مجله مهندسی مالی و مدیریت اوراق بهادار
شماره بیست و پنجم / زمستان ۱۳۹٤

استفاده از روش هیبرید انتخاب ویژگی و الگوریتم نزدیکترین همسایگی برای پیش‌بینی جهت حرکتی روزانه شاخص۵۰ شرکت فعال‌تر بورس و اوراق بهادار تهران

احمد پویان‌فر[۱]
سعید فلاح پور[۲]
عیسی نوروزی یان لکوان[۳]
امیرحسین فرهادی شولی[۴]

تاریخ دریافت: ۹۴/۲/۱۵ تاریخ پذیرش: ۹۴/۶/۷

چکیده

پیش‌بینی بازار سهام به علت پر سود بودن معاملات سهام همواره مورد توجه معامله‌گران و سرمایه-گذاران می‌باشد. یک معامله موفق سهام در خرید و یا فروش در نزدیکی نقاطی که روند قیمت تغییر می‌یابد، اتفاق می‌افتد. بنابراین پیش‌بینی شاخص بازار سهام و تحلیل آن برای تشخیص اینکه آیا قیمت بسته شدن سهام در روز بعد افزایش خواهد یافت و یا کاهش، بسیار مهم است. در این پژوهش از روش طبقه‌بندی نزدیکترین همسایگی بر پایه روش ترکیبی انتخاب ویژگی برای پیش‌بینی جهت حرکتی شاخص ۵۰ شرکت فعال تر بورس اوراق بهادار تهران استفاده شده است. این روش هیبرید انتخاب ویژگی، که ترکیبی از روش تجزیه و تحلیل اجزای اساسی و الگوریتم ژنتیک می‌باشد از مزایای هر دو نوع روش پوشش دهنده و فیلترکننده انتخاب ویژگی، برای انتخاب یک زیرمجموعه بهینه از بین فضای کل ویژگی‌ها برخوردار می‌باشد. عملکرد روش ترکیبی پیشنهادی با روش‌های متداول انتخاب ویژگی که عبارت است از: زنجیره اطلاعات، رلیف و روش آنالیز اجزای اساسی که جزو روش‌های فیلتر هستند و روش الگوریتم ژنتیک که از خانواده روش‌های پوشش‌دهنده می‌باشد، با استفاده از آزمون مقایسات زوجی مقایسه گردیده و نتایج حاصل نشان می‌دهد که روش ترکیبی ارائه شده از عملکرد بالاتری نسبت به دیگر روش‌های استفاده شده، در پیش‌بینی جهت حرکتی روزانه شاخص ۵۰ شرکت فعال‌تر بورس اوراق بهادار تهران برخوردار می‌باشد.

واژه‌های کلیدی: آنالیز اجزای اساسی، الگوریتم ژنتیک، انتخاب ویژگی، پیش‌بینی روند، پوشش‌دهنده، فیلترکننده، نزدیکترین همسایگی.

۱- دکتری مدیریت مالی، دانشکده مدیریت، دانشگاه تهران، apouyanfar@gmail.com
۲- دانشیار، دکتری مدیریت مالی، دانشکده مدیریت، دانشگاه تهران، falahpor@ut.ac.ir
۳- دانشجوی کارشناسی ارشد مهندسی مالی، دانشکده مدیریت، دانشگاه تهران (نویسنده مسئول)، norouzian@ut.ac.ir
۴- کارشناسی ارشد مهندسی پزشکی، دانشکده برق و کامپیوتر، دانشگاه تهران، a.farhadi@ut.ac.ir

Archive of SID
www.SID.ir

۱- مقدمه

پیش‌بینی دقیق بازار سهام به دلیل وضعیت سیاسی، شرایط اقتصاد جهانی و سایر عوامل معمولاً کار بسیار دشواری است. بنابراین محققین همواره به دنبال معرفی و ابداع تکنیک‌هایی بوده‌اند که بتوانند بر این محدودیت‌ها فائق آمده و قیمت سهام را پیش‌بینی کنند. به همین دلیل تاکنون روش‌ها و تکنیک‌های بسیاری ارائه گردیده است. این روش‌ها از شاخصه‌های تحلیل تکنیکال که از ساده‌ترین تکنیک‌ها هستند شروع و شامل تکنیک‌های پیشرفته‌تری از قبیل شبکه عصبی مصنوعی، رگرسیون خطی و غیرخطی، الگوریتم ژنتیک، ماشین بردار پشتیبان و غیره می‌باشند (هانگ ، یانگ و چانگ، ۲۰۰۸).

با عنایت به عدم پیش‌بینی پذیری اندازه حرکت یا تغییر قیمت دارائی در اکثر بازارها، و در برخی موارد مهم نبودن اندازه حرکت آتی، محققین به سراغ روش‌هایی رفتند که صرفا جهت حرکت قیمت دارائی را پیش‌بینی نماید. این روش‌ها شامل روش‌های سنتی از قبیل لاجیت ، پروبیت، گشت تصادفی و روش‌های جدید تر از قبیل شبکه عصبی مصنوعی، الگوریتم نزدیکترین همسایگی و ماشین بردار پشتیبان می باشند. پژوهش‌های انجام شده در زمینه پیش‌بینی جهت حرکتی شاخص و قیمت سهام نشان‌دهنده این است که روش‌های جدید داده‌کاوی و مدل‌های غیرخطی در مقایسه با روش‌های پیش‌بینی سنتی دقت بالاتری داشته و از عملکرد بهتری برخوردار می‌باشند (فلاح‌پور، گل ارضی، فتوره چیان، ۱۳۹۲، ص ۲۷۰).

با پیشرفت‌های اخیر در حوزه هوش مصنوعی روش‌های جدیدتری برای پیش‌بینی ارائه گردیده است. تحقیقات بسیاری در زمینه پیش‌بینی جهت حرکتی قیمت با به کارگیری از محاسبات نرم[1] و روش‌های مختلف داده‌کاوی صورت گرفته که یکی از جدیدترین روش ها، روش‌های انتخاب ویژگی[2] برای انتخاب ویژگی‌های مهم در پیش‌بینی می‌باشند (آتسالکیس و والوانیس،۲۰۰۹، ص ۵۹۳۱). در زمینه پیش‌بینی قیمت سهام و یا شاخص بازار، انتخاب ویژگی، انتخاب یک زیرمجموعه بهینه از ویژگی‌های بنیادی و یا تکنیکال خواهد بود.

در پیش‌بینی جهت حرکتی قیمت سهام و شاخص، دقت پیش‌بینی تحت تأثیر تعداد ویژگی‌ها و متغیرهای ورودی مدل قرار می‌گیرد. استفاده از بردارهای ویژگی با ابعاد بزرگ هزینه محاسباتی زیاد و ریسک بیش برازشی مدل را به همراه دارد. انتخاب ویژگی با تعیین یک زیرمجموعه بهینه از ویژگی‌هایی که بیشترین تأثیر را در دقت پیش‌بینی دارند باعث کاهش ابعاد بردار ویژگی‌ها می‌شود. روش‌های انتخاب ویژگی به طور به دو دسته، روش‌های فیلترکننده و روش‌های پوشش‌دهنده تقسیم می‌شوند. **از** روش‌های فیلترکننده می‌توان به روش‌های آنالیز اجزای اساسی، زنجیره اطلاعات و رلیف که جزو معروف‌ترین روش-های فیلترکننده هستند، اشاره نمود. الگوریتم روش‌های فیلترکننده بر اساس قوانین آماری طراحی شده‌اند و همچنین از روش‌های پوشش‌دهنده نیز می‌توان به الگوریتم ژنتیک، الگوریتم کلونی مورچگان و سایر الگوریتم‌های زیستی دیگر اشاره کرد که بر اساس قوانین نشأت گرفته از طبیعت عمل می‌کنند. روش‌های ذکر شده از مهمترین و پرکاربردترین روش‌هایی می‌باشند که محققین و پژوهشگران از آن‌ها در زمینه پیش‌بینی جهت حرکتی شاخص بازار و سهام استفاده کرده‌اند. هر کدام از این دو نوع رویکرد انتخاب ویژگی

