ISSN: 2538-3663
http://www.Joas.ir

پژوهش های کاربردی در مدیریت و حسابداری (سال سوم)
شماره ۱۰ / تابستان ۱۳۹۷ / ص ۱۵۱ – ۱۳۷

مدل سازی پیش بینی قیمت سهام با استفاده از تئوری مجموعه راف و الگوریتم ژنتیک

مجتبی صدیقی [۱]، حسین جهانگیرنیا [۲]
[۱] باشگاه پژوهشگران جوان و نخبگان، واحد قم، دانشگاه آزاد اسلامی، قم، ایران.
[۲] استادیار گروه مالی و حسابداری، دانشگاه آزاد اسلامی، قم، ایران.

نام نویسنده مسئول:
حسین جهانگیرنیا

چکیده

این مقاله مدل جدیدی را مطرح می کند که بر مبنای چهار روش جدید CDPA, MEPA RST و GA می باشد تا به بهبود عملکرد پیش بینی بازار سهام بپردازد. تجزیه و تحلیل تکنیکال به عنوان روشی کارآمد برای پیش بینی قیمت های سهام در بازار سرمایه می باشد. مدل هیبریدی پیش بینی کننده قیمت های سهام با استفاده از ترکیب چهار روش ذکر شده و شاخص های تکنیکال چندگانه برای پیش بینی دقیق روند بازار سهام ارائه می شود.

کارایی مدل مطرح شده بر مبنای دو نوع ارزیابی عملکرد، صحت و درآمد حاصل از سهام در مجموعه داده TEPIX برای یک دوره زمانی هفت ساله برای ۵۰ شرکت برتر بورس اوراق بهادار تهران محاسبه شده و به عنوان مجموعه داده آزمایشی می باشد. نتایج تجربی نشان می دهد که مدل مطرح شده برتر از مدل های پیش بینی شده دیگر، از نقطه نظر دقت بوده و ارزیابی مربوط به سود سرمایه نشان می دهد که منافع ایجاد شده توسط مدل پیشنهادی بیشتر از سه مدل دیگر می باشد.

واژگان کلیدی: شاخص های تکنیکال، قیمت سهام، نظریه مجموعه راف، الگوریتم ژنتیک، روش توزیع احتمال تجمّعی، روش اصل آنتروپی کمینه.

ISSN: 2538-3663
http://www.Joas.ir

مقدمه

در بازار سهام، بسیار مشکل می باشد تا به پیش بینی روند سهام بپردازیم که این به دلیل فاکتورهای پیچیده ای می باشد که بازارهای سهام و روابط غیر خطی را تحت تاثیر قرار داده که در بین دوره های متفاوتی از قیمت سهام مد نظر قرار می گیرد. روش های پیش بینی بیشماری به کار گرفته شده است تا به پیش بینی قیمت سهام بپردازد. [1,2]

در حوزه پیش بینی بازار سهام، روش تحلیل تکنیکال به عنوان یکی از روش های تحلیلی مهمی می باشد که توسط سرمایه گذاران برای گرفتن تصمیمات سرمایه گذاری مورد استفاده قرار می گیرد، و بسیاری از محققان تمرکز خود را بر روی تحلیل تکنیکال به منظور بالا بردن سود سرمایه گذاری شان قرار می دهند. علاوه بر این روش تحلیل تکنیکال دارای قابلیت پیش بینی مسیر قیمت های آینده با بررسی داده های بازاری در گذشته، به ویژه قیمت سهام و مقدار آن می باشد. روش تحلیل تکنیکال بر این فرض می باشد که قیمت سهام و مقدار آن به عنوان دو فاکتور مهم در تعیین مسیر آینده در رفتار سهام خاص یا بازار و شاخص های فنی که حاصل فرمول های ریاضی بر مبنای قیمت سهام و مقدار آن می باشد، که می تواند برای پیش بینی نوسانات قیمت و همچنین ایجاد داده برای سرمایه گذاران، و توانمند ساختن آن ها برای تعیین زمان بندی خرید یا فروش سهام به کار گرفته شود. [3-5]

علاوه بر روش های تحلیل تکنیکال، بسیاری از مدل های پیش بینی عددی متعارف توسط محققان مالی مطرح شده است که شامل مدل ناهماهنگی پراکنش مشروط اتورگرسیو انگل (ARCH)، مدل عمومی ARCH (GARCH) بولراسلو، مدل میانگین متحرک اتورگرسیو (ARMA) باکس و جنکین، و مدل میانگین متحرک یکپارچه اتورگرسیو (ARIMA) می باشند. [6-8]

در دهه های اخیر، بسیاری از محققان از روش های دیگری برای پیش بینی مالی استفاده می کنند: که شامل الگوریتم های هوش مصنوعی در سال ۱۹۹۰ می باشد، کینوتو و همکارانش با استفاده از شبکه های عصبی، سیستم پیش بینی را برای بازار سهام ایجاد کرده اند. نیکولوپولوس و فلراس به ادغام الگوریتم های ژنتیک و شبکه عصبی برای ایجاد یک سیستم کارشناسی هیبریدی برای تصمیمات سرمایه گذاری پرداخته اند. کیم و هان یک روش الگوریتم ژنتیک را به منظور مد نظر قرار دادن گسسته سازی و تعیین میزان ارتباط برای شبکه های عصبی مصنوعی به منظور پیش بینی شاخص قیمت سهام مطرح کرده اند. هارنگ و یو یک شبکه عصبی توسعه پسین را برای ایجاد روابط فازی در سری زمانی فازی برای پیش بینی قیمت های سهام به کار گرفته اند. رو به ادغام شبکه عصبی و مدل سری زمانی برای پیش بینی فراریت شاخص قیمت سهام پرداخته است. [9-14]

از تحقیقاتی که در بالا مطرح شده است، به هر حال سه مانع در روش های پیش بینی و مدل ها مد نظر قرار می گیرد:۱) تحلیلگران بازار سهام و مدیران سرمایه شاخص های فنی مختلفی را برای پیش بینی روند بازار سهام بر مبنای تجربه شخصی شان بکار می گیرند، که منجر به قضاوت های نادرست علائم موجود در بازار می گردد. ۲) در ارتباط با اکثر روش های آماری، فرضیاتی در ارتباط با متغیرهای بکارگرفته شده در تحلیل ها وجود دارد، که نمی تواند برای مجموعه داده هایی بکارگرفته شود که توزیع آمار را دنبال نمی کنند. و ۳) شبکه های عصبی مصنوعی به عنوان روش جعبه سیاه می باشند، و قوانین حاصل شده از ان ها به آسانی قابل درک نمی باشد. [15,16]

