روش نوین خوشه‌بندی ترکیبی با استفاده از سامانهٔ ایمنی مصنوعی و سلسله‌مراتبی

احمد رضا جعفریان مقدم[1*]، فرناز برزین‌پور[2] و محمد فتحیان[3]

[1*] دانشکده عمران و حمل و نقل، دانشگاه اصفهان، اصفهان ، ایران

[2، 3] دانشکده مهندسی صنایع، دانشگاه علم و صنعت ایران، تهران ، ایران

چکیده

سامانه ایمنی مصنوعی (AIS) یکی از مهم‌ترین الگوریتم‌های متاهیوریستیک به‌منظور حل مسائل بسیار پیچیده است. از این الگوریتم می‌توان در تحلیل خوشه‌بندی داده‌ها استفاده کرد. علی‌رغم اینکه AIS قادر است پیکربندی فضای جستجو را به‌خوبی نمایش دهد؛ اما تعیین خوشه‌های داده‌ها به‌طور مستقیم با استفاده از خروجی آن بسیار مشکل است. بر این اساس در این مقاله الگوریتم دومرحله‌ای پیشنهاد شده است. در مرحله نخست با استفاده از الگوریتم AIS پیشنهادی، فضای جستجو مورد بررسی قرار گرفته و پیکربندی فضا تعیین می‌شود و در مرحله دوم با استفاده از روش خوشه‌بندی سلسله‌مراتبی، خوشه‌ها و تعداد آنها مشخص می‌شود. در انتها الگوریتم پیشنهادی بر روی نمونه واقعی متشکل از داده‌های زلزله در ایران پیاده‌سازی و با نتایج الگوریتم مشابه مقایسه شده است. نتایج نشان داد که الگوریتم پیشنهادی توانسته است، نقایص موجود در AIS و روش خوشه‌بندی سلسله‌مراتبی را پوشش دهد و از طرفی از دقت و سرعت قابل قبولی برخوردار است.

واژگان کلیدی: تحلیل خوشه‌بندی؛ سامانه ایمنی مصنوعی (*AIS*)؛ خوشه‌بندی سلسله‌مراتبی.

New Clustering Technique using Artificial Immune System and Hierarchical technique

Ahmad Reza Jafarian-Moghaddam[1*], Farnaz Barzinpour[2] & Mohammad Fathian[3]

[1*] Faculty of Civil Engineering and Transportation, University of Isfahan, Isfahan, Iran

[2, 3] School of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran[2]

Abstract

Artificial immune system (AIS) is one of the most meta-heuristic algorithms to solve complex problems. With a large number of data, creating a rapid decision and stable results are the most challenging tasks due to the rapid variation in real world. Clustering technique is a possible solution for overcoming these problems. The goal of clustering analysis is to group similar objects.

AIS algorithm can be used in data clustering analysis. Although AIS is able to good display configure of the search space, but determination of clusters of data set directly using the AIS output will be very difficult and costly. Accordingly, in this paper a two-step algorithm is proposed based on AIS algorithm and hierarchical clustering technique. High execution speed and no need to specify the number of clusters are the benefits of the hierarchical clustering technique. But this technique is sensitive to outlier data.

So, in the first stage of introduced algorithm using the proposed AIS algorithm, search space was investigated and the configuration space and therefore outlier data are determined. Then in second phase, using hierarchical clustering technique, clusters and their number are determined. Consequently, the first

* نویسندهٔ عهده‌دار مکاتبات

[DOI: 10.18869/acadpub.jsdp.13.4.133]

stage of proposed algorithm eliminates the disadvantages of the hierarchical clustering technique, and AIS problems will be resolved in the second stage of the proposed algorithm.
In this paper, the proposed algorithm is evaluated and assessed through two metrics that were identified as (i) execution time (ii) Sum of Squared Error (SSE): the average total distance between the center of a cluster with cluster members used to measure the goodness of a clustering structure. Finally, the proposed algorithm has been implemented on a real sample data composed of the earthquake in Iran and has been compared with the similar algorithm titled Improved Ant System-based Clustering algorithm (IASC). IASC is based on Ant Colony System (ACS) as the meta-heuristics clustering algorithm. It is a fast algorithm and is suitable for dynamic environments. Table 1 shows the results of evaluation.

Table 4: Compare the two algorithms

Alg.	IASC	Proposed algorithm
Execution time (s)	18	12
SSE	9/4	5/3

The results showed that the proposed algorithm is able to cover the drawbacks in AIS and hierarchical clustering techniques and the other hand has high precision and acceptable run speed.

Keywords: **Clustering Analysis; Artificial immune system (AIS); Hierarchical Clustering.**

۱ـ مقدمه

سامانهٔ ایمنی طبیعی (NIS[1]) شبکه پیچیده‌ای از بافت‌ها، سلول‌ها و گلبول‌ها است که وظیفه‌اش حفظ بدن در مقابل عوامل بیماری‌زا و همچنین کاهش صدمات وارده به بدن و اطمینان از عملکرد پیوسته اجزای خود است. بدین منظور سامانه ایمنی یک سامانه خودکنترل محسوب می‌شود. این سامانه بسته به نوع حملاتی که توسط عوامل بیماری‌زا به بدن صورت می‌گیرد، دارای عملکرد متفاوتی بوده و برای این منظور دارای سطوح مختلف عملکردی است. AIS[2] دارای تعاریف مختلفی متناسب با زمینه‌های کاربرد آن است. در [2]، AIS را یک روش‌شناسی هوشمند برگرفته از NIS می‌داند که قادر است، مسائل جهان واقعی را حل کند. در این تعریف AIS به‌عنوان یک روش‌شناسی در نظر گرفته شده است. [3] AIS را یک سامانه محاسباتی برگرفته از NIS تشریح کرده است. تعریف کامل‌تر برای AIS در [4] ارائه شده است. براساس این تعریف AIS یک سامانه تطابقی برگرفته از نظریه ایمنی‌شناسی است که وظایف، فرآیندها، مدل‌ها و اصول NIS را مورد توجه قرار می‌دهد. از AIS در زمینه‌های مختلفی چون مسائل یادگیری ماشین [5]، مسائل بهینه‌سازی [6] ،[7]، داده‌کاوی [8]، ایمنی رایانه‌ها [11]-[9]و غیره استفاده شده است.

روش خوشه‌بندی از مهم‌ترین روش‌های داده‌کاوی است که امروزه اهمیت آن در دنیای واقعی بر کسی پوشیده نیست. با بزرگ‌ترشدن بانک‌های داده‌ای، تلاش پژوهش‌گران به‌منظور یافتن روش‌های خوشه‌بندی کارا و مؤثر متمرکز شده است تا بدین طریق بتوانند زمینه تصمیم‌گیری سریع و منطبق با واقعیت را فراهم آورند. تحلیل خوشه‌بندی شاخه‌ای از تحلیل آماری چندمتغیره بوده و روشی به‌منظور گروه‌بندی داده‌های مشابه در خوشه‌های یکسان است [12]، [13]. این روش روشی بر پایه داده است که سعی دارد با جستجوی مشابهت‌های بین داده‌ها، آنها را خوشه‌بندی کند [12]،[13]. این شباهت براساس قوانین نظیر نزدیک‌ترین فاصله، دورترین فاصله و غیره می‌تواند باشد. روش‌های خوشه‌بندی سعی دارند با کشف روابط موجود در بین داده‌های جدید، روش خوشه‌بندی خود را بهبود بخشند ازاین‌رو روش‌های خوشه‌بندی به روش‌های یادگیرنده نیز شهرت یافته‌اند [14]،[15] به‌نحوی که قادرند پس از تعیین خوشهٔ داده‌های مختلف، خوشهٔ داده جدید که به مجموعه اضافه می‌شود با صرف کم‌ترین زمان مشخص کند [16]. انواع روش خوشه‌بندی عبارتند از: خوشه‌بندی سلسله‌مراتبی[3] [14]،[17]، خوشه‌بندی مختلط[4] [18]،[19]، خوشه‌بندی شبکه یادگیرنده[5] [23]-[20]، خوشه‌بندی براساس تابع هدف[6] و خوشه‌بندی افراز[7] [24]،[25].

