ارائهٔ روشی ترکیبی برای افزایش دقت پیش‌بینی در کاهش داده با استفاده از مدل مجموعهٔ راف و هوش تجمعی

الهه میرزائی و منصور اسماعیل پور*
دانشکده مهندسی کامپیوتر، واحد همدان، دانشگاه آزاد اسلامی، همدان، ایران

چکیده

نیاز به طراحی سامانه‌هایی که قادر به اکتشاف سریع اطلاعات مورد علاقه کاربران با تأکید بر کمینهٔ مداخله انسانی باشند از یک سو و روی‌آوردن به روش‌های تحلیل متناسب با حجم داده‌های حجیم ازسوی دیگر، در دنیای امروزی به‌خوبی احساس می‌شود؛ ازاین‌رو بهره‌گیری از قدرت فرآیند داده‌کاوی جهت شناسایی الگوها و مدل‌ها و نیز ارتباط عناصر مختلف در پایگاه داده جهت کشف دانش نهفته در داده‌ها روزبه‌روز ضروری‌تر می‌شود. از سوی دیگر تئوری مجموعه راف را به‌عنوان یک ابزار برای کشف وابستگی داده‌ها و کاهش خصیصه‌های موجود در یک مجموعه داده، تنها با استفاده از داده‌ها و بدون نیاز به اطلاعات می‌توان اضافی برشمرد. در این پژوهش جهت بهبود روند انتخاب ویژگی‌های اصلی و بهبود تئوری مجموعه راف، از ترکیب الگوریتم مورچگان و تئوری مجموعه راف جهت یافتن زیرمجموعه ویژگی‌های اصلی و حذف اطلاعات غیر مفید با ازدست‌رفتن کمینه اطلاعات استفاده شده است. نتایج حاصل از این ترکیب در ارزیابی داده‌های قیمت نفت نشان می‌دهد که ترکیب الگوریتم مورچگان و تئوری مجموعه راف در انتخاب ویژگی‌های مفید و بهینه، عملکرد مناسب‌تری نسبت به مدل‌های اخیر دارد.

واژگان کلیدی: تئوری مجموعه‌های راف، هوش تجمعی، الگوریتم مورچگان، انتخاب ویژگی، پیش‌بینی.

A New Hybrid Method to Increase the Prediction in Data Reduced Using Rough Set and Swarm Intelligence Model

Elaheh Mirzaee & Mansour Esmaeilpour*
Department of Computer Engineering, Hamedan Branch, Islamic Azad University, Hamedan, Iran

Abstract

Designing a system with an emphasis on minimal human intervention helps users to explore information quickly. Adverting to methods of analyzing large data is compulsory as well. Hence, utilizing power of the data mining process to identify patterns and models become more essential from aspect of relationship between the various elements in the database and discover hidden knowledge. Therefore, Rough set theory can be used as a tool to explore data dependencies and reducing features outlined in a data set. The main purpose of the rough theory is to obtain approximate concepts of acquired data. This theory is a powerful mathematical tool for arguing in ambiguous and indeterminate terms that provides methods for remove and reduce unrelated or excessive knowledge information on the data sets. This process of data reduction is based on the main task of the system, and without losing the basic data of the data sets. Rough set theory can play a very effective role to support decision-making systems, but in some cases, with increasing data volumes, there are inconsistent or collisional results which using swarm intelligence-based methods can choose the best of the contradictory, effectless or dummy data. This will bring interesting, unexpected and valuable structures from within a wide range of data. Since the ant colony optimization compares all the

* نویسندهٔ عهده‌دار مکاتبات
* Corresponding author

exploratory paths generated by each ant and the best route is selected from the existing paths, so considering the improvement of the selecting the main features and improving the theory of the Rough set, paths are not eliminated from the possible paths. In this research, the combination of the ant colony optimization and rough set theory have been used to find the subset of the main features and to delete the inappropriate information with the loss of the minimum information. This research will improve the features reduction technique employment Rough set theory and ant colony optimization. The gist of this research is removing useless information with minimal information loss. The results on petroleum prices data evaluation demonstrate that the hybrid method is more efficient than recent methods.

Keywords: Rough Set Theory, Swarm Intelligence, Ant Colony Optimization, Feature Reduction.

۱- مقدمه

هدف اصلی از تحلیل مجموعه راف[1] به‌دست‌آوردن مفاهیم تقریبی از داده‌های اکتسابی است. این تئوری، یک ابزار قدرتمند ریاضی برای استدلال در موارد ابهام و نامعین است که روش‌هایی را برای زدودن و کاستن اطلاعات دانش نامربوط یا مازاد بر نیاز از پایگاه‌های داده‌ها مهیا می‌سازد. این فرایند حذف داده‌های زائد، بر مبنای آموزش وظیفه اصلی سامانه و بدون ازدست‌دادن داده‌های اساسی پایگاه داده‌ها صورت می‌پذیرد [1]؛ درنتیجه تقلیل اطلاعات، مجموعه‌ای از قواعد تلخیص‌شده و پرمعنا حاصل می‌شود که کار تصمیم‌گیرنده را بسیار ساده‌تر می‌کند [2]. درحقیقت می‌توان گفت که مجموعه راف با کاهش فضای داده‌ها و برگزیدن نمونه داده های مهم، یک نگاشت از فضای داده‌های خام و نمونه داده‌ها را به فضای مسئله انجام می‌دهد؛ لذا، با توجه به رشد انفجاری حجم اطلاعات، مجموعه راف می‌توانند نقش بسیار مؤثری در سامانه‌های پشتیبان تصمیم‌گیری داشته باشد؛ اما در برخی موارد با بالارفتن حجم اطلاعات، به نتایج متناقض یا متلاقی در این تئوری می‌رسیم که با استفاده از روش‌های مبتنی بر هوش تجمعی از بین داده‌های متناقض یا متلاقی بهترین مورد را انتخاب می‌توان کرد؛ یا آنهایی را که در نتیجهٔ کار تأثیری ندارند، حذف کرد و درنتیجه ساختارهای جالب توجه، غیرمنتظره و باارزش را از داخل مجموعه وسیعی از داده‌ها کشف کرد و با اطمینان خاطر بیشتر دست به پیش‌بینی و تصمیم‌گیری زد [3]. استفاده از روش کلی گفته‌شده در تئوری مجموعه‌های راف، در بسیاری از مجموعه داده‌ها غیر عملی است [4].

از آنجایی‌که در الگوریتم مورچگان، تمام مسیرهای اکتشافی که توسط هر مورچه ایجاد می‌شود، با یکدیگر مقایسه شده و بهترین مسیر از بین مسیرهای موجود انتخاب می‌شود؛ بنابراین مسیری از بین مسیرهای ممکن حذف نمی‌شود؛ ازاین‌رو جهت بهبود روند انتخاب ویژگی‌های اصلی و بهبود تئوری مجموعه راف، در این پژوهش از ترکیب الگوریتم مورچگان و تئوری مجموعه راف جهت یافتن زیرمجموعه ویژگی های اصلی و حذف اطلاعات غیر مفید با ازدست‌رفتن کمینه اطلاعات استفاده شده است.

در بخش بعدی این پژوهش راجع به تئوری مجموعه‌های راف و روش استخراج ویژگی در این تئوری توضیح داده شده است، در بخش ۳ شرح مختصری از الگوریتم‌های کاهش ویژگی که در این پژوهش به‌منظور استخراج ویژگی استفاده شده آورده شده است؛ در بخش ۴ الگوریتم ارائه‌شده در این پژوهش شرح داده و در بخش ۵ نتایج تجربی پژوهش، شامل ارزیابی عملکرد و مقایسه دقت پیش‌بینی روش‌های ذکر شده آورده شده و بخش ۶ شامل نتیجه‌گیری است.