برای بهبود مدل های پیش بینی گذشته، یک مدل اصلاح شده می بایست دارای این قابلیت باشد تا بر روی موانعی که شامل مدل های قبلی می باشد؛ غلبه کرده و می بایست روش های مناسبی را ارائه دهد که به اسانی توسط سرمایه گذار مورد استفاده قرار می گیرد. بنابراین این مقاله یک مدل پیش بینی کننده هیبریدی را برای اصلاح روش های گذشته در پیش بینی قیمت سهام مطرح کرده و چهار روش جدید را در پیش بینی مراحل ایجاد می کند: ۱) انتخاب شاخص های فنی ضروری با استفاده از ماتریس همبستگی؛ ۲) استفاده از CPDA (روش توزیع احتمال جمع پذیر) و MEPA (روش اصولی انتروپی حداقل برای تشخیص مختصه های شرطی) و خصوصیات تصمیم گیری (نوسانات قیمت روزانه)؛ ۳) به کارگیری نظریه مجموعه ها برای ایجاد قوانینی حاصل از ارزش های زبانی شاخص های فنی؛ و ۴) بکارگیری الگوریتم های ژنتیک (GAs) برای اصلاح قوانین حاصل شده برای بهبود دقت پیش بینی و سود سهام.[17]

از لحاظ تجربی، این مقاله دو نوع از پایگاه داده سهام را (شاخص سهام و قیمت سهام مجزا) بر مبنای پایگاه داده آزمایشی به کار می گیرد. با تطبیق مدل، می توان نشان داد که مراحل اصلاحی در بهبود دقت پیش بینی موثر بوده و بر مبنای شواهد، یک سرمایه گذار و تحلیل گر سهام می تواند مراحل اصلاحی مطرح شده را در این مقاله برای بهبود مدل ها و ابزارهای پیش بینی، به کار گیرد.

۱. استراتژی اجزای مختلف مدل

این بخش به مرور فعالیت های مربوطه برای تحلیل فنی، روش توزیع احتمال جمع پذیر، روش اصلی انتروپی حداقل، مجموعه نظریه ها، و الگوریتم های ژنتیک می پردازد.

ISSN: 2538-3663
http://www.Joas.ir

۱-۱ تحلیل تکنیکال

تحلیل تکنیکال تلاشی به منظور پیش بینی جا به جایی قیمت سهام آینده با تجزیه و تحلیل توالی گذشته از قیمت های سهام می باشد. ان تکیه ای بر روی نمودارها داشته و به دنبال پیکره بندی خاصی می باشد که به نظر می رسد دارای ارزش پیش بینی باشد. تحلیل گران تمرکزشان را بر روی روان شناسی سرمایه گذار قرار می دهند، که واکنش های معمول سرمایه گذار را نسبت به شکل گیری قیمت خاص و جا به جایی های قیمت برای تحلیل نوسانات بازار سهام، نشان می دهند. قیمتی که سرمایه گذاران تمایل دارند خرید و فروش انجام دهند بستگی به انتظارات فردی دارد. درک این مفهوم بسیار مهم می باشد که متقاضیان بازار رشد اینده را پیش بینی کرده و اقدامات به موقعی انجام می دهند که منجر به جا به جایی قیمت می گردد. از این رو بسیاری از محققان تمرکزشان را بر روی تحلیل تکنیکال به منظور بهبود منافع سرمایه گذاری انجام می دهند.[18,19]

۱-۲روش توزیع احتمالی جمع پذیر (CDPA)

احتمال اشاره ای به بررسی تصادفی بودن و عدم قطعیت دارد. در هر شرایط، یکی از چند احتمال روی می دهد. نظریه احتمال روش هایی را برای تعیین شانس، احتمالات، ایجاد کرده که مرتبط به نتایج مختلف می باشد. چون توزیع احتمال بر روی خط حقیقی به صورت فاصله نیم باز p(a, b] میباشد، بنابراین $F(b) - F(a) \text{ if } a < b$ می باشد. توزیع احتمال متغیرهای تصادفی با مقدار حقیقی X کاملا توسط تابع توزیع جمع پذیر (CDF) مشخص می گردد. برای هر عدد X، CDF هر X توسط فرمول زیر بدست می آید: [20]

$$x \to F_X(x) = P(X \leqslant x) \quad \forall x \in \Re \qquad (۱)$$

که در این فرمول بخش سمت راست نشان دهنده احتمال (p) می باشد که متغیر تصادفی X مقداری کمتر یا برابر با x می پذیرد. F با حرف بزرگ برای نشان دادن تابع توزیع جمع پذیر مورد استفاده قرار می گیرد که در مقایسه با f حروف کوچک که تابع چگالی احتمال و تابع مقدار احتمال را نشان می دهد، قرار می گیرد. CDF یا x از نقطه نظر تابع چگالی احتمال f به صورت زیر تعریف می گردد: [21,22]

$$F(x) = P[X \leqslant x] = \int_{\infty}^{x} f(t)dt \qquad (۲)$$

معکوس CDF نرمال توسط پارامترهای μ و σ با مد نظر قرار گرفتن احتمالات نظری به نظیر در p محاسبه می شود، به صورتی که μ نشان دهنده میانگین و σ نشان دهنده انحراف معیار داده می باشد. تابع معکوس استاندارد از نقطه نظر CDF استاندارد به صورت زیر تعریف می گردد:[23,24]

$$x = F^{-1}(p|\mu, \sigma) = \{x : F(x|\mu, \sigma) = p\} \qquad (۳)$$

که

$$p = F(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty^e}^{x} \frac{-(t-\mu)^2}{2\sigma^2} dt \qquad (۴)$$

احتمال جمع پذیر توزیع نرمال برای تعیین فواصل مورد استفاده قرار می گیرد.