AIS یکی از مهم‌ترین الگوریتم‌های فرا ابتکاری[8] به منظور حل مسائل بسیار پیچیده است که از مهم‌ترین کابردهای آن می‌توان در تحلیل خوشه‌بندی داده‌ها بیان

[1] Natural Immune System
[2] Artificial Immune System

[3] hierarchical clustering
[4] mixture-model clustering
[5] learning network clustering
[6] objective-function-based clustering
[7] partition clustering
[8] Meta-heuristics

[DOI: 10.18869/acadpub.jsdp.13.4.133]

نمود. علی‌رغم اینکه AIS قادر است پیکربندی فضای جستجو را به‌خوبی نمایش دهد، اما تعیین خوشه‌های داده‌ها به‌طورمستقیم با استفاده از خروجی آن بسیار مشکل و هزینه‌بر خواهد بود. این موضوع تاکنون در مطالعات گذشته مورد توجه قرار نگرفته است.

بر این اساس در این مقاله یک الگوریتم دو مرحله‌ای پیشنهاد شده است. در مرحله نخست با استفاده از الگوریتم AIS پیشنهادی، فضای جستجو مورد بررسی قرار گرفته و پیکربندی فضا تعیین می‌شود و در مرحله دو با استفاده از روش خوشه‌بندی سلسله‌مراتبی، خوشه‌ها و تعداد آنها مشخص می‌شود. از مزایای روش خوشه‌بندی سلسله‌مراتبی عدم نیاز به تعیین تعداد خوشه‌ها است، ولی از طرفی این روش به داده‌های پرت حساس است و قادر به تعیین ساختار داده‌ها در بانک داده‌های بزرگ نیست [26]. بنابراین در مرحله نخست الگوریتم با استفاده از الگوریتم پیشنهادی AIS با تعیین پیکربندی فضای داده‌ها، ضعف روش خوشه‌بندی سلسله‌مراتبی در تعیین داده‌های پرت و ساختار داده‌ها مرتفع خواهد شد و از طرفی در مرحله دوم قادر خواهیم بود با استفاده از روش خوشه‌بندی سلسله‌مراتبی نقایص موجود در AIS در تعیین خوشه‌ها و تعداد آن را پوشش دهیم. در انتهای مقاله، الگوریتم پیشنهادی بر روی نمونه واقعی متشکل از داده‌های زلزله در ایران پیاده‌سازی شده است. نتایج نشان داد که الگوریتم پیشنهادی توانسته است نقایص موجود در AIS و روش خوشه‌بندی سلسله‌مراتبی را پوشش دهد و از طرفی از دقت و سرعت بالا در اجرا برخوردار است.

ادامه مقاله به‌صورت زیر بخش‌بندی شده است؛ بررسی مطالعاتی که از الگوریتم‌های فرابتکاری به‌منظور خوشه‌بندی داده‌ها استفاده کرده‌اند، در بخش دوم مورد توجه قرار می‌گیرد. بخش سوم مقاله مروری بر AIS و انواع الگوریتم‌های آن دارد. الگوریتم پیشنهادی در بخش چهارم ارائه شده است و در بخش پنجم داده‌های مرتبط با نمونه واقعی متشکل از داده‌های زلزله در ایران با استفاده از الگوریتم پیشنهادی خوشه‌بندی می‌شود. نتایج حاصل از اجرای این الگوریتم در بخش پایانی مقاله مورد توجه قرار گرفته است.

۲- مروری بر ادبیات موضوع

همان‌طور که درقبل اشاره شد، روش خوشه‌بندی از مهم‌ترین روش‌های داده‌کاوی است که با بزرگ‌ترشدن بانک‌های داده‌ای، تلاش پژوهش‌گران به‌منظور یافتن روش‌های خوشه‌بندی کارا و مؤثر متمرکز شده است. بر این اساس اخیرا استفاده از الگوریتم‌های فرا ابتکاری به‌منظور ارائه الگوریتم‌های خوشه‌بندی سریع و با دقت مناسب بسیار رواج یافته است.

در [28]،[27] شبکهٔ عصبی مصنوعی با بردار موقعیت دوبعدی به‌منظور تعیین تعداد خوشه‌ها در بانک داده ترکیب شده است. مدل ارائه‌شده در [29] شبکه مصنوعی مبتنی بر تئوری تشدید انطباقی (ART) است که قادر است تعداد خوشه‌های واقعی را تعیین کند.

در [7] یک الگوریتم k میانگین مبتنی بر الگوریتم ژنتیک پیشنهاد شده است و در [30] الگوریتم ارائه‌شده در [7] به‌منظور خوشه‌بندی با استفاده از تئوری زنجیرهٔ مارکو بهبود یافته است و توانسته جواب بهینهٔ عمومی را ارائه دهد.

در [22] الگوریتم خوشه‌بندی مبتنی بر ACS پیشنهاد داده‌اند که در آن از استراتژی مورچه‌های مطلوب استفاده شده است. در این مقاله از الگوریتم SA[1] به‌منظور کاهش تعداد شهرهای بازدیدشده توسط مورچه‌ها و از استراتژی انتخاب مسابقه[2] به‌منظور یافتن بهترین مسیرها بهره گرفته شده تا الگوریتم سریع‌تری ارائه شود. خوشه‌بندی در این الگوریتم با استفاده از فرومون‌های در نظر گرفته شده برای هر شهر براساس بازدید مورچه‌های مطلوب صورت می‌گیرد. در [31] الگوریتمی به‌منظور خوشه‌بندی داده‌های ارائه شده است که در آن هر شهر (داده) بیان‌گر یک مورچه است و از طرفی مورچه‌ها دارای خصوصیات و ویژگی‌های متفاوتی هستند. در [32] الگوریتم خوشه‌بندی بر پایه سامانه مورچگان ارائه شده است و در [33] الگوریتم ارائه‌شده در [32] با الگوریتم خوشه‌بندی k میانگین ترکیب شده تا الگوریتم قوی‌تری ارائه شود. در [34] الگوریتم‌های ارائه‌شده در [32] و [33] را مورد توجه قرار داده و یک الگوریتم دومرحله‌ای را ارائه کرده است. در مرحله نخست الگوریتم، داده‌ها با استفاده از الگوریتم خوشه‌بندی بر پایه سامانه مورچگان و الگوریتم k میانگین مورچگان، خوشه‌بندی شده و در مرحله دوم با استفاده از الگوریتم مبتنی بر ACS قوانین انجمنی در هر یک از خوشه‌ها را استخراج کرده است. در [11]-[9]، [37]-[35] به موضوع استفاده از الگوریتم‌های ایمنی مصنوعی در خوشه‌بندی پرداخته است. مقایسات نتایج الگوریتم‌های مختلف فرا ابتکاری به‌منظور خوشه‌بندی داده‌ها نشان از سرعت بسیار بالای اجرا الگوریتم‌های ایمنی

[1] Simulated Annealing
[2] Tournament selection strategy

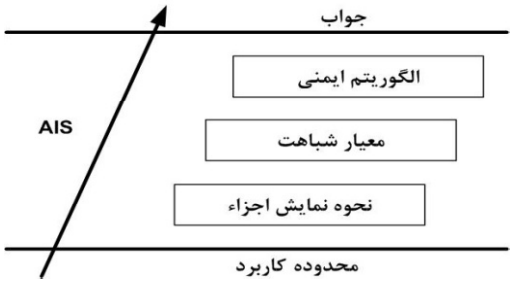
[DOI: 10.18869/acadpub.jsdp.13.4.133]

مصنوعی دارد. در این مقاله نیز با توجه به اینکه الگوریتم‌های ایمنی مصنوعی سرعت بسیار بالا و دقت مناسبی در اجرا دارند. از این الگوریتم‌ها به‌منظور خوشه‌بندی داده‌ها استفاده شده است. از طرفی یکی از مهم‌ترین مزایای روش‌های خوشه‌بندی سلسله‌مراتبی، سرعت اجرای بالای آنها و عدم نیاز به تعیین تعداد خوشه است. بر این اساس مرحله دوم الگوریتم پیشنهادی در این مقاله، به‌منظور افزایش سرعت خوشه‌بندی از این روش استفاده کرده است. الگوریتم ایمنی مصنوعی با تعیین تعدادی پادتن که بسیار کمتر از پادزاها هستند، فضای جستجو را مورد بررسی قرار می‌دهند و الگوریتم سلسله‌مراتبی پادتن‌ها و درواقع فضای جستجو را خوشه‌بندی می‌کند.

