

یک روش ترکیبی برای خوشه بندی داده ها (کلونی مورچه ها + اتوماتای یادگیر سلولی)

سیمین قدیری گرجان^۱؛ دکتر محمدرضا میبدی^۲

چکیده

خوشه بندی داده ها یکی از مهمترین مراحل بازشناخت الگو می باشد که تاکنون در دامنه وسیعی از مسایل از جمله آنالیز داده ها، داده کاوی و قسمت بندی تصاویر بکار گرفته شده است. در سالهای اخیر کارهای تحقیقاتی متعددی در زمینه خوشه بندی داده ها مبتنی بر سیستم کلونی مورچه ها انجام شده است. در برخی از روش های ارایه شده مورچه ها به عنوان آیتم های داده در سطح یک شبکه سلولی بصورت تصادفی حرکت می کنند تا زمانی که مورچه های نماینده آیتم های مشابه در کنار یکدیگر قرار گیرند. در این مقاله یک روش ترکیبی خوشه بندی داده ها مبتنی بر سیستم کلونی مورچه ها و اتوماتای یادگیر سلولی ارایه شده است. در روش پیشنهادی مورچه های پراکنده در سطح شبکه حرکت نموده و با استفاده از رد پای فرمون، مورچه های مشابه را می یابند و اتوماتاهای یادگیر، حرکت آنها را با مورچه های مشابه موجود در همسایگی هم جهت می نمایند. روش پیشنهادی با الگوریتم های ASM و LF مقایسه شده است و نتایج حاصله حاکی از سرعت و کارایی بیشتر آن نسبت به الگوریتم های یاد شده می باشد.

کلمات کلیدی

اتوماتاهای یادگیر سلولی، خوشه بندی، سیستم کلونی مورچه ها

A Data Clustering Algorithm based on Ant Colony System and Cellular Learning Automata

Simin Ghadiri Garjan; Mohammad Reza Meybodi

ABSTRACT

Clustering is one of the most important stages of pattern recognition, that is used in variety of problems such as data analysis, data mining, and image segmentation. In recent years, several researches have been performed on Ant_based clustering. In some approaches ants move in the space as data objects until similar ants gather in one location. In this paper we propose a hybrid clustering algorithm based on cellular learning

^۱ دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه امیرکبیر، تهران، ایران، mmeybodi@aut.ac.ir

^۲ دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه امیرکبیر، تهران، ایران، mmeybodi@aut.ac.ir

automata and ant colony system. In the proposed approach, ants move on the cellular space by using pheromone trails to find similar ants, and the learning automata of each ant, coordinates its movements with adjacent similar ants. The proposed data clustering algorithm is tested on several standard data sets and then the results are compared with the results obtained for ASM and M-CLA algorithms. The results of comparisons show the superiority of the proposed clustering algorithm over these algorithms.

KEYWORDS

ant colony system, clustering, cellular learning automata

۱. مقدمه

به عمل گروه‌بندی بر روی یک مجموعه داده چند بعدی، به نحوی که داده‌های مشابه در کنار هم در قالب چند خوشه، قرار گیرند، خوشه‌بندی^۱ گفته می‌شود [۱۲]. تا کنون روش‌های متعددی برای خوشه‌بندی داده‌ها ارائه شده است. یک دسته از انواع روش‌های موجود برای خوشه‌بندی، روش‌های مبتنی بر کلونی مورچه‌ها می‌باشد. در اغلب روش‌های خوشه‌بندی توسط کلونی مورچه‌ها از قالب یک شبکه سلولی برای جایی داده‌ها استفاده شده است. در برخی روش‌های ارائه شده مورچه‌ها داده‌های پراکنده در محیط را جابه‌جا می‌کنند تا تمامی داده‌های مشابه در کنار یکدیگر قرار گیرند و در برخی دیگر، مورچه‌ها خود به عنوان آیت‌های داده در سطح محیط حرکت می‌کنند تا در جای مناسب مستقر شوند. روش‌های نوع دوم کارایی بیشتری را از نظر زمان و پیچیدگی محاسباتی نسبت به روش‌های نوع اول دارند. از این دسته می‌توان به الگوریتم‌های ASM که ترکیب اتوماتای سلولی و مدل خواب مورچگان می‌باشد [۷] و Cellular Ant که ترکیب خوشه‌بندی مورچگان بر اساس فرومون و اتوماتای سلولی می‌باشد [۱۰]، اشاره نمود.