۱-۳ روش اصلی انتروپی حداقل (MEPA)

هدف تجزیه و تحلیل انتروپی حداقل تعیین محتوای اطلاعاتی در پایگاه داده مورد نظر می باشد. انتروپی توزیع احتمال بر مبنای اندازه گیری مبهم بودن توزیع می باشد. برای تقسیم کردن داده به تابع عضویت، ایجاد نقاط تقطیع شده بین مجموعه داده ها مورد نیاز می باشد. نقاط قطعه بندی شده از طریق روش گزینش حداقل انتروپی مد نظر قرار می گیرد؛ سپس مراحل قطعه بندی را با تقسیم کردن ان به دو دسته آغاز کنید. بنابراین، پارتیشن بندی تکراری با توجه به مقدار محاسبه نقاط تقسیم شده این امکان را به ما می دهد تا به پارتیشن بندی پایگاه داده به تعدادی از مجموعه های فازی بپردازیم. در سال های اخیر، MEPA در مشکلات مرتبط به پیش بینی مورد استفاده قرار گرفته است. [25,25,26]

فرض کنید مقدار نقطه تقسیم برای نمونه هایی در محدوده بین x_1 و x_2 مد نظر قرار می گیرد. معادله انتروپی برای مناطق $[x,x_1]$ و $[x,x_2]$ نوشته شده، و نشان دهنده اولین منطقه p و دومین منطقه q می باشد. انتروپی هر مقدار x به صورت زیر نوشته می شود: [27,28]

$$S(x) = p(x)S_p(x) + q(x)S_q(x) \qquad (۵)$$

ISSN: 2538-3663
http://www.Joas.ir

که

$$S_p(x) = -[p_1(x)\ln p_1(x) + p_2(x)\ln p_2(x)] \quad (۶)$$

$$S_q(x) = -[q_1(x)\ln q_1(x) + q_2(x)\ln q_2(x)] \quad (۷)$$

و به ترتیبی که $p_k(x)$ و $q_k(x)$ بر مبنای احتمالات شرطی می باشند که نمونه k به ترتیب در منطقه $[x_1,x_1+x]$ و $[x_1+x,x_2]$ می باشد، و $p(x)$ و $q(x)$ بر مبنای احتمالاتی می باشد که تمام نمونه ها به ترتیب در منطقه $[x_1,x_1+x]$ و $[x_1+x,x_2]$ می باشند. [29,30]

$$p(x) + q(x) = 1 \quad (۸)$$

مقدار x که جداقل انتروپی را نشان می دهد بر مبنای مقدار بهینه هر بخش می باشد. تخمین انتروپی $pk(x)$ و $qk(x)$ و $p(x)$ و $q(x)$ به صورت زیر محاسبه می گردد.[31]

$$pk(x) = \frac{n_k(x)+1}{n(x)+1} \quad (۹)$$

$$qk(x) = \frac{N_k(x)+1}{N(x)+1} \quad (۱۰)$$

$$p(x) = \frac{n(x)}{n} \quad (۱۱)$$

$$q(x) = 1 - p(x) \quad (۱۲)$$

که در این فرمول $nk(x)$ تعداد نمونه دسته k می باشد که واقع در $[x_1,x_1+x]$ است، $n(x)$ تعداد کل نمونه های واقع در $[x_1,x_1+x]$ می باشد. $Nk(x)$ تعداد نمونه های دسته k می باشد که واقع در $[x_1+x,x_2]$ بوده، $N(x)$ تعداد کل نمونه هایی می باشد که واقع در $[x_1+x,x_2]$ است و n تعداد کل نمونه ها در $[x_1,x_2]$ می باشد. [32,33]

۱-۴ نظریه مجموعه ها

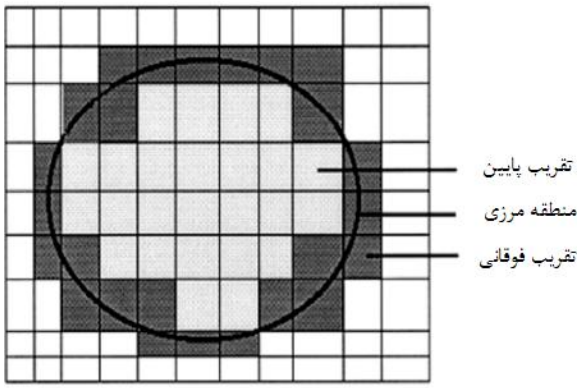
نظریه مجموعه ها (RST) توسط پاواک در سال ۱۹۸۲ مطرح شده است. در سال های اخیر RST در پیش بینی های مالی و اقتصادی مورد استفاده قرار گرفته اند. بسیاری از محققان از RST برای کشف قوانین تجارت استفاده می کنند. مفهوم RST بر مبنای این فرضیه می باشد که با مد نظر قرار دادن اهداف مربوطه موضوع بحث، بعضی از اهداف اطلاعاتی مشخص شده توسط اطلاعات مشابه از نقطه نظر اطلاعات در دسترس مرتبط به ان ها غیرقابل تشخیص می باشد. هر مجموعه از تمام اهداف غیر قابل تشخیص بر مبنای مجموعه اولیه ای می باشد و اطلاعاتی را در مورد این بحث ها شکل می دهند. هر واحد از مجموعه های مقدماتی به نام مجموعه مشخصی می باشد؛ در غیر اینصورت این مجموعه به صورت ناهنجار می باشد.

با در نظر گرفتن این مجموعه های ناهنجار، زوجی از این مجموعه های دقیق، به نام تخمین سطح بالا یا سطح پایین می باشد. $\underline{B}X = \{x|[x]_B \subseteq X\}$ و $\overline{B}X = \{x|[x]_B \cap X \neq \emptyset\}$ از مجموعه ناهنجار، مرتبط به یکدیگر می باشند. تخمین سطح پایین شامل تمام اهدافی می باشد که مطمئنا به آن مجموعه تعلق دارد، و تخمین سطح بالا شامل تمام اهدافی می باشد که احتمالا به ان مجموعه تعلق دارند. تفاوت بین تخیمنی پایین و بالا شامل تمام اهدافی می باشند که متعلق به مجموعه هستند. تفاوت بین تخمین بالا و پایین منطقه کرانی $BN_B(x) = \overline{B}X - \underline{B}X$ مجموعه ناهنجار را می سازد. مجموعه X، با توجه به اطلاعات مربوط به B ناهنجار (یا به سختی قابل تعریف) می باشد، اگر منطقه کران غیر تهی باشد. مفهوم اصلی در مجموعه های ناهنجار در شکل ۱ نشان داده شده است. [34,35]

RST به عنوان مجموعه ای از روش استدلال منطقی می باشد که برای تجزیه و تحلیل سیستم های اطلاعاتی مورد استفاده قرار می گیرد. سیستم اطلاعاتی به عنوان یک جدول تصمیم گیری مد نظر قرار می گیرد، که به صورت $S = (U, A, C, D)$ تعریف می گردد، که در ان U به عنوان موضوع بحث می باشد. A به عنوان مجموعه ای از مشخصه اولیه می باشد، و $C, D \subset A$ به عنوان دو مجموعه از مشخصه ها

ISSN: 2538-3663
http://www.Joas.ir

می باشند، با این فرض که $A = C \cup D$ و $C \cap D = \emptyset$، که در این فرمول C به عنوان مشخصه مشروط و D به عنوان مشخصه تصمیم می باشد. اندازه گیری برای توصیف نادرستی طبقه بندی تخمین به نام کیفیت تخمین X بر B می باشد. [36,37,38]



شکل ۱. مفهوم اصلی مجموعه های ناهنجار

آن نشان دهنده درصد اهدافی می باشد که به طور صحیح به دسته X تقسیم می شوند، که مشخصه B را بکار می گیرد:

$$\gamma_B(A) = \frac{\sum card(\underline{B}X_i)}{card(U)} \qquad (۱۳)$$

اگر $\gamma_B(A) = 1$ باشد، به این ترتیب جدول تصمیم گیری پایدار می باشد؛ در غیر اینصورت ناپایدار است.

