

یک مدل همکارانه جدید برای خوشبندی داده‌ها

دانیال یزدانی^۱، فرهاد محمدکاظمی^۲، محمدرضا میبدی^۳

^۱دانشگاه آزاد اسلامی، واحد شیروان، ایران
d.yazdani@IEEE.org

^۲دانشگاه پیام نور، مشهد، ایران
fmkazemi@pnu.ac.ir

^۳دانشکده مهندسی کامپیوتر، فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران
mmeybodi@aut.ac.ir

چکیده

الگوریتم دسته ماهی‌های مصنوعی (AFSA) یکی از الگوریتم‌های هوش جمعی است که بر اساس جمعیت و جستجوی تصادفی کار می‌کند و کاربرد اصلی آن در حل مسائل بهینه‌سازی است. این الگوریتم در کاربردهای مختلفی از جمله خوشبندی داده‌ها، آموزش شبکه‌های عصبی، بهینه‌سازی توابع غیرخطی و غیره به کار رفته است. خوشبندی داده‌ها در زمینه‌های زیادی از جمله یادگیری ماشین، داده‌کاوی، شبکه‌های حسگر بی‌سیم و تشخیص الگو، بکار رفته است. یکی از معروف‌ترین روش‌های خوشبندی، روش K-means می‌باشد که به طور مؤثری در بسیاری از مسائل خوشبندی به کار رفته است ولی این الگوریتم دارای مشکلاتی از جمله همگرایی به مینیمم محلی و حساسیت به نقاط ابتدایی می‌باشد. در این مقاله یک روش خوشبندی همکارانه براساس AFSA و K-means پیشنهاد می‌گردد. در الگوریتم پیشنهادی از توانایی بالای AFSA در جستجوی سراسری در کنار توانایی بالای K-means در جستجوی محلی به صورت همکارانه استفاده شده است. الگوریتم پیشنهادی بر روی پنج مجموعه داده استاندارد تست شده است و کارایی آن با الگوریتم‌های PSO، K-means و KPSO و AFSA مقایسه شده است. نتایج آزمایشات نشان می‌دهند روش پیشنهادی دارای کارایی مناسب و قابل قبولی در خوشبندی داده است.

کلمات کلیدی

الگوریتم دسته ماهی‌های مصنوعی، خوشبندی داده‌ها، K-means، روش‌های همکارانه، بهینه‌سازی دسته ذرات.

یکی از خوشبندی‌ها تخصیص خواهد داد، بر این اساس که داده‌های تخصیص داده شده به یک خوشبندی، نسبت به داده‌هایی که در خوشبندی دیگر هستند، به هم شبیه‌تر باشند. الگوریتم K-means یکی از معروف‌ترین روش‌های خوشبندی است که در بسیاری از مسائل به کار رفته است [۶][۷][۸]. الگوریتم K-means با k مرکز خوشبندی تصادفی شروع می‌شود و مجموعه‌ای از اشیاء را به k زیر مجموعه تقسیم می‌کند. این روش یکی از محبوب‌ترین و پر استفاده‌ترین تکنیک‌های خوشبندی است زیرا به سادگی قابل درک و اجرا است و همچنین دارای پیچیدگی زمانی خطی است. اما الگوریتم K-means چندین مشکل اساسی است. از جمله این مشکلات می‌توان به گیرافتادن در بهینه‌های محلی و حساس بودن به مقادیر اولیه مراکز خوشبندی اشاره کرد [۶].

خوشبندی داده‌ها جز مسائل NP می‌باشد. یافتن راه حل برای مسائلی که از نوع NP می‌باشند بسیار مشکل است. الگوریتم‌هایی از جمله

۱- مقدمه

خوشبندی داده‌ها، دارای کاربردهای فراوانی در داده کاوی [۱]، شبکه‌ای حسگر بی‌سیم [۲][۳]، تشخیص الگو [۴] و یادگیری ماشین [۵] می‌باشد. اهمیت خوشبندی در علوم مختلف و همچنین نوع داده‌های مورد استفاده، سرعت خوشبندی، دقت و بسیاری پارامترهای دیگر باعث معرفی روش‌ها و الگوریتم‌های متنوعی از خوشبندی داده‌ها شده است. خوشبندی یک تکنیک دسته‌بندی بدون نظارت است که در آن مجموعه‌های داده‌ها که معمولاً بردارهایی در فضای چندبعدی می‌باشند، بر اساس یک معیار شباهت یا عدم شباهت، به تعداد مشخصی خوشبندی تقسیم می‌شوند. وقتی که تعداد خوشبندی‌ها برابر K باشد و n داده بعدی داشته باشیم، الگوریتم خوشبندی، هر یک از این داده‌ها را به