۳- مروری بر AIS و الگوریتم‌های آن

همان‌طورکه درقبل اشاره شد AIS یکی از مهم‌ترین الگوریتم‌های فراابتکاری به‌منظور حل مسائل بسیار پیچیده است که یک سامانه تطابقی برگرفته از نظریه ایمنی‌شناسی است که وظایف، فرآیندها، مدل‌ها و اصول NIS را مورد توجه قرار می‌دهد [38]. اجزای اصلی NIS عبارتند از: پادزا (Ag[1])، پادتن (Ab[2])، تکثیر سلولی[3] و لیمفوسیت (ALC[4]) که نوعی گلبول سفید است و دارای سلول‌های ایمنی B و T است.

در NIS میزان ارتباط بین یک پادتن و یک پادزا را نزدیکی (affin[5]) می‌نامند. هر چه مقدار نزدیکی کمتر باشد، گلبول‌های سفید بیشتر برانگیخته خواهند شد. بر این اساس نحوه عمل NIS به این صورت است: با حملهٔ عامل خارجی به بدن پادزاها تولید می‌کند؛ پادزاها شناسایی شده و به ALC اطلاع داده می‌شود؛ ALC برانگیخته شده و با استفاده از سلول‌های B و T خود برای نابودی پادزا اقدام می‌کند؛ سلول‌های B تکثیر شده و برخی بدون تغییر مانده و برخی به‌صورت پلاسما[6] تغییر می‌یابند. پلاسماها تولید پادتن کرده و پادتن‌ها مانع رشد پادزاها شده و از طرفی باعث تقویت سامانه ایمنی می‌شوند (فرایند یادگیرنده). عملکرد AIS بسیار شبیه NIS است. به‌منظور طراحی یک الگوریتم AIS، سه پارامتر مهم را باید مورد توجه قرار داد. نمودار (۱) این پارامترها و ارتباط آنها را نشان داده است.



(نمودار-۱): پارامترها در طراحی الگوریتم AIS
(Figure-1): AIS algorithm design parameters

الگوریتم‌های AIS به دو دسته کلی تقسیم می‌شوند که عبارتند از: ۱- الگوریتم‌های مبتنی بر جمعیت[7]، در این الگوریتم‌ها شبکه ایمنی مورد توجه قرار نمی‌گیرد. یکی از مهم‌ترین این الگوریتم‌ها تئوری انتخاب تکثیر سلولی[8] است [39]. از دیگر الگوریتم‌های مطرح در این زمینه CLONAG است که در [40] ارائه شده است. الگوریتم دیگر در این زمینه تئوری انتخاب منفی[9] است [11] که این فرآیند را فرآیند حذف سلول‌ها و تشخیص خودی از غیر خودی می‌توان نام نهاد. در [6]،[35] از این فرآیند برای حل مسائل بهینه‌سازی استفاده شده است. ۲- الگوریتم‌های مبتنی بر شبکه[10]، این الگوریتم‌ها برگرفته از تئوری شبکه سامانه ایمنی بدن می‌باشند. از مهم‌ترین این الگوریتم‌ها تئوری شبکه ایمنی[11] است [36] که در تحلیل خوشه‌بندی داده‌ها بسیار کاربرد دارد. یکی دیگر از مشهورترین الگوریتم‌های AIS بر پایه شبکه ایمنی aiNET نام دارد که در [37] ارائه شده است.

الگوریتم ارائه شده در روش پیشنهادی در این مقاله، بسیار شبیه aiNET است که برخی اصلاحات بر روی آن صورت گرفته است. از جمله مشکلات aiNET، تعیین پادزاها و تعداد آنها، شرط توقف الگوریتم و تعیین خوشه‌ها و تعداد خوشه‌های نهایی است. در این مقاله در ابتدا سعی شده است، بهبودهایی در الگوریتم aiNET صورت گیرد و با استفاده از روش خوشه‌بندی سلسله‌مراتبی در مرحله دوم روش پیشنهادی، سایر نقایص aiNET مرتفع شود.

[1] Antigens
[2] Antibodies
[3] Clone
[4] Lymphocytes
[5] Affinity
[6] Plasma
[7] Population base
[8] Clonal selection theory
[9] Negative selection theory
[10] Network base
[11] Immune network theory

[DOI: 10.18869/acadpub.jsdp.13.4.133]

۴- الگوریتم پیشنهادی

در این بخش الگوریتم پیشنهادی در دو مرحله ارائه می‌شود. متغیرها و پارامترهای مورد نیاز برای این الگوریتم در خلال ارائه مراحل الگوریتم، تشریح شده‌اند. لازم به ذکر است که قبل از اجرای الگوریتم داده‌ها باید بین اعداد صفر و یک نرمال شوند.

مرحله نخست الگوریتم (نمودار ۲)

۱- تمام داده‌های موجود در بانک داده را به‌عنوان پادزاها در نظر بگیر (N : تعداد آنتی ژن).

۲- به تعداد N_1 از پادتن به‌طور تصادف ایجاد کن. بعد Ab_i و Ag_j باید برابر باشد. ($i=1, \ldots, N_1$)

۳- مراحل زیرا را تا رسیدن به شرط توقف اجرا کن.

۳-۱- مراحل زیر را برای هر کدام از Ag_j تکرار کن: ($j=1, \ldots, N$)

۳-۱-۱- میزان نزدیکی (f_{ij}) هر یک از Ab_i را با Ag_j محاسبه کن.

- f_{ij} برابر است با معکوس فاصله اقلیدسی Ab_i با Ag_j ($f_{ij}=1/D_{ij}$)
- فاصله‌های اقلیدسی در مجموعه D ذخیره می‌شود.

۳-۱-۲- پادتن‌ها را براساس f_{ij} به‌صورت نزولی مرتب کن.

۳-۱-۳- به تعداد n تا از پادتن‌ها را با بالاترین f_{ij} انتخاب کن. ($Ab_{\{n\}}$)

۳-۱-۴- تعداد تکثیر سلولی برای هر یک از پادتن‌های انتخاب‌شده را با استفاده از معادله (۱) [۳۷] تعیین کن.

$$N_c = \sum_{i=1}^{n} round(N - D_{ij}.N) \qquad (1)$$

- تعداد تکثیر متناسب با میزان نزدیکی است. هر چه میزان نزدیکی یک پادتن به پادزا بیشتر باشد، تعداد بیشتری از پادتن تکثیر می‌شود.
- در معادله (۱)، N تعداد پادتن‌ها است و عمل‌گر round(a) عدد صحیح کوچک‌تر یا مساوی a را برمی‌گرداند.

۳-۱-۵- مجموعه C را از پادتن‌های تکثیرشده ایجاد کن.

۳-۱-۶- عمل جهش را بر روی سلول‌های تکثیرشده در مجموعه C با استفاده از معادله (۲) [37] انجام بده و مجموعه C^* را ایجاد کن.

$$C_k^* = C_k + rand(0,1) \times D_{kj}.mi\left(Ag_j - C_k\right) \qquad (2)$$

- در معادله (۲)، k (k=1, ..,K) برابر تعداد سلول‌های تکثیرشده در مجموعه C است. mi نرخ یادگیری نام دارد و یک پارامتر ورودی است. عمل‌گر *rand(0, 1)* یک عدد تصادفی بین صفر و یک تولید می‌کند. D_{kj} متناسب با فاصله والد سلول k از Ag_j است.