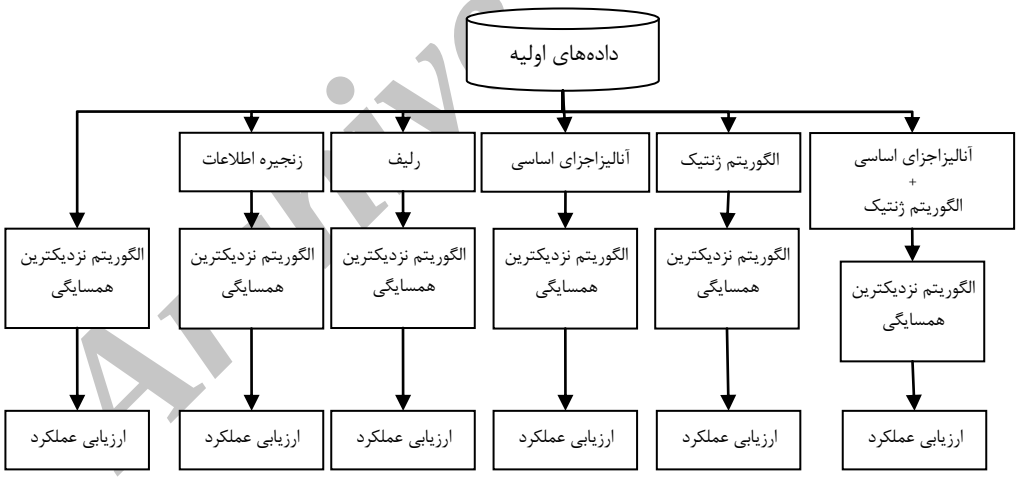
Archive of SID

www.SID.ir

مزایا و معایبی دارند که با طراحی مدلی که تشکیل شده از این دو نوع روش باشد می‌توان از مزایای این دو روش بهره گرفت و معایب آنها را حذف کرد. تاکنون تحقیقات کمی در زمینه ترکیب این دو نوع روش انتخاب ویژگی با همدیگر در حوزه پیش‌بینی جهت حرکتی شاخص و سهام صورت گرفته است. در این پژوهش یک مدل جدید بر اساس الگوریتم نزدیکترین همسایگی و ترکیب آن با یک روش هیبرید انتخاب ویژگی برای پیش‌بینی جهت حرکتی شاخص ۵۰ شرکت برتر بورس اوراق بهادار تهران ارائه گردیده است. این روش هیبرید انتخاب ویژگی از ترکیب یک روش فیلترکننده انتخاب ویژگی و یک روش پوشش‌دهنده انتخاب ویژگی برای پیدا کردن زیرمجموعه بهینه از ویژگی‌ها از میان ویژگی‌ها ورودی اولیه تشکیل شده است.

مدل‌های ارائه شده در این پژوهش، از نوع مدل‌های ترکیبی[۳] انتخاب ویژگی با مدل طبقه‌بندی کننده[۴] می‌باشند. معیارهای تحلیل تکنیکال به عنوان ویژگی‌ها، ورودی مدل هستند که از آنها برای پیش‌بینی جهت حرکتی استفاده شده است. بدین ترتیب عملکرد روش هیبرید[۵] آنالیز اجزای اساسی و الگوریتم ژنتیک به عنوان یک روش ترکیبی جدید با روش‌های انتخاب ویژگی آنالیزاجزای اساسی[۶]، زنجیره‌ی اطلاعات[۷] و رلیف[۸] که از نوع روش‌های فیلترکننده[۹] می‌باشند و همچنین الگوریتم ژنتیک که از نوع روش‌های پوشش دهنده[۱۰] می‌باشد مقایسه شده است.

در شکل شماره ۱. کلیه تکنیک‌ها و روش‌های استفاده شده در این پژوهش ارائه شده است.



شکل شماره۱. روش های مختلف به کاربرده شده

نتایج حاصله حاکی از این واقعیت می باشد که با استفاده از روش هیبرید آنالیز اجزای اساسی و الگوریتم ژنتیک و ترکیب آن با الگوریتم نزدیکترین همسایگی می‌توان جهت حرکتی شاخص ۵۰ شرکت فعال‌تر را با دقت بالاتری نسبت به دیگر روش‌های استفاده شده در این پژوهش پیش‌بینی نمود.

Archive of SID

www.SID.ir

ادامه مقاله به صورت زیر می‌باشد. پس از مقدمه حاضر مروری بر مطالعات انجام شده در حوزه انتخاب ویژگی خواهیم داشت. در بخش سوم مدل تحقیق و در بخش چهارم روش تحقیق و سپس یافته‌های آن در بخش پنجم ارائه می گردد. نهایتاً در بخش ششم به نتیجه‌گیری خواهیم پرداخت.

۲- مبانی نظری و مروری بر پیشینه پژوهش

پیش‌بینی جهت حرکتی قیمت سهام و یا شاخص کار بسیار مشکلی می‌باشد. بر اساس فرضیه معروف بازار کارا که توسط فاما (۱۹۷۰) مطرح گردیده قیمت‌ها به سرعت نسبت به اطلاعات در دسترس واکنش نشان می‌دهند و تنها چیزی که باعث تغییر قیمت می‌شود اطلاعات جدید می‌باشد. بنابراین، چون اخبار جدید را نمی‌توان پیش‌بینی کرد، جهت حرکت قیمت سهام کاملا تصادفی بوده و بر اساس این فرضیه کسب بازده غیرعادی و اضافی در این بازار ممکن نمی‌باشد. بنابراین بهترین استراتژی در این بازار خرید و نگه‌داری سهام می‌باشد

در طی چندین دهه گذشته، تحقیقات بی شماری در ارتباط با اعتبارسنجی این فرضیه صورت گرفت. که برخی از آنها این فرضیه را تایید (فاما،۱۹۷۰) و برخی آن را رد نموده‌اند (هاگن، ۱۹۹۹و لاس،۲۰۰۰، تکسیرا و الیویرا، ۲۰۱۰). تکنیک‌های بررسی فرضیه فوق، یا به عبارتی دیگر فرضیه پیش‌بینی پذیری قیمت دارائی‌های مالی، از آزمون‌های آماری بسیار ساده همانند آزمون گردش شروع و تکنیک های بسیار پیشرفته اقتصادسجی و مدل های کمی همانند شبکه‌های عصبی را شامل می شود. به عنوان مثال وانستون و تان (وانستون و تان، ۲۰۰۳) کارهای صورت گرفته در این زمینه را بررسی نمودند و آن‌ها را به چند گروه به شرح زیر تقسیم نمودند: سری‌های زمانی (ویو، لیو و وانگ، ۲۰۰۵و کاو و تای، ۲۰۰۳)، طبقه‌بندی و کشف الگو (بااو و یانگ، ۲۰۰۸و نانی ۲۰۰۶و سای و یوان،۲۰۰۷)، بهینه‌سازی(چانگ و هسیو، ۲۰۰۷) و روش‌های هیبرید (آفولابی و اولاد ۲۰۰۷، کیم و شین، ۲۰۰۷، کوان و مون، ۲۰۰۷).

هر چند مدل‌های آماری توانسته‌اند پیش‌بینی‌های خوبی را در زمینه پیش‌بینی شاخص و قیمت سهام از خود نشان دهند ولی وجود مفروضات محدودکنندهٔ برخی از این مدل‌ها که بر اثربخشی آنها موثر بوده است، باعث شد به تدریج روش‌های دیگری برای مقابله با این محدودیت‌ها و بهبود عملکرد پیش‌بینی‌ها معرفی شوند (فلاح‌پور، گل ارضی، فتوره چیان، ۱۳۹۲، ص ۲۷۳). در ادامه نمونه‌هایی از تحقیقاتی که بیشترین ارتباط را با موضوع این پژوهش دارند، ذکر خواهد گردید.

محققین و پژوهشگران همواره به دنبال افزایش دقت پیش‌بینی می‌باشند بنابراین با توسعه علم، روش-های جدید به وجود آمده مورد آزمون قرار می‌گیرند تا عملکرد و مزایا و معایب آن‌ها مورد بررسی قرار گیرند. اولین پژوهشی که با استفاده از هوش مصنوعی انجام شد، مربوط به وایت بود که در سال ۱۹۸۸ منتشر گردید. او سعی کرد تا با استفاده از شبکه عصبی قوانینی را در سری‌های زمانی کشف نماید. در سال ۱۹۹۰ توسط کینوتو و همکارانش با استفاده ازشبکه‌های عصبی سیستم پیش‌بینی قیمت را توسعه داد. نیکولوپلاس و فلرات (۱۹۹۴) از ترکیب الگوریتم ژنتیک و شبکه عصبی یک مدل پیش‌بینی کننده برای

Archive of SID

www.SID.ir

کمک به تصمیمات سرمایه‌گذاری ارائه نمودند. کیم و هان (۲۰۰۰) از الگوریتم ژنتیک برای گسسته‌سازی ویژگی و تعیین وزن‌های پیوسته برای شبکه عصبی مصنوعی به منظور پیش‌بینی شاخص بورس استفاده کردند.

آبراهام و همکارانش در سال ۲۰۰۱ با استفاده از شبکه عصبی فازی BPNN[11]جهت حرکتی شاخص نزدک را پیش‌بینی کردند. آنها در تحقیق خود از معیارهای تحلیل بنیادی استفاده نمودند که با استفاده از روش انتخاب ویژگیPCA[12] سعی در انتخاب یک زیر مجموعه بهینه از ویژگی‌ها نمودند تا بتوانند توسط آن بالاترین دقت را حاصل نمایند.