یک اتوماتای یادگیر سلولی یک اتوماتای سلولی است که هر سلول آن به یک یا چند اتوماتای یادگیر^۲ (LA) که وظیفه تعیین حالت آن سلول را به عهده دارند، تجهیز شده است [۲]. در اتوماتای یادگیر سلولی هر سلول بر اساس اتوماتای یادگیر موجود در آن عملی را انتخاب نموده و بر پایه اعمال انتخابی همسایگان سلول و قانون محلی حاکم بر همسایگی، عمل انتخابی سلول پاداش و یا جریمه دریافت می‌نماید. در [۱۱] مدل جدیدی از اتوماتای یادگیر سلولی با نام اتوماتای یادگیر سلولی متحرک معرفی گردیده است که در آن اتوماتاهای یادگیر در محیط اتوماتای سلولی، توانایی حرکت دارند.

در [۱۱] یک الگوریتم خوشه‌بندی مبتنی بر اتوماتای یادگیر سلولی متحرک ارائه شده است که در آن، به ازای هر داده یک اتوماتای یادگیر که در حکم یک عامل می‌باشد، در نظر گرفته شده است. از تعامل هر اتوماتای یادگیر با همسایگانش، اتوماتاهای یادگیری که شباهت بیشتری داشته باشند، در یک راستای خاص حرکت نموده که با افزایش تعداد تکرار، هر اتوماتای یادگیر جهت مناسب حرکت خود که در راستای همسایگان مشابه می‌باشد را یاد می‌گیرد.

در روش ارائه شده در [۱۱] عامل‌ها تا زمان یافتن همسایه‌های مشابه در سطح شبکه به صورت تصادفی حرکت می‌کنند که این باعث کندی الگوریتم می‌شود. همچنین بعلاوه وجود قانونی برای از حرکت باز ایستادن عامل‌ها، حرکت عامل‌ها بصورت پراکنده ادامه می‌یابد و مانع از ایجاد خوشه‌های ثابت از داده‌های مشابه در کنار یکدیگر در شبکه می‌شود.

در این مقاله یک روش خوشه‌بندی ترکیبی مبتنی بر سیستم کلونی مورچه‌ها و اتوماتای یادگیر سلولی ارائه شده است. در روش پیشنهادی مورچه‌های پراکنده در سطح شبکه با استفاده از ردپای فرومون، مورچه‌های مشابه را می‌یابند و اتوماتاهای یادگیر، حرکت آنها را با مورچه‌های مشابه موجود در همسایگی هم جهت می‌نمایند. استفاده از فرومون باعث می‌شود که مورچه‌ها با استفاده از اطلاعات سایر مورچه‌های عبور کننده از هر سلول، با سرعت بیشتری در جهت صحیح در سطح شبکه حرکت نمایند. همچنین با استفاده از شبکه سلولی بدون مرز دوری و ایجاد قوانینی برای از حرکت باز ایستادن مورچه‌ها، پس از تعدادی تکرار، الگوریتم به ایجاد خوشه‌های ثابت و مجزا منجر می‌شود. علاوه بر آن در این روش نرخ یادگیری اتوماتای یادگیر سلولی بصورت متغیر و براساس تعداد مورچه‌های مشابه موجود در همسایگی، الگوریتم در مراحل پایانی یادگیری که تعداد مورچه‌های مشابه همسایه بتدریج بیشتر می‌شود، با سرعت بیشتری به سمت پاسخ بهینه همگرا می‌گردد.

ادامه این گزارش به صورت زیر سازماندهی شده است. در بخش دوم اتوماتای یادگیر و اتوماتای یادگیر سلولی معرفی شده است، در بخش سوم مروری بر روش‌های خوشه‌بندی مبتنی بر سیستم کلونی مورچه‌ها و اتوماتای یادگیر سلولی و در بخش چهارم الگوریتم ترکیبی پیشنهادی ارائه شده‌اند. در بخش پنجم روش ارائه شده مورد ارزیابی قرار گرفته است و بخش پایانی نتیجه‌گیری می‌باشد.

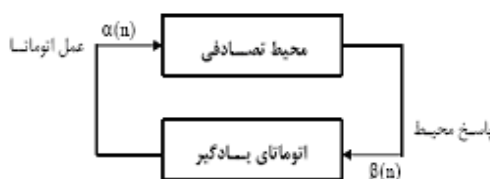
۲. اتوماتای یادگیر و اتوماتای یادگیر سلولی

در این بخش شرح مختصری درباره اتوماتاهای یادگیر و اتوماتای یادگیر سلولی ارائه می گردد.