۲- تئوری مجموعه‌های راف

یافتن یک واژه معادل برای عبارت "ROUGH SETS" کار مشکلی است. در فرهنگ لغت برای کلمه "ROUGH" معادل‌هایی مانند زبر، درشت، تقریبی، بی‌ادب، متلاطم و ناصاف در نظر گرفته شده است [5] که در میان این واژه‌ها کلمه تقریبی، شباهت بیشتری با مفهوم مورد نظر بنیان‌گذار این نظریه دارد؛ اما هیچ‌یک از این کلمات بار معنایی خود واژه لاتین را ندارند؛ به همین علت در این پژوهش عبارت "مجموعه‌های راف" به‌عنوان معادل آن به کار گرفته خواهد شد. تئوری مجموعه‌های راف در اوایل سال ۱۹۸۰میلادی توسط پروفسور زدیسلاو پاولاک پایه‌گذاری شد. این تئوری

[1] Rough set

با تحلیل جدول‌های داده سروکار دارد. در این تئوری جدول‌های داده می‌توانند توسط اندازه‌گیری یا افراد متخصص و آگاه (خبره) به‌دست آمده باشد. هدف اصلی از تحلیل مجموعه راف به‌دست‌آوردن مفاهیم تقریبی از داده‌های اکتسابی است. این تئوری، یک ابزار قدرتمند ریاضی برای استدلال در موارد ابهام و نامعین است که روش هایی را برای زدودن و کاستن اطلاعات دانش نامربوط یا مازاد بر نیاز از پایگاه های داده ها مهیا می سازد. این فرایند حذف داده های زائد، بر مبنای آموزش وظیفه اصلی سامانه، و بدون ازدست‌دادن داده‌های اساسی پایگاه داده‌ها صورت می‌پذیرد. درنتیجهٔ تقلیل اطلاعات، مجموعه‌ای از قواعد تلخیص‌شده و پرمعنا حاصل می‌شود که کار تصمیم‌گیرنده را بسیار ساده‌تر می‌کند [6] و [7]؛ لذا، با توجه به رشد انفجاری حجم اطلاعات، مجموعه راف نقش بسیار مؤثری را در سامانه‌های پشتیبان تصمیم‌گیری می‌توانند داشته باشند [8]. تئوری مجموعه راف نقاط اشتراک زیادی با تئوری مجموعه‌های فازی، تئوری شهود، روش‌های استدلال بولی و تحلیل تفکیکی دارد؛ اما با این وجود به‌عنوان یک تئوری مستقل در نظر گرفته می‌شود.

۱-۲- مجموعه‌های راف در حالت تئوری

در جدول اطلاعات (داده‌ها) هر زیرمجموعه از ویژگی‌ها یک رابطه هم‌ارزی فراهم می‌کند [9]. اگر مجموعه B یک زیرمجموعهٔ غیر تهی از دو شئ (موضوع) v وu (غیر قابل تشخیص توسط B) متعلق به U باشند، خواهیم داشت:
برای هر ویژگی A_i در B داریم:

$$u \equiv v(mod\ B)\ \ if\ \rho(u,A_i)= \rho(v,A_i) \qquad (۱)$$

علامت ≡ نشان‌دهندهٔ رابطه هم‌ارزی است و به آن رابطه تعادل نیز می‌گویند و آن را با IND(B) نمایش می‌دهند. طبقه‌های هم‌ارزی شامل U را به‌وسیله [u]IND(B) یا ساده‌تر به‌صورت [u]B نشان می‌دهند که B می‌تواند یک مجموعه منفرد به‌صورت B={Ai} باشد.

دراین حالت به‌سادگی رابطه هم‌ارزی به‌صورت IND(Ai) را تعریف می‌کنیم و خواهیم داشت:

$$IND(B)=\cap\{IND(A_i) : A_i\ in\ B\} \qquad (۲)$$

رابطه بالا بیان می‌کند که طبقه‌های هم‌ارزی مربوط بـه IND(B) همه اشتراک‌های ممکن از طبقه‌های هم‌ارزی مربوط به $IND(A_i)$ را شامل می‌شود؛ بنابراین در حالت کلی از رابطه IND(P) که به‌عنوان رابطه نامشخصه‌ها هم خوانده می‌شود، به‌منظور نشان‌دادن نامشخصه بر روی مجموعه مرجع U استفاده می‌توان کرد.

$$IND(P) = \{(x, y)\in \mathrm{U}\ 2 \mid \forall a \in P, a(x) = a(y)\} \qquad (۳)$$

طبقهٔ هم‌ارزی رابطه نامشخصه های P با $[x]_P$ نمایش داده می‌شود و تقریب پایین X بر اساس فرمول (۴) به‌دست می‌آید.

$$\underline{P}X= \{x \mid [x]_p \subseteq X\} \qquad (۴)$$

و رابطه هم‌ارزی P و Q بر روی U در نواحی مثبت براساس فرمول‌های (۴و ۵) تعریف می‌شود.

$$\mathrm{pos}_p(\mathrm{Q}) = \mathrm{U}_{\mathrm{x}\in \mathrm{U}|Q}\, PX \qquad (۵)$$

و درنهایت درجه وابستگی ویژگی را می‌توان با استفاده از فرمول (۶) نمایش داد که بیان‌گر این است که: Q روی P وابسته و درجه وابستگی آن برابر $K\ (0 \le k \le 1)$ است ($\gamma_P(Q)$):

$$\mathrm{K}=\gamma_P(\mathrm{Q})= \frac{|POS_p(Q)|}{|U|} \qquad (۶)$$

ترکیب مفاهیم مجموعه راف در ساخت درخت تصمیم‌گیری برای طبقه‌بندی داده‌های مبهم و نامشخص نتایج بهتر دارد؛ اما مجموعه راف براساس طبقه‌بندی برای داده‌های خیلی حجیم، گاهی اوقات به یک مشکل عمده به نام Over-fitting (مدلی با ساختار محدود) منجر خواهد شد[10] .

۳- الگوریتم‌های کاهش ویژگی

جستجو به‌منظور یافتن یک زیرمجموعهٔ بهینه شامل یافتن ویژگی‌هایی است که به‌شدت با قابلیت تصمیم‌گیری در ارتباط بوده، ولی از یکدیگر مجزا هستند [11]. انتخاب زیرمجموعهٔ بهینه بسته به مشکلی که بررسی می‌شود، متفاوت است؛ به‌طورمعمول الگوریتم‌های کاهش ویژگی[1] مبتنی بر روش‌های اکتشافی یا جستجوی تصادفی برای کاهش درجه پیچیدگی است و درنهایت به زیرمجموعه‌ای

[1] FS(Feature Selection)

کاهش‌یافته از ویژگی‌ها ختم می‌شود. از آنجایی که انتخاب زیرمجموعه بهینه از بین مجموعه ویژگی‌ها را می‌توان مهم‌ترین قسمت در فرآیند داده‌کاوی برشمرد ازاین‌رو در ادامه ۳ الگوریتم کاهش ویژگی معتبر به‌منظور انتخاب ویژگی ذکر شده است؛ هر یک از این روش‌ها به‌عنوان یک روش مناسب و بهبودیافته معرفی شده‌اند و دارای قابلیت کشف زیرمجموعه کیمنه یا بهینه هستند [12].

۱-۳- الگوریتم جانسون بر پایه مجموعه‌های راف به‌منظور کاهش ویژگی

این الگوریتم با دادن مقداردهی اولیه R (نامزد کاهش) با یک مجموعه تهی شروع و سپس هر ویژگی در تابع نامشخصه‌ها به‌وسیله اندازه‌گیری اکتشافی ارزیابی می‌شود [13]. الگوریتم استاندارد جانسون به‌طورعمومی تعداد ویژگی‌هایی را که درون عبارت ظاهر می‌شوند، می‌شمارد و ویژگی‌هایی را که بیشتر ظاهر می شوند، مهم‌تر در نظر می‌گیرد، ویژگی‌هایی که بالاترین مقدار اکتشافی را دارند به نامزدهای کاهش اضافه می‌شوند و تمام عبارت‌هایی که در تابع نامشخصه‌ها شامل این ویژگی هستند، حذف می‌شوند [14]. به‌محض اینکه تمام عبارات حذف شوند، الگوریتم خاتمه یافته و مقدار R را برمی‌گرداند. R یافته‌شده تضمین می‌کند که کاهش، انجام شده وتمام عبارت‌های موجود در تابع نامشخصه‌ها از بین رفته‌اند [15].

۲-۳- الگوریتم بهینه‌سازی ذرات[1] بر پایه طبقه‌بندی cf subset eval

الگوریتم PSO یک الگوریتم جستجوی اجتماعی است که از روی رفتار اجتماعی دسته‌های پرندگان مدل شده است. اساس کار PSO بر این اصل استوار است که در هر لحظه هرذره مکان خود را در فضای جستجو با توجه به بهترین مکانی که تاکنون در آن قرار گرفته است و بهترین مکانی که در کل همسایگی‌اش وجود دارد، تنظیم می‌کند [16].

