

یک الگوریتم دسته‌مایه‌های مصنوعی بهبودیافته برای خوشبندی داده‌ها

دانیال یزدانی^۱، برات سامان^۲ و محمد رضا میبدی^۳

^۱دانشگاه آزاد اسلامی، واحد شیراز، ایران، daniel.yazdani@yahoo.com

^۲دانشگاه آزاد اسلامی، واحد شیراز، ایران، barat_saman@yahoo.com

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

چکیده- 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 Kmeans می‌باشد ولی این الگوریتم دارای مشکلاتی از جمله همگرایی به مینیمم محلی و حساسیت به نقاط ابتدایی می‌باشد. الگوریتم دسته‌مایه‌های مصنوعی (AFSA) جزء الگوریتم‌های هوش جمعی می‌باشد و کاربرد عمده آن در حل مسائل بهینه‌سازی است. از جمله ویژگی‌های این الگوریتم می‌توان به سرعت همگرایی بالا و حساس نبودن به مقادیر اولیه اشاره کرد. اما الگوریتم AFSA با تعریف پارامترهای اولیه، تنها قادر است یا جستجوی سراسری را به خوبی انجام دهد یا جستجوی محلی را به درستی انجام دهد. در این مقاله یک روش خوشبندی براساس الگوریتم دسته‌مایه‌های مصنوعی بهبودیافته پیشنهاد می‌گردد. در این AFSA بهبودیافته، میان جستجوی سراسری و جستجوی محلی موازن برقرار شده است. نتایج آزمایشات بر روی مجموعه داده‌های استاندارد و دستی نشان‌دهنده صحت و کارایی روش پیشنهادی است.

کلید واژه- الگوریتم دسته‌مایه‌های مصنوعی، خوشبندی، بهینه‌سازی، Kmeans، جستجوی سراسری، جستجوی محلی.

مجموعه‌های از اشیاء را به k زیر مجموعه تقسیم می‌کند. این

- مقدمه-

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 Kmeans دارای 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 جزء مسائل NP می‌باشد. یافتن راه حل برای مسائلی که از نوع NP می‌باشند بسیار مشکل است. الگوریتم‌هایی از جمله الگوریتم‌های هوش جمعی تا حدی این مشکل را حل نموده‌اند. توسط این نوع الگوریتم‌ها راه حل‌هایی پیدا می‌شوند که تقریباً به جواب نزدیکند. یکی از معروف‌ترین و پرکاربردترین الگوریتم‌های هوش جمعی، الگوریتم PSO می‌باشد که در سال ۱۹۹۵ توسط کندی و ابرهارت^[۹] ارائه شد. این الگوریتم یک تکنیک کارا برای حل مسائل بهینه‌سازی است که بر مبنای قوانین احتمال و بر اساس جمعیت کار می‌کند. یکی دیگر از الگوریتم‌های هوش جمعی که تا کنون ارائه شده است الگوریتم دسته‌مایه‌های مصنوعی^۱ (AFSA) می‌باشد. گونه‌هایی از الگوریتم‌های برگرفته شده از طبیعت می‌باشد که

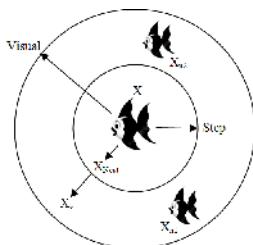
خوشبندی داده‌ها، دارای کاربردهای فراوانی در داده کاوی^[۱]، شبکه‌های حسگر بیسیم^{[۲][۳]}، تشخیص الگو^[۴] و یادگیری ماشین^[۵] می‌باشد. اهمیت خوشبندی در علوم مختلف و همچنین نوع داده‌های مورد استفاده، سرعت خوشبندی، دقت و بسیاری پارامترهای دیگر باعث معرفی روش‌ها و الگوریتم‌های متنوعی از خوشبندی داده‌ها شده است. خوشبندی یک تکنیک دسته‌بندی بدون نظارت است که در آن مجموعه‌داده‌ها که معمولاً بردارهایی در فضای چندبعدی می‌باشند، بر اساس یک معیار شباهت یا عدم شباهت، به تعداد مشخصی خوشبندی تقسیم می‌شوند. وقتی که تعداد خوشبندی K باشد و n داده^m بعدی داشته باشیم، الگوریتم خوشبندی، هر یک از این داده‌ها را به 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 Kmeans یکی از معروف‌ترین روش‌های خوشبندی است که در بسیاری از مسائل به کار رفته است^{[۶][۷][۸]}. الگوریتم Kmeans با k مرکز خوشبندی تصادفی شروع می‌شود و

بخش پنجم نتایج آزمایشات بررسی و تحلیل می‌شوند و بخش پایانی به بیان نتیجه‌گیری می‌پردازد.

۲- الگوریتم دسته‌ماهی‌های مصنوعی

در دنیای زیر آب، ماهی‌ها می‌توانند مناطقی را پیدا کنند که دارای غذای بیشتری است، که این امر با جستجوی فردی یا گروهی ماهی‌ها محقق می‌شود. مطابق با این ویژگی، مدل ماهی مصنوعی با رفتارهای جستجوی غذا، حرکت گروهی و دنباله‌روی ارائه شده است که به وسیله آنها فضای مسئله جستجو می‌شود [۱۵]. محیطی که ماهی مصنوعی در آن زندگی می‌کند اساساً فضای راه حل و حوزه‌های ماهی‌ها مصنوعی دیگر است. درجه تراکم غذا در منطقه آبی تابع هدف AFSA می‌باشد. در نهایت، ماهی‌های مصنوعی به مکانی می‌رسند که درجه تراکم و غلظت غذا در آنجا بیشترین (بهینه سراسری) باشد. همان‌طور که در شکل ۱ مشاهده می‌شود، ماهی مصنوعی مقاومت خارجی را از طریق بینایی درک می‌کند [۱۶].