یک مسئله مهم در RST کاهش نسبت می باشد که در آن مجموعه کاهش یانتروپیه ای از نسبت ها، کیفیت مشابهی را از مجموعه تخمین ها ایجاد می کنند. دو مفهوم بنیادین در ارتباط با این کاهش مشخصه وجود دارد. کاهش B از A که توسط RED(B) تعریف می گردد، به عنوان زیرمجموعه حداقل A می باشد، که کیفیت مشابهی از تخمین اهداف را در دسته های اولیه B به عنوان مشخصه کل A ایجاد می کند. هسته B از Aف هسته (B)، به عنوان بخش مهم A می باشد، که بدون برهم زدن قابلیت برای دسته بندی اهداف در دسته مقدماتی B کنار گذاشته نمی شود. آن فصل مشترک تمام کاهش ها می باشد. [39]

$$CORE(B) = \bigcap_{R_i \in RED(B)}, \quad i = 1, 2, \ldots \qquad (۱۴)$$

با استفاده از الگوریتم کاهشی، قوانین از طریق تعیین مقدار مشخصه تصمیم گیری بر مبنای مقادیر نسبت داده شده شرطی مد نظر قرار می گیرند. بنابراین قوانین در شرایط IF و سپس در قالب تصمیم گیری نشان داده می شوند. مفهوم جدول تصمیم گیری در این بررسی مد نظر قرار می گیرد تا قوانینی را از روابط فازی ایجاد کند، که قوانین را برای نتایج پیش بینی بهتر ایجاد می کند. [40,41]

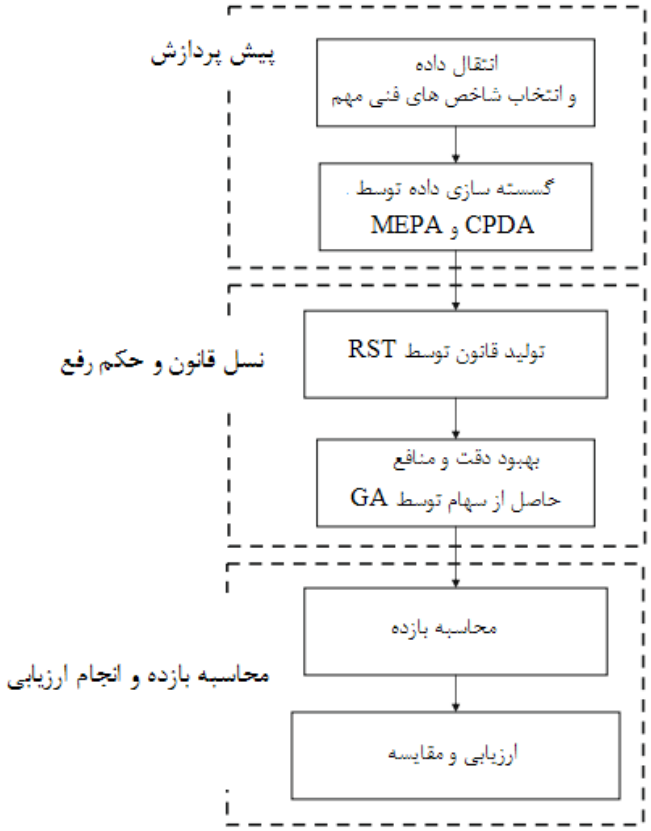
۱-۵ الگوریتم های ژنتیک

الگوریتم های ژنتیک (GAs) توسط هالند در سال ۱۹۷۵ مطرح شد و توسط گولدبرگ در سال ۱۹۸۹ بسط پیدا کرد. GAs به عنوان الگوریتم های جستجو می باشند که از طریق ارزیابی تحریک شده و در جستجو برای بهینه سازی کلی به منظور کاربردهای زیاد بکار گرفته شدند. علاوه بر این GAs به طور موفقیت امیزی در پیش بینی های مالی و اقتصادی بکار گرفته شدند. این الگوریتم ها به رمزگذاری راه حل های بلقوه برای مشکلات خاص در ساختار داده به شکل کروموزوم پرداخته و ترکیب مجدد اپراتورها را در این ساختارها برای حفظ اطلاعات مهم بکار می گیرند. مراحل GAs در مدل مطرح شده، بر مبنای گفته های گولدبرگ برای این پژوهش مد نظر قرار می گیرد.[42-45]

۱-۶ مدل پیشنهادی

بر مبنای مفاهیم مطرح شده در بالا، چارچوب کلی مدل هیبریدی پیشنهادی در شکل ۲ نشان داده شده است که به ادغام چهار روش داده کاوی جدید (همانند CPDA, MEPA, RST و GAs) در مراحل پیش بینی پرداخته و سه فاز اصلی را که در زیر بیان شده (شامل مرحله) مد نظر قرار می دهد.