بنابراین هر ماهی مصنوعی دارای $K \times D$ بعد خواهد بود. شکل ۱، بردار یک ماهی مصنوعی را نشان می‌دهد که شامل K مرکز خوشة D بُعدی می‌باشد.

$$[Z_{1,1}, Z_{1,2}, \dots, Z_{1,D}, Z_{2,1}, Z_{2,2}, \dots, Z_{2,D}, \dots, Z_{K,1}, Z_{K,2}, \dots, Z_{K,D}]$$

شکل(۱) : ساختار موقعیت یک ماهی مصنوعی در فضای مسئله خوشه‌بندی.

در الگوریتم پیشنهادی ابتدا ماهی‌های مصنوعی دسته را به صورت تصادفی مقادره‌ی اولیه می‌کنیم، بدین ترتیب هر یک از ماهی‌های مصنوعی شامل K مرکز خوشة اولیه تصادفی می‌باشند. سپس با توجه به اختصاص بردارهای داده به هر یک از خوشه‌ها در ماهی‌های مصنوعی بر اساس فاصله اقلیدوسی و تابع شایستگی رابطه ۱، رفتارهای الگوریتم دسته ماهی‌های مصنوعی را برای ماهی‌های مصنوعی اجرا می‌کنیم.

برای مسئله خوشه‌بندی، AFSA دارای یک ضعف عمدی است. در AFSA، پارامترهای میدان دید و طول گام، دو پارامتر عددی هستند و یک مقدار دارند. خوشه‌بندی داده‌ها از جمله مسائلی است که برای حل آن نیاز به ماهی‌های مصنوعی با ابعاد بالا دارد. در صورتی که بازه فضای مسئله برای تمام این متغیرها برابر باشد، می‌توان مقدار مناسبی را برای پارامترهای میدان دید و طول گام تعیین کرد. اما در برخی مسائل خوشه‌بندی، ابعاد مختلف بردارهای داده بازه‌های متفاوتی دارند. برای مثال فرض کنید بعد از تمام بردارهای داده در بازه [۱۲، ۴۰] قرار دارد و بردار زام این بردارهای داده در بازه [۱۰۰، ۴۰] قرار دارد. در این حالت اگر مقدار پارامترهای میدان دید و طول گام را کوچک در نظر بگیریم تا مناسب بُعد آن باشد، جستجو در بُعد زام دچار مشکل می‌شود به این دلیل که مقدار پارامترهای میدان دید و طول گام کوچک است سرعت همگرایی بسیار پایین می‌آید و احتمال افتادن در بهینه‌های محلی در این بُعد بسیار بالا می‌رود. اگر مقدار پارامترهای میدان دید و طول گام را بزرگ در نظر بگیریم تا مناسب بُعد زام باشد، جستجو در بُعد آن دچار مشکل می‌شود زیرا در این حالت مقدار پارامترهای میدان دید و طول گام برای جستجوی بُعد آن بسیار زیاد است و این امر باعث جستجو در فضایی خارج از بازه مقادیر مربوط به این بُعد می‌شود و احتمال یافتن مقادیر با شایستگی بهتر در این بُعد بسیار پایین می‌آید. بدین ترتیب نمی‌توان مقدار پارامترهای میدان دید و طول گام را طوری تعیین کرد که جستجو در همه ابعاد با بازه‌های متفاوت به خوبی انجام شود.

برای رفع این مشکل پارامترهای میدان دید و طول گام را به صورت برداری در نظر می‌گیریم. تعداد ابعاد پارامترهای میدان دید و طول گام برابر ابعاد ماهی‌های مصنوعی است. بدین ترتیب مقدار پارامترهای میدان دید و طول گام در بُعد آن بر اساس بازه تغییرات این بُعد در فضای مسئله تعیین می‌شود.

