایت بخش ۵ برخی نکات نتیجه گیری شده را نشان می دهد.

## ۲- مبنای نظری و پیشینه پژوهش

### ۲-۱- قوانین پیوستگی کمی

قوانین پیوستگی برای نمایش و شناسایی پیوستگی ها بین ارقام موجود در یک پایگاه داده استفاده می شوند.<sup>[۲۰]</sup> همانطور که قبلاً گفتیم، آنها بیانی از نوع  $X \rightarrow Y$  که در آن  $X$  و  $Y$  مجموعه ارقام و  $X \cap Y = \emptyset$  می باشند. بسیاری از مطالعات پیشین برای کاوش قوانین پیوستگی بر روی مجموعه داده ها با مقادیر دوتایی و یا گسسته تمرکز داشتند، اما داده ها در برنامه های دنیای واقعی معمولاً شامل مقادیر کمی هستند. بنابراین، طراحی الگوریتم های داده کاوی با قابلیت کار با انواع مختلف داده، یک چالش در این زمینه است. یک روش معمول برای به کار بردن بازه های پیوسته در کاوش قوانین پیوستگی، طبقه بندی بازه ویژگی ها به صورت فواصل است. این نوع قوانین به عنوان قوانین پیوستگی کمی شناخته می شوند<sup>[۱۷]</sup>. برای مثال، یک قانون پیوستگی می تواند به صورت زیر باشد:

$$\in [3,4] \text{ تعداد ماشین} \rightarrow \in [3,4] \text{ درآمد}, [30,52] \in \text{سن}$$

تایید و اطمینان رایج ترین معیارها برای ارزیابی قوانین پیوستگی هستند. این معیارها برای قانون  $X \rightarrow Y$  به صورت زیر تعریف می شوند:

$$\text{Support}(X \rightarrow Y) = \frac{\text{SUP}(XY)}{|D|} \quad (1)$$

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{SUP}(XY)}{\text{SUP}(X)} \quad (2)$$

که  $\text{SUP}(XY)$  تعداد الگوهای مجموعه داده است که توسط مقدم و تالی قانون پوشش داده شده است،  $\text{SUP}(X)$  تعداد الگوهای مجموعه داده است که توسط مقدم قانون پوشش داده شده است و  $|D|$  تعداد الگوها در مجموعه داده است.<sup>[۱۶]</sup>

تکنیکهای کلاسیک برای کاوش قوانین پیوستگی سعی در کشف قوانینی دارند که تایید و اطمینان آنها بیشتر از آستانه های حداقل تایید ( $\text{minSup}$ ) و حداقل اطمینان ( $\text{minConf}$ ) پیش فرض کاربر است. هرچند،

برخی مولفین به اشکالاتی از این چارچوب که منجر به پیدا کردن تعداد بیش از حد قوانین می شود اشاره کرده اند.<sup>[۱۵، ۴]</sup> برای مثال، اطمینان قادر به شناسایی استقلال آماری و یا همبستگی منفی بین اقلام نیست زیرا تایید نتیجه را به حساب نمی آورد. علاوه بر این، اقلام با تایید بسیار بالا منبعی از قوانین گمراه کننده هستند زیرا در تراکنش های بسیاری ظاهر می شوند و بنابراین هر مجموعه اقلام (برخلاف معنی آن) یک پیش بینی کننده خوب از وجود مجموعه اقلام با تایید بالا به نظر می رسد. در سال های اخیر، برخی محققین معیارهای دیگری برای انتخاب و رتبه بندی ویژگی ها براساس جالب بودن بالقوه آنها برای کاربر ارائه کرده اند.<sup>[۲۵، ۹، ۱۴]</sup>

## ۲-۲- تحلیل تکنیکی

تحلیل تکنیکی به معنی بررسی نوسانات قیمت سهام در گذشته است که به کمک نمودار و به منظور پیش بینی حرکت آتی بازار انجام می گیرد. در این نوع تحلیل سعی می شود از روند قیمت یک سهم در گذشته، آینده آن پیش بینی شود. اساس این نوع تحلیل ریاضیات می باشد که در نمودارها و نشانگرها خلاصه می شود. یک تحلیلگر تکنیکی معتقد است که همه اطلاعات یک سهم در سابقه قیمت آن نهفته است و از روند حرکت قیمت، می توان آینده آن را پیش بینی نمود. تحلیل تکنیکی سرعت عمل بالایی داشته و به ما این امکان را می دهد که در مدت کوتاه تری نسبت به تحلیل بنیادی، تعداد خیلی بیشتری سهام را مورد مطالعه و بررسی قرار بدهیم.<sup>[۱۵]</sup> تحلیل تکنیکی بر اساس سه پیش فرض بنا شده است:

- ۱) همه چیز در نوسان قیمت منعکس می شود.
- ۲) حرکت قیمت همواره دارای روند است.
- ۳) گذشته همیشه سعی در تکرار خود دارد.

نشانگر تکنیکی مجموعه ای از نقاط اطلاعاتی است که از فرمول بندی داده های قیمت یک سهم حاصل می شود. داده های قیمت، باز شدن، بالاترین قیمت، پایین ترین قیمت و قیمت بسته شدن (قیمت پایانی) می باشد. برخی نشانگرها ممکن است فقط قیمت بسته شدن را به کار ببرند؛ درحالی که برخی دیگر حجم سهام معامله شده و قیمت باز شدن را هم در فرمول وارد می کنند. در واقع فرمول هر نشانگر مانند یک تابع عمل می کند، به این صورت که داده های قیمت، ورودی این تابع و نقاط نمودار نشانگر، خروجی آن می باشد. از وصل کردن نقاط خروجی، نمودار نشانگر به دست می آید به منظور تسهیل انجام تجزیه و تحلیل، نشانگرهای تکنیکی معمولاً به صورت یک نمودار، بالا یا پایین نمودار قیمت نشان داده می شوند. این کار سبب می شود مقایسه نمودار قیمت و نشانگر ساده تر انجام پذیرد.<sup>[۱۵]</sup> نشانگرها سه وظیفه مهم هشدار دادن، تایید کردن و پیش بینی را دارند. در این تحقیق سعی شده است که رایج ترین و موثرترین نشانگرها مورد استفاده قرار گیرند، به گونه ای که با در کنار هم قرار گرفتنشان، یکدیگر را تکمیل و تایید کنند.