ISSN: 2538-3663
http://www.Joas.ir



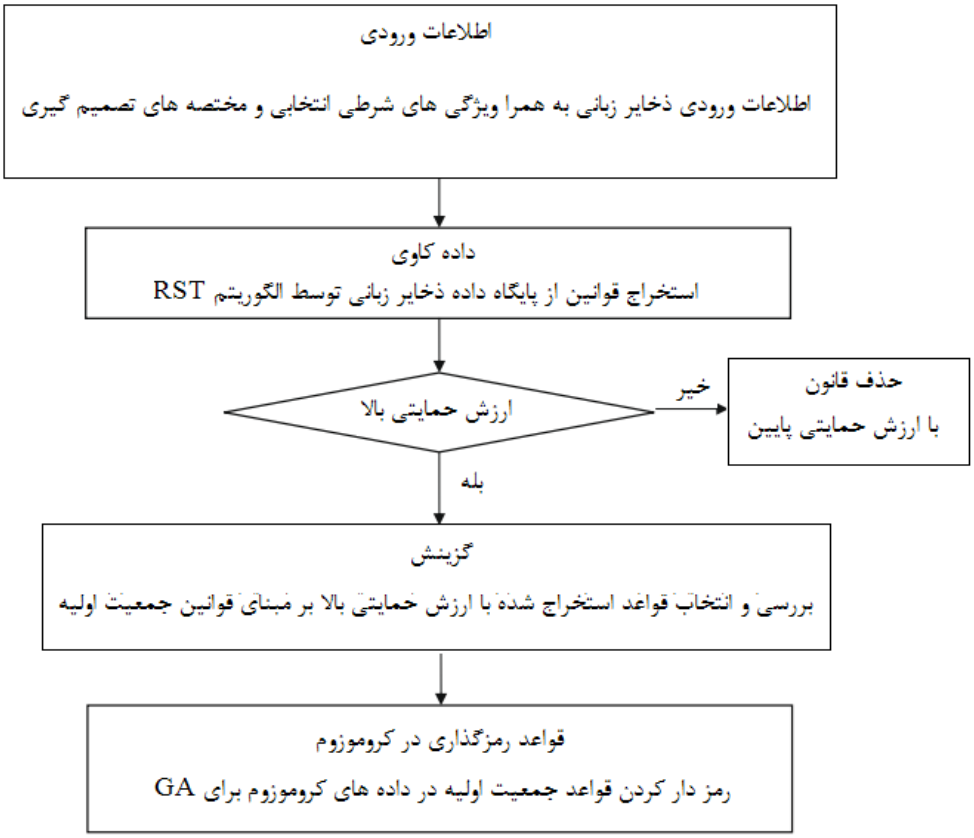
شکل ۲. چارچوب مدل پیشنهادی

برای شرح مدل مطرح شده، هر مرحله از مدل مطرح شده به صورت زیر توصیف می گردد:

مرحله ۱: تبدیل داده و انتخاب شاخص های فنی ضروری

در این مرحله، شاخص های فنی بر مبنای مختصات شرایط مورد استفاده قرار می گیرند، و نوسانات قیمت روز بعد، که در معادله ۱۶ تعریف شده است، بر مبنای مشخصات تصمیم گیری مورد استفاده قرار می گیرد. شاخص های فنی حاصل از پنج کمیت اصلی می باشد (قیمت اغازین، بالاترین قیمت، پایین ترین قیمت، قیمت نهایی، حجم تجارت)

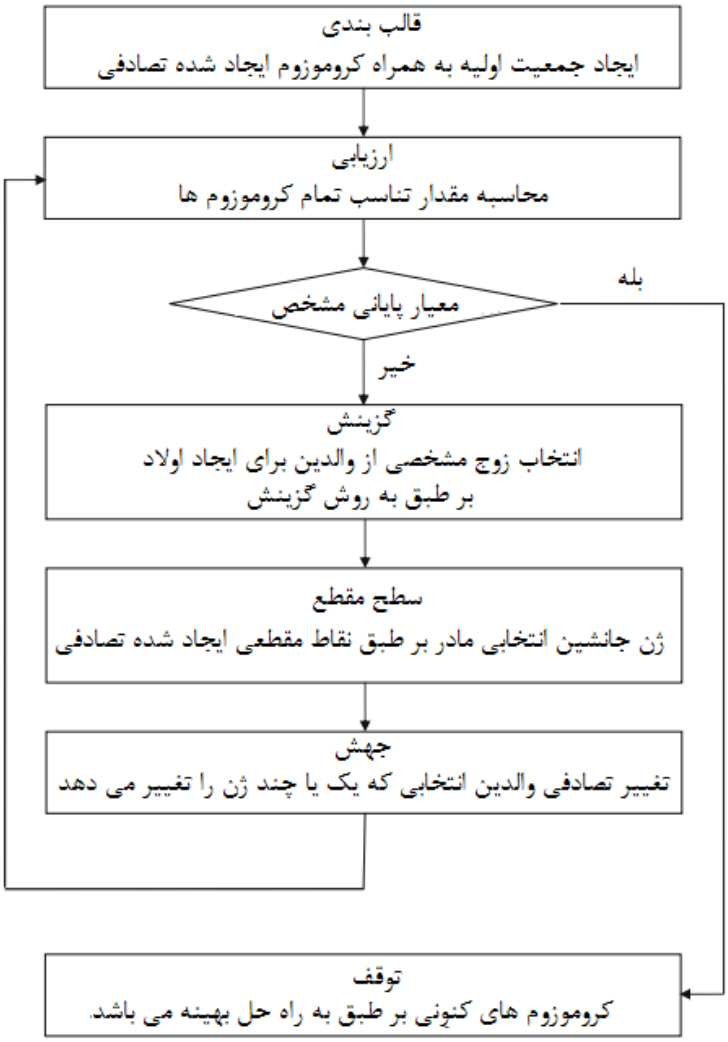
ISSN: 2538-3663
http://www.Joas.ir



شکل ۳. ایجاد قوانین توسط RST

شکل ۳ مراحل ایجاد قوانین بر مبنای RST را نشان می دهد. این روش مجموعه ابزارهایی را بر مبنای [38,53] RST (LEM2) بکار می گیرد، تا به استخراج قوانین فازهای از پیش پردازش شده از مجموعه داده ذخیره زبانی استخراج کرده و به انتخاب قوانین استخراجی با ارزش حمایتی بالا به صورت جمعیت اولیه برای عملیات GA بپردازد.[46-48]

ISSN: 2538-3663
http://www.Joas.ir



شکل ۴. بهبود دقت و منافع حاصل از سهام توسط عملیات GA

شکل ۴ مراحل دقت و بهبود منافع حاصل از سهام توسط GA را نشان می دهد. این روش از GA برای اصلاح قوانین استخراج شده استفاده می کند، بنابراین باعث بهبود دقت در رده بندی و منافع حاصل از سهام می گردد.

به منظور انتخاب این شاخص های فنی ضروری بر مبنای مشخصات شرایط که در سطح بالایی مرتبط به نوسانات قیمت روز بعد می باشد، ماتریس همبستگی به کار گرفته می شود تا شاخص های ضروری را از چند شاخص رایج انتخاب کند.

نوسان قیمت روزانه (t+1) = قیمت نهایی (t+1) – قیمت آغازی (t+1)

مرحله ۲: کسستگی داده توسط CPDA و MEPA

در این مرحله، CPDA به کار گرفته می شود تا به تفکیک نوسان قیمت روز بعد (مشخصه تصمیم) بپردازد، و MEPA برای تفکیک شاخص های ضروری (خصوصیات فنی مشروط) مورد استفاده قرار می گیرد. چون نوسان قیمت در بازار سهام و منافع حاصل از سهام به دقت توزیع نرمال را دنبال می کند، CPDA به عنوان روش مناسبی برای تفکیک داده قیمت سهام می باشد. در این مدل، نوسان قیمت روز بعد، توسط سه ارزش زبانی تعریف می گردد: بالا، متوسط، پایین. علاوه بر این به منظور پارتیشن بندی موضوع هر شاخص فنی مهم بر مبنای خصوصیات داده، این مرحله از MEPA برای تشخیص خصوصیات شرایط استفاده می کند. خصوصیات دسته بندی شده به عنوان یک مرحله مهم برای مراحل داده کاوی به ویژه RST بوده، و توسط این مرحله، ابعاد اطلاعاتی مربوط به پایگاه داده کاهش یانتروپیه و ساده می گردد.