وضعیت فعلی ماهی مصنوعی توسط بردار $X = (x_1, x_2, \dots, x_n)$ نشان داده می‌شود. Visual برابر میدان دید ماهی مصنوعی می‌باشد و X_v موقعیتی در میدان دید است که ماهی مصنوعی می‌خواهد به آنجا برود. حال اگر وضعیت X_v از نظر تراکم غذایی بهتر از وضعیت فعلی باشد، یک گام به جلو در جهت آن پیش می‌رویم که باعث تغییر وضعیت ماهی مصنوعی از X به X_{next} می‌شود، ولی اگر وضعیت فعلی بهتر از X_v باشد، به گشتزنی در محدوده میدان دید ادامه می‌دهیم. Step برابر حداقل طول گام حرکت است، فاصله بین دو ماهی مصنوعی که در وضعیت‌های X_i و X_j هستند با $\|X_i - X_j\|_{eu}$ نشان داده می‌شود(فاصله اقلیدوسی).



شکل ۱: ماهی مصنوعی و محیط پیرامون آن.

مدل ماهی مصنوعی شامل دو بخش متغیرها و توابع است که متغیرها شامل X (موقعیت فعلی ماهی مصنوعی)، Step (حداقل طول گام حرکت)، Visual (میدان دید)، try_number (سقف تعداد آزمایش و سعی کردن) و فاکتور شلوغی δ ($0 < \delta < 1$) می‌باشند [۱۱].

توسط دکتر Lei Xiao در سال ۲۰۰۲ ارائه شده است [۱۰]. این الگوریتم تکنیکی مبتنی بر رفتارهای جمعی است که از رفتارهای اجتماعی دسته‌ماهی‌ها 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. این الگوریتم در کاربردهای بهینه‌سازی از جمله تخمین پارامترها [۱۱، ۱۲]، بهینه‌سازی ترکیبی [۱۳]، یادگیری شبکه‌های عصبی پیشرو [۱۴] و ... به کار رفته است. در AFSA دو پارامتر مسافتی به نام‌های میدان دید و طول گام وجود دارد. ماهی‌های مصنوعی در فضای میدان دید خود به جستجوی فضای مسئله می‌پردازند و به اندازه درصدی تصادفی از طول گام به AFSA سمت هدف خود در هر تکرار حرکت می‌کنند. در AFSA استاندارد، تعیین مقدار اولیه این دو پارامتر تأثیر بسزایی در کیفیت نتیجه نهایی دارد. مقدار این دو پارامتر تا انتهای اجرای الگوریتم ثابت و برابر با همان مقدار اولیه آنها باقی می‌ماند. در صورتی که مقدار اولیه این دو پارامتر را بزرگ در نظر بگیریم دسته‌ماهی‌های مصنوعی می‌توانند با سرعت بیشتری به سمت بهینه سراسری حرکت کنند و توانایی بیشتری برای عبور از بهینه‌های محلی خواهند داشت و در صورتی که مقدار این دو پارامتر را کوچک در نظر بگیریم الگوریتم می‌تواند با دقت بیشتری جستجوی محلی را انجام دهد. بنابراین با تعیین مناسب مقدار اولیه این پارامترها، الگوریتم می‌تواند کی از جستجوهای محلی و یا جستجوی سراسری را به طور مطلوب انجام دهد ولی نمی‌تواند هر دوی آنها را در یک اجرا به خوبی انجام دهد.

در این مقاله یک روش خوشبندی بر اساس الگوریتم AFSA بهبودیافته به نام IAFSA پیشنهاد می‌گردد. در بهبود AFSA یافته میان جستجوی سراسری و جستجوی محلی موازنه برقرار شده است. الگوریتم پیشنهادی دارای سرعت همگرایی بالایی می‌باشد و توانایی بالایی در عبور از بهینه‌های محلی دارد. نتایج آزمایشات الگوریتم‌های Kmeans، PSO، AFSA استاندارد و روش پیشنهادی AFSA بهبودیافته بر روی مجموعه داده‌های استاندارد Wine، Irish و دو مجموعه داده دستی دو و سه بعدی نشان می‌دهند که الگوریتم پیشنهادی از کارایی بالاتری نسبت به دیگر الگوریتم‌های آزمایش شده برخوردار است. ادامه این مقاله بدین صورت پیکربندی شده است: در بخش دوم AFSA شرح داده می‌شود، در بخش سوم الگوریتم Kmeans تشریح می‌شود. بخش چهارم به شرح الگوریتم پیشنهادی اختصاص دارد. در

$(Y_j > Y_i)$ نشدیم، رفتار حرکت آزادانه برای ماهی مصنوعی آم اجرا می‌شود.

همچنین توابع شامل رفتار جستجوی (شکار) غذا، رفتار حرکت گروهی و رفتار دنباله‌روی (پیروی) می‌باشدند.

۳-۲- رفتار حرکت گروهی ماهی‌ها

فرض کنیم X_c موقعیت مرکزی (X_{Center}) را نشان می‌دهد که برابر مرکز ثقل بردارهای اعضای دسته است و n_f برابر تعداد همسایگان موجود در میدان دید X_c است ($d_{cx} \leq Visual$) و n برابر تعداد کل ماهی‌های مصنوعی می‌باشد. حال اگر Y_i تراکم غذا در موقعیت مرکزی باشد و Y_i تراکم غذا در وضعیت فعلی باشد، اگر $Y_i > Y_c$ باشد و $\delta = n_f/n$ باشد یعنی موقعیت مرکزی دارای تراکم غذایی بالاتری نسبت به وضعیت کنونی است و تراکم جمعیت در آن بالا نیست، پس یک گام به سمت میدان مرکزی پیش می‌رویم که با استفاده از فرمول ۴ محاسبه می‌گردد [۱۰ و ۱۷]:

$$X_i^{(t+1)} = X_i^{(t)} + \frac{X_c - X_i^{(t)}}{\|X_c - X_i^{(t)}\|} \cdot Step \cdot Rand(0,1) \quad (4)$$

حال اگر $n_f = 0$ باشد یا شرط حرکت به موقعیت مرکزی ارضانشود، رفتار جستجوی غذا اجرا می‌گردد.