تاکنون الگوریتم‌های متعددی بر پایه PSO به‌منظور کاهش ویژگی‌ها ارائه شده است که از آن جمله می‌توان به جستجوی دودویی، جستجوی برداری، بهینه‌سازی ذرات بر پایه SVM، Gain ratio، Gain Attribute، cf subset eval و.... اشاره کرد. پژوهش‌های انجام‌شده در این زمینه و مقایسه‌های انجام‌گرفته در پژوهش‌های پیشین، حاکی از این است که انتخاب ویژگی بر پایه طبقه‌بندی cf subset eval [17] نتایج بهتری در بر داشته و ویژگی‌های بهینه‌تری نسبت به سایر روش‌ها انتخاب شده است. بنابراین در این پژوهش مبنای استخراج ویژگی بر اساس PSO مبتنی بر cf subset eval است.

۳-۳- الگوریتم جستجوی هارمونیک[2]

بسیاری از استراتژی‌های جستجو برای کار انتخاب ویژگی، در تلاش برای شناسایی زیرمجموعه‌هایی با کیفیت بهتر و کم‌حجم‌تر هستند. HS یک روش کلی است که می‌تواند در رابطه با بسیاری از روش‌های ارزیابی زیرمجموعه مورد استفاده قرار گیرد. سادگی الگوریتم HS باعث کاهش پیچیدگی کلی فرایند جستجو می‌شود. این روش قادر به فرار از راه‌حل‌های محلی و شناسایی راه‌حل‌های متعدد با توجه به ماهیت تصادفی الگوریتم HS است [18]. روش کار این الگوریتم کنترل ویژگی‌های اضافی معرفی‌شده به‌منظور کاهش ویژگی‌ها است. این الگوریتم با سایر روش‌های استخراج ویژگی تکرارشونده ترکیب شده است و سعی در کشف زیرمجموعه‌هایی با کیفیت بهتر دارد [19].

۴- الگوریتم ترکیبی مورچگان و تئوری مجموعه‌های راف

۱-۴- مروری بر الگوریتم مورچگان

الگوریتم کلونی مورچه‌ها برای نخستین‌بار توسط دوریگو برای حل مسائل مشکل بهینه‌سازی مطرح شد. الگوریتم از رفتار جستجوی غذای برخی گونه‌های مورچه در طبیعت الهام گرفته شده است. مورچه‌ها زمان جستجوی غذا ابتدا نواحی اطراف لانه را با یک رفتار تصادفی کاوش می‌کنند؛ به‌محض اینکه مورچه‌ای منبع غذایی را پیدا کرد آن را ارزیابی می‌کند و مقداری از آن را به سمت لانه برمی‌گرداند [20]. در مسیر برگشت، مورچه دنباله‌ای از فرومون[3] را بر روی زمین بر جای می‌گذارد. این فرومون برجای مانده که مقدار آن به کیفیت و اندازه منبع غذایی بستگی دارد، مورچه‌های دیگر را به سمت آن هدایت می‌کند. ارتباط غیرمستقیم بین مورچه‌ها یعنی استفاده از دنباله فرومون ، آنها را قادر به پیداکردن کوتاه‌ترین مسیر بین لانه و غذا می‌کند [21].

[1] PSO(Particle Swarm Optimization)

[2] HS(Harmonic Search)

[3] Pheromone

۱-۱-۴- شرح الگوریتم مورچگان

پس از تعیین پارامترها و مقداردهی اولیهٔ فرومون ، الگوریتم در یک حلقه اصلی تکرار می‌شود، در هر تکرار همه مورچه‌ها راه‌حل‌ها را می‌سازند، این راه‌حل‌ها در یک جستجوی محلی تقویت می‌شوند و درنهایت مقدار فرومون به‌روزرسانی می‌شود [22]. شبه‌کد الگوریتم بهینه‌سازی کلونی مورچه درزیر آورده شده است [23].

Input:
$P(a_1,...,a_k)$ //A population of K ants
$C=\{c_{ij}\}$ //A set of solution components or edges to be traversed
//A set of pheromone values where each τ_{ij} is associated with the solution component c_{ij} δ
//a stopping criterion

Output:
X_{best} //best solution
Begin
Initialise T //initialise pheromone values in T
InitialiseX_{best} // Initialise best solution
While δ not reached do
For each ant K in P do
X_p=construct solution
//Optionally apply local search to improve X_p
If cost(X_l) is better than cost(X_{best})
$X_{best}= X_l$ //update best solution as required
Update pheromone(T)
End for
End while
Return X_{best}
End

(الگوریتم-۱): بهینه سازی کلونی مورچه
(Algorithm-1): Ant colony optimization

مجموعه P نشان‌دهندهٔ مجموعه مورچگان است که تعداد آنها K تا در نظر گرفته شده است و از a_1 تا a_k شماره‌گذاری شده‌اند.

مجموعه C نشان‌دهنده راه‌حل‌هایی است که تاکنون ساخته شده است یا گره‌هایی که پیمایش شده‌اند. مجموعه T نشان‌دهندهٔ میزان فرومون است که میزان فرومون بین دوگره با τ_{ij} نمایش داده می‌شود. یا به عبارت دیگر، میزان فرومون بین هر دو گره i و j که توسط راه‌حل C_{ij} حاصل شده است.

δ بیان‌گر شرط توقف است و Xp مسیر پیمایش‌شده توسط راه حل جدید را نگهداری می‌کند، به‌منظور اینکه هرگره فقط یک‌بار پیمایش شود؛ X' جستجویی محلی بین راه‌حل‌های ساخته‌شدهٔ تاکنون را نگهداری می‌کند.X_{best} مقدار بهترین راه‌حل است.

پارامترهای α وβ مقادیر اختیاری دارند و در پیاده‌سازی الگوریتم مقادیر دلخواهی برای آنها می‌توان در نظر گرفت، در بیشتر پژوهش‌های صورت گرفته [24]، مقدار پارامتر α نزدیک به ۰/۵ و مقدار پارامترβ نزدیک به یک فرض شده است. پارامترρ مقدار محوشدن مورد استفاده برای شبیه‌سازی تبخیر فرومون است که$\rho \in [0,1]$ در نظر گرفته می‌شود. و τ_{ij} مقدار اولیه فرومون است، در ابتدا مقدار بسیار ناچیزی برای آن در نظر در نظر گرفته می‌شود.

۲-۴- الگوریتم مورچگان بر اساس تئوری مجموعه‌های راف (روش پیشنهادی)

به‌دلیل عدم کارایی الگوریتم های یادگیری استاندارد در مواجهه با مسایل نامتوازن، باید اقدام به اصلاح الگوریتم‌های حاضر کرد [25] که در این مقاله الگوریتم مورچگان با تئوری مجموعه راف تلفیق شده تا بتواند نتایج قابل قبولی ارائه کند که در ادامه به شرح آن پرداخته می‌شود.

در الگوریتم ارائه‌شده هر ویژگی به‌عنوان یک گره شناخته می‌شود؛ و هر مورچه به‌صورت مجزا راه‌حل خود را می‌سازد؛ سپس راه‌حل‌های ساخته‌شده با یکدیگر مقایسه شده و بهترین راه‌حل به‌عنوان زیرمجموعه بهینه انتخاب می‌شود. در حالت کلی ساخت راه‌حل در این الگوریتم شامل چهار گام اصلی است.

در گام نخست: یک ویژگی به تصادف انتخاب شده و مقادیر احتمال برای انتخاب ویژگی بعدی از بین ویژگی‌های انتخاب‌نشده محاسبه می‌شود، مقادیر احتمال بر اساس فرمول (۷) به‌دست می‌آید [26]:

$$p_{ij}^{k}(t)=\frac{\tau_{ij}^{\alpha}(t)*\eta_{ij}^{\beta}(t)}{\sum_{i\in u}\tau_{ij}^{\alpha}(t)*\eta_{ij}^{\beta}(t)} \qquad (۷)$$

هر ویژگی که مقدار احتمال بیشتری را داشته باشد، به‌عنوان ویژگی بعدی انتخاب می‌شود. در الگوریتم مورچگان که هدف، یافتن کوتاه‌ترین مسیر است مقدار η_{ij} برابر با عکس فاصله بین گره‌ها در نظر گرفته می‌شود و $\tau_{ij}(t)$ بیان‌گر مقدار اولیه فرومون است؛ اما در این پژوهش به‌دلیل اینکه هدف، یافتن زیرمجموعه‌های بهینه بر پایه مجموعه‌های راف است، این مقدار برابر درجه وابستگی i وj است، که بر اساس رابطه (۸) محاسبه می‌شود:

$$K=\gamma_p(Q)=\frac{|POS_p(Q)|}{|U|} \quad (۸)$$

که در آن $|POS_p(Q)|$ بیان‌گر رابطه هم‌ارزی P و Q بر روی مجموعه مرجع و در نواحی مثبت و $\gamma_p(Q)$ بیان‌گر میزان وابستگی P به Q است، که نحوه محاسبه آن در بخش ۲-۱ آورده شده است.