در این مقاله، ما به تفکیک هر شرایط با استفاده از ۱۵ اصطلاح زبانی می پردازیم، زیرا با استفاده از TSMC به عنوان پایگاه داده آزمایشی، صحت مدل مطرح شده با استفاده از ۱۵ اصطلاح زبانی به بهترین وجه به اجرا در می آید که د مقایسه با اصطلاحات زبانی ۲، ۳ و ۷ قرار می گیرد. [49,50]

ISSN: 2538-3663
http://www.Joas.ir

مرحله ۳: ایجاد قوانین توسط RST

این مرحله از الگوریتم RST برای استخراج قوانین از مجموعه داده های ذخیره شده آموزشی استفاده می کند. قوانین استخراج شده به شکل شرطی با توجه به ارزش شرایطی خاص و ارزش مشخصات تصمیم گیری می باشند که به صورت زیر است:

جدول ۱. صحت مدل مطرح شده با اصطلاحات زبانی مشخصات شرطی

دوره های زبانی مشخصه تصمیم گیری	دوره زبانی مشخصه شرطی	۲	۳	۷	۱۵
	بالا، شکست، پایین	۰,۳۲۲	۰,۴۰۵	۰,۶۱	۰,۶۸

جدول ۲. صحت مدل مطرح شده با اصطلاحات زبانی مختلف مشخصات شرطی برای TEPIX

دوره های زبانی مشخصه تصمیم گیری	دوره زبانی مشخصه شرطی	۲	۳	۷	۱۵
	بالا، شکست، پایین	۰,۴۷۲	۰,۶۳	۰,۶۹	۰۷۷

اگر RSI=L8 و با در نظر گرفتن حجم، به این ترتیب داریم $\mathrm{DPF}_{next\ day} = Up$

این قانون بیا ن می کند که اگر ارزش زبانی RSI به صورت L8 (متوسط) باشد، و ارزش زبانی حجم برابر با L9 (کم و بیش بالا)، به این ترتیب نوسان قیمت روز بعد بالا می باشد.

مرحله ۴. افزایش صحت و منافع حاصل از سهام توسط GA

این مرحله از GA برای اصلاح قوانین استخراج شده توسط RST از مرحله ۳ برای بهبود دقت پیش بینی و منافع حاصل از سهام استفاده می کند.

مرحله ۵: محاسبه منافع

این مرحله به محاسبه منافع حاصل از سهام با استفاده از قوانین مرحله ۴ در پیش بینی قیمت سهام می پردازد. روش استنباطی بر مبنای قوانین اصلاح شده در الگوریتم زیر شرح داده می شود:

جدول ۳. ساختار کروموزوم ها برای مدل طرح شده

نام ویژگی	F1	F2	…	F_{n-1}	F_n	DPF_{nd}
ارزش کدگذاری	0-15	0-15	…	0-15	0-15	1-3

جدول ۴. تنظیمات پارامتر الگوریتم ژنتیک

اندازه جمعیت	۱۰۰۰
تعداد نسل ها	۱۰۰
روش ارزش دهی	روش صحیح کد گذاری
درصد الیت	۰,۲
روش گزینش	انتخاب مسابقات
روش متقاطع	متقاطع یکنواخت
نسبت متقاطع	۰,۰۸
روش جهش	جهش تک نقطه
نسبت جهش	۰,۰۵

$$\mathrm{Return}_{\mathrm{Up}}(t+1) = (\text{closing price}(t+1) - \text{opening price}(t+1)) \times \text{unit} \qquad (۱۵)$$

ISSN: 2538-3663
http://www.Joas.ir

$$\text{Return}_{\text{Down}}(t+1) = (\text{opening price}(t+1) - \text{closing price}(t+1)) \times \text{unit} \quad (۱۶)$$

مرحله ۶: ارزیابی و مقایسه.

در این مرحله، برای ارزیابی دقیق عملکرد پیش بینی، صحت و منافع حاصل از سهام بر مبنای شاخص عملکردی می باشند. این مرحله به محاسبه منافع بر مبنای سه استراتژی اقتصادی بالا پرداخته و به محاسبه منافع حاصل از سهام تمام معاملات می پردازد. صحت مدل داده کاوی در معادله ۲۰ تعریف می گردد. [51,52]

$$\text{Accuracy } (T) = \frac{\sum_{i=1}^{|T|} \text{correct classification}(t_i)}{|T|}, \quad t_i \in T \quad (۱۷)$$

$$\text{Correct classification}(t_i) = \begin{cases} 1, & \text{if classify } (t_i) = \text{actual price fluctuation } (t_i) \\ 0, & \text{otherwise} \end{cases}$$

که در اینجا T مجموعه ای از نمونه داده می باشد که توسط مدل پیشنهادی دست بندی می شود، $t_i \in T$: نوسان قیمت واقعی (t_i) به عنوان دسته بندی هر نمونه داده در ارتباط با نوسان قیمت واقعی بوده که به عنوان یکی از مقادیر زبانی می باشد، بالا (L_3)، متوسط(L_2)، و پایین(L_1)؛ و به تنظیم (t_i) منافع دسته بندی نمونه داده t_i توسط مدل پیشنهادی می پردازد.

۳. ارزیابی و اعتبار سنجی مدل

برای تایید مدل پیشنهادی، این بخش دو نوع ارزیابی عملکرد را انجام می دهد: ۱) پیش بینی ارزیابی دقیق: که دقت مدل پیشنهادی را بر مبنای معادله ۲۰ ایجاد کرده و دو مدل مقایسه ای دیگر را نیز ایجاد می کند.، که تنها از یک روش داده کاوی استفاده می کند، RST یا GA، که پیش بینی هایی را تحت شرایط از پیش پردازش شده بر مبنای مدل مورد نظر می پردازد؛

و ۲) ارزیابی منافع حاصل از سهام: سود سهام را بر مبنای معادله ۱۸ و ۱۹ ایجاد کرده و سه مدل مقایسه ای دیگر را نیز مد نظر قرار می دهد: که شامل خرید و نگهداشت، RST و GA، می باشند.