۱.۲. اتوماتای یادگیر

اتوماتای یادگیر [۴] یک ماشین با حالات محدود است که می تواند تعداد محدودی عمل را انجام دهد. هر عمل انتخاب شده، توسط محیطی احتمالی ارزیابی می گردد و پاسخی به اتوماتای یادگیر داده می شود. اتوماتای یادگیر از این پاسخ استفاده نموده و عمل بعدی را انتخاب می کند. در طی این فرایند، اتوماتای یادگیر، یاد می گیرد که چگونه بهترین عمل را انتخاب نماید. بهترین عمل، عملی است که احتمال دریافت پاداش از محیط را به حداکثر برساند.

کارکرد اتوماتای یادگیر در تعامل با محیط، در شکل (۱) مشاهده می شود.



شکل ۱ - اتوماتای یادگیر و محیط

محیط را می توان توسط سه تایی $E \equiv \{\alpha, \beta, c\}$ نشان داد که در آن $\alpha \equiv \{\alpha_1, \alpha_2, \dots, \alpha_r\}$ مجموعه ورودی های محیط، $\beta \equiv \{\beta_1, \beta_2, \dots, \beta_r\}$ مجموعه خروجی های محیط و $c \equiv \{c_1, c_2, \dots, c_r\}$ مجموعه احتمال های جریمه می باشند. ورودی محیط یکی از r عمل انتخاب شده توسط اتوماتای یادگیر است. خروجی (پاسخ) محیط برای هر عمل i توسط β_i مشخص می شود. اگر β_i یک پاسخ دودویی باشد، محیط نوع P نامیده می شود. در چنین محیطی $\beta_i(n) = 1$ بعنوان پاسخ نامطلوب یا شکست و $\beta_i(n) = 0$ بعنوان پاسخ مطلوب یا موفقیت در نظر گرفته می شوند. در محیط نوع Q ، $\beta_i(n)$ شامل تعداد محدودی از مقادیر موجود در بازه $[0, 1]$ می باشد. در محیط نوع S مقادیر $\beta_i(n)$ یک متغیر تصادفی در بازه $[0, 1]$ می باشد. مجموعه C احتمالات جریمه (شکست) پاسخ های محیط را مشخص می کند.

الگوریتم بیان شده در روابط (۱) و (۲) بیانگر نمونه ای از الگوریتم خطی یادگیری در محیط از نوع P می باشد.

$$p_i(n+1) = p_i(n) - \beta(n)b.p_i(n) + (1 - \beta(n))a.[1 - p_i(n)] \quad (1)$$

$$p_j(n+1) = p_j(n) + \beta(n).(\frac{b}{r-1} - b.p_j(n)) - (1 - \beta(n)).a.(1 - p_j(n)) \quad \forall j, j \neq i \quad (2)$$

در روابط فوق r تعداد اقدام های اتوماتا، $\beta(n)$ سیگنال تقویتی تولی محیط در بازه $[0, 1]$ ، پارامتر پاداش و b پارامتر جریمه می باشد. در محیط از نوع S هر عمل بر مبنای میزان $\beta(n)$ ، توأمان پاداش و جریمه دریافت می نماید.

۲.۲. اتوماتای یادگیر سلولی

اتوماتای یادگیر سلولی مجموعه ای متشکل از اجزاء ساده بوده که رفتار هر جزء بر اساس تجربیات گذشته و رفتار همسایگانش تعیین و اصلاح می شود [۲]. هر اتوماتای یادگیر سلولی، از یک اتوماتای سلولی تشکیل شده است که هر سلول آن به یک یا چند اتوماتای یادگیر مجهز می باشد. همانند اتوماتای سلولی، قانون محلی در محیط حاکم بوده، و این قانون تعیین می کند که عمل انتخاب شده توسط یک اتوماتا در سلول باید پاداش داده شود و یا اینکه جریمه دریافت نماید.

در اتوماتای یادگیر سلولی می توان از ساختارهای مختلفی برای همسایگی استفاده نمود. معمول ترین آنها همسایگی اسمیت، کول، ون نیومن و مور می باشد. عملکرد اتوماتای یادگیر سلولی را می توان به شرح زیر بیان نمود. در هر لحظه هر اتوماتای یادگیر در اتوماتای یادگیر

سلولی یک عمل از مجموعه اعمال خود را انتخاب می‌کند. این عمل می‌تواند بر اساس مشاهدات قبلی و یا به صورت تصادفی انتخاب شود. عمل انتخاب شده با توجه به اعمال انتخاب شده توسط سلول‌های همسایه و قانون حاکم بر اتوماتای یادگیر سلولی پاداش و یا جریمه داده می‌شود. با توجه به اینکه عمل انتخاب شده پاداش گرفته و یا جریمه شده است، اتوماتا رفتار خود را تصحیح کرده و ساختار داخلی اتوماتا به‌نگام می‌گردد. بعد از بروزرسانی، هر اتوماتای یادگیر در اتوماتای یادگیر سلولی دوباره یک عمل از مجموعه اعمال خود را انتخاب می‌نماید. فرآیند انتخاب عمل و دادن پاداش و یا جریمه تا زمانی که سیستم به حالت پایدار برسد و یا یک معیار از قبل تعریف شده‌ای برقرار شود، ادامه می‌یابد.