۴- رفتار دنباله‌روی

در فرآیند حرکت دسته‌ای ماهی‌ها، هنگامی که یک ماهی یا تعدادی از آنها غذا پیدا کنند، همسایگان و ماهی‌های نزدیک به آنها بدنبال آنها می‌روند و به سرعت به غذا می‌رسند. اگر X_i وضعیت فعلی ماهی مصنوعی باشد، همسایه X_j را کاوش می‌کند و $Y_j < Y_i$ را می‌بیند، اگر $\delta = (n_f/n) \leq Visual$ یعنی وضعیت X_j دارای تراکم غذایی بیشتری نسبت به وضعیت فعلی است و تراکم جمعیتی در نزدیکی آن زیاد نمی‌باشد، پس یک گام به سمت X_j می‌رویم که با استفاده از فرمول ۵ محاسبه می‌گردد [۱۰]:

$$X_i^{(t+1)} = X_i^{(t)} + \frac{X_j - X_i^{(t)}}{\|X_j - X_i^{(t)}\|} \cdot Step \cdot Rand(0,1) \quad (5)$$

در غیر این صورت رفتار جستجوی غذا را اجرا می‌کند.

۵- روند اجرای AFSA

در بخش قبل چهار رفتار جستجوی غذا، حرکت جمعی، دنباله روی و حرکت آزادانه در AFSA تشریح شد. برای ماهی‌های مصنوعی، رفتارهای جستجوی غذا و حرکت آزادانه رفتارهای انفرادی هستند، رفتار دنباله روی یک رفتار گروهی است که هر یک از ماهی‌های مصنوعی به صورت محلی بر اساس

۱- رفتار حرکت آزادانه

در طبیعت، هنگامی که ماهی‌ها موفق به یافتن غذا نشوند، به صورت آزادانه حرکت می‌کنند. در AFSA هنگامی که ماهی مصنوعی نتواند به سمت موقعیتی با مقدار غذای بیشتر حرکت کند، یک گام به صورت تصادفی با استفاده از فرمول ۱ در فضای مسئله حرکت می‌کند.

$$X(t+1) = X(t) + Step \times Rand(-1,1) \quad (1)$$

که در آن X_i برابر بردار موقعیت ماهی مصنوعی آم در فضای D بعدی است.تابع $Rand$ یک بردار آبجعی از اعداد تصادفی جداگانه را در بازه $[1, -1]$ با توزیع یکنواخت تولید می‌کند.

۲- رفتار جستجوی غذا

اگر X_i وضعیت فعلی ماهی مصنوعی باشد، یک وضعیت Z_i را به طور تصادفی در میدان دید ماهی مصنوعی انتخاب می‌کنیم، $Y=f(X)$ برابر تراکم و غلظت غذا (مقدار تابع هدف) می‌باشد. وضعیت Z_i با استفاده از فرمول ۲ بدست می‌آید [۱۰]:

$$X_j = X_i + Visual \cdot Rand(-1,1) \quad (2)$$

که در آن تابع $Rand$ یک عدد تصادفی با توزیع یکنواخت در بازه $[1, -1]$ تولید می‌کند. حال تراکم غذایی در Z_i را با تراکم غذایی در وضعیت فعلی مقایسه می‌کنیم، اگر $Z_i < Y_i$ باشد (در مسائلی بیشینه‌سازی)، ماهی مصنوعی مورد نظر از وضعیت فعلی اش یک گام به جلو در جهت Z_i حرکت می‌کند که با استفاده از فرمول ۳ انجام می‌شود:

$$X_i^{(t+1)} = X_i^{(t)} + \frac{Z_i - X_i^{(t)}}{\|Z_i - X_i^{(t)}\|} \cdot Step \cdot Rand(0,1) \quad (3)$$

اما در صورتی که $Z_i < Y_i$ نباشد، یک وضعیت Z_i دیگر را با استفاده از فرمول ۱ انتخاب می‌کنیم و دوباره تراکم غذایی آن را می‌سنجیم تا بفهمیم که آیا شرط پیش‌روی را ارضاء می‌کند یا نه، اگر پس از `try_number` بار موفق به ارضاء کردن شرط پیش‌روی

```

Blackboard = arg min f(Xi)
repeat:
    for each Artificial Fish i ∈ [1 .. N]
        Perform Swarm Behavior on Xi(t) and Compute
        Xi,swarm
        Perform Follow Behavior on Xi(t) and Compute
        Xi,follow
        if f(Xi(swarm)) ≥ f(Xi(follow))
            then Xi(t+1) = Xi,follow
        else
            Xi(t+1) = Xi,swarm
        endfor
        if f(XBest-AF) ≤ f(Blackboard)
            then Blackboard = XBest-AF
    until stopping criterion is met

```

شکل ۲: شه کد الگوریتم AFSA

۲- خوشبندی و الگوریتم Kmeans

خوشبندی در فضای D بعدی اقلیدسی، فرآیندی است که در طی آن یک مجموعه N عضوی به K گروه یا خوشه بر اساس یک معیار شباهت تقسیم می‌شوند. روش‌های خوشبندی متعددی تا کنون ارائه شده است. اساس کار الگوریتم‌های خوشبندی سنجش شباهت بین داده‌ها است و با یکتابع، مشخص می‌شود که دو بردار داده چقدر به هم شبیه هستند. الگوریتم Kmeans یکی از قدیمی‌ترین و مشهورترین روش‌های خوشبندی است [۱۸]. این روش، بردارهای داده در فضای D بعدی را در خوشه‌هایی که از قبیل تعداد آنها مشخص است دسته‌بندی می‌کند، این دسته‌بندی براساس فاصله اقلیدسی بین داده‌ها و مراکز خوشه است که به عنوان معیار شباهت در نظر گرفته می‌شوند. فاصله اقلیدسی بین بردارهای داده عضو یک خوشه با مرکز آن خوشه، نسبت به فاصله اقلیدسی آنها با مراکز خوشه‌های دیگر کمتر است. الگوریتم Kmeans استاندارد به شکل زیر است.

- موقعیت اولیه K مرکز خوشه به صورت تصادفی مشخص می‌شوند.

مراحل زیر تکرار می‌شوند:

(a) برای هر بردار داده: بردار به خوشه‌ای تخصیص می‌یابد که فاصله اقلیدسی آن از مرکز آن خوشه نسبت به دیگر مراکز خوشه کمتر باشد.

