



کاربرد الگوریتم ایمنی مصنوعی در خوشه یابی داده ها

سیدحمید ظهیری

گروه مهندسی الکترونیک و مخابرات، دانشکده مهندسی، دانشگاه بیرجند

shzahiri@yahoo.com

1- مقدمه

الگوریتم ایمنی مصنوعی¹ به عنوان یک روش جستجو و بهینه سازی کارآمد در بین روشهای دیگری مانند الگوریتم ژنتیک² و هوش جمعی³ از جایگاه ویژه ای برخوردار است. الگوریتم ایمنی مصنوعی روشی بر گرفته از نحوه عملکرد سیستم ایمنی بدن در مواجهه با عوامل بیماریزای خارجی است. ظرافت و دقت این سیستم دانشمندان را بر آن داشت که با شبیه سازی عملکرد سیستم ایمنی موجودات روش بهینه سازی ایمنی مصنوعی را طراحی کنند به گونه ای که قادر به حل مسایل دشوار مهندسی باشد. حجم و تنوع تحقیقاتی که در زمینه حل مسایل مختلف مهندسی با استفاده از این روش گزارش شده اند، نشانی بر توانایی الگوریتم ایمنی مصنوعی در حل مسایل پیچیده مهندسی است ([1] تا [3]).

در تحقیق حاضر استفاده از الگوریتم ایمنی مصنوعی در مسئله خوشه یابی⁴ - یکی از مسایل مهم در علم بازشناسی الگو⁵ - مورد نظر قرار گرفته است. مسئله خوشه یابی عبارت است از

چکیده: در این تحقیق از الگوریتم ایمنی مصنوعی برای پایه ریزی یک روش جدید خوشه یابی بهره گیری شده است. مولفه های سلولهای دفاعی به شکل مختصات مراکز خوشه ها تعریف شده در یک روند تکاملی «قربت» خود را با آنتی ژنها به وسیله عملگر جهش بهبود می بخشند. کمیت قربت در سیستم ایمنی شبیه سازی شده به شکل عکس خطای کوانتیزه کردن تعریف شده است. روش ارائه شده بر روی انواع مختلفی از داده های مصنوعی و داده های مشهور در پردازش الگو (با تنوع در ابعاد فضای ویژگی و تعداد نمونه ها) آزمایش شده است. نتایج به دست آمده ضمن موفقیت روش پیشنهادی در خوشه یابی مجموعه داده های مختلف، برتری نسبتاً قابل توجهی را در عملکرد این روش نسبت به روش k-means و روش خوشه یابی ژنتیک نشان می دهند.

واژه های کلیدی: خوشه یابی داده، الگوریتم ایمنی مصنوعی، بازشناسایی الگو.

Artificial immune algorithm¹

Genetic algorithm²

Swarm intelligence³

Clustering⁴

Pattern recognition⁵

پردازش داده ها در فضای ویژگی به گونه ای که نمونه های شبیه به هم در خوشه هایی جداگانه گرد هم آیند. این مسئله در موضوعات مهمی همچون داده کاوی⁶ [4] و تقسیم بندی داده های تصویر [5] کاربرد دارد و روشهای مختلفی نیز به این منظور پیشنهاد شده است که از آن میان می توان به روشهای *k-means* و *ISODATA* [6] و اخیراً روش خوشه یابی ژنتیک [7] و [8] و روش خوشه یابی گروه ذرات [9] اشاره کرد.

در اغلب روشهای خوشه یابی ژنتیک کروموزومها به صورت مراکز خوشه ها تعریف و سپس بیان باینری آنها مورد استفاده قرار گرفته تحت عملگرهای جهش و برش به سمت مقدار بهینه میل می کنند. در این روش تابع برازندگی با مقدار خطای کوانتیزه مرتبط است [7].

در روشهای هوش جمعی مانند روش بهینه سازی گروه ذرات مکان هر ذره به شکل مختصات مراکز خوشه تعریف شده با معادلات تغییر مکان بهترین مختصات مراکز خوشه محاسبه می شود [9].

در روش خوشه یابی پیشنهاد شده در این تحقیق سلولهای دفاعی به شکل بردارهایی که شامل مختصات مراکز خوشه ها هستند، تعریف می شود. «قربانیت»⁷ سلولهای دفاعی با آنتی ژنها به صورت عکس خطای کوانتیزه کردن تعریف می گردد و آنتی بادیهای با قربانیت کمتر به وسیله عملگر جهش⁸ ساختار خود را برای دستیابی به قربانیت بیشتر با آنتی ژنها بهبود می دهند.