اینس و ترافایلس در سال ۲۰۰۴ با استفاده از شاخصه‌های تحلیل تکنیکال و مدل‌های پیش‌بینی کننده SVM[13]وBPNN شاخص نزدک را پیش‌بینی نمودند. همچنین آنها برای کاهش ابعاد بردار ویژگی ورودی مسئله از روش‌های انتخاب ویژگیFA[14]و PCA استفاده نمودند.

سال ۲۰۰۸ توسط لی و کیو پژوهشی در راستای پیش‌بینی جهت حرکتی شاخص بازار بورس تایوان با استفاده از مدل SOM+BPNN به عنوان مدل پیش‌بینی کننده و روشDWT[15]استفاده نمودند. ورودی مدل آنها شاخصه‌های تحلیل تکنیکال بود که با استفاده از روش DWT سعی کردند موثرترین شاخصه‌های تحلیل تکنیکال را برای رسیدن به بالاترین دقت شناسایی کرده و آنها را به کار گیرند.

لین و همکارانش در سال ۲۰۰۹ با استفاده از شاخصه‌های تحلیل تکنیکال جهت حرکتی شاخص S&p 500 را پیش‌بینی کردند. آنها در تحقیق خود از روش PCA به عنوان یک روش فیلترکننده انتخاب ویژگی و از مدل‌هایESN[16]، RNN[17]وBPNN به عنوان مدل‌های پیش‌بینی کننده استفاده کردند.

لای و همکارانش در سال ۲۰۰۹ تحقیقی در زمینه پیش‌بینی جهت حرکتی شاخص بورس انجام دادند. مورد مطالعاتی آنها شاخص بورس تایوان بود آنها از ۷ شاخصه تحلیل تکنیکال به عنوان ویژگی‌های ورودی مدل خود استفاده کردند.آنها برای انتخاب موثرترین ویژگی‌ها از روش آماری رگرسیون استفاده نمودند و از درخت تصمیم فازی برای پیش‌بینی جهت حرکتی شاخص بورس تایوان استفاده کردند.

در سال ۲۰۰۹ هانگ و تسای با به کارگیری ۱۳ معیار تحلیل تکنیکال و یک روش انتخاب ویژگی فیلترکننده و همچنین با به کارگیری و مدل SVR[19]-SOFM[18]به عنوان مدل پیش‌بینی کننده سعی در پیش‌بینی جهت حرکتی شاخص آتی بورس تایوان نمودند.

در سال ۲۰۱۰ تکسیرا و الیویرا با استفاده از الگوریتم نزدیکترین همسایگی و ۲۲ شاخص تحلیل تکنیکال به عنوان ویژگی‌های ورودی مدل خود در پیش‌بینی قیمت سهام در بازار بورس اوراق بهادار برزیل استفاده نمودند. آنها یک سیستم معاملاتی بر مبنای مدل پیش‌بینی کننده خود طراحی نمودند.

نایر، مهنداس و ساکتوهیل در سال ۲۰۱۰ برای پیش‌بینی روند حرکتی بازار سهام، از الگوریتم ژنتیک بر پایه درخت تصمیم که برای پیش‌بینی از ماشین بردار پشتیبان بهره گرفته شد، استفاده نمودند. نتایج بدست آمده از آن نشان داد، مدل ترکیبی ذکر شده از عملکرد بالاتری نسبت به شبکه عصبی مصنوعی برخوردار است. در این پژوهش از شاخصه‌های تحلیل تکنیکال به عنوان متغیرهای ورودی مدل استفاده شد.

Archive of SID

www.SID.ir

نتایج بدست آمده از پژوهشی که در سال ۲۰۱۱ توسط یاکوب، ملک و عمر در کشور ترکیه انجام شد، نشان داد شبکه عصبی مصنوعی از دقت بالاتری نسبت به ماشین بردار پشتیبان در پیش‌بینی روند حرکتی شاخص بازار سهام استانبول برخوردار است. آنها از ۱۰ معیار شاخص تکنیکال برای پیش‌بینی روند حرکتی شاخص استفاده نمودند..

در ایران نیز تحقیقاتی در این زمینه صورت گرفته است. تحقیق انجام شده توسط فلاح شمس و همکارانش (۱۳۸۸) نشان داد، که روش شبکه عصبی خطای RMSE به میزان قابل توجهی کمتر از RMSE روشهای دیگر است و در بازار بورس اوراق بهادار تهران پیش‌بینی کوتاه مدت با فاصله زمانی کمتر، مناسب‌تر از پیش‌بینی بلند مدت با فاصله زمانی طولانی‌تر است.

در پژوهشی منجمی، ابزری و رعیتی شوازی، (۱۳۸۸)، قیمت سهام، در بازار بورس اوراق بهادار، به کمک شبکه عصبی فازی و الگوریتم ژنتیک پیشبینی شد و نتایج آن با روش شبکه عصبی مصنوعی مقایسه گردید و نتایج نشان داد شبکه عصبی مصنوعی در ترکیب با الگوریتم ژنتیک نسبت به شبکه عصبی مصنوعی منفرد دقت و سرعت بالاتری دارد.

عبادی (۱۳۸۸) در پژوهش خود به پیش‌بینی شاخص کل بورس اوراق بهادار پرداخت و نشان داد، شبکه عصبی در برآوردن شاخص کل قیمت سهام بورس اوراق بهادار تهران از کارایی بالایی برخوردار است.

هاشمی (۱۳۸۹) با استفاده از شبکه عصبی رگرسیونی جلوسو تاثیر فاکتورهای رفتاری بر پیش‌بینی قیمت سهام را برای ۱۰ شرکت بررسی نمود و به این نتیجه رسید که فاکتورهای رفتاری در پیش‌بینی قیمت سهام نه شرکت از ده شرکت موثر هستند و دقت پیش‌بینی را به طرز قابل توجهی افزایش می‌دهد.

۳- مدل و الگوریتم پژوهش و متغیرهای آن

در این پژوهش از الگوریتم نزدیکترین همسایگی به عنوان مدل پیش‌بین استفاده گردیده است که از ترکیب آن با روش‌های مختلف انتخاب ویژگی سعی در افزایش دقت آن نموده‌ایم. الگوریتم نزدیکترین همسایگی نسبت به دیگر روش‌های هوش مصنوعی مانند شبکه عصبی و ماشین بردار پشتیبان از سرعت بالاتری برخوردار است، به همین دلیل در طراحی بسیاری از سیستم‌های معاملاتی هوشمند که به دلیل تولید سیگنال‌های خرید و فروش در بازه‌های زمانی کوتاه نیاز به سرعت بالایی در پیش‌بینی دارند، می‌تواند بسیار پرکاربرد باشد.

۳-۱- انتخاب ویژگی

به طور کلی الگوریتم‌های انتخاب ویژگی بر اساس معیارهای مختلف ارزیابی طراحی شده‌اند و به دو گروه اصلی تقسیم می‌شوند:

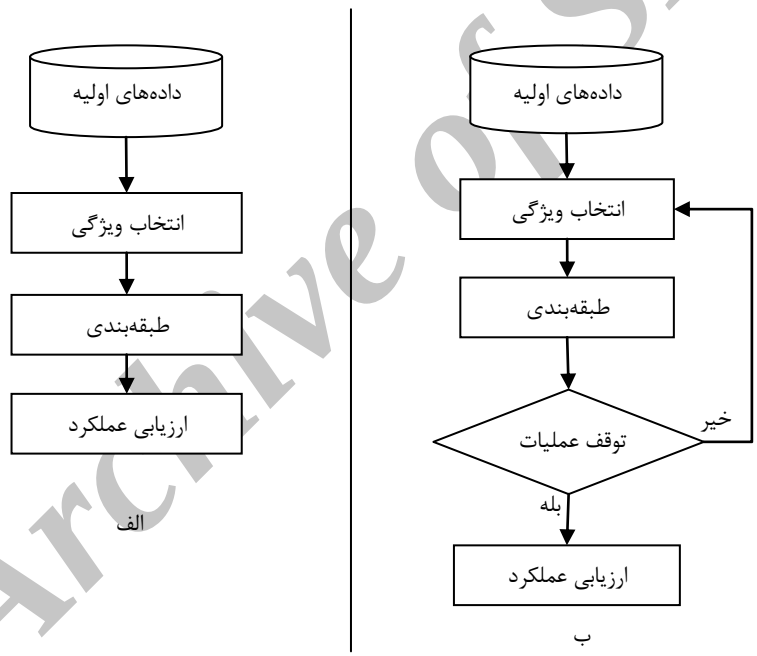
۱) روش‌های فیلترکننده[۲۰](داش و همکارانش،۲۰۰۲، هال، ۲۰۰و یو ولیو ۲۰۰۳)

۲) روش‌های پوشش‌دهنده[۲۱](دی و برودلی،۲۰۰۰ و کوهاوی و جوهان، ۱۹۹۷)