گام دوم: احتمالات به‌دست‌آمده در گام دوم با یکدیگر مقایسه و ویژگی‌ای که بیشترین مقدار را داشته به‌عنوان ویژگی بعدی انتخاب می‌شود.

گام سوم: با استفاده از رابطه(۹) مقدار فرومون به‌روزرسانی می‌شود. که در این رابطه، $\tau_{ij}(t)$ بیان‌گر مقدار پیشین فرومون و $\tau_{ij}(t+1)$ مقدار فرومون جدید است.

$$\tau_{ij}(t+1)=\rho\tau_{ij}(t)+(1-\rho)\eta_{ij} \quad (۹)$$

پس از به‌دست‌آوردن مقدار جدید فرومون، در مراحل بعدی مقدار جدید فرومون در رابطه (۷) جایگزین و انتخاب ویژگی جدید بر اساس آن تعیین می‌شود.

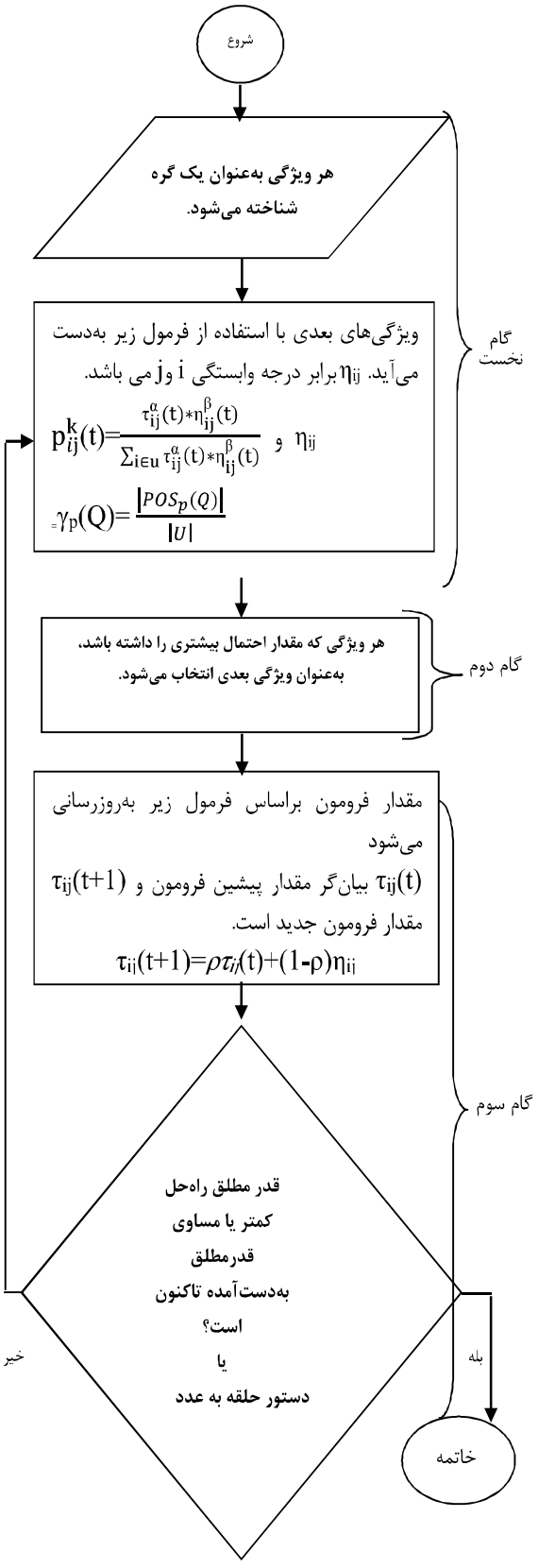
گام چهارم: در این گام شرایط توقف بررسی می‌شود. در الگوریتم ترکیبی آورده‌شده در این پژوهش زمانی که یکی از دو شرط زیر اتفاق بیفتد روند ساخت راه‌حل خاتمه می‌یابد:

- قدر مطلق راه‌حل کمتر یا مساوی قدرمطلق به‌دست‌آمده تاکنون باشد.
- دستور حلقه به عدد بیشینهٔ خود رسیده باشد.

شرط نخست به این معنی است که بهترین راه‌حل ساخته شده است و دومین شرط نشان می‌دهد که راه‌حل بهتری نمی‌تواند ساخته شود.

اگر هنوز شرط خاتمه تحقق نیافته باشد دوباره باید از گام نخست شروع کرد البته در این تکرار مقدار فرومون به‌دست‌آمده در گام سوم را باید جایگزین کرد. در بررسی شرط توقف، با توجه به اینکه در پایان تکرار نخست فقط مورچه نخست تمام ویژگی‌ها را انتخاب می‌کند؛ سایر مورچه ها نیز باید این مسیر را طی تا تمام تکرارها (یعنی تعداد مورچه‌ها) از این مسیر حرکت نمایند.

۱-۲-۴- روندنمای الگوریتم پیشنهادی



(شکل-۱):روندنمای الگوریتم پیشنهادی
(Figure-1): Proposed method flowchart

۵- نتایج تجربی

۵-۱- داده‌ها

در این پژوهش از ۴۷ متغیر کلیدی که از لحاظ تئوری می‌توانستند بر قیمت جهانی نفت تأثیرگذار باشند، استفاده شده است. فاصله زمانی مورد استفاده در این پژوهش از سال ۱۹۸۱ تا سال ۲۰۰۹ میلادی به‌صورت سالانه است. داده‌های موردنظر از طریق پایگاه اینترنتی اداره اطلاعات انرژی وزارت انرژی ایالات متحده امریکا[1]، سازمان بین‌المللی انرژی[2]، شرکت نفتی BP[3]، سازمان همکاری اقتصادی و توسعه (OECD)[4] و بانک جهانی[5] به‌دست آمده‌اند. لازم به ذکر است که عوامل استخراجی با مطالعه بازار نفت و عوامل تأثیرگذار بر آن، پژوهش‌های صورت گرفته در این زمینه و نیز استفاده از نظرات مشاوره‌ای اساتید خبره اقتصادی و کارشناسان بازار بین‌المللی نفت انتخاب شده است. دلیل استفاده از قیمت جهانی نفت خام در این پژوهش به‌عنوان مجموعه داده‌ها بر این اساس بوده است که نوسانات قیمت در این بازار زیاد بوده و عوامل زیادی به‌عنوان ویژگی‌های تأثیرگذار بر این بازار معرفی شده‌اند، که درنتیجه روند پیش‌بینی را نسبت به سایر مجموعه داده‌هایی که بیشتر از ۵ یا ۶ ویژگی تأثیرگذار را ندارند پیچیده‌تر می‌کند؛ درضمن بازه زمانی درنظرگرفته‌شده در این پژوهش از سال ۱۹۸۱ تا سال ۲۰۰۹ است؛ اگر چه داده‌های سال‌های آینده نیز در دسترس هستند، اما با بررسی‌ها و پژوهش‌های انجام‌شده مشخص شد در این دوره زمانی بازار نفت دست‌خوش تغییرات غیر قابل پیش‌بینی بوده و درنتیجه کمتر روشی توانسته است، قیمت نفت را در سال ۲۰۱۰ پیش‌بینی کند. ۴۷ ویژگی مورد استفاده در این پژوهش شامل موارد زیر است:

ذخایر اثبات‌شده نفت آمریکا، مصرف نفت OECD، میزان تولید نفت جهان، نفت OECD، نفت گروه ۷ و نفت آمریکا، توان عملیاتی پالایشگاه‌های نفت OECD، ذخایر اثبات‌شده گازطبیعی جهان، مصرف زغال‌سنگ جهان، OECD، آمریکا و چین، رشد اقتصادی امریکا، گروه ۷، چین و رشد اقتصادی جهان، ذخایر اثبات‌شده نفت آمریکای شمالی، آمریکای جنوبی، آمریکای مرکزی، اروپا، شوروی سابق، خاورمیانه، آفریقا، آسیا، منطقه پاسیفیک و کشورهای OPEC، OECD، Non-OPEC، عوامل سیاسی، رشد اقتصادی، آب‌وهوا، بازار فراورده‌ها، حمل‌ونقل نفت خام، تصمیمات OPECT، بازار آمریکا، ضریب اعتماد مصرف‌کنندگان، نرخ بهره، سفارش و تولیدات صنعتی، رونق ساخت‌وساز در بخش مسکن، نرخ بیکاری، تقویم تعطیلات، طرح‌های اقتصادی، موامل اجتماعی، تحریم‌های اقتصادی کشورهای تولیدکننده نفت خام، اعتصابات کارگری، ناآرامی‌های سیاسی، جنگ، انقلاب‌ها و عملیات‌های تروریستی.