۳. کارهای گذشته

در این بخش الگوریتم‌های LF، ASM و M-CLA که الگوریتم پیشنهادی مقایسه خواهند شد، به اجمال شرح داده می‌شوند.

روش خوشه بندی مبتنی بر کلونی مورچه‌های LF: گونه‌هایی از مورچه‌ها، لاشه‌های جمع‌آوری شده را با هدف مرتب‌سازی لانه خود خوشه بندی می‌کنند. اولین مدل با استفاده از این ویژگی مورچه‌ها توسط Deneubourg ارائه شد، که در آن جمعیتی از مورچه‌ها که بصورت تصادفی بر روی یک شبکه حرکت می‌کنند، اجازه برداشتن و گذاشتن آیت‌های داده را برای خوشه بندی داده‌ها دارند. Lumer and Faieta در [۶] این مدل را عمومی‌تر نموده و الگوریتم جدیدی با نام Ant Colony Clustering ارائه دادند که برای تحلیل داده و جستجو مورد استفاده قرار گرفت. این مدل با نام LF مشهور شد و در بررسی‌های بعدی مورد استفاده قرار گرفت. در هر دو روش ارائه شده، حرکت‌های داده از طریق حرکت مورچه‌ها پیاده‌سازی می‌شد و مورچه‌های بدون داده نیز در سطح شبکه حرکت می‌کردند، بنابراین نیاز به فضای حافظه اضافی و محاسبات بیشتر وجود داشت.

روش خوشه بندی مبتنی بر کلونی مورچه‌های ASM: Chen در [۷] مدل مورچه‌های خواب‌رونده^۳ را معرفی نمود. در این مقاله مقدار شایستگی مکان برای هر مورچه بصورت شباهت مورچه با مورچه‌های همسایگی تعریف می‌شود. وقتی شایستگی مکان یک مورچه کوچک باشد، احتمال فعال شدن و حرکت آن بیشتر می‌شود. این روش براساس Artificial Ant Sleeping Model (ASM) می‌باشد و با نام الگوریتم A^*C ^۴ ارائه شده است. در این روش هر آیت‌های داده به یک عامل (مورچه) تخصیص داده می‌شود. بنابراین هر مورچه حاوی اطلاعات یک آیت‌های داده و نیز برچسب کلاس نسبت داده شده به آن داده می‌باشد. در ابتدا مورچه‌ها به صورت تصادفی در سطح جدول دوبعدی اتوماتای سلولی قرار می‌گیرند و در طول الگوریتم براساس یک استراتژی تعیین مسیر به صورت خانه به خانه بر روی این شبکه حرکت می‌کنند. هر مورچه دارای دو وضعیت می‌باشد: فعال، و در حال استراحت. هر مورچه پس از قرار گرفتن در یک موقعیت جدید یک تابع تناسب را براساس شباهت داده مربوط به خود با داده‌های موجود در همسایگی و براساس آن میزان احتمال سکون یک عامل در مکان فعلی خود و یا حرکت عامل به همسایگی‌های مجاورش را محاسبه می‌نماید.

روش خوشه بندی مبتنی بر اتوماتای یادگیر سلولی: در روش M-CLA^۵ ارایه شده در [۱۱] برای خوشه‌بندی داده‌ها، مفهوم جدیدی به عنوان اتوماتای یادگیر سلولی متحرک ارائه شده است که در آن هر اتوماتای یادگیر می‌تواند بر اساس بردار احتمال جهات حرکتش در سطح اتوماتای سلولی به جهتی حرکت می‌نماید. در هر مرحله اگر عامل مورد نظر در همسایگی خود عاملی را مشاهده نماید، از جهت حرکت عامل همسایه برای بهبود جهت حرکت خود استفاده می‌نماید. یادگیری هر عامل بر اساس تشابه و یا عدم تشابه با عامل‌های همسایه به صورت زیر انجام می‌شود:

اگر عامل همسایه غیر مشابه باشد، جهت حرکت بیشینه عامل همسایه (جهت حرکت با احتمال بیشتر) در عامل مورد بررسی جریمه دریافت می‌نماید. اگر عامل همسایه مشابه باشد، جهت حرکت بیشینه عامل همسایه در عامل مورد بررسی پاداش دریافت می‌نماید.