## ۲-۳- پیشینه پژوهش

استخراج قوانین پیوستگی کمی برای اولین بار توسط آگراوال<sup>۱۳</sup> و اسریکانت<sup>۱۴</sup> در سال ۱۹۹۴ مطرح شد که در آن الگوریتم قوانین پیوستگی را با قسمت بندی کردن دامنه مشخصه ها و ترکیب قسمت های مجاور مسئله به حالت دوتایی تبدیل می نمود. این کار باعث تاکید بیش از حد و یا نادیده گرفتن داده هایی می شد که در مرز بین فاصله ها قرار داشتند.<sup>[۱]</sup>

در سال ۲۰۰۵، یو پینگ هوانگ<sup>۱۵</sup> و لی یین کاو<sup>۱۶</sup> به منظور حل مشکل فوق الگوریتمی ارائه کردند که در آن ابتدا مشخصه های کمی به چندین مشخصه فازی تبدیل می شد. کاوش قوانین پیوستگی در این رویکرد بر اساس الگوریتم Apriori بود. نتایج نشان دهنده اثر بخشی این روش در معاملات و بهبود نتایج قبلی بود.<sup>[۱۱]</sup>

یو تینگ<sup>۱۷</sup> و همکاران در سال ۲۰۰۷ رویکردی ارائه کردند که در آن داده های کمی بازار بورس را با استفاده از نمودار های آماری و تحلیل تکنیکی به رشته های نمادی تبدیل می کرد تا با کاوش قوانین پیوستگی جهت حرکت قیمت آینده سهام را پیش بینی کند. نتایج نشان داد این روش در پیش بینی جهت حرکت قیمت و استخراج روابط بین سهام موثر می باشد.<sup>[۱۸]</sup>

گاش<sup>۱۸</sup> و نث<sup>۱۹</sup> در سال ۲۰۰۴ برای اولین بار استفاده از چندین هدف را جهت استخراج قوانین پیوستگی مطرح کردند. آن ها معیارهایی از جمله تایید، قابلیت درک و جالب بودن را به صورت اهداف مختلف برای ارزیابی قوانین حاصله به کار بردند. آن ها از الگوریتم ژنتیک مبتنی بر پارتو جهت استخراج قوانین پیوستگی موثر استفاده کردند. در این روش بر خلاف روش های قبلی به تعیین آستانه حدود برای معیار های تایید و اطمینان نیاز نبود. نتایج حاکی از آن بود که این روش برای مجموعه داده های بزرگ بسیار مناسب است.<sup>[۱۰]</sup>

در سال ۲۰۰۷، آلتاس<sup>۲۰</sup> و همکاران، الگوریتم تکاملی تفاضلی چندهدفه با نام MODENAR را به عنوان یک استراتژی کاوش قوانین پیوستگی ارائه دادند. این الگوریتم به منظور جستجو ما بین فاصله های مشخصه های کمی و یافتن قوانین پیوستگی میان آنها طراحی شده بود. نتایج، بهبودی نسبت به نتایج قبلی نشان نمی داد ولی استفاده از این الگوریتم آسان و قابل فهم تر بود.<sup>[۳]</sup>

یان<sup>۲۱</sup> و همکاران در سال ۲۰۰۹ الگوریتمی مبتنی بر الگوریتم ژنتیک تک هدفه با نام ARMGA/EARMGA مطرح نمودند که در این الگوریتم قوانین پیوستگی بدون تعیین حدود آستانه تایید توسط کاربر شناسایی می شوند. نتایج نشان می دهد این الگوریتم دارای کارایی بالایی می باشد و از آن می توان برای داده های جهان واقعی نیز به کار برد.<sup>[۱۹]</sup>

چانگ چین<sup>۲۲</sup> و چن<sup>۲۳</sup> در سال ۲۰۱۰ روش جدیدی برای طبقه بندی ارائه نمودند که بر اساس قوانین پیوستگی بود و می توانست از داده های کمی و روابط بین آنها استفاده کند. در این تحقیق داده های بازار بورس به کار برده شده و توسط الگوریتمی بر مبنای الگوریتم ژنتیک، مدل طبقه بندی کننده ای ایجاد شده

که می تواند با استفاده از نشانگرهای مالی استراتژی معاملاتی را شناسایی کند. نتایج حاکی از آن بود که این رویکرد یک روش طبقه بندی موثر بوده و دارای دقت بالایی می باشد.<sup>[۶]</sup>

مارتین<sup>۲۴</sup> و همکاران در سال ۲۰۱۳ از قوانین پیوستگی با مقادیر کمی استفاده کردند و یک الگوریتم تکاملی چند هدفه جدید با نام QAR-CIP-NSGA II را برای کاوش مجموعه ای از قوانین پیوستگی کمی ارائه کردند که سه معیار قابلیت درک، جالب بودن و کارایی را با تبادل خوبی بین قابلیت تفسیر و دقت، حداکثر می کرد. نتایج به دست آمده بر روی داده ها در دنیای واقعی، اثربخشی روش ارائه شده را نشان می دهد.<sup>[۱۳]</sup>

### ۳- روش شناسی پژوهش

در این پژوهش، ما یک مدل تکاملی چند هدفه جدید برای کاوش مجموعه ای از قوانین پیوستگی کمی ارائه می دهیم که سه هدف را حداکثر می کند: قابلیت درک، جالب بودن و کارایی. این مدل استفاده از دو الگوریتم NSGA II و NPGA در کارایی فراگیری تکاملی در قوانین در نظر می گیرد<sup>[۱۲]</sup> و دو مؤلفه جدید برای مدل تکاملی خود معرفی می کند: یک جمعیت خارجی (EP) و یک فرایند شروع مجدد. در ادامه، جزئیات تمام ویژگی های آنها را شرح خواهیم داد.

#### ۳-۱- مدل چندهدفه تکاملی

در این پژوهش، مدل ارائه شده QAR-CIP، دو الگوریتم معروف NSGA II و NPGA را گسترش و یک جمعیت خارجی (EP) و یک فرایند شروع مجدد برای مدل به ترتیب به منظور ذخیره تمام قوانین نامغلوب یافته شده، ایجاد تنوع در جمعیت و بهبود پوشش مجموعه داده تکاملی ارائه می دهد. EP تمام قوانین نامغلوب یافته شده را ذخیره می کند و در پایان هر نسل با قوانین نامغلوب جمعیت فعلی به روز می شود. قوانین نامغلوب اضافه، به منظور جلوگیری از تداخل (اشتراک) قوانین از EP حذف می شوند. یک قانون وقتی اضافه به حساب می آید که فواصل تمامی متغیرهای آن در فواصل متغیرهای یک قانون دیگر وجود داشته باشد. علاوه بر این، اندازه EP به منظور به دست آوردن تعداد زیادتری از قوانین جبهه پارتو و به منظور کاهش اندازه جمعیت (مستقل از اندازه مسئله)، محدود نشده است که به کنترل بهتر همگرایی روش کمک می کند.