داده های گوناگون مصنوعی و مشهور برای محک روش ارائه شده و مقایسه نتایج به دست آمده با نتایج حاصل از سایر روشهای مرسوم خوشه یابی انتخاب شده اند. مجموعه های داده انتخاب شده دارای تنوع گوناگون در تعداد نمونه ها، ابعاد بردارهای ویژگی و تعداد کلاسه های مرجع می باشند.

پیکر بندی این مقاله به این ترتیب است که در بخش 2 به مفاهیم اساسی الگوریتم ایمنی مصنوعی اشاره مختصری

خواهد شد. مطالب این بخش ما را آماده می کند که روش پیشنهادی خود را با نام «خوشه یابی بر مبنای الگوریتم ایمنی مصنوعی» در بخش 3 ارائه کنیم. پس از توضیحات تفصیلی درباره روش خوشه یابی ارائه شده در بخش 4 به ارزیابی عملکرد آن در مواجهه با انواع مجموعه های داده و مقایسه آن با عملکرد سایر روشها پرداخته خواهد شد. بحث و نتیجه گیری نهایی در بخش 5، مقاله را خاتمه خواهد داد.

2- الگوریتم ایمنی مصنوعی

هنگامی که عوامل بیماریزای خارجی-مانند ویروسها و باکتریها و در یک قالب کلی آنتی ژنها- ارگانیسم موجودات زنده را مورد تهاجم قرار می دهند، ضمن تخریب سلولها به تکثیر نیز می پردازند. یکی از مکانیسمهای جالب توجه سیستم تدافعی موجودات زنده در مقابله با این تهاجم تکثیر سریع سلولهای تدافعی است که توفیق لازم در شناسایی آنتی ژنها و نابود کردن آنها را دارا هستند. جالب اینجاست که میزان تکثیر سلولهای تدافعی و آنتی بادی به میزان موفقیت آنها در نابود کردن فاکتورهای تهاجمی وابسته است. یعنی سیستم ایمنی سلولهای تدافعی با عملکرد بهتر را بیشتر و آنهایی که دارای قابلیت کمتری هستند را کمتر تکثیر می کند. میزان تشخیص یک آنتی ژن به وسیله سلولهای تدافعی به وسیله فاکتوری به نام «قربانیت» شناخته می شود. سلولهای تدافعی با قربانیت کمتر باید متحمل عملگر زیستی به نام «جهش» شوند تا با تغییرات ساختاری بتوانند قربانیت خود را با عوامل بیماریزا بیشتر کرده عملکرد دفاعی خویش را بهبود بخشند. میزان جهش برای عوامل دفاعی با قربانیت بیشتر، کمتر است و بالعکس.

با توجه به توضیحات فوق چهار مفهوم مهم «قربانیت»، «انتخاب»، «جهش»، و «تکثیر» دخالت عمده ای در عملکرد سیستم ایمنی موجودات زنده دارند. لذا در الگوریتم ایمنی مصنوعی نیز همین مفاهیم باید نقش عمده را ایفا کنند. در واقع برای پیاده سازی یک الگوریتم ایمنی مصنوعی مراحل زیر اجتناب ناپذیر است:

1) مقدمه سازی:

در این مرحله جمعیتی تصادفی از سلولهای ایمنی تشکیل

⁶ Data mining

⁷ Affinity

⁸ Mutation

می شوند.

2) حلقه جستجو

در این حلقه مراحل زیر در خصوص هر آنتی ژن انجام

می گیرد:

1-2) ارزیابی قرابت

سلولهای تدافعی جدید مورد ارزیابی قرار می گیرند.

2-2) انتخاب

سلولهای ایمنی با قرابت بیشتر نسبت به آنتی ژنها انتخاب

می شوند.

z_{11}	z_{12}	\vdots	z_{1n}	z_{21}	z_{22}	\vdots	z_{2n}	\dots	\dots	z_{K1}	z_{K2}	\dots	z_{Kn}
----------	----------	----------	----------	----------	----------	----------	----------	---------	---------	----------	----------	---------	----------

شکل 1- ساختار سلولهای دفاعی

2-3) تکثیر و تغییر حالت ساختاری

سلولهای ایمنی تکثیر می شوند. تشخیص آنتی ژن و

قرابت بیشتر با آن یعنی تکثیر بیشتر سلول تدافعی

متناظر با آن. سلولهای با قرابت کمتر تحت عملگر

جهش تغییر ساختاری پیدا می کنند. میزان اعمال

عملگر جهش با میزان قرابت نسبت عکس دارد.