عامل‌های همسایه هر عامل به صورت مجزا بر روی جهت حرکت عامل مورد بررسی تأثیر گذار می‌باشند. اگر تشابه دو عامل زیاد باشد، عامل مورد بررسی با دریافت پاداش با عامل همسایه همسو شده و در صورت تشابه کم، عامل مورد بررسی با دریافت جریمه با عامل همسایه غیر همسو می‌شود. بدین ترتیب در طول الگوریتم یادگیری، عامل‌های مشابه بتدریج با یکدیگر همسو شده و در کنار یکدیگر قرار می‌گیرند.

۴. روش ترکیبی پیشنهادی برای خوشه بندی داده‌ها

در این بخش ابتدا روش پیشنهادی را بصورت کلی مطرح نموده و سپس به شرح جزئیات الگوریتم خوشه بندی می‌پردازیم. هر داده‌ای نرمال شده به یک مورچه اختصاص می‌یابد. مورچه‌ها در سطح یک شبکه سلولی حرکت می‌نمایند و در مسیر خود مقداری ردپای فرمون^۵ بر جای می‌گذارند. در ابتدا تمامی مورچه‌ها بصورت تصادفی در سطح شبکه پراکنده هستند و تا زمانی که در همسایگی مورچه مشابهی قرار نگرفته‌اند^۱ براساس ردپای فرمون موجود در شبکه حرکت می‌نمایند. به محض یافتن اولین همسایه مشابه اتوماتای یادگیر مربوط به مورچه فعال

شده و براساس جهت حرکت مورچه های مشابه و غیر مشابه موجود در همسایگی، جهت حرکت بعدی مورچه را تعیین می نماید. بدین ترتیب احتمال انتخاب جهت های یکسان در مورچه های مشابه بیشتر شده و در نهایت تمامی مورچه های مشابه که نماینده داده های مربوط به یک خوشه هستند در کنار یکدیگر در سطح شبکه جمع می شوند.

جزئیات الگوریتم خوشه بندی به شرح زیر می باشد: در این روش ابتدا به هر کدام از داده های موجود در مجموعه داده ای که باید عمل خوشه بندی بر روی آن انجام گیرد به یک مورچه اختصاص می یابد. مورچه ها بصورت تصادفی بر روی یک شبکه مربعی سلولی قرار می گیرند. این شبکه مربعی دارای مرز دوری نمی باشد، به عبارت دیگر بالا و پایین، و چپ و راست شبکه به هم متصل نیست. همسایگی هر سلول بصورت همسایگی ون نیومن^۶ می باشد و جهت حرکت هر مورچه براساس این همسایگی انتخاب می گردد. به هر مورچه یک اتوماتای یادگیر نیز اختصاص می یابد که در ابتدا غیر فعال می باشد. این الگوریتم دارای تعدادی دور یا تکرار می باشد که هریک از مورچه ها در هر تکرار اقداماتی را که در ادامه می آید انجام می دهند. هر مورچه براساس تعداد همسایه های مشابه خود در مورد ایستادن و یا حرکت خود تصمیم می گیرد. به همسایه ای که فاصله آن با مورچه مورد نظر کمتر از یک مقدار معین باشد، همسایه مشابه گفته می شود. در صورت عدم مشاهده همسایه مشابه، مورچه براساس ردپای فرمون موجود در همسایگی خانه بعدی حرکت خود را انتخاب می نماید. ردپای فرمون موجود در هر سلول دارای اطلاعات مورچه گذارنده ردپا و زمان گذشتن مورچه از سلول مذکور می باشد. بدین ترتیب مورچه ردپای مورچه هایی را دنبال می کند که با مورچه مشابه باشند و زمان زیادی از عبور آنها از سلول نگذشته باشد. احتمال انتخاب هر سلول همسایگی براساس رابطه (۳) محاسبه می گردد.

$$P_i = \frac{x_i}{\sum_{Ph_i} (t_{current} - t_{Phi})^{1/3}} \quad (3)$$

اگر مورچه گذارنده فرمون Ph_i با مورچه مورد نظر مشابه باشد $x_i = 0$ و در غیر اینصورت $x_i = 1$ می باشد. هر مورچه با احتمال ۰.۹ خانه دارای بیشترین مقدار P_i و با احتمال ۰.۱ یک خانه تصادفی را برای حرکت بعدی انتخاب می نماید.