Archive of SID

www.SID.ir

روش‌های فیلترکننده دقت پیش‌بینی و یا طبقه‌بندی را بر اساس یک معیار غیر مستقیم ارزیابی می‌کنند مانند معیار فاصله که اشاره به این دارد که چقدر خوب کلاس‌ها از هم جدا شده‌اند (هیوانگ و همکاران، ۲۰۰۷). این روش‌ها در انتخاب ویژگی بدون در نظر گرفتن دقت پیش‌بینی عمل می‌کنند و سرعت بالایی دارند. بر خلاف روش‌های فیلترکننده، روش‌های پوشش‌دهنده کاملا به مدل طبقه‌بندی کننده وابسته هستند و الگوریتم بر اساس دقت بدست آمده از مدل طبقه‌بندی کننده زیرمجموعه بهینه را تعیین می‌کند و معیار انتخاب آن دقت بدست آمده می‌باشد و هر زیرمجموعه‌ای که دقت بالاتری را حاصل کند انتخاب می‌گردد (مین ولی، ۲۰۰۵ و لاورنس و همکاران، ۱۹۹۷). این روش‌ها به دلیل اینکه در هر بار انتخاب زیرمجموعه برای ارزیابی آن مدل طبقه‌بندی را فراخوانی می‌کنند نسبت به روش‌های فیلترکننده سرعت کمتری دارند.



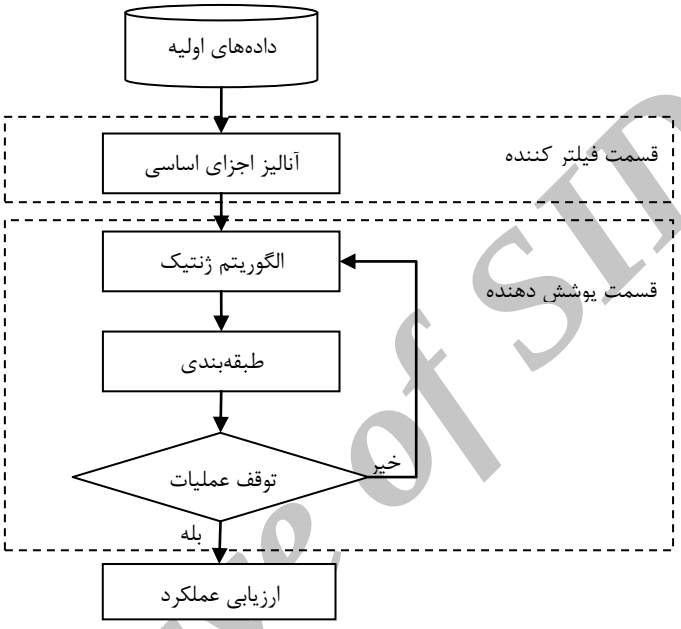
شکل شماره ۱. مقایسه روش‌های پوشش‌دهنده (الف) و فیلترکننده برای انتخاب ویژگی(ب)

علی‌رغم دقت بالای روش‌های پوشش‌دهنده، این روش‌ها به دلیل پیچیدگی محاسباتی که دارند دچار محدودیت استفاده هستند. بنابراین دقت بالا از مزایای روش‌های پوشش‌دهنده بوده و سرعت بالا نیز از مزایای روش‌های فیلترکننده می‌باشد. در این پژوهش برای بهره‌گیری از سرعت بالای روش‌های فیلترکننده و اجتناب از پیچیدگی محاسباتی روش‌های پوشش‌دهنده یک روش هیبرید انتخاب ویژگی ارائه شده‌است

Archive of SID

www.SID.ir

که این روش از ترکیب روش‌آنالیز اجزای اساسی به عنوان یک روش‌فیلترکننده و الگوریتم ژنتیک به عنـوان یک روش پوشش‌دهنده بدست آمده است.



شکل شماره ۲. فلوچارت روش هیبرید انتخاب ویژگی

۳-۲- زنجیره اطلاعات

این روش ارتباط بین ویژگی‌ها را بر اساس زنجیره‌ی اطلاعات در نظر گرفته و با توجه به طبقه‌ی هر ویژگـی به ویژگی‌ها وزن می‌دهد. X , Y نشان دهنده ویژگی می‌باشد.

$$InfoGain = H(Y) - H(Y|X) \quad (۱)$$

$$H(Y) = -\sum_{y \epsilon Y} P(y) log_2(p(y)) \quad (۲)$$

$$H(Y|X) = -\sum_{x \epsilon X} P(x) \sum_{y \epsilon Y} (y|x) \log(p(y|x)) \quad (۳)$$

Archive of SID

www.SID.ir

۳-۳- رلیف

ویژگی‌ها را بر اساس نمونه‌برداری و جاگذاری با در نظرگرفتن ارزش داده شده به ویژگی بـا نزدیکتـرین جایگزین با طبقه یکسان یا متفاوت مشخص می‌شود. اهمیت داده شده به ویژگی‌ها در این موارد باید مرتـب بروزآوری شده و مجدداً امتیازدهی شود (نی و همکاران، ۲۰۱۱).

۳-۴- آنالیز اجزای اساسی

یک روش ریاضی که با استفاده از تبدیل متعامد، یک مجموعه‌ای از مشاهدات احتمالا همبسته را به یک مجموعه ارزش از مشاهدات ناهمبسته که به آن مولفه‌های اصلی می‌گویند، تبدیل می‌کند. تعداد مولفه‌هـای اساسی کوچکتر یا مساوی تعداد ویژگی‌های ورودی می‌باشد. اولین جزء اصلی به گونه‌ای انتخاب می‌شود که بالاترین واریانس را داشته باشد. هر جزئی که واریانس بالایی داشته باشد، می‌توان آن را ناهمبسته از اجزای دیگر دانست و ویژگی‌ها بر اساس واریانس خود مرتب می‌شوند.

اجزای اساسی می‌توانند به صورت زیر تعریف شوند:

(۴)

$$Y_1 = \vec{a}_1^{\mathrm{T}}.\vec{\mathrm{x}} = a_{11}\mathrm{x}_1 + a_{12}\mathrm{x}_2 + \cdots + a_{1\mathrm{n}}\mathrm{x}_\mathrm{n}$$

$$Y_2 = \vec{a}_2^{\mathrm{T}}.\vec{\mathrm{x}} = a_{21}\mathrm{x}_1 + a_{22}\mathrm{x}_2 + \cdots + a_{2\mathrm{n}}\mathrm{x}_\mathrm{n}$$

$$\cdots$$

$$Y_n = \vec{a}_n^{\mathrm{T}}.\vec{\mathrm{x}} = a_{\mathrm{n}1}\mathrm{x}_1 + a_{\mathrm{n}2}\mathrm{x}_2 + \cdots + a_{\mathrm{nn}}\mathrm{x}_\mathrm{n}$$

جایی که به ترتیب $\mathrm{x_i}, i = 1,2,3 \ldots, n$ متغیرهای اولیه ، Y_i اجزای اساسی و $\vec{a}_\mathrm{i}$ بردار ضرایب می‌باشد. $\vec{a}_\mathrm{i}$ می‌توانـد تخمـین زده شـود بـه وسـیله ماکسـیمم سـازی $\mathrm{var}(Y_\mathrm{i})$ بـا شـرایط محـدودیت $\vec{a}_\mathrm{i}^{\mathrm{T}}.\vec{a}_\mathrm{i} = 1$ و $\mathrm{cov}(Y_\mathrm{i}, Y_\mathrm{j}) = \vec{a}_\mathrm{i}.\sum \vec{a}_\mathrm{i} = 0, \mathrm{j} = 1,2 \ldots \mathrm{i} - 1$ جایی که $\sum = (\sigma_{ij})_{n\times n}$ کواریانس ماتریس $\vec{\mathrm{x}}$ است.

انتخاب اجزای اساسی: ماتریس کواریانس $\sum = (\sigma_{ij})_{n\times n}$ ، $\vec{\mathrm{x}} = (\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n})^{\mathrm{T}}$ یک ماتریس متقـارن نامنفی می‌باشد که شامل n ریشه $\lambda_1, \lambda_2 \ldots . \lambda_n$ و n بردار ویژگی می‌باشد.