۵-۲- ارزیابی بهینه‌بودن روش ارائه‌شده از طریق انتخاب ویژگی

برای پیاده‌سازی از الگوریتم‌های جانسون به‌عنوان الگوریتمی که به‌تنهایی از روش رافست به‌منظور استخراج ویژگی استفاده می‌کند و الگوریتم بهینه‌سازی ذرات که از روش‌های هوش تجمعی است و جستجوی هارمونیک که بر مبنای الگوریتم تپه‌نوردی است، درنهایت الگوریتم مورچگان بر پایه مجموعه راف که ترکیبی از روش‌های مجموعه‌های راف و هوش تجمعی است، ویژگی‌های تأثیرگذار استخراج و در جدول (۱) آورده شده است.

با پژوهش و بررسی بازار نفت مشخص شد که زیرمجموعه خروجی از مدل بهینه‌سازی کلونی مورچگان بر پایه مجموعه‌های راف، به‌عنوان زیرمجموعه بهینه می‌تواند انتخاب شود؛ زیرا ویژگی‌های استخراجی از این مدل تأثیرگذاری بیشتری در زمینه عرضه و تقاضا بر بازار نفت دارد. این ویژگی‌ها به‌تنهایی قیمت نفت را در یک سال می‌توانند تغییر دهند. حال آنکه برخی از ویژگی‌های استخراجی توسط روش‌های هارمونیک، جانسون و PSO گرچه بر بازار نفت تأثیرگذار بوده‌اند، اما میزان تأثیرگذاری به این اندازه نیست که به‌تنهایی بتوانند قیمت نفت را در یک سال تحت‌الشعاع خود قرار دهند و تغییراتی در آن ایجاد کنند. از جهت دیگر با مشاهده نتایج می‌توان نتیجه گرفت که روش جانسون که بر پایه تئوری مجموعه‌های راف است نسبت به سایر روش‌ها کم‌ترین تعداد ویژگی‌ها را استخراج کرده است و معیار کمینه‌بودن زیرمجموعه انتخابی در آن لحاظ شده است. بنابراین با استفاده از روش تئوری مجموعه‌های راف به زیرمجموعه‌ای کمینه می‌توان دست یافت و با ترکیب آن با الگوریتم مورچگان معیار بهینگی هم برآورده می‌شود.

[1] http://www.eia.gov/
[2] http://www.iea.org/
[3] http://www.bp.com/
[4] http://www.oecd.org/
[5] http://www.worldbank.com/

(جدول-۱): تعداد ویژگی‌های استخراج‌شده توسط مدل‌های پژوهش

(Table-1): Number of features extracted by research models

الگوریتم مورداستفاده	تعداد کل ویژگی‌ها	تعداد ویژگی‌های استخراج‌شده	ویژگی‌های استخراج‌شده
جانسون	47	4	ذخایر اثبات‌شده نفت آمریکا، میزان تولید نفت کشورهای عضو اوپک ، مصرف زغال OECD و رشد اقتصادی آمریکا
بهینه‌سازی جمعی ذرات	47	7	ذخایر اثبات‌شده نفت آمریکا، میزان مصرف نفت OECD، ظرفیت پالایشگاه‌های نفت OECD، ذخایر گاز طبیعی جهان، مصرف زغال سنگ آمریکا، مصرف زغال‌سنگ جهان، رشد اقتصادی آمریکا و رشد اقتصادی گروه۷
جستجوی هارمونیک	47	15	ذخایر اثبات‌شده نفت آمریکا، مصرف نفت OECD، میزان تولید نفت جهان، میزان تولید نفت OECD، میزان تولید نفت گروه۷، میزان تولید نفت آمریکا، توان عملیاتی پالایشگاه‌های نفت OECD، ذخایر اثبات‌شده گازطبیعی جهان، مصرف زغال‌سنگ جهان، مصرف زغال‌سنگ OECD، مصرف زغال‌سنگ آمریکا، مصرف زغال‌سنگ چین، رشد اقتصادی امریکا، رشد اقتصادی گروه۷ + چین و رشد اقتصادی جهان
بهینه‌سازی کلونی مورچگان بر پایه مجموعه‌های راف	47	6	ذخایر اثبات‌شده نفت OECD، مصرف نفت آمریکا، مصرف نفت جهان، رشد اقتصادی آمریکا و رشد اقتصادی گروه۷

۳-۵- ارزیابی بهینه‌بودن روش ارائه‌شده از طریق پیش‌بینی

اگر چه ویژگی‌های استخراجی توسط روش ارائه‌شده بیان‌گر عملکرد مناسب الگوریتم است؛ اما به‌منظور ارزیابی دقیق‌تر کارایی الگوریتم‌های مورد استفاده در این پژوهش از سه معیار بازخوانی (Recall)، دقت (Precision) و معیار F (F-measure) که در بیش‌تر پژوهش‌ها به‌عنوان معیار ارزیابی به کار می‌روند و به‌صورت فرمول‌های ۱۰ تا ۱۲ تعریف می‌شوند، استفاده و نتایج حاصل از آن در جدول (۲) ارائه شده است.

$$\Pr ecision = \frac{\sum \text{True positive}}{\sum \text{Test outcome positive}} \quad (۱۰)$$

$$\text{Re}call = \frac{\sum \text{True positive}}{\sum \text{Condition positive}} \quad (۱۱)$$

$$F - measure = 2 * \frac{\Pr ecision * \text{Re}call}{\Pr ecision + \text{Re}call} \quad (۱۲)$$

(جدول-۲): ارزیابی کارایی الگوریتم‌های مورد استفاده در پژوهش

(Table-2): Efficiency Evaluation of the algorithms used in the research

الگوریتم مورد استفاده	Precision	Recall	F_ Measure
جانسون	0.667	0.769	0.714
بهینه‌سازی جمعی ذرات	0.857	0.923	0.889
جستجوی هارمونیک	0.867	1	0.929
بهینه‌سازی کلونی مورچگان بر پایه مجموعه‌های راف	0.929	1	0.963

نتایج حاصل از ارزیابی بیان‌گر میزان دقت الگوریتم بهینه سازی کلونی مورچگان بر پایه مجموعه‌های راف در مقایسه با سایر الگوریتم‌های مورد استفاده در این پژوهش است و در هر سه معیار مورد بررسی مقادیر نزدیک به یک که مقدار ایده‌آل است به‌دست آورده است. اگر چه سایر الگوریتم‌های مورداستفاده برای مقایسه هم دارای عملکرد

مناسبی بوده و نتایجی بالاتر از ۰/۵ را به‌دست آورده‌اند؛ اما مقادیر حاصل از روش ترکیبی مورچگان بر پایه مجموعه‌های راف دقیق‌تر بوده و درنتیجه می‌توان مطمئن شد که روش ارائه‌شده هم از نظر ویژگی‌های استخراجی و هم از نظر کارایی بهینه‌تر از سایر الگوریتم‌ها است.