در صورتیکه تعداد مورچه های مشابه موجود در همسایگی بیش از صفر و کمتر از ۴ باشد، اتوماتای یادگیر مربوط به مورچه فعال شده و وظیفه انتخاب جهت حرکت را مشابه روش M-CLA بر عهده می گیرد. بدین ترتیب که هر مورچه بر اساس جهت های تعریف شده در همسایگی اش و همچنین احتمال حرکت جهات مختلف، یک جهت را برای حرکت انتخاب می نماید. پس از حرکت اگر مورچه مورد نظر در همسایگی خود مورچه ای را مشاهده نماید، از جهت حرکت مورچه همسایه برای بهبود جهت حرکت خود استفاده می نماید. یادگیری هر مورچه بر اساس تشابه و یا عدم تشابه با مورچه های همسایه به صورت زیر انجام می شود: اگر مورچه همسایه غیر مشابه باشد: جهت حرکت پیشینه مورچه همسایه (جهت حرکت با احتمال بیشتر) در مورچه عامل مورد بررسی جریحه دریافت می نماید. اگر مورچه همسایه مشابه باشد: جهت حرکت پیشینه مورچه همسایه در عامل مورد بررسی پاداش دریافت می نماید.

در این الگوریتم از یادگیری مدل P استفاده شده است (رابطه ۱ و ۲). در این روابط مقادیر نرخ پاداش و جریحه براساس تعداد مورچه های مشابه موجود در همسایگی محاسبه می شود. رابطه (۴) و (۵) نحوه محاسبه این مقادیر را نشان می دهد.

$$b = 0.1 + 0.1 * (N_{Like} / N_{Tot}) \quad (4) \quad a = 0.05 + 0.05 * (N_{Like} / N_{Tot}) \quad (5)$$

N_{Like} نشان دهنده تعداد مورچه های مشابه موجود در همسایگی و N_{Tot} نشان دهنده تعداد همسایه های مشابه است. تمامی ۸ همسایگی سلول بعنوان N_{Tot} در نظر گرفته می شود. با این کار در طول الگوریتم به تدریج با جمع شدن مورچه های مشابه در کنار یکدیگر نرخ یادگیری افزایش یافته و مورچه ها به سمت جهت پیشینه همگرا می شوند.

اگر مورچه بیش از ۴ همسایه مشابه و حداقل یک همسایه غیر مشابه داشت نیز اقدام به حرکت می نماید و در صورتی که هیچ یک از حالت های فوق اتفاق نیفتاد در جای خود باقی می ماند.

۵. ارزیابی مدل پیشنهادی

نتایج پیاده سازی مدل ارائه شده بر روی سه نوع مجموعه داده از مجموعه داده های نرمال شبیه سازی شده و مجموعه داده های واقعی مورد بررسی قرار گرفته است.

یک مجموعه داده شامل ۲۰۰ آیت دو بعدی متعلق به ۴ کلاس که از توزیع نرمال $N(\mu, \sigma^2)$ تبعیت می کنند ایجاد گردیده است.

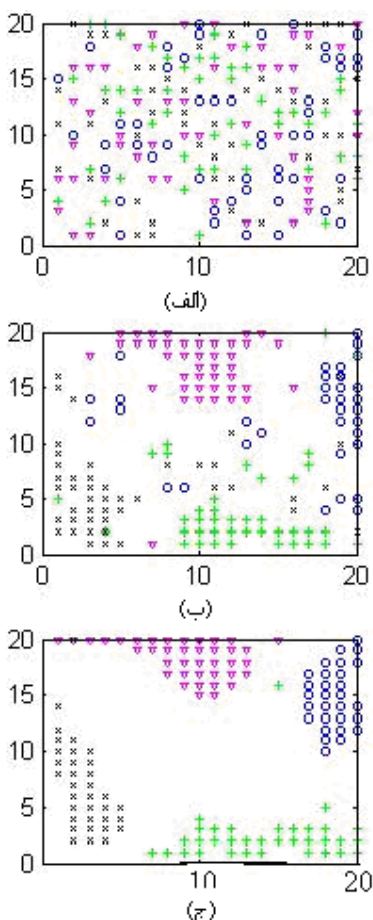
توزیع نرمال ۴ نوع داده به صورت $(N(0.2, 0.1^2), N(0.8, 0.1^2))$ ، $(N(0.2, 0.1^2), N(0.8, 0.1^2))$ ، $(N(0.2, 0.1^2), N(0.8, 0.1^2))$ و $(N(0.8, 0.1^2), N(0.8, 0.1^2))$ می باشد.