3) بستن حلقه برای تعداد تکرار پیش فرض

مرحله 2 برای تعداد پیش تعریف شده ای از تکرار حلقه و

یا نیل به همگرایی (از میان رفتن همه آنتی ژنها) ادامه می

یابد.

بوده و به خاطر عملکرد خوب با اقبال بیشتری نسبت به سایر

روشها مواجه شده است. امروزه با افزایش حجم اطلاعات،

شبکه های اطلاع رسانی و بانکهای

داده نیاز به روشهای موثرتری جهت خوشه یابی در فضاهای با

ابعاد زیاد و در حضور حجم بالا نقاط داده است. از این رو

محققان اقدام به معرفی روشهای جدیدی در این خصوص

کرده اند. به طور مثال در [7] و [8] با استفاده از روش ژنتیک

روشهایی تازه به این منظور پیشنهاد شده است که در مقدمه به

اجمال معرفی شد.

در این بخش با استفاده از الگوریتم ایمنی مصنوعی روشی

جدید برای خوشه یابی پیشنهاد می شود.

با توجه به توضیحات ارائه شده در بخش 2 ابتدا لازم است

پارامترهای حاضر در مسئله خوشه یابی با عوامل موجود در

الگوریتم ایمنی مصنوعی معرفی شوند.

الف) سلولهای دفاعی

سلولهای دفاعی در بر دارنده مرکز خوشه ها می باشند. مثلاً

برای تعداد K خوشه در فضای n بعدی ساختار یک سلول

دفاعی به شکل زیر است:

در شکل 1، $Z_j = (z_{j1}, z_{j2}, \dots, z_{jn})$ ، $j=1, 2, \dots, K$

مرکز خوشه j ام هستند. لازم به توضیح است که تعداد خوشه

ها در این روش به صورت پیش فرض معلوم است.

3- خوشه یابی بر مبنای الگوریتم ایمنی مصنوعی

روشهای مختلفی برای حل مسئله خوشه یابی پیشنهاد شده

است. که در مقدمه به آنها اشاره شد از میان آنها روش خوشه

بندی k -means از شهرت بیشتری برخوردار بوده استفاده

گسترده تری نسبت به سایر موارد دارد [6]. در این روش ابتدا

به تعداد پیش فرض مراکز خوشه ها به صورت کاملاً تصادفی

از میان نقاط داده انتخاب می شوند و سپس عمل تشکیل

خوشه ها با تعلق دادن نقاط داده به نزدیکترین مرکز صورت

می گیرد. در مرحله بعدی مراکز جدید خوشه ها به مقدار

میانگین همه نقاط موجود در یک خوشه تغییر مقدار می دهند

و مجدداً عمل تعلق داده ها در فضای ویژگی به خوشه ها

تکرار می شود. این فرایند آنقدر تکرار می شود که مراکز

خوشه ها ثابت باقی بمانند. شاید بتوان گفت روش k -means

به دلیل ساختار ساده آن یکی از بهترین روشهای خوشه یابی

سلولهای دفاعی با ساختار فوق به تعداد لازم و به طور کاملاً تصادفی به عنوان جمعیت اولیه در مرحله مقدمه سازی تولید می شوند.

(ب) قرابت

در روش پیشنهاد شده میزان قرابت هر سلول دفاعی با رابطه زیر سنجیده می شود.

$$Affinity(a) = \left[\frac{\sum_{j=1}^K \sum_{\forall X \in C_j} d(X, Z_j) / |C_j|}{N} \right]^{-1} \quad (1)$$

خوشه مذکور می باشد و N تعداد کل داده ها در فضای ویژگی است.

z_{11}	z_{12}	$\mathbf{1}$	z_{1n}	z_{21}	z_{22}	$\mathbf{1}$	z_{2n}	\dots	\dots	z_{K1}	z_{K2}	$\mathbf{1}$	z_{Kn}
----------	----------	--------------	----------	----------	----------	--------------	----------	---------	---------	----------	----------	--------------	----------

(الف)

z_{11}	z_{12}^*	$\mathbf{1}$	z_{1n}^*	z_{21}	z_{22}	$\mathbf{1}$	z_{2n}^*	\dots	\dots	z_{K1}	z_{K2}^*	$\mathbf{1}$	z_{Kn}
----------	------------	--------------	------------	----------	----------	--------------	------------	---------	---------	----------	------------	--------------	----------

(ب)

شکل 2 - نحوه اعمال عملگر جهش.