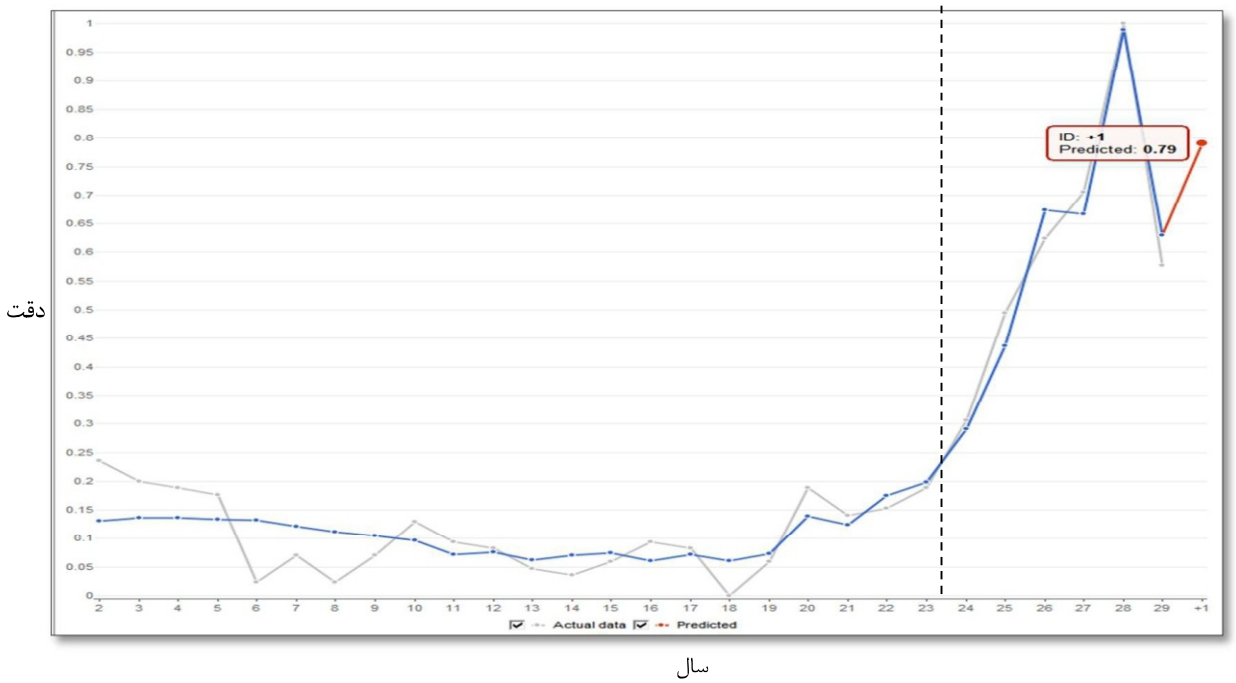
به‌منظور اثبات مطالب گفته‌شده و مشاهده بهتر عملکرد مناسب الگوریتم ارائه‌شده نسبت به روش‌های جانسون، بهینه‌سازی جمعی ذرات، جستجوی هارمونیک و شبکه عصبی پرسپترون با چهار لایه میانی (که از نرم‌افزار WEKA برای شبیه‌سازی استفاده شده است) نتایج حاصل از پیش‌بینی قیمت نفت خام در سال آینده با استفاده از ویژگی‌های استخراجی و پس از تبدیل داده‌های نرمال‌شده در جدول (۳) ارائه شده است.

نتایج نشان داده‌شده در جدول (۳) قدرت پیش‌بینی مدل ارائه‌شده در مقایسه با سه مدل دیگر را نشان می‌دهد و بیان می‌کند که الگوریتم ترکیبی مورچگان و مجموعه‌های راف بالاترین دقت را داشته، سپس مدل هارمونیک و پس از آن مدل جانسون و درنهایت مدل PSO.

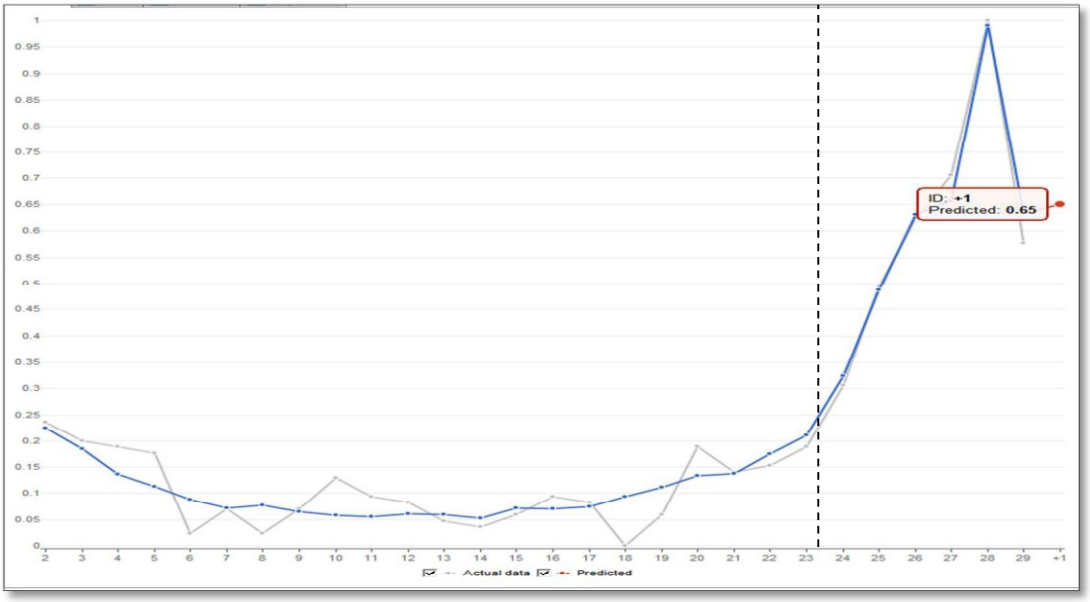
نتایج حاصل از پیاده‌سازی ویژگی‌های استخراجی در چهار مدل مطرح‌شده به‌صورت نمودار در ادامه ارائه شده است.

(جدول-۳): مقایسه دقت پیش‌بینی توسط مدل‌های پژوهش
(Table-3): Compare prediction accuracy of the algorithms used in the research

الگوریتم مورداستفاده	قیمت پیش‌بینی‌شده	قیمت واقعی
جانسون	83.14	79.61
بهینه‌سازی جمعی ذرات	67.61	79.61
جستجوی هارمونیک	87.06	79.61
شبکه عصبی پرسپترون	78.20	79.61
بهینه‌سازی کلونی مورچگان بر پایه مجموعه‌های راف	79.56	79.61

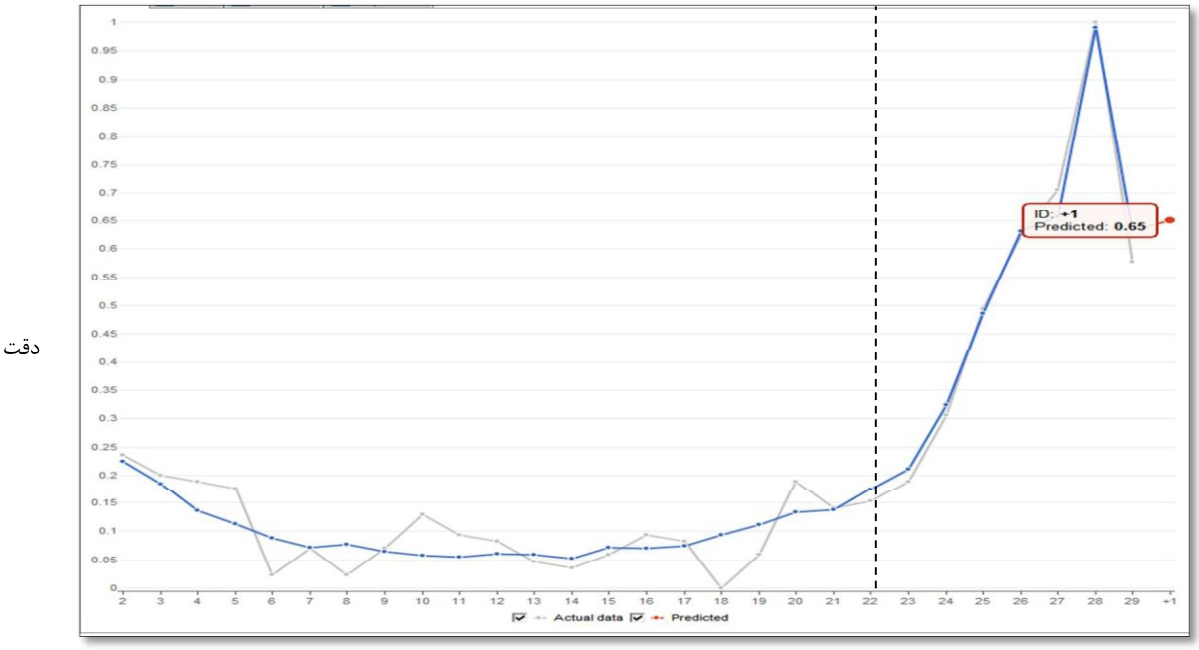


(شکل-۲): روند پیش‌بینی با روش مورچگان بر پایه مجموعه‌های راف
(Figure-2): Forecasting process using Ant method using rough set