فرض کنید $\lambda_1 \geq \lambda_2 \geq \cdots . \geq \lambda_n \geq 0$ و $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_\mathrm{n}$ بردارهای ویژه دستگاه متعامد باشند آنگاه iامین عنصر اساسی $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n}$ به صورت زیر بیان می‌شود.

$$\mathrm{Y_i} = \mathrm{e_{i1}x_1} + \mathrm{e_{i2}x_2} \ldots \mathrm{e_{in}x_n}, \quad i = 1,2,3 \ldots ., n \qquad (۵)$$

$\mathrm{cov}(Y_i, Y_j) = \vec{e}_i^{\mathrm{T}}.\sum.\vec{e}_i = 0, \mathrm{j} = \mathrm{i}$ و $\vec{e}_i^{\mathrm{T}}.\sum.\vec{e}_i = \lambda_\mathrm{i}\mathrm{var}(\mathrm{Y_i})$=با

نخستین p اجزای اساسی از رابطه زیر بدست می‌آید.

$$\mathrm{ACR(P)} = \sum_{i=1}^{p}\lambda_\mathrm{i} / \sum_{i=1}^{n}\lambda_\mathrm{i} \quad (۶)$$

که نشان دهنده قدرت توضیحی برای داده‌های اصلی از اجزای اصلی استخراج شده از روش PCA می‌باشد.

Archive of SID

www.SID.ir

۵-۳- الگوریتم ژنتیک

الگوریتم ژنتیک یکی از مشهورترین الگوریتم‌های بهینه‌سازی برای حل مسائل پیچیده با فضای حل وسیع می‌باشد. این الگوریتم برگرفته از مفاهیم زیستی می‌باشد (شواف و فوستر، ۱۹۹۶). از اصلی‌ترین ویژگی‌های الگوریتم ژنتیک موارد زیر می باشند:

الگوریتم ژنتیک روی مجموعه جواب‌هایی از فضای قابل قبول جستجو می‌کند. وسعت جستجوی عملگرهای الگوریتم ژنتیک کمک می‌کند تا به طور موثری جواب‌های کشف نشده در فضای جستجو شناسایی و تست شوند. احتمالی بودن ساختار مسئله با عملگرهای الگوریتم ژنتیک باعث می‌شود یک جواب بهتر کشف وارائه شود. وجود جمعیت‌های مختلف باعث می‌شود که احتمال گیرافتادن الگوریتم در یک نقطه بهینه محلی کاهش یابد. گام‌های مختلف الگوریتم ژنتیک به گونه‌ای است که بعد ازاجرای تمامی گام‌ها دوباره به اولین قدم برگشته و تا رسیدن به یک جواب مطلوب این عملیات تکرار می‌شود. گام اول در الگوریتم ژنتیک، تولید یک سری جواب‌های تصادفی است که جمعیت یا نسل اولیه نام دارد (تان ، کیوک و سی، ۲۰۰۷، اورسکی و همکاران، ۲۰۱۲). کیفیت نسل اولیه انتخاب شده، یک نقش بسیار مهم در کیفیت جواب نهایی مسئله دارد (یو و لیو، ۲۰۰۳). هر حلقه از الگوریتم ژنتیک شامل مراحل زیر است:

۱) کنترل کردن اینکه آیا تعداد نسل‌های تولید شده بزرگتر از محدودیت تعیین شده است یا نه. اگر بزرگتر بود با بهترین جواب بدست آمده از الگوریتم خارج می‌شود اگر نه ادامه می‌دهد.
۲) ارزیابی ارزش هر یک از افراد مطابق با تابع هدف.
۳) کنترل کردن هر یک از جواب‌های بدست آمده به صورت انفرادی که آیا به عنوان یک جواب بهینه، رضایت بخش هستند یا خیر. اگر رضایت بخش بودند دستور خروج و اگر نه ادامه می‌دهد.
۴) انتخاب بهترین جواب‌های فردی از میان جمعیت تولید شده و ساختن نسل بعدی از آنها برای اجرای عملگر الگوریتم ژنتیک که عبارتند از: تولید دوباره، احتمال جهش و احتمال تغییر در ساختار ژنتیکی برای تولید نسل بعدی صورت می‌گیرد. عملگر الگوریتم ژنتیک می‌تواند عملیات تولید نسل جدید را از ترکیب دو فرد و یا از یک فرد به تنهایی انجام دهد.
۵) عملگر تولید مجدد، از بهترین جواب‌های نسل قبلی برای تولید نسل بعدی استفاده می‌کند.
۶) احتمال جهش، در واقع احتمال تولید افراد جدید با تغییر اطلاعات از والدین انتخاب شده، برای به وجود آوردن یک فرد(فرزند) مناسب‌تر می‌باشد.
۷) یک نسل جدید تولید می‌گردد که جایگزین نسل قبلی شده و دوباره به گام اول برمی‌گردد و تمامی دستورات در حلقه اجرا شده و ادامه پیدا می‌کند.

الگوریتم ژنتیک از نوع روش‌های پوشش‌دهنده می‌باشد که با استفاده از آن می‌توان به یک جواب بهینه مطلق و یا یک جواب نزدیک به جواب بهینه رسید.

Archive of SID

www.SID.ir

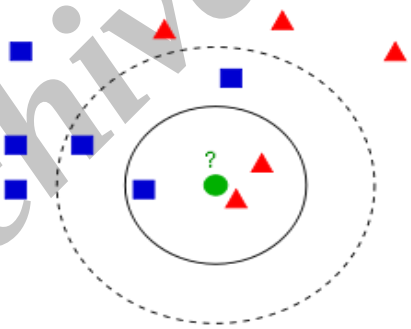
۶-۳- الگوریتم نزدیکترین همسایگی

درتشخیص الگو، الگوریتم نزدیک‌ترین همسایگی و یا K-NN یک روش غیر پارامتری مورد استفاده برای طبقه‌بندی و رگرسیون است. در هر دو مورد، ورودی شامل K نزدیک‌ترین نمونه‌های آموزشی در فضای ویژگی می‌باشد. خروجیK-NN به اینکه برای طبقه‌بندی استفاده می‌شود و یا رگرسیون بستگی دارد.

خروجی طبقه‌بندی K-NN، عضویت در کلاس است. یک شی است که با رأی اکثریت همسایگان خود طبقه‌بندی می‌شود، با هدف که هر شی به رایج‌ترین کلاس در میان K نزدیک‌ترین همسایگان خود اختصاص داده می‌شود (k یک عدد صحیح مثبت، به طور معمول کوچک است). اگر k=1 باشد، پس از آن شی به سادگی به کلاس نزدیک‌ترین همسایه واحد خود اختصاص داده می‌شود. در رگرسیون K-NN، خروجی یک مقدار ارزش (ضریب) برای شی مورد نظر می‌باشد که مقدار آن برابر با میانگین ارزش‌های k نزدیکترین همسایه تعیین شده برای آن است.

K-NN یک نوع یادگیری بر اساس مثال و یا یادگیری تنبل[۲۲] است که در آن تابع فقط به صورت محلی تقریب ومحاسبه می‌گردد. الگوریتم K-NN یکی ازساده‌ترین الگوریتم های یادگیری ماشینی می‌باشد.

در هر دو مورد طبقه‌بندی ورگرسیون، می‌توان به کمک همسایگان وزن‌دهی را انجام داد، به طوری که همسایه نزدیکتر اهمیت بیشتری نسبت به همسایه دورتر دارد. به عنوان مثال یک طرح وزن معمول عبارت است ازدادن وزن 1/d به هر یک از همسایگان، که در آن d فاصله به همسایه است. شکل شماره ۴. تصویری از نحوه عملکرد الگوریتم همسایگی را نشان می‌دهد (تکسیرا،۲۰۱۰).



شکل شماره ۴. نحوه عملکرد الگوریتم نزدیکترین همسایگی

۷-۳- متغیرهای پژوهش

هدف اصلی از این تحقیق استفاده از روش هیبرید انتخاب ویژگی و الگوریتم نزدیکترین همسایگی برای افزایش دقت پیش‌بینی جهت حرکتی شاخص ۵۰ شرکت فعال‌تر بورس اوراق بهادار تهران می‌باشد. برای انجام این کار از چندین متغیر توضیحی (شاخصه‌های تحلیل تکنیکال) برای پیش‌بینی جهت حرکتی شاخص در یک روز آینده استفاده شده است.

Archive of SID

www.SID.ir

پس از مطالعه و بررسی متغیرهای توضیحی مورد استفاده در تحقیقات گذشته در نهایت تعداد ۲۸ متغیر به عنوان متغیرهای ورودی مدل تعیین گردید. معیار مهم در انتخاب این متغیرها مفید بودن آنها و تاثیر آنها در پیش‌بینی جهت حرکت شاخص بوده است. جدول شماره ۱ نشان دهنده کلیه متغیرهای استفاده شده در این پژوهش می‌باشد.