مجموعه داده های Iris مشتمل ۱۵۰ داده آموزشی با بعد ورودی چهار، شامل ۳ کلاس می باشد. تنها یک کلاس از این سه کلاس بصورت خطی از دو کلاس دیگر جدایی پذیر است.

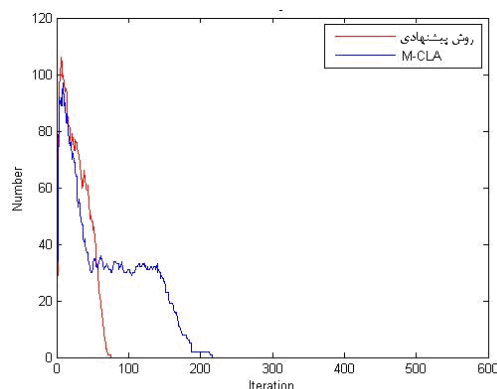
مجموعه داده های Wine شامل ۱۷۸ نمونه می باشد که متعلق به سه کلاس هستند. تعداد ویژگی های هر داده برابر ۱۳ ویژگی می باشد. ۵۹ داده متعلق به کلاس ۱، ۷۱ داده متعلق به کلاس ۲ و ۴۸ داده متعلق به کلاس ۳ می باشد.

در شکل (۲) توزیع اولیه داده ها و مراحل خوشه بندی مجموعه داده های اول نمایش داده شده است. داده ها در یک شبکه سلولی 20×20 قرار گرفته اند. مقدار آستانه شباهت داده ها برابر ۰.۴ در نظر گرفته شده است. همانگونه که مشاهده می شود، داده های مربوط به هر خوشه در طول الگوریتم به تدریج یک جهت حرکت را برای خود انتخاب نموده و پس از ۴۰۰ تکرار، داده ها در ۴ خوشه مجزا در کنار یکدیگر قرار می گیرند.

نمودار شکل (۳) خطای خوشه بندی حاصل از دو الگوریتم M-CLA و روش ترکیبی پیشنهادی را بر روی مجموعه داده های دسته اول نشان می دهد تعداد خطای خوشه بندی در روش پیشنهادی سریعتر کاهش یافته و به صفر رسیده است و به عبارت دیگر الگوریتم سریعتر همگرا شده است. علت این نتیجه استفاده از ردپای فرمون در یافتن مسیر حرکت مورچه های مشابه و همچنین تعیین پویای مقادیر نرخ جریمه و پاداش می باشد.



شکل ۲ - مراحل خوشه بندی مجموعه داده های نرمال دسته اول (الف) توزیع اولیه داده ها (ب) پراکندگی پس از ۲۰۰ تکرار (ج) پراکندگی پس از ۴۰۰ تکرار



شکل ۳- خطای خوشه بندی مجموعه داده های نرمال دسته اول توسط الگوریتم های پیشنهادی و M-CLA

نتایج خوشه بندی سه مجموعه داده های نام برده شده در جدول (۱) مقایسه شده اند. این جدول نتایج بدست آمده را برای ۲۰ آزمایش با مکان های ابتدایی تصادفی برای مجموعه داده های فوق نمایش می دهد.

همانگونه که مشاهده می شود روش پیشنهادی در خوشه بندی مجموعه داده های اول و داده های Wine از نظر کیفیت خوشه بندی و تعداد تکرار بهتر عمل می نماید. در داده های Iris میانگین خطاها کمتر از مدل LF و بیشتر از مدل ASM و M-CLA می باشد. اما بیشترین تعداد خطاها در این روش از هر سه روش بهتر است.

جدول ۱ - مقایسه نتایج خوشه بندی

مجموعه داده	الگوریتم	بیشترین تعداد تکرار	خطای خوشه بندی		
			میانگین	بیشترین	کمترین
مجموعه داده های اول	LF	۱۰۰۰۰۰	۱	۳	۰
	ASM	۵۰۰	۱	۳	۰
	M-CLA	۹۰	۰	۰	۰
	روش پیشنهادی	۴۰	۰	۰	۰
Iris	LF	۱۰۰۰۰۰	۶۶۸	۱۳	۳
	ASM	۵۰	۴۳۹	۸	۲
	M-CLA	۱۵۰	۵۶	۸	۲
	روش پیشنهادی	۹۷۰	۶۰۱	۷	۳
Wine	LF	۱۰۰۰۰۰	۱۵	۱۸	۱۲
	ASM	۵۰۰	۱۳۶	۱۵	۱۲
	M-CLA	۱۵۰	۱۴	۱۸	۱۰
	روش پیشنهادی	۱۲۰	۱۰۳	۱۶	۸