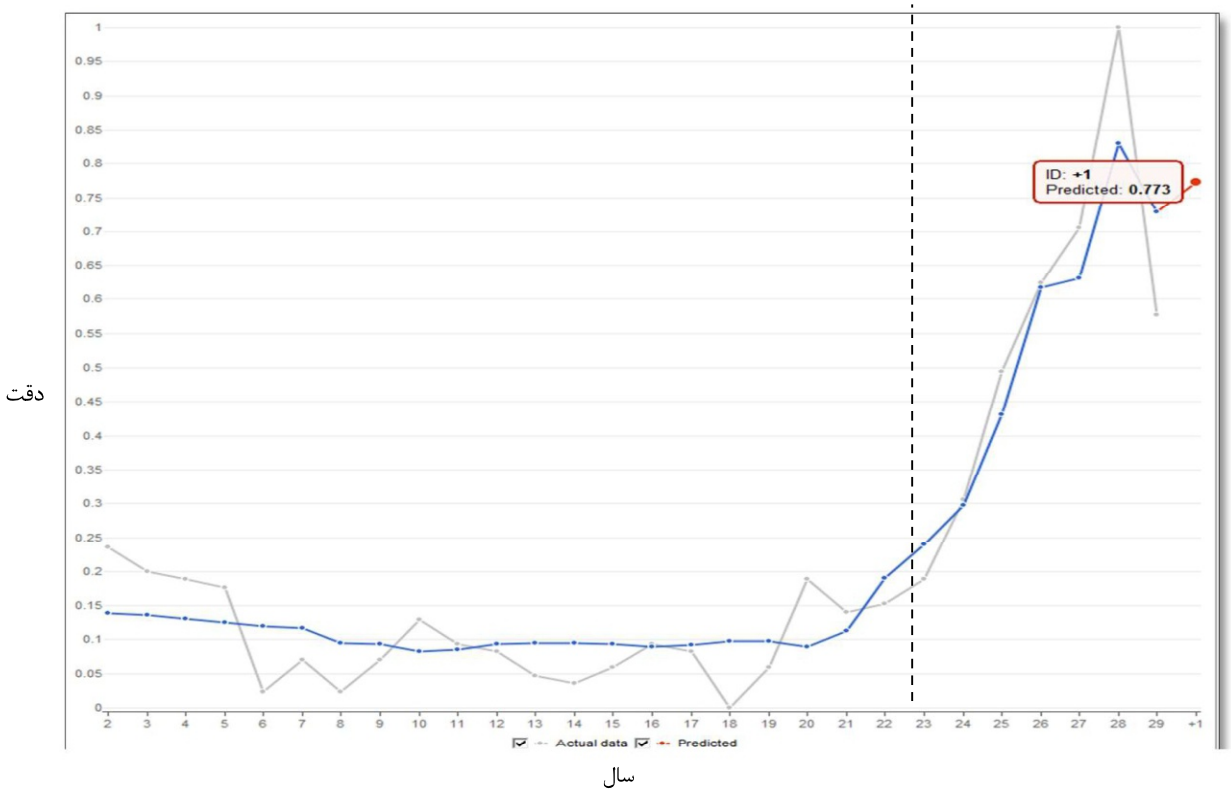
(شکل-۳): روند پیش‌بینی با روش جانسون

(Figure-3): Forecasting process using Johanson Algorithm



(شکل-۴): روند پیش‌بینی با روش PSO

(Figure-4): Forecasting process using PSO



(شکل-۵): روند پیش‌بینی با روش Harmonic Search

(Figure-5): Forecasting process using Harmonic Search

در روند پیش‌بینی حدود ۷۰٪ داده‌ها به بازه **تعلیم** و ۳۰٪ به بازه **آزمایش** تخصیص یافته است و درنهایت بر مبنای این بازه‌ها مقدار در یک سال آینده پیش‌بینی شده است؛ بنابراین در نمودارها تا سال بیست و دوم به‌عنوان بازه **تعلیم** مشخص شده است که به‌وسیله خطوط نقطه‌چین عمودی در شکل مشخص شده‌اند. بنابراین در ابتدای ترسیم، نمودار مربوط به هر مدل با نمودار واقعی تطبیق کمتری دارد؛ اما به‌مرور و با افزایش بازه زمانی نمودار مدل تطبیق بیش‌تری یافته و همان‌گونه که در شکل‌ها مشخص است، نمودار مجموعه‌های راف بر پایه مورچگان نسبت به سایر مدل‌ها سریع‌تر منطبق و مقادیر مناسب‌تری را ارائه داده است و به‌طورتقریبی از سال نوزدهم با نمونه واقعی تطبیق یافته و روند کاهش یا افزایش را مدل‌سازی کرده است؛ اما در سایر الگوریتم‌ها این تطبیق به‌طورتقریبی از سال بیست و دوم یا سوم آغاز شده است و این گونه استنباط می‌توان کرد که الگوریتم ارائه‌شده با روند اصلی سریع‌تر انطباق پیدا می‌کند و سریع‌تر به مرحله پیش‌بینی می‌رسد. پس از استخراج مقادیر پیش‌بینی‌شده توسط هر الگوریتم، عکس روند نرمال‌سازی بر روی مقادیر انجام و نتایج در جدول (۳) درج شده است.

همان‌گونه که در شکل‌ها مشخص است هر چهار روش مورداستفاده در این پژوهش افزایش قیمت را پیش‌بینی کرده‌اند و این بیان‌گر قدرت مناسب تمامی این الگوریتم‌ها است؛ اما با مقایسه دقیق‌تر می‌توان نتیجه گرفت که، نمودار مجموعه‌های راف بر پایه مورچگان نسبت به سایر مدل‌ها سریع‌تر منطبق و مقادیر مناسب‌تری را ارائه داده و به‌طورتقریبی از سال نوزدهم با نمونه واقعی تطبیق یافته و روند کاهش یا افزایش را مدل‌سازی کرده است؛ اما در سایر الگوریتم ها این تطبیق تقریبا از سال بیست و دوم یا سوم آغاز شده است و می‌توان این گونه استنباط کرد که الگوریتم ارائه‌شده با روند اصلی سریع‌تر انطباق پیدا می‌کند و سریع‌تر به مرحله پیش‌بینی می‌رسد.

۶- نتیجه‌گیری

در این مقاله، مفاهیم اولیه و زمینه‌های تئوری داده‌کاوی و کاربرد مجموعه‌های راف در زمینه استخراج دانش از مجموعه داده‌ها بیان شده است و می‌توان نتیجه گرفت که این تئوری یک ابزار قدرتمند در انتخاب زیرمجموعه‌های کمینه و حذف داده‌های زائد از پایگاه داده‌ها است؛ اما به