فاصله تا مرکز خوشه براساس رابطه (۶) محاسبه می‌شود.

$$Dis(X_p, Z_j) = \sqrt{\sum_{i=1}^D (X_{pi} - Z_{ji})^2} \quad (6)$$

همسایگان خود آنرا اجرا می‌کند و رفتار حرکت گروهی یک رفتار گروهی است که تمامی ماهی‌های مصنوعی به صورت سراسری آنرا اجرا می‌کنند. رفتار جستجوی غذا در صورتی که یک ماهی مصنوعی با اجرای رفتار دنباله روی و یا حرکت گروهی موفق به حرکت به سمت موقعیت بهتر نشد اجرا می‌شود و در صورتی که یک ماهی مصنوعی در یافتن موقعیت بهتر با رفتار جستجوی غذا موفق نبود، رفتار حرکت آزادانه را اجرا می‌کند. در واقع رفتارهای جستجوی غذا و حرکت آزادانه رفتارهایی هستند که به طور مستقل توسط ماهی‌های مصنوعی اجرا نمی‌شوند و تنها زمانی اجرا می‌شوند که یک ماهی مصنوعی نتواند با استفاده از رفتارهای دنباله روی و حرکت جمعی به سمت یک موقعیت بهتر حرکت کند. در هر گام از فرآیند بهینه‌سازی، تمامی ماهی‌های مصنوعی به صورت موازی روند یکسانی را طی می‌کنند.

فرض کنیم موقعیت ماهی مصنوعی آم در زمان t برابر X_i(t) باشد. ماهی مصنوعی آم یکبار رفتار دنباله روی را از X_i(t) اجرا می‌کند و باعث تغییر مکان ماهی به X_{i(follow)} می‌شود. ماهی مصنوعی آم پس از انجام رفتار دنباله روی، یکبار از همان موقعیت X_{i(t)} رفتار حرکت جمعی را اجرا می‌کند که باعث حرکت ماهی مصنوعی آم به مکان X_{i(swarm)} می‌شود. پس از اجرای دو رفتار دنباله روی و حرکت جمعی که هر دو بر اساس موقعیت (t) اجرا شده اند، موقعیت بعدی ماهی مصنوعی آم یعنی X_{i(t+1)} با استفاده از رابطه ۱۶ بدست می‌آید.

$$X_i(t+1) = \begin{cases} X_i(follow) & \text{if } f(X_i(follow)) \leq f(X_i(swarm)) \\ X_i(swarm) & \text{if } f(X_i(follow)) > f(X_i(swarm)) \end{cases} \quad (16)$$

از یک تخته سیاه برای ثبت بهترین نقطه‌ای که تا AFSA کنون توسط کل اعضای دسته پیدا شده است استفاده می‌شود. AFSA، پس از اجرای رفتارهایی در هر تکرار از اجرای توسط تمام ماهی‌های مصنوعی و انتقال آنها به مکان‌های جدید، مقدار شایستگی بهترین ماهی مصنوعی با مقدار ثبت شده بر روی تخته سیاه مقایسه می‌شود. در صورتی که مقدار شایستگی بهترین ماهی مصنوعی دسته بهتر از مقدار شایستگی نقطه ثبت شده بر روی تخته سیاه بود، موقعیت بهترین ماهی دسته بر روی تخته سیاه به عنوان بهترین نقطه یافت شده تا در شکل ۲ نشان داده شده کنون ثبت می‌شود. شبه کد است.

```

for each Artificial Fish i ∈ [1 .. N]
    initialize xi
endfor

```

پارامترهای میدان دید و طول گام را بزرگ در نظر می‌گیریم تا توانایی جستجوی سراسری الگوریتم بالا باشد. بدین ترتیب دسته با سرعت بیشتری به سمت بهینه سراسری حرکت می‌کند و از بهینه‌های محلی با توانایی بیشتری عبور می‌کند. سپس همزمان با حرکت ماهی‌های مصنوعی به سمت بهینه سراسری، از مقدار این دو پارامتر کاسته می‌شود تا هنگامی که ماهی‌های مصنوعی به نزدیکی بهینه سراسری رسیدند، بتوانند محدوده آن را با دقت بیشتری برای رسیدن به نتایج بهتر کاوش کنند. در این مقاله یک روش خود تطبیقی برای کاهش این دو پارامتر در طول اجرای الگوریتم پیشنهاد می‌گردد. در این روش مقدار این دو پارامتر به صورت پویا و بر اساس موقعیت دسته در فضای مسئله تغییر می‌کند. در این روش برای تنظیم پارامترهای میدان دید و طول گام از نسبت مقدار شایستگی بهترین ماهی‌های مصنوعی به مقدار متوسط شایستگی کل اعضای گروه استفاده می‌شود. برای بدست آوردن مقدار این پارامترها در هر تکرار از فرمول پیشنهادی ۸ و ۹ استفاده می‌شود:

$$Visual_{itr} = Visual_{itr-1} - \left(\frac{BestAFFitValue}{SwarmFitValueAvr} \times L \times Visual_{itr-1} \right) \quad (8)$$

$$Step_{itr} = Step_{itr-1} - \left(\frac{BestAFFitValue}{SwarmFitValueAvr} \times L \times Step_{itr-1} \right) \quad (9)$$

در روابط ۸ و ۹، مقدار پارامترهای میدان دید و طول گام در تکرار itr به اندازه درصدی از مقدار این پارامترها در تکرار قبل از آن بدست می‌آید. در روابط ۸ و ۹ نسبت مقدار شایستگی بهترین ماهی‌های مصنوعی (BestAF.FitValue) به متوسط مقدار شایستگی کل اعضای گروه (Swarm.FitValueAvr) عددی در بازه صفر و یک است. هرچه این مقدار بزرگتر باشد نشان می‌دهد که اعضای گروه در نزدیکی یکدیگرند و گروه متراکم می‌باشد بنابراین در این حالت اعضای گروه در نزدیکی بهینه سراسری جمع شده‌اند پس برای بالا بردن توانایی جستجوی محلی آنها باید مقدار پارامترهای میدان دید و طول گام با شدت بیشتری کاهش یابند و هر چه این مقدار این نسبت کوچکتر باشد، نشان می‌دهد که اعضای گروه دور از یکدیگر هستند و تراکم گروه در فضای مسئله پایین است و ماهی‌های مصنوعی در حال حرکت به سمت بهینه سراسری هستند بنابراین نباید مقدار پارامترهای میدان دید و طول گام را کاهش داد زیرا با کاهش مقدار این دو

در رابطه (۶)، X_p مشخص کننده p امین بردار داده است Z_j مشخص کننده مرکز خوشة زام است و D تعداد خصوصیات (ابعاد) داده‌ها و مراکز خوشه‌ها می‌باشد.