به منظور ایجاد تنوع در جمعیت، هنگامی که تعداد افراد جدید در جمعیت یک نسل کمتر از  $\alpha\%$  اندازه جمعیت فعلی باشد ( $\alpha$  توسط کاربر تعیین می شود و معمولاً  $5\%$  است) از فرایند شروع مجدد استفاده می شود. در این حالت، نمونه های تحت پوشش قوانین در EP علامت گذاری شده و فرایند مقداردهی اولیه جمعیت دوباره از نمونه های پوشش داده نشده اعمال می شود که به ما اجازه می دهد یک بررسی خوب از فضای جستجو انجام دهیم. در نهایت، EP با جمعیت جدید به روز می شود. توجه داشته باشید که هر دو عضو، مکمل یکدیگرند. فرایند شروع مجدد از نمونه های پوشش داده نشده توسط قوانین EP برای تولید

جمعیت جدید استفاده می‌کند. علاوه بر این، EP تمام قوانین نامغلوب یافته شده تا آخرین لحظه را نگه می‌دارد که از حذف قوانین هنگامی که فرایند شروع مجدد کل جمعیت را مجدداً شروع می‌کند، جلوگیری می‌کند.

با این اصلاحات، مدل تکاملی به شرح زیر خواهد بود: نخست، یک جمعیت اولیه تولید می‌کند و EP را با قوانین نامغلوب از جمعیت اولیه مقدار دهی اولیه می‌کند. سپس جمعیت فرزندان از جمعیت فعلی با انتخاب، تقاطع و جهش تولید می‌شود. جمعیت بعدی از جمعیت فعلی و فرزندان ساخته می‌شود، EP با جمعیت فعلی به روز شده و در صورت لزوم، فرآیند شروع مجدد اعمال می‌شود. هنگامی که تعداد افراد جدید در جمعیت بعدی کمتر از  $\alpha\%$  است، فرآیند شروع مجدد اعمال می‌شود. این فرایند تا زمانی که یک شرط توقف ارضا شود، تکرار می‌شود. الگوریتم NSGA-II و NPGA دو ویژگی دارند که آنها را الگوریتم‌های فراابتکاری تکاملی با عملکرد بالا می‌سازد. یکی ارزیابی برازش هر جواب بر اساس رتبه بندی پارتو و یک معیار ازدحام است، و دیگری یک روش به روز رسانی نسل نخبه‌گرا است.

هر جواب در جمعیت فعلی به صورت زیر ارزیابی می‌شود. ابتدا، رتبه ۱ به همه جواب‌های نامغلوب در جمعیت فعلی اختصاص داده می‌شود. تمام جواب‌های با رتبه ۱ موقتاً از جمعیت فعلی حذف می‌شوند. سپس، رتبه ۲ به تمام جواب‌های نامغلوب در جمعیت کاهش یافته فعلی اختصاص داده می‌شود. تمام جواب‌های با رتبه ۲ موقتاً از جمعیت فعلی حذف می‌شوند. این فرایند تا زمانی که تمام جواب‌ها موقتاً از جمعیت فعلی حذف شوند، تکرار می‌شود. در نتیجه، رتبه‌ای متفاوت به هر جواب اختصاص داده می‌شود. جواب‌های با رتبه‌های کوچکتر بهتر از جواب‌های با رتبه‌های بزرگتر محسوب می‌شوند.

در میان جواب‌های با رتبه مشابه، شرط دیگری به نام معیار ازدحام در نظر گرفته می‌شود. معیار ازدحام برای یک جواب، فاصله بین جواب‌های مجاور با رتبه مشابه در فضای هدف را محاسبه می‌کند. جواب‌های با ازدحام کمتر با مقادیر بزرگ‌تر معیار ازدحام، بهتر از جواب‌های با ازدحام بیشتر با مقادیر کوچکتر معیار ازدحام محسوب می‌شوند.

یک جفت از جواب‌های والد از جمعیت فعلی با تورنامنت دوتایی در الگوریتم NSGA-II و چرخ رولت در الگوریتم NPGA بر اساس رتبه بندی پارتو و معیار ازدحام انتخاب می‌شوند. هنگامی که باید جمعیت بعدی ساخته شود، جمعیت فعلی و فرزندان به یک جمعیت ادغام شده ترکیب می‌شوند. هر جواب در جمعیت ادغام شده به همان شیوه‌ای که در مرحله انتخاب جواب‌های والد با استفاده از رتبه بندی پارتو و معیار ازدحام استفاده شد، ارزیابی می‌شوند. جمعیت بعدی با انتخاب یک عدد معین (مثلاً اندازه جمعیت) از بهترین جواب‌های جمعیت ادغام شده ساخته می‌شود.

### ۳-۲- کدگذاری و ژن های اولیه

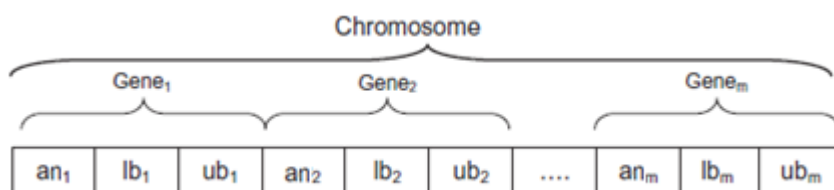
هر کروموزوم برداری از ژن هاست که نشان دهنده ویژگی ها و فواصل قانون است. در اینجا ما از یک کدگذاری مکانی استفاده کرده ایم که در آن ویژگی  $i$  ام استفاده شده در ژن  $i$  ام، کدگذاری شده است. هر ژن شامل سه بخش است:

- بخش اول (an) نشان دهنده این است که یک ژن در قانون نقش دارد یا نه. هنگامی که این بخش صفر است، این ویژگی در قانون نقش ندارد، و هنگامی که این بخش ۱ است، این ویژگی بخشی از مقدم قانون بوده است.
- بخش دوم نشان دهنده کران پایین (lb) فواصل ویژگی هاست.
- بخش سوم نشان دهنده کران بالا (ub) فواصل ویژگی هاست.