جهش یافته می باشند. مکان این مولفه ها به صورت کاملاً تصادفی، با توزیع یکنواخت در بازه $(0, K*n)$ انتخاب می شوند. تعداد مولفه های جهش یافته به نرخ عملگر جهش وابسته است و همانطور که در بخش قبل ذکر شد، نرخ عملگر جهش با میزان قرابت یک سلول دفاعی نسبت معکوس دارد. مقدار یک مولفه جهش یافته از دستور

$$z^* = z + N(0, S) \quad (2)$$

در رابطه فوق $N(0, S)$ بیانگر توزیع نرمال با واریانس تطبیقی S و میانگین صفر است. با این توضیحات روش پیشنهادی تحت عنوان خوشه یابی مبتنی بر الگوریتم ایمنی مصنوعی به وسیله روال معرفی شده در شکل 3 ارائه می شود.

با این تعریف سلولهای دفاعی که دربر دارنده مراکز یی نزدیکترین فاصله نسبت به نمونه های خود میباشند دارای قرابت بیشتری هستند.

تعلق یک نمونه به خوشه C_j در ابتدای کار (پس از ساختن سلولهای دفاعی تصادفی اولیه) به راحتی با استفاده از روش نزدیکترین همسایه صورت می گیرد.

(ج) انتخاب

عملگر انتخاب در مسئله مورد نظر به روش «چرخ رولت»⁹ صورت می گیرد. پس احتمال انتخاب یک سلول دفاعی کاملاً متناسب با مقدار قرابت آن است.

(د) جهش

عملگر جهش از نوع گوسی است. نحوه اعمال این روش در شکل 2 نشان داده شده است.

در شکل 2 مولفه هایی که دارای علامت * هستند مولفه های

شکل 3- روال خوشه یابی به وسیله الگوریتم ایمنی مصنوعی.

- Initialization

-Get (K); K is the number of clusters

-Generate (Anti-Bodies); for K clusters

-Assign each data point to one cluster using this

rule:

$$X \in C_i \quad \text{if } \|X - Z_i\| < \|X - Z_j\| \quad \text{for all } j \neq i$$

-Update Anti-Bodies using this relation:

$$Z_i = \frac{1}{N_j} \sum_{X \in C_j} X$$

- Basic Search Loop:

While (Termination-Condition):

-Assign each data point to one cluster using this

مصنوعی» بر روی مجموعه داده های دو بعدی آزمایش می شود. داده های مورد استفاده مشتمل بر ترکیبات متفاوتی از داده های تصادفی با توزیعهای یکنواخت و نرمال است. شکل 4 (الف تا د) ضمن نمایش این داده ها نتیجه خوشه یابی با الگوریتم پیشنهادی را نشان می دهند. داده های نشان داده شده در این اشکال با استفاده از ترکیبی توابع تصادفی یکنواخت و نرمال و با تغییر واریانس و مقدار میانگین آنها صورت گرفته است.

در این آزمایشات جمعیت اولیه برابر با 20 و نرخ جهش در سلولهای دفاعی متناسب با عکس قرابت آنها با آنتی ژنها در نظر گرفته شده است. تعداد تکرار الگوریتم به صورت پیش فرض 500 در نظر گرفته شده است.

از این شکل معلوم می شود تا زمانیکه دسته های مربوط به داده های ایجاد شده دارای انسجام لازم هستند، روش خوشه یابی پیشنهادی آنها را به خوبی با بهترین تعداد خوشه ها (یعنی 4 خوشه) طبقه بندی و مراکز آنها را مطابق شکل 4-الف و ب به دست می دهد. رفته رفته با افزایش پراکندگی داده ها و تداخل بیش از حد داده های «.» و «+» الگوریتم پیشنهادی داده ها را در سه خوشه با مراکز نشان داده شده در شکل 4-ج طبقه بندی می کند. و بالاخره زمانی که پراکندگی بیش از حد، دو دسته «مربع» و «مثلث» را در یکدیگر ادغام می کند، روش پیشنهاد شده داده ها را در دو خوشه (که البته تعداد خوشه ها به وسیله کاربر تعیین می شود) دسته بندی و