(b) مراکز خوشه‌ها با استفاده از رابطه (۷) بروزرسانی می‌شوند.

$$Z_j = \frac{1}{n_j} \left[\sum_{\forall X_p \in C_j} X_p \right] \quad (7)$$

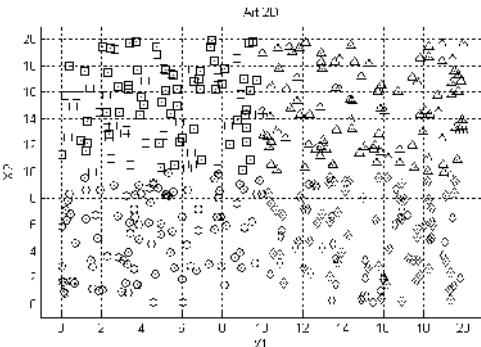
در رابطه (۷) n_j تعداد بردارهای داده مربوط به خوشه زام است و C_j زیرمجموعه‌ای از کل بردارهای داده است که خوشه زام را تشکیل می‌دهند و عضو آن هستند. مراحل a و b تا هنگامی که شرط توقف برقرار گردد تکرار می‌شوند [۱۸].

۴- مدل پیشنهادی

در این بخش یک الگوریتم خوشبندی بر پایه AFSA ارائه می‌شود. در AFSA ماهی‌های مصنوعی در فضای میدان دید خود با اجرای رفتارهای AFSA به جستجوی فضای مسئله می‌پردازند و به اندازه درصدی تصادفی از طول گام به سمت هدف خود در هر تکرار حرکت می‌کنند. در AFSA استاندارد، تعیین مقدار اولیه این دو پارامتر تأثیر بسزایی در کیفیت نتیجه نهایی دارد. مقدار این دو پارامتر تا انتهای اجرای الگوریتم ثابت و برابر با همان مقدار اولیه آنها باقی می‌ماند. در صورتی که مقدار اولیه این دو پارامتر را بزرگ در نظر بگیریم دسته ماهی‌های مصنوعی می‌توانند با سرعت بیشتری به سمت بهینه سراسری حرکت کنند زیرا ماهی‌های مصنوعی می‌توانند فضای بزرگتری را در اطراف خود کاوش کنند و با طول گام بزرگتری در هر تکرار حرکت کنند. در این شرایط دسته ماهی‌های مصنوعی توانایی بیشتری برای عبور از بهینه‌های محلی خواهد داشت. اما در این حالت الگوریتم جستجوی سراسری را به خوبی انجام می‌دهد و از انجام یک جستجوی محلی قابل قبول عاجز است. در صورتی که مقدار این دو پارامتر را کوچک در نظر بگیریم الگوریتم می‌تواند با پایداری و دقت بیشتری جستجوی محلی را انجام دهد ولی در این حالت ماهی‌ها با سرعت کمتری به سمت هدف حرکت می‌کنند و دسته آنها جمع‌تر و فشرده‌تر می‌شود که در نتیجه توانایی گذشتن از بهینه‌های محلی آنها کاهش می‌یابد. بنابراین لازم است برای بهبود نتایج، میان فرآیندهای جستجوی سراسری و جستجوی محلی موازن برقار شود تا الگوریتم بتواند هر دوی آنها را به طور قابل قبول انجام دهد. برای این منظور ابتدا مقدار

آزمایشات بر روی چهار مجموعه داده انجام شده است که از میان آنها، دو مجموعه داده واقعی شامل Iris و Wine هستند که از مجموعه داده های استاندارد UCI انتخاب شده اند و دو مجموعه داده به صورت دستی تهیه شده اند که در ادامه خصوصیات هر یک از آنها شرح داده شده است:

- 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 دسته بندی شده اند. در این مجموعه داده هر نمونه دارای ۱۳ ویژگی می باشد.
- ۳- Art 2D: در این مجموعه داده دستی، ۴ کلاس وجود دارد و هر کلاس شامل ۱۰۰ نمونه می باشد که هر نمونه دارای ۲ ویژگی است. نحوه چینش نمونه های این مجموعه داده در فضای دوبعدی با توزیع یکنواخت انجام شده است. نمونه های این مجموعه داده در شکل ۳ نشان داده شده است. همانطور که مشاهده می شود، در این مجموعه داده، نمونه های مشترک مرزی وجود دارد.



شکل ۳: مجموعه داده Art 2D

- ۴- Art 3D: در این مجموعه داده دستی، ۵ کلاس وجود دارد و هر کلاس شامل ۵۰ نمونه می باشد که هر نمونه دارای ۳ ویژگی است. نحوه توزیع ویژگی های مختلف نمونه های هر کلاس به صورت Class1~Uniform(25,40)، Class2~Uniform(40,55)، Class3~Uniform(55,70) و Class4~Uniform(70,85)، Class5~Uniform(85,100) می باشد. این مجموعه داده در شکل ۴ نشان داده شده است.

پارامتر توانایی جستجوی سراسری و گذر از بهینه های محلی کاهش می یابد.

ثابت L برابر با حداکثر درصدی از میدان دید یا طول گام است که در هر تکرار می تواند نسبت به تکرار قبل کاهش یابد. بدین ترتیب بر اساس ترکیب و قرار گیری اعضای گروه در فضای مسئله، مقدار پارامترهای میدان دید و طول گام در هر تکرار به طور پویا کاهش می یابند.