۳-۱-۷- میزان نزدیکی سلول‌های موجود در مجموعه C^* را با Ag_j محاسبه کن.

۳-۱-۸- qi درصد از بهترین سلول‌ها در مجموعه C^* را براساس میزان نزدیکی محاسبه شده، انتخاب کن.

۳-۱-۹- سلول‌های انتخاب‌شده را در مجموعه M_j قرار بده.

۳-۱-۱۰- سلول‌های موجود در مجموعه M_j که فاصله آنها با Ag_j بیشتر از مقدار tp می‌باشد از مجموعه M_j حذف کن.

- tp آستانه هرس[1] نام دارد و یک پارامتر ورودی برای الگوریتم است.
- این مرحله تنها پادتن‌هایی را باقی می‌گذارد که دارای بیشترین شباهت به پادزا مورد نظر باشند.

۳-۱-۱۱- میزان نزدیکی سلول‌های موجود در مجموعه M_j را نسبت به یکدیگر محاسبه کن.

۳-۱-۱۲- سلول‌هایی را از مجموعه M_j که میزان فاصله آنها کمتر از مقدار ts است از مجموعه M_j حذف کن.

- ts آستانه مرگ[2] نام دارد که یک پارامتر ورودی برای الگوریتم است.
- میزان فاصله متناسب با فاصله اقلیدسی است.
- در این مرحله پادتن‌های نزدیک به یکدیگر حذف و بر این اساس شبکه بین پادتن‌ها تشکیل می‌شود.

۳-۱-۱۳- مجموعه M^* را از جمع پادتن‌ها و سلول‌های موجود در مجموعه M_j تشکیل بده.

۳-۲- میزان نزدیکی سلول‌های موجود در M^* را نسبت به یکدیگر محاسبه کن.

۳-۲-۱- سلول‌هایی را از مجموعه *M که میزان فاصله آنها کمتر از مقدار ts است از مجموعه M_j حذف کن.

- در این مرحله تعداد شبکه موجود در فضای مورد بررسی، تعیین شده و معادله بین شبکه‌هایی که از یکدیگر دور هستند، قطع می‌شود. در نتیجه هر مجموعه شبکه از پادتن‌ها یک خوشه را ایجاد خواهد کرد.

۳-۳- تعداد پادتن جدید به‌صورت تصادفی ایجاد کن.

۳-۴- *gen=gen+1*

۳-۴- پادتن‌های جدید تولید شده و پادتن‌های موجود در مجموعه M^* را به‌عنوان مجموعه پادتن‌ها در نظر بگیر.

[1] Prune threshold
[2] Suppression threshold

[DOI: 10.18869/acadpub.jsdp.13.4.133]

۴- بررسی شرط توقف.

- در این مرحله شرط توقف به‌صورت زیر در نظر گرفته می‌شود. تازمانی‌که تعداد تولید از پادتن‌ها (gen) به تعداد تولید مورد نظر (Maxgen) نرسیده باشد و میزان بهترین نزدیکی به پادزا بیشتر از یک مقدار مشخص (sc=0.01) باشد، اجرای الگوریتم ادامه خواهد داشت.

۵- ارائه مجموعه M^* به‌عنوان خروجی الگوریتم

- مجموعه M^* مجموعه‌ای از پادتن‌ها خواهد بود که پیکربندی فضای مورد نظر را به خوبی تعیین خواهند کرد.

مرحله دوم الگوریتم

۱- دریافت مجموعه M^* به‌عنوان ورودی در این مرحله

۲- تعیین فاصله بین پادتن‌های موجود در M^*.

۳- تعیین سطح برش برای روش خوشه‌بندی سلسله‌مراتبی.

- در این مرحله سطح برش (cut) در نمودار سلسله‌مراتبی تعیین می‌شود که 0<cut<1 است.

۴- اجرای روش خوشه‌بندی سلسله‌مراتبی.

- در این روش از روش نزدیک‌ترین فاصله بین داده‌های خوشه‌ها[1] استفاده می‌شود؛ زیرا این روش در مواردی که خوشه‌ها دارای شکل‌های نامنظم هستند، بسیار کارا عمل می‌کند و از طرفی خوشه‌های بزرگ را به خوشه‌های کوچک‌تر تبدیل نمی‌کند. با این وجود این روش به داده‌های پرت حساس است که مرحله نخست این نقص را برطرف می‌کند.

۵- ارائه خروجی الگوریتم.

- خروجی الگوریتم خوشه‌ها و تعداد خوشه‌ها خواهد بود.

بنابراین با اجرای مرحله نخست الگوریتم پادتن‌هایی به دست خواهند آمد که فضای مورد بررسی را به‌خوبی نمایش خواهند داد. مزیت استفاده از این مرحله آن است که به‌منظور شناخت پیکربندی فضا جستجو، نیازی به بررسی تمام داده‌ها نیست؛ بلکه تنها با بررسی پادتن‌ها به‌راحتی و با سرعت و دقت بیشتر می‌توان فضای مورد نظر و وضعیت داده‌های پرت را بررسی، و مورد تجزیه و تحلیل قرار داد [38]. از طرفی استفاده مستقیم از پادتن‌ها به‌منظور خوشه‌بندی داده‌ها، بسیار هزینه‌بر و زمان‌بر خواهد بود. بنابراین در مرحله دوم الگوریتم پیشنهادی، با استفاده از روش خوشه‌بندی سلسله‌مراتبی، به‌سرعت خواهیم توانست به خوشه‌ها دست یابیم. از مهم‌ترین مشکلات روش خوشه‌بندی سلسله‌مراتبی حساسیت به داده‌های پرت و عدم تشخیص صحیح ساختار سلسله‌مراتبی در بانک داده‌های بزرگ است [26]. بنابراین اجرای مرحله نخست الگوریتم این روش را در شناسایی ساختار بانک داده و داده‌های پرت یاری می‌رساند.