در این بررسی، دوره هفت ساله از شاخص سهام TEPIX ، از سال ۱۳۹۰ تا ۱۳۹۶ بر مبنای پایگاه داده آزمایشی انتخاب می گردد.

صحت این سه مدل (RST, GAs و مدل پیشنهادی) در جدول ۵ فهرست شده اند و مقایسه آن نشان می دهد که مدل پیشنهادی دارای عملکرد بهتری نسبت به دو مدل دیگر می باشد.

جدول ۵. مقایسه دقت بین مدل ها (TEPIX)

سال	تئوری مجموعه‌های راف	الگوریتم ژنتیک	مدل پیشنهادی
۱۳۹۰	۰,۷۱۲	۰,۶۲۱	۰,۷۶۰
۱۳۹۱	۰,۶۸۸	۰,۶۲۶	۰,۷۲۲
۱۳۹۲	۰,۵۹۲	۰,۵۲۴	۰,۶۱۳
۱۳۹۳	۰,۵۵۴	۰,۵۵۱	۰,۶۰۵
۱۳۹۴	۰,۵۲۳	۰,۵۰۲	۰,۵۴۰
۱۳۹۵	۰,۵۱۸	۰,۴۹۲	۰,۵۹۴
۱۳۹۶	۰,۵۱۴	۰,۵۰۳	۰,۵۶۶

جدول ۶. مقایسه سود سهام برای چهار مدل

سال	استراتژی خرید و فروش	تئوری مجموعه‌های راف	الگوریتم ژنتیک	مدل پیشنهادی
۱۳۹۰	-۳۰,۲۴۵	۵۵,۳۱۷	۸۸,۲۷۵	۱۳۵,۵
۱۳۹۱	-۲۴,۷۱۳	۹۸,۸	۱۲۵,۲	۱۸۴,۷
۱۳۹۲	۶۷,۳	۱۰۳,۷۴	۱۳۶,۲۱	۱۵۴,۹
۱۳۹۳	۲۴,۳۷	۷۲,۳۶۱	۹۹,۸۱۴	۶۵,۳۰۵

ISSN: 2538-3663
http://www.Joas.ir

۱۳۹۴	-۱۵,۲۷۱	۳۳,۳۳۵	۶۵,۲۱۴	۷۷,۵۱۴
۱۳۹۵	۲۴,۷۳۰	۴۷,۳۵۱	۶۸,۲۷۸	۸۸,۳۶۶
۱۳۹۶	-۱۲,۱۵۰	۳۶,۲۸۱	۵۹,۶۲۴	۹۳,۲۴۷

با استفاده از مقایسه سود سهام که در جدول ۶ نشان داده شده است، مشخص است که مدل پیشنهادی برتر از سه مدل دیگر در هر مرحله از تست به جز سال ۱۳۹۳ می باشد. TEPIX در سال ۱۳۹۳ دارای نوسان شدیدی بوده ارزیابی عملکرد نشان می دهد که بدترین شرایط در مورد مدل پیشنهادی برای سال ۱۳۹۳ به وقوع پیوسته است ارزیابی مربوط به سود سهام به اثبات عملکرد برجسته مدل پیشنهادی می پردازد.

ISSN: 2538-3663
http://www.Joas.ir

نتیجه گیری

این مقاله مدل هیبریدی جدیدی را مطرح می کند، که بر مبنای چهار روش جدید (CDPA, MEPA RST و GA) می باشد، تا به بهبود عملکرد پیش بینی بازار سهام بپردازد. با استفاده از داده ارزیابی عملکرد در بخش بالا، می توانیم به این نتیجه گیری برسیم که هدف اصلی این مقاله حاصل شده است. علاوه بر این با بررسی دقیق داده عملیاتی، می توانیم ثابت کنیم که مدل پیشنهادی سود سهام مثبتی را حاصل می کند. شواهد اشاره به توانایی استثنایی مدل پیشنهادی برای مد نظر قرار دادن الگوی قیمت صحیح در بازار سهام دارد. ادغام RST و GA در فرایند پیش بینی تاثیر مثبتی را ایجاد کرده، و عملکرد مدل را بالاتر می برد. با انجام این بررسی، دو مزیت برای مدل پیشنهادی حاصل می شود: ۱) مدل پیشنهادی می تواند قواعد قابل درک و منطقی تری را ایجاد کند، زیرا قوانین شرطی ایجاد شده توسط RST می تواند به مدلسازی جنبه کیفی دانش بشری بپردازد؛ و ۲) مدل پیشنهادی می تواند برای سرمایه گذاران سهام پیشنهادات هدفمندی را به منظور تصمیم گیری برای سرمایه گذاری در بازار سهام ایجاد کند، زیرا مدل پیشنهادی قوانین پیش بینی شده ای را بر مبنای داده های سهام عینی نسبت به قضاوت ذهنی انسان ایجاد کند.