حال به شرح و بیان طرز کار الگوریتم پیشنهادی می پردازیم. در این روش، برای یافتن مقادیر بهینه مراکز خوش (که تعداد آنها از پیش تعیین شده است)، از رابطه 10×10 استفاده می شود. در واقع رابطه 10×10 ،تابع شایستگی مورد نظر می باشد که قرار است بهینه سازی (کمینه سازی) شود و فاصله های درون خوش (Intra cluster distances) را محاسبه می کند [۱۸].

$$J(C_1, C_2, \dots, C_K) = \sum_{i=1}^K \left(\sum_{X_j \in C_i} \|Z_i - X_j\| \right) \quad (10)$$

در رابطه 10×10 ، مجموع فاصله اقلیدوسی تمام بردارهای داده از مرکز خوش ای که عضو آن هستند محاسبه و جمع می شود. در این رابطه K خوش (C) داریم که هر یک از N بردار داده (X) بر اساس فاصله ای که از هر یک از این مراکز خوش (Z) دارند، دسته بندی می شوند و در یکی از خوشها قرار می گیرند. بنابراین هدف، تعیین مراکز خوش ای است که رابطه 10×10 را کمینه کنند، بدین ترتیب مراکز خوش بهینه تعیین می شوند. از آنجایی که داده ها D بعدی بوده و تعداد K خوش وجود دارد، بنابراین هر ماهی مصنوعی دارای $K \times D$ بعد خواهد بود. شکل ۲، بردار یک ماهی مصنوعی را نشان می دهد که شامل K مرکز خوش بُعدی می باشد.

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

شکل ۲: ساختار موقعیت یک ماهی مصنوعی در فضای مسئله خوش بندی.

در الگوریتم پیشنهادی ابتدا ماهی های مصنوعی گروه را به صورت تصادفی مقداردهی اولیه می کنیم، بدین ترتیب هر یک از ماهی های مصنوعی شامل K مرکز خوش اولیه تصادفی می باشند. سپس با توجه به اختصاص بردارهای داده به هر یک از خوشها در ماهی های مصنوعی بر اساس فاصله اقلیدوسی و تابع شایستگی رابطه 10×10 ، رفتارهای الگوریتم دسته ماهی های مصنوعی را برای ماهی های مصنوعی اجرا می کنیم.

۵- نتایج آزمایشات

در نهایت به نتایج بهتری دست پیدا می‌کند. علت بهتر بودن نتایج حاصل از الگوریتم پیشنهادی در برابر دیگر الگوریتم‌های آزمایش شده برقراری توازن و تعادل میان جستجوی سراسری و جستجوی محلی به شکل مطلوب است.

جدول ۱: بهترین، متوسط و انحراف معیار فاصله درون خوشه‌ای و نرخ خطای حاصل از ۳۰ بار اجرای الگوریتم‌های AFSA، Kmeans و PSO و

روش پیشنهادی IAFSA بر روی مجموعه داده Iris.

الگوریتم	معیارها	فاصله درون خوشه‌ای	نرخ خطای
بهترین		۹۷,۳۲۵۹	۱۰,۶۶۶
Kmeans	متوسط	۱۰,۵۸۶	۱۶,۰۵۳
انحراف معیار		۱۱,۳۳۷۸	۱۰,۹۹۸
بهترین		۹۶,۹۱۵۸	۱۰,۳۳۳
StdAFSA	متوسط	۱۱۲,۲۲۱۴	۲۷,۸۸۸
انحراف معیار		۵,۴۶۳۹	۹,۰۰۹۲
بهترین		۹۷,۱۰۴۴	۱۰
PSO	متوسط	۱۰,۲۶۱۷	۱۰,۵۴۴
انحراف معیار		۵,۸۱۲۳	۴,۵۰۲۸
بهترین		۹۶,۶۶۲۳	۱۰
IAFSA	متوسط	۹۶,۶۸۸۹	۱۰,۱۵۵
انحراف معیار		۰,۰۲۸۷	۰,۲۸۶

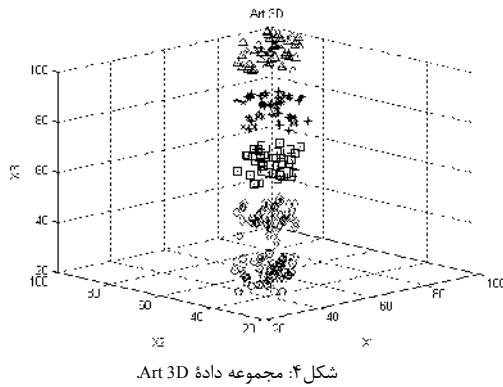
جدول ۲: بهترین، متوسط و انحراف معیار فاصله درون خوشه‌ای و نرخ خطای حاصل از ۳۰ بار اجرای الگوریتم‌های AFSA، Kmeans، PSO و روش

پیشنهادی IAFSA بر روی مجموعه داده Wine.

الگوریتم	معیارها	فاصله درون خوشه‌ای	نرخ خطای
بهترین		۱۶۵۵۵,۶۷۹۴	۲۹,۷۷۵۲
Kmeans	متوسط	۱۷۶۶۲,۷۲۸۳	۳۴,۳۸۲
انحراف معیار		۱۸۷۸۰,۶۷۶۹	۶,۰۸۳۷
بهترین		۱۶۶۹۵,۴۷۱۸	۲۸,۱۰۹۹
StdAFSA	متوسط	۱۶۸۴۵,۳۹۳۳	۲۹,۱۳۸۶
انحراف معیار		۵۵,۷۱۷۰	۰,۴۳۶
بهترین		۱۶۳,۷,۱۶۲۲	۲۸,۰۸۹۸
PSO	متوسط	۱۶۳۲۰,۵۶۷۲	۲۸,۷۴۵۳
انحراف معیار		۹,۵۲۷۶	۰,۳۹۲۶
بهترین		۱۶۳۰,۷,۱۰۵۹	۲۸,۰۸۹۸
IAFSA	متوسط	۱۶۳۱۰,۳۴۳۹	۲۸,۴۰۸۲
انحراف معیار		۱,۵۵۹۶	۰,۲۸۳۱