در نهایت، یک کروموزوم  $C_T$  به صورت زیر کدگذاری می شود، که در آن  $m$  تعداد ویژگی ها در مجموعه داده است (که در اینجا با توجه به تعداد نشانگرهای تکنیکی استفاده شده، برابر ۱۰ است).

$$Gene_i = (an_i, lb_i, ub_i), \quad i=1, \dots, m$$

$$C_T = Gene_1 Gene_2 \dots Gene_m$$



شکل ۱- نمایی از یک کروموزوم

به منظور اجتناب از افزایش فواصل بیش از اندازه کل بازه، دامنه<sup>۲۵</sup> را به عنوان حداکثر اندازه ای که فاصله یک ویژگی تعیین شده می تواند داشته باشد، تعریف می کنیم. به این ترتیب، دامنه ویژگی  $i$  ام به صورت زیر تعریف می شود:

$$Amplitude_i = (Max_i - Min_i) / \delta \quad (۳)$$

که در آن  $Max_i$  و  $Min_i$  به ترتیب مقادیر حداکثر و حداقل بازه ویژگی  $i$ ،  $\delta$  یک مقدار داده شده توسط متخصص سیستم است که تعیین کننده تعامل بین تعمیم<sup>۲۶</sup> و اختصاصیت<sup>۲۷</sup> از قوانین است. جمعیت اولیه از یک مجموعه قانون با پوشش خوبی از مجموعه داده ها تشکیل می شود. از آنجا که در این تحقیق، بخش تالی این قوانین همان سیگنال های خرید و فروش مورد نظر ماست، ما تنها ویژگی ها را در بخش مقدم قوانین در نظر می گیریم. برای تولید جمعیت اولیه، ابتدا به طور تصادفی ویژگی هایی را که بخشی از مقدم قوانین هستند را انتخاب می کنیم (حداقل یک ویژگی برای مقدم انتخاب خواهد شد).



سپس به طور تصادفی یک الگوی بدون علامت از مجموعه داده را انتخاب کرده و فاصله هر ویژگی را با اندازه ای برابر ۵۰٪ دامنه  $i$  هر ویژگی و با مقادیر الگوی انتخاب شده در مرکز هر کدام از آنها تولید می کنیم. اگر برخی از کرانه های فواصل از بازه ویژگی تجاوز کنند، با کرانه بازه جایگزین خواهد شد. در نهایت، الگوهای پوشش داده شده برای این قانون در مجموعه داده علامت گذاری می شوند. این فرایند تا زمانی که جمعیت اولیه کامل شود، تکرار می شود. توجه داشته باشید که، اگر همه الگوها علامت گذاری شده و جمعیت اولیه کامل نشده باشد، تمام الگوها دوباره بی علامت می شوند و این فرایند تا زمانی که جمعیت اولیه کامل شود، تکرار خواهد شد.

برای مثال، یک مجموعه داده ساده با سه ویژگی  $X_1$ ،  $X_2$  و  $X_3$ ، شش الگو و  $\delta=2$  را در نظر بگیرید. جدول ۱، شش الگو، کران بالا و پایین بازه ویژگی و ۵۰٪ دامنه  $i$  هر ویژگی را نشان می دهد. فرض کنید به طور تصادفی، الگوی ID3 و ویژگی  $X_1$  و  $X_2$  را برای مقدم قانون انتخاب کرده ایم.

جدول ۱) شش الگو در این مثال

| ID              | $X_1$ | $X_2$ | $X_3$ |
|-----------------|-------|-------|-------|
| ID 1            | ۰,۳   | ۲۳    | ۳,۲   |
| ID 2            | ۱,۰   | ۳۸    | ۲۰,۵  |
| ID 3            | ۰,۹   | ۱۲    | ۱۰,۱  |
| ID 4            | ۰,۰   | ۲۰    | ۵۰,۳  |
| ID 5            | ۰,۴   | ۵۰    | ۷۰,۸  |
| ID 6            | ۰,۲   | ۱۰    | ۵,۹   |
| کران پایین بازه | ۰,۰   | ۱۰    | ۳,۲   |
| کران بالای بازه | ۱,۰   | ۵۰    | ۷۰,۸  |
| ۵۰٪ دامنه $i$   | ۰,۲۵  | ۱۰    | ۱۶,۹  |

فواصل به صورت زیر محاسبه می شوند:

$$lb_1 = \max\left\{0.9 - \frac{0.25}{2}, 0.0\right\} = 0.77$$

$$ub_1 = \min\left\{0.9 + \frac{0.25}{2}, 1.0\right\} = 1.0$$

$$lb_2 = \max\left\{12 - \frac{10}{2}, 10\right\} = 10$$

$$ub_2 = \min\left\{12 + \frac{10}{2}, 50\right\} = 17$$

$$lb_3 = \max\left\{10.1 - \frac{16.9}{2}, 3.2\right\} = 3.2$$

$$ub_3 = \min \left\{ 10.1 + \frac{16.9}{2}, 70.8 \right\} = 18.55$$

۵۰٪ دامنه  $z$  برای قرار گرفتن مقدار الگوی ID3 در مرکز فواصل تولید شده، تقسیم بر ۲ می شود. توجه داشته باشید که،  $ub_1$  برابر ۱،۰ است زیرا هنگامی که کرانه فواصل از بازه ویژگی تجاوز کند، با کرانه بالا/پایین بازه جایگزین خواهد شد. شکل ۲ کروموزوم تولید شده در این مثال را نشان می دهد. در این تکرار، قانون زیر تولید می شود:

$$X_1 \in [0.77, 1.0] \text{ و } X_2 \in [10, 17] \rightarrow Y.$$

این قانون، الگوی ID1 و ID3 در جدول ۱ را پوشش می دهد. در این حالت، این الگوها در مجموعه داده علامت گذاری می شوند. EP با قوانین نامغلوب جمعیت اولیه، مقداردهی می شود.

| $X_1$  |      |     | $X_2$  |    |    | $X_3$  |     |       |
|--------|------|-----|--------|----|----|--------|-----|-------|
| Gene 1 |      |     | Gene 2 |    |    | Gene 3 |     |       |
| an     | lb   | ub  | an     | lb | ub | an     | lb  | ub    |
| 1      | 0.77 | 1.0 | 1      | 10 | 17 | 0      | 3.2 | 18.55 |

شکل ۲- کروموزوم تولید شده در این مثال

### ۳-۳- اهداف

سه هدف در این مسئله حداکثر می شوند: قابلیت درک، جالب بودن و کارایی. **قابلیت درک:** این معیار تلاش می کند تا سهولت درک قانون را به صورت کمی اندازه گیری کند. هرچه تعداد ویژگی های نقش دار قانون کمتر باشد، آن قانون برای کاربر قابل فهم تر است. در اینجا، قابلیت درک قانون  $X \rightarrow Y$  به صورت زیر تعریف می شود:

$$\text{Comprehensibility}(X \rightarrow Y) = 1/\text{Attr}_{X \rightarrow Y}$$

که  $\text{Attr}_{X \rightarrow Y}$  تعداد ویژگی های نقش دار در مقدم قانون است.