جدول شماره ۱. متغیرهای اولیه ورودی مدل

نام متغیر	شماره متغیر	نام متغیر	شماره متغیر
Directnl movement +ID	v_{15}	Open price	v_1
-Directnl movement ID	v_{16}	Close prise	v_2
Money flow index	v_{17}	High price	v_3
RSI	v_{18}	Low price	v_4
TRIX	v_{19}	volume	v_5
CCL-Equis	v_{20}	Bollinger band 1	v_6
Volume ROC	v_{21}	Bollinger band 0	v_7
Accumulation/distribution	v_{22}	Bollinger band -1	v_8
Chakina/d oscillator	v_{23}	MACD	v_9
Price ROC	v_{24}	Momentume	v_{10}
Dynamice momentum index	v_{25}	Williams acc/dist	v_{11}
Moving average	v_{26}	Wilders smoothing	v_{12}
Stochastic oscillator	v_{27}	Volume oscillator	v_{13}
Stochastic momentum	v_{28}	Stochastic Example of hhv() function	v_{14}
		۲۸	تعداد کیله متغیرها

شاخص مورد بررسی شاخص ۵۰ شرکت فعال‌تر بورس اوراق بهادار تهران از تاریخ ۱۳۸۸/۰۱/۰۵ تا ۱۳۹۳/۰۲/۰۹ می‌باشد. در این پژوهش برای پیش‌بینی مسیر حرکتی روزانه شاخص از اعداد "۱" و "۱-" که به ترتیب نشان‌دهنده افزایش و کاهش (کمتر یا برابر بودن) عدد فردا نسبت به امروز به عنوان برچسب استفاده شده است.

۴- فرضیه‌های پژوهش

با توجه به نتایج بدست آمده از تحقیقات گذشته، این پژوهش شامل دو فرضیه می‌باشد:

"دقت کلی روش هیبرید آنالیز اجزای اساسی و الگوریتم ژنتیک بر پایه الگوریتم نزدیکترین همسایگی از لحاظ آماری با سطح معناداری ۰/۹۵ از عملکرد بالاتری نسبت به روش‌های انتخاب ویژگی فیلترکنندهٔ زنجیره اطلاعات، رلیف و روش آنالیز اجزای اساسی برخوردار می‌باشد"

Archive of SID

www.SID.ir

"دقت کلی روش هیبرید آنالیز اجزای اساسی و الگوریتم ژنتیک بر پایه الگوریتم نزدیکترین همسایگی از لحاظ آماری با سطح معناداری ۰/۹۵ از عملکرد بالاتری نسبت به روش انتخاب ویژگی پوشش دهندهٔ الگوریتم ژنتیک برخوردار می‌باشد"

۵- نتایج پژوهش

در این قسمت یافته‌های تحقیق و تحلیل آنها ارائه می‌شود. کلیه نتایج بدست آمده برای دقت پیش‌بینی مدل‌های بکار برده شده در جدول‌های زیر نشان داده شده‌اند. جدول شماره ۲. نشان دهنده نتایج حاصل شده از روش انتخاب ویژگی زنجیره اطلاعات و الگوریتم نزدیکترین همسایگی می‌باشد.

جدول شماره ۲- نتایج حاصل از روش رنجیره اطلاعات و الگوریتم نزدیکترین همسایگی

جهت حرکتی شاخص		واقعیت	
		افزایشی	کاهشی
پیش‌بینی	افزایشی	۴۵۸	۲۰۳
	کاهشی	۲۱۱	۳۳۹

جدول شماره ۳. نشان دهنده نتایج حاصل شده از روش انتخاب ویژگی رلیف و الگوریتم نزدیکترین همسایگی می‌باشد.

جدول شماره ۳- نتایج حاصل از روش رلیف و الگوریتم نزدیکترین همسایگی

جهت حرکتی شاخص		واقعیت	
		افزایشی	کاهشی
پیش‌بینی	افزایشی	۴۵۲	۲۰۶
	کاهشی	۲۱۷	۳۳۶

جدول شماره ۴. نشان دهنده نتایج حاصل شده از روش انتخاب ویژگی آنالیز اجزای اساسی و الگوریتم نزدیکترین همسایگی می‌باشد.

جدول شماره ۴. نتایج حاصل از روش آنالیز اجزای اساسی الگوریتم نزدیکترین همسایگی

جهت حرکتی شاخص		واقعیت	
		افزایشی	کاهشی
پیش‌بینی	افزایشی	۴۷۳	۲۰۷
	کاهشی	۱۹۶	۳۳۵

Archive of SID

www.SID.ir

جدول شماره ۵. نشان دهنده نتایج حاصل شده از روش انتخاب ویژگی الگوریتم ژنتیک و الگوریتم نزدیکترین همسایگی می‌باشد.

جدول شماره ۵- نتایج حاصل از روش الگوریتم ژنتیک و الگوریتم نزدیکترین همسایگی

جهت حرکتی شاخص		واقعیت	
		افزایشی	کاهشی
پیش‌بینی	افزایشی	۴۷۱	۲۰۰
	کاهشی	۱۹۸	۳۴۲

جدول شماره ۶. نشان دهنده نتایج حاصل شده از روش انتخاب ویژگی هیبرید و الگوریتم نزدیکترین همسایگی می‌باشد.

جدول شماره ۶- نتایج حاصل از روش هیبرید و الگوریتم نزدیکترین همسایگی

جهت حرکتی شاخص		واقعیت	
		افزایشی	کاهشی
پیش‌بینی	افزایشی	۴۸۵	۱۸۷
	کاهشی	۱۸۴	۳۵۵

جدول شماره ۷. نشان دهنده نتایج حاصل شده از الگوریتم نزدیکترین همسایگی بدون انتخاب ویژگی می‌باشد.

جدول شماره ۷- نتایج حاصل از الگوریتم نزدیکترین همسایگی

جهت حرکتی شاخص		واقعیت	
		افزایشی	کاهشی
پیش‌بینی	افزایشی	۴۶۱	۲۱۷
	کاهشی	۲۰۸	۳۲۵

برای اینکه دقت پیش‌بینی مدل‌های استفاده شده به درستی با هم مقایسه شوند، مقادیر $\alpha(1)$، $\alpha(-1)$، $\beta(1)$ و $\beta(-1)$ برای هر یک از مدل‌ها به صورت جداگانه محاسبه شد و نتایج حاصل شده از آن در جدول شماره ۸. ارائه گردید. همچنین جدول شماره ۹. دقت کلی بدست آمده از هر یک از مدل‌های استفاده شده در این پژوهش را نشان می‌دهد.

Archive of SID

www.SID.ir

$$\alpha(1)=\frac{\text{تعداد روزهایی که شاخص جهت حرکتی افزایشی داشته و درست پیش‌بینی شده است}}{\text{تعداد کل روزهایی که شاخص جهت حرکتی افزایشی داشته}}$$

$$\alpha(-1)=\frac{\text{تعداد روزهایی که شاخص جهت حرکتی کاهشی داشته و درست پیش‌بینی شده است}}{\text{تعداد کل روزهایی که شاخص جهت حرکتی کاهشی داشته}}$$

$$\beta(1)=\frac{\text{تعداد روزهایی که شاخص جهت حرکتی افزایشی داشته و درست پیش‌بینی شده است}}{\text{تعداد کل روزهایی که جهت حرکتی شاخص افزایشی پیش‌بینی شده است}}$$

$$\beta(-1)=\frac{\text{تعداد روزهایی که شاخص جهت حرکتی کاهشی داشته و درست پیش‌بینی شده است}}{\text{تعداد کل روزهایی که جهت حرکتی شاخص کاهشی پیش‌بینی شده است}}$$

جدول شماره ۸- نتایج همه جانبه بدست آمده برای هر یک از مدل‌ها

روش انتخاب ویژگی	مدل طبقه‌بندی کننده	$\alpha(1)\times100$	$\alpha(-1)\times100$	$\beta(1)\times100$	$\beta(-1)\times100$
زنجیره اطلاعات	الگوریتم نزدیکترین همسایگی	۶۸/۴۶	۶۲/۵۵	۶۹/۲۹	۶۴/۶۱
رلیف	الگوریتم نزدیکترین همسایگی	۶۷/۵۶	۶۱/۹۹	۶۸/۶۹	۶۰/۷۶
آنالیز اجزای اساسی	الگوریتم نزدیکترین همسایگی	۷۰/۷۰	۶۱/۸۱	۶۹/۵۶	۶۳/۰۹
الگوریتم ژنتیک	الگوریتم نزدیکترین همسایگی	۷۰/۴۰	۶۳/۱۰	۷۰/۱۹	۶۳/۳۳
هیبرید	الگوریتم نزدیکترین همسایگی	۷۲/۵۰	۶۵/۵۰	۷۲/۱۷	۶۵/۸۶
——	الگوریتم نزدیکترین همسایگی	۶۸/۹۱	۵۹/۹۶	۶۷/۹۹	۶۰/۹۸

جدول شماره ۹- دقت کلی حاصل شده از مدل‌های مختلف به کار برده شده

روش انتخاب ویژگی	الگوریتم طبقه‌بندی کننده	میانگین دقت کلی
زنجیره اطلاعات	الگوریتم نزدیکترین همسایگی	۸۲/۶۵
رلیف	الگوریتم نزدیکترین همسایگی	۶۵/۰۸
آنالیز اجزای اساسی	الگوریتم نزدیکترین همسایگی	۶۶/۷۷
الگوریتم ژنتیک	الگوریتم نزدیکترین همسایگی	۱۳/۶۷
آنالیز اجزای اساسی+الگوریتم ژنتیک(هیبرید)	الگوریتم نزدیکترین همسایگی	۳۷/۶۹
—	الگوریتم نزدیکترین همسایگی	۵۷/۶۳
بالاترین دقت بدست آمده		**۳۷/۶۹**

Archive of SID

www.SID.ir

همانطور که جدول شماره ۸. نشان می‌دهد، روش هیبرید انتخاب ویژگی و الگوریتم نزدیکترین همسایگی در تمامی موارد نسبت به دیگر روش‌های استفاده شده، از دقت بالاتری برخوردار است. همچنین نتایج بدست آمده نشان می‌دهد، تمامی مدل‌های استفاده شده در این پژوهش جهت حرکتی افزایشی شاخص ۵۰ شرکت فعالتر را با دقت بالاتری نسبت به جهت حرکتی کاهشی آن پیش‌بینی می‌کنند. جدول شماره ۹. که در آن دقت کلی پیش‌بینی هر یک از مدل‌ها ارائه گردیده است، نشان از برتری روش هیبرید انتخاب ویژگی و الگوریتم نزدیکترین همسایگی دارد.