مطابق با جدول ۳ و شکل ۳، در مجموعه داده Art 2D، بدليل وجود نمونه‌های مرزی متعدد، مقداری خطای وجود دارد ولی مطابق با جدول ۴ و شکل ۴، در مجموعه داده Art 3D، نمونه‌های مختلف موجود در خوشه‌های مختلف، از یکدیگر تفکیک شده‌اند، بنابراین همانطور که در جدول ۴ مشاهده می‌شود، تمامی الگوریتم‌های آزمایش شده در طی ۳۰ بار اجرا، حداقل یک بار به خوشه‌بندی بدون خطای دست یافته‌اند اما الگوریتم پیشنهادی IAFSA، در تمامی اجرای خود موفق شده است مجموعه داده Art 3D را بدون خطای خوشه‌بندی کند. در مجموع نتایج

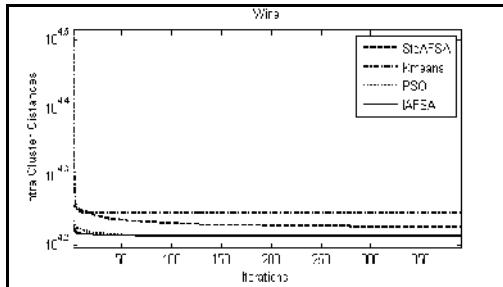
در آزمایشات انجام شده، جمعیت در AFSA استاندارد، الگوریتم پیشنهادی IAFSA و PSO، مساوی ۵ برابر تعداد ابعاد فضای مسئله است. ابعاد فضای مسئله برای هر یک از مجموعه داده‌ها برابر حاصل ضرب تعداد کلاس‌های آن مجموعه داده در تعداد ویژگی‌های نمونه‌های آن مجموعه داده است. حداکثر تکرار برای خوشه‌بندی هر یک از مجموعه داده‌ها مساوی ۱۰ برابر تعداد ابعاد فضای مسئله در نظر گرفته شده [۱۸]. در AFSA استاندارد و IAFSA، مقدار فاکتور شلوغی برابر ۰/۵، حداکثر تعداد تلاش‌ها برابر ۱۰، 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 PSO مقدار c1 و c2 برابر ۲ در نظر گرفته شده است و وزن اینرسی از رابطه $W = rand/2 + 0.5$ بدست می‌آید [۱۹]. تابع شایستگی برابر فاصله درون خوشه‌ای است و با رابطه $10 \times \text{محاسبه می‌شود}$. آزمایش ۳۰ بار تکرار شده‌اند و بهترین، متوسط و انحراف معیار نتایج بهینه‌سازی فاصله درون خوشه‌ای و خطای خوشه‌بندی با الگوریتم‌های AFSA استاندارد، PSO، Kmeans و الگوریتم پیشنهادی IAFSA در جداول ۱ تا ۴ بر روی مجموعه داده‌های نامبرده نشان داده شده است. در شکل‌های ۵ تا ۸، نمودار مقدار متوسط تابع شایستگی (فاصله درون خوشه‌ای) در طول اجرای الگوریتم‌ها در ۳۰ بار اجرا نشان داده شده است.



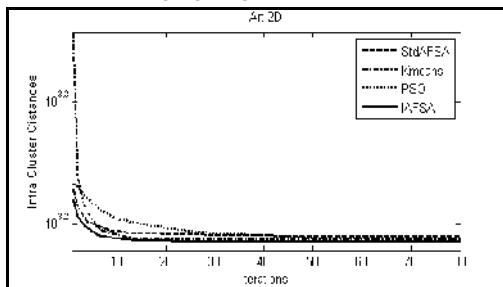
شکل ۴: مجموعه داده Art 3D.

همانطور که در شکل‌های ۵ تا ۸ نشان داده شده است، الگوریتم Kmeans دارای سرعت همگرایی بالاتری نسبت دیگر الگوریتم‌های آزمایش شده بر اساس تعداد تکرار است اما این الگوریتم دچار همگرایی زودرس می‌شود و در بهینه‌های محلی گیر می‌کند. همانطور که در اشکال ۵ تا ۸ دیده می‌شود، الگوریتم پیشنهادی IAFSA دارای سرعت همگرایی بالاتری نسبت به AFSA استاندارد و PSO بر اساس تعداد تکرار است و بهتر از دیگر الگوریتم‌های آزمایش شده از بهینه‌های محلی عبور می‌کند و

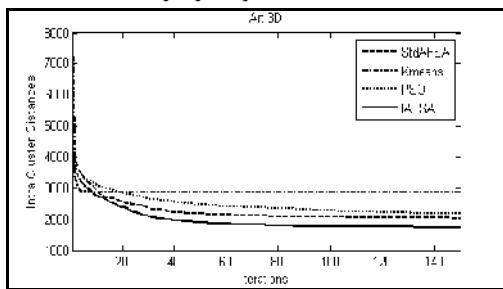
شکل ۶: نمودار متوسط فاصله درون خوشه‌ای در طول اجرای الگوریتم‌های استاندارد، *PSO* و روش پیشنهادی *IAFSA*، *Kmeans* داده Wine در ۳۰ بار اجرا.



شکل ۷: نمودار متوسط فاصله درون خوشه‌ای در طول اجرای الگوریتم‌های استاندارد، *PSO* و روش پیشنهادی *IAFSA*، *Kmeans* داده Art 2D در ۳۰ بار اجرا.



شکل ۸: نمودار متوسط فاصله درون خوشه‌ای در طول اجرای الگوریتم‌های استاندارد، *PSO* و روش پیشنهادی *IAFSA*، *Kmeans* داده Art 3D در ۳۰ بار اجرا.



انحراف معیار مربوط به فاصله درون خوشه‌ای که در جداول ۱ تا ۴ نشان داده شده است نشان می‌دهند که همگرایی به سمت بهینه سراسری در الگوریتم پیشنهادی به خوبی تضمین شده است.