**جالب بودن:** این معیار میزان جالب بودن قانون را اندازه می گیرد که به ما اجازه می دهد فقط آن دسته از قوانینی را استخراج کنیم که مورد علاقه کاربران خواهد بود. در این جا ما معیار پیشرفت را به عنوان معیار جالب بودن در نظر می گیریم. بنابراین جالب بودن یک قانون به صورت زیر تعریف می شود:

$$\text{Interestingness}(X \rightarrow Y) = \frac{\text{Support}(XY)}{\text{Support}(X)\text{Support}(Y)}$$

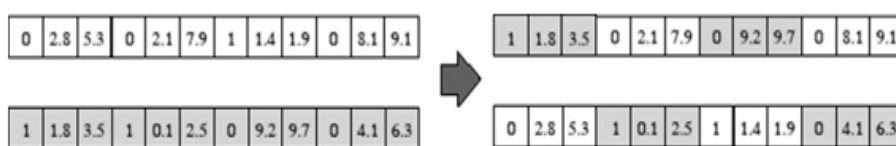
**کارایی:** این معیار حاصل ضرب فاکتور قطعیت (CF) و تایید است که به ما اجازه می دهد قوانینی دقیق و تبادل خوبی بین قوانین محلی و کلی به دست آوریم. توجه داشته باشید که ما تنها به قوانین بسیار قوی که نشان دهنده یک همبستگی مثبت بین اقلام و حل اشکال تایید می باشند علاقه مند هستیم<sup>[۹]</sup>. کارایی قانون  $X \rightarrow Y$  به صورت زیر تعریف می شود:

$$\text{Performance}(X \rightarrow Y) = \text{CF}(X \rightarrow Y) \times \text{Support}(X \rightarrow Y)$$

این معیار می تواند مقادیری در بازه [۰ و ۱] به دست آورد. یک قانون با مقدار کارایی نزدیک به ۱ احتمالا برای کاربر مفیدتر است.

### ۳-۴- عملگرهای ژنتیک

**عملگر تقاطع،** یک جفت از جواب های والد از جمعیت فعلی با تورنامنت دوتایی (در الگوریتم NSGA II) یا چرخ رولت (در الگوریتم NRGA) بر اساس رتبه بندی پارتو و معیار ازدحام انتخاب می شوند. با جابه جا کردن ژن های والدین به طور تصادفی، دو فرزند تولید می کند. شکل ۳ یک مثال ساده از عملکرد این عملگر را نشان می دهد.



شکل ۳- یک مثال ساده از عملکرد عملگر تقاطع

**عملگر جهش،** شامل تغییر تصادفی وقفه ها (lb و ub) و an یک ژن به طور تصادفی انتخاب شده است. این عملگر به صورت تصادفی یکی از کران های وقفه را انتخاب کرده و به صورت تصادفی مقدار آن را افزایش یا کاهش می دهد. باید کاملا مراقب باشیم که از مقدار ثابت دامنه تجاوز نکنیم. در آن صورت، روشی که وقفه ها را تغییر می دهیم، شبیه به روش محاسبه شده در فرایند مقاردهی اولیه است. مقدار an از مجموعه {۰ و ۱} انتخاب می شود. بنابراین در صورت ۰ بودن، این مقدار به ۱ و در صورت ۱ بودن، به ۰ تبدیل می شود.

### ۳-۵- فرآیند شروع مجدد

برای دور شدن از بهینه محلی، این الگوریتم از یک فرآیند شروع مجدد استفاده می کند. این رویکرد، فرایند مقاردهی اولیه را مجدداً برای جمعیت اعمال می کند. علاوه بر این، EP با جمعیت جدید، طبق

شرایط نامغلوب، به روز می شود. فرآیند شروع مجدد هنگامی اعمال می شود که تعداد افراد جدید در جمعیت یک نسل کمتر از  $\alpha\%$  اندازه جمعیت فعلی باشد.

### ۳-۶- نمودار جریان مدل

با توجه به توضیحات بالا، الگوریتم پیشنهادی برای کاوش قوانین پیوستگی کمی می تواند به صورت زیر خلاصه شود:

**ورودی:** اندازه جمعیت  $N$ ، تعداد ارزیابی ها  $nTrials$ ، احتمال جهش  $P_{mut}$ ، فاکتور دامنه برای هر ویژگی مجموعه داده  $\delta$ ، آستانه تفاوت  $\alpha$ .

**خروجی:** EP

**گام ۱:** مقداردهی اولیه

الف) تولید جمعیت اولیه  $(P_t)$  با  $N$  کروموزوم

ب) ارزیابی جمعیت اولیه

پ) تولید تمام جبهه های نامغلوب  $F = (F_1, F_2, \dots)$  جمعیت اولیه و محاسبه فاصله ازدحام در  $F_i$ .

ت) مقداردهی اولیه EP

**گام ۲:** تولید جمعیت فرزندان  $(Q_t)$  به صورت زیر:

الف) انتخاب یک جفت جواب والد با انتخاب تورنمنت دوتایی در NSGA-II و چرخ رولت در

NRGA بر اساس رتبه بندی پارتو و معیار ازدحام

ب) تقاطع هر جفت و تولید دو فرزند (این عملگر ژن های والدین را جابه جا می کند). سپس،

اعمال عملگر جهش بر روی دو فرزند.