الگوریتم‌های هوش جمعی تا حدی این مشکل را حل نموده‌اند. توسط این نوع الگوریتم‌ها راه حل‌هایی پیدا می‌شوند که تقریباً به جواب نزدیکند. الگوریتم دسته ماهی‌های مصنوعی (AFSA)^(۱) یکی از الگوریتم‌های برگرفته شده از طبیعت و هوش جمعی می‌باشد که توسط دکتر Lei Xiao در سال ۲۰۰۲ ارائه شده است [۹]. این الگوریتم تکنیکی مبتنی بر رفتارهای جمعی است که از رفتارهای اجتماعی دسته ماهی‌ها در طبیعت الهام گرفته شده است. این الگوریتم دارای ویژگی‌های سرعت همگرایی بالا، غیرحساس بودن به مقادیر اولیه، انعطاف‌پذیری و تحمل پذیری خطای بالایی است. از این الگوریتم در کاربردهای بهینه‌سازی از جمله یادگیری شبکه‌های عصبی پیشرو [۱۰]، خوشه‌بندی داده‌ها [۱۱]، داده‌کاوی [۱۲]، بهینه‌سازی توابع غیرخطی [۱۳]، بهینه‌سازی ترکیبی [۱۵] و ... استفاده شده است.

در این مقاله یک الگوریتم همکارانه جدید بر پایه K- means و AFSA ارائه می‌شود. AFSA وظیفه جستجوی سراسری و یافتن بهینه‌ها را بر عهده دارد و K-means وظیفه جستجوی محلی و افزایش دقت جواب‌ها را انجام می‌دهد. الگوریتم پیشنهادی به همراه الگوریتم‌های بهینه‌سازی دسته ذرات (PSO) [۱۷]، K-means [۱۸] و AFSA [۱۸] استاندارد بر روی پنج مجموعه داده استاندارد آزمایشات نشان می‌دهند الگوریتم پیشنهادی از کارایی قابل قبولی برخوردار است.

ادامه این مقاله بدین ترتیب سازماندهی شده است. بخش دوم الگوریتم پیشنهادی را تشریح می‌کند. در بخش سوم نتایج آزمایشات مورد بررسی قرار می‌گیرند و بخش نهایی به بیان نتیجه‌گیری می‌پردازد.

۲- الگوریتم پیشنهادی

در الگوریتم پیشنهادی، برای یافتن مقادیر بهینه مرکز خوشه (که تعداد آنها از پیش تعیین شده است)، از رابطه ۱ استفاده می‌شود. در واقع رابطه ۱، تابع شایستگی مورد نظر می‌باشد که قرار است بهینه‌سازی (کمینه‌سازی) شود و فاصله‌های درون خوشه‌ای (Intra cluster distances) را محاسبه می‌کند [۲۰].

$$J(C_1, C_2, \dots, C_K) = \sum_{i=1}^K \left(\sum_{X_j \in C_i} \|Z_i - X_j\| \right) \quad (1)$$

در رابطه ۱، مجموع فاصله اقلیدوسی تمام بردارهای داده از مرکز خوشه‌ای که عضو آن هستند محاسبه و جمع می‌شود. در این رابطه K خوشه (C) داریم که هر یک از N بردار داده (X) بر اساس فاصله‌ای که از هر یک از این مرکز خوشه (Z) دارند، دسته‌بندی می‌شوند و در یکی از خوشه‌ها قرار می‌گیرند. بنابراین هدف، تعیین مرکز خوشه‌ای است که رابطه ۱ را کمینه کنند، بدین ترتیب مرکز خوشه بهینه تعیین می‌شوند. از آنجایی که داده‌ها D بُعدی بوده و تعداد K خوشه وجود دارد،

```

for each Artificial Fish  $i \in [1 .. N]$ 
    initialize  $x_i$ 
endfor
Blackboard =  $\arg \min_{X_i} f(X_i)$ 
repeat:
    for each Artificial Fish  $i \in [1 .. N]$ 
        Perform Swarm Behavior on  $X_i(t)$  and Compute  $X_{i,swarm}$ 
        Perform Follow Behavior on  $X_i(i)$  and Compute  $X_{i,follow}$ 
        if  $f(X_{i,swarm}) \geq f(X_{i,follow})$ 
            then  $X_i(t+1) = X_{i,follow}$ 
        else
             $X_i(t+1) = X_{i,swarm}$ 
        endifor
        if  $f(X_{Best-AF}) \leq f(Blackboard)$ 
            then  $Blackboard = X_{Best-AF}$ 
    if swarm is converged then
        Execute k-means on  $X_{Best-AF}$  Until stopping criterion is met
        if  $f(X_{k-means}) \leq f(Blackboard)$  then
             $Blackboard = X_{k-means}$ 
        Reinitialize swarm //end of IF body
    until stopping criterion is met

```

شکل(۲) : شبه کد الگوریتم پیشنهادی.