[1] Linkage (Min) method

۵- اجرای الگوریتم پیشنهادی

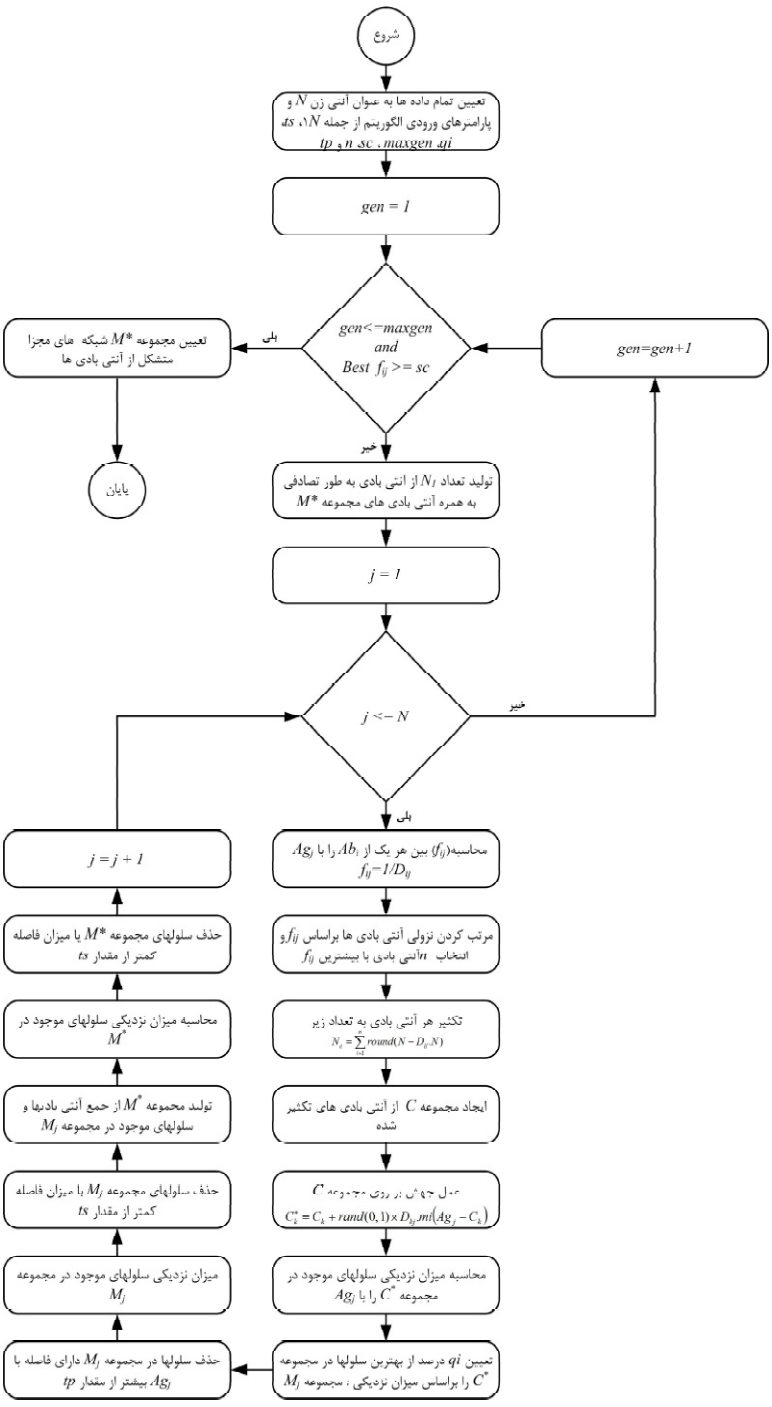
در این بخش طی دو مرحله عملکرد الگوریتم پیشنهادی مورد بررسی قرار می‌گیرد. در ادامه ابتدا با ارائه تحلیل حساسیت پارامترهای مرحله نخست الگوریتم، مقادیر پیشنهادی پارامترها تعیین شده است (نمودارهای ۳ تا ۵). از آنجا که مرحله نخست الگوریتم پیشنهادی شبیه الگوریتم aiNET [37] است، تحلیل حساسیت پارامترهای پیشنهادی با استفاده از داده‌های مرجع [37] صورت گرفته است؛ سپس با استفاده از مقادیر تعیین‌شده برای پارامترها، نتایج اجرای الگوریتم پیشنهادی بر روی داده‌های واقعی جداول (۱) و (۲) ارائه شده است. نمونهٔ واقعی شامل داده‌های زلزله در ایران است که این داده‌ها از تارنماهای ngdir.ir، bhrc.ac.ir و geophysics.ut.ac.ir جمع‌آوری شده است و دارای ۵۱۴ داده بوده که داده‌ها دارای هفت ویژگی شامل سال وقوع زلزله، شدت زلزله براساس مقیاس‌های امواج درونی (Mb)، امواج سطحی (Ms)، گشتاوری (Mw) و مقیاس محلی (Ml) و همچنین طول و عرض موقعیت جغرافیایی زمین لرزه است. الگوریتم پیشنهادی در نرم‌افزار MATLAB پیاده‌سازی و تحت سامانه با مشخصات Intel® Core™2 Duo 2.93Ghz و 2 GB RAM اجرا شد. در ادامه نتایج به‌دست‌آمده با نتایج حاصل از اجرای الگوریتم کلونی مورچگان بر روی داده‌های واقعی بالا مقایسه شد.

۵-۱- تحلیل حساسیت پارامترهای الگوریتم

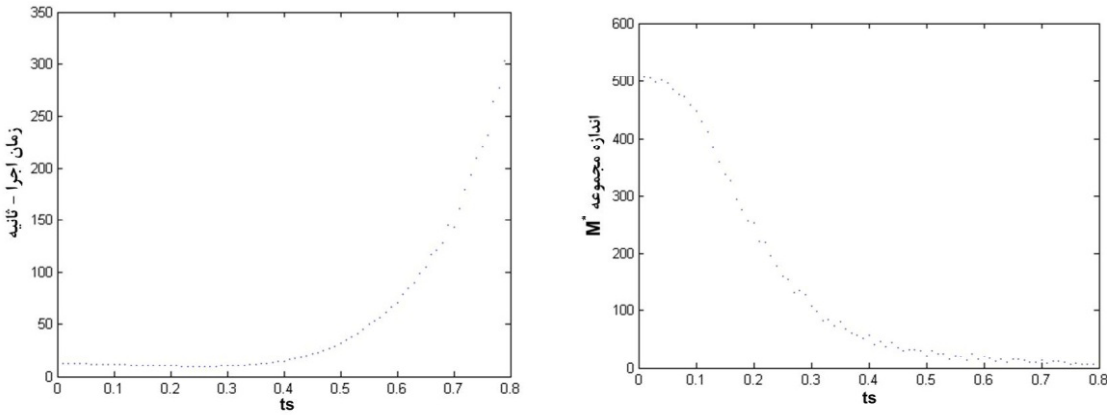
نمودارهای (۳) تا (۵) میزان حساسیت الگوریتم پیشنهادی را نسبت به پارامترهای ts، qi و Maxgen نشان می‌دهند. در هر کدام از نمودارهای بالا یک پارامتر متغیر و سایر پارامترها ثابت فرض شده است. نمودار (۳) نشان می‌دهد که اندازه M^* و زمان اجرای الگوریتم نسبت به تغییرات ts (آستانه مرگ) حساس است. با این وجود تعیین بازه [۰/۴، ۰/۲] برای پارامتر ts منطقی است؛ زیرا در این بازه زمان اجرای الگوریتم در کم‌ترین مقدار قرار دارد و همچنین اندازه M^* تعداد معقولی را نشان می‌دهد.

[DOI: 10.18869/acadpub.jsdp.13.4.133]

نمودارهای (۴) و (۵) بیان‌گر این نتیجه هستند که با افزایش مقدار qi و Maxgen زمان اجرای الگوریتم نیز افزایش می‌یابد. افزایش زمان اجرا با افزایش Maxgen رابطه خطی و با افزایش qi رابطه نمایی دارد. بنابراین افزایش مقدار qi تأثیر بیشتری در زمان اجرای الگوریتم خواهد داشت. نمودار (۵) حساسیت بسیار کم خروجی الگوریتم پیشنهادی نسبت به تغییرات Maxgen را نشان داده است و از طرفی انتخاب بازه [۰/۵، ۰/۱۰] برای qi منطقی به‌نظر می‌رسد؛ زیرا در این بازه طبق نمودار (۴) زمان کمی به منظور اجرای الگوریتم نیاز است و همچنین در این بازه M^* کمترین مقدار خود را خواهد داشت. از طرفی نتایج اجرای این مرحله نشان می‌دهد که الگوریتم پیشنهادی سرعت اجرای بسیار بالاتری نسبت به الگوریتم ارائه شده در [37] دارد.