پ) محاسبه افراد جدید

**گام ۳:** تولید جمعیت بعدی  $(P_{t+1})$  به صورت زیر:

الف) ایجاد جمعیت ادغام شده با  $P_t$  و  $Q_t$

ب) تولید تمام جبهه های نامغلوب  $F = (F_1, F_2, \dots)$  جمعیت ادغام شده و محاسبه فاصله ازدحام

در  $F_i$

پ) ایجاد  $P_{t+1}$  با بهترین کروموزوم ها از جمعیت ادغام شده با استفاده از جبهه های نامغلوب و

فاصله ازدحام:

- شامل جبهه نامغلوب  $i$  ام در  $P_{t+1}$
- بررسی جبهه بعدی برای ورود (inclusion?)
- مرتب سازی به ترتیب نزولی با استفاده از عملگر مقایسه ازدحام
- انتخاب اولین جزء  $(N - |P_{t+1}|)$  از  $F_i$

**گام ۴:** به روز رسانی EP طبق شرایط نامغلوب

**گام ۵:** اگر تفاوت بین جمعیت فعلی و جمعیت قبلی کمتر از  $\alpha\%$  باشد، شروع مجدد جمعیت

**گام ۶:** اگر بیشترین تعداد ارزیابی بدست نیامد، رفتن به گام ۲

**گام ۷:** حذف اضافات در EP، پاک کردن کروموزوم‌هایی که زیر کروموزوم بقیه هستند. یک زیر کروموزوم، کروموزومی است که در آن فواصل تمامی ژن‌ها در فواصل ژن‌های کروموزوم دیگری وجود داشته باشد.

**گام ۸:** EP به دست می‌آید.

#### ۴- یافته‌های پژوهش

در این پژوهش، نمونه‌ها بر اساس روش نمونه‌گیری خوشه‌ای از بین شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران (داده‌های داخلی) و شرکت‌های پذیرفته شده در بورس نیویورک (داده‌های خارجی) انتخاب شده‌اند. داده‌های مورد استفاده در پژوهش حاضر از نوع داده‌های سری زمانی می‌باشند؛ داده‌های سری زمانی داده‌هایی هستند که در طی یک دوره زمانی جمع‌آوری می‌شوند. در این پژوهش از داده‌های روزانه شامل بالاترین و پایین‌ترین قیمت، قیمت پایانی و حجم معامله مربوط به سهام ۵ شرکت داخلی در سال ۹۴ و سهام ۵ شرکت خارجی در سال ۲۰۱۵ استفاده شده است. سپس از این داده‌ها برای محاسبه ۱۰ تا از رایجترین نشانگرهای مالی استفاده می‌شود. این نشانگرها عبارتند از:

میانگین متحرک ساده<sup>۲۸</sup>، میانگین متحرک نمایی<sup>۲۹</sup>، شاخص قدرت نسبی<sup>۳۰</sup>، همگرایی و اگرایی میانگین متحرک<sup>۳۱</sup>، نرخ تغییر<sup>۳۲</sup>، شاخص کانال کالا<sup>۳۳</sup>، نوسانگر تصادفی<sup>۳۴</sup>، شاخص دامنه درصد ویلیامز<sup>۳۵</sup>، جریان پول چایکین<sup>۳۶</sup> و نوسانگر درصد نقدی<sup>۳۷</sup>.

هریک از الگوریتم‌های QAR-CIP-NSGA II و QAR-CIP-NRGA بر روی هر سهم ۱۰ بار اجرا و نتایج حاصل ثبت شد. پس از حصول شرط خاتمه که در این پژوهش ۱۰۰ بار تکرار الگوریتم می‌باشد، از بین جواب‌هایی با تابع برازش بالا، به ازای هر کلاس (۱ و ۰ و -۱) سه جواب با بهترین تابع برازش را به عنوان قوانین طبقه‌بندی کننده انتخاب می‌کنیم. کلاس ۱ به معنی افزایش قیمت طی ۱۰ روز آینده، کلاس ۰- به معنی کاهش قیمت طی ۱۰ روز آینده و کلاس ۰ به معنی عدم تغییر قیمت طی ۱۰ روز آینده می‌باشد.

سپس بازده حاصل از سرمایه‌گذاری با استفاده از مدل‌های پیشنهادی با بازده استراتژی خرید و نگه‌داری سهم به مدت یک سال مقایسه گردید. همچنین دقت سیگنال‌های تولید شده در پیش‌بینی درست‌افزایش یا کاهش قیمت سهام محاسبه گردید. نتایج حاصل از این مقایسه در جدول ۲ به طور خلاصه آمده است.

جدول ۲- خلاصه نتایج حاصل از بازده مدل

| دقت پیش بینی |                 | بازده        |                 | بازده استراتژی خرید و نگهداری | نام سهم |
|--------------|-----------------|--------------|-----------------|-------------------------------|---------|
| QAR-CIP-NRGA | QAR-CIP-NSGA II | QAR-CIP-NRGA | QAR-CIP-NSGA II |                               |         |
| ٪۹۴,۳۸       | ٪۹۷,۱۸          | ٪۱۹۳,۴۱      | ٪۱۹۷,۶۹         | ٪-۵۶,۷۹                       | تاپیکو  |
| ٪۸۷,۶۷       | ٪۹۷,۲۲          | ٪۹۷,۳۷       | ٪۹۵,۹۴          | ٪-۵۵,۴۲                       | فولاد   |
| ٪۸۶,۱۸       | ٪۸۳,۷۷          | ٪۲۰۷,۳۳      | ٪۲۰۷,۰۳         | ٪-۵۲,۹۰                       | وغدیر   |
| ٪۹۶,۴۷       | ٪۹۸,۹۴          | ٪۱۱۸,۲۹      | ٪۱۱۳,۶۳         | ٪-۵۰,۷۲                       | همراه   |
| ٪۹۲,۱۱       | ٪۹۴,۹۳          | ٪۸۴,۶۳       | ٪۸۴,۵۲          | ٪-۳۵,۵۱                       | فارس    |
| ٪۸۱,۸۱       | ٪۸۱,۸۶          | ٪۱۶۸,۴۱      | ٪۱۷۰,۰۹         | ٪۱۱,۱۱                        | BAC     |
| ٪۸۲,۲۲       | ٪۸۹,۶۴          | ٪۱۶۸,۸۱      | ٪۱۶۷,۶۳         | ٪-۳۷,۹۲                       | FCX     |
| ٪۹۰,۳۲       | ٪۹۳,۹۰          | ٪۱۰۱,۷۴      | ٪۱۰۲,۵۲         | ٪۱۴,۷۵                        | GM      |
| ٪۸۹,۵۰       | ٪۸۷,۴۷          | ٪۲۱۴,۹۲      | ٪۲۱۳,۶۴         | ٪۲۶,۴۸                        | CHK     |
| ٪۹۰,۹۸       | ٪۹۶,۰۷          | ٪۴۲۱,۲۸      | ٪۴۴۱,۱۴         | ٪-۴۵,۱۹                       | PBR     |