آزمایشات در بخش بعد ارائه شده است.

۳- نتایج آزمایشات

برای ارزیابی صحت و کارایی الگوریتم پیشنهادی، آزمایشات بر روی پنج مجموعه داده استاندارد انجام شده است که از آنها برای تعیین صحت و درستی الگوریتم های خوشبندی استفاده می شود. این مجموعه داده ها شامل Iris ، Glass ، Wine ، Sonar و WDBC هستند که از مجموعه داده های استاندارد UCI انتخاب شده اند که در ادامه خصوصیات هر یک از آنها شرح داده شده است [۱۹]:

۱- Iris: این مجموعه داده مربوط به شناخت گل های زنبق می باشد که در آن سه کلاس مختلف وجود دارد و هر کلاس شامل ۵۰ نمونه می باشد. هر نمونه دارای چهار ویژگی است.

۲- Wine: این مجموعه داده مربوط به شناخت نوشیدنی ها می باشد که در مجموع دارای ۱۷۸ نمونه است که به ترتیب در سه کلاس مختلف شامل ۵۹، ۷۱ و ۴۸ نمونه دستributed شده اند. در این مجموعه داده هر نمونه دارای ۱۳ ویژگی می باشد.

۳- Glass Identification Database: این مجموعه داده مربوط به اطلاعات چندین نوع شیشه است که در مجموع ۲۱۴ نمونه دارد که در ۶ کلاس مختلف کلاس بندی شده اند. تعداد نمونه کلاس های ۱ تا ۶ به ترتیب ۲۹، ۱۳، ۷۶، ۱۷، ۷۰ و ۹ است که هر یک از آنها دارای ۹ ویژگی با مقادیر پیوسته می باشد.

۴- Wisconsin Diagnostic Breast Cancer (WDBC): این مجموعه داده مربوط به اطلاعات غده سرطانی بیماران مبتلا به

در AFSA موقعیت بهترین نقطه ای که تاکنون توسط اعضای دسته یافته شده در تخته سیاه [۹] ذخیره می شود و ماهی مصنوعی که آنرا یافته است ممکن است با اجرای یک حرکت آزادانه حتی به موقعیت های بدتر حرکت کند. بنابراین ماهی های مصنوعی نمی توانند از بهترین تجربه گروهی خود در راستای افزایش سرعت همگرایی استفاده کنند زیرا تنها آنرا در تخته سیاه ذخیره می کنند. از طرفی انجام حرکت آزادانه در دسته یک امر حیاتی برای حفظ تنوع در دسته است. در این مقاله برای رفع این مشکل، حرکت آزادانه را تنها برای بهترین ماهی مصنوعی دسته انجام نمی دهیم. بدین ترتیب بهترین ماهی مصنوعی دسته تنها در صورتی حرکت می کند که به سمت موقعیت های بهتر باشد. در نتیجه اعضای دیگر دسته می توانند با اجرایی رفتارهای دنباله روی و حرکت جمعی به سمت بهترین موقعیت یافت شده حرکت کنند.

در الگوریتم پیشنهادی AFSA اجرا می شود و پس از اینکه دسته ماهی های مصنوعی همگرا شد، موقعیت بهترین ماهی مصنوعی دسته که شامل K مرکز خوشی است به الگوریتم K-means داده می شود تا این الگوریتم بتواند با توانایی بالایی که در جستجوی محلی دارد، موقعی مراکز خوشی بدست آمده از AFSA را بهبود بخشد. پس از اینکه موقعیت بهترین ماهی مصنوعی دسته به الگوریتم k-means داده شد، ماهی های مصنوعی در فضای مسئله مقداردهی اولیه می شوند و یک جستجوی جدید را در فضای مسئله آغاز می کنند. هنگامی دسته همگرا شده که مقدار اختلافتابع شایستگی بهترین ماهی مصنوعی آن پس از چند تکرار معین، از یک حد آستانه ای کوچکتر باشد.