۶- نتیجه‌گیری

در این مقاله یک مدل جدید برای خوشه‌بندی داده‌ها بر پایه الگوریتم دسته‌های ماهی‌های مصنوعی پیشنهاد گردید. در مدل پیشنهادی، میان جستجوی سراسری و جستجوی محلی *AFSA* به صورت خودطبیقی توازن برقرار شده و ساختار ماهی‌های مصنوعی در *AFSA* برای عمل خوشه‌بندی پیکربندی شده است. نتایج آزمایشات برای بهینه‌سازیتابع شایستگی مربوط به فاصله درون خوشه‌ای نشان داد که الگوریتم پیشنهادی به خوبی از بهینه‌های محلی عبور می‌کند و به

آزمایشات نشان می‌دهند الگوریتم پیشنهادی دارای کارایی بهتری در مقایسه با *AFSA* استاندارد و دیگر الگوریتم‌های آزمایش شده است.

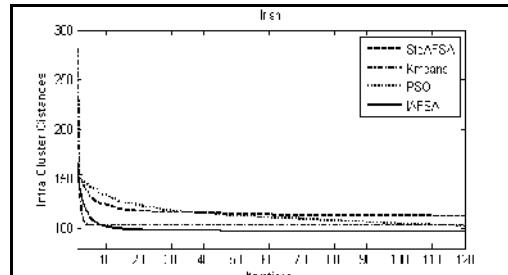
جدول ۳: بهترین، متوسط و انحراف معیار فاصله درون خوشه‌ای و نرخ خطای حاصل از ۳۰ بار اجرای الگوریتم‌های *AFSA*، *Kmeans* استاندارد، *PSO* و روش پیشنهادی *IAFSA* بر روی مجموعه داده Iris 2D

الگوریتم	معیارها	فاصله درون خوشه‌ای	نرخ خطای
بهترین		۱۵۲۵,۶۵۶۹	
متوسط		۱۵۴۰,۲۱۸۲	
انحراف معیار		۲۶,۸۷۲۱	
بهترین		۱۵۴۱,۵۹۷۶	
متوسط		۱۵۴۸,۶۵۳۰	
انحراف معیار		۳,۶۵۰۳	
بهترین		۱۵۲۵,۳۰۱۹	
متوسط		۱۵۳۹,۶۹۶۴	
انحراف معیار		۴,۲۲۹۰	
بهترین		۱۵۳۴,۸۶۷۴	
متوسط		۱۵۳۵,۱۵۵۹	
انحراف معیار		۰,۵۷۲۷	

جدول ۴: بهترین، متوسط و انحراف معیار فاصله درون خوشه‌ای و نرخ خطای حاصل از ۳۰ بار اجرای الگوریتم‌های *AFSA*، *Kmeans* استاندارد، *PSO* و روش پیشنهادی *IAFSA* بر روی مجموعه داده Iris 3D

الگوریتم	معیارها	فاصله درون خوشه‌ای	نرخ خطای
بهترین		۱۷۴۲,۷۰۴۶	
متوسط		۲۸۵۵,۲۴۷۸	
انحراف معیار		۶۰,۹۶۹,۰۹	
بهترین		۱۹۴۴,۶۳۱۲	
متوسط		۲۰۳۲,۴۱۵۵	
انحراف معیار		۵۵,۹۹,۰۹	
بهترین		۱۸۰,۳,۴۳۷۴	
متوسط		۲۱۶۹,۶۳۳۴	
انحراف معیار		۳۰,۰,۴۹۸	
بهترین		۱۷۴۱,۱۸۸۶	
متوسط		۱۷۴۲,۸۸۴۹	
انحراف معیار		۱,۰۳۰	

شکل ۵: نمودار متوسط فاصله درون خوشه‌ای در طول اجرای الگوریتم‌های *AFSA*، *Kmeans* استاندارد، *PSO* و روش پیشنهادی *IAFSA* بر روی مجموعه داده Iris در ۳۰ بار اجرا.



- In Proc. of New Weaponry Technology & Application, 2006.
- [14] 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.
- [15] J. M. Xiao, X. M. Zheng and X. H. Wang, "A Modified Artificial Fish Swarm Algorithm", In Proc. of the IEEE 6th World Congress on Intelligent Control and Automation, Dalian, 2006.
- [16] M. Jiang, Y. Wang, F. Rubio and D. Yuan, "Spread Spectrum Code Estimation by Artificial Fish Swarm Algorithm", In Proc. IEEE International Symposium on Intelligent Signal Processing (WISP), 2007.
- [17] M. Y. Jiang and D. F. Yuan, "Artificial Fish Swarm Algorithm and Its Applications", In Proc. of the International Conference on Sensing, Computing and Automation (ICSCA'2006), 2006.
- [18] Y. T. Kao, E. Zahara and I. W. Kao, "A Hybridized Approach to Data Clustering", Journal of Elsevier on Expert Systems with Applications, Vol. 34, pp. 1754-1762, 2008.
- [19] C. Y. Chen and F. Ye, "Particle Swarm Optimization Algorithm and its Application to Clustering Analysis", In Proc. of the IEEE International Conference on Networking, Sensing and Control, Taipei, Taiwan, 2004.

¹ Artificial Fish Swarm Algorithm

سمت بهینه سراسری همگرا می شود و در نهایت با افزایش توانایی جستجوی محلی، الگوریتم محدوده بهینه سراسری را برای یافتن نتایج بهتر با دقت بیشتری کاوش می کند. در مجموع نتایج آزمایشات نشان داد که الگوریتم پیشنهادشده دارای کارایی بهتری نسبت به Kmeans و PSO 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.
- [9] J. Kennedy, R. C. Eberhart, "Particle Swarm Optimization", in IEEE International Conference on Neural Network, pp. 1942-1948, Perth, 1995.
- [10] 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.
- [11] M. Jiang, Y. Wang, S. Pfletschinger, M. A. Lagunas and D. Yuan , "Optimal Multiuser Detection with Artificial Fish Swarm Algorithm", In Proc. of International Conference on Intelligent Computing (ICIC), 2007.
- [12] G. Cui, X. Cao, J. Zhou and Y. Wang, "The Optimization of DNA Encoding Sequences Based on Improved AFS Algorithms", In Proc. of IEEE International Conference on Automation and Logistics, Jinan, 2007.
- [13] H. Chen, S. Wang, J. Li and Y. Li, "A Hybrid of Artificial Fish Swarm Algorithm and Particle Swarm Optimization for Feed-forward Neural Network Training",