برای بررسی فرضیه‌های پژوهشی مبنی برعملکرد بهتر روش هیبرید آنالیز اجزای اساسی و الگوریتم ژنتیک بر پایه الگوریتم نزدیکترین همسایگی نسبت به روش‌های الگوریتم ژنتیک، زنجیره اطلاعات، رلیف و آنالیز اجزای اساسی از آزمون مقایسات زوجی استفاده گردیده است. در این آزمون عملکرد بهتر روش هیبرید از نظر آماری بررسی شده است. جدول شماره ۱۰. نشان دهنده نتایج بدست آمده از آزمون مقایسات زوجی می‌باشد.

جدول شماره ۱۰- نتایج بدست آمده از آزمون مقایسات زوجی

روش انتخاب ویژگی	الگوریتم ژنتیک		آنالیز اجزای اساسی		رلیف		زنجیره اطلاعات	
روش هیبرید	بررسی فرضیه پژوهشی	p-value	بررسی فرضیه پژوهشی	p-value	بررسی فرضیه پژوهشی	p-value	بررسی فرضیه پژوهشی	p-value
	پذیرفته می‌شود	۰/۰۴۸	پذیرفته می‌شود	۰/۰۲۳	پذیرفته می‌شود	۰/۰۰۸	پذیرفته می‌شود	۰/۰۱۱

همانطور که جدول شماره ۱۰. نشان می‌دهد، می‌توان گفت میانگین دقت کلی روش ترکیبی آنالیز اجزای اساسی و الگوریتم ژنتیک (روش هیبرید) در ترکیب با الگوریتم نزدیکترین همسایگی با سطح اطمینان ۹۵ درصد، از دقت بالاتری نسبت به روش‌های زنجیره اطلاعات، رلیف و آنالیز اجزای اساسی که از خانواده روش‌های فیلترکننده بوده و روش الگوریتم ژنتیک که جزو روش‌های پوشش‌دهنده است، برخوردار می‌باشد. بنابراین فرضیه‌های پژوهشی پذیرفته می‌شوند.

۶- نتیجه‌گیری و بحث

سرمایه‌گذاری در بازار سهام یکی از جذاب‌ترین فعالیت‌های سرمایه‌گذاری در جهان می‌باشد، اگرچه پیش‌بینی دقیق بازار سهام به دلیل وضعیت سیاسی، اقتصاد جهانی و سایر عوامل معمولاً کار بسیار دشواری است ولی بدون داشتن توانایی در پیش‌بینی قیمت سهام سرمایه‌گذاری و کسب سود کار بسیار سختی خواهد بود. پژوهشگران، سرمایه‌گذاران و معامله‌گران همواره علاقه‌مند به پیش‌بینی جهت حرکتی قیمت سهام و یا شاخص‌های مختلف هستند که به همین دلیل تاکنون روش‌ها و تکنیک‌های بسیاری با هدف افزایش دقت پیش‌بینی ارائه گردیده است. با پیشرفت‌های اخیر در حوزه هوش مصنوعی و به وجود آمدن

Archive of SID

www.SID.ir

الگوریتم‌ها و روش‌های جدید در زمینه پیش‌بینی، افزایش دقت برای رسیدن به یک دقت مطلوب میسر شده است.

تحقیقات بسیاری در زمینه پیش‌بینی جهت حرکتی قیمت با به کارگیری از محاسبات نرم و روش‌های مختلف داده‌کاوی صورت گرفته که یکی از جدیدترین روش‌ها، استفاده از روش‌های انتخاب ویژگی هستند که برای انتخاب تأثیرگذارترین ویژگی‌ها در افزایش دقت پیش‌بینی می‌باشند زیرا در مسائل پیش‌بینی جهت حرکتی قیمت سهام و شاخص دقت پیش‌بینی تحت تأثیر تعداد ویژگی‌ها و متغیرهای ورودی مدل قرار می‌گیرد. از شاخصه‌های تحلیل تکنیکال و تحلیل بنیادی به صورت گسترده‌ای در پیش‌بینی جهت حرکتی قیمت سهام و شاخص‌های مختلف بازار پول و سرمایه استفاده می‌شود و پژوهشگران همواره به دنبال انتخاب شاخصه‌هایی می‌باشند که با کمک آنها بتوانند دقت پیش‌بینی را افزایش دهند. تحقیقات صورت گرفته توسط محققان نشان می‌دهد، استفاده از یک روش انتخاب ویژگی مناسب در ترکیب با یک مدل طبقه‌بندی کننده می‌تواند افزایش قابل توجهی در دقت پیش‌بینی حاصل نماید.

می‌توان به برخی از مزایای استفاده از انتخاب ویژگی به صورت زیر اشاره نمود.

۱) حذف ویژگی‌های غیرمرتبط و زائد و همچنین انتخاب مهمترین ویژگی‌ها که مدل طبقه‌بندی کننده از آن‌ها برای پیش‌بینی استفاده می‌نماید.
۲) استفاده از انتخاب ویژگی همچنین می‌تواند بالاترین دقت پیش‌بینی را حاصل نماید.
۳) استفاده از انتخاب ویژگی باعث آموزش مدل طبقه‌بندی کننده با استفاده از ویژگی‌هایی که از حساسیت بالاتری در پیش‌بینی برخوردار هستند، می‌شود.
۴) انتخاب ویژگی با کاهش ابعاد مسئله و در نتیجه باعث کاهش پیچیدگی‌های محاسباتی می‌گردد.

بیشتر تحقیقات انجام شده تنها از یک روش انتخاب ویژگی برای تعیین زیرمجموعه بهینه از ویژگی‌ها به منظور پیش‌بینی قیمت سهام استفاده می‌کنند. در پژوهش حاضر یک روش ترکیبی جدید انتخاب ویژگی که ترکیبی از دو روش آنالیز اجزای اساسی و الگوریتم ژنتیک می‌باشد، ارائه گردید و عملکرد آن با روش‌های زنجیره اطلاعات، رلیف، آنالیز اجزای اساسی و الگوریتم ژنتیک مقایسه شد. نتایج حاصله حاکی از این واقعیت می‌باشد که با استفاده از روش هیبرید آنالیز اجزای اساسی و الگوریتم ژنتیک و ترکیب آن با الگوریتم نزدیکترین همسایگی می‌توان جهت حرکتی شاخص ۵۰ شرکت فعال‌تر را با دقت بالاتری نسبت به دیگر روش‌های استفاده شده در این پژوهش پیش‌بینی نمود.

به عنوان پیشنهادی برای تحقیقات آتی می‌توان از روش شبکه عصبی، ماشین بردار پشتیبان و یا رگرسیون بردار پشتیبان به عنوان مدل طبقه‌بندی کننده و ترکیب آن با روش جدید انتخاب ویژگی استخراج شده از این پژوهش برای پیش‌بینی قیمت سهام، شاخص کل بورس اوراق بهادار تهران و پرتفلیو-های مختلف استفاده نمود.