پس از هر بار همگرایی دسته، موقعیت بهترین ماهی مصنوعی دسته به الگوریتم k-means سپرده می شود و دسته در فضای مسئله مقداردهی اولیه می شود. الگوریتم k-means تا زمانی که موقعیت مراکز خوشی بدست آمده در دو تکرار متوالی یکسان نباشد اجرا می شود. پس از پایان اجرای k-means فاصله درون خوشه ای بدست آمده با فاصله درون خوشه ای ثبت شده در تخته سیاه می شود. در صورتی که مقدار فاصله درون خوشه ای بدست آمده بهتر از مقدار فاصله درون تخته سیاه باشد، مراکز خوشی بدست آمده به همراه مقدار فاصله درون خوشه ای مربوط به آن در تخته سیاه به جای مقادیر قبلی ذخیره می شوند. برای پایان الگوریتم می توان تعداد تکرار معینی را از اجرای AFSA تعیین کرد. پس از پایان اجراء، نتیجه نهایی در تخته سیاه ذخیره شده است.

در الگوریتم پیشنهادی AFSA وظیفه انجام جستجوی سراسری و means وظیفه انجام جستجوی محلی را دارد. در طول اجرا ممکن است AFSA قله های متفاوتی را در فضای مسئله بیابد و جستجوی محلی آنرا به k-means بسپارد. بدین ترتیب دو الگوریتم با هم برای رسیدن به مقادیر بهتر فاصله درون خوشه ای همکاری می کنند. شبه کد الگوریتم پیشنهادی در شکل ۲ نشان داده شده است.

پارامترهای برداری میدان دید، در هر یک از ابعاد برابر ۰.۲۰٪ از محدوده تغییر مقادیر بردار نمونه‌ها در همان بعد و مقدار پارامتر برداری طول-گام در هر بعد برابر نصف مقدار میدان دید در همان بعد در نظر گرفته شده است. در PSO مقدار c_1 و c_2 برابر ۲ در نظر گرفته شده است و وزن اینترسی در هر تکرار از رابطه $W = \text{rand}/2 + 0.5$ بدست می-آید [۱۷]. مقادیر پارامترها بر اساس آزمایشات بسیار زیاد و مراجع مربوطه تنظیم شده‌اند.تابع شایستگی برابر فاصله درون خوش‌های است و با رابطه ۱ محاسبه می‌شود. آزمایش‌ها ۳۰ بار تکرار شده‌اند و بهترین، متوسط و انحراف معیار نتایج بهینه‌سازی فاصله درون خوش‌های با الگوریتم‌های AFSA استاندارد، Kmeans، PSO و KPSO پیشنهادی در جدول ۱ بر روی مجموعه داده‌های نامبرده نشان داده شده است و بهترین نتیجه پررنگ شده است. جدول ۲ متوسط و انحراف معیار خطای خوش‌بندی پنج الگوریتم را بر روی مجموعه داده‌های نامبرده نشان می‌دهد.

همانطور که در جدول ۱ مشاهده می‌شود، الگوریتم پیشنهادی به نتایج بهتری نسبت به دیگر الگوریتم‌ها دست یافته است و مطابق با جدول ۲، خوش‌بندی را با خطای کمتری نسبت به دیگر الگوریتم‌ها انجام داده است. مطابق با مقادیر نشان داده در جدول ۱، متوسط نتیجه بدست آمده از الگوریتم پیشنهادی بسیار نزدیک به بهترین نتیجه است که مقدار کم انحراف معیار الگوریتم پیشنهادی نیز این امر را تأیید می‌کند. دلیل پایداری نتایج الگوریتم پیشنهادی در اجراهای مختلف این است که AFSA در یک بار اجرا بارها بهینه‌هایی را پیدا می‌کند و پس از سپردن نتایج به k-means در فضای مسئله پراکنده می‌شود و از اول جستجو را آغاز می‌کند. این امر باعث شده تا الگوریتم در یک بار اجرا با احتمال خیلی بالا بتواند قله بهینه را بیابد و در بهینه‌های محلی گیر نکند. الگوریتم‌های دیگر به دلیل افتادن در بهینه‌های محلی به نتایج قابل قبولی در مقایسه با الگوریتم پیشنهادی دست نیافرته‌اند.