از نتایج به دست آمده می توان نتیجه گرفت در هر دو الگوریتم QAR-CIP-NSGA II و QAR-CIP-NRGA بازده مدل نسبت به بازده استراتژی خرید و نگه داری سهم با اختلاف زیادی بیشتر است. برای ارزیابی عملکرد الگوریتم های فراابتکاری چندهدفه، ۵ معیار ارائه شده است. این معیارها عبارتند از: تعداد جواب های پارتو<sup>۳۸</sup> (N)، فاصله از نقطه ایده آل<sup>۳۹</sup> (MID)، یکنواختی<sup>۴۰</sup> (S)، بیشترین گسترش<sup>۴۱</sup> (MD) و زمان اجرای الگوریتم<sup>۴۲</sup> (T).

برای مقایسه دو الگوریتم از آزمون فرض برابری میانگین دو جامعه مستقل استفاده کردیم. به طور کلی فرضیه های مورد آزمون به شرح زیر هستند:

$$\begin{cases} H_0 = \mu_{\text{QAR-CIP-NSGA II}}(X) = \mu_{\text{QAR-CIP-NRGA}}(X) \\ H_1 = \mu_{\text{QAR-CIP-NSGA II}}(X) \neq \mu_{\text{QAR-CIP-NRGA}}(X) \end{cases}$$

نتایج آزمون های فرض مقایسه میانگین دو الگوریتم پیشنهادی در سطح  $\alpha=0,05$  به طور خلاصه در جدول ۳ آمده است.

جدول ۳- خلاصه نتایج آزمون های فرض مقایسه دو الگوریتم پیشنهادی

| ردیف | معیار                          | الگوریتم | میانگین | معناداری ۹۵٪ | الگوریتم برتر   |
|------|--------------------------------|----------|---------|--------------|-----------------|
| ۱    | تعداد جواب های آرشیو پارتو (N) | NSGA II  | ۱۹,۰۶   | ۰,۰۰۸        | QAR-CIP-NSGA II |
|      |                                | NRGA     | ۱۷,۰۷   |              |                 |
| ۲    | فاصله از نقطه ایده آل (MID)    | NSGA II  | ۱,۴۰۳۸  | ۰,۳۹۷        | تفاوتی ندارند   |
|      |                                | NRGA     | ۱,۴۱۱۱  |              |                 |
| ۳    | یکنواختی (S)                   | NSGA II  | ۰,۱۲۹۶  | ۰,۱۶۲        | تفاوتی ندارند   |
|      |                                | NRGA     | ۰,۱۴۵۰  |              |                 |
| ۴    | بیشترین گسترش (MD)             | NSGA II  | ۱,۱۱۴۳  | ۰,۷۴۳        | تفاوتی ندارند   |
|      |                                | NRGA     | ۱,۱۰۴۴  |              |                 |
| ۵    | زمان اجرای الگوریتم (T)        | NSGA II  | ۸۲,۱۹۱۱ | ۰,۰۰۱        | QAR-CIP-NSGA II |
|      |                                | NRGA     | ۸۵,۳۲۷۸ |              |                 |

#### ۵- نتیجه گیری و بحث

در این پژوهش، سعی شده است که با استفاده از الگوریتم های فرا ابتکاری، یک سیستم سیگنال دهی خودکار به منظور پشتیبانی از تصمیمات سرمایه گذاران ایجاد شود که با بررسی نشانگرهای تکنیکی و کاوش قوانین پیوستگی کمی بین داده ها، سیگنال های معاملاتی تولید کند تا سرمایه گذار بتواند با استفاده از این سیستم برای تصمیم گیری، سود حاصل از سرمایه گذاری خود را بیشینه کند. بدین منظور، یک مدل تکاملی چند هدفه جدید با استفاده از دو الگوریتم ژنتیک چندهدفه NSGA II و NRGA ارائه شد که سه هدف قابلیت درک، جالب بودن و کارایی را حداکثر می کند.

پس از اجرای هر الگوریتم و ثبت نتایج حاصل از آن، از بین قوانین پیوستگی تولید شده، برای هر سهم به ازای هر کلاسه سه تا از بهترین قوانین طبقه بندی کننده که منجر به تولید سیگنال خرید یا فروش می شوند را انتخاب می کنیم. بازده حاصل از سرمایه گذاری با استفاده از مدل های پیشنهادی با بازده استراتژی خرید و نگه داری سهم به مدت یک سال مقایسه می گردد. استفاده از این مدل در شرایطی معقول خواهد بود که بازده حاصل از آن بیشتر از بازده حاصل از استراتژی خرید و نگه داری باشد. با توجه به نتایج بدست آمده در جدول ۲، می توان نتیجه گرفت در هر دو الگوریتم QAR-CIP-NSGA II و QAR-CIP-NRGA بازده مدل نسبت به بازده استراتژی خرید و نگه داری سهم با اختلاف زیادی بیشتر است. بنابراین، با در نظر گرفتن دقت بالای پیش بینی، مدل فوق می تواند به عنوان ابزاری برای تصمیم گیری در معاملات بازار سرمایه جهت افزایش میزان سوددهی در بازارهای سرمایه، در همه چهارچوب های زمانی مورد استفاده قرار گیرد.