Archive of SID

www.SID.ir

فهرست منابع

* سینایی حسنعلی، مرتضویسعید الله، تیموری اصل یاسر. (۱۳۸۴)، پیش‌بینی شاخص بورس اوراق بهادار تهران با استفاده از شبکه‌های عصبی، بررسی های حسابداری و حسابرسی، سال دوازدهم، شماره۴۱، صص ۵۹-۸۳.
* عبادی، ا. (۱۳۸۸). "پیش‌بینی قیمت شاخص کل سهام در بازار بورس تهران با استفاده از شبکه‌های عصبی مصنوعی. پایان‌نامه کارشناسی ارشد، دانشکده اقتصاد و علوم اجتماعی دانشگاه بوعلی سینا، همدان.
* عبده تبریزی، حسین. جوهری، هادی. (۱۳۷۵). بررسی کارامدی شاخص بورس اوراق بهادار تهران، تحقیقات مالی، ۳ (۲):۴۷-۶۱.
* فلاح‌پور سعید، گل ارضی غلامحسین، فتوره چیان ناصر. (۱۳۹۲). پیش‌بینی روند حرکتی قیمت سهام با استفاده از ماشین بردار پشتیبان بر پایه ژنتیک در بورس اوراق بهادار تهران، تحقیقات مالی، ۱۵(۲)، صص ۲۶۹ -۲۸۸.
* هاشمی احمد. (۱۳۸۹). تاثیر فاکتورهای رفتاری بر پیش‌بینی قیمت سهام با استفاده از مدل شبکه‌های عصبی رگرسیونی جلوسو، پایان نامه کارشناسی ارشد. دانشکده صنایع دانشگاه علم و فرهنگ، تهران.
* منجمی، سید امیر حسین، ابزری مهدی، رعیتی شوازی علیرضا.(۱۳۸۸). پیش‌بینی قیمت سهام در بازار بورس اوراق بهادار با استفاده از شبکه عصبی مصنوعی. فصلنامه اقتصاد مالی، ۶ (۳)،۱-۲۶.
* میرفیض فلاح شمس، دلنواز اصغری، بیتا. (۱۳۸۸)، پیش‌بینی شاخص بورس اوراق بهادار تهران با استفاده از شبکه های عصبی، فراسوی مدیریت، سال سوم، شماره ۹، صص، ۱۹۱-۲۱۲.
* Afolabi, M. O. Olude, O. (2007). "Predicting stock prices using a hybrid self organizing map (SOM)", In Proceedings of the 40th Hawaii international conference on system sciences (p.48).
* Altman, N. S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression",The American Statistician, **46** (3), PP. 175–185.
* Anna, B. (2007)."Should normal distribution be normal? The Student's T alternative", ComputerInformation Systems and Industrial Management Applications, PP. 3–8.
* Atsalakis, G. S., Valavanis, K. P. (2009). "Surveying stock market forecasting techniques – Part II: Softcomputing methodsT",Expert Systems with Applications, 36(3), PP. 5932–5941.
* Bao, D. Yang, Z. (2008). "Intelligent stock trading system by turning point confirming and probabilistic reasoning",Expert Systems with Applications, 34, PP. 620–627.
* Bollerslev, T. (1986), "Generalized autoregressive conditional heteroscedasticity", Journal of Econometrics, 31, PP. 307–327.
* Box, G. Jenkins, G. (1976), "Time Series Analysis: Forecasting and Control, Holden-Day", San Francisco.
* Cao, L. J. Tay, F. E. H. (2003), "Support vector machine with adaptive parameters infinancial time series forecasting",IEEE Transactions on Neural Networks, 14(6), PP. 1506–1518.
* Dash, M. et al., (2002),"Feature selection for clustering – a filter solution", In Proceedings of the second international conference on data mining , PP. 115–122.

Archive of SID

www.SID.ir

* Dy, J. G.Brodley, C. E. (2000),"Feature subset selection and order identification for unsupervised learning". In Proceedings of the 17th international conference on machine learning, PP. 247–254.
* Engle, R.F. (1982),"Autoregressive conditional heteroscedasticity with estimator of the variance of United Kingdom inflation", Econometrica, 50 (4), PP. 987–1008.
* Fama, E. F. (1970),"Efficient capital markets: A review of theory and empirical work,Proceedings of the Twenty-Eighth Annual Meeting of the American Finance Association",Journal of Finance, 25(2), PP. 383–417.
* GharehMohammadi, F. SanieeAbadeh, M. (2014)."Image steganalysis using a bee colony based feature selection algorithm". Engineering Applications of Artificial Intelligence, 31, PP. 35–43.
* Hall, M. A. (2000),"Correlation-based feature selection for discrete and numeric class machine learning",In Proceedings of the 17th international conference on machine learning, PP. 359–366.
* Huang, C. Yang, D. Chuang, Y. (2008)."Application of wrapper approach and composite classifier to the stock trend prediction", Expert System with Application, 34, PP. 2870–2878.
* Kim, K. Han, I. (2000), "Genetic algorithms approach to feature discretization in artificial neural networks for prediction of stock index", Expert System with Application, 19, PP. 125–132.
* Kim, H. Shin, K. (2007),"A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets",Applied Soft Computing, 7, 569–576.
* Kimoto, T. Asakawa, K. Yoda, M. Takeoka, M. (1990),"Stock market prediction system with modular neural network", in: Proceedings of the International Joint Conference on Neural Networks, San Diego, California, PP. 1–6.
* Kohavi, R. John, G. H. (1997),"Wrappers for feature subset selection. Artificial Intelligence", 97(1–2), PP. 273–324.
* Kwon, Y. Moon, B. (2007),"A hybrid neurogenetic approach for stock forecasting", IEEE Transactions on Neural Networks, 18(3), PP. 851–864.
* Lawrence, S. Giles, C. L. Tsoi, A. C. (1997),"Lessons in neural network training: Overfitting may be harder than expected",In Proceedings of the fourteenth national conference on artificial intelligence, AAAI-97, PP. 540–545.
* Los, C. A. (2000), "Nonparametric efficiency testing of Asian markets using weekly Data",Advances in Econometrics, 14, PP. 329–363.
* Min, J. H. Lee, Y.-C. (2005),"Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters",Expert Systems with Applications, 28(4), PP. 603–614.
* Nagarajan, V. Wu, Y. Liu, M., & Wang, Q. (2005),"Forecast studies for financial markets using technical analysis",In International Conference on Control and Automation *(ICCA)*, PP. 259–264.
* Nair, B.B., Mohandas, V.P. & Sakthivel, N.R. (2010). "A Genetic Algorithm Optimized Decision Tree-SVM based Stock Market Trend Prediction System", (IJCSE) International Journal on Computer Science and Engineering, 2 (9): 2981-2988.
* Nanni, L. (2006),"Multi-resolution subspace for financial trading", Pattern Recognition Letters, 27, PP. 109–115.
* Nikolopoulos, C. Fellrath, P. (1994),"A hybrid expert system for investment advising", Expert Systemswith applications, 11 (4), PP. 245–250.
* Ni,L. g. Ni, Z. Gao, W. Y. (2011), "Stock trend prediction based on fractal feature selection and support vector machine", Expert Systems with Applications, 38, PP. 5569-5576.

Archive of SID

www.SID.ir

* Oreski, S, Oreski, D.Oreski, G. (2012),"Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment", Expert systems with applications, 39(16), PP. 12605-12617
* Sai, Y. Yuan, Z. (2007),"Mining stock market tendency by RS-based support vector machines",IEEE International Conference on Granular Computing, 659–664.
* Shoaf, J. S. Foster, J. A. (1996),"A genetic algorithm solution to the efficient set problem: A technique for portfolio selection based on the Markowitz model", In Proceedings of the decision sciences institute annual meeting, Orlando, Florida, PP.571573.
* Tan, T. Z. Quek, C. See, Ng. G. (2007),"Biological brain-inspired genetic complementarylearning for stock market and bank failure prediction",Computational Intelligence, 23(2), PP. 236 -261.
* Teixeira, L.A. Oliveira.A .L. (2010), "A method for automatic stock trading combining technical analysis and nearest neighbor classification".Expert Systems with Applications, 37 (2010), PP. 6885–6890.
* Tsai, C. F. Hsiao, Y. C. (2010),"Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches", Decision Support Systems, 50, PP. 258-269.
* Vanstone, B. Tan, C. (2003),"A survey of the application of soft computing to investment and financial trading",In Proceedings of the Australian and New Zealand intelligent information systems conference, PP. 211–216.
* White, H. (1988),"Economic prediction using neural networks: A case of IBM daily stock returns". IEEE International Conference on Neural Networks, 2, PP. 451–458.
* Yu, L. Liu, H., (2003), "Feature selection for high-dimensional data: A fast correlation-based filter solution",In Proceedings of the 20th international conference on machine learning, PP. 856–863.

یادداشت‌ها

1-Feature Selection
2- Soft Computing
3- Combined Model
4- Classifier
5- Hybrid
6-Principal Component Analysis
7- Information Gain
8- Relif
9- Filter Method
10- Wrapper Method
11-BPNN: Back-propagation neural network.
12- PCA: Principal Component Analysis
13- SVM: Support vector machine.
14- FA: Factor Analysis.
15- DWT: Discrete wavelet transform.
16-ESN: Echo state network.
17- RNN: Recurrent neural network.
18-SOFM: Self-organizing feature map.
19-SVR: Support vector regression.
20- Filter methods
21-Wrapper methods
22- Lazy learning

Archive of SID

www.SID.ir