سرطان سینه است که در دانشگاه Wisconsin گردآوری شده است. در این مجموعه داده ۵۶۹ نمونه در دو کلاس خوش‌خیم (۳۵۷ نمونه) و بدخیم (۲۱۲ نمونه) کلاس‌بندی شده است. هر یک از نمونه‌های این مجموعه داده دارای ID بیمار است و ۳۰ ویژگی می‌باشد که یکی از ویژگی‌ها شماره ID بیمار است و ۳۰ ویژگی دیگر مربوط به مشخصات غده سرطانی مانند بافت، فشردگی، مساحت و غیره می‌باشند که همگی عدد حقیقی می‌باشند و دسته‌بندی بر اساس همین ۳۰ ویژگی انجام می‌شود.

- ۵. Sonar, Mines vs. Rocks: این مجموعه داده برای کلاس‌بندی سیگنال‌های Sonar در زیردریایی‌ها می‌باشد. سیگنال‌های Sonar در دو کلاس تشخیص مین‌های دریایی و صخره‌ها طبقه‌بندی شده‌اند. در این مجموعه داده ۲۰۸ نمونه وجود دارد که کلاس مین‌های دریایی شامل ۱۱۱ نمونه و کلاس صخره‌ها شامل ۹۷ نمونه می‌باشد. هر یک از این نمونه‌ها دارای ۶۰ ویژگی پیوسته هستند.

در آزمایشات انجام‌شده، جمعیت در AFSA استاندارد، PSO و KPSO، مساوی ۵ برابر تعداد ابعاد فضای مسئله است. در الگوریتم پیشنهادی جمعیت برابر با دو برابر تعداد ابعاد فضای مسئله خواهد بود. ابعاد فضای مسئله برای هر یک از مجموعه داده‌ها برابر حاصلضرب تعداد کلاس‌های آن مجموعه داده در تعداد ویژگی‌های نمونه‌های آن مجموعه داده است. حداکثر تکرار برای خوش‌بندی هر یک از مجموعه داده‌ها مساوی ۱۰ برابر تعداد ابعاد فضای مسئله در نظر گرفته شده [۲۰]. در AFSA استاندارد و الگوریتم پیشنهادی، مقدار فاکتور شلوغی برابر ۰/۵ و حداکثر تعداد تلاش‌ها برابر ۱۰ می‌باشد. در AFSA استاندارد مقدار پارامتر میدان دید برابر ۰/۲۰٪ از محدوده تغییر مقادیر ابعاد نمونه‌ها و مقدار پارامتر طول گام برابر نصف مقدار میدان دید در نظر گرفته شده است. در الگوریتم پیشنهادی، مقدار

جدول (۱): مقایسه فاصله درون خوش‌های ۵ الگوریتم بر روی ۵ مجموعه داده استفاده شده در این مقاله.