4- نتایج عملی

این بخش به بررسی نتایج آزمایشات مختلفی که بر روی روش پیشنهاد شده در مواجهه با داده های متنوع صورت گرفته است، می پردازد. این نتایج در دوبخش آمده است. در بخش نخست مجموعه های متفاوتی از داده ها در فضای ویژگی دوبعدی شبیه سازی شده اند و روش پیشنهادی برای خوشه یابی آنها به کار گرفته شده است. این بخش بدون انجام هرگونه مقایسه عملکرد این روش با سایر روشهای مرسوم خوشه یابی، صرفاً برای اثبات توانایی آن در خوشه یابی داده های ساده، با مشاهده نتایج خوشه یابی آمده است. در بخش دوم داده های انتخابی مشتمل بر مجموعه های شناخته شده ای در پردازش الگو هستند که ضمن برخورداری از پیچیدگی لازم دارای تنوع در تعداد خوشه ها و ابعاد بردارهای ویژگی نیز می باشند. در بخش دوم ضمن بررسی عملکرد روش خوشه یابی پیشنهاد شده، به مقایسه آن با سایر روشهای مرسوم نیز خواهیم پرداخت. همانطور که قبلاً نیز ذکر شد تعداد خوشه ها به وسیله کاربر تعیین می شود.

1-4- بررسی عملکرد روش پیشنهاد شده در مواجهه با

داده های دوبعدی

در این بخش روش «خوشه یابی بر مبنای الگوریتم ایمنی

مراکز هر خوشه را با موفقیت محاسبه می کند.

4-2- مقایسه عملکرد روش پیشنهادی با سایر روشهای

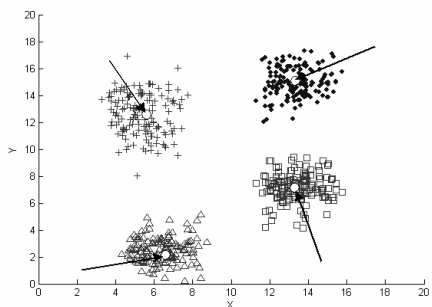
مرسوم خوشه یابی

این بخش حاوی نتایج مقایسه ای روش خوشه یابی پیشنهادی، روش k -means، و روش جدید «خوشه یابی ژنتیک»¹⁰ می باشد. روش k -means یک روش خوشه یابی با الگوریتمی ساده و تکرار گرا است که به صورت وسیعی در «پردازش الگو» مورد استفاده واقع شده است [6]. در کنار این روش، آنچه در تحقیقات جدید جلب توجه می کند، استفاده از روشهای تکاملی برای طراحی روشهای کارا و جدید خوشه یابی است که منجر به ارائه روش «خوشه یابی ژنتیک» شده است [7] و [8].

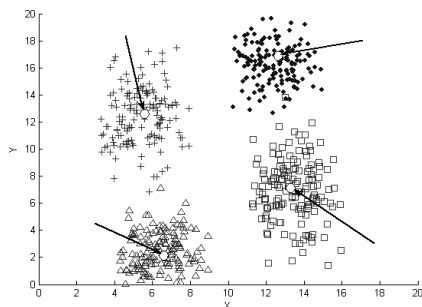
در این بخش روش خوشه یابی پیشنهادی که مبتنی بر الگوریتم ایمنی مصنوعی است در کنار روشهای فوق الذکر بر روی مجموعه های متنوعی از داده ها با ابعاد و پیچیدگیهای مختلف، آزمایش و نتایج آنها به صورت مقایسه ای گزارش شده است.

ابتدا درباره مجموعه داده های مورد استفاده اطلاعاتی ارائه می شود:

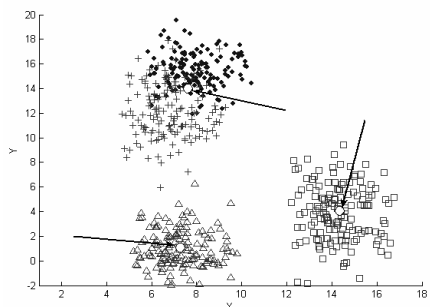
داده های گلهای زنبق - داده های گلهای زنبق شامل 50 اندازه گیری از چهار ویژگی سه نوع گل زنبق به نامهای *Setosa*، *Versicolor* و *Vergenica* است [10]. این ویژگیها عبارتند از طول کاسبرگ، عرض کاسبرگ، طول گلبرگ و عرض گلبرگ. بنابراین در این مسئله سه کلاس موجود است و هرکلاس دارای 50 الگو با بردارهای ویژگی چهار بعدی است.