نتیجه مقایسه عملکرد دو الگوریتم QAR-CIP-NSGA II و QAR-CIP-NRGA نشان می دهد که جز در تعداد جواب های پارتو و زمان اجرای الگوریتم که در آن الگوریتم QAR-CIP-NSGA II بهتر عمل می

کند، تفاوت معناداری بین عملکرد این دو الگوریتم وجود ندارد. معیار زمان، در دوره های زمانی طولانی تر که تعداد داده ها بیشتر است، اهمیت می یابد. همچنین به نظر می رسد با توجه به اینکه الگوریتم QAR-CIP-NSGA II تعداد جواب (قوانین) بیشتری نسبت به الگوریتم QAR-CIP-NRGA تولید می کند، دارای دقت پیش بینی بالاتری باشد. صحت این امر در جدول ۲ نیز تایید می گردد. بنابراین به نظر می رسد استفاده از الگوریتم QAR-CIP-NSGA II برای کاوش قوانین پیوستگی کمی در بازار سهام مناسب تر باشد. به منظور افزایش کارایی مدل و کاربردی شدن آن، موارد زیر پیشنهاد می گردد:

- ۱) استفاده از سایر نشانگرهای تکنیکی برای افزایش کارایی مدل، افزایش تعداد نشانگرها یا تعویض آن ها با دیگر نشانگرهای در صورت افزایش کارایی.
- ۲) استفاده از نشانگرهایی که سیگنال های متفاوت و هم پوشانی کمتری داشته باشند.
- ۳) استفاده از توابع هدف متفاوت در صورت افزایش کارایی
- ۴) تخصیص وزن برای نشانگرها (باتوجه به اهمیت هر کدام) برای محاسبه اهداف مسئله

#### فهرست منابع

- [1] Agrawal, R., Srikant, R., Fast algorithms for mining association rules. In Proc. 20th Int. Conf. Very Large Data Bases, September 1994, Vol. 1215, pp. 487-499
- [2] Ahn K.-I., Kim J.-Y., Efficient mining of frequent itemsets and a measure of interest for association rule mining, Journal of Information & Knowledge Management 3, pp. 245-257, 2004
- [3] Alatas B., Akin E., Karci A. MODENAR: Multi-objective differential evolution algorithm for mining numeric association rules. Applied Soft Computing, 8(1), pp. 646-656, 2008
- [4] Berzal F., Blanco I., Sanchez D., Vila M., Measuring the accuracy and interest of association rules: a new framework, Intelligent Data Analysis 6, pp. 221-235, 2002
- [5] Brin S., Motwani R., Ullman J., Tsur S., Dynamic itemset counting and implication rules for market basket data, ACM SIGMOD Record 26 pp.255-264, 1997
- [6] Chang Chien Y.W., Chen Y.L., Mining associative classification rules with stock trading data-A GA-based method. Knowledge-Based Systems, 23(6), pp. 605-614. 2010
- [7] Coello C., Lamont G., Veldhuizen D.V., Evolutionary Algorithms for Solving Multi-Objective Problems. Kluwer Academic Publishers, 2002
- [8] Deb K., Multi-Objective Optimization using Evolutionary Algorithms, Kluwer Academic, 2001
- [9] Geng L., Hamilton H., Interestingness measures for data mining: a survey, ACM Computing Surveys 38, pp. 1-32, 2006
- [10] Ghosh A., Nath B., Multi-objective rule mining using genetic algorithms. Information Sciences 163 (1-3) pp. 123-133, 2004
- [11] Huang Y.P., Kao L.J., "A Novel Approach to Mining Inter-Transaction Fuzzy Association Rules from Stock Price Variation Data". The 14th IEEE International Conference, 2005
- [12] K. Deb, S. Agrawal, A. Pratab, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions Evolutionary Computation 6 (2) (2002) 182-197.
- [13] Martín D., Rosete A., Alcalá-Fdez J., Herrera F. QAR-CIP-NSGA-II: A new multi-objective evolutionary algorithm to mine quantitative association rules. Information Science, 2013



- [14] Ramaswamy S., Mahajan S., Silberschatz A., "On the discovery of interesting patterns in association rules", 24rd International Conference on Very Large Data Bases, San Francisco, CA, USA: 1998
- [15] Shannon, B. Technical Analysis Using Multiple Timeframes, Life Vest, 2008
- [16] Silverstein C., Brin S., Motwani Beyond R., market baskets: generalizing association rules to dependence rules, Data Mining and Knowledge Discovery 2, pp. 39–68, 1998
- [17] Srikant R., Agrawal R., Mining quantitative association rules in large relational tables, in: ACM SIGMOD International Conference on Management of data (SIGMOD96), Montreal, Quebec, Canada, 1996.
- [18] Ting J., Fu T.C., Chung F.L., Mining of stock data: intra-and inter-stock pattern associative classification. Threshold, 5(100), pp. 5-99, 2006
- [19] Yan X., Zhang C., Zhang S., Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support, Expert Systems with Applications 36 (2) (2009) 3066–3076.
- [20] Zhang C., Zhang S., Association Rule Mining: Models and Algorithms. Lecture Notes in Computer Science, LNAI 2307, Springer-Verlag, Berlin: 2002

## یادداشت‌ها

- <sup>1</sup>. Association Rules
- <sup>2</sup>. Metaheuristics
- <sup>3</sup>. Support
- <sup>4</sup>. Confidence
- <sup>5</sup>. Comprehensibility
- <sup>6</sup>. Interestingness
- <sup>7</sup>. Performance
- <sup>8</sup>. Non-dominated Sorting Genetic Algorithm
- <sup>9</sup>. Non-dominated Ranked Genetic Algorithm
- <sup>10</sup>. External Population
- <sup>11</sup>. Restarting Process
- <sup>12</sup>. Technical Andicators
- <sup>13</sup>. Agrawall
- <sup>14</sup>. Srikant
- <sup>15</sup>. Yo-Ping Huang
- <sup>16</sup>. Li-Jen Kao
- <sup>17</sup>. Jo Thing
- <sup>18</sup>. Ghosh
- <sup>19</sup>. Nath
- <sup>20</sup>. Alatas
- <sup>21</sup>. Yan
- <sup>22</sup>. Chang Chien
- <sup>23</sup>. Chen
- <sup>24</sup>. Martin
- <sup>25</sup>. Amplitude
- <sup>26</sup>. Generalization
- <sup>27</sup>. Specificity
- <sup>28</sup>. Simple Moving Average (SMA)

- <sup>29</sup>. Exponential Moving Average (EMA)
- <sup>30</sup>. Relative Strength Index (RSI)
- <sup>31</sup>. Moving Average Convergence-Divergence (MACD)
- <sup>32</sup>. Rate Of Change (ROC)
- <sup>33</sup>. Commodity Channel Index (CCI)
- <sup>34</sup>. Stochastic Oscillator (K,D)
- <sup>35</sup>. Williams' Percent Range (%R)
- <sup>36</sup>. Chaikin Money Flow (CMF)
- <sup>37</sup>. Percentage Price Oscillator (PPO)
- <sup>38</sup>. Number Of Solutions
- <sup>39</sup>. Mean Ideal Distance
- <sup>40</sup>. Spacing
- <sup>41</sup>. Maximum Diversity
- <sup>42</sup>. Time