K-means	Std-AFSA	PSO	KPSO	الگوریتم پیشنهادی	معیارها	مجموعه داده
۹۷.۲۲	۹۶.۹۱	۹۷.۱۰	۹۶.۷۸	۹۶۶۵	بهترین	Irish
۱۰.۲۵۷	۱۱۲.۳۲	۱۰۲.۲۶	۹۹.۶۱	۹۶۶۵	متوسط	
۱۱.۳۴	۵.۴۶	۵.۸۱	۷.۲۱	۰.۰۰	انحراف معیار	
۱۶۵۵۵.۶۸	۱۶۶۹۵.۴۷	۱۶۳۰.۷۱۶	۱۶۲۹۸.۹۲	۱۶۴۹۳.۴	بهترین	Wine
۱۷۶۶۲.۷۳	۱۶۸۴۴.۱۰	۱۶۳۰.۶۷	۱۶۳۰.۷۵۸	۱۶۴۹۳.۴۸	متوسط	
۱۸۷۸.۰۷	۵۷.۱۹	۹.۵۳	۷.۲۳	۰.۰۶	انحراف معیار	
۲۱۳۴۲	۳۴۸.۹۳	۲۳۰.۶۴	۲۱۲.۰۳	۲۱۰.۰۸	بهترین	Glass
۲۴۱۰.۳	۳۶۷.۵۲	۲۵۸.۰۲	۲۳۳.۲۸	۲۱۳.۹۵	متوسط	
۲۵۰.۲۲	۸.۷۲	۱۲.۲۴	۱۴.۰۵	۱.۴۶	انحراف معیار	
۱۵۲۶۴۷.۲۵	۱۵۲۹۴۶.۱۶	۱۴۹۵۳۷.۷۳	۱۴۹۴۸۰.۹۳	۱۴۹۴۷۴.۳۵	بهترین	WDBC
۱۷۹۷۹۴.۴۵	۱۵۳۳۸۸.۹۸	۱۴۹۸۳۰.۸۷	۱۴۹۵۹۴.۰۵	۱۴۹۴۸۰.۴۰	متوسط	
۵۵۲۲۲.۱۷	۲۱۱.۵۹	۳۶۴.۷۳	۱۹۸.۳۱	۲.۲۶	انحراف معیار	
۲۳۴.۷۷	۳۹۹.۶۶	۲۷۱.۸۳	۲۳۴.۶۵	۲۲۳.۷۵	بهترین	Sonar
۲۳۵.۰۶	۴۳۲.۸۹	۲۷۶.۶۸	۲۳۴.۹۲	۲۲۳.۷۵	متوسط	
۰.۱۵	۱۵.۴۷	۳.۷۹	۰.۲۲	۰.۰۰	انحراف معیار	

جدول (۲): مقایسه متوسط و انحراف معیار نرخ خطای خوشبندی ۵ الگوریتم بر روی ۵ مجموعه داده استفاده شده در این مقاله.

K-means	Std-AFSA	PSO	KPSO	الگوریتم پیشنهادی	مجموعه داده
۱۶.۰۵±۱۰.۱۰	۲۷.۸۹±۹.۰۱	۱۰.۶۴±۴.۵۰	۱۲.۵۸±۷.۶۷	۱۰۰.۰±۰.۰۰	Irish
۳۴.۳۸±۶.۰۸	۲۹.۱۱±۰.۴۳	۲۸.۷۴±۰.۳۹	۲۸.۵۹±۰.۴۷	۲۸.۰۵±۰.۰۰	Wine
۴۸.۷۰±۳.۱۴	۵۴.۰۷±۳.۲۴	۴۸.۷۲±۱.۳۴	۴۷.۸۰±۱.۹۸	۴۷.۷۵±۲.۹۴	Glass
۱۹.۱۲±۹.۲۲	۱۳.۲۸±۰.۲۸	۱۳.۱۸±۰.۰۰	۱۳.۱۸±۰.۰۰	۱۳.۱۸±۰.۰۰	WDBC
۴۴.۹۵±۰.۹۷	۴۶.۶۳±۰.۰۰	۴۶.۶۰±۰.۴۲	۴۴.۸۹±۰.۸۴	۴۴.۸۵±۰.۰۹	Sonar