بالارفتن حجم اطلاعات و افزایش ویژگی‌ها به تناقضاتی برخورد و همین امر موجب حذف ناخواسته برخی از ویژگی‌های با ارزش و مؤثر می‌شود؛ درنتیجه برای رفع این مشکل از الگوریتم مورچگان که از جمله روش‌های هوش تجمعی بوده و عملکرد مناسبی در انتخاب بهترین‌ها از بین تمام مسیرهای ممکن دارد، به‌منظور رفع اشکال تئوری راف استفاده شده است و به‌منظور بررسی عملکرد روش ارائه‌شده از داده‌ها مربوط به قیمت نفت خام و ۴۷ ویژگی تأثیرگذار بر قیمت در بین سال‌های ۱۹۸۱تا ۲۰۰۹ استفاده شده و در مرحله نخست بر اساس چهار مدل بیان شده ویژگی‌های مؤثر استخراج و مورد ارزیابی قرار گرفته‌اند که مقایسه نتایج حاصل از ارزیابی حاکی از عملکرد مناسب الگوریتم مورچگان بر پایه مجموعه‌های راف دارد و در مرحله دو به‌منظور اثبات دقت الگوریتم ارائه‌شده قیمت نفت خام با استفاده از ویژگی‌های استخراجی در سال ۲۰۱۰ مورد پیش‌بینی قرار گرفته است با بررسی‌ها و پژوهش‌های انجام‌شده مشخص شد در این دوره زمانی بازار نفت دست‌خوش تغییرات غیر قابل پیش‌بینی بوده و درنتیجه کمتر روشی توانسته است قیمت نفت را در سال ۲۰۱۰ پیش بینی کند. قیمت واقعی نفت در سال ۲۰۱۰، به مبلغ ۷۹/۶۱ دلار به‌ازای هر بشکه بوده است. پس از پیش‌بینی قیمت با چهار مدل ذکرشده نتایج به‌ترتیب دقت برای الگوریتم ترکیبی مورچگان و مجموعه‌های راف، جستجوی هارمونیک، الگوریتم جانسون و الگوریتم بهینه‌سازی جمعی ذرات به این صورت استخراج شدند: ۷۹/۵۶، ۷۸/۰۶، ۸۳/۱۴ و ۶۷/۶۱ دلار در هر بشکه که با مقایسه مبالغ استخراجی مشاهده می‌شود که الگوریتم ارائه‌شده در این پژوهش نزدیک‌ترین مبلغ به مقدار واقعی را پیش‌بینی کرده است و بالاترین دقت را در بین سایر روش‌های بیان‌شده دارد.

۷- مراجع 7-References

[1] J. Wang, "Reduction Algorithms Based on Discernibility Matrix: The Ordered Attributes Method," *Journal of Computer Science & Technology*, Vol. 16, No. 6, pp. 489–504, 2001.

[2] H. Nakayama, Y. Hattori and R. Ishii, "Rule Extraction Based on Rough Set Theory and Its Application to Medical Data Analysis," *IEEE Conference on Data Analysis* 0-7803-5731-0/99, pp.924-929, 1999.

[3] Z. Pawlak, "Rough Sets: Theoretical Aspects of Reasoning about Data," *Kluwer Academic Publishers, Dordrecht, Boston*, London, 1991.

[4] Z. Pawlak, "Rough Sets and Data Analysis," *IEEE Conference Intelligent Processing Systems*, Beijing, China, October 28-31, 1996.

[5] A.S. Honby, Oxford Advanced Learners Dictionary of Current English," Oxford University Press, UK, 1974.

[6] Q. Zhang, Z. Han and F. Wen, "A New Approach for Fault Diagnosis in Power Systems Based on Rough Set Theory," *4th international Conference on Power Systems Control*, Operation and Management, APSCOM-97, Hong Kong, November 1997, pp.597-602, 1997.

[7] M. Toshinori, "Rough Control Application of Rough Set Theory to Control," *Computer and Information Science Department Cleveland State University*, 1996.

[8] W. Ziarko, "The Discovery, Analysis and Representation of Data Dependencies in Databases," *Knowledge Discovery in Databases*, AAAI MIT Press, Cambridge, MA, pp.213-228, 1993.

[9] L. Xiaolei and W. Xiaobing, "The Application of Rough Set Theory in Vehicle Transmission System Fault Diagnosis," *IEEE International Vehicle*, ISSN:0-7803-5296-3/99,1999, pp.240-242, 1999.

[10] S. Surekha, and G. Jaya Suma, "Swarm Intelligence and Variable Precision Rough Set Model: A Hybrid Approach for Classification." *Computational Intelligence Techniques in Health Care Part of the series Springer Briefs in Applied Sciences and Technology*, pp. 83-94, 2016.

[11] P. Langley, "Selection of relevant features in machine learning." *AAAI Fall Symposium on Relevance*, pp. 1-5, 1994.

[12] W. Siedlecki and J. Sklansky, "On automatic feature selection," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 2, No. 2, pp. 197-220, 1988.

[13] A. Raze and G. Nasajyan, "Application of Rough Theory in Decision-Making Theory", Conference on Electrical Engineering, Gonabad Branch, Islamic Azad University, 2016.

[14] D. Chouchoulas and Q. Shen, "Rough set-aided keyword reduction for text cat-egorisation." *Applied Artificial Intelligence*, Vol. 15, No.9, pp. 843-873, 2001.

[15] T. Beaubouef and R. Lang, "Rough Set Techniques for Uncertainty Management in Automated Story Generation," *36th Annual*

Conference on Southeast Regional Conference, April, pp.326-331, 1998.

[16] P. Hongxia, M. Qingfeng and W. Xiuye, "Research on Fault Diagnosis of Gearbox Based on Particle Swarm Optimization Algorithm," *IEEE 3rd International Conference on Mechatronics*, pp. 228-231, 2006.

[17] B. Xue, M. Zhang and W.N. Browne, "Particle Swarm Optimization for Feature election in Classification: A Multi-Objective Approach," *IEEE Transactions on Cybernetics*, Vol. 43, No. 6, pp. 56-71, 2013.

[18] C.O. Caio, D. Ramos, N.S. André, G. Chiachia, X. Alexandre and J.P. Papad, "A novel algorithm for feature selection using Harmony Search and its application for non-technical losses detection," *Computers & Electrical Engineering*, Vol. 37, pp. 886–894, 2011.

[19] R. Diao and Q. Shen, "Feature Selection with Harmony Search," IEEE Systems, *Man and Cybernetics Society*, pp.128-135, 2012.

[20] H. Tovhidi, H. Nezamabadi and S. sarozadi, "eature Selection Using Binary Ant colony Population Algorithm", Fisrt international conference on fuzzy systems, 2004.

[21] S. Kyanfar and M.R. Meybodi, "Provides an adaptive ant colony algorithm for solving continuous optimization problems", Fifth National Conference on Command and Control, 2010.

[22] K. Socha, and M. Dorigo, "Ant colony optimization for continuous domains," *European Journal of Operational Research*. Vol. 185, No. 3, pp. 1155-1173, 2008.

[23] B. De la Iglesia, "Evolutionary computation for feature selection in classification problems." *Data Mining and Knowledge Discovery*, Vol. 3, No. 6, pp. 381–407, 2013.

[24] D. Jia, X. Duan and M.K. Khan, "Binary Artificial Bee Colony optimization using bitwise operation." *Computer Industrial Engineering*, Vol. 76, pp. 360–365, 2014.

[۲۵] مهدی‌زاده محبوبه، افتخاری مهدی، ارائه روش جدید مبتنی بر برنامه‌نویسی ژنتیک برای وزن‌دهی قوانین فازی در طبقه‌بندی نامتوازن. پردازش علائم و داده‌ها. ۱۳۹۳؛ ۱۱ (۲): ۱۱۱-۱۲۵.

[25] M. Mahdizadeh and M. Eftekhari, "A new fuzzy rules weighting approach based on Genetic Programming for imbalanced classification." *JSDP*. Vol. 11, No 2, pp. 111-125, 2015.

[26] M. Dorigo and G.D. Caro, "Ant colony optimization: A new meta-heuristic," *Congress on Evolutionary Computing*, pp. 17-26, 1999.

الهه میرزائی مدرک کارشناسی و کارشناسی ارشد خود را در رشته مهندسی کامپیوتر-نرم‌افزار از دانشگاه آزاد اسلامی واحد همدان در سال های ۱۳۸۸ و ۱۳۹۳ دریافت کرد. زمینهٔ پژوهشی مورد علاقه ایشان داده‌کاوی و سامانه‌های یادگیر است.

نشانی رایانامه ایشان عبارت است از:

mirzaee634@gmail.com

منصور اسماعیل‌پور مدرک کارشناسی و کارشناسی ارشد خود را در رشته مهندسی کامپیوتر-نرم‌افزار در سال‌های ۱۳۸۱ و ۱۳۸۳ دریافت کرد. مدرک دکترا را در رشته مهندسی کامپیوتر-هوش مصنوعی از دانشگاه ملی مالزی در سال ۲۰۱۲ اخذ و از سال ۲۰۱۲ تا ۲۰۱۴ به‌عنوان فوق دکترا در همان دانشگاه به پژوهش پرداخت. وی از سال ۱۳۸۱ تاکنون عضو هیأت علمی گروه مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد همدان است. زمینه پژوهشی مورد علاقه ایشان داده‌کاوی، فرآیندکاوی، یادگیری ماشین و سامانه‌های یادگیر است.

نشانی رایانامه ایشان عبارت است از:

esmaeilpour@iauh.ac.ir