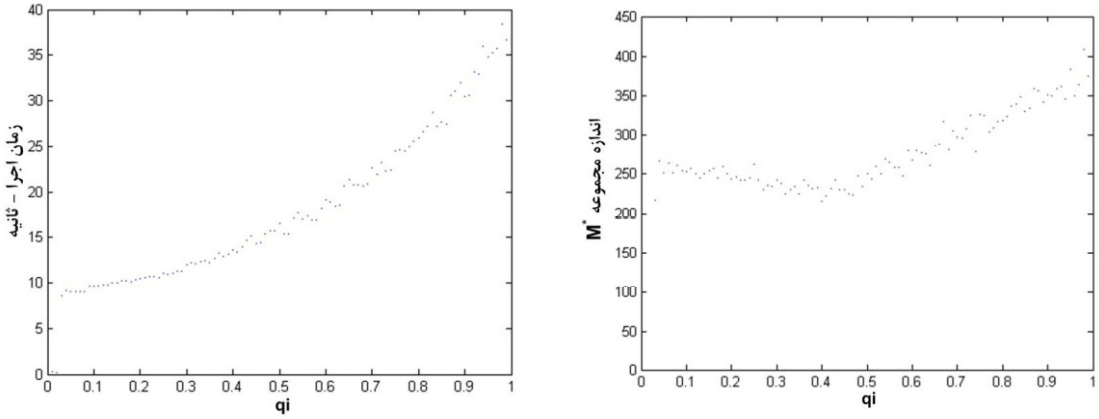


(نمودار-۲): مرحله نخست الگوریتم
(Figure-2): The first stage of the algorithm



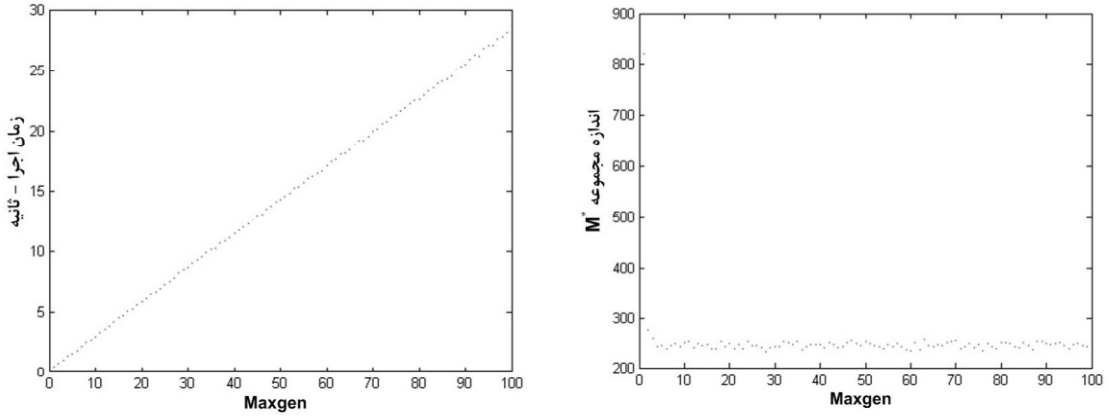
(نمودار-۳): تغییرات اندازه M^* و زمان اجرا با تغییرات ts (qi=0.15 و Maxgen=35)

(Figure-3): M* and runtime changes vs. ts changes (Maxgen=35 and qi=0.15)



(نمودار-۴): تغییرات اندازه M^* و زمان اجرا با تغییرات qi (ts=0.2 و Maxgen=35)

(Figure-4): M* and runtime changes vs. qi changes (Maxgen=35 and ts=0.2)



(نمودار-۵): تغییرات اندازه M^* و زمان اجرا با تغییرات Maxgen (qi=0.15 و ts=0.2)

(Figure-5): M* and runtime changes vs. Maxgen changes (ts=0.2 and qi=0.15)

۵-۲-نتایج اجرای الگوریتم پیشنهادی و مقایسه آن با الگوریتم کلونی مورچگان

جدول (۱) اجرای الگوریتم پیشنهادی برای بانک داده واقعی مورد نظر (داده‌های زلزله در ایران) را براساس مقادیر مختلف برای پارامترهای الگوریتم ارائه کرده است. همان‌طورکه از نتایج جدول (۱) مشخص است، اجرای الگوریتم در مرحله نخست بسیار سریع است. همچنین نتایج اجرای مرحله نخست نشان داد که با افزایش در مقادیر پارامترهای qi، ts و

[DOI: 10.18869/acadpub.jsdp.13.4.133]

Maxgen زمان اجرا الگوریتم اندکی افزایش می‌یابد و از طرفی با افزایش در مقادیر پارامترهای qi و ts اندازه ماتریس M^* بسیار کاهش می‌یابد. به‌عبارتی با افزایش در پارامترهای qi و ts به پادتن کمتری برای شناسایی فضای جستجو نیاز است. در اجرای سوم تعداد پادتن در مقابل اندازه بانک داده کوچک است که این امر نشان از آن دارد که با انتخاب مقادیر پارامترها در اجرای سوم می‌توان به‌دقت و سرعت اجرای بالاتر در الگوریتم پیشنهادی دست یافت.

(جدول-۱): پارامترهای الگوریتم و نتایج حاصل از اجرای مرحله اول الگوریتم پیشنهادی

(Table-1): Algorithm parameters and results of the first phase of the proposed algorithm

	ts	tp	n	N_1	qi (درصد)	Maxgen	زمان s	اندازه M^*
اجرای اول	0.07	1	4	10	10	30	9/01	479
اجرای دوم	0.15				15	35	10/15	332
اجرای سوم	0.20				20	40	11/49	237
اجرای چهارم	0/07	1	8	20	10	30	11/5	485
اجرای پنجم	0/15				15	35	13/15	355
اجرای ششم	0/20				20	40	14/93	249
اجرای هفتم	0/07	1	12	30	10	30	13/85	491
اجرای هشتم	0/15				15	35	16	355
اجرای نهم	0/20				20	40	18/75	263

(جدول-۲): نتایج حاصل از اجرای مرحله دوم الگوریتم پیشنهادی

(Table-2): The results of the second phase of the proposed algorithm

cut		0/1	0/2	0/3	0/4
اجرای اول	تعداد خوشه	363	363	363	363
	زمان اجرا (S)	0/067	0/053	0/046	0/053
اجرای دوم	تعداد خوشه	246	246	246	246
	زمان اجرا (S)	0/047	0/038	0/029	0/029
اجرای سوم	تعداد خوشه	187	187	187	187
	زمان اجرا (S)	0/028	0/021	0/02	0/02
اجرای چهارم	تعداد خوشه	485	485	485	485
	زمان اجرا (S)	0/04	0/03	0/03	0/03
اجرای پنجم	تعداد خوشه	355	355	355	355
	زمان اجرا (S)	0/02	0/02	0/02	0/02
اجرای ششم	تعداد خوشه	249	249	249	249
	زمان اجرا (S)	0/02	0/02	0/02	0/01
اجرای هفتم	تعداد خوشه	491	491	491	491
	زمان اجرا (S)	0/04	0/04	0/03	0/03
اجرای هشتم	تعداد خوشه	355	355	355	355
	زمان اجرا (S)	0/02	0/02	0/02	0/02
اجرای نهم	تعداد خوشه	263	263	263	263
	زمان اجرا (S)	0/02	0/02	0/02	0/02

[DOI: 10.18869/acadpub.jsdp.13.4.133]

مرحلهٔ دوم الگوریتم نیز در نرم‌افزار MATLAB پیاده‌سازی و تحت سامانه قبلی اجرا شد. نمودار جدول (۲) نتایج حاصل از اجرای مرحله دوم الگوریتم پیشنهادی با مقادیر مختلف سطح برش (cut) را نشان داده است.

طبق نتایج جدول (۲) الگوریتم در اجراهای مختلف داده‌ها را در تعداد گروه‌های مختلف، خوشه‌بندی کرده است. از طرفی زمان اجرا نشان از سرعت بسیار بالای اجرای الگوریتم دارد. همچنین براساس سطح برش‌های مختلف در هر یک از اجراهای الگوریتم پیشنهادی تعداد خوشه یکسان بوده و همچنین تفاوتی در داده‌های خوشه‌ها وجود ندارد. این موضوع دقت اجرای الگوریتم و عدم وابستگی آن به پارامتر cut و همچنین شناخت صحیح فضای داده‌ها ازجمله شناسایی داده‌های پرت را نشان می‌دهد. از طرفی با توجه به نتایج اجراهای چهارم به بعد که باعث افزایش تعداد خوشه‌ها شده است و از طرفی با توجه به نتایج اجرای مرحله نخست الگوریتم که بهترین خروجی در اجرای سوم به‌دست آمده است، تعیین مقدار $N_1 = ۱۰$ توصیه می‌شود تا بتوان به تعداد خوشه بهتری دست یافت.

در ادامه این بخش نتایج الگوریتم پیشنهادی در مقاله حاضر با نتایج الگوریتم ارائه‌شده توسط مینایی و همکاران [1] مقایسه می‌شود. مینایی و همکاران [1] الگوریتم خوشه‌بندی بهبود یافته مبتنی بر کلونی مورچگان (IASC[1]) توسعه داده و نتایج اجرای آن را بر روی داده‌های مشابه در این مقاله ارائه کرده‌اند. جدول (۳) نتایج اجرای الگوریتم IASC را نشان داده است.

براساس نتایج ارائه‌شده در جدول (۳)، الگوریتم IASC داده‌ها را در تعداد خوشه بسیار کمتری نسبت به اجرای سوم الگوریتم پیشنهادی در این مقاله (جدول ۲) گروه‌بندی کرده است؛ اما زمان اجرای الگوریتم IASC بسیار بیشتر از الگوریتم پیشنهادی مبتنی بر AIS است.

به‌منظور تعیین بهترین الگوریتم از بین الگوریتم‌های پیشنهادی مبتنی بر AIS و الگوریتم IASC، از دو پارامتر زیر استفاده شده است:

۱: زمان اجرای الگوریتم: الگوریتمی بهتر است که زمان اجرای کمتری داشته باشد.

۲: SSE:[2] هر چه خوشه فشرده‌تر باشد آن خوشه و یا الگوریتم بهتر خواهد بود؛ زیرا SSE پارامتری است که بیان‌کننده میزان فشردگی خوشه است. SSE بیان‌گر میانگین فاصلهٔ مرکز یک خوشه با اعضای خوشه مورد نظر است.

(جدول-۳): نتایج اجرای الگوریتم IASC
(Table-3): Results of IASC algorithm

زمان اجرا (s)	تعداد خوشه	تعداد مورچه
18	6	20
21/6	11	30
34/15	13	40
29/92	17	50
26/71	19	60
38/11	21	70

جدول (۴) مقایسه دو الگوریتم را براساس پارامترهای مختلف نشان داده است.

(جدول-۴): مقایسه دو الگوریتم
(Table-4): Compare the two algorithms

مبتنی بر AIS	مبتنی بر IASC	الگوریتم
12	18	زمان اجرا (s)
5/3	9/4	SSE

با توجه به نتایج جدول (۴)، الگوریتم پیشنهادی دارای زمان اجرا و میزان خطای SSE کمتری نسبت به الگوریتم IASC است. بنابراین پارامترهای بالا برتری الگوریتم پیشنهادی در این مقاله را نشان می‌دهند؛ اما باید توجه داشت که تعداد خوشه‌های ارائه‌شده توسط الگوریتم IASC بسیار کمتر از تعداد خوشه‌های الگوریتم پیشنهادی است. هدف اصلی این مقاله بررسی الگوریتم ترکیبی AIS و سلسله‌مراتبی بوده است؛ به‌نحوی که در این مقاله الگوریتم پیشنهادی با استفاده از داده‌های واقعی مورد ارزیابی قرار گرفت. ارزیابی الگوریتم با استفاده از اجراهای مختلف الگوریتم و شناسایی بهترین مقدار پارامترهای انجام شود.

۶- نتیجه‌گیری

در این مقاله دو روش مطرح به‌منظور خوشه‌بندی داده‌ها مورد استفاده قرار گرفت. روش خوشه‌بندی AIS مبتنی بر شبکه ایمنی اصلاح‌شده که یکی از مهم‌ترین الگوریتم‌های فرا ابتکاری به‌منظور حل مسائل بسیار پیچیده است، به‌خوبی قادر است پیکربندی فضای جستجو را نمایش دهد و داده‌های پرت، دسته‌بندی داده‌ها و ساختار داده‌ها را شناسایی کند [38]. اما این روش در تعیین خوشه داده‌ها و تعداد آنها عاجز است. از طرفی روش خوشه‌بندی سلسله‌مراتبی، خوشه‌ها و تعداد آنها را با اجرای سریع خود، به دست می‌دهد؛ ولی از مهم‌ترین مشکلات روش خوشه‌بندی سلسله‌مراتبی حساسیت به داده‌های پرت و عدم

[1] Improved Ant System-based Clustering algorithm
[2] Sum of Squared Error

[DOI: 10.18869/acadpub.jsdp.13.4.133]

تشخیص صحیح ساختار سلسله‌مراتبی در بانک داده‌های بزرگ است [26].

بنابراین الگوریتم دومرحله‌ای پیشنهادی با اجرای مرحلهٔ نخست خود ساختار داده‌ها را تعیین و مشکلات روش خوشه‌بندی را برطرف می‌کند و در مرحلهٔ دوم با اجرای روش تحلیل سلسله‌مراتبی، نقایص موجود در مرحلهٔ نخست را رفع خواهد کرد. از دیگر مزایای الگوریتم پیشنهادی عدم نیاز به بررسی کل داده‌ها به‌منظور خوشه‌بندی فضای مورد بررسی است. با تعیین بهترین پادتن‌ها، خواهیم توانست درک درستی از فضای جستجو و نحوه پیکربندی آن داشته باشیم. در این مقاله به‌منظور بررسی عملکرد الگوریتم پیشنهادی تحلیل حساسیت پارامترهای الگوریتم و همچنین نمونهٔ واقعی متشکل از داده‌های زلزله در ایران مورد توجه قرار گرفت. نتایج اجرای الگوریتم نشان داد که الگوریتم پیشنهادی به پارامترهای ورودی خود حساسیت بسیار کمی دارد و توانسته است نقایص موجود در AIS و روش خوشه‌بندی سلسله‌مراتبی را پوشش دهد و از طرفی از دقت و سرعت بسیار بالایی در اجرا برخوردار است.

۷- مراجع 7-refrences

[1] مینایی، بهروز، فتحیان، محمد، جعفریان مقدم، احمد رضا، و نصیری، مهدی، "استفاده از تکنیک خوشه‌بندی سیستم کلونی مورچگان بهبود یافته به منظور خوشه‌بندی داده‌های زلزله ایران"، *نشریه تخصصی مهندسی صنایع*، صفحات ۲۲۱-۲۲۷، ۱۳۹۱.

[1] Minaei, B., Fathian, M., Jafarian-Moghaddam, A. R., & Nasiri, M.. Clustering Iran Earthquake Data using Improved Ant System-Based Clustering Algorithm. Journal of Industrial Engineering, 221-227, 2011.

[2] D. Dasgupta, *Artificial immune systems and their applications*. Springer, 1999.

[3] J. Timmis, M. Neal, and J. Hunt, "An artificial immune system for data analysis," Biosystems, vol. 55, pp. 143–150, 2000

[4] L. N. De Castro and J. I. Timmis, *Artificial immune systems: a new computational intelligence approach*. Springer, 2002

[5] U. Aicklein, P. Bentley, S. Cayser, K. Jungwon, and J. McLeod, "Danger theory: The link between AIS and IDS," In *ICARIS International Conference on Artificial Immune Systems*, LNCS Springer, 2003, pp. 147–155.

[6] X. Cao, H. Qiao, and T. Xu, "Negative selection based immune optimization," *Advances in Engineering Software*, vol. 38, pp. 649–656, 2003.

[7] H. Maulik and S. Bandyopadhyay, "Genetic algorithm-based clustering technique," *Pattern Recognition*, vol. 33, pp. 1455–1465, 2000.

[8] A. J. Graaff and A. P. Engelbrecht, "Clustering data in an uncertain environment using an artificial immune system," *Pattern Recognition Letters*, pp. 342-351, 2011.

[9] J. C. L. Pinto and F. J. Von Zuben, "Fault detection algorithm for telephone systems based on the danger theory," In *ICARIS International Conference on Artificial Immune Systems*, LNCS Springer, 2003, pp. 418–431.

[10] P. Matzinger, "Tolerance, danger and the extended family," *Annual Reviews of Immunology*, vol. 12, pp. 991–1045, 1994.

[11] S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri, "Self-nonself discrimination in a computer," In *IEEE Symposium on Research in Security and Privacy*, Los Alamos, CA. IEEE Pres, 1994.

[12] H. Kamber, *Data Mining: concepts and technique*. Second Edition, Elsevier, 2006.

[13] M. S. Aldenderfer and R. K. Blashfield, *Cluster analysis*. Newbury Park: Sage Publications, 1986.

[14] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. Wiley, New York, 1990.

[15] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis*. Wiley, New York, 1973.

[16] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 4–37, 2000.

[17] J. A. Hartigan, *Clustering Algorithms*. Wiley, New York, 1975.

[18] G. J. McLachlan and K. E. Basford, *Mixture models: inference and applications to clustering*. Marcel Dekker, New York, 1988.

[19] G. J. McLachlan and T. Krishnan, *The EM algorithm and extensions*. Wiley, New York, 1997.

[20] S. Grossberg, "Adaptive pattern classification and universal recoding I: Parallel development and coding of neural feature detectors," *Biological Cybernetics*, vol. 23, pp. 121–134, 1976.

[21] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE Transactions on*

[DOI: 10.18869/acadpub.jsdp.13.4.133]

Acoustics, Speech, Signal Processing, pp. 4–22, 1987.

[22] C. F. Tsai, H. C. Wu, and C. W. Tsai, "A new clustering approach for data mining in large databases," In *Proceedings of the international symposium on parallel architectures, algorithms and networks (ISPAN'02)*, IEEE Computer Society, 2002, pp. 1087–4089.

[23] T. Kohonen, *Self-Organizing maps*. third ed. Springer-Verlag, Berlin, 2001.

[24] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithm*. Plenum Press, New York, 1981.

[25] M. S. Yang, "A survey of fuzzy clustering," *Mathematical and Computer Modelling*, vol. 18, pp. 1–16, 1993.

[26] P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*. Addison-Wesley, 2005.

[27] U. Fayyad, "Data mining and knowledge discovery in databases: implications for scientific databases," In *Scientific and statistical database management, 1997 proceedings, ninth international conference*, 1997, pp. 2–11.

[28] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in database," *American Association for Artificial Intelligence*, pp. 37–54, 1996.

[29] G. A. Carpenter and S. Grossberg, "ART2: self-organization of stable category recognition codes for analog input pattern," *Applied Optics*, vol. 26, pp. 4919–4930, 1987.

[30] K. Krishna and M. Murty, "Genetic K-means algorithm," *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 29, pp. 433–439, 1993.

[31] X. B. Yang, J. G. Sun, and D. Huang, "A new clustering method based on ant colony algorithm," In *Proceedings of the 4th world congress on intelligent control and automation*, 2002, pp. 2222–2226.

[32] R. J. Kuo, H. S. Wang, T. L. Hu, and S. H. Chou, "Application of Ant K-Means on clustering analysis," *Computers and Mathematics with Applications*, vol. 50, pp. 1709-1724, 2005.

[33] R. J. Kuo, S. Y. Lin, and C. W. Shih, "Mining association rules through integration of clustering analysis and ant colony system for health insurance database in Taiwan," *Expert Systems with Applications*, vol. 33, pp. 794–808, 2007.

[34] R. J. Kuo and C. W. Shih, Association rule mining through the ant colony system for National Health Insurance Research Database in Taiwan," *Computers and Mathematics with Applications*, vol. 54, pp. 1303–1318, 2007.

[35] X. Z. Gao, S. J. Ovaska, and X. Wang, "A GA-based negative selection algorithm," *International Journal on Innovative Computing Information and Control*, vol. 4, pp. 971–979, 2008.

[36] N. K. Jerne, "Toward a network theory of the immune system," In *Annales d'Immunologic (Institute Pasteur)*, Paris, France, 1976, 373–389.

[37] L. N. De Castro and F. J. Von Zuben, "aiNET: an artificial immune network for data analysis," In *Data mining: A heuristic approach*, H. Abbas, R. Sarker, C. Newton, Eds. Idea Group Publishing, 2001, pp. 231– 259.

[38] E. L. Talbi, *Metaheuristics: from design to implementation*. John Wiley and Sons, Inc., Hoboken, New Jersey, 2009.

[39] F. M. Burnet, *The clonal selection theory of acquired immunity*. University Press, Cambridge, 1959.

[40] L. N. De Castro and F. J. Von Zuben, "The clonal selection algorithm with engineering applications," In *Workshop on Artificial Immune Systems and Their Applications, (GECCO'00)*, Las Vegas, NV, 2000. Pp. 36–37.

احمد رضا جعفریان مقدم عضو هیئت علمی دانشکده مهندسی عمران و حمل و نقل در دانشگاه اصفهان است. مدارک کارشناسی و کارشناسی ارشد خود را در رشته مهندسی حمل و نقل ریلی به‌ترتیب در سال‌های ۱۳۸۵ و ۱۳۸۶ از دانشگاه علم و صنعت ایران دریافت کرده است. وی در سال ۱۳۹۴ در دوره دکترای رشته مهندسی صنایع (مدیریت سیستم و بهره‌وری) در دانشگاه علم و صنعت ایران فارغ التحصیل شده است. ایشان تاکنون سه کتاب و بیش از پانزده مقاله را به چاپ رسانده است. زمینه‌های مورد علاقه وی داده کاوی، مدل‌سازی شبکه‌های حمل و نقل، برنامه‌ریزی حمل و نقل، مدیریت و کنترل پروژه، ارزیابی اقتصادی پروژه‌ها و مدیریت استراتژیک است.

نشانی رایانامه ایشان عبارت است از:

ar.jafarian@trn.ui.ac.ir

[DOI: 10.18869/acadpub.jsdp.13.4.133]

فرناز برزین پور عضو هیئت علمی دانشکده مهندسی صنایع در دانشگاه علم و صنعت ایران است. مدرک کارشناسی خود را در سال ۱۳۷۵ در رشته مهندسی صنایع از دانشگاه علم و صنعت ایران و مدارک کارشناسی ارشد و دکترا را نیز در رشته مهندسی صنایع از دانشگاه تربیت مدرس به‌ترتیب در سال‌های ۱۳۷۷ و ۱۳۸۳ دریافت کرده است. ایشان تاکنون مقالات و کتب مختلفی را به چاپ رسانده است. مهم‌ترین زمینه‌های مورد علاقه وی بهینه‌سازی و الگوریتم‌های فراابتکاری، زمان‌بندی و مدیریت پروژه و مدیریت دانش و فناوری اطلاعات است.

نشانی رایانامه ایشان عبارت است از:
barzinpour@iust.ac.ir

محمد فتحیان عضو هیئت علمی دانشکده مهندسی صنایع در دانشگاه علم و صنعت ایران و دارای درجه استادی است. مدرک کارشناسی را در رشته مهندسی الکترونیک از دانشگاه خواجه نصیر الدین طوسی، مدرک کارشناسی ارشد و دکترای خود را در رشته مهندسی صنایع از دانشگاه علم و صنعت ایران به‌ترتیب در سال‌های ۱۳۷۰، ۱۳۷۶ و ۱۳۸۱ دریافت کرده است. تاکنون پنج کتاب و بیش از سی مقاله را به چاپ رسانده است. زمینه‌های مورد علاقه وی مدیریت و مهندسی فناوری اطلاعات و مدیریت و مهندسی صنایع است.

نشانی رایانامه ایشان عبارت است از:
fathian@iust.ac.ir

[DOI: 10.18869/acadpub.jsdp.13.4.133]