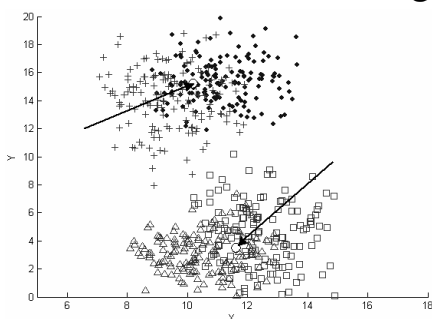
(الف)



(ب)



(ج)



(د)

شکل 4- نتایج حاصل از خوشه یابی بر روی داده های نشان داده شده به ازای پراکندگی های مختلف نمونه ها.

داده های انواع شراب – این مجموعه متشکل از آنالیز شیمیایی عناصر رشد داده شده شرابهایی یکسان در شرایط مساوی است [11]. بردارهای ویژگی دارای 13 مولفه هستند. در مجموع 178 الگو از سه کلاس مجزا در این مجموعه داده وجود دارد.

داده های سرطان سینه – این داده ها به تعداد 683 نمونه مربوط به به توده های موجود در سینه بیماران مراجعه کننده به بیمارستان دانشگاه Wisconsin است. نمونه ها دارای ابعاد 9 بعدی بوده در دو کلاس خوش خیم و بد خیم دسته بندی شده اند [12].

مجموعه داده های انتخابی دارای تنوع قابل قبولی در ابعاد فضای ویژگی، تعداد نمونه ها، تعداد کلاسها و تداخل آنها می باشند و از این رو می توانند مسایل مناسبی برای بررسی عملکرد روش خوشه یابی پیشنهادی به شمار آیند. در تمامی آزمایشات تعداد سلولهای دفاعی اولیه در روش پیشنهادی 20 در نظر گرفته شده است. همین تعداد کروموزوم در روش خوشه یابی ژنتیک به عنوان جمعیت اولیه به صورت تصادفی تولید می شود. احتمال جهش در روش خوشه یابی ایمنی مصنوعی متناسب با عکس «قربت» بوده از 0/001 تا 0/5 تغییر می کند. در روش خوشه یابی ژنتیک احتمال جهش مقدار ثابت 0/001 و احتمال بازترکیب¹¹ مقدار ثابت 0/8 (همان مقادیری که در [7] گزارش شده اند) در نظر گرفته می شود. برای هر سه روش (در صورت عدم همگرایی) حداکثر 1000 تکرار در فرایند جستجوی مراکز خوشه ها در نظر گرفته شده است.

با توجه به ماهیت تصادفی هر سه روش خوشه یابی ایمنی مصنوعی، k-means و ژنتیک، آزمایشات برای هر مجموعه داده به تعداد 20 بار انجام شده است و مقادیری که گزارش خواهند شد میانگین نتایج حاصل از این آزمایشات است. خطای کوانتیزه کردن¹²، معیار برگزیده برای مقایسه عملکرد سه روش خوشه یابی می باشد. خطای کوانتیزه رایج ترین

معیار بوده دارای تعریفی به شکل عکس آنچه در مورد «قربت» در رابطه (1) آمده است، می باشد.

جدول 1 نتایج به دست آمده از اعمال سه روش خوشه یابی ایمنی مصنوعی، k-means و ژنتیک را بر روی مجموعه داده های انتخابی در بر دارد. با ملاحظه این مقادیر، بهبود فرایند خوشه یابی در روش پیشنهاد شده بر مبنای الگوریتم ایمنی مصنوعی نسبت به دو روش دیگر در اغلب موارد به چشم می خورد. خطای کوانتیزه کردن در روش پیشنهادی در دو مجموعه داده گلپای زنبق و شراب از هر دو روش دیگر بهتر است. تنها در داده های سرطان سینه خطای خوشه یابی ژنتیک به مقدار ناچیزی از روش پیشنهاد شده بهتر است. میزان خطای کوانتیزه کردن تا حدودی می تواند بیانگر تاثیر داده های نویزی در مجموعه داده های فوق الذکر نیز باشد.