ISSN: 2538-3663
http://www.Joas.ir

منابع و مراجع

[1] P.J. Acklam, An algorithm for computing the inverse normal cumulative distribution function, (2004), Available from: http://home.online.no/~pjacklam/notes/invnorm.
[2] F. Allen, R. Karalainen, Using genetic algorithms to find technical trading rules, Journal of Financial Economics 51 (1999) 245–271.
[3] B. Anna, Should normal distribution be normal? The Student'sTalternative, Computer Information Systems and Industrial Management Applications (2007) 3–8.
[4] M.E. Azo, Neural Network Time Series Forecasting of Financial Markets, Wiley, New York, 1994.
[5] T. Blickle, L. Thiele, A mathematical analysis of tournament selection, in: L.J. Eshelman (Ed.), Proceedings of the 6th International Conference on Genetic Algorithms, 1995, pp. 506–511.
[6] T. Bollerslev, Generalized autoregressive conditional heteroscedasticity, Journal of Econometrics 31 (1986) 307–327.
[7] G. Box, G. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, 1976.
[8] A.P. Chen, Y.H. Chang, Using extended classifier system to forecast S&P futures based on contrary sentiment indicators, Evolutionary Computation (2005) 2084–2090.
[9] A.P. Chen, Y.C. Chen, W.C. Tseng, Applying extending classifier system to develop an option-operation suggestion model of intraday trading – An example of Taiwan index option, Lecture Notes in AI (2005) 27–33.
[10] J.S. Chen, C.H. Cheng, Extracting classification rule of software diagnosis using modified MEPA, Expert Systems with Applications 34 (1) (2008) 411– 418.
[11] C.H. Cheng, J.R. Chang, C.A. Yeh, Entropy-based and trapezoid fuzzification-based fuzzy time series approaches for forecasting IT project cost, Technological Forecasting & Social Change 73 (2006) 524–542.
[12] C.H. Cheng, J.S. Chen, Extracting classification rule of software diagnosis using modified MEPA, Expert Systems with Applications 34 (2008) 411–418.
[13] S.C. Chi, W.L. Peng, P.T. Wu, M.W. Yu, The study on the relationship among technical indicators and the development of stock index prediction system,Fuzzy Information Processing Society (2003) 291–296.
[14] R. Christensen, Entropy minimax sourcebook, Fundamentals of Inductive Reasoning, vol. 1, Entropy Ltd., Lincoln, MA, 1980.
[15] N. Clarence, W. Tan, A hybrid financial trading system incorporating chaos theory, statistical and artificial intelligence/soft computing methods, in: Queensland Finance Conference, School of Information Technology, Bond University, 1999.
[16] A.I. Dimitras, R. Slowinski, R. Susmaga, C. Zopounidis, Business failure prediction using rough sets, European Journal of Operational Research 114 (1999) 263–280.
[17] R.F. Engle, Autoregressive conditional heteroscedasticity with estimator of the variance of United Kingdom inflation, Econometrica 50 (4) (1982) 987-1008.
[18] L.J. Eshelman, R.A. Caruana, J.D. Schaffer, Biases in the crossover landscape, in: Proceedings of the 11th International Joint Conference on Artificial Intelligence, Morgan Kaufmann, San Mateo, 1989, pp. 10–19.
[19] E. Faerber, All About Stocks: From the Inside Out, Probus, Chicago, 1995.
[20] E.H. Tay* Francis, S. Lixiang, Economic and financial prediction using rough sets model, European Journal of Operational Research 141 (2002) 641–659.
[21] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison Wesley, Reading, MA, 1989.
[22] D.E. Goldberg, K. Deb, A comparative analysis of selection schemes used in genetic algorithm, in: G. Rawlins (Ed.), Foundation of Genetic algorithms, Morgan Kaufmann, 1991, pp. 69-93.
[23] S. Greco, B. Matarazzo, R. Slowinski, A new rough set approach to evaluation of bankruptcy risk, in: C. Zopounidis (Ed.), Operational Tools in the Management of Financial Risks, Kluwer Academic Publishers, Dordrecht, 1998, pp. 121-136.
[24] J.H. Holland, Adaptation in Nature and Artificial Systems, University of Michigan Press, 1975.

ISSN: 2538-3663
http://www.Joas.ir

[25] K. Huarng, H.K. Yu, The application of neural networks to forecast fuzzy time series, Physica A 363 (2006) 481–491.
[26] S.H. Irwin, C.H. Park, What do we know about the profitability of technical analysis?, Journal of Economic Surveys 21 (4) (2007) 786–826 1628
[27] K. Kim, I. Han, Genetic algorithms approach to feature discretization in artificial neural networks for prediction of stock index, Expert System with Application 19 (2000) 125–132.
[28] M.J. Kim, S.H. Min, I. Han, An evolutionary approach to the combination of multiple classifiers to predict a stock price index, Expert Systems with Applications 31 (2006) 241-247..
[29] T. Kimoto, K. Asakawa, M. Yoda, M. Takeoka, Stock market prediction system with modular neural network, in: Proceedings of the International Joint Conference on Neural Networks, San Diego, California, 1990, pp. 1–6.
[30] J.B. Li, Y.T. Yu, A.P. Chen, Integration of group decisions and XCS in Intelligent financial decision support system – An example of Taiwan index, Evolutionary Computation, IEEE Congress (2006) 2389–2396.
[31] H. Liu, F. Hussain, C. Tan, M. Dash, Discretization: An enabling technique, Data Mining and Knowledge Discovery 6 (4) (2002) 393–423.
[32] C. Nikolopoulos, P. Fellrath, A hybrid expert system for investment advising, Expert Systems 11 (4) (1994) 245–250.
[33] Z. Pawlak, Rough sets, International Journal of Computational Information Science (1982) 341–356.
[34] Z. Pawlak, Rough sets and intelligent data analysis, Information Sciences 147 (2002) 1–12.
[35] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (2007) 3–27.
[36] Z. Pawlak, A. Skowron, Rough sets: Some extensions, Information Sciences 177 (2007) 28–40.
[37] Z. Pawlak, A. Skowron, Rough sets and Boolean reasoning, Information Sciences 177 (2007) 41–73.
[38] B. Predki, R. Slowinski, J. Stefanowski, R. Susmaga, Sz. Wilk, ROSE – Software implementation of the rough set theory, in: L. Polkowski, A. Skowron (Eds.), Rough sets and current trends in computing, Lecture Notes in Artificial Intelligence, vol. 1424, Springer-Verlag, Berlin, 1998, pp. 605–608.
[39] M.J. Pring, Technical Analysis, New York, 1991.
[40] F.K. Reilly, Investment Analysis and Porfolio Management, Dryden Press, Chicago, 1989.
[41] J.B. Richard, R.D. Julie, Technical Market Indicators: Analysis & Performance, John Wiley, New York, 1999.
[42] T.H. Roh, Forecasting the volatility of stock price index, Expert Systems with Applications 33 (2007) 916–922.
[43] T.J. Ross, Fuzzy Logic with Engineering Applications, McGraw-Hill, USA, 2000.
[44] H. Shimodaira, A newgenetic algorithm using large mutation rates and population-elitist selection (GALME), in: Proceedings of 8th IEEE Conference on Tools with Artificial Intelligence, 1996, pp. 25–32.
[45] C Cheng, T Chen, L Wei, A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting, Journal of Information Sciences,(2010), Volume 180, Issue 9, Pages 1610-1629
[46] G. Syswerda, Uniform crossover in genetic algorithms, in: Proceeding of the 3rd International Conference on Genetic Algorithms, 1989, pp. 2–9.
[47] R.R. Trippi, D.D. Sieno, Trading equity index futures with a neural network, Journal of Portfolio Management (1992) 27–33.
[48] C. Tuncay, D. Stauffer, Power laws and Gaussians for stock market fluctuations, Physica A (2007) 325–330.
[49] Y.F. Wang, Mining stock price using fuzzy rough set system, Expert Systems with Applications 24 (2003) 13–23.
[50] L. William, P. Russell, M.R. James, Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support, Decision Support Systems 32 (2002) 361–377.

ISSN: 2538-3663
http://www.Joas.ir

[51] C. Xipin, J. Min, X. Bin, C. Jie, Application of elitist preserved genetic algorithms on fuzzy controller, in: Proceedings of the 3th World Congress on Intelligent Control and Automation, Hefei, PR China, 2000.
[52] R. Yager, D. Filev, Template-based fuzzy system modeling, Intelligent and Fuzzy System 2 (1994) 39–54.