- [9] L. X. Li, Z. J. Shao and J. X. Qian, "An Optimizing Method based on Autonomous Animals: Fish Swarm Algorithm", In Proc. of Systems Engineering Theory & Practice, 2002.
- [10] C. R. Wang, C. L. Zhou and J. W. Ma, "An Improved Artificial Fish Swarm Algorithm and Its Application in Feed-Forward Neural Networks", In Proc. of 4th International Conference on Machine Learning and Cybernetics, Guangzhou, 2005.
- [11] S. Hi, N. Belacel, H. Hamam and Y. Bouslimani, "Fuzzy Clustering with Improved Artificial Fish Swarm Algorithm", In International Joint Conference on Computational Sciences and Optimization 09, Vol. 2, pp. 317-321, Hainan, China, 2009.
- [12] M. Zhang, C. Shao, M. Li and J. Sun, "Mining Classification Rule with Artificial Fish Swarm", in 6th World Congress on Intelligent Control and Automation, Vol. 2, pp. 5877-5881, Dalian, 2006.
- [13] D. Yazdani and A. T. Haghigat, "A Modified Artificial Fish Swarm Algorithm with New Behavior for Avoiding Trap in Local Optima", in 17th Iranian Conference on Electrical Engineering(ICEE2009), Tehran, Iran, 2009.
- [14] D. Yazdani, Adel Nadjaran and M. R. Meybodi, "Fuzzy Adaptive Artificial Fish Swarm Algorithm", in 23th international conference on artificial intelligence, Adelaide, Australia, 2010.
- [15] D. Yazdani, S. Golyari and M. R. Meybodi, "A New Hybrid Algorithm for Optimization Based on Artificial Fish Swarm Algorithm and Cellular Learning Automata", in 5th International Conference on Telecommunication (IST2010), Tehran, Iran, 2010.
- [16] D. Yazdani and M. R. Meybodi, "AFSA-LA: A New Model for Optimization", in 15th Conference of Computer Society of Iran (CSICC2010), Tehran, Iran, 2010.
- [17] A. A. A. Esmin, D. L. Pereira and F. Araujo, "Study of Different Approach to Clustering Data by Using the Particle Swarm Optimization Algorithm", in IEEE Congress on Evolutionary Computation, pp. 1817-1822, Hong Kong, June 2008.
- [18] D. W. van der Merwe and A. P. Engelbrecht, "Data Clustering Using Particle Swarm Optimization", in the 2003 Congress on Evolutionary Computation, Vol. 1, pp. 215-220, December 2003.
- [19] <http://archive.ics.uci.edu/ml/>
- [20] Y. T. Kao, E. Zahara and I. W. Kao, "A Hibridized Approach to Data Clustering", in Elsevier Journal on Expert System with Applications, pp. 1754-1762, 2008.

زیرنویس‌ها

^۱ Artificial Fish Swarm Algorithm

۴- نتیجه‌گیری

در این مقاله یک الگوریتم همکارانه بر پایه الگوریتم دسته ماهی‌های مصنوعی و k-means ارائه گردید. در الگوریتم پیشنهادی ماهی‌های مصنوعی وظیفه انجام جستجوی سراسری را بر عهده دارد و پس از همگرایی، وظیفه جستجوی محلی بر عهده الگوریتم k-means مقداردهی اولیه می‌شود و مجددًا جستجو را در فضای مسئله کند. بدین ترتیب الگوریتم پیشنهادی از توانایی‌های هر دو الگوریتم استفاده می‌کند و ضعفهای آنها را برطرف می‌کند. نتایج آزمایشات نشان داد الگوریتم پیشنهادی از نتایج قابل قبولی برخوردار است.

مراجع

- [1] C. F. Tsai, H. C. Chang and C. W. Tsai, "A New Data Clustering Approach for Data Mining in Large Databases", in Proc. Of Parallel Architectures, Algorithms and Networks, Makati City, Metro Manila, 2002.
- [2] M. Kumar, S. Verma and P. P. Sigah, "Data Clustering in Sensor Networks Using ART", in 4th International Conference on Wireless Communication and Sensor Networks, Allahabad, India, 2008.
- [3] T. Wang and Z. Yang, "A Location-Aware-Based Data Clustering Algorithm in Wireless Sensor Networks", in 11th IEEE Singapore International Conference on Communication Systems, Guangzhou, 2008.
- [4] A. K. C. Wong and G. C. L. Li, "Simultaneous Pattern and Data Clustering for Pattern Cluster Analysis", IEEE Transactions on Knowledge and Data Engineering, Vol. 20, PP. 911-923, Los Angeles, USA, 2008.
- [5] Y. Kao and S. Y. Lee, "Combining K-means and Particle Swarm Optimization for Dynamic Data Clustering Problems", in IEEE International Conference on Intelligent Computing and Intelligent Systems, Shanghai, China, 2009.
- [6] D. Malyszko and S. T. Wierzchon, "Standard and Genetic K-means Clustering Techniques in Image Segmentation", in 6th International Conference on Computer Information Systems and Industrial Management Applicatins, Minneapolis, 2007.
- [7] B. Abolhassani and J. E. Salt, "A Simplex K-means Algorithm for Radio-Port Placement in Cellular Networks", in Canadaian Conference on Electrical and Computer Engineering, Saskatoon, 2005.
- [8] T. W. Chen and S. Y. Chien, " Bandwidth Adaptive Hardware Architecture of K-Means Clustering for Video Analysis", IEEE Transactions on VLSI Systems, pp. 1-10, 2009.