۶. نتیجه

در این مقاله یک روش ترکیبی خوشه بندی بدون ناظر داده ها برپایه سیستم کلونی مورچه ها و اتوماتای یادگیر سلولی ارائه شد. در روش پیشنهادی مورچه های پراکنده در سطح شبکه با استفاده از ردپای فرمون، مورچه های مشابه را می یابند و اتوماتاهای یادگیر، حرکت آنها را با مورچه های مشابه موجود در همسایگی هم جهت می نمایند. همچنین بعلا تعیین نرخ یادگیری اتوماتای یادگیر سلولی بصورت پویا، الگوریتم در مراحل پایانی یادگیری که تعداد مورچه های مشابه همسایه بتدریج بیشتر می شود، با سرعت بیشتری به سمت پاسخ بهینه همگرا می گردد. نتایج حاصله از مقایسه الگوریتم پیشنهادی با سایر الگوریتم های خوشه بندی حاکی از سرعت و کارایی بیشتر آن نسبت به الگوریتم های یاد شده می باشد.

- [۱] S. Wolfram, “Cellular Automata”, Los Alamos Science, Vol. ۹, pp. ۲-۲۱, Fall ۱۹۸۳.
- [۲] H. Beigy and M. R. Meybodi, “A Mathematical Framework for Cellular Learning Automata”, Advances in Complex Systems, Vol. ۷, Nos. ۳-۴, pp. ۲۹۵-۳۲۰, September/December ۲۰۰۴.
- [۳] H. Beigy and M. R. Meybodi, “Open Synchronous Cellular Learning Automata”, ۸th World Multi-conference on Systemics- Cybernetics and Informatics(SCI۲۰۰۴) , USA, pp. ۹-۱۵, July ۲۰۰۴.
- [۴] M. A. L. Thathachar and P. S. Sastry, “Varieties of Learning Automata: An Overview”, IEEE Transaction on Systems, and Cybernetics-Part B: Cybernetics Vol. ۳۲, No. ۶, pp. ۷۱۱-۷۲۲, ۲۰۰۲.
- [۵] M. Dorigo and L. M. Gambardella, “Ant Colony System: A Cooperative Learning Approach to The Traveling Salesman Problem”, IEEE Transactions on Evolutionary Computation, Vol. ۱, pp. ۵۳-۶۶, ۱۹۹۶.
- [۶] E. D. Lumer and B. Faieta, “Diversity and Adaptation in Populations of Clustering Ants”, International Conference on Simulation of Adaptive Behaviour: From Animals to Animats, vol.۳, MIT Press/Badford Books, pp. ۵۰۱-۵۰۸, ۱۹۹۴.
- [۷] X. Zhang, H. Peng and Q. Zheng, “A Novel Ant Colony Optimization Algorithm for Clustering”, ۸th International Conference on Signal Processing(ICSP), vol. ۳, IEEE Computer Society Press, ۲۰۰۶.
- [۸] A. Ghosh and A. Halder, “Aggregation Pheromone Density Based Data Clustering”, Elsevier, Information Science, No ۱۷۸, pp. ۲۸۱۶-۲۸۳۱, ۲۰۰۸.
- [۹] J. Handl and B. Meyer, “Ant-Based and Swarm-Based Clustering”, Swarm Intelligence, No. ۱, pp. ۹۵-۱۱۳, November, ۲۰۰۷.
- [۱۰] A. V. Moere and J. J. Clayden, “Cellular Ants: Combining Ant-based Clustering with Cellular Automata”, ۱۷th IEEE international conference on tools with artificial intelligence, ICTAI ۲۰۰۵, Piscataway: IEEE Press, pp. ۱۷۷-۱۸۴, ۲۰۰۵.
- [۱۱] M.Hosseini Sedehi and M. R. Meybodi, “A Data Clustering Algorithm Based on Cellular Learning Automata“, Iranian Data Mining Confrence(IDMC۰۷), AmirKabir University, Tehran, Iran,Nov ۲۰۰۷.
- [۱۲] A.K. Jain, M. N. Murty and P. J. Flynn, “Data Clustering: a Review”, ACM Computing Surveys Vol. ۳۱, No. ۳, pp. ۲۶۴-۳۲۳. ۱۹۹۹.

- ^۱ Clustering
- ^۲ Learning Automata
- ^۳ Ant Sleeping Model
- ^۴ Adaptive Artificial Ant Clustering Algorithm
- ^۵ Pheromone Trail
- ^۶ Von Neumann