جدول 1 – مقادیر مقایسه ای خوشه یابی مجموعه داده های گلپای زنبق، شراب و سرطان سینه برای سه روش خوشه یابی ایمنی مصنوعی، k-means و ژنتیک

مجموعه داده	روش خوشه یابی	خطای کوانتیزه کردن
گلپای زنبق	روش ایمنی مصنوعی	0.635
	k-means	0.701
	روش ژنتیک	0.674
شراب	روش ایمنی مصنوعی	1.004
	k-means	1.201
	روش ژنتیک	1.105
سرطان سینه	روش ایمنی مصنوعی	1.990
	k-means	2.005
	روش ژنتیک	1.987

5- نتیجه گیری

در این تحقیق یک روش جدید خوشه یابی بر مبنای الگوریتم ایمنی مصنوعی ارائه شد. خوشه یابی مجموعه داده های متنوع در این مقاله در اغلب موارد عملکرد بهتری را برای روش پیشنهاد شده در مقایسه با روش k-means (به عنوان یک روش رایج و قدیمی) و خوشه یابی ژنتیک (به عنوان روشی

¹¹ Crossover

¹² Quantization error

& Sons, 1975.

[7] U. Maulik and S. Bandyopadhyay, "Genetic algorithm based clustering technique," Journal of the Pattern Recognition, vol. 33 pp. 1455-1465, 2000.

[8] L.Y. Tseng, and S.B. Yang, "A genetic approach to the automatic clustering problem," Journal of the Pattern Recognition, vol. 34 pp. 415-424, 2001.

[9] DW. Van der Merwe and AP. Engelbrecht, " Data clustering using particle swarm optimization," In Proc. of the 2003 Congress on Evolutionary Computation, pp. 215-220, 2003.

[10] R. A. Fisher, "The use of multiple measurements in taxonomic problems," Ann. Eugen, vol. 7, pp. 179-188, 1936.

[11] University of California, Irvine, via anonymous ftp [ftp ftp.ics.uci.edu/pub/machine-learning-databases](ftp://ftp.ics.uci.edu/pub/machine-learning-databases).

[12]<http://www.ics.uci.edu/~mllearn/MLRepository.html>
1.

تکاملی با سابقه کمتر) گزارش می کنند.

نکته مهمی که نباید از نظر دور داشت این است که اگرچه در غالب اوقات عملکرد روش خوشه یابی ایمنی مصنوعی نسبت به دو روش دیگر دارای بهبود نسبتا قابل توجهی است، اما نرخ پایین همگرایی آن نسبت به روش *k-means* باعث افزایش هزینه محاسباتی آن می شود. البته مطابق نتایج به دست آمده همگرایی سریع روش *k-means* را فقط می توان به صورت گرفتاری آن در یک پاسخ محلی¹³ تعبیر کرد. به عبارت بهتر، به ازای هزینه محاسباتی بیشتر، روش خوشه یابی ایمنی مصنوعی به سمت پاسخ کلی¹⁴ میل می کند.

مراجع

[1] S. M. Thayer, S. P. N. Singh, "Development of an immunology-based multi-robot coordination algorithm for exploration and mapping domains," In the Proceedings of Intelligent Robots and System, International Conference (IEEE/RSJ 2002), vol. 3, pp. 2735 -2739. September 30 - October 5, 2002.

[2] D. R. Carvalho and A. A. Freitas, "An immunological algorithm for discovering small-disjunct rules in data mining. GECCO'2001. (Workshop Program.) San Francisco," California. July 7, 2001.

[3] R. Canham, A.H. Jackson, A. Tyrrell, "Robot Error Detection using an Artificial Immune System. Evolvable Hardware, Proceedings. NASA/DoD Conference," pp 199 -207, July 9-11, 2003.

[4] I.E Evangelou, D.G. Hadjimitsis, A.A. Lazakidou, and C. Clayton, "Data mining and knowledge discovery in complex image data using artificial neural networks," Workshop on Complex Reasoning and Geographical Data, Cyprus, 2001.

[5] T. Lillesand, and R. Keifer, Remote Sensing and Image Interpretation, John Wiley & Sons, 1994.

[6] J.A. Hartigan, Clustering Algorithms, John Wiley

¹³ Local

